Randomness in the Limit

Henry Towsner

July 13, 2022

DRAFT

Contents

Preface

The core topic of the book is this cluster of methods, lying at the intersection of extremal combinatorics, probability theory, and mathematical logic, which allow us to approach finite counting problems by using the tools of probability in a genuinely infinite setting.

A motivating example is Szemerédi's Theorem:

Theorem (Szemerédi's Theorem)**.** *For every $\epsilon > 0$ and every k, there is an N so that whenever $n \geq N$ and $A \subseteq \{1, 2, \ldots, n\}$ is a set with $\frac{|A|}{n} \geq \epsilon$, there is an $a \in A$ and a $d > 0$ such that*

$$a, a + d, a + 2d, \ldots, a + (k - 1)d \in A.$$

The set $\{a, a + d, a + 2d, \ldots, a + (k - 1)d\}$ is an *arithmetic progression of length k*, so this theorem says that whenever A is a "dense set"—a set which contains at least ϵ of the points in the interval $\{1, 2, \ldots, n\}$—A must contain an arithmetic progression of length k.

This looks like a statement about counting finite sets of numbers, but most of the work in this book will be about developing probability theory for certain kinds of infinite graphs and hypergraphs*.

We will present a proof of this theorem, eventually: the case when $k = 3$ will be proven in Chapter 6, and the full result will wait until Chapter 8. Instead of working directly towards this proof, the book develops the machinery needed in a general way and takes many digressions to discuss related ideas.

There has been a great deal of work in this area in the last twenty years, and only a small fragment is covered here; the remarks at the end of each chapter give a very non-comprehensive list of further developments and applications in the literature.

*This is far from the only way to prove Szemerédi's Theorem, which has a rather large number of proofs—e.g. [13, 65, 66, 71, 75, 76, 80, 87, 121, 134, 145, 149, 150, 165], though these proofs are far from completely distinct.

The proximate motivation for this book was a conversation with a co-author. I told him that the answer to a technical question he'd asked was "obvious" and outlined an argument. When he asked where this was written down, I admitted that the proof was scattered across three or four different papers, with perhaps one or two others needed to translate between the different formalisms being used so the ideas from different papers could be combined.

Nonetheless, I insisted, the result was "obvious"—if you just happened to already be thinking about the problem in precisely the right way.

My guiding light while writing this book has been to try to make obvious the things that I think should be obvious: to attempt to illustrate the way my preferred formalism falls naturally out of the questions we are investigating, and to illustrate why this way of thinking about limits of sequences of large graphs gives us the insight needed to understand them.

Books in mathematics have to choose a position somewhere between serving as an introduction to a topic and serving as a reference. There is some tension between these roles—should I state the most general form of a theorem, or the more concrete version needed? Should I give the most elegant proof, or the most convenient one at the moment?

I wrote this book thinking of it as an introduction. I have tolerated some repetition of proofs, proving the special case when it is needed and waiting to state the general result later, even when it means repeating the proof. I (mostly) resisted the temptation to comprehensively list corollaries and applications from the literature; instead, many chapters end with stray tidbits I like and references to the literature where those directions are further explored.

Acknowledgements: This book benefitted immeasurably from many conversations with colleagues too numerous to mention.

I am grateful to my Math 571 students—Koby Frank, Stephen Gillen, Perry Hart, Mary-Frances Jagod, Alec Korman, Martin Rubin, Anschel Schaffer-Cohen, Owain West, and Yuqing Zhao—for many helpful suggestions and comments made while suffering through a course based on an early draft.

This book would not have been possible without the patient support of the NSF, through grants DMS-1600263 and DMS-2054379, as well the American Institute of Mathematics and the Institut Henri Poincaré.

Finally, I am grateful to my family for their support. My husband, Jon, helped remind me so many times that this was worth finishing even when the amount left to do seemed unsurmountable. This book is dedicated to my son, Solomon, who was willing (occasionally) to sleep quietly in my lap while wrote a section or two.

Chapter 1

Random and Quasi-random Graphs

We will begin with the notion of a randomly generated finite graph. We are interested in how, faced with a particular graph, we can decide whether or not it is a "typical" example of a randomly generated graph. This will lead us to distill the notion of a *quasi-random graph* and investigate the ways in which any quasi-random graph has many of the properties we would expect, with high probability, of a graph we generated randomly.

1.1 Random Finite Graphs

We start by investigating what it means for a graph to be random.

Definition 1.1. When V is a set, we write $\binom{V}{k}$ for the set of subsets of V of size exactly k.

When V is a non-empty set, a *graph on* V is a set $E \subseteq \binom{V}{2}$. The elements of E are the *edges* of the graph and the elements of V are the *vertices*.

This definition excludes the possibility of "loops" (edges between a vertex and itself): by definition, an edge is a pair of distinct vertices. Similarly, this definition excludes directed graphs—E is a set of unordered pairs, so there is no difference between saying $\{v, w\} \in E$ and $\{w, v\} \in E$. We also prohibit a graph with no vertices at all (though we allow graphs with no edges).

We frequently write "$G = (V, E)$ is a graph" to mean that V is the set of vertices of a graph and $E \subseteq \binom{V}{2}$ is the set of edges, but we will also sometimes refer to E by itself as a graph with the set of vertices implied.

When V is a finite set—usually a large finite set—we want to consider a random graph on V. Informally speaking, this is the graph we obtain by flipping a fair coin for each pair $\{v, w\} \in \binom{V}{2}$ and placing an edge between v and w if the coin comes up heads.

Since we will deal with this graph repeatedly, we give it a name: we will call this random graph $\mathbf{R}_{1/2}$, or $\mathbf{R}_{1/2}(V)$ if we wish to be explicit about the set of vertices V. We will abuse notation to write $\mathbf{R}_{1/2}$ for both the graph and the set of edges. Note that we follow the convention that random variables are written in bold.

One of the basic questions we'll be concerned with is which graphs "look like" random graphs. Of course, a random graph *could* look like anything—a random graph on 100 vertices could end up having no edges at all if every coin comes up tails. With a probability 2^{-4950}, this isn't *likely*, however. So we want to ask which properties a random graph will *probably* have.

A good place to start is observing that it's unlikely for a random graph to have no edges at all. More specifically, there are $\binom{|V|}{2}$ pairs which might be edges, and a random graph ought to have just about half of them. We make this concrete by showing that, when $|V|$ is large, $\mathbf{R}_{1/2}$ is very likely to have close to half the edges.

Theorem 1.2. *For every $\epsilon > 0$ and every $\delta > 0$, whenever V is sufficiently large,*

$$\mathbb{P}\left(\left||\mathbf{R}_{1/2}| - \frac{1}{2}\binom{|V|}{2}\right| < \epsilon\binom{|V|}{2}\right) \geq (1 - \delta).$$

It will be more natural for us to think in terms of "densities" rather than quantities: rather than looking at the size of $\mathbf{R}_{1/2}$, we will look at $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$, which represents the fraction of "possible edges" which are present in $\mathbf{R}_{1/2}$. So we will show the equivalent statement

$$\mathbb{P}\left(\left|\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}} - \frac{1}{2}\right| < \epsilon\right) \geq 1 - \delta.$$

Proof. The idea is that each edge is placed in $\mathbf{R}_{1/2}$ independently, so $|\mathbf{R}_{1/2}|$ is a sum of independent random variables: for each $\{v, w\}$, let $\mathbf{1}_{\{v,w\}}$ be the random variable which is 1 if $\{v, w\} \in \mathbf{R}_{1/2}$ and 0 if $\{v, w\} \notin \mathbf{R}_{1/2}$. Then $|\mathbf{R}_{1/2}| = \sum_{\{v,w\} \in \binom{V}{2}} \mathbf{1}_{\{v,w\}}$. Since each of these random variables $\mathbf{1}_{\{v,w\}}$ is chosen independently, it is very likely that close to half of them are 1 while the other half are 0.

Slightly more formally, what $\mathbf{1}_{\{v,w\}}$ means is that we first pick a particular pair of vertices, $\{v, w\}$, and then generate the set of edges $\mathbf{R}_{1/2}$ by flipping

coins, and $\mathbf{1}_{\{v,w\}}$ is 1 in the event that the potential edge we picked in advance actually turns up in our final graph. In particular, $\mathbf{1}_{\{v,w\}}$ is simply the result of a single coin flip, so its expected value, $\mathbb{E}(\mathbf{1}_{\{v,w\}})$, is $1/2$, since half the time the coin comes up heads and $\mathbf{1}_{\{v,w\}} = 1$, and the other half of the time the coin comes up tails and $\mathbf{1}_{\{v,w\}} = 0$.

The expected value of $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$ is also $1/2$: by the the linearity of expected value

$$\begin{aligned}\mathbb{E}(\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}) &= \mathbb{E}\left(\frac{\sum_{\{v,w\}\in\binom{V}{2}} \mathbf{1}_{\{v,w\}}}{\binom{|V|}{2}}\right) \\ &= \frac{1}{\binom{|V|}{2}} \sum_{\{v,w\}\in\binom{V}{2}} \mathbb{E}(\mathbf{1}_{\{v,w\}}) \\ &= \frac{1}{\binom{|V|}{2}} \sum_{\{v,w\}\in\binom{V}{2}} \frac{1}{2} \\ &= \frac{1}{2}.\end{aligned}$$

That is, before we actually flip the coins, we expect the average value of $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$ to be $1/2$. But we must consider the possibility that we reach this average because the graphs that have almost all the edges are cancelled out by graphs with very few edges. What remains is to show that the distribution is narrow: that most of these graphs have close to the right number of edges.

This should happen because $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$ is the sum of a large number of independent random variables. The *Hoeffding inequality* covers precisely this situation.

Theorem (Hoeffding Inequality)**.** *If* $\mathbf{X} = \frac{1}{k}\sum_{i\leq k} \mathbf{X}_i$ *where the* $\mathbf{X}_i$ *are independent random variables such that* $0 \leq \mathbf{X}_i \leq 1$ *always holds then*

$$\mathbb{P}(|\mathbf{X} - \mathbb{E}(\mathbf{X})| \geq \epsilon) \leq 2e^{-2k\epsilon^2}.$$

We take $\mathbf{X}$ to be the random variable $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$, so the Hoeffding inequality says that

$$\mathbb{P}\left(\left|\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}} - \frac{1}{2}\right| \geq \epsilon\right) \leq 2e^{-2\binom{|V|}{2}\epsilon^2}.$$

So by choosing $|V|$ sufficiently large (on the order of $-\sqrt{\ln \delta}/\epsilon$), we can make the bound on the right smaller than δ. □

Of course, this property—having roughly half the possible edges—is not unique to random graphs: it is not difficult to produce examples which have the same number of edges as a random graph, but are quite clearly non-random.

Example 1.3. The *complete bipartite graph,* $K_{n,n}$, is the graph $(V \cup W, E)$ where $|V| = |W| = n$, V and W are disjoint sets, and E consists of all pairs of vertices between V and W. Then $|E| = n^2$ while $\frac{1}{2}\binom{|V\cup W|}{2} = \frac{1}{2}\frac{2n(2n-1)}{2} = n^2 - \frac{n}{2}$.

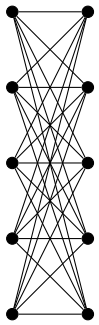


Figure 1.1: $K_{5,5}$

In this example, $|E|$ is not quite identical to half the possible edges, but the error—$n/2$—is small relative $\binom{|V\cup W|}{2}$. Specifically, for any $\epsilon > 0$, when n is sufficiently large we have $\frac{n/2}{\binom{|V\cup W|}{2}} < \epsilon$, so the error is within the margins given by the preceding theorem. (We could eliminate this error entirely by removing $n/2$ edges—say, choosing a single vertex in V and removing half its edges.)

It seems clear that generating a graph randomly would be very unlikely to produce a complete bipartite graph. But we would like to prove this by identifying some property that random graphs are likely to have but which $K_{n,n}$ does not.

One observation is that while $K_{n,n}$ has the "correct" number of edges, it has no triangles—there are no triples $\{v_0, v_1, v_2\} \in V \cup W$ with all three edges (v_0, v_1), (v_0, v_2), and (v_1, v_2) present in $K_{n,n}$.

This behavior seems non-random, and we will now set out to prove that it is indeed extremely improbable in a random graph.

1.2 Subgraph Density

First, we must identify how many triangles a random graph *should* have. It will not be much more complicated to ask a more general question: if H is *any* finite graph, we can ask how many different ways H appears as a subgraph. More precisely, in keeping with our preference for a probabilistic perspective, we will ask what fraction of the possible copies of H are actually present.

Definition 1.4. When $H = (W, F)$ and $G = (V, E)$ are graphs, a *copy of H in G* is a function $\pi : W \to V$ such that, for each edge $\{w, w'\} \in F$, $\{\pi(w), \pi(w')\} \in E$.

A *potential copy of H in V* is a function $\pi : W \to V$.

We define $t_H(G)$ to be the fraction of potential copies of H which are actual copies:

$$t_H(G) = \frac{|\{\pi : W \to V \mid \pi \text{ is a copy of } H \text{ in } G\}|}{|V|^{|W|}}.$$

We call $t_H(G)$ the *subgraph density of H in G*.

A potential copy doesn't really depend on the edges: it's just a function mapping the vertices of W to the vertices of V. A potential copy is an actual copy if every edge of H is mapped to an edge of G.

We think of G being a graph on n vertices where n is large, and H as a small graph like a triangle. There are two subtleties to note in the definition of $t_H(G)$. To see the first, consider the case where H is a triangle—the graph we call C_3 (a cycle of length 3).

Definition 1.5. C_3 is the graph $(\{0, 1, 2\}, \binom{\{0,1,2\}}{2})$—that is, the triangle with three vertices and all three edges: 

Then a potential copy of C_3 is an *ordered* choice of 3 vertices *allowing repetition**. That means that any time we have three distinct vertices in V, we count that as 6 potential triangles, one for each order of the three vertices.

This doesn't make much difference in the calculation of $t_{\mathrm{C}_3}(G)$: since $t_{\mathrm{C}_3}(G)$ is a fraction, the factor of 6 appears in both the numerator and the denominator and therefore cancels out. This also means that the denominator includes some cases where we choose three vertices but at least two are the same; we will sometimes call these "degenerate" triangles. However there aren't very many of these—there are only $O(n^2)$ cases where we have repeated vertices—so when n is large (say, much larger than $1/\epsilon$), these degenerate triangles will get absorbed into the error terms of our calculations.

The second subtlety is that while we require that edges in W map to edges in V, we do not require that *non-edges* get mapped to non-edges: we

*This is sometimes called counting "labeled triangles". This come from the view that we are counting, not just triangles, but specifically three vertices labeled "0", "1", or "2", and we consider it a different triangle if we choose the same three vertices but with different labels.

still consider π to represent a copy of H even if $\pi(W)$ contains extra edges. For instance, suppose G is the graph on 4 vertices arranged like  . If H is the cycle on four vertices,  , then by our definition, $t_H(G)$ is positive: we count the copies where π maps the four vertices of H to the four vertices of G, and the extra edge in G is no obstacle.

This is consistent with the usual definition of a subgraph in graph theory. The stricter notion, where $\pi(W)$ should have exactly the same edges as W, is called an *induced subgraph*, and there is a corresponding variant of $t_H(G)$.

Definition 1.6. When $H = (W, F)$ and $G = (V, E)$ are graphs, an *induced copy of H in V* is a possible copy $\pi : W \to V$ such that, for each pair $\{w, w'\} \in \binom{W}{2}$,

$$\{w, w'\} \in F \text{ if and only if } \{\pi(w), \pi(w')\} \in E.$$

We define $t_H^{\text{ind}}(G)$ to be the fraction of potential copies of H which are induced copies:

$$t_H^{\text{ind}}(G) = \frac{|\{\pi : W \to V \mid \pi \text{ is an induced copy of } H \text{ in } G\}|}{|V|^{|W|}}.$$

We call $t_H(G)$ the *induced subgraph density of H in G*.

For the triangle these definitions are the same, but as soon as $F \subsetneq \binom{W}{2}$, we can have $t_H^{\text{ind}}(G) < t_H(G)$.

We can now set out to show that, with high probability, a random graph has the "right" number of copies of each small graph. First we need to figure out what the right number of copies is—that is, what the expected value of $t_H(\mathbf{R}_{1/2})$ is.

Suppose we set out to generate a random graph $\mathbf{R}_{1/2}(V)$ on the set of vertices V. If $H = (W, F)$ and we pick in advance a potential copy $\pi : W \to V$, we can take $\mathbf{1}_\pi$ to be the random variable which is 1 if π turns out to be an actual copy of H—that is, if, for each $\{w, w'\} \in F$, $\{\pi(w), \pi(w')\} \in \mathbf{R}_{1/2}(V)$. $\mathbf{1}_\pi$ is 0 if π is not an actual copy of H in $\mathbf{R}_{1/2}(V)$. As long as π is injective, for each pair $\{w, w'\} \in F$, there is a $1/2$ chance that $\{\pi(w), \pi(w')\}$ ends up being put into $\mathbf{R}_{1/2}(V)$ and each edge is determined independently, so $\mathbb{E}(\mathbf{1}_\pi) = 2^{-|F|}$—we flip $|F|$ coins, one for each edge of H, and $\mathbf{1}_\pi$ is 1 if all these coins come up heads.

When π is not injective, the issue is messier, but we include this case in the error term: there are fewer than $|W| \cdot |V|^{|W|-1}$ non-injective functions $\pi : W \to V$. When we divide by $|V|^{|W|}$, these terms will contribute at most $\frac{|W|}{|V|}$ to $t_H(G)$—that is, an error term on the order $O(\frac{1}{|V|})$.

Since $t_H(\mathbf{R}_{1/2}(V)) = \frac{1}{|V|^{|W|}} \sum_\pi \mathbf{1}_\pi$, the linearity of expectation says that

$$\mathbb{E}(t_H(\mathbf{R}_{1/2}(V))) = \frac{1}{|V|^{|W|}} \sum_\pi \mathbb{E}(\mathbf{1}_\pi) = 2^{-|F|} + O(\frac{1}{|V|}).$$

In particular,

$$\lim_{|V|\to\infty} \mathbb{E}(t_H(\mathbf{R}_{1/2}(V))) = 2^{-|F|}.$$

(In the next two chapters we will actually pass to the limit, allowing us to dispense with error terms entirely.)

Of course, we should not be surprised that we have to worry a little about the size of V: after all, if W were larger than V, we wouldn't expect there to be any copies of H in $\mathbf{R}_{1/2}(V)$. We only expect $\mathbf{R}_{1/2}(V)$ to have the right number of copies of H when V is much larger than W.

Once again, we must rule out the possibility that this average is the result of having some cases where the random graph contains too many copies of H being canceled out by cases where there are too few.

Theorem 1.7. *For every $\epsilon > 0$, every $\delta > 0$, and every finite graph $H = (W, F)$, whenever V is sufficiently large,*

$$\mathbb{P}\left(\left|t_H(\mathbf{R}_{1/2}) - 2^{-|F|}\right| < \epsilon\right) \geq (1 - \delta).$$

Proof. The idea is similar to the proof of Theorem 1.2 above: we want to argue that $t_H(\mathbf{R}_{1/2})$ is the sum of a large number of separate events, and therefore it is likely that the sum comes close to the average. Unfortunately, the various random variables $\mathbf{1}_\pi$ are no longer independent: if $\pi(H)$ and $\pi'(H)$ share an edge, $\mathbf{1}_\pi$ and $\mathbf{1}_{\pi'}$ are correlated.

However the edges are still independent, and each edge only appears in a small fraction of the potential copies of H. This means that, although $t_H(\mathbf{R}_{1/2})$ is no longer a sum of many independent random variables, $t_H(\mathbf{R}_{1/2})$ is a *function* of many independent random variables where each individual random variable (that is, each edge) only has a small impact on the value of the function. This is precisely the situation to which McDiarmid's inequality applies.

Theorem (McDiarmid's inequality)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_k$ be independent random variables and let $f(x_1, \ldots, x_k)$ be a function with the following property: for each $i \leq k$ there is a $c_i \geq 0$ such that, for any values $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$ and any x_i, x_i',*

$$\left|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_k)\right| < c_i.$$

Then

$$\mathbb{P}\left(|\mathbb{E}(f(\mathbf{X}_1, \ldots, \mathbf{X}_k)) - f(\mathbf{X}_1, \ldots, \mathbf{X}_k)| \geq \epsilon\right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^k c_i^2}}.$$

The Hoeffding and McDiarmid inequalities belong to a general family of results in probability theory called *concentration inequalities* which describe the way that random variables "concentrate" near their expected value under suitable assumptions.

The condition

$$\left|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_k)\right| < c_i$$

says that no individual random variable has a disproportionate impact on the function: we can change any individual value x_i and not change the value of the function by much.

In this case, our independent random variables are the random variables $\mathbf{1}_{\{v,w\}}$ for the individual edges, and the function $f(\{\mathbf{1}_{\{v,w\}}\}_{\{v,w\} \in \binom{V}{2}})$ is $t_H(\mathbf{R}_{1/2})$, the density of copies of H in the randomly generated graph. The difference

$$\left|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_k)\right| < c_i$$

in the statement of McDiarmid's inequality is then asking how much $t_H(\mathbf{R}_{1/2})$ can change if a single edge flips from present to absent or vice versa. The edge $\{v, w\}$ only matters to those potential copies of H which contain both vertices.

Starting with an edge $\{v, w\} \in \binom{V}{2}$, how many potential copies $\pi : W \to V$ of H contain it? There are $|W| \cdot (|W| - 1)$ ways to pick one vertex from W to map to v and then one to map to w, and then $|V|^{|W|-2}$ ways to assign the remaining $|W| - 2$ vertices, so there are $\leq |W|^2 |V|^{|W|-2}$ potential copies containing this edge. (In fact, this is slightly over-counting, because the

copies where multiple vertices from W get mapped to either v or w get counted more than once.) So, at worst, flipping the edge $\{v, w\}$ could mean losing or gaining $|W|^2|V|^{|W|-2}$ copies of H. This means that the density t_H changes by at most $\frac{|W|^2|V|^{|W|-2}}{|V|^{|W|}} = \frac{|W|^2}{|V|^2}$.

So the values c_i we use for McDiarmid's inequality are each bounded by $\frac{|W|^2}{|V|^2}$, so $\sum_{i=1}^{k} c_i^2 = \binom{|V|}{2}\frac{|W|^4}{|V|^4} \leq \frac{|W|^4}{|V|^2}$.

If V is big enough that $\left|\mathbb{E}(t_H(\mathbf{R}_{1/2})) - 2^{-|F|}\right| < \epsilon/2$ then McDiarmid's inequality says that

$$\begin{aligned}\mathbb{P}\left(\left|2^{-|F|} - t_H(\mathbf{R}_{1/2})\right| \geq \epsilon\right) &\leq \mathbb{P}\left(\left|\mathbb{E}(t_H(\mathbf{R}_{1/2})) - t_H(\mathbf{R}_{1/2})\right| \geq \epsilon/2\right) \\ &\leq 2e^{-\frac{\epsilon^2}{|W|^4}|V|^2}.\end{aligned}$$

Once we pick an ϵ, the value $\frac{\epsilon^2}{2|W|^4}$ is fixed, so by choosing V large enough, we can make this bound as small as we like, and in particular smaller than δ. □

The same ideas apply to $t_H^{\text{ind}}(\mathbf{R}_{1/2})$; in this case when we fix an individual potential copy $\pi : W \to V$, if π is injective then the probability that π becomes an induced copy is $2^{-\binom{|W|}{2}}$, irrespective of how many edges H has: in order for π to be an induced copy, each pair $\{\pi(w), \pi(w')\}$ has to do exactly the right thing—become an edge if $\{w, w'\} \in F$ or a non-edge if $\{w, w'\} \notin F$—and each pair has a $1/2$ chance of doing that. Other than that, the arguments go through unchanged:

Theorem 1.8. *For every $\epsilon > 0$, every $\delta > 0$, and every finite graph H, whenever V is sufficiently large,*

$$\mathbb{P}\left(\left|t_H^{\text{ind}}(\mathbf{R}_{1/2}) - 2^{-\binom{|W|}{2}}\right| < \epsilon\right) \geq (1 - \delta).$$

In particular, we can now conclude that it is very unlikely for a randomly generated graph to look like the bipartite graph $\mathrm{K}_{n,n}$: in a random graph, about $\frac{1}{8}$ of the potential triangles will almost certainly be actual triangles, while in $\mathrm{K}_{n,n}$ none of them will be.

However it turns out that there are other, more complicated graphs, which do a better job of imitating a random graph: they have $1/2$ the edges and also have $\frac{1}{8}$ of the potential triangles. Before building these, however, it will be useful to develop some tools for calculating $t_H(G)$.

1.3 $t_H(G)$ as an Integral

It will be helpful to introduce a measure-theoretic notation for counting things like subgraph densities.

Definition 1.9. Let V be a finite set of vertices. For each k, we write μ_k for the *counting measure* on V^k given by

$$\mu_k(S) = \frac{|S|}{|V|^k}$$

for every $S \subseteq V^k$.

For example, the set of triples $(v_0, v_1, v_2) \in V^3$ which are triangles in E is a set, and $t_{\mathrm{C}_3}(G)$ is precisely the measure of this set under μ_3.

Note that we are counting ordered pairs here—in particular, strictly speaking E is not a subset of V^2 (the set of ordered pairs) because E is a subset of $\binom{V}{2}$ (the set of unordered pairs). However there is a closely related set of ordered pairs—$\{(v,w) \mid \{v,w\} \in E\}$—and $|\{(v,w) \mid \{v,w\} \in E\}| = 2|E|$, so $\mu_2(\{(v,w) \mid \{v,w\} \in E\})$ is close to $\frac{|E|}{|\binom{V}{2}|}$ (but slightly smaller, because of the presence of repetitive ordered pairs like (v,v) in the denominator $|V^2|$).

It will be convenient to identify $t_H(G)$ with an integral.

Definition 1.10. When E is a graph on V, $\chi_E : V^2 \to \{0,1\}$, the *characteristic function of* E, is the function given by:

$$\chi_E(v,w) = \begin{cases} 1 & \text{if } (v,w) \in E \\ 0 & \text{otherwise} \end{cases}$$

We will sometimes abuse notation and write $t_H(E)$ or $t_H^{\mathrm{ind}}(E)$ if the vertex set is clear.

This lets us write, for instance,

$$t_{\mathrm{C}_3}(G) = \iiint \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\, d\mu_3.$$

Definition 1.11. K_2 is the graph with two elements and an edge between them.

That is, K_2 is the graph consisting of a single edge. Then

$$t_{\mathrm{K}_2}(G) = \mu_2(\{(v,w) \mid \{v,w\} \in E\}) = \frac{\{(v,w) \mid \{v,w\} \in E\}}{V^2} = \frac{2|E|}{|V|^2},$$

which differs from $\frac{|E|}{\binom{|V|}{2}}$ by an amount on the order of $1/|V|$, which we can treat as negligible when V is large enough. Going forward, we will often focus on the quantity $t_{\mathrm{K}_2}(G)$ rather than $\frac{|E|}{\binom{|V|}{2}}$, and call this the *edge density* of G.

More generally, we have

Theorem 1.12. *For any graph $H = (W, F)$ with $W = \{w_1, \ldots, w_k\}$ and any graph $G = (V, E)$,*

$$t_H(G) = \int \prod_{1 \le i < j \le k, \{w_i, w_j\} \in F} \chi_E(v_i, v_j) d\mu_k.$$

Note that, since these are finite spaces, integrals are really averages:

$$\begin{aligned} &\int \prod_{1 \le i < j \le k, \{w_i, w_j\} \in F} f(v_i, v_j) d\mu_{|W|} \\ =&\frac{1}{|V|^{|W|}} \sum_{\{v_1, \ldots, v_{|W|}\} \in V^{|W|}} \prod_{1 \le i < j \le k | \{w_i, w_j\} \in F} f(v_i, v_j). \end{aligned}$$

Motivated by this, we can define t_H for functions, not just graphs:

Definition 1.13. We say $f : V^2 \to \mathbb{R}$ is *symmetric* if, for all $(v, w) \in V^2$, $f(v, w) = f(w, v)$.

When $f : V^2 \to \mathbb{R}$ is symmetric, we define

$$t_H(f) = \int \prod_{1 \le i < j \le k, \{w_i, w_j\} \in F} f(v_i, v_j) d\mu_{|W|}.$$

This precisely generalizes our definition for graphs: when $G = (V, E)$, $t_H(G) = t_H(\chi_E)$.

The reason we demand symmetry is:

Lemma 1.14. *When f is symmetric, $t_H(f)$ does not depend on the ordering of vertices $W = \{w_1, \ldots, w_k\}$.*

If f were not symmetric, the behavior of $t_H(f)$ could be strange, since the definition only includes $f(v_i, v_j)$ in the product when $i < j$. By swapping the order of two vertices in the enumeration of W, we could replace some $f(v_i, v_j)$ with $f(v_j, v_i)$; if f is not symmetric, this would change the value of the product.

For instance,

$$t_{\mathrm{C}_3}(f) = \int f(x,y)f(x,z)f(y,z)\,d\mu_3.$$

But, since we do not attach any significance to the order of the vertices, this should be equal to

$$\int f(x,y)f(z,x)f(y,z)\,d\mu_3.$$

One advantage of this notation is that it gives us a quick way to calculate things like $\mathbb{E}(t_H(\mathbf{R}_{1/2}))$: for the purposes of calculating expected subgraph density, an edge which exists half the time (when the corresponding coin is heads) is equivalent to a "weighted edge" which is always equal to 1/2. So instead of looking at subgraph density in a random graph, we can look at subgraph density in a function which is constantly equal to 1/2.

Theorem 1.15. *Let $f : V^2 \to \mathbb{R}$ be the function which is constantly equal to 1/2. Then for each $H = (W, F)$, there is a constant C so that*

$$\left|\mathbb{E}(t_H(\mathbf{R}_{1/2})) - t_H(f)\right| < \frac{C}{|V|}.$$

Proof. Using the linearity of expectation, $\mathbb{E}(t_H(\mathbf{R}_{1/2})) = \frac{1}{|V|^{|W|}} \sum_{\pi:W\to V} \mathbb{E}(\mathbf{1}_\pi)$ where $\mathbf{1}_\pi$ is the indicator variable which is 1 if π is a copy of H.

When π is injective,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_\pi) &= 1 \cdot \mathbb{P}(\pi \text{ is a copy of } H) + 0 \cdot \mathbb{P}(\pi \text{ is not a copy of } H)\\ &= 2^{-|F|}\\ &= \prod_{1\le i<j\le k, \{w_i,w_j\}\in F} f(v_i, v_j).\end{aligned}$$

So $\mathbb{E}(t_H(\mathbf{R}_{1/2}))$ and $t_H(f)$ are both averages which agree on all the injective π; since the non-injective π contribute less than $\frac{C}{|V|}$, we have

$$\left|\mathbb{E}(t_H(\mathbf{R}_{1/2})) - t_H(f)\right| < \frac{C}{|V|}.$$

□

1.4 Counting Triangles

We now return to the question of when a graph "looks random". Having shown that a random graph should have about 1/8 of the possible triangles, we would like to show that this is not enough to identify a graph as random: that there are graphs which have 1/2 of the possible edges, 1/8 of the possible triangles, and are still non-random.

Our approach will be to start with two graphs, both of which have the right number of edges, but where one has too many triangles and the other has too few. Then we'll interpolate between these graphs to find one with the right number of triangles.

We've already seen a graph with too few triangles: $\mathrm{K}_{n,n}$ has about half the edges, but no triangles at all. For a graph with too many triangles, the "complement" of $\mathrm{K}_{n,n}$, which we will call $\overline{\mathrm{K}_{n,n}}$, works: the graph $(V \cup W, \binom{V}{2} \cup \binom{W}{2})$.

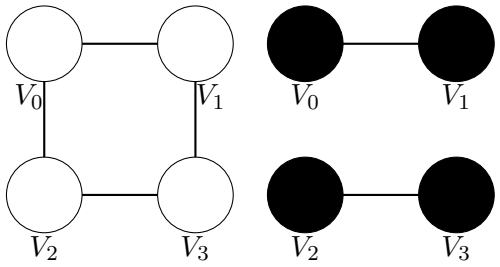
$\overline{\mathrm{K}_{n,n}}$ has two pieces V and W and all edges within V, all edges within W, and none between the two parts. We could think of $\overline{\mathrm{K}_{n,n}}$ as the disjoint union of the complete graph on V with the complete graph on W.

$\overline{\mathrm{K}_{n,n}}$ has roughly half the edges, but about 1/4 of the possible triangles: an ordered triple (v_0, v_1, v_2) is a triangle in this graph as long as both v_1 and v_2 are in the same part as v_0, so if we select a possible triangle $\pi : \{0, 1, 2\} \to V \cup W$, for any choice $v_0 = \pi(0)$, half the choices for $\pi(1)$ and half the choices for $\pi(2)$ will give us an actual triangle.

We'll combine these as follows. First, fix some value of n. We'll work in a graph with $4n$ vertices V divided into *four* disjoint equally sized sets of vertices, $V = V_0 \cup V_1 \cup V_2 \cup V_3$, each with n vertices. We'll define two graphs on these vertices:

- E_0 consists of all pairs with one vertex in $V_0 \cup V_3$ and one vertex in $V_1 \cup V_2$, and

- $E_1 = \binom{V_0 \cup V_1}{2} \cup \binom{V_2 \cup V_3}{2}$—all pairs with both vertices in $V_0 \cup V_1$, or all pairs with both vertices in $V_2 \cup V_3$.



E_0 E_1

In these pictures, the lines mean that we have all edges between the two parts, and the filled circles mean we have all edges within that part. There are no edges within the empty circles, and no edges between parts with no line between them.

Notice that $G_0 = (V, E_0)$ is really $\mathrm{K}_{n,n}$—the two parts are $V_0 \cup V_3$ and $V_1 \cup V_2$. Similarly, $G_1 = (V, E_1)$ is really $\overline{\mathrm{K}_{n,n}}$, except the parts are $V_0 \cup V_1$ and $V_2 \cup V_3$.

Next we define a family of partially random graphs interpolating between these. For the purposes of this section, we will call these graphs $\mathbf{G}_p$ for $p \in [0, 1]$. $\mathbf{G}_p = (V, \mathbf{E}_p)$ will be a graph on the same set of vertices $V = V_0 \cup V_1 \cup V_2 \cup V_3$. For each pair $\{v, w\}$, we flip a weighted coin which comes up heads with probability p and tails with probability $1 - p$. If the coin comes up heads, we place the edge in $\mathbf{E}_p$ if it's present in E_1. If the coin comes up tails, we place the edge in $\mathbf{E}_p$ if it's present in E_0. As usual, all the coins are flipped independently.

It is convenient to represent $\mathbf{G}_p$ with a grid

	V_0	V_1	V_2	V_3
V_0	p	1	1-p	0
V_1	1	p	0	1-p
V_2	1-p	0	p	1
V_3	0	1-p	1	p

which indicates that we have all the edges between V_0 and V_1 (because those edges are present in both G_0 and G_1) but none of the edges between V_1 and V_2 (because those are absent in both G_0 and G_1). Each edge within V_0 has probability p of being included—that edge is only present in G_1, so only appears when the coin for that pair comes up heads, while each edge between V_0 and V_2 has probability $1 - p$ of being included, because it only appears when the coin for that pair comes up tails.

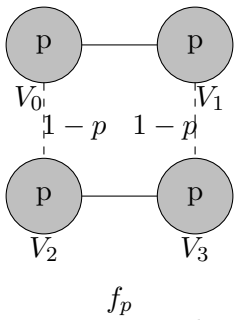
Analogously to the way we counted densities in the random graph using a function constantly equal to $1/2$, we can count densities in $\mathbf{G}_p$ using a function f_p which reflects the grid above.

Theorem 1.16. *Let $\mathbf{G}_p$ be the randomly generated graph defined above. Let $f_p : V^2 \to [0,1]$ be the function given by*

$$f_p(v,w) = \begin{cases} 1 & \text{if } (v,w) \in (V_0 \times V_1) \cup (V_1 \times V_0) \cup (V_2 \times V_3) \cup (V_3 \times V_2) \\ p & \text{if } (v,w) \in (V_0 \times V_0) \cup (V_1 \times V_1) \cup (V_2 \times V_2) \cup (V_3 \times V_3) \\ 1-p & \text{if } (v,w) \in (V_0 \times V_2) \cup (V_2 \times V_0) \cup (V_1 \times V_3) \cup (V_3 \times V_1) \\ 0 & \text{if } (v,w) \in (V_0 \times V_3) \cup (V_3 \times V_0) \cup (V_1 \times V_2) \cup (V_2 \times V_1) \end{cases}.$$

Then for any $H = (W,F)$, there is a C so that $|\mathbb{E}(t_H(\mathbf{G}_p)) - t_H(f_p)| < \frac{C}{|V|}$.

The function f_p looks like this:



f_p

where the gray circles indicate that the vertices in these are present with probability p, edges between the parts connected by dashed lines are present with probability $1-p$.

Proof. The method is the same as the one we used in the previous section:

$$\mathbb{E}(t_H(\mathbf{G}_p)) = \frac{1}{|V|^{|W|}} \sum_{\pi : W \to V} \mathbb{E}(\mathbf{1}_\pi)$$

where $\mathbf{1}_\pi$ is 1 if π is a copy of H and 0 otherwise.

When π is injective, $\mathbb{E}(\mathbf{1}_\pi) = \prod_{\{w,w'\} \in F} \mathbb{E}(\mathbf{1}_{\{\pi(w),\pi(w')\}})$ where $\mathbf{1}_{\{\pi(w),\pi(w')\}}$ is 1 if $\{\pi(w),\pi(w')\}$ is an edge in $\mathbf{G}_p$ and 0 otherwise. Since f_p is exactly defined so that $\mathbb{E}(\mathbf{1}_{\{v,w\}}) = f_p(v,w)$, we have

$$\begin{aligned} \mathbb{E}(t_H(\mathbf{G}_p)) &= \frac{1}{|V|^{|W|}} \sum_{\pi : W \to V, \pi \text{ injective}} \prod_{\{w,w'\} \in F} f_p(\pi(w), \pi(w')) + O(\frac{1}{|V|}) \\ &= t_H(f_p) + O(\frac{1}{|V|}). \end{aligned}$$

□

We also have

Theorem 1.17. *For every $\epsilon > 0$, every $\delta > 0$, and every finite graph $H = (W, F)$, whenever n is sufficiently large,*

$$\mathbb{P}\left(|t_H(\mathbf{G}_p) - \mathbb{E}(t_H(\mathbf{G}_p))| < \epsilon\right) \geq (1 - \delta).$$

Proof. This is the same argument using McDiarmid's inequality as for the random graph: the quantity $t_H(\mathbf{G}_p)$ is a function of the random variables $\mathbf{1}_{\{v,w\}}$ with the property that changing any single edge can only change $t_H(\mathbf{G}_p)$ by at most $\frac{|W|^2|V|^{|W|-2}}{|V|^{|W|}} = \frac{|W|^2}{|V|^2}$.

So, by McDiarmid's inequality, the probability that $t_H(\mathbf{G}_p)$ differs from $\mathbb{E}(t_H(\mathbf{G}_p))$ by more than ϵ is at most

$$2e^{-\frac{2\epsilon^2}{\binom{|V|}{2}\left(\frac{|W|^2}{|V|^2}\right)^2}} \leq 2e^{-\frac{\epsilon^2|V|^2}{|W|^4}}.$$

In particular, when $n = |V|/4$ is sufficiently large, the probability that $|t_H(\mathbf{G}_p) - \mathbb{E}(t_H(\mathbf{G}_p))| \geq \epsilon$ is $< \delta$. □

Theorem 1.18. *For any $\epsilon > 0$, when n is sufficiently large, with probability $\geq (1 - \epsilon)$:*

- $|t_{\mathrm{K}_2}(\mathbf{G}_p) - 1/2| < \epsilon$, *and*
- $|t_{\mathrm{C}_3}(\mathbf{G}_p) - \frac{1}{8}(p^3 - 2p^2 + 3p)| < \epsilon$.

Proof. Combining the previous two theorems, it suffices to show that $t_{\mathrm{K}_2}(f_p) = 1/2$ and $t_{\mathrm{C}_3}(f_p) = \frac{1}{8}(p^3 - 2p^2 + 3p)$.

For the first claim, $t_{\mathrm{K}_2}(f_p) = \int f_p(v, w)d\mu_2$. The four components V_0, V_1, and so on are symmetric, so

$$\begin{aligned}
\int f_p(v,w)d\mu_2 &= 4\int_{V_0\times V} f_p(v,w)d\mu_2 \\
&= 4\Big(\int_{V_0\times V_0} f_p(v,w)d\mu_2 + \int_{V_0\times V_1} f_p(v,w)d\mu_2 \\
&\quad + \int_{V_0\times V_2} f_p(v,w)d\mu_2 + \int_{V_0\times V_3} f_p(v,w)d\mu_2\Big) \\
&= 4(p\mu(V_0\times V_0) + \mu(V_0\times V_1) + (1-p)\mu(V_0\times V_2) + 0\mu(V_0\times V_3)) \\
&= 4(\frac{p}{4} + \frac{1}{4} + \frac{1-p}{4}) \\
&= 1/2.
\end{aligned}$$

The calculation of $t_{C_3}(f_p)$ is similar but more complicated. Again, the symmetry of the four components means it suffices to consider the case where the first vertex is in V_0; this leaves 16 cases, which is tedious but not infeasible, especially when we combine symmetric cases:

$$\begin{aligned}
t_{C_3}(f_p) &= \int f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&= 4\int_{V_0\times V\times V} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&= 4\Bigg(\int_{V_0\times V_0\times V_0} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 + 2\int_{V_0\times V_0\times V_1} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&\quad + 2\int_{V_0\times V_0\times V_2} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 + 2\int_{V_0\times V_0\times V_3} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&\quad + \int_{V_0\times V_1\times V_1} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 + 2\int_{V_0\times V_1\times V_2} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&\quad + 2\int_{V_0\times V_1\times V_3} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 + \int_{V_0\times V_2\times V_2} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 \\
&\quad +2\int_{V_0\times V_2\times V_3} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3 + \int_{V_0\times V_3\times V_3} f_p(u,v)f_p(u,w)f_p(v,w)d\mu_3\Bigg) \\
&= 4\frac{1}{4^3}(p^3 + 2p + 2p(1-p)^2 + 0 + p + 0 + 0 + p(1-p)^2 + 0 + 0) \\
&= \frac{1}{8}(p^3 - 2p^2 + 3p).
\end{aligned}$$

□

Of course, $\mathbf{G}_1$ should just be the graph G_1 we started with, so we are not surprised that $t_{C_3}(\mathbf{G}_1)$ is 1/4; similarly, $\mathbf{G}_0$ should just be the graph G_0, so $t_{C_3}(\mathbf{G}_0) = 0$.

But since $\mathbb{E}(t_{C_3}(\mathbf{G}_p))$ is continuous in p, there must be some value $p^* \in (0,1)$ so that $\mathbb{E}(t_{C_3}(\mathbf{G}_{p^*})) = 1/4$.

So the graph $\mathbf{G}_{p^*}$ more closely resembles a random graph. But it still has some distinctly "non-random" features: the components V_0, V_1, and V_3 are all large sets, each with a quarter of the total vertices, but there are no edges at all between V_0 and V_3, while every edge between V_0 and V_1 is present.

This is suspicious, and indeed, we can rule it out:

Theorem 1.19. *For each $\epsilon > 0$ and $\delta > 0$ there is a C so that, when V is sufficiently big, with probability $\geq (1-\delta)$, for* all *subsets $X \subseteq V$, $Y \subseteq V$ with $|X| \geq C\ln|V|$ and $|Y| \geq C\ln|V|$,*

$$\left|\frac{|\mathbf{R}_{1/2} \cap (X\times Y)|}{|X|\cdot|Y|} - \frac{1}{2}\right| < \epsilon.$$

This says that in a random graph, we expect to have the property that not only are about half the edges present, but whenever we look at subsets X and Y which aren't too tiny, the "edge density" between X and Y is also close to one half.

Note that by $\mathbf{R}_{1/2} \cap (X \times Y)$, we mean the set of *ordered* pairs (x, y) such that $x \in X$, $y \in Y$, and $\{x, y\} \in \mathbf{R}_{1/2}$. In particular, if $x, y \in X \cap Y$, this means the pair (x, y) should be counted twice—once for the order (x, y), and once for the order (y, z).

Proof. The idea is that if we pick a single pair of sets X and Y in advance, we expect close to half the edges in $X \times Y$ will belong to $\mathbf{R}_{1/2}$. If we then look at many pairs of sets, each pair of sets has a small probability of "being defective"—of having either too many or too few edges between them. Exactly how small that probability is depends on the sizes of X and Y—it is some value $p(|X|, |Y|)$. (More precisely, the probability depends not just on $|X|$ and $|Y|$, but on $|X \cap Y|$. We will take $p(|X|, |Y|)$ to be the maximum of this probability over possible values of $|X \cap Y|$.)

So we could fix sizes x and y, consider all pairs of sets where $|X| = x$ and $|Y| = y$, and ask what the probability that there is *at least one* "defective" pair X and Y with $|X| = x$ and $|Y| = y$. At worst, this is $\binom{|V|}{x}\binom{|V|}{y}p(x, y)$—that is, at worst, the probability that there is at least one defective pairs is the number of pairs times the probability that an individual pair is defective. (This is called the *union bound*.) There is a bit of a tradeoff: when we make x and y are small relative to $|V|$, $\binom{|V|}{x}$ and $\binom{|V|}{y}$ get small, but $p(x, y)$ gets bigger.

Furthermore, we want to consider all pairs $x, y \geq C \ln |V|$, which we do by using the union bound again: $\binom{|V|}{x}\binom{|V|}{y}p(x, y)$ is an upper bound on the probability that there is a defective pair with $|X| = x$ and $|Y| = y$, so (using the union bound again)

$$\sum_{x \geq C \ln |V|, y \geq C \ln |V|} \binom{|V|}{x}\binom{|V|}{y}p(x, y)$$

is an upper bound on the probability that there is any defective pair. The bound $C \ln |V|$ comes from doing out the calculations and figuring out how big x and y need to be to keep this sum small.

To find the actual value of $p(x, y)$, suppose we have arbitrary sets X and Y with $|X| = x$ and $|Y| = y$. By the same arguments we used for the whole

graph $\mathbf{R}_{1/2}$, we see that

$$\begin{aligned}\mathbb{E}\left(\frac{|\mathbf{R}_{1/2}\cap(X\times Y)|}{xy}\right) &= \frac{1}{xy}\sum_{(x,y)\in X\times Y}\mathbf{1}_{\{x,y\}}\\ &= 1/2.\end{aligned}$$

The quantity $\frac{|\mathbf{R}_{1/2}\cap(X\times Y)|}{xy}$ is a function of at least $\frac{x(y-1)}{2}\geq\frac{xy}{4}$ random variables (accounting for the worst case where $X=Y$); changing a single edge affects the total by at most $\frac{2}{xy}$ (since the edge is counted at most twice). So by McDiarmid's inequality,

$$\mathbb{P}\left(\left|\frac{|\mathbf{R}_{1/2}\cap(X\times Y)|}{xy}-\frac{1}{2}\right|\geq\epsilon\right)\leq 2e^{-2\epsilon^2 xy}.$$

So the probability that there is at least one "defective pair" in $\mathbf{R}_{1/2}$—that is, the probability that there are some sets X and Y with $|X|\geq C\ln|V|$ and $|Y|\geq C\ln|V|$ so that

$$\left|\frac{|\mathbf{R}_{1/2}\cap(X\times Y)|}{xy}-\frac{1}{2}\right|\geq\epsilon$$

is at most

$$\sum_{x\geq C\ln|V|,y\geq C\ln|V|}\binom{|V|}{x}\binom{|V|}{y}2e^{-2\epsilon^2xy}.$$

Using *Stirling's Approximation*, we have the bound $\binom{n}{k}\leq\left(\frac{ne}{k}\right)^k$, so

$$\binom{|V|}{x}\binom{|V|}{y}\leq e^{x\ln|V|+y\ln|V|+x+y}.$$

Then the probability that there is at least one "defective pair" is at most

$$\begin{aligned}&\sum_{x\geq C\ln|V|,y\geq C\ln|V|}e^{x\ln|V|+y\ln|V|+x+y}2e^{-2\epsilon^2xy}\\ =&\sum_{x\geq C\ln|V|,y\geq C\ln|V|}2e^{x(\ln|V|+1)+y(\ln|V|+1)-2\epsilon^2xy}\\ =&\sum_{x\geq C\ln|V|,y\geq C\ln|V|}2e^{\epsilon^2(x(\ln|V|+1-y)+y(\ln|V|+1-x)}.\end{aligned}$$

By taking C a bit larger than $1/\sqrt{2}\epsilon$, we can bound this by:

$$\begin{aligned}
&\sum_{x\geq C\ln|V|,y\geq C\ln|V|} 2e^{\epsilon^2(x(\ln|V|+1-y)+y(\ln|V|+1-x)} \\
&\qquad\leq \sum_{x\geq C\ln|V|,y\geq C\ln|V|} 2e^{-\frac{\epsilon}{\sqrt{2}}\ln|V|(x+y)} \\
&\qquad\leq \sum_{x\geq C\ln|V|,y\geq C\ln|V|} 2e^{-\sqrt{2}\epsilon C\ln^2|V|} \\
&\qquad\leq \sum_{x\geq C\ln|V|,y\geq C\ln|V|} 2e^{-\ln^2|V|} \\
&\qquad\leq 2|V|^2e^{-\ln^2|V|}.
\end{aligned}$$

In particular, when V is large enough, the probability is less than δ. □

So the graph $\mathbf{G}_{p^*}$ still does not resemble $\mathbf{R}_{1/2}$: $\frac{|\mathbf{G}_{p^*}\cap(V_0\times V_3)|}{|V_0|\cdot|V_3|}=0$, which would be very unlikely in $\mathbf{R}_{1/2}$.

1.5 Counting Cycles of Length 4

Another way to try to distinguish $\mathbf{G}_{p^*}$ from $\mathbf{R}_{1/2}$ would be to investigate other subgraph densities. The next natural density to find is the graph C_4, the cycle of length 4:

Definition 1.20. C_4 is the graph with 4 vertices and 4 edges arranged in a cycle:  .

We could also call this graph $\mathrm{K}_{2,2}$, and sometimes that perspective is more useful: there are two pairs of vertices, $\{x,x'\}$ and $\{y,y'\}$, and the edges are exactly those with one vertex from each pair.

Theorem 1.21. *For each $\epsilon>0$ and $\delta>0$, when n is sufficiently large, with probability $\geq 1-\delta$, $|t_{\mathrm{C}_4}(\mathbf{G}_p)-\frac{1}{8}(p^4-2p^3+3p^2-2p+1)|<\epsilon$ (where $\mathbf{G}_p$ is the graph from the previous section).*

Sketch. Again, it suffices to show that $t_{\mathrm{C}_4}(f_p)=\frac{1}{8}(p^4-2p^3+3p^2-2p+1)$. The calculation is still a little tedious: since

$$t_{\mathrm{C}_4}(f_p)=\int f_p(x,y)f_p(y,x')f_p(x',y')f_p(y',x)d\mu_4,$$

there are 4^4 cases to consider, depending on which of the four parts the four vertices belong to. Using symmetry, it suffices to consider only the case where $u \in V_0$:

$$t_{\mathrm{C}_4}(f_p) = 4 \int_{V_0 \times V^3} f_p(x,y) f_p(y,x') f_p(x',y') f_p(y',x) d\mu_4,$$

which leaves us with "only" 64 cases to consider. Considering each of these cases in turn will give the stated polynomial.

□

As it happens, that means that $t_{\mathrm{C}_4}(\mathbf{G}_{p^*}) \approx 0.07123$, which is not the $1/16$ that we would expect in a random graph. Indeed, there is no value of p which gives $\mathbb{E}(t_{\mathrm{C}_4}(\mathbf{G}_p)) = 1/16$: the polynomial $\frac{1}{8}(p^4 - 2p^3 + 3p^2 - 2p + 1)$ achieves its minimum when $p = 1/2$, and even then, it is equal to $9/128$. So no matter what p is, $\mathbf{G}_p$ will (with high probability) have too many cycles of length 4 to be a truly random graph.

This turns out to be general: there are no graphs with too few copies of C_4.

Lemma 1.22. *For any graph G,*

$$t_{\mathrm{C}_4}(G) \geq (t_{\mathrm{K}_2}(G))^4.$$

Proof. This follows by a couple applications of a simple form of the Cauchy-Schwarz inequality.

Theorem (Cauchy-Schwarz)**.**

$$\left| \int f(x)\, d\mu \right|^2 \leq \int |f(x)|^2\, d\mu.$$

$$\begin{aligned}
(t_{\mathrm{K}_2}(G))^4 &= \left(\int \chi_E(x,y)\, d\mu_2\right)^4 \\
&= \left(\int (\int \chi_E(x,y) d\mu(y))\, d\mu(x)\right)^4 \\
&\leq \left(\int \left(\int \chi_E(x,y) d\mu(y)\right)^2 d\mu(x)\right)^2 \\
&= \left(\int\int \chi_E(x,y) d\mu(y) \int \chi_E(x,y') d\mu(y') d\mu(x)\right)^2 \\
&= \left(\int \chi_E(x,y)\chi_E(x,y') d\mu_3\right)^2 \\
&= \left(\int \left(\int \chi_E(x,y)\chi_E(x,y') d\mu(x)\right) d\mu_2(y,y')\right)^2 \\
&\leq \int \left(\int \chi_E(x,y)\chi_E(x,y') d\mu(x)\right)^2 d\mu_2(y,y') \\
&= \int \chi_E(x,y)\chi_E(x,y')\chi_E(x',y)\chi_E(x',y') d\mu_4 \\
&= t_{\mathrm{C}_4}(G).
\end{aligned}$$

□

In graphs with very few edges (for instance, a "perfect matching" with $2n$ vertices and n edges where each vertex has exactly one neighbor), this is a little counterintuitive—it depends on the fact that our definition of $t_{\mathrm{C}_4}(G)$ includes "degenerate cycles" in which $x = x'$ and $y = y'$. However this only matters in very sparse graphs (the perfect matching has edge density $t_{\mathrm{K}_2}(G) = \frac{1}{4n}$). In the case we are mostly interested in, where $t_{\mathrm{K}_2}(G)$ is some fixed real number $p \in (0,1)$ and n is quite large, there are at least $p^4 n^4$ copies of C_4, of which at most around n^3 are degenerate, so the degenerate copies contribute only a small error term to $t_{\mathrm{C}_4}(G)$.

This lemma rules out any hope of repeating what we did for triangles: we can't find a graph with too few copies of C_4 to balance against a graph with too many.

Instead, we'll take the idea that graphs with the correct number of copies of C_4 really are special in some way, and our goal for the rest of the chapter will be to explore what properties they have. Towards this, we define:

Definition 1.23. A graph $G = (V, E)$ is *ϵ-quasirandom* if

$$|t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4| < \epsilon.$$

We have written this with an absolute value to emphasize that the point is that $t_{C_4}(G)$ is close to $(t_{K_2}(G))^4$; however, because of the previous lemma, it is equivalent just to have $t_{C_4}(G) - (t_{K_2}(G))^4 < \epsilon$.

Although we have only discussed $\mathbf{R}_{1/2}$ so far, it is not a big jump to generalize to the graph $\mathbf{R}_p$, which is generated by independently flipping, for each pair $\{v, w\}$, a weighted coin* which comes up heads with probability p and including this edge if the coin comes up heads.

Our choice of the name "quasirandom" indicates that when $t_{C_4}(G) \approx (t_{K_2}(G))^4$, $t_{K_2}(G) = p$, and G has a large number of vertices, G is supposed to resemble $\mathbf{R}_p$, a claim we will justify below.

Note that being quasirandom is a very different sort of property than being random. $\mathbf{R}_{1/2}$ is not a particular graph, nor even a property of a graph: it is a method of producing a graph. If we encounter a graph "in the wild", it is not meaningful to ask "is this graph $\mathbf{R}_{1/2}$?"; all we can ask is the question we have been asking: how does this graph resemble, or fail to resemble, a typical graph generated according to $\mathbf{R}_{1/2}$.

By contrast, quasirandomness is a conventional property of graphs; a given graph either does or does not have the property of being ϵ-quasirandom.

At a minimum if the terminology is reasonable, when a graph is generated randomly, it should, with high probability, be quasirandom.

Theorem 1.24. *For every $\epsilon > 0$, every $\delta > 0$, and every $p \in (0, 1)$, when V is sufficiently large, $\mathbf{R}_p$ is ϵ-quasirandom with probability $\geq (1 - \delta)$.*

Proof. We can choose ϵ' so that if $|a - p| < \epsilon'$ then $|a^4 - p^4| < \epsilon/2$. By the same argument using McDiarmid's inequality as in Theorem 1.7, when V is big enough, the probability that both

$$|t_{K_2}(\mathbf{R}_p) - \mathbb{E}(t_{K_2}(\mathbf{R}_p))| < \epsilon'$$

and

$$|t_{C_4}(\mathbf{R}_p) - \mathbb{E}(t_{C_4}(\mathbf{R}_p))| < \epsilon/2$$

is $\geq (1 - \delta)$.

Let f_p be the function which is constantly equal to p. Then

$$\mathbb{E}(t_{K_2}(\mathbf{R}_p)) = t_{K_2}(f_p) = p$$

*A theoretical weighted coin, since such a contraption is not possible in practice, as discussed in [69]

and

$$\mathbb{E}(t_{C_4}(\mathbf{R}_p)) = t_{C_4}(f_p) = \int f_p(x,y)f_p(y,x')f_p(x',y')f_p(y',x)d\mu_4 = p^4.$$

Therefore, with probability $\geq 1-\delta$

$$\left|t_{C_4}(\mathbf{R}_p) - (t_{K_2}(\mathbf{R}_p))^4\right| \leq \left|t_{C_4}(\mathbf{R}_p) - p^4\right| + \left|(t_{K_2}(\mathbf{R}_p))^4 - p^4\right| < \epsilon/2 + \epsilon/2 = \epsilon.$$

□

1.6 Quasirandom Graphs

We now turn to justifying the name "quasirandom", showing that quasirandom graphs really do resemble random ones.

First, as something of a warm-up, we show that in a quasirandom graph, the edges are evenly distributed—most vertices have the same number of neighbors.

Definition 1.25. When $G = (V, E)$ is a graph and $x \in V$, $E_x = \{y \in x \mid \{x, y\} \in E\}$ is the *neighborhood* of x.

The (normalized) degree of x in $G = (V, E)$ is $\deg_G(x) = \frac{|E_x|}{|V|}$.

Note that $\deg_G(x) = \int \chi_E(x, y)d\mu$. (Usually the degree would be $|E_x|$, but the normalized version will be more useful for us.)

Theorem 1.26. *For every $\epsilon > 0$ there is a δ so that if G is δ-quasirandom and $t_{K_2}(G) = p \in [0, 1]$ then the set of x such that*

$$|\deg_G(x) - p| \geq \epsilon$$

has measure $< \epsilon$.

Proof. We'll prove the contrapositive. Suppose the conclusion fails, so $t_{K_2}(G) = p$ but there is a set $S \subseteq V$ such that $\mu(S) \geq \epsilon$ and, for each $x \in S$, $|\deg_G(x) - p| \geq \epsilon$. Then we wish to show that the graph has too many copies of C_4. We will do this by first counting an intermediate shape, the V-shaped graph  consisting of three vertices $\{1, 2, 3\}$ with $\{1, 2\}$ and $\{1, 3\}$ as the only edges. We will show that we have too many—that is, more than p^2—copies of this V-shaped graph, and then use Cauchy-Schwarz to conclude that we have more than p^4 copies of C_4.

Our integral notation helps us focus on the "deviation" from the copies that would be present in a random graph. We define $f(x, y) = \chi_E(x, y) -$

p, the "balanced" version of χ_E, so that, in particular, $\int f(x,y)\,d\mu_2 = \int \chi_E(x,y) - p\,d\mu_2 = 0$. Then, writing $\chi_E(x,y) = p + f(x,y)$, we can calculate

$$\begin{aligned}\int \chi_E(x,y)\chi_E(x,z)d\mu_3 &= \int (p+f(x,y))(p+f(x,z))d\mu_3 \\ &= p^2 + 2p\int f(x,y)d\mu_2 + \int f(x,y)f(x,z)d\mu_3.\end{aligned}$$

The p^2 term accounts for all the copies of the V-shaped graph which we "should" have in a quasirandom graph. The second term vanishes since $\int f(x,y)d\mu_2 = \mu(E) - p = 0$. So it suffices to show that our assumption about the set S will force $\int f(x,y)f(x,z)d\mu_3$ to be non-zero.

We can give a bound

$$\begin{aligned}\int f(x,y)f(x,z)d\mu_3 &= \int (\int f(x,y)d\mu(y))^2 d\mu(x) \\ &= \int (\int \chi_E(x,y) - p\,d\mu(y))^2 d\mu(x) \\ &= \int (\deg_G(x) - p)^2 d\mu \\ &\geq \int_S (\deg_G(x) - p)^2 d\mu \\ &\geq \mu(S)\epsilon^2 \\ &\geq \epsilon^3.\end{aligned}$$

Putting these together, we have

$$\int \chi_E(x,y)\chi_E(x,z)d\mu_3 \geq p^2 + \epsilon^3.$$

Now we use Cauchy-Schwarz:

$$\begin{aligned}t_{C_4}(G) &= \int \left(\int \chi_E(x,y)\chi_E(x,z)d\mu(x)\right)^2 d\mu_2(y,z) \\ &\geq \left(\int \chi_E(x,y)\chi_E(x,z)d\mu_3\right)^2 \\ &\geq (p^2+\epsilon^3)^2 \\ &\geq p^4 + \epsilon^6.\end{aligned}$$

So we see that G cannot be ϵ^6-quasirandom. □

Using this, we will now show that quasirandom graphs satisfy a weakened form of Theorem 1.19. That theorem showed that when X and Y are sufficiently large subsets of V (of size $\geq C \ln |V|$ for some constant C), $\frac{|\mathbf{R}_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|}$ is approximately $1/2$. We will show the same property in quasirandom graphs, but only when X and Y are much larger.

Definition 1.27. When $G = (V, E)$ is a graph and $X \subseteq V, Y \subseteq V$, the *edge density between X and Y* is $d_E(X, Y) = \frac{|E \cap (X \times Y)|}{|X| \cdot |Y|}$.

Theorem 1.28. *For every $\epsilon > 0$ there is a $\delta > 0$ so that whenever $G = (V, E)$ is δ-quasirandom with $t_{\mathrm{K}_2}(G) = p$, $X \subseteq V$, $Y \subseteq V$, and $\frac{|X|}{|V|} \geq \epsilon$ and $\frac{|Y|}{|V|} \geq \epsilon$,*

$$|d_E(X, Y) - p| < \epsilon.$$

Proof. Suppose G is δ-quasirandom for a sufficiently small δ, and consider any sets $X \subseteq V$ and $Y \subseteq V$ with $\mu(X) \geq \epsilon$ and $\mu(Y) \geq \epsilon$.

Then, as in the previous proof, we will show that there are too many copies of C_4. Again, we evaluate $t_{\mathrm{C}_4}(G)$ by looking at the deviation of χ_E from the function which is constantly equal to p. Let $f(x, y) = \chi_E(x, y) - p$. Then

$$\begin{aligned}
t_{\mathrm{C}_4}(G) &= \int \chi_E(x,y)\chi_E(y,x')\chi_E(x',y')\chi_E(y',x)d\mu_4 \\
&= \int (f(x,y)-p)(f(y,x')-p)(f(x',y')-p)(f(y',x)-p)d\mu_4 \\
&= p^4 + 4p\int f(x,y)f(y,x')f(x',y')d\mu_4 + 4p^2\int f(x,y)f(y,x')d\mu_3 \\
&\quad + 2p^2\int f(x,y)f(x',y')d\mu_4 + 4p^3\int f(x,y)d\mu_2 + t_{\mathrm{C}_4}(f).
\end{aligned}$$

Again, p^4 accounts for all the copies of C_4 we should have. $t_{\mathrm{C}_4}(f)$ will have to be non-negative (it is an integral of the square of a quantity), so it will suffice to show that it is large, and that the middle terms are all small enough that they do not cancel out $t_{\mathrm{C}_4}(f)$.

Consider a typical middle term,

$$\int f(x,y)f(y,x')f(x',y)d\mu_4 = \int f(x,y)f(y,x')(\deg_G(x') - p)d\mu_3.$$

Using the previous theorem, since G is δ-quasirandom for some small enough δ, we can ensure that there is a set S such that $\mu(S) < \epsilon^{12}/56$ and, for $x \notin S$, $|\deg_G(x) - p| < \epsilon/56$, so $|\int f(x,y)f(y,x')(\deg_G(x') - p)d\mu_3| < \epsilon^{12}/28$.

The same holds for the other middle terms, so we have

$$|t_{C_4}(G) - (p^4 + t_{C_4}(f))| < \epsilon^{12}/2.$$

What remains is showing that $t_{C_4}(f)$ is large. For convenience, let us abbreviate

$$F(x, x', y, y') = f(x, y)f(y, x')f(x', y')f(y', x).$$

First, we split into four cases, based on whether x and x' are in the set X that we began with:

$$\begin{aligned} t_{C_4}(f) &= \int F(x, x', y, y')d\mu_4 \\ &= \int F(x, x', y, y')\chi_X(x)\chi_X(x')d\mu_4 \\ &\quad + 2\int F(x, x', y, y')\chi_X(x)\chi_{V\setminus X}(x')d\mu_4 \\ &\quad + \int F(x, x', y, y')\chi_{V\setminus X}(x)\chi_{V\setminus X}(x')d\mu_4. \end{aligned}$$

Since

$$\int F(x, x', y, y')\chi_{V\setminus X}(x)\chi_{V\setminus X}(x')d\mu_4 = \int (\int f(x, y)f(y, x')d\mu)^2\chi_{V\setminus X}(x)\chi_{V\setminus X}(x')d\mu_2$$

and

$$\int F(x, x', y, y')\chi_X(x)\chi_{V\setminus X}(x')d\mu_4 = \int (\int f(x, y)f(y, x')d\mu)^2\chi_X(x)\chi_{V\setminus X}(x')d\mu_2,$$

we have

$$t_{C_4}(f) \geq \int F(x, x', y, y')\chi_X(x)\chi_X(x')d\mu_4$$

because the other two integrals are integrals of squares, and therefore non-negative.

We then do the same split on the other variables, based on whether y and y' are in the set Y, and then use Cauchy-Schwarz twice:

$$\begin{aligned} t_{C_4}(f) &\geq \int F(x, x', y, y')\chi_X(x)\chi_X(x')d\mu_4 \\ &\geq \int F(x, x', y, y')\chi_X(x)\chi_X(x')\chi_Y(y)\chi_Y(y')d\mu_4 \\ &\geq \left(\int f(x, y)\chi_X(x)\chi_Y(y)d\mu\right)^4 \\ &\geq ((d_E(X, Y) - p)\mu(X)\mu(Y))^4 \\ &\geq \epsilon^8 d_E(X, Y)^4. \end{aligned}$$

Therefore

$$t_{\mathrm{C}_4}(G) \geq p^4 + t_{\mathrm{C}_4}(f) - \epsilon^{12}/2 \geq p^4 + \epsilon^8 d_E(X,Y)^4 - \epsilon^{12}/2.$$

In particular, since G is δ-quasirandom, when $\delta \leq \epsilon^{12}/2$, we must have $\epsilon^8 d_E(X,Y)^4 - \epsilon^{12}/2 < \epsilon^{12}/2$ and therefore $d_E(X,Y) < \epsilon$. □

This property is useful enough to merit a name.

Definition 1.29. $G = (V,E)$ is *ϵ-regular* if whenever $X \subseteq V$, $Y \subseteq V$ with $\frac{|X|}{|V|} \geq \epsilon$ and $\frac{|Y|}{|V|} \geq \epsilon$,

$$|d_E(X,Y) - p| < \epsilon$$

where $p = t_{\mathrm{K}_2}(G)$.

So Theorem 1.28 can be rephrased

Theorem 1.28, Rephrased. *For every $\epsilon > 0$ there is a $\delta > 0$ so that if G is δ-quasirandom then G is ϵ-regular.*

δ-regularity turns out to be a useful property to work with. In fact, we can use it to show that if a graph has roughly the same number of copies of C_4 as a random graph then it has roughly the same number of *every* small graph.

We illustrate the main idea by counting triangles.

Theorem 1.30. *For every $\epsilon > 0$ and every $p \in [0,1]$ there is a $\delta > 0$ so that whenever $G = (V,E)$ is δ-quasirandom with $t_{\mathrm{K}_2}(G) = p$ and V is sufficiently large, $|t_{\mathrm{C}_3}(G) - p^3| < \epsilon$.*

Proof. Suppose G is δ-quasirandom for some δ small enough (based on the calculations to follow), and we assume that ϵ is much smaller than p (otherwise we can replace it with a smaller ϵ and obtain a stronger conclusion).

For each x, let $E_x = \{y \mid \{x,y\} \in E\}$, so $\deg_G(x) = \mu(E_x)$. Then, for most x, $\mu(E_x) \approx p$ by Theorem 1.26. Furthermore, by Theorem 1.28, $d_E(E_x, E_x) \approx p$. But a triangle whose first vertex is x is exactly an edge between E_x and itself, so there are about p^3 triangles whose first vertex is x. Since this holds for most vertices, it should give us the right number of triangles.

More precisely, by Theorem 1.26, there is a set S with $\mu(S) < \epsilon/2$ such that, $x \notin S$, $|\deg_G(x) - p| < \epsilon/6p$. We can divide $t_{\mathrm{C}_3}(G)$ into those triangles

whose first vertex is in S and those whose first vertex isn't: So

$$\begin{aligned} t_{C_3}(G) &= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)d\mu_3 \\ &= \int \chi_{V\setminus S}(x)\chi_E(x,y)\chi_E(x,z)\chi_E(y,z)d\mu_3 + \int \chi_S(x)\chi_E(x,y)\chi_E(x,z)\chi_E(y,z)d\mu_3 \\ &= \int \chi_{V\setminus S}(x)\mu(E_x)^2 d_E(E_x,E_x)d\mu_3 + \int \chi_S(x)\chi_E(x,y)\chi_E(x,z)\chi_E(y,z)d\mu_3. \end{aligned}$$

Since S is small, there aren't many triangles whose first vertex is in S:

$$\int \chi_S(x)\chi_E(x,y)\chi_E(x,z)\chi_E(y,z)d\mu_3 \le \mu(S) < \epsilon/2.$$

On the other hand, by Theorem 1.28 with $X = Y = E_x$, we also have $|d_E(E_x,E_x) - p| < \epsilon/6p$. So

$$\begin{aligned} \left|\int \chi_{V\setminus S}(x)\mu(E_x)^2 d_E(E_x,E_x)d\mu_3 - p^3\right| &< \left|(p+\epsilon/6p)^2(p+\epsilon/6p) - p^3\right| \\ &\le 3p\epsilon/6p \\ &\le \epsilon/2. \end{aligned}$$

So

$$\left|t_{C_3}(G) - p^3\right| < \epsilon.$$

□

This generalizes to all subgraphs, which perhaps explains what the quasirandom graphs look sufficiently random to merit the name.

Theorem 1.31. *For every finite graph $H = (W,F)$, each $\epsilon > 0$, there is a $\delta > 0$ so that if $G = (V,E)$ is δ-quasirandom with $t_{K_2}(G) = p$, $|t_H(G) - t_{K_2}(G)^{|F|}| < \epsilon$.*

The main idea has already appeared in the previous proof: we want to prove this by induction on H, showing that quasirandom graphs have the right number of copies of larger and larger graphs. The difficulty is that the induction hypothesis needs to be strengthened: it isn't enough to know that G has the right number of copies of H: we need to be able to start with a copy of some subgraph of H and be able to extend it to the right number of copies of H. We already saw this issue in the previous proof: it wasn't enough to know that there are the right number of edges; we needed to know that most vertices x have the right number of neighbors.

Rather than work out this notationally messy induction now, we will postpone the proof until much later when we will have some additional tools to simplify it.

While we used quasirandomness in the proof, we could actually have used regularity instead. In particular, δ-regularity for sufficiently small δ implies $t_{C_4}(G) \approx p^4$, which means

Theorem 1.32. *For every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $G = (V, E)$ is δ-regular and V is sufficiently large, G is ϵ-quasirandom.*

This means that quasirandomness and regularity are, in some sense, equivalent. The proof is essentially a step in the proof of Theorem 1.31, so we defer it to later as well.

1.7 Sequences of Graphs

One inconvenience of ϵ-quasirandomness is that it's always an approximate notion—a finite graph can be ϵ-quasirandom, but never just quasirandom.

We can address this by considering, instead of individual graphs, sequences of graphs:

Definition 1.33. A sequence of graphs $\langle G_n \rangle$ is *quasirandom* if, for each $\epsilon > 0$, there is an m so that whenever $n \geq m$, G_n is ϵ-quasirandom.

So a quasirandom sequence of graphs is a sequence that is "getting more and more random". Since ϵ-quasirandomness can't be small, this implies that the graphs in the sequence must be getting larger and larger.

Then the results in the previous section can be stated:

Theorem 1.34. *Let $\langle G_n \rangle$ be a sequence of graphs with $p = \lim_{n\to\infty} t_{K_2}(G_n)$. The following are equivalent:*

- *the sequence is quasirandom,*
- *for every finite graph $H = (W, F)$, $\lim_{n\to\infty} t_H(G_n) = p^{|F|}$,*
- $\lim_{n\to\infty} t_{C_4}(G_n) = p^4$.

1.8 Digression: Spectral Graph Theory

The previous section suggested two perspectives on quasirandom graphs: the counting perspective based on $t_{C_4}(G)$, and an "equidistribution" perspective based on showing that edges are evenly distributed, in the sense that $d_E(X, Y) \approx p$ whenever X and Y are large sets.

We now describe a third perspective, in terms of eigenvalues associated to a graph.

When E is a graph on V, we can associate a symmetric $|V| \times |V|$ matrix, the *adjacency matrix*, whose value at (x, y) is 1 when $\{x, y\} \in E$, and 0 otherwise. We can describe the same idea in more abstract terms.

The space of functions from V to $\mathbb{R}$ is a $|V|$ dimensional vector space. Furthermore, it has a natural choice of inner product and a corresponding norm:

Definition 1.35. If $f, g : V \to \mathbb{R}$ then $\langle f, g \rangle = \int f(x)g(x)d\mu$ and $||f||_{L^2} = \sqrt{\langle f, f \rangle} = \sqrt{\int (f(x))^2 d\mu}$.

Then a graph—and, more generally, a symmetric function—gives us a linear transformation on this space:

Definition 1.36. When $h : V^2 \to \mathbb{R}$ is a symmetric function, we associate a linear transformation T_h: given $f : V \to \mathbb{R}$, we define $T_h f : V \to \mathbb{R}$ by

$$(T_h f)(x) = \int h(x, y) f(y) d\mu(y).$$

When $h = \chi_E$, we write T_E instead of T_{χ_E}.

$T_h f$ has an asymmetry between the variables, but this vanishes when we look at the inner product:

$$\langle T_h f, g \rangle = \int h(x, y) f(x) g(y) d\mu_2.$$

The symmetry of h means that T_h is self-adjoint:

Lemma 1.37. *T_h is self-adjoint; that is, $\langle T_h f, g \rangle = \langle f, T_h g \rangle$.*

This means the Spectral Theorem applies.

Theorem (Spectral Theorem)**.** *Let T be a self-adjoint operator on an n-dimensional vector space. Then there exist eigenvalues (not necessarily distinct) $\lambda_1, \ldots, \lambda_n$ and vectors $f_1, \ldots, f_n$ such that:*

- *for each $i \leq n$, $||f_i||_{L^2} = 1$,*
- *for each $i \leq n$, $Tf_i = \lambda_i f_i$,*
- *for each $i < j \leq n$, $\langle f_i, f_j \rangle = 0$.*

Furthermore, the values λ_i are unique and, for each λ, the space generated by $\{f_i \mid \lambda_i = \lambda\}$ is uniquely determined.

In particular, the vectors $f_1, \ldots, f_n$ form an orthonormal basis.

For example, recall our bipartite graph $K_{n,n}$, consisting of two pieces, say X and Y, with n vertices each, and exactly the edges between those two parts. Then λ_1 turns out to be within $O(1/|V|)$ of $1/2$, and the corresponding f_1 is the function which is constantly equal to 1.

The next eigenfunction explains the bipartite structure: f_2 is the function which is 1 on one of the pieces and -1 on the other. Then

$$\begin{aligned}\lambda_2 &= \langle T_E f_2, f_2\rangle \\ &= \int \chi_E(x,y) f_2(x) f_2(y) d\mu_2 \\ &= \int_{X\times X} 0 \cdot 1 \cdot 1 d\mu_2 + \int_{X\times Y} 1 \cdot 1 \cdot -1 d\mu_2 + \int_{Y\times X} 1 \cdot -1 \cdot 1 d\mu_2 + \int_{Y\times Y} 0 \cdot -1 \cdot -1 d\mu_2 \\ &= -\frac{1}{2}.\end{aligned}$$

In fact, these two eigenvalues completely explain the space—all other eigenvalues are 0. To see this, observe that we can write

$$\chi_E(x,y) = \frac{1}{2} f_1(x) f_1(y) - \frac{1}{2} f_2(x) f_2(y).$$

The fact that we can write χ_E as a weighted sum of the eigenvectors is not a coincidence.

Theorem 1.38. *When $\lambda_1, \ldots, \lambda_n$ and $f_1, \ldots, f_n$ are the eigenvalues and eigenvectors associated to T_h, $h(x,y) = \sum_{i\le n} \lambda_i f_i(x) f_i(y)$.*

Proof. First we show that h and $\sum_{i\le n} \lambda_i f_i(x) f_i(y)$ give the same operator.

Consider any $g : V \to \mathbb{R}$. Since the $\{f_i\}$ form an orthonormal basis, we can write $g(x) = \sum_{i\le n} c_i f_i(x)$ for some unique sequence of coefficients c_i. Then we can check that

$$\begin{aligned}\int h(x,y) g(y) d\mu(y) &= \int h(x,y) \sum_{i\le n} c_i f_i(x) d\mu(y) \\ &= \sum_{i\le n} c_i \int h(x,y) f_i(x) d\mu(y) \\ &= \sum_{i\le n} c_i \lambda_i f_i(x)\end{aligned}$$

while, similarly,

$$\begin{aligned}\int \left(\sum_{i\leq n} \lambda_i f_i(x) f_i(y)\right) g(y) d\mu(y) &= \int \left(\sum_{i\leq n} \lambda_i f_i(x) f_i(y)\right) (\sum_{i\leq n} c_i f_i(y)) d\mu(y) \\ &= \sum_{i\leq n}\sum_{j\leq n} \int \lambda_i f_i(x) f_i(y) c_j f_j(y) d\mu(y) \\ &= \sum_{i\leq n}\sum_{j\leq n} c_j \lambda_i f_i(x) \int f_i(y) f_j(y) d\mu(y).\end{aligned}$$

But $\int f_i(y) f_j(y) d\mu$ is 0 if $i \neq j$ and 1 if $i = j$, so this last line is equal to

$$\sum_{i\leq n} c_i \lambda_i f_i(x)$$

as well.

The equality as transformations is really the important fact, but we can use this to show that these are literally the same as functions as well. Start with any $y_0 \in V$ and consider the function $\chi_{\{y_0\}}$ which is 1 on y_0 and 0 everywhere else. Then

$$(T_h \chi_{\{y_0\}})(x) = \int h(x,y) \chi_{\{y_0\}}(y) d\mu(y) = \frac{1}{|V|} h(x, y_0).$$

But that means that

$$\int \left(\sum_{i\leq n} \lambda_i f_i(x) f_i(y)\right) \chi_{\{y_0\}}(y) d\mu(y) = \frac{1}{|V|} h(x, y_0)$$

as well, which means that $\sum_{i\leq n} \lambda_i f_i(x) f_i(y_0) = h(x, y_0)$ for every x. Since this holds for every y_0,

$$\sum_{i\leq n} \lambda_i f_i(x) f_i(y) = h(x,y).$$

□

This gives a convenient proof of the standard result relating the trace of a matrix to the sum of its eigenvalues:

Theorem 1.39. *If $h : V^2 \to \mathbb{R}$ is symmetric and $\lambda_1, \ldots, \lambda_n$ are the associated eigenvalues,*

$$\int h(x,x) d\mu = \sum_{i\leq n} \lambda_i.$$

Proof. This follows from the previous theorem since

$$\begin{aligned}\int h(x,x)d\mu &= \int \sum_{i\le n} \lambda_i f_i(x) f_i(x) d\mu \\ &= \sum_{i\le n} \lambda_i \int (f_i(x))^2 d\mu \\ &= \sum_{i\le n} \lambda_i\end{aligned}$$

since each f_i has norm 1. □

It is also not a coincidence that the largest eigenvalue of $\mathrm{K}_{n,n}$ was equal to 1/2—it will always be the case that when $\deg_G(x)$ is constant, the largest eigenvalue is equal to this constant. More generally, we obtain a lower bound on the largest eigenvalue:

Lemma 1.40. *When λ_1 is the largest eigenvalue (in absolute value) of T_h, $|\lambda_1| \ge t_{\mathrm{K}_2}(h)$.*

Proof. Consider the function $g(x)$ which is constantly equal to 1. Then

$$\langle T_h g, g\rangle = \int h(x,y)g(x)g(y)d\mu_2 = \int h(x,y)d\mu_2 = t_{\mathrm{K}_2}(h).$$

We can stop here if we invoke a standard fact about eigenvalues, that the largest eigenvalue λ_1 satisfies $|\lambda_1| = \max_u \langle T_h u, u\rangle$ where u ranges over the unit vectors: since g is a unit vector, $|\lambda_1| \ge \langle T_h g, g\rangle = t_{\mathrm{K}_2}(h)$.

If we do not wish to invoke this, we can show that $|\lambda_1| \ge \langle T_h g, g\rangle$ by writing g in terms of the orthonormal basis $f_1, \ldots, f_n$: for some choice of constants c_i, $g(x) = \sum_{i\le n} c_i f_i(x)$, and we have

$$\begin{aligned}1 &= \int (g(x))^2 d\mu \\ &= \int \left(\sum_{i\le n} c_i f_i(x)\right)\left(\sum_{j\le n} c_j f_j(x)\right) d\mu \\ &= \sum_{i\le n}\sum_{j\le n} c_i c_j \int f_i(x) f_j(x) d\mu \\ &= \sum_{i\le n} c_i^2.\end{aligned}$$

Then

$$\begin{aligned}
t_{\mathrm{K}_2}(h) &= \int h(x,y)g(x)g(y)d\mu_2 \\
&= \int \left(\sum_{i\le n} \lambda_i f_i(x)f_i(y)\right)\left(\sum_{j\le n} c_j f_j(x)\right)\left(\sum_{k\le n} c_k f_k(y)\right) d\mu_2 \\
&= \sum_{i\le n} \lambda_i \left(\int \sum_{j\le n} c_j f_i(x) f_j(x) d\mu\right)\left(\int \sum_{k\le n} c_k f_i(x) f_k(y) d\mu\right) \\
&= \sum_{i\le n} \lambda_i c_i^2 \\
&\le |\lambda_1| \sum_{i\le n} c_i^2 \\
&= |\lambda_1|.
\end{aligned}$$

□

The connection to quasirandomness comes from the observation that

$$t_{\mathrm{C}_4}(h) = \int h(x,y)h(y,x')h(x',y')h(y',x)d\mu_4 = \int h^4(x,x)d\mu$$

where $h^4(x,z) = \int h(x,y)h(y,x')h(x',y')h(y',z)d\mu_3(x',y,y')$, and therefore $t_{\mathrm{C}_4}(h)$ will be the sum of the eigenvalues of h^4. We call this function h^4 because it corresponds to the fourth power of the matrix corresponding to h, and we have another standard fact about the relationship between the eigenvalues of h and the powers of h:

Theorem 1.41. • *When $h : V^2 \to \mathbb{R}$ is symmetric, $(T_h)^2 = T_{h^2}$ where $h^2 : V^2 \to \mathbb{R}$ is a symmetric function given by*

$$h^2(x,y) = \int h(x,z)h(z,y)d\mu(z).$$

- *If $\lambda_1, \dots, \lambda_n$ and $f_1, \dots, f_n$ are the eigenvalues and eigenvectors associated with h then $\lambda_1^2, \dots, \lambda_n^2$ and $f_1, \dots, f_n$ are the eigenvalues and eigenvectors associated with h^2.*

Proof. For the first part, observe that for any $f : V \to \mathbb{R}$,

$$\begin{aligned}((T_h)^2 f)(x) &= \int h(x,y)\left[(T_h f)(y)\right] d\mu(y) \\ &= \int h(x,y)\left[\int h(y,z) f(z) d\mu(z)\right] d\mu(y) \\ &= \int \left(\int h(x,y)h(y,z) d\mu(y)\right) f(z) d\mu(z) \\ &= \int h^2(x,z) f(z) d\mu(z).\end{aligned}$$

For the second part, observe that for each eigenvalue f_i of h,

$$\begin{aligned}((T_h)^2 f_i)(x) &= \int h^2(x,y) f_i(y) d\mu(y) \\ &= \int \left(\int h(x,z)h(z,y) d\mu(z)\right) f_i(y) d\mu(y) \\ &= \int h(x,z)h(z,y) f_i(y) d\mu_2(y,z) \\ &= \int h(x,z)(T_h f_i)(z) d\mu(z) \\ &= \lambda_i \int h(x,z) f_i(z) d\mu(z) \\ &= \lambda_i (T_h f_i)(x) \\ &= \lambda_i^2 f_i(x).\end{aligned}$$

□

This allows us to prove that a graph is quasirandom exactly when all the eigenvalues other than the largest eigenvalue are small. (Where λ_1 is always the largest eigenvalue, and we always mean largest in the sense of absolute value.)

Theorem 1.42. *$G = (V, E)$ is δ-quasirandom if and only if*

$$(\lambda_1 - t_{\mathrm{K}_2}(G))^4 + \sum_{1<i\le n} \lambda_i^4 < \delta.$$

Proof. Observe that

$$\begin{aligned}t_{\mathrm{C}_4}(G) - t_{\mathrm{K}_2}(G)^4 &= \sum_{i\le n} \lambda_i^4 - t_{\mathrm{K}_2}(G)^4 \\ &= (\lambda_1^4 - t_{\mathrm{K}_2}(G)^4) + \sum_{1<i\le n} \lambda_i^4.\end{aligned}$$

□

We are usually not concerned with the exact quantity $(\lambda_1 - t_{K_2}(G))^4 + \sum_{1<i\leq n} \lambda_i^4$—the point is that G is δ-quasirandom exactly when two things happen: first, λ_1 is close to $t_{K_2}(G)$, and second, the remaining eigenvalues are all small.

1.9 An Example: The Payley Graph

It is possible to have graphs constructed in a completely deterministic way which are nonetheless quasirandom.

Definition 1.43. When q is a prime with $q \equiv 1 \mod 4$, the *Paley graph* Q_q is the graph with vertices $\{0, 1, \ldots, q-1\}$ where $\{x, y\}$ is an edge precisely when $x - y$ is a quadratic residue modulo q (that is, $x - y \neq 0$ and there is a k so that $k^2 \equiv x - y \mod q$).

It is a standard fact about quadratic residues that when q is a prime with $q \equiv 1 \mod 4$, -1 is a quadratic residue, and therefore $x - y$ is a quadratic residue exactly when $y - x$ is. (This is not the case when $q \equiv 3 \mod 4$, so we need $q \equiv 1 \mod 4$ to make Q_p symmetric.)

Half the elements of $\{1, \ldots, q-1\}$ are quadratic residues, so each element of Q_q has $\frac{q-1}{2}$ neighbors. In particular, this graph has density

$$t_{K_2}(Q_q) = \int \chi_{Q_q}(x, y) d\mu = \frac{1}{2} - \frac{1}{2q},$$

which approaches $1/2$ when q is large.

Theorem 1.44. *For every $\delta > 0$ there is a q_0 so that when $q > q_0$ is a prime with $q \equiv 1 \mod 4$, Q_q is δ-quasirandom.*

Proof. Let us write $Q_q = (V, E)$, so $V = \{1, \ldots, q-1\}$. If Q_q is going to be quasirandom, each pair x, y should have about $q/4$ neighbors in common. Recall that, for $x \in V$, E_x is the set of neighbors of x.

Rather than looking directly at $E_x \cap E_y$, we can look at those z which are neighbors to either both x and y or neighbors to neither

$$C(x, y) = \{z \mid z \in E_x \cap E_y \text{ or } z \in \overline{E_x} \cap \overline{E_y}\}.$$

That is, $C(x, y) = \{z \mid \chi_{E_x}(z) + \chi_{E_y}(z) \equiv 0 \mod 2\}$.

The product of two quadratic residues is also a quadratic residue, and the product of *non*-quadratic residues is also a quadratic residue. (To see this, recall that $Q_q^\times = \{1, \ldots, q-1\}$ is a group under multiplication, and the

quadratic residues form a subgroup of index 2, so the non-quadratic residues are the other conjugacy class.)

So, given vertices x and y, the value $\frac{z-x}{z-y}$ is a quadratic residue if either $z-x, z-y$ are both quadratic residues, or neither is—that is, $\frac{z-x}{z-y}$ is a qudratic residue exactly when $z \in C(x,y)$.

Assume x and y are distinct. There are $\frac{1}{2}(q-1)$ quadratic residues a. If $a=1$, we cannot have $\frac{z-x}{z-y}=a$. If $a \neq 1$, $\frac{z-x}{z-y}=a$ exactly when $z=\frac{x-ay}{1-a}$. So for each of the $\frac{1}{2}(q-1)-1$ choices of a to be a quadratic residue other than 1, there is a corresponding vertex which is either a neighbor of both x and y, or a neighbor of neither.

So $|C(x,y)| = \frac{q-1}{2} - 1$. Since $|E_x| = |E_y| = \frac{q-1}{2}$, we can conclude that $|E_x \cap E_y| = \frac{q-5}{4}$. Therefore

$$t_{C_4}(Q_q) = \int \left(\int \chi_E(x,z)\chi_E(y,z) d\mu(z) \right)^2 d\mu_2 = \int \frac{|E_x \cap E_y|^2}{q^2} d\mu_2 = \frac{1}{16} + O(\frac{1}{q}).$$

□

So—despite being completely deterministic in their construction—the Paley graphs are quasirandom.

A finer analysis, which we will not consider here, distinguishes the Paley graph from truly random graphs. For example, for infinitely many q, there is a set X with with $|X| \geq \log q \log\log q$ which is a clique in Q_q [79]: every pair of distinct elements of X is an edge. That means that $d_E(X,X)$ is very close to 1, but Theorem 1.19 says that in $\mathbf{R}_{1/2}$, whenever $|X|$ has size at least $C \log q$, $d_E(X,X)$ should be close to $1/2$.

1.10 Remarks

The theory of random graphs goes far beyond what we have touched on here. We have only considered one specific model of a random graph, in which the edges are generated independently and where the probability of an edge being present is some fixed value p independent of the size of the graph—that is, where the graphs generated are *dense*, so the number of edges is typically ϵn^2 for some $\epsilon > 0$. Many investigations of random graphs consider the case where the probability of an edge being present depends on n, so that the typical graph has, say, $Cn^{3/2}$ edges. Other models weaken or modify the assumption that edges are present independently. Bollobás' book [18] is a canonical reference, especially when supplemented by more recent books [7, 63, 92].

The study of quasirandom graphs was introduced by Chung, Graham, and Wilson in [30], and most of this chapter is drawn from that paper. Since then, a number of additional characterizations of quasirandomness have been investigated [120, 123, 136, 137, 142, 143, 144, 163]. The notion is strikingly robust, and the results in later chapters will shed some light on why.

Nonetheless, as the Payley graph illustrates, quasirandom graphs can still be distinguished from random ones. Stricter notions than quasirandomness, like the notion of jumbledness introduced by Thomason [156, 157] have also been investigated [104]. A theme we will see again later is that quasirandomness captures the "dense" part of randomness—the properties which only consider what happens in dense sets—but that stronger notions are needed if one wants to consider sets $X \subseteq V$ with $|X|$ much smaller than $|V|$.

There are others graphs H with the property that $t_H(G) \approx (t_{K_2}(G))^k$ (for suitable k) implies that G is quasirandom. Graphs with this property are called *forcing*. Which graphs (or, more generally, sets of graphs) are forcing has been studied extensively [37, 131, 144]; many graphs in addition to C_4 are known to be forcing, but it is not known in general which graphs are forcing. The question of which bipartite graphs are forcing is closely related to Sidorenko's conjecture [140], which concerns a generalization of the Cauchy-Schwarz calculations we were using above [82].

The quantity $\max_{X,Y \subseteq V} |d_E(X,Y)|X| \cdot |Y| - |E||$ (or the restriction where $X = Y$) is called the *discrepancy* of a graph. Erdős and Spencer [50] showed that there is a constant so that every graph with $|V|$ has discrepancy at least $cn^{3/2}$, and that random graphs have close to this minimum discrepancy. The behavior of the discrepancy and related quantities in random, quasirandom, and other graphs has also been further studied; Chazelle's book [21] gives an introduction to this subject.

The investigation of graphs in terms of their eigenvalues and eigenvectors is its own subject—spectral graph theory—with a standard reference by Chung [33].

Chapter 2

Ultraproducts

We develop an infinitary setting, the ultraproduct, which allows us to identify sequences of graphs with a single limit object, which will be an infinite (indeed, uncountably infinite) graph.

2.1 Convergent Subsequences

Keeping track of the ϵ's and δ's and worrying about things which approach 0 as the number of vertices approaches infinity gets increasingly messy as we move to more intricate arguments, so we'd like to move to a different, infinitary setting where these terms disappear.

Suppose we have a sequence of graphs $G_n = (V_n, E_n)$ where $|V_n|$ is approaching infinity. We would like to assemble the graphs G_n into a limiting graph—temporarily, we might call this hypothetical graph $\lim_{n\to\infty} G_n$—which should be an infinite graph which somehow captures the "eventual" properties of the graphs G_n.

For example, we should have the property that for any finite graph H,

$$\lim_{n\to\infty} t_H(G_n) = t_H(\lim_{n\to\infty} G_n).$$

This immediately points out a potential pitfall: the sequence $\langle t_H(G_n)\rangle_{n\in\mathbb{N}}$ need not be a convergent sequence of real numbers. For example, suppose that whenever n is even, G_n is $\mathrm{K}_{n/2,n/2}$, but when n is odd, G_n is the complete graph $(\{1,\ldots,n\}, \binom{\{1,\ldots,n\}}{n})$. Then the sequence $t_{\mathrm{C}_3}(G_n)$ is the sequence $\langle 1,0,1,0,1,0,\ldots\rangle$.

In order to obtain a limit, we will need to pass to a subsequence—we need to decide to either "concentrate" on the case where n is even, or the case where n is odd.

Before considering graphs further, let us recall some facts about the convergence of sequences of real numbers.

Theorem 2.1 (Bolzano–Weierstrass)**.** *If $\langle r_n \rangle_{n\in\mathbb{N}}$ is a sequence of real numbers in some bounded interval $[a, b]$ then there is a* convergent subsequence*: there is a sequence $n_1 < n_2 < \cdots$ so that $\langle r_{n_k} \rangle_{k\in\mathbb{N}}$ converges to some value in the interval $[a, b]$.*

It is worth reviewing the proof, which will be a model for later arguments.

Proof. For notational convenience, let us assume that $[a, b] = [0, 1]$. (We can obtain the general case from this by working with the sequence $s_n = \frac{r_n - a}{b-a}$, since $\langle s_{n_k} \rangle_{k\in\mathbb{N}}$ converges exactly when $\langle r_{n_k} \rangle_{k\in\mathbb{N}}$ does.)

We will construct the sequence $n_1 < n_2 < \cdots$ iteratively. Divide the interval $[0, 1]$ in half—$[0, 1] = [0, 1/2] \cup (1/2, 1]$. Consider the two sets

$$\{n \mid r_n \in [0, 1/2]\} \text{ and } \{n \mid r_n \in [1/2, 1]\}.$$

Every $n \in \mathbb{N}$ is in at least one of these, so at least one of these two sets must be infinite. We pick one of these two halves which is infinite, and choose n_1 so that r_{n_1} belongs to the chosen half.

Now consider the half we have chosen, and divide it in half again—for instance, if we chose $r_{n_1} \in [0, 1/2]$, next write $[0, 1/2] = [0, 1/4] \cup (1/4, 1/2]$. Again, one of the sets

$$\{n \mid r_n \in [0, 1/4]\} \text{ and } \{n \mid r_n \in [1/4, 1/2]\}$$

must be infinite because their union is $\{n \mid r_n \in [0, 1/2]\}$ which (since we only would have picked $r_1 \in [0, 1/2]$ if $\{n \mid r_n \in [0, 1/2]\}$ was infinite) must be infinite. Again, we pick a half which is infinite and chose $n_2 > n_1$ so that r_{n_2} belongs to the chosen half.

We repeat this process: after k stages, we have chosen $n_1 < n_2 < \cdots < n_k$ and an interval $[\frac{a}{2^k}, \frac{a+1}{2^k}]$ so that there are infinitely many n with $r_n \in [\frac{a}{2^k}, \frac{a+1}{2^k}]$. We divide the interval in half—

$$[\frac{a}{2^k}, \frac{a+1}{2^k}] = [\frac{2a}{2^{k+1}}, \frac{2a+1}{2^{k+1}}] \cup (\frac{2a+1}{2^{k+1}}, \frac{2a+2}{2^{k+1}}],$$

choose a subinterval containing infinitely many r_n, and choose $n_{k+1} > n_k$ so that $r_{n_{k+1}}$ belongs to the chosen subinterval.

We have chosen n_k so that for all $m \geq k$, $r_{n_m} \in [\frac{a}{2^k}, \frac{a+1}{2^k}]$. In particular, that means that when $m \geq k$, $|r_{n_m} - r_{n_k}| \leq 2^k$. That means that the sequence $\langle r_{n_k} \rangle_{k\in\mathbb{N}}$ converges. □

In general, it might be the case that both subintervals contain infinitely many r_n, and this might happen for many values of k, so we would have to make an arbitrary choice many times. This reflects the fact that $\langle r_n \rangle_{n \in \mathbb{N}}$ could have many different convergent subsequences which converge to many different values, and there is not generally a "best" convergent subsequence.

The idea of a convergent subsequence is useful enough to warrant a definition of its own, and it will be convenient for us to view the indices as an infinite set.

Definition 2.2. Let $K \subseteq \mathbb{N}$ be an infinite set. We say $\langle r_n \rangle_{n \in \mathbb{N}}$ *converges to* r *on* K, written

$$\lim_{n \to K} r_n = r$$

if, for every $\epsilon > 0$,

$$\{n \in K \mid |r - r_n| \geq \epsilon\}$$

is finite. We say $\langle r_n \rangle$ *converges on* K, or $\lim_{n \to K} r_n$ *exists*, if there is some r so that $\lim_{n \to K} r_n = r$.

This definition generalizes ordinary convergence, which is the case where $K = \mathbb{N}$, by "concentrating" on a particular set K—deciding that only elements of K matter, while ignoring $\mathbb{N} \backslash K$. Every subsequence of a convergent sequence also converges—that is, if $J \subseteq K$ is infinite and $\lim_{n \to K} r_n = r$ then also $\lim_{n \to J} r_n = r$—so by "concentrating" on a smaller set K, we make more sequences converge.

Consider the sequence $\langle r_n \rangle_{n \in \mathbb{N}}$ where $r_n = n \mod 2$—that is, the sequence $\langle 1, 0, 1, 0, 1, 0, \ldots \rangle$. Then if we choose a K so that $\lim_{n \to K} r_n$ converges, K cannot contain both infinitely many even numbers and infinitely many odd numbers—either K consists of infinitely many even numbers and finitely many odd numbers, in which case $\lim_{n \to K} r_n = 0$, or K consists of infinitely many odd numbers and finitely many even numbers and $\lim_{n \to K} r_n = 1$.

If we next consider the sequence $\langle s_n \rangle_{n \in \mathbb{N}}$ which repeats in the pattern $\langle 0, 2, 1, 3, 0, 2, 1, 3, 0, 2, 1, 3, \ldots \rangle$, we have no reason to think that $\lim_{n \to K} s_n$ also exists. However we can apply the argument of Bolzano–Weierstrass again to obtain a $J \subseteq K$ so that both $\lim_{n \to J} r_n$ and $\lim_{n \to J} s_n$ exist. Furthermore, our choice of $\lim_{n \to K} r_n$ constrains the possible values for $\lim_{n \to J} s_n$—if $\lim_{n \to K} r_n = 0$ then, since $J \subseteq K$, $\lim_{n \to J} s_n \in \{2, 3\}$.

Once $\lim_{n \to J} r_n$ and $\lim_{n \to J} s_n$ both exist, we can conclude that other related limits exist—for instance,

$$\lim_{n \to J} (r_n + s_n) = (\lim_{n \to J} r_n) + (\lim_{n \in J} s_n).$$

Suppose now that we have many sequences—say, for each $i \in \mathbb{N}$ we have a sequence $\langle r^i_n \rangle_{n\in\mathbb{N}}$. Then we can ask for all these sequences to converge simultaneously: we can ask for a single set K so that, for each i, $\langle r^i_n \rangle_{n\in\mathbb{N}}$ converges on K. If there were only finitely many sequences we could find a simultaneously convergent subsequence by simply iterating Bolzano–Weierstrass finitely many times: each time we use it, Bolzano–Weierstrass further thins out our subsequence, but ensure that one more sequence converges.

With infinitely many sequences, however, this approach no longer works: each time we apply Bolzano–Weierstrass, we might lose elements from our set, and if we're not careful, after infinitely many applications there might be no elements left.

Theorem 2.3. *Suppose that, for each $i \in \mathbb{N}$, $\langle r^i_n \rangle_{n\in\mathbb{N}}$ is a sequence of real numbers in a bounded interval $[a_i, b_i]$. Then there is a set K and values $r^i \in [a_i, b_i]$ so that, for every i, $\lim_{n\to K} r^i_n = r^i$.*

Proof. The idea is to repeatedly apply Bolzano–Weierstrass, but remember to save an element each time.

Let $J_1 = \mathbb{N}$ and let $n_1 = 1$, the least element of J_1. By Bolzano–Weierstrass, we can choose $J_2 \subseteq J_1$ so that $\langle r^1_n \rangle_{n\in\mathbb{N}}$ converges on J_2, and let n_2 be the smallest element of J_2 larger than n_1.

In general, at the k-th stage we have a set J_k so that, for each $i < k$, $\langle r^i_n \rangle_{n\in\mathbb{N}}$ converges on J_k, and we have chosen $n_1 < \cdots < n_k$. Choose $J_{k+1} \subseteq J_k$ so that $\langle r^k_n \rangle_{n\in\mathbb{N}}$ converges on J_{k+1} and let n_{k+1} be the smallest element of J_{k+1} larger than n_k.

Let $K = \{n_1 < n_2 < \cdots\}$ and consider any $\langle r^k_n \rangle_{n\in\mathbb{N}}$. For each $i \geq k$, $n_i \in J_{k+1}$, and therefore $\{n_k < n_{k+1} < \cdots\} \subseteq J_{k+1}$. Since $\langle r^k_n \rangle_{n\in\mathbb{N}}$ converges on J_{k+1}, $\langle r^k_n \rangle_{n\in\mathbb{N}}$ also converges on $\{n_k < n_{k+1} < \cdots\}$, and therefore $\langle r^k_n \rangle_{n\in\mathbb{N}}$ converges on K (which differs only on the finitely many elements $\{n_1 < \cdots < n_{k-1}\}$).

So, for every k, $\langle r^k_n \rangle_{n\in\mathbb{N}}$ converges on K. □

Once we have arranged for these sequences to converge simultaneously, we can work with all sorts of combinations. For instance, we have

$$\lim_{n\to K} \sum_i \frac{1}{2^i \max\{|a_i|, |b_i|\}} r^i_n = \sum_i \frac{1}{2^i \max\{|a_i|, |b_i|\}} \lim_{n\to K} r^i_n$$

(having chosen the denominator precisely to ensure that the sum is finite).

2.2 Ultralimits

We would like to go a step further: we would like to choose an infinite set K on which *every single sequence* converges simultaneously. This is too much to ask, so we will have to weaken our notion of what it means for sequences to converge simultaneously.

We would at least like to choose, for every sequence $\langle r_n \rangle_{n \in \mathbb{N}}$, a limiting value $\lim_? r_n$, which we temporarily denote with a question mark since we have not identified this prospective notion of convergence. Our choice of $\lim_? r_n$ should certainly be a plausible limit of the sequence—that is, there should be some set K witnessing the convergence, so that $\lim_? r_n = \lim_{n \to K} r_n$.

We need the choice of limiting values to be compatible—for instance, we should have

$$\lim_?(r_n + s_n) = (\lim_? r_n) + (\lim_? s_n).$$

Recall the sequences $\langle 1, 0, 1, 0, \ldots \rangle$ and $\langle 0, 2, 1, 3, 0, 2, 1, 3, \ldots \rangle$ from above. If we decide that $\lim_?\langle 1, 0, 2, 0, \ldots \rangle = 0$, we should commit to having

$$\lim_?\langle 0, 2, 1, 3, 0, 2, 1, 3, \ldots \rangle \in \{2, 3\}.$$

That is, the decision that $\lim_?\langle 0, 1, 0, 1, \ldots \rangle = 0$ means that we must be "concentrating" on the even indices, and our choices for all other sequences should reflect that the even indices matter while the odd ones do not.

If we have chosen J so that $\lim_? r_n = \lim_{n \to J} r_n$ and K so that $\lim_? s_n = \lim_{n \to K} s_n$, this amounts to requiring that $J \cap K$ also be infinite, so that we can take $\lim_?(r_n + s_n) = \lim_{n \to J \cap K}(r_n + s_n)$.

This suggests that we can work with the collection of witnessing sets: we will let $\mathcal{F}$ be some collection of infinite sets, and we can define $\lim_? r_n = \lim_{n \to K} r_n$ for some set $K \in \mathcal{F}$ such that $\langle r_n \rangle_{n \in \mathbb{N}}$ converges on K. In order for sequences to converge simultaneously, we need to require that when $J \in \mathcal{F}$ and $K \in \mathcal{F}$, also $J \cap K \in \mathcal{F}$.

This quickly leads us to the notion of a free filter.

Definition 2.4. A collection $\mathcal{F}$ of subsets of $\mathbb{N}$ is a *free filter* if:

- whenever $J, K \in \mathcal{F}$, also $J \cap K \in \mathcal{F}$,
- whenever $J \in \mathcal{F}$ and $J \subseteq K$, also $K \in \mathcal{F}$,
- $\emptyset \notin \mathcal{F}$, and

- if $\mathbb{N} \setminus K$ is finite (that is, if K is *cofinite*) then $K \in \mathcal{F}$.

These conditions imply that every set in $\mathcal{F}$ is infinite. (We will not need the slightly more general notion of a *filter*, which weakens the last condition to merely $\mathbb{N} \in \mathcal{F}$.)

While the definition of a free filter only considers intersections of two sets, induction says that free filters contain intersections of finitely many sets:

Lemma 2.5. *If $\mathcal{F}$ is a free filter and $K_1, \ldots, K_n \in \mathcal{F}$ then $K_1 \cap \cdots \cap K_n \in \mathcal{F}$.*

Proof. By induction on n. For $n = 1$ this is tautological and for $n = 2$ this is part of the definition of a free filter.

Suppose the claim holds for n and $K_1, \ldots, K_n, K_{n+1} \in \mathcal{F}$. Then, by the inductive hypothesis, $K_1 \cap \cdots \cap K_n \in \mathcal{F}$, and by the definition of a free filter, also $(K_1 \cap \cdots \cap K_n) \cap K_{n+1} \in \mathcal{F}$. □

Definition 2.6. If $\mathcal{F}$ is a free filter and $\langle r_n \rangle_{n \in \mathbb{N}}$ is a sequence, we can define $\lim_{n \to \mathcal{F}} r_n = r$ if, for each $\epsilon > 0$, $\{n \mid |r_n - r| < \epsilon\} \in \mathcal{F}$.

When K is an infinite set, the collection of all sets J such that $K \setminus J$ is finite forms a free filter $\mathcal{F}_K$, and $\lim_{n \to \mathcal{F}_K} r_n = \lim_{n \to K} r_n$ (where one side exists exactly when the other does) for any sequence $\langle r_n \rangle_{n \in \mathbb{N}}$.

So convergence on a free filter is a generalization of the idea of convergence of on a set. We still need a free filter with the property that *every* sequence converges. For that, we need one more condition.

Definition 2.7. An *ultrafilter* is a free filter $\mathcal{U}$ such that, for every set $K \subseteq \mathbb{N}$, either $K \in \mathcal{U}$ or $(\mathbb{N} \setminus K) \in \mathcal{U}$. *

We sometimes refer to this additional property as the "ultra" property of an ultrafilter.

The properties of ultrafilters combine to give the useful property that whenever we have a finite union of sets in $\mathcal{U}$, it must be because one of these sets is in $\mathcal{U}$.

Lemma 2.8. *If $\mathcal{U}$ is an ultrafilter and $K_1 \cup K_2 \cup \cdots \cup K_n \in \mathcal{U}$ then there is an $i \leq n$ so that $K_i \in \mathcal{U}$.*

Proof. Suppose not, so $K_i \notin \mathcal{U}$ for all $i \leq n$. By the ultra property, each $(\mathbb{N} \setminus K_i) \in \mathcal{U}$. So also the intersection $\bigcap_{i \leq n} (\mathbb{N} \setminus K_i)) \in \mathcal{U}$, which is a contradiction, since this intersection is the empty set.

□

*Technically we have defined a *nonprincipal ultrafilter*, but since this is the main case, and the only case we are interested in, we will omit the word "nonprincipal".

Ultrafilters have exactly the property we want: they make every sequence converge to a unique value.

Theorem 2.9. *For every ultrafilter $\mathcal{U}$ and every sequence $\langle r_n \rangle_{n\in\mathbb{N}}$ in a bounded interval $[a, b]$, there is a unique r so that $\lim_{n\to\mathcal{U}} r_n = r$.*

Proof. The proof is similar to the proof of Bolzano–Weierstrass. Each time we split the interval in half, it could be that there are infinitely many r_n in both halves, and therefore we have to choose which half to continue in. In the proof of Bolzano–Weierstrass, we could choose arbitrarily because we only cared about showing that there was some convergent subsequence.

When we have an ultrafilter, it forces a choice on us—exactly one of the two halves is consistent with the ultrafilter. Indeed this is precisely what an ultrafilter does: whenever $\mathbb{N} = J \cup K$, an ultrafilter tells us to concentrate on exactly one of J or K.

Again, assume that $[a, b] = [0, 1]$. We will construct a sequence $\langle s_n \rangle_{n\in\mathbb{N}}$ which converges in the usual sense to a real number $r = \lim_{n\to\infty} s_n$ while ensuring that, for each $\epsilon > 0$, $\{n \mid |r_n - r| < \epsilon\} \in \mathcal{U}$.

Divide the interval $[0, 1]$ in half, so $[0, 1] = [0, 1/2] \cup (1/2, 1]$. Consider $\{n \mid r_n \in [0, 1/2]\}$; if this set is in $\mathcal{U}$ then we take $s_1 = 1/4$ and promise that, for all $n > 1$, $s_n \in [0, 1/2]$. In particular, that will mean $r \in [0, 1/2]$, and so

$$\{n \mid |r_n - r| \leq 1/2\} \subseteq \{n \mid r_n \in [0, 1/2]\} \in \mathcal{U}.$$

Otherwise, by the "ultra" property of the ultrafilter, $\mathbb{N} \setminus \{n \mid r_n \in [0, 1/2]\} = \{n \mid r_n \in (1/2, 1]\}$ must belong to $\mathcal{U}$, and we will take $s_1 = 3/4$ and promise that, for all $n > 1$, $s_n \in (1/2, 1]$. In particular, that will mean $r \in [1/2, 1]$, and so

$$\{n \mid |r_n - r| \leq 1/2\} \subseteq \{n \mid r_n \in [1/2, 1]\} \in \mathcal{U}.$$

Next we split the interval in half again. For instance, if $s_1 = 1/4$ then we split $[0, 1/2] = [0, 1/4] \cup (1/4, 1/2]$. If $\{n \mid r_n \in [0, 1/4]\} \in \mathcal{U}$ then we take $s_2 = 1/8$ and promise that, for all $n > 2$, $s_n \in [0, 1/4]$, which will mean that $r \in [0, 1/4]$, and therefore that

$$\{n \mid |r_n - r| \leq 1/4\} \setminus \{n \mid r_n \in [0, 1/4]\} \in \mathcal{U}.$$

Otherwise $\{n \mid r_n \in (1/4, 1]\} \in \mathcal{U}$, so also

$$\{n \mid r_n \in (1/4, 1]\} \cap \{n \mid r_n \in [0, 1/2]\} = \{n \mid r_n \in (1/4, 1/2]\} \in \mathcal{U}.$$

In this case we take $s_2 = 3/8$ and promise that, for all $n > 2$, $s_n \in (1/4, 1/2]$. This means that $r \in [1/4, 1/2]$, and therefore that

$$\{n \mid |r_n - r| \leq 1/4\} \subseteq \{n \mid r_n \in (1/4, 1/2]\} \in \mathcal{U}.$$

In general, after k stages we have chosen $s_1, \ldots, s_k$ and an interval $I = [\frac{a}{2^k}, \frac{a+1}{2^k}]$ so that:

- when $i < j$, $|s_i - s_j| \leq \frac{1}{2^i}$,
- $s_k \in I$,
- $\{n \mid r_n \in I\} \in \mathcal{U}$.

We divide this interval in half, as

$$[\frac{a}{2^k}, \frac{a+1}{2^k}] = [\frac{2a}{2^{k+1}}, \frac{2a+1}{2^{k+1}}] \cup (\frac{2a+1}{2^{k+1}}, \frac{2a+2}{2^{k+1}}]$$

and observe that exactly one of the sets

$$\{n \mid r_n \in [\frac{2a}{2^{k+1}}, \frac{2a+1}{2^{k+1}}]\} \text{ and } \{n \mid r_n \in (\frac{2a+1}{2^{k+1}}, \frac{2a+2}{2^{k+1}}]\}$$

belongs to $\mathcal{U}$. We choose s_{k+1} to be the midpoint of the corresponding interval.

The sequence $\langle s_n \rangle_{n \in \mathbb{N}}$ is certainly convergent since, whenever $i < j$, we have $|s_i - s_j| \leq \frac{1}{2^i}$. Letting $r = \lim_{n \to \infty} \langle s_n \rangle$, for any $\epsilon > 0$ we can choose k large enough so that $\frac{1}{2^k} < \epsilon$, note that $|s_k - r| \leq \frac{1}{2^k}$, so

$$\{n \mid |r_n - r| < \epsilon\} \subseteq \{n \mid |r_n - r| \leq \frac{1}{2^{k+1}}\} \subseteq \{n \mid |r_n - s_k| \leq \frac{1}{2^k}\} \in \mathcal{U}.$$

□

Definition 2.10. When $\mathcal{U}$ is an ultrafilter and $\langle r_n \rangle_{n \in \mathbb{N}}$ is a sequence of real numbers in $[a, b]$, we call $\lim_{n \to \mathcal{U}} r_n$ the *ultralimit* (with respect to $\mathcal{U}$) of $\langle r_n \rangle_{n \in \mathbb{N}}$.

Once we have an ultrafilter, we no longer need to worry about convergence issues: every bounded sequence converges with respect to that ultrafilter.

2.3 Ultrafilters

We defined ultrafilters as collections of sets satisfying certain properties, and then derived the striking conclusion that they make every sequence converge. We should worry that this is so strong precisely because no such collections exist. Free filters certainly exist: let $\mathcal{F}_0$ be the collection of all cofinite sets. This is called the *Fréchet filter*, and we can see that it has the properties of a free filter.

- If J and K are cofinite, so $\mathbb{N} \setminus J$ and $\mathbb{N} \setminus K$ are both finite, then $\mathbb{N} \setminus (J \cap K) = (\mathbb{N} \setminus J) \cup (\mathbb{N} \setminus K)$ is also finite, and therefore $J \cap K \in \mathcal{F}_0$.
- If J is cofinite and $J \subseteq K$ then $\mathbb{N} \setminus K \subseteq \mathbb{N} \setminus J$ is also finite, so $K \in \mathcal{F}_0$.

Indeed, $\mathcal{F}_0$ is the smallest possible free filter. We will try to expand $\mathcal{F}_0$ into an ultrafilter. If we have any free filter $\mathcal{F}$ which is *not* an ultrafilter, so there is some set with both $K \notin \mathcal{F}$ and $\mathbb{N} \setminus K \notin \mathcal{F}$, then we can at least extend the free filter to contain one of these sets.

Lemma 2.11. *Let $\mathcal{F}$ be a free filter such that $\mathbb{N} \setminus K \notin \mathcal{F}$. Then there is a free filter $\mathcal{F}' \supseteq \mathcal{F} \cup \{K\}$.*

Proof. There is a unique choice of a minimal $\mathcal{F}' \supseteq \mathcal{F} \cup \{K\}$—$\mathcal{F}'$ needs to be the collection of all I such that, for some $J \in \mathcal{F}$, we have $K \cap J \subseteq I$. Certainly $\mathcal{F}' \supseteq \mathcal{F}$ since for any $J \in \mathcal{F}$, $J \supseteq K \cap J$ so $J \in \mathcal{F}'$. Also $\mathbb{N} \in \mathcal{F}$ so $K \supseteq K \cap \mathbb{N}$, so $K \in \mathcal{F}'$. The closure properties of a free filter force $\mathcal{F}'$ to contain all these sets, so we just need to prove that (under the assumption that $\mathbb{N} \setminus K \notin \mathcal{F}$) this is a free filter.

Suppose $I_0, I_1 \in \mathcal{F}'$, so $I_0 \supseteq K \cap J_0$ an $I_1 \supseteq K \cap J_1$ with $J_0, J_1 \in \mathcal{F}$. Then $I_0 \cap I_1 \supseteq (K \cap J_0) \cap (K \cap J_1) = K \cap (J_0 \cap J_1)$ and since $J_0 \cap J_1 \in \mathcal{F}$, also $I_0 \cap I_1 \in \mathcal{F}'$.

If $I_0 \in \mathcal{F}'$ and $I_0 \subseteq I_1$ then $I_1 \supseteq I_0 \supseteq K \cap J_0$ so $I_1 \in \mathcal{F}'$.

If $\emptyset \in \mathcal{F}'$ then $\emptyset \supseteq K \cap J$ for some $J \in \mathcal{F}$. But then $J \subseteq (\mathbb{N} \setminus K)$, so $\mathbb{N} \setminus K \notin \mathcal{F}$, which contradicts our assumption. Since every cofinite set is in $\mathcal{F}$ and $\mathcal{F} \subseteq \mathcal{F}'$, every cofinite set is in $\mathcal{F}'$.

Since $\mathcal{F}' \supseteq \mathcal{F}$, it is the desired free filter. □

So we can now imagine how we would obtain an ultrafilter: we begin with the Fréchet filter and successively extend this filter over and over again using Lemma 2.11, once for each subset of $\mathbb{N}$, until we obtain an ultrafilter. At each step we can consider one set $K \subseteq \mathbb{N}$ and extend our filter, if necessary, to ensure that it either contains K or $\mathbb{N} \setminus K$.

This process is very non-canonical. For example, if we begin with $\mathcal{F}_0$, we might decide to add either the even numbers or the odd numbers to give the next filter. Both choices are reasonable—they give us valid free filters—but they lead to different ultrafilters. So this approach suggests that if we obtain an ultrafilter, it will only be because there are many ultrafilters and we happen to have found one of them.

If there were countably many subsets of $\mathbb{N}$, this approach would simply work; the problem is that there are uncountably many subsets of $\mathbb{N}$, so in order to have enough steps to consider every subset of $\mathbb{N}$, we need an uncountably long construction.

In particular, we will encounter the situation where we have long "chains" of filters—where we have chosen a long sequence of filters $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ where this sequence is infinitely long and we need to combine these into a single filter.

Lemma 2.12. *Let $(L, \leq)$ be a non-empty ordered set, and suppose that, for each $i \in L$, $\mathcal{F}_i$ is a free filter so that when $i \leq j$, $\mathcal{F}_i \subseteq \mathcal{F}_j$. Then there is a free filter $\mathcal{F}$ such that, for all $i \in L$, $\mathcal{F}_i \subseteq \mathcal{F}$.*

Proof. We simply take $\mathcal{F} = \bigcup_{i \in L} \mathcal{F}_i$.

Suppose $J, K \in \mathcal{F}$. Then there are i, j so that $J \in \mathcal{F}_i$ and $K \in \mathcal{F}_j$. Either $i \leq j$ or $j \leq i$; without loss of generality, assume $j \leq i$, so $\mathcal{F}_j \subseteq \mathcal{F}_i$, so $K \in \mathcal{F}_i$. Then $J \cap K \in \mathcal{F}_i$, so also $J \cap K \in \mathcal{F}_i$.

If $J \in \mathcal{F}$ and $J \subseteq K$ then $J \in \mathcal{F}_i$, so $K \in \mathcal{F}_i$, so $K \in \mathcal{F}$.

Since $\emptyset \notin \mathcal{F}_i$ for all i, $\emptyset \notin \mathcal{F}$.

If K is cofinite then $K \in \mathcal{F}_i$ for any i, so $K \in \mathcal{F}$. □

There are different ways to put these lemmata together to construct an ultrafilter. They all depend, in an essential way, on some use of the Axiom of Choice, and differ mostly in how they phrase the use of the Axiom of Choice. One way is to place the subsets of $\mathbb{N}$ in an order so that we can use transfinite recursion to construct an ultrafilter—that is, arrange a sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ so that, at stage i, we ensure that either $K_i \in \mathcal{F}_{i+1}$ or $\mathbb{N} \setminus K_i \in \mathcal{F}_{i+1}$. Doing this requires some use of Axiom of Choice (the version usually called the "well-ordering principle") to obtain an ordering of the subsets of $\mathbb{N}$ on which transfinite recursion works.

A different form of the Axiom of Choice, Zorn's Lemma, suggests a more abstract approach.

Theorem 2.13 (Zorn's Lemma)**.** *Let $\mathcal{P}$ be a set partially ordered by $\subseteq$ so that whenever $(L, \leq)$ is a totally ordered set and $f : I \to \mathcal{P}$ is an order preserving function (so $i \leq j$ implies $f(i) \subseteq f(j)$), there is a $P \in \mathcal{P}$ such that, for every $i \in I$, $f(i) \subseteq P$. Then there is a maximal element of $\mathcal{P}$—a $P \in \mathcal{P}$ such that, for any $Q \in \mathcal{P}$ such that $P \leq Q$, $P = Q$.*

Zorn's Lemma is exactly suited to the situation we are in. One interpretation of Lemma 2.11 is that it says that ultrafilters are precisely the maximal free filters—if $\mathcal{F}$ is a maximal free filter and $\mathbb{N} \setminus K \notin \mathcal{F}$ then there is a $\mathcal{F}' \supseteq \mathcal{F} \cup \{K\}$ which, by maximality, must already be $\mathcal{F}$, so $K \in \mathcal{F}$.

But what if there were no maximal free filters? One way we might fail to have a maximal free filter is if there were chains that got larger and larger—some sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ (potentially infinitely or even uncountably infinitely long) of larger and larger free filters that never ends and never concludes with some object above all of them. Zorn's Lemma says that this is the only obstacle to finding a maximal object: if every sequence has a top then there must actually be a maximal element.

Theorem 2.14. *There is an ultrafilter.*

Here it really matters that we mean a nonprincipal ultrafilter (that is, we require that all sets in the ultrafilter be infinite). The existence of principal ultrafilters is fairly trivial—the collection of all sets containing 3 is one.

Proof. Let $\mathcal{P}$ be the set of free filters, ordered by $\subseteq$. Using Lemma 2.12 and Zorn's Lemma, there must be some maximal free filter $\mathcal{F}$. As shown above, $\mathcal{F}$ is an ultrafilter.

□

The use of the Axiom of Choice (in the form of Zorn's Lemma) further suggests that the choice of an ultrafilter is non-canonical. Indeed, there is no "best" or "unique" ultrafilter, and no way to construct one concretely. One way to say this formally is to observe that in the axioms of ZF—that is, set theory without the Axiom of Choice—one cannot prove that an ultrafilter exists [17, 54].*

*It is not quite true that having an ultrafilter *requires* the Axiom of Choice, since the existence of an ultrafilter follows from weaker axioms, though ones that still go beyond ZF. In this sense the existence of an ultrafilter is essentially a *weak* form of the Axiom of Choice. The relationship between various many Axiom of Choice-like principles has been extensively studied [88].

One might think that, since the existence of ultrafilters depends on (a weak form of) the Axiom of Choice, the proofs we will give using ultrafilters depend on the Axiom of Choice as well. In fact, this will not be true: despite using ultrafilters in our construction, we will ultimately give proofs about finite graphs which do not require the use of the Axiom of Choice. This is a subtle point which we will return to later: the use of ultrafilters, in the applications we are interested in, can be systematically removed by purely syntactic manipulations on the proofs.

2.4 Products

We now return to our original concern: how to take the limit of sequences of graphs.

We will suppose we have a sequence of graphs $G_n = (V_n, E_n)$ where $|V_n|$ is approaching infinity, and we set out to define a graph which serves as a limit of the sequence G_n. For concreteness, let us take an example: G_n will be the bipartite graph $\mathrm{K}_{n,n}$. Specifically, we will take V_n to be the set of integers $\{1, 2, \ldots, 2n\}$ and E_n will consist of all pairs $\{i, j\}$ where exactly one element is even, so the two parts of G_n are the even vertices and the odd vertices.

A vertex in our limit graph will be based on a sequence of vertices from the graphs G_n. As a first attempt, we will consider the product of the sets of vertices: the set of vertices will be $\prod_{n\in\mathbb{N}} V_n$, which consists of sequences $\langle v_n\rangle_{n\in\mathbb{N}}$ such that, for each n, $v_n \in V_n$. (Recall that we required that graphs have at least one vertex, so this product is always non-empty—for each n there is at least one possible choice of v_n.)

When we consider two sequences $\langle v_n\rangle_{n\in\mathbb{N}}$ and $\langle w_n\rangle_{n\in\mathbb{N}}$, we need to decide whether to put an edge between them. The sequence $\langle v_n\rangle_{n\in\mathbb{N}}$ could be a mix of even and odd values, as could $\langle w_n\rangle_{n\in\mathbb{N}}$—it could be that, for some value of n, v_n and w_n have the same parity, and therefore $\{v_n, w_n\} \notin E_n$, while for other values of n, v_n and w_n have opposite parity, and therefore $\{v_n, w_n\} \in E_n$.

This leads to two sets which partition $\mathbb{N}$:

$$\mathbb{N} = \{n \mid \{v_n, w_n\} \in E_n\} \cup \{n \mid \{v_n, w_n\} \notin E_n\},$$

and this perhaps makes clear where our ultrafilter will come in. Exactly one of these two sets belongs to the ultrafilter, so we will place an edge between $\langle v_n\rangle_{n\in\mathbb{N}}$ and $\langle w_n\rangle_{n\in\mathbb{N}}$ exactly if $\{n \mid \{v_n, w_n\} \in E_n\}$ belongs to $\mathcal{U}$. That is, some values of n think there should be an edge while others thing

there should not be, and we go with the choice of "most" n, where "most" is determined by the ultrafilter. In formal notation, we define:

$$[E_n]_{\mathcal{U}} = \{\{\langle v_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\} \mid \{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}.$$

Note that it is very important that our choice of "most" n is "coherent". Suppose we take three elements of $\prod_{n\in\mathbb{N}} V_n$, $\langle u_n \rangle_{n\in\mathbb{N}}$, $\langle v_n \rangle_{n\in\mathbb{N}}$, and $\langle w_n \rangle_{n\in\mathbb{N}}$. It could be that, for some values of n, $\{u_n, v_n\} \in E_n$, and also for some values of n, $\{u_n, w_n\} \in E_n$, and for yet other values of n, $\{v_n, w_n\} \in E_n$. Any two of these possibilities are consistent, but there are no values of n for which all three happen at once. We want to make sure that we do not add edges between all three sequences in E—there are no triangles in any (V_n, E_n), so there should be no triangles in $(\prod_{n\in\mathbb{N}} V_n, [E_n]_{\mathcal{U}})$.

But our decision about which pairs to put in $[E_n]_{\mathcal{U}}$ is made pair by pair—we place $\{\langle u_n \rangle_{n\in\mathbb{N}}, \langle v_n \rangle_{n\in\mathbb{N}}\} \in [E_n]_{\mathcal{U}}$ if $\{n \mid \{u_n, v_n\} \in E_n\} \in \mathcal{U}$, and, separately, we place $\{\langle v_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\} \in [E_n]_{\mathcal{U}}$ if $\{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}$, and similarly for $\{\langle u_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\} \in [E_n]_{\mathcal{U}}$. Yet these three decisions cannot be independent, because we cannot place all three pairs in $[E_n]_{\mathcal{U}}$.

Coordinating all the different choices is the job of the ultrafilter $\mathcal{U}$. (This is the *simultaneous* part of asking that all sequences converge simultaneously.) Suppose all three pairs did somehow end up in E; that would mean

$$\{n \mid \{u_n, v_n\} \in E_n\}, \{n \mid \{u_n, w_n\} \in E_n\}, \{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}.$$

But since the intersection of elements of the ultrafilter is also in the ultrafilter,

$$\{n \mid \{u_n, v_n\} \in E_n\} \cap \{n \mid \{u_n, w_n\} \in E_n\} \cap \{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}.$$

That is impossible, because there are no such n and $\emptyset \notin \mathcal{U}$, so this cannot happen.

At least in this (very simple) case, the object we get is a plausible limit: the limit of larger and larger finite bipartite graphs is an infinite bipartite graph.

Theorem 2.15. *There is a partition $\prod_{n\in\mathbb{N}} V_n = V_1 \cup V_2$ such that if $v, w \in V$, $\{v, w\} \in [E_n]_{\mathcal{U}}$ if and only if v and w are in different parts.*

Proof. Let us take V_1 to consist of those sequences $\langle v_n \rangle_{n\in\mathbb{N}}$ such that $\{n \mid v_n \text{ is even}\} \in \mathcal{U}$, and V_2 to consist of all other sequences. For any sequence $\langle v_n \rangle_{n\in\mathbb{N}}$, we have

$$\mathbb{N} = \{n \mid v_n \text{ is even}\} \cup \{n \mid v_n \text{ is odd}\},$$

so if $\langle v_n \rangle_{n\in\mathbb{N}} \notin V_1$, we must have $\{n \mid v_n \text{ is odd}\} \in \mathcal{U}$. So V_1 is the sequences which are "mostly even" and V_2 is the sequences which are "mostly odd".

Consider a pair with $\langle v_n \rangle_{n\in\mathbb{N}} \in V_1$ and $\langle w_n \rangle_{n\in\mathbb{N}} \in V_2$. Then

$$\{n \mid \{v_n, w_n\} \in E_n\} \supseteq \{n \mid v_n \text{ is even}\} \cap \{n \mid w_n \text{ is odd}\} \in \mathcal{U},$$

so $\{\langle v_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\} \in [E_n]_{\mathcal{U}}$.

Similarly, if $\langle v_n \rangle_{n\in\mathbb{N}} \in V_1$ and $\langle w_n \rangle_{n\in\mathbb{N}} \in V_1$ (the case where both are in V_2 is similar), then

$$\{n \mid \{v_n, w_n\} \notin E_n\} \supseteq \{n \mid v_n \text{ is even}\} \cap \{n \mid w_n \text{ is even}\} \in \mathcal{U},$$

so $\{\langle v_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\} \notin [E_n]_{\mathcal{U}}$. □

Let us consider a second example, which will highlight a problem with simply using the product. Let G_n be the complete graph on n vertices: $V_n = \{1, 2, \ldots, n\}$ and $E_n = \binom{V_n}{2}$. The limit should be a complete infinite graph.

Suppose we take the same definition: the vertices are $\prod_{n\in\mathbb{N}} V_n$, the set of all sequences $\langle v_n \rangle_{n\in\mathbb{N}}$ such that $v_n \in V_n$ for all n, and $[E_n]_{\mathcal{U}}$ consists of all pairs $\{\langle v_n \rangle_{n\in\mathbb{N}}, \langle w_n \rangle_{n\in\mathbb{N}}\}$ such that $\{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}$.

Consider two elements of $\prod_{n\in\mathbb{N}} V_n$ which differ at only one index—say, the sequence $\langle v_n \rangle_{n\in\mathbb{N}}$ where $v_n = 1$ for all n, and the sequence $\langle w_n \rangle_{n\in\mathbb{N}}$ where $w_2 = 2$ but $w_n = 1$ for all $n \neq 2$. Then $\{n \mid \{v_n, w_n\} \in E_n\} = \{2\}$, which is a finite set and therefore not in $\mathcal{U}$.

If we believe—and we do—that the limit of complete graphs should be a complete graph, this is a problem, because we have found two sequences in $\prod_{n\in\mathbb{N}} V_n$ which are not adjacent.

The problem is that these two sequences are too similar—for "most" values of n we have $v_n = w_n$. We should accept majority rule for $=$ as well as for E: since most indices think these sequences are equal, we should decide they actually are equal.

Definition 2.16. We write $\langle v_n \rangle_{n\in\mathbb{N}} \sim_{\mathcal{U}} \langle w_n \rangle_{n\in\mathbb{N}}$ if $\{n \mid v_n = w_n\} \in \mathcal{U}$.

Theorem 2.17. *$\sim_{\mathcal{U}}$ is an equivalence relation.*

Proof. Reflexivity holds because, for any sequence $\langle v_n \rangle_{n\in\mathbb{N}}$, $\{n \mid v_n = v_n\} = \mathbb{N} \in \mathcal{U}$.

Symmetry follows from the symmetry of equality, since if $\langle v_n \rangle_{n\in\mathbb{N}} \sim_{\mathcal{U}} \langle w_n \rangle_{n\in\mathbb{N}}$ then

$$\{n \mid w_n = v_n\} = \{n \mid v_n = w_n\} \in \mathcal{U}.$$

And transitivity holds because if $\langle v_n \rangle_{n \in \mathbb{N}} \sim_{\mathcal{U}} \langle w_n \rangle_{n \in \mathbb{N}}$ and $\langle v_n \rangle_{n \in \mathbb{N}} \sim_{\mathcal{U}} \langle x_n \rangle_{n \in \mathbb{N}}$ then

$$\{n \mid v_n = x_n\} \supseteq \{n \mid v_n = w_n\} \cap \{n \mid w_n = x_n\} \in \mathcal{U}.$$

□

We will decide that two sequences represent the same vertex of our graph if they are equivalent to each other using $\sim_{\mathcal{U}}$. That is, we will use the *quotient* $[V_n]_{\mathcal{U}} = \prod_{n \in \mathbb{N}} V_n / \sim_{\mathcal{U}}$.

Formally, a vertex of $[V_n]_{\mathcal{U}}$ is an equivalence class of sequences—that is, it is a set of sequences which are all $\sim_{\mathcal{U}}$ equivalent to each other.

Definition 2.18. When $\langle v_n \rangle_{n \in \mathbb{N}}$, we will write $[v_n]_{\mathcal{U}}$ for the equivalence class of $\langle v_n \rangle_{n \in \mathbb{N}}$ in $[V_n]_{\mathcal{U}}$.

We call $\langle v_n \rangle_{n \in \mathbb{N}}$ a *representative* of the equivalence class $[v_n]_{\mathcal{U}}$.

It would not be unreasonable to write $\lim_{n \to \mathcal{U}} v_n = [v_n]_{\mathcal{U}}$, and we will sometimes view $[v_n]_{\mathcal{U}}$ as a sort of limit of the sequence $\langle v_n \rangle_{n \in \mathbb{N}}$.

When we want to talk about an element $v \in [V_n]_{\mathcal{U}}$, we will often pick some sequence $\langle v_n \rangle_{n \in \mathbb{N}}$ such that $[v_n]_{\mathcal{U}} = v$. The notation $[v_n]_{\mathcal{U}}$ is supposed to remind us that we are working with one of the sequences which represents the equivalence class.

Note that $[v_n]_{\mathcal{U}} = [w_n]_{\mathcal{U}}$ exactly when $\langle v_n \rangle_{n \in \mathbb{N}} \sim_{\mathcal{U}} \langle w_n \rangle_{n \in \mathbb{N}}$, and therefore exactly when $\{n \mid v_n = w_n\} \in \mathcal{U}$—two sequences represent the same equivalence class when they are equivalent.

There are many different sequences representing an equivalence class, so when working with the notation $[v_n]_{\mathcal{U}}$, we have to be careful that we really are talking about the equivalence class, not the sequence—that is, we need to be sure that we would get the same result if we used a different representative.

For instance, we need to revisit our definition of $[E_n]_{\mathcal{U}}$: we defined adjacency for sequences, not equivalence classes. Suppose $[v_n]_{\mathcal{U}} = [v'_n]_{\mathcal{U}}$ and $[w_n]_{\mathcal{U}} = [w'_n]_{\mathcal{U}}$, and $\{\langle v_n \rangle_{n \in \mathbb{N}}, \langle w_n \rangle_{n \in \mathbb{N}}\} \in [E_n]_{\mathcal{U}}$. We have to worry about the possibility that $\{\langle v'_n \rangle_{n \in \mathbb{N}}, \langle w'_n \rangle_{n \in \mathbb{N}}\} \notin [E_n]_{\mathcal{U}}$—in other words, that we could have equivalent sequences which disagree about whether or not the pair belongs to $[E_n]_{\mathcal{U}}$.

Again, the fact that ultrafilters are closed under intersection comes to our rescue: if $\{\langle v_n \rangle_{n \in \mathbb{N}}, \langle w_n \rangle_{n \in \mathbb{N}}\} \in [E_n]_{\mathcal{U}}$ then

$$\{n \mid \{v'_n, w'_n\} \in E_n\} \supseteq \{n \mid \{v_n, w_n\} \in E_n\} \cap \{n \mid v_n = v'_n\} \cap \{n \mid w_n = w'_n\} \in \mathcal{U},$$

so also $\{\langle v'_n\rangle_{n\in\mathbb{N}}, \langle w'_n\rangle_{n\in\mathbb{N}}\} \in [E_n]_\mathcal{U}$.

So we can define an edge relation on $\prod_{n\in\mathbb{N}} V_n/\sim_\mathcal{U}$—which we will also call $[E_n]_\mathcal{U}$—by saying $\{[v_n],[w_n]\} \in [E_n]_\mathcal{U}$ when $\{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}$.

This fixes the example that gave us trouble. If G_n is the complete graph $(\{1,2,\ldots,n\}, \binom{\{1,2,\ldots,n\}}{2})$ then $\{[v_n],[w_n]\} \in E_\mathcal{U}$ if and only if

$$\{n \mid v_n \neq w_n\} = \{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U},$$

so exactly when $[v_n]_\mathcal{U} \neq [w_n]_\mathcal{U}$. So, with this modification, the limiting graph is the complete graph on the set $[V_n]_\mathcal{U}$.

2.5 Ultraproducts

We can assemble what we said in the previous section into the full definition of our limit objects.

Definition 2.19. If $G_n = (V_n, E_n)$ is a sequence of graphs and $\mathcal{U}$ is an ultrafilter, the *ultraproduct*, written $\prod_{n\to\mathcal{U}} G_n$ or $[G_n]_\mathcal{U}$ is the graph $([V_n]_\mathcal{U}, [E_n]_\mathcal{U})$ where:

- $[V_n]_\mathcal{U} = \prod_n V_n/\sim_\mathcal{U}$—that is, $[V_n]_\mathcal{U}$ consists of equivalence classes $[v_n]_\mathcal{U}$ where $\langle v_n\rangle_{n\in\mathbb{N}}$ is a sequence with $v_n \in V_n$ for all n and $[v_n]_\mathcal{U} = [w_n]_\mathcal{U}$ if $\{n \mid v_n = w_n\} \in \mathcal{U}$, and
- $[E_n]_\mathcal{U}$ consists of pairs $\{[v_n]_\mathcal{U}, [w_n]_\mathcal{U}\}$ such that $\{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}$.

We call the graphs G_n the *ground graphs* of $\prod_{n\to\mathcal{U}} G_n$.

The point is that, in many ways, the ultraproduct $\prod_{n\to\mathcal{U}} G_n$ will capture the "limiting" behavior of the ground graphs. Understanding how the ultraproduct resembles the ground graphs will concern us throughout the rest of the book. Although the ultraproduct can depend on the particular ultrafilter $\mathcal{U}$, we will generally work with an arbitrary ultrafilter and focus on the relationship between the ground graphs and the ultraproduct.

Clearly there are ways that the ultraproduct differs from the ground structures; for example, even if the G_n are finite, $[G_n]_\mathcal{U}$ is typically infinite, and indeed, uncountably infinite.

For our purposes, the only sizes are finite ones, countably infinite, and uncountably infinite. Recall that a set S is countable (that is, either finite or countably infinite) if there is surjective function $f : \mathbb{N} \to S$, and countably infinite if there is a bijection $f : \mathbb{N} \to S$.

There are many different incfinite cardinalities, but we will not be concerned with the distinctions among them.

Theorem 2.20. *Suppose that* $\lim_{n\to\infty} |V_n| = \infty$. *Then* $[V_n]_{\mathcal{U}}$ *is uncountably infinite.*

Proof. Suppose not—that is, suppose $[V_n]_{\mathcal{U}}$ were countable. Then there would be a surjective function $v : \mathbb{N} \to [V_n]_{\mathcal{U}}$. We can choose a representative for each element: $v(i) = [v_n^i]_{\mathcal{U}}$. (We do not need to be particular about this—any representatives will do.) We need to find a sequence $\langle w_n \rangle_{n\in\mathbb{N}}$ so that $[w_n]_{\mathcal{U}} \neq [v_n^i]_{\mathcal{U}}$ for all i.

For each n, choose $w_n \in V_n \setminus \{v_n^1, \ldots, v_n^{|V_n|-1}\}$. Since we are only excluding at most $|V_n| - 1$ values form V_n, we know some choice of w_n is possible. (Note that, when we work index-wise, considering each index n individually, we also only consider the n-th terms of the representatives—we are considering v_n^i because this is an element of V_n.)

For any i, we need to show that $[w_n]_{\mathcal{U}} \neq [v_n^i]_{\mathcal{U}}$. This is the same as showing that $\{n \mid w_n \neq v_n^i\} \in \mathcal{U}$. We chose w_n to avoid v_n^i when $i < |V_n|$. Therefore

$$\{n \mid w_n \neq v_n^i\} \supseteq \{n \mid |V_n| > i\}.$$

Since $\lim_{n\to\infty} |V_n| = \infty$, $\{n \mid |V_n| > i\}$ is cofinite and therefore in $\mathcal{U}$, so $\{n \mid w_n \neq v_n^i\} \in \mathcal{U}$.

Therefore, for each i, $[w_n]_{\mathcal{U}} \neq [v_n^i]_{\mathcal{U}}$, contradicting the surjectivity of v.

So whenever we have a countable list of elements of V, we can obtain a new element different from all of them. Therefore V is uncountably infinite. □

We said at the beginning of the chapter that we want the subgraph density of the limit to be the limit of the subgraph densities; in our new terminology, we can say that we want

$$t_H(\prod_{n\to\mathcal{U}} G_n) = \lim_{n\to\mathcal{U}} t_H(G_n).$$

Working with subgraph *density* will have to wait until the next chapter, when we develop a probability measure in ultraproducts, but we can at least show that the presence of finite graphs in the ultraproduct reflects their presence in the ground graphs.

Theorem 2.21. *Let H be a finite graph. There is a copy of H in $[G_n]_{\mathcal{U}}$ if and only if*

$$\{n \mid \text{there is a copy of } H \text{ in } G_n\} \in \mathcal{U}.$$

Proof. First, suppose there is a copy of $H = (W, F)$ in $[G_n]_{\mathcal{U}}$. Recall, that means there is a function $\pi : W \to V$ such that, for each edge $\{w, w'\} \in F$, $\{\pi(w), \pi(w')\} \in [E_n]_{\mathcal{U}}$. For each $w \in W$, pick a representative $\pi(w) = [v_n^w]_{\mathcal{U}}$. Then, for each $\{w, w'\} \in F$, we must have a set

$$K_{w,w'} = \{n \mid \{v_n^w, v_n^{w'}\} \in E_n\} \in \mathcal{U}$$

in the ultraproduct which witnesses the presence of the edge between $\pi(w) = [v_n^w]_{\mathcal{U}}$ and $\pi(w') = [v_n^{w'}]_{\mathcal{U}}$.

Since W, and therefore F, is finite, the intersection of all these sets, $\bigcap_{\{w,w'\}\in F} K_{w,w'}$ is also in $\mathcal{U}$. For any $n \in \bigcap_{\{w,w'\}\in F} K_{w,w'}$, we claim there is a copy of H in G_n: take $\pi_n(w) = v_n^w$. Then for any $\{w, w'\} \in F$, $\{\pi_n(w), \pi_n(w')\} = \{v_n^w, v_n^{w'}\} \in E_n$ because $n \in K_{w,w'}$.

For the converse, suppose that there is a set $K \in \mathcal{U}$ such that, for every $n \in K$, we have a copy of H in V_n—that is, a function $\pi_n : W \to V_n$ so that, whenever $\{w, w'\} \in F$, $\{\pi_n(w), \pi_n(w')\} \in E_n$. For each $n \notin K$, define $\pi_n : W \to V_n$ arbitrarily—we do not require that π_n be a copy of H when $n \notin K$.

Then we define $\pi : W \to V$ by assigning $\pi(w)$ to be $[\pi_n(w)]_{\mathcal{U}}$. For any $\{w, w'\} \in F$, $\{n \mid \{\pi_n(w), \pi_n(w')\} \in E_n\} \supseteq K$, so $\{\pi(w), \pi(w')\} \in [E_n]_{\mathcal{U}}$. Therefore π is an actual copy of H. □

The same argument would also show that there is an *induced* copy of H in $[G_n]_{\mathcal{U}}$ if and only if $\{n \mid \text{there is an induced copy of } H \text{ in } G_n\} \in \mathcal{U}$.

Properties like this will be a recurring theme: properties $\mathfrak{P}$ such that $\mathfrak{P}$ is true of the ultraproduct exactly when $\{n \mid \mathfrak{P} \text{ is true of } G_n\} \in \mathcal{U}$. We give one more (somewhat artificial) example right now to illustrate the technique, and to get more practice working with ultraproducts.

Let us temporarily say that a graph $G = (V, E)$ has an *isolated triangle* if it contains three vertices u, v, w with $\{u, v\}, \{u, w\}, \{v, w\} \in E$ where none of the vertices u, v, w has any neighbors other than the other vertices in the triangle. (That is, the triangle is a connected component of the graph.) We consider this example because it is essentially the simplest property we can ask about which is not just the presence or absence of a finite graph.

Theorem 2.22. *The ultraproduct $[G_n]_{\mathcal{U}}$ has an isolated triangle if and only if $\{n \mid G_n$ has an isolated triangle$\} \in \mathcal{U}$.*

Proof. First, suppose $[G_n]_{\mathcal{U}}$ has an isolated triangle, u, v, w. We can choose representatives $u = [u_n]_{\mathcal{U}}$, $v = [v_n]_{\mathcal{U}}$, and $w = [w_n]_{\mathcal{U}}$. We claim that the set of n such that u_n, v_n, w_n is an isolated triangle belongs to $\mathcal{U}$.

Let $K_1 = \{n \mid \{u_n, v_n\} \in E_n\}$, $K_2 = \{n \mid \{u_n, w_n\} \in E_n\}$, and $K_3 = \{n \mid \{v_n, w_n\} \in E_n\}$; since u, v, w is a triangle in the ultraproduct, each of these sets belongs to $\mathcal{U}$, so also $K = K_1 \cap K_2 \cap K_3$ belongs to $\mathcal{U}$.

Let J_1 be the set of n such that u_n has a neighbor other than v_n or w_n, let J_2 be the set of n such that v_n has a neighbor other than u_n or w_n, and let J_3 be the set of n such that w_n has a neighbor other than u_n or v_n.

For each $n \in J_1$, there is a vertex x_n with $\{u_n, x_n\} \in E_n$, $x_n \neq v_n$, and $x_n \neq w_n$. If J_1 were in $\mathcal{U}$ then we could take a vertex $x = [x_n]_{\mathcal{U}}$ (where x_n is any element of V_n for $n \notin J_1$). But then we would have $\{x, u\} \in [E_n]_{\mathcal{U}}$, $x \neq v$, and $x \neq w$, contradicting the fact that we started with an isolated triangle.

By the same argument, neither J_2 nor J_3 can be in $\mathcal{U}$. Since $\mathcal{U}$ is an ultrafilter, that means their complements $\mathbb{N} \setminus J_i$ must be in $\mathcal{U}$, so

$$K \cap (\mathbb{N} \setminus J_1) \cap (\mathbb{N} \setminus J_2) \cap (\mathbb{N} \setminus J_3) \in \mathcal{U}.$$

But for any n in this set, u_n, v_n, w_n is an isolated triangle.

To prove the converse, let K be the set of n such that G_n contains an isolated triangle and suppose that $K \in \mathcal{U}$. Then, for each $n \in K$, we can choose a particular isolated triangle u_n, v_n, w_n. We can take the elements $[u_n]_{\mathcal{U}}, [v_n]_{\mathcal{U}}, [w_n]_{\mathcal{U}}$ of $[V_n]_{\mathcal{U}}$ (where, once again, we take u_n, v_n, w_n to be any element of V_n we like when $n \notin K$). For every $n \in K$ we have $\{u_n, v_n\}, \{u_n, w_n\}, \{v_n, w_n\} \in E_n$, so $[u_n]_{\mathcal{U}}, [v_n]_{\mathcal{U}}, [w_n]_{\mathcal{U}}$ form a triangle in $[G_n]_{\mathcal{U}}$.

Suppose this triangle is not isolated, so there is some vertex $x = [x_n]_{\mathcal{U}}$ adjacent to one of the vertices in the triangle but not equal to any of the vertices in the triangle; we may as well assume $\{[x_n]_{\mathcal{U}}, [u_n]_{\mathcal{U}}\} \in [E_n]_{\mathcal{U}}$, $[x_n]_{\mathcal{U}} \neq [v_n]_{\mathcal{U}}$, and $[x_n]_{\mathcal{U}} \neq [w_n]_{\mathcal{U}}$. That means $\{n \mid \{x_n, u_n\} \in E_n\} \in \mathcal{U}$, $\{n \mid x_n \neq v_n\} \in \mathcal{U}$, and $\{n \mid x_n \neq w_n\} \in \mathcal{U}$. Therefore also

$$K \cap \{n \mid \{x_n, u_n\} \in E_n\} \cap \{n \mid x_n \neq v_n\} \cap \{n \mid x_n \neq w_n\} \in \mathcal{U}.$$

Since $\emptyset \notin \mathcal{U}$, there must be an n in all these sets. But this gives a contradiction: we chose u_n, v_n, w_n to be an isolated triangle in V_n, and x_n contradicts that isolation.

So $[u_n]_{\mathcal{U}}, [v_n]_{\mathcal{U}}, [w_n]_{\mathcal{U}}$ does form an isolated triangle in $[G_n]_{\mathcal{U}}$. □

On the other hand, there will also be important properties which do *not* pass from the ground graphs to the ultraproduct.

Theorem 2.23. *There is a sequence of finite* connected *graphs $G_n = (V_n, E_n)$ such that $[G_n]_\mathcal{U}$ is not connected.*

Recall that a graph is connected if for any vertices v, w, there is a finite path $v = v_1, v_2, \ldots, v_n = w$ so that, for each $i < n$, $\{v_i, v_{i+1}\} \in E$.

Proof. Take G_n to be the path of length n: $V_n = \{1, 2, \ldots, n\}$ and $\{i, j\} \in E_n$ exactly when $|i - j| = 1$—that is, the graph  with n vertices.

To see that the ultraproduct is not connected, consider the vertex $[v_n]_\mathcal{U}$ where $v_n = 1$ for all n, and the vertex $[w_n]_\mathcal{U}$ where $w_n = n$ for all n. Suppose there were a finite path between these vertices; let us write this path $u^1, u^2, \ldots, u^k$ where $u^1 = [v_n]_\mathcal{U}$ and $u^k = [w_n]_\mathcal{U}$ and, for each $i < k$, $\{u^i, u^{i+1}\} \in E$. Pick representatives $u^i = [u^i_n]_\mathcal{U}$ where $u^1_n = v_n$ and $u^k_n = w_n$.

For each $i < k$, let $K_i = \{n \mid \{u^i_n, u^{i+1}_n\} \in E_n\}$, so $K_i \in \mathcal{U}$. Then also $\bigcap_{i<k} K_i \in \mathcal{U}$. In particular, $\bigcap_{i<k} K_i$ must be infinite, so we can choose $n \in \bigcap_{i<k} K_i$ with $n > k$. Then we have $u^1_n = v_n = 1$ adjacent to u^2_n, so $u^2_n = 2$. Since u^2_n is adjacent to u^3_n, we must have $u^3_n = 3$ or $u^3_n = 1$. Continuing in this way, we must have $u^i_n \leq i$ for each i, which would mean that $k \geq u^k_n = w_n = n$, contradicting the choice that $k < n$.

Therefore there cannot be a finite path from $[v_n]_\mathcal{U}$ to $[w_n]_\mathcal{U}$. □

This example is worth investigating slightly further. Recall that when a graph (V, E) is not connected, we can partition $V = X \cup Y$ so that X and Y are non-empty but there are no edges between X and Y. The ultraproduct in the previous example is disconnected; one fairly natural choice of a disconnected partition is to take X to be all the vertices represented by constant sequences—the vertices of the form $[v_n]_\mathcal{U}$ where $v_n = c$ for $c \leq n$ and v_n arbitrary for $n < c$; we abbreviate these $[c]_\mathcal{U}$. Then Y can be all other vertices. The argument in the proof shows that these sets are not adjacent—that the vertices in X are only adjacent to other vertices in X (indeed, $[c]_\mathcal{U}$ is adjacent exactly to $[c+1]_\mathcal{U}$ and $[c-1]_\mathcal{U}$).

The sets X and Y have an important property: they cannot be described "coordinate-wise" in terms of subsets of the V_n.

Definition 2.24. When $X_n \subseteq V_n$ is a sequence of sets, let $[X_n]_\mathcal{U}$ consist of all points $[v_n]_\mathcal{U}$ such that $\{n \mid v_n \in X_n\} \in \mathcal{U}$.

Note that, despite being stated in terms of coordinates, this definition does not depend on the choice of representatives for the point $[v_n]_\mathcal{U}$: if $[v_n]_\mathcal{U} = [v'_n]_\mathcal{U}$ and $[v_n]_\mathcal{U} \in [X_n]_\mathcal{U}$ then

$$\{n \mid v'_n \in X_n\} \supseteq \{n \mid v_n \in X_n\} \cap \{n \mid v_n = v'_n\} \in \mathcal{U},$$

so we also have $[v'_n]_{\mathcal{U}} \in [X_n]_{\mathcal{U}}$.

Lemma 2.25. *In the ultraproduct $[G_n]_{\mathcal{U}} = ([V_n]_{\mathcal{U}}, [E_n]_{\mathcal{U}})$ where G_n is the path of length n, the set $X = \{[c]_{\mathcal{U}} \mid c \in \mathbb{N}\}$ is not equal to $[X_n]_{\mathcal{U}}$ for any sequence of sets $X_n \subseteq X$.*

Proof. Consider some sequence of sets $X_n \subseteq V_n$. We will show that $X \neq [X_n]_{\mathcal{U}}$. Assume $[X_n]_{\mathcal{U}} \supseteq X$, and we will construct a point $[v_n]_{\mathcal{U}} \in [X_n]_{\mathcal{U}} \setminus X$. For each n, take v_n to be the largest element of X_n if $X_n \neq \emptyset$, and $v_n = 1$ if $X_n = \emptyset$. We have

$$\{n \mid v_n \in X_n\} = \{n \mid X_n \neq \emptyset\} \supseteq \{n \mid 1 \in X_n\}$$

where $\{n \mid 1 \in X_n\} \in \mathcal{U}$ since $[1]_{\mathcal{U}} \in X \subseteq [X_n]_{\mathcal{U}}$.

For any c, we have $[c+1]_{\mathcal{U}} \in X \subseteq [X_n]_{\mathcal{U}}$, so

$$\{n \mid v_n \neq c\} \supseteq \{n \mid v_n > c\} \supseteq \{n \mid c+1 \in X_n\} \in \mathcal{U},$$

which shows that $[v_n]_{\mathcal{U}} \neq [c]_{\mathcal{U}}$.

Since $[v_n]_{\mathcal{U}} \neq [c]_{\mathcal{U}}$ for every c, we conclude that $[v_n]_{\mathcal{U}} \in [X_n]_{\mathcal{U}} \setminus X$. □

2.6 A Few Examples

Before going on, we should consider a few examples of ultraproducts of graphs. Of course, since we are only beginning to investigate the properties of ultraproducts, we will be limited in what we can say about them, but we will be able to flesh them out later.

First, consider the ultraproduct $\mathrm{K}_{\infty,\infty} = [\mathrm{K}_{n,n}]_{\mathcal{U}}$.

Lemma 2.26. *$\mathrm{K}_{\infty,\infty} = (V, E)$ is a complete bipartite graph with two infinite parts.*

Proof. For each n, we have $\mathrm{K}_{n,n} = V_n^0 \cup V_n^1$ with $|V_n^0| = |V_n^1|$ and $E_n = (V_n^0 \times V_n^1) \cup (V_n^1 \times V_n^0)$.

For each $x = [x_n]_{\mathcal{U}} \in V$, we have $\mathbb{N} = \{n \mid x_n \in V_n^0\} \cup \{n \mid x_n \in V_n^1\}$, so exactly one of these sets belongs to $\mathcal{U}$. Naturally, we take V_0 to be those $[x_n]_{\mathcal{U}}$ where $\{n \mid x_n \in V_n^0\} \in \mathcal{U}$ and V_1 to be those $[x_n]_{\mathcal{U}}$ where $\{n \mid x_n \in V_n^1\} \in \mathcal{U}\}$.

If $x = [x_n]_{\mathcal{U}}$ and $y = [y_n]_{\mathcal{U}}$ then $\{x, y\} \in E$ if and only if $\{n \mid \{x_n, y_n\} \in E_n\} \in \mathcal{U}$, which happens if and only if $\{n \mid x_n$ and y_n are in different parts$\} \in \mathcal{U}$, which happens if and only if x and y are in different parts. □

We could also consider variants, like $\mathrm{K}_{\infty,\infty^2} = [\mathrm{K}_{n,n^2}]_{\mathcal{U}}$. By the same argument, we can see that

Lemma 2.27. $\mathrm{K}_{\infty,\infty^2} = (V, E)$ *is a complete bipartite graph with two infinite parts.*

It will be slightly harder to make sense of the idea that, in $\mathrm{K}_{\infty,\infty}$, the two parts have the same size, while in $\mathrm{K}_{\infty,\infty^2}$ one part is much larger than the other. Indeed, as graphs, $\mathrm{K}_{\infty,\infty}$ and $\mathrm{K}_{\infty,\infty^2}$ are actually isomorphic. However we will show later that only in $K_{\infty,\infty}$ can we find a bijection between the two parts which "respects the ultraproduct" in a certain sense.

If, for each n, we take $G_n = \mathbf{R}_p(\{1, \ldots, n\})$, a random graph on n vertices, we get an ultraproduct we will call $\mathbf{R}_{p,\mathcal{U}} = [G_n]_{\mathcal{U}}$. Instead of talking about what happens "with high probability", the interesting properties of $\mathbf{R}_{p,\mathcal{U}}$ will turn out to be mostly settled up to probability 1.

2.7 Internal Sets

Sets which can be represented coordinate-wise will be central to our understanding of ultraproducts and their uses.

Definition 2.28. We say a set $X \subseteq [V_n]_{\mathcal{U}}$ is *internal* if there is some sequence $X_n \subseteq V_n$ such that $X = [X_n]_{\mathcal{U}}$ (as sets—that is, $v \in X$ if and only if $v \in [X_n]_{\mathcal{U}}$).

More generally, we will speak of internal subsets of the product set $[V_n]^k_{\mathcal{U}} = [V_n^k]_{\mathcal{U}}$ for $k > 1$ in the same way—$X \subseteq [V_n]^k_{\mathcal{U}}$ is internal if there are sets $X_n \subseteq V_n^k$ so that X is precisely the set of tuples $([v_n^1]_{\mathcal{U}}, \ldots, [v_n^k]_{\mathcal{U}})$ such that

$$\{n \mid (v_n^1, \ldots, v_n^k) \in X_n\} \in \mathcal{U}.$$

For example, the set $\{(v, w) \mid \{v, w\} \in [E_n]_{\mathcal{U}}\}$ is an internal subset of $[V_n]^2_{\mathcal{U}}$. (Indeed, as the notation suggests, we could think of $[E_n]_{\mathcal{U}}$ itself as an internal subset of $\binom{[V_n]_{\mathcal{U}}}{2}$.)

An internal set X has many representations in the form $[X_n]_{\mathcal{U}}$, but these representations are as similar as we could hope for—they are equal almost everywhere.

Lemma 2.29. *If $[X_n]_{\mathcal{U}} = [Y_n]_{\mathcal{U}}$ (as sets—any element of one set is also an element of the other) then $\{n \mid X_n = Y_n\} \in \mathcal{U}$.*

Proof. We prove the contrapositive. Suppose $\{n \mid X_n = Y_n\} \notin \mathcal{U}$. Then

$$\{n \mid X_n \neq Y_n\} = \{n \mid X_n \setminus Y_n \neq \emptyset\} \cup \{n \mid Y_n \setminus X_n \neq \emptyset\} \in \mathcal{U}.$$

So one of these sets must in $\mathcal{U}$; without loss of generality, assume $K = \{n \mid X_n \setminus Y_n \neq \emptyset\} \in \mathcal{U}$. Then, for each $n \in K$, take $a_n \in X_n \setminus Y_n$. Letting a_n be arbitrary for $n \notin K$, we have $\{n \mid a_n \in X_n\} \supseteq K \in \mathcal{U}$ and $\{n \mid a_n \notin Y_n\} \supseteq K \in \mathcal{U}$, so $[a_n]_\mathcal{U} \in [X_n]_\mathcal{U} \setminus [Y_n]_\mathcal{U}$. □

In the previous section we showed that a particular set X coming from a partition of an ultraproduct into disconnected components was external. That example reflects a general phenomenon: an ultraproduct of connected graphs may not be connected, but it is "internally connected", in the sense that it cannot be partitioned into non-empty internal sets without edges between them.

Theorem 2.30. *Suppose that $\{n \mid G_n$ is connected$\} \in \mathcal{U}$. Then for any internal set $X \subseteq [V_n]_\mathcal{U}$ such that X and $[V_n]_\mathcal{U} \setminus X$ are non-empty, there is a $v \in X$ and a $w \in [V_n]_\mathcal{U} \setminus X$ with $\{v, w\} \in [E_n]_\mathcal{U}$.*

Proof. Since X is internal, we have $X = [X_n]_\mathcal{U}$ for some sequence $X_n \subseteq V_n$. Let $I = \{n \mid X_n \neq \emptyset\}$ and let $J = \{n \mid V_n \setminus X_n \neq \emptyset\}$. Since X and $[V_n]_\mathcal{U} \setminus X$ are both non-empty, $I, J \in \mathcal{U}$. Also, let K be $\{n \mid G_n$ is connected$\}$, which is also in $\mathcal{U}$.

For any $n \in I \cap J \cap K$, we know that G_n is connected, there is a vertex in X_n, and a vertex in $V_n \setminus X_n$. There must be a path between these vertices, so by starting with the vertex in X_n and following the path until we leave X_n, we must find a point $v_n \in X_n$ which is adjacent to some $w_n \in V_n \setminus X_n$—that is, $\{v_n, w_n\} \in E_n$. (Choose v_n and w_n arbitrarily for $n \notin I \cap J \cap K$.)

Since $I \cap J \cap K \in \mathcal{U}$, we can conclude that all three properties pass up to the ultraproduct: $[v_n]_\mathcal{U} \in [X_n]_\mathcal{U}$, $[w_n]_\mathcal{U} \in V \setminus [X_n]_\mathcal{U}$, and $\{[v_n]_\mathcal{U}, [w_n]_\mathcal{U}\} \in E_\mathcal{U}$. □

This sort of phenomenon will be common: when we consider only internal sets, rather than all sets, the ultraproduct will often closely resemble the ground graphs.

Analogous to internal sets, we have internal functions.

Definition 2.31. A function $f : [V_n]^k_\mathcal{U} \to [V_n]^m_\mathcal{U}$ is *internal* if there is a sequence of functions $f_n : V_n^k \to V_n^m$ such that, for any $[\vec{v}_n]_\mathcal{U} \in [V_n]^k$,

$$f([\vec{v}_n]_\mathcal{U}) = [f_n(\vec{v}_n)]_\mathcal{U}.$$

This lets us characterize the difference between $\mathrm{K}_{\infty,\infty}$ and $\mathrm{K}_{\infty,\infty^2}$: in $\mathrm{K}_{\infty,\infty}$, there is an internal bijection between the two parts, but not in $\mathrm{K}_{\infty,\infty^2}$.

The internal sets have some natural closure properties. They form an *algebra*.

Definition 2.32. If $\mathcal{B} \subseteq \mathcal{P}(V)$ (the power set of V), we say $\mathcal{B}$ is an *algebra* if:

- $\emptyset \in \mathcal{B}$ and $V \in \mathcal{B}$,
- whenever $B \in \mathcal{B}$, $V \setminus B \in \mathcal{B}$,
- whenever $B_0, B_1 \in \mathcal{B}$, also $B_0 \cup B_1 \in \mathcal{B}$.

These properties immediately imply that if $\mathcal{B}$ is an algebra and $B_0, B_1 \in \mathcal{B}$ then also $B_0 \cap B_1 \in \mathcal{B}$.

Lemma 2.33. *For each k, the internal subsets of $[V_n]^k_{\mathcal{U}}$ form an algebra.*

Proof. $\emptyset = [\emptyset]_{\mathcal{U}}$ and $[V_n]^k_{\mathcal{U}} = [V_n^k]_{\mathcal{U}}$.

Let $B \subseteq [V_n]^k_{\mathcal{U}}$ be internal, so $B = [B_n]_{\mathcal{U}}$. Then

$$\begin{aligned} [x_n]_{\mathcal{U}} \in [V_n]^k_{\mathcal{U}} \setminus B \text{ if and only if } & [x_n]_{\mathcal{U}} \notin B \\ \text{if and only if } & \{n \mid x_n \in B_n\} \notin \mathcal{U} \\ \text{if and only if } & \{n \mid x_n \in V_n^k \setminus B_n\} \in \mathcal{U} \\ \text{if and only if } & [x_n]_{\mathcal{U}} \in [V_n^k \setminus B_n]_{\mathcal{U}}, \end{aligned}$$

so $[V_n]^k_{\mathcal{U}} \setminus B = [V_n^k \setminus B_n]_{\mathcal{U}}$.

If $B^0, B^1 \subseteq [V_n]^k_{\mathcal{U}}$ are internal, so $B^0 = [B^0_n]_{\mathcal{U}}$ and $B^1 = [B^1_n]_{\mathcal{U}}$,

$$\begin{aligned} [x_n]_{\mathcal{U}} \in B^0 \cup B^1 \text{ if and only if } & [x_n]_{\mathcal{U}} \in B^0 \text{ or } [x_n]_{\mathcal{U}} \in B^1 \\ \text{if and only if } & \{n \mid x_n \in B^0_n\} \in \mathcal{U} \text{ or } \{n \mid x_n \in B^1_n\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid x_n \in B^0_n \cup B^1_n\} \in \mathcal{U} \\ \text{if and only if } & [x_n]_{\mathcal{U}} \in [B^0_n \cup B^1_n]_{\mathcal{U}}, \end{aligned}$$

so $B^0 \cup B^1 = [B^0_n \cup B^1_n]_{\mathcal{U}}$. □

Since the union (or intersection) of two internal sets is internal, induction shows that the union (or intersection) of finitely many internal sets is internal. The union (or intersection) of *countably* many internal sets is almost never internal, however, as we will see later.

Internal sets have an additional important closure property. They are closed under projections:

Lemma 2.34. *If $B \subseteq [V_n]_{\mathcal{U}}^{k+1}$ is internal then*

$$\{\vec{b} \mid \exists x\, (\vec{b}, x) \in B\}$$

and

$$\{\vec{b} \mid \forall x\, (\vec{b}, x) \in B\}$$

are internal.

Proof. Since $\{\vec{b} \mid \forall x\, (\vec{b}, x) \in B\}$ is

$$[V_n]_{\mathcal{U}}^{k} \setminus \{\vec{b} \mid \forall x\, (\vec{b}, x) \in [V_n]^{k+1} \setminus B\},$$

it suffices to show only the first case and use the closure under complements we have already shown.

If $B \subseteq [V_n]_{\mathcal{U}}^{k+1}$ is internal, so $B = [B_n]_{\mathcal{U}}$, for each n let $C_n \subseteq V_n^k$ be given by $C_n = \{\vec{b} \mid \exists x \in V_n\, (\vec{b}, x) \in B_n\}$. Then

$$\begin{aligned}
\exists x\, ([\vec{b}_n]_{\mathcal{U}}, x) \in B &\text{ if and only if there is a } [x_n]_{\mathcal{U}} \text{ such that } ([\vec{b}_n]_{\mathcal{U}}, [x_n]_{\mathcal{U}}) \in B \\
&\text{if and only if there is a } [x_n]_{\mathcal{U}} \text{ such that } \{n \mid (\vec{b}_n, x_n) \in B_n\} \in \mathcal{U} \\
&\text{if and only if } \{n \mid \vec{b}_n \in C_n\} \in \mathcal{U} \\
&\text{if and only if } [\vec{b}_n]_{\mathcal{U}} \in [C_n]_{\mathcal{U}}.
\end{aligned}$$

□

Analogously to internal sets, we could speak of "internal statements"—that is, satements $\mathfrak{P}$ such that $\mathfrak{P}$ is true in $[V_n]_{\mathcal{U}}$ if and only if $\{n \mid \mathfrak{P} \text{ is true in } V_n\} \in \mathcal{U}$. For instance, Theorem 2.21 shows that "there exists a copy of H" is an "internal statement" in this sense.

To make this more rigorous, we would have to precisely formulate what we mean by a "statement". The right choice would be first-order logic: the closure properties above are are exactly what we need to align the internal sets with the definable sets of first-order logic. Indeed, the lemmata above immediately give the following:

Corollary 2.35. *Let $[V_n]_{\mathcal{U}}$ be an ultraproduct $\mathcal{L}$ a language of first-order logic, and suppose that each predicate symbol in $\mathcal{L}$ is interpreted by an internal set and each function symbol in $\mathcal{L}$ is interpreted by an internal function. Then all definable sets are internal.*

More concretely, we have the following theorem directly relating first-order logic to what happens in ultraproducts. In particular, if each V_n is the universe of an $\mathcal{L}$-structure $\mathfrak{M}_n$, the ultraproduct $[V_n]_{\mathcal{U}}$ is immediately an $\mathcal{L}$-structure $[\mathfrak{M}_n]_{\mathcal{U}}$ with all symbols interpreted by internal sets and functions. (For instance, for each predicate symbol R, $R^{[\mathfrak{M}_n]_{\mathcal{U}}} = [R_n^{\mathfrak{M}_n}]_{\mathcal{U}}$). Then all the definable sets are internal, and furthermore we have

Theorem 2.36 (Łoś's Theorem)**.** *Let $\mathcal{L}$ be a language of first-order logic. Suppose that, for each n, V_n is the universe of an $\mathcal{L}$-structure $\mathfrak{M}_n$.*

Then:

- *$[V_n]_{\mathcal{U}}$ is a $\mathcal{L}$-structure $[\mathfrak{M}_n]_{\mathcal{U}}$ with all symbols interpreted by internal sets and functions,*
- *for any formula ϕ,*

$$\{\vec{a} \in [V_n]_{\mathcal{U}}^k \mid [\mathfrak{M}_n]_{\mathcal{U}} \vDash \phi(\vec{a})\} = [\{\vec{a} \in V_n \mid \mathfrak{M}_n \vDash \phi(\vec{a})\}]_{\mathcal{U}}.$$

We will mostly not need this perspective, since we will work directly with the internal sets, but the connection between internal and definable sets is a substantive one. In fact, one can think of the internal sets as exactly those sets which are "potentially definable"—the sets definable, not necessarily in the current language, but in any extended language. To make this precise, one identifies first-order logic with the functors from structures to sets which respect ultraproducts [113].

We will be particularly interested in the following question: *how do properties of ultraproducts correspond to the properties of the ground structures?* Łoś' Theorem gives the beginning of an answer.

Corollary 2.37. *Suppose $\mathfrak{P}$ is a property which can be represented by a first-order sentence—that is, there is a formula σ with no free-variables such that a graph G has property $\mathfrak{P}$ if and only if the formula σ is true in G.*

Then $[G_n]_{\mathcal{U}}$ has the property $\mathfrak{P}$ if and only if $\{n \mid G_n$ has the property $\mathfrak{P}\} \in \mathcal{U}$.

In particular, properties like "contains a triangle" (or, more generally, contains a copy of the finite graph H) are first-order—that is, we have the following two equivalent facts:

- there is a sentence σ (in the language of graphs) so that G contains a triangle exactly when $G \vDash \sigma$,
- $[G_n]_{\mathcal{U}}$ contains a triangle if and only if $\{n \mid G_n$ contains a triangle $\} \in \mathcal{U}$.

2.8 Saturation

Every finite set is internal (the set $\{a^1, \ldots, a^k\}$ can be defined as $\{x \in V \mid x = a^1 \vee x = a^2 \vee \cdots \vee x = a^k\}$) and it is easy to produce uncountably infinite internal sets (like $[V_n]_{\mathcal{U}}$ itself). But there are no countably infinite internal sets. This follows from a much more general compactness property of ultraproducts called saturation.

Theorem 2.38 (Saturation)**.** *Suppose that, for each $i \in \mathbb{N}$, $X^i = [X^i_n]_{\mathcal{U}}$ is an internal set, and that, for every k, $\bigcap_{i \leq k} X^i$ is non-empty. Then $\bigcap_{i \in \mathbb{N}} X^i$ is non-empty.*

Proof. For each n, let k be maximal so that $\bigcap_{i \leq k} X^i_n$ is non-empty, and choose $v_n \in \bigcap_{i \leq k} X^i_n$. (If X^1_n is empty, choose v_n arbitrarily.)

For each i,

$$\{n \mid v_n \in X^i_n\} \supseteq \{n \mid i \leq n \text{ and } \bigcap_{i \leq k} X^i_n \text{ is non-empty}\} \in \mathcal{U},$$

so $[v_n]_{\mathcal{U}} \in [X^i_n]_{\mathcal{U}} = X^i$. Since this holds for every i, $[v_n]_{\mathcal{U}} \in \bigcap_i X^i$. □

This immediately implies that there can be no countably infinite internal set: if X is internal and $v^1, \ldots, v^i, \ldots$ are distinct elements of X then we can take $X^i = X \setminus \{v^1, \ldots, v^i\}$. $v^{k+1} \in \bigcap_{i \leq k} X^i$, so by saturation, there is some $v \in \bigcap_i X^i$, so $v \notin \{v^1, \ldots, v^i, \ldots\}$.

Saturation also implies the following, which we will need later.

Theorem 2.39. *Suppose that, for each i, X^i is internal. If $\bigcup_i X^i$ is internal then there is a k so that $\bigcup_i X^i = \bigcup_{i \leq k} X^i$.*

Proof. Suppose $X = \bigcup_i X^i$ were internal. Then the sets $Y^i = X \setminus X^i$ would also be internal. There can be no $v \in \bigcap_i Y^i = \bigcap_i (X \setminus X^i) = \emptyset$, so by the contrapositive of saturation, there must be some k so that $\bigcap_{i \leq k} Y^k = \emptyset$. Therefore $X = \bigcup_{i \leq k} X^i$. □

Saturation is one of the essential properties that characterizes how ultraproducts behave.

Definition 2.40. If $G = (V, E)$ is a graph and $v, w \in V$, the *distance between v and w* is the smallest k such that there is a path $v = v_0, v_1, \ldots, v_k = w$ such that $\{v_i, v_{i+1}\} \in E$ for all $i < k$, or ∞ if there is no such path.

G has *finite diameter* if there is a single value of k such that, for every $v, w \in V$, the distance between v and w is $\leq k$.

Theorem 2.41. *$[G_n]_{\mathcal{U}}$ is connected if and only if $[G_n]_{\mathcal{U}}$ has finite diameter.*

Proof. A graph with finite diameter is connected by definition. For the other direction, suppose $[G_n]_{\mathcal{U}} = ([V_n]_{\mathcal{U}}, [E_n]_{\mathcal{U}})$ does not have finite diameter. For each $i \in \mathbb{N}$, let $X^k = \{(v, w) \in V_{\mathcal{U}}^2 \mid \text{the distance between } v \text{ and } w \text{ is greater than } k\}$.

X^k is internal. One way to see this is to note that

$$\{(v, w, y_1, \ldots, y_k) \mid \{v, y_1\} \in E \text{ and } \{y_1, y_2 \in E\} \text{ and } \cdots \text{ and } \{y_{k-1}, w\} \in E\}$$

is internal, so the projection Y^k—onto those (v, w) so that there is a path of length exactly k between them—is also internal. Then $X^k = \bigcap_{j \leq k}([V_n]_{\mathcal{U}}^2 \setminus Y_j)$ is a finite intersection of the complements of these sets: X^k is those (v, w) so that, for each $j \leq k$, there is not a path of length j.

Since $[G_n]_{\mathcal{U}}$ does not have finite diameter, each $X^k = \bigcap_{i \leq k} X^j$ is non-empty. Therefore, by saturation, $\bigcap_i X^i$ is non-empty—there are a pair of vertices (v, w) such that the distance between v and w is larger than k for every k, and therefore there is no path between v and w, so $[G_n]_{\mathcal{U}}$ cannot be connected. □

This example is typical of the behavior of ultraproducts. Saturation forces a great deal of uniformity on an ultraproduct: if every pair of vertices has a finite distance (that is, if the graph is connected) then there must be a *uniform* bound on that distance—that is, the graph must have finite diameter.

We will later need an additional saturation-like property. Suppose we have internal sets $A^1 \subseteq A^2 \subseteq A^3 \subseteq \cdots$. We have already seen that, unless the sequence stops growing at some finite A^i, the union $\bigcup_{i \in \mathbb{N}} A^i$ is not internal. However we can find internal sets which *contain* the union and are somehow related to the sequence.

This is an instance of a more general property of ultraproducts called *sequential comprehensiveness* which, roughly speaking, lets us extend countable sequences.*

Pick representations $A^i = [A_n^i]_{\mathcal{U}}$. Then there is a "diagonal" set $A^* = [\bigcup_{i \leq n} A_n^i]_{\mathcal{U}}$. (One might wonder why we use $\bigcup_{i \leq n} A_n^i$ rather than just A_n^n, since $\bigcup_{i \leq n} A^i = A^n$. However we only know that when $i \leq j$, $A_n^i \subseteq A_n^j$ for

*At a technical level, sequential comprehensiveness is more powerful than saturation—that is, there exist structures which are saturated but do not have sequential comprehensiveness. For our purposes, where we just work in ultraproducts, we do not notice the difference because ultraproducts have both properties, but the distinction comes up in axiomatic approaches to nonstandard analysis; see [70], for instance. This distinction is reflected in the proof: we cannot use saturation, but must instead return again to looking at sets in terms of their sequences of representatives.

many n, with the set of n depending on i and j. So we might not have $A^i_n \subseteq A^n_n$ when $i \leq n$.)

Certainly $\bigcup_{i\in\mathbb{N}} A^i \subseteq A^*$: if $[v_n]_{\mathcal{U}} \in \bigcup_{i\in\mathbb{N}} A^i$ then there is some i so that $\{n \mid v_n \in A^i_n\} \in \mathcal{U}$, and therefore $\{n \mid i \leq n \text{ and } v_n \in A^i_n\} \in \mathcal{U}$ as well, so $[v_n]_{\mathcal{U}} \in A^*$.

Note that this is a coordinate-wise definition which *does* depend on the specific choice of representatives—the set A^* is not canonically associated with the sequence $A^1 \subseteq A^2 \subseteq \cdots$. Indeed, there are many sets with the same properties as A^*.

More generally, whenever $f : \mathbb{N} \to \mathbb{N}$ is a function, we can define an internal set $A^f = [\bigcup_{i\leq f(n)} A^i_n]_{\mathcal{U}}$. As long as $\{n \mid i \leq f(n)\} \in \mathcal{U}$, we will have $A^i \subseteq A^f$. (Indeed, more generally, if $\{n \mid g(n) \leq f(n)\} \in \mathcal{U}$ then we have $A^g \subseteq A^f$.) By choosing the function f carefully, we can sometimes construct internal sets with additional properties we need.

An application we need in the next chapter is showing that if we have an increasing sequence of internal sets below a decreasing sequence of internal sets, we can find an internal set C in between the two sequences.

Theorem 2.42. *Suppose that*

$$A^1 \subseteq A^2 \subseteq A^3 \subseteq \cdots \subseteq B^3 \subseteq B^2 \subseteq B^1$$

are internal sets. Then there is an internal set C such that, for every i, $A^i \subseteq C \subseteq B^i$.

Proof. Fix representatives $A^i = [A^i_n]_{\mathcal{U}}$ and $B^i = [B^i_n]_{\mathcal{U}}$. Let $C = A^f$ where $f(n)$ is the smallest j such that, for all $i, i' \leq j$, $A^i_n \subseteq B^{i'}_n$. (If $A^1_n \not\subseteq B^1_n$ then $f(n) = 0$, which is fine, since the set of n on which this happens is not in $\mathcal{U}$.)

For each i, i', let $K_{i,i'} = \{n \mid A^i_n \subseteq B^{i'}_n\} \in \mathcal{U}$. Since $A^i \subseteq B^{i'}$, $K_{i,i'} \in \mathcal{U}$. Therefore, for each j, $\bigcap_{i\leq j, i'\leq j} K_{i,i'} \in \mathcal{U}$. But if $n \in \bigcap_{i\leq j,i'\leq j} K_{i,i'}$ and $j \leq n$ then $f(n) \geq j$. In particular, for each j, $\{n \mid f(n) \geq j\} \in \mathcal{U}$, so $A^j \subseteq C$.

Similarly, whenever $f(n) \geq j$, $\bigcup_{i\leq f(n)} A^i_n \subseteq B^j_n$. Since $\{n \mid f(n) \geq j\} \in \mathcal{U}$, we also have $C \subseteq B^j$. □

2.9 Related Topic: Arrow's Theorem

When we build an ultraproduct, we have a sequence of graphs G_n, and the graphs "vote" on what should be true in the ultraproduct: for instance, $\{[a_n]_{\mathcal{U}}, [b_n]_{\mathcal{U}}\} \in E_{\mathcal{U}}$ when $\{n \mid \{a_n, b_n\} \in E_n\} \in \mathcal{U}$, which we can think of as saying that a "majority" of the n's voted for a_n and b_n to be adjacent.

This perspective can be taken somewhat literally, as a digression to voting theory shows: we can use ultraproducts to prove Arrow's famous theorem that there is no voting system satisfying certain reasonable properties.

Suppose there is a finite list of candidates, $\mathcal{C}$, running for some office. We have a set of voters, N, and each voter has a list of preferences among this candidates—that is, for each $n \in N$, there is a linear ordering $\prec_n$ on $\mathcal{C}$. (We assume voters always have a linear ordering of preferences.)

A *preference aggregation rule* is a function F which produces a single linear ordering $\prec = F(\{\prec_n\}_{n\in N})$ on $\mathcal{C}$ depending on the voter preferences. (Note that we are requiring that F produce a full, linear ranking on $\mathcal{C}$, not merely that it select the winner.)

For example, a familiar voting system is that each voter votes for their first choice—that is, chooses the top ranked candidate in $\prec_n$—and the ordering $\prec$ is given by the number of votes cast for each candidate, so the candidate with the most votes is highest in $\prec$, then the candidate with the next most votes, and so on. (We will assume, here and throughout this section, that ties never come up, since ties raise minor notational complications without changing the main ideas.)

This particular example ignores how the voters compare candidates other than their first choice. A different function might be to have each voter assign points to the candidates: perhaps each voter n gives $|\mathcal{C}|$ points to the top candidate in $\prec_n$, $|\mathcal{C}| - 1$ points to the second highest candidate in $\prec_n$, and so on, and then the aggregate ordering $\prec$ orders candidates by the number of points received.

Theorem 2.43 (Arrow's Impossibility Theorem)**.** *If N is finite and there are at least three candidates, there is no preference aggregation rule such that:*

(1) *If $a \succ_n b$ for every $n \in N$ then $a \succ b$.*

(2) *Suppose that $\{\prec_n\}_{n\in N}$ and $\{\prec'_n\}_{n\in N}$ are two different lists of preferences, $a, b \in \mathcal{C}$, and for every $n \in N$,*

$$a \prec_n b \Leftrightarrow a \prec'_n b.$$

Then also

$$a \prec b \Leftrightarrow a \prec' b.$$

(3) *There is no $n_0 \in N$ such that $\succ$ is always equal to $\succ_{n_0}$.*

The first requirement seems natural enough: it says that if every voter prefers a to b then a gets ranked above b; this is called the "unanimity" requirement.

The second requirement is more subtle. It is called "independence of irrelevant alternatives". It says that whether a is ranked above b depends only on which voters prefer a to b, and not on how other candidates compare to a and b. It will be useful to think of $\prec_n$ and $\prec'_n$ as "before" and "after" situations: an initial poll shows that the each voter n has the view $\prec_n$, which would give the outcome $\prec$. Later, the voters have changed their views about other candidates, and have the new views $\prec'_n$, but no one has changed their mind about a and b—each voter who thought $a \prec_n b$ still thinks $a \prec'_n b$ and vice versa. Then the outcome between a and b still hasn't changed: if $a \prec b$, so b was ahead of a in the earlier poll, then $a \prec' b$, so b is still ahead of a.

Both the examples above violate this second requirement. To see why the example where we base $\prec$ on the number of first-place votes violates it, suppose that there are three candidates $\{a, b, c\}$ and that, initially, 60% of the voters have the view $a \succ_n c \succ_n b$ while 40% have the view $b \succ_n a \succ_n c$. Then a gets more first place votes than b, so $a \succ b$. But then suppose c runs an effective campaign and half of the first group of voters change their mind about a and c: now 30% of voters have $c \succ_n a \succ_n b$, 30% still have $a \succ_n c \succ_n b$, and 40% still have $b \succ_n a \succ_n c$. Now we have $b \succ a$. This sort of "spoiler" effect is exactly what the second requirement is trying to prevent.

The third requirement says there is no dictator: there is no single voter whose preferences just get imposed.

Proof. We suppose there is such a preference aggregation rule satisfying the first two requirements, and we show that there is a dictator. We will do this by showing that a preference aggregation rule satisfying the first two conditions gives rise to a collection of sets resembling an ultrafilter.

Let us say that a set $K \subseteq N$ is *victorious* for a over b such that for any set of preferences $\{\prec_n\}_{n\in N}$, if $a \succ_n b$ for every $n \in K$ then $a \succ b$: if a can beat b with every voter in K then a will beat b in the final result.

In fact, being a victorious set does not depend on the particular candidates a and b: we first show that if K is victorious for some pair a over b then K is victorious for every pair of candidates.

Suppose K is victorious for a over b. Then we also show that, for any c, K is victorious for a over c. Suppose the voters have views $\{\prec_n\}_{n\in N}$ such that, for every $n \in K$, $a \succ_n c$. Whatever their views of b are, let us have them change their mind about b to have preferences $\{\prec'_n\}_{n\in N}$ so that every voter has $b \succ'_n c$, and in particular, every voter $n \in K$ has $a \succ'_n b \succ'_n c$. Since $a \succ'_n b$ for every $n \in K$, $a \succ' b$. Since $b \succ'_n c$ for every n, by unanimity we have $b \succ' c$. Since $\succ'$ is a linear ordering, $a \succ' c$, and then by independence

of irrelevant alternatives, $a \succ c$ as well. This works whenever $a \succ_n c$ for every $n \in K$, so K is victorious for a over c.

By a symmetric argument, K is victorious for every d over b. Combining these steps, K is victorious for every d over every c.

So let us just speak of victorious sets. By unanimity, N is victorious. From the definition, if K is victorious and $K \subseteq J$ then J is victorious.

Suppose J and K are both victorious and consider some situation $\{\prec_n\}_{n\in N}$ where $a \succ_n b$ for every $n \in J \cap K$. Choose a third outcome c (we assumed that a third outcome exists). Suppose voters change their minds about c to have preferences $\{\prec'_n\}_{n\in N}$ as follows. For each $n \in J \cap K$, we have $a \succ'_n c \succ'_n b$. For each $n \in J \setminus K$, we have $c \succ'_n b$. For each $n \in K \setminus J$, we have $a \succ'_n c$.

Then, since J is victorious and for every $n \in J$ we have $c \succ'_n b$, we must have $c \succ' b$. For every $n \in K$ we have $a \succ'_n c$, so $a \succ' c$. Therefore $a \succ' b$ and, by independence of irrelevant alternatives, also $a \succ b$. This works whenever $a \succ_n b$ for every $n \in J \cap K$, so $J \cap K$ is victorious.

We have shown that the victorious sets form a filter.*

Next, suppose K is *not* victorious. That means there is some list of preferences $\{\prec^-_n\}_{n\in N}$ so that $a \succ^-_n b$ for every $n \in K$, but $b \succ^- a$. We will show that $N \setminus K$ is victorious. Consider a third candidate c, and any situation $\{\prec_n\}_{n\in N}$ so that $b \succ_n c$ for every $n \in (N \setminus K)$. We will change the voter preferences about a to a third situation, $\{\prec'_n\}_{n\in N}$, which incorporates some information from $\{\prec^-_n\}_{n\in N}$: for every $n \in K$, a will become the voters first choice (so, in particular, both $a \succ'_n b$ and $a \succ'_n c$), while for every $n \in N \setminus K$, we will have $a \succ'_n c$ and will also have $a \succ'_n b$ if and only if $a \succ^-_n b$. The voters views of a and b in $\{\prec'_n\}_{n\in N}$ are the same as in $\{\prec^-_n\}_{n\in N}$, so, by independence of irrelevant alternatives, since $b \succ^- a$ also $b \succ' a$. For every $n \in N$ we have $a \succ'_n c$, so by unanimity, $a \succ' c$. Therefore $b \succ' c$.

Then, by independence of irrelevant alternatives, we must have $b \succ c$. Since this works whenever $c \prec_n b$ for every $n \in N \setminus K$, so $N \setminus K$ is victorious for b over c, and therefore is victorious.

Therefore the victorious sets have the "ultra" property. Therefore the victorious sets are an ultrafilter.† But, since N is finite, $N = \{n_1\} \cup \cdots \{n_k\}$ for some finite k, and therefore there is some i so that $\{n_i\} \in N$. But then n_i is a dictator: whenever $a \succ_{n_i} b$, we have $a \succ b$. □

Indeed, this argument shows that when N is not finite, if F satisfies the

*Though not necessarily a free filter—we have not promised that every cofinite set is victorious, and indeed, N is finite, so every set is cofinite.

†In the conventional definition, where we work over filters rather than free filters.

first two requirements then the victorious sets form an ultrafilter on N.

2.10 Remarks

We considered ultrafilters over $\mathbb{N}$—that is, ultrafilters which consist of subsets of $\mathbb{N}$—but we can consider an ultrafilter made out of subsets of any sets, or, more generally ultrafilters in any partially ordered set. This leads to more general ultraproducts whose ground models are indexed by sets other than $\mathbb{N}$.

If one replaces the last property of a free filter—that it contains all cofinite sets—with the weaker assumption that the collection is non-empty (and therefore contains $\mathbb{N}$), one obtains a filter. Adding the ultra property to a filter gives the conventional definition of an ultrafilter. The only difference is that the definition of an ultrafilter given in this chapter excludes the *principal ultrafilters*: for each n, the collection of all sets containing n is a principal ultrafilter. Since the principal ultrafilters are a degenerate case, we have simply excluded them from the definition rather than repeatedly specifying "non-principal" everywhere.*

The space of ultrafilters has a structure of its own, as a topological semi-group—indeed, with a discrete space like $\mathbb{N}$, the collection of ultrafilters coincides with the maximal compactification of the space, the *Stone-Čech* compactification. Ultrafilters have a number of direct (that is, without constructing an ultraproduct) applications in mathematics [16, 84], and it is an interesting open question whether all the uses of ultrafilters can be channeled through ultraproducts [42].

The ultraproduct construction applies to structures much more general than graphs. Indeed, the ultraproduct construction applies to a vast array of mathematical objects—it is usually defined for structures of first-order logic [49], which includes combinatorial structures like directed graphs and hypergraphs; algebraic structures like groups, rings, fields, and so on; and also models of set theory. Ultraproducts of the natural numbers and the reals have become a standard way to approach non-standard analysis [41, 70]. Ultraproducts also pay an important role in set theory [93, 105], where one can even consider ultraproducts of the entire universe of sets. The ultraproduct construction, with a suitable modification, also applies to

*To paraphrase an old joke, if the reader is ever stranded on a desert island, I recommend giving a talk to the ocean about ultrafilters without specifying that you mean non-principal ultrafilters. Someone will promptly arrive to remind you that, surely, you meant to specify the non-principal ultrafilters.

structures in continuous logic [14], including Banach spaces and C*-algebras, and normed spaces more generally [83]. (Indeed, one of the motivations for the development of continuous logic was to explain the already-observed existence of ultraproducts in non-first-order structures like Banach spaces.)

Chapter 3

Density as Probability

3.1 Density is not Cardinality

We now turn to making sure that for any H, the density t_H has the property that

$$t_H([G_n]_{\mathcal{U}}) = \lim_{n \to \mathcal{U}} t_H(G_n).$$

We have only even defined $t_H(G)$ when G is finite, so we need to figure out how to define the left side of this equation. First, we should notice that we *can't* just define $t_H([G_n]_{\mathcal{U}})$ as a ratio of cardinalities the way we did for finite graphs.

Consider a sequence of graphs with few (but not zero) edges: let G_n be the graph C_n consisting of n vertices arranged in a cycle. In a cycle, every vertex has exactly two neighbors, and this is a property that passes to the ultraproduct—observe that $[x_n]_{\mathcal{U}}$ has exactly two neighbors if and only if $\{n \mid x_n \text{ has exactly two neighbors}\} \in \mathcal{U}$.

For any fixed finite cycle k, $T_{\mathrm{C}_k}(\mathrm{C}_n) = \emptyset$ once $n > k$—that is, there are no small loops. So, for each k, $\{n \mid T_{\mathrm{C}_k}(\mathrm{C}_n) = \emptyset\} \in \mathcal{U}$, so $[G_n]_{\mathcal{U}}$ does not contain any copies of C_k for any k.

So each vertex in $G_{\mathcal{U}}$ has two neighbors, each of which has an additional neighbor, each of which has an additional neighbor, and since we cannot loop, we get infinite chains that look like ⋯●—●—●—●⋯. A chain like this has countably many vertices, so $[G_n]_{\mathcal{U}}$ must contain uncountably many of these chains. This gives a complete description of $[G_n]_{\mathcal{U}}$ as a graph: uncountably many chains which stretch forever in both directions, and therefore uncountably many edges (indeed, the same uncountable cardinality as the set of vertices).

However $\lim_{n\to\mathcal{U}} t_{K_2}(G_n) = 0$: the graph G_n has n vertices and n edges, so $t_{K_2}(G_n) = \frac{n}{n^2} = \frac{1}{n}$. So even though $[G_n]_\mathcal{U}$ has uncountably many edges, the "density" of these edges should still be 0.

On the other hand, if G'_n is the complete graph on n vertices, $[G'_n]_\mathcal{U}$ is a complete graph on uncountably many vertices. Since $t_{K_2}(G'_n) = 1$ for every n, the density of edges in $[G'_n]_\mathcal{U}$ should also be 1.

So $[G_n]_\mathcal{U}$ and $[G'_n]_\mathcal{U}$ have the same cardinality of both vertices and edges, but very different densities. Infinite cardinality is a much coarser notion than finite cardinality, and we will need are more refined notion to make sense of density on ultraproducts.

3.2 Probability on Internal Sets

Our approach will be to turn $[G_n]_\mathcal{U} = ([V_n]_\mathcal{U}, [E_n]_\mathcal{U})$ into a probability space. When V is a set, a probability measure is a function μ which assigns, to subsets of V, a probability in the interval $[0, 1]$.

When $X \subseteq [V_n]_\mathcal{U}$ is internal, there is a natural choice for the probability of X: since X is internal, $X = [X_n]_\mathcal{U}$ and we should have

$$\mu(X) = \lim_{n\to\mathcal{U}} \frac{|X_n|}{|V_n|}.$$

To make this even more explicit, recall that in Section 1.3 we defined a counting measure: when $S \subseteq V_n$, let us write $\mu_n(S) = \frac{|S|}{|V_n|}$ for the counting measure on the n-th ground model. Then our definition is simply

$$\mu(X) = \lim_{n\to\mathcal{U}} \mu_n(X_n).$$

This is a coordinate-wise definition, so we should make sure it doesn't depend on the representation of X we choose.

Lemma 3.1. *If $[X_n]_\mathcal{U} = [Y_n]_\mathcal{U}$ then*

$$\lim_{n\to\mathcal{U}} \mu_n(X_n) = \lim_{n\to\mathcal{U}} \mu_n(Y_n).$$

Proof. If $[X_n]_\mathcal{U} = [Y_n]_\mathcal{U}$ then there is a $K \subseteq \mathcal{U}$ so that, for all $n \in K$, $X_n = Y_n$. Therefore, for each $n \in K$, $\mu_n(X_n) - \mu(Y_n) = 0$, and therefore $\lim_{n\to\mathcal{U}}(\mu_n(X_n) - \mu(Y_n)) = 0$, so also

$$\lim_{n\to\mathcal{U}} \mu_n(X_n) = \lim_{n\to\mathcal{U}} \mu_n(Y_n).$$

□

For example, suppose that each G_n is a complete bipartite graph where the sides have *almost* the same size: take $G_n = K_{n,n+1} = (V_n, E_n)$, where $|V_n| = 2n + 1$ and we have $V_n = X_n \cup Y_n$ where $|X_n| = n + 1$ and $|Y_n| = n$, and $E_n = \{\{x, y\} \mid x \in X_n \text{ and } y \in Y_n\}$. Then, as in the previous chapter, $[E_n]_{\mathcal{U}}$ will be a complete bipartite graph on $[V_n]_{\mathcal{U}}$. $[X_n]_{\mathcal{U}}$ and $[Y_n]_{\mathcal{U}}$ are subsets of $[V_n]_{\mathcal{U}}$ and

$$\mu([X_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} \mu_n(X_n) = \lim_{n\to\mathcal{U}} \frac{|X_n|}{|V_n|} = \lim_{n\to\mathcal{U}} \frac{n+1}{2n+1} = \frac{1}{2}.$$

In the finite graphs $K_{n,n+1}$, X_n is *approximately* half the vertices (at least when n is large), but in the ultraproduct, the measure of $[X_n]_{\mathcal{U}}$ is *exactly* one half. In Chapter 1, dealing with the finite case, we sometimes noted that we were disregarding error terms that were sufficiently small—specifically, erros which were less than $\epsilon|V_n|$ for each $\epsilon > 0$ (and when n was much larger than $1/\epsilon$). In the ultraproduct, these error terms are literally 0.

This demonstrates that the relationship between statements about measure in the ground models and the corresponding statements in ultraproducts is slightly more complicated than the straightforward relationship given by first-order formulas. In this example, $\{n \mid \mu_n(X_n) > 1/2\} = \mathbb{N} \in \mathcal{U}$, but $\mu([X_n]_{\mathcal{U}}) \leq 1/2$.

However this is the most that $\mu([X_n]_{\mathcal{U}})$ can deviate from the behavior of the $\mu_n(X_n)$:

Lemma 3.2. *Whenever $X \subseteq [V_n]_{\mathcal{U}}$ is internal, taking $X = [X_n]_{\mathcal{U}}$ we have:*

- *if $\{n \mid \mu_n(X_n) \leq c\} \in \mathcal{U}$ then $\mu(X) \leq c$,*
- *if $\mu(X) \leq c$ then, for every $\epsilon > 0$, $\{n \mid \mu_n(X_n) < c + \epsilon\} \in \mathcal{U}$,*
- *if $\{n \mid \mu_n(X_n) \geq c\} \in \mathcal{U}$ then $\mu(X) \geq c$,*
- *if $\mu(X) \geq c$ then, for every $\epsilon > 0$, $\{n \mid \mu_n(X_n) > c - \epsilon\} \in \mathcal{U}$.*

All parts of the lemma follow from the definition of $\lim_{n\to\mathcal{U}}$.

It will be convenient, once in a while, to have approximations of μ which are internal.

Definition 3.3. $\mu([X_n]_{\mathcal{U}}) \gtrsim \epsilon$ holds exactly when $\{n \mid \mu(X_n) \geq \epsilon\} \in \mathcal{U}$.

The lemma above then says that

$$\mu([X_n]_{\mathcal{U}}) > \epsilon \;\Rightarrow\; \mu([X_n]_{\mathcal{U}}) \gtrsim \epsilon \;\Rightarrow \mu([X_n]_{\mathcal{U}}) \geq \epsilon.$$

μ inherits many of the rules we expect of probability from the ground models:

Lemma 3.4. *If X and Y are internal sets then $\mu(X \cup Y) = \mu(X) + \mu(Y) - \mu(X \cap Y)$.*

Proof. Pick representatives $X = [X_n]_{\mathcal{U}}$ and $Y = [Y_n]_{\mathcal{U}}$. Observe that $X \cup Y = [X_n \cup Y_n]_{\mathcal{U}}$:

$$\begin{aligned}[v_n]_{\mathcal{U}} \in X \cup Y &\text{ if and only if } [v_n]_{\mathcal{U}} \in X \text{ or } [v_n]_{\mathcal{U}} \in Y \\ &\text{if and only if } \{n \mid v_n \in X_n\} \in \mathcal{U} \text{ or } \{n \mid v_n \in Y_n\} \in \mathcal{U} \\ &\text{if and only if } \{n \mid v_n \in X_n \text{ or } v_n \in Y_n\} \in \mathcal{U} \\ &\text{if and only if } [v_n]_{\mathcal{U}} \in [X_n \cup Y_n]_{\mathcal{U}}.\end{aligned}$$

Similarly, $X \cap Y = [X_n \cap Y_n]_{\mathcal{U}}$:

$$\begin{aligned}[v_n]_{\mathcal{U}} \in X \cup Y &\text{ if and only if } [v_n]_{\mathcal{U}} \in X \text{ and } [v_n]_{\mathcal{U}} \in Y \\ &\text{if and only if } \{n \mid v_n \in X_n\} \in \mathcal{U} \text{ and } \{n \mid v_n \in Y_n\} \in \mathcal{U} \\ &\text{if and only if } \{n \mid v_n \in X_n \text{ and } v_n \in Y_n\} \in \mathcal{U} \\ &\text{if and only if } [v_n]_{\mathcal{U}} \in [X_n \cap Y_n]_{\mathcal{U}}.\end{aligned}$$

Therefore

$$\begin{aligned}\mu(X \cup Y) &= \lim_{n \to \mathcal{U}} \mu_n(X_n \cup Y_n) \\ &= \lim_{n \to \mathcal{U}} (\mu_n(X_n) + \mu_n(Y_n) - \mu_n(X_n \cap Y_n)) \\ &= \lim_{n \to \mathcal{U}} \mu_n(X_n) + \lim_{n \to \mathcal{U}} \mu_n(Y_n) - \lim_{n \to \mathcal{U}} \mu_n(X_n \cap Y_n) \\ &= \mu(X) + \mu(Y) - \mu(X \cap Y).\end{aligned}$$

□

More generally, we have finite additivity.

Corollary 3.5. *Suppose $B^1, \ldots, B^k$ are pairwise disjoint internal sets. Then $\mu(\bigcup_{i \leq k} B^i) = \sum_{i \leq k} \mu(B^i)$.*

Proof. By induction on k, using the previous lemma □

The conventional setting for probability theory is a *σ-algebra*, in which we have not only finite unions and intersections, but countable ones.

Definition 3.6. We say $\mathcal{B}$ is a *σ-algebra* if $\mathcal{B}$ is an algebra and, additionally, whenever $B_i \in \mathcal{B}$ for every $i \in \mathbb{N}$, $\bigcup_{i \in \mathbb{N}} B_i \in \mathcal{B}$.

When $\mathcal{B} \subseteq \mathcal{P}(V)$ is a σ-algebra, a *probability measure* on $(V, \mathcal{B})$ is a function $\mu : \mathcal{B} \to [0, 1]$ such that:

- $\mu(V) = 1$,
- if $B_i \in \mathcal{B}$ for each i and $B_i \cap B_j = \emptyset$ whenever $i \neq j$ then $\mu(\bigcup_{i\in I} B_i) = \sum_{i\in I} \mu(B_i)$.

A *probability measure space* is a triple $(V, \mathcal{B}, \mu)$ where $\mathcal{B} \subseteq \mathcal{P}(V)$ is a σ-algebra and μ is a probability measure.

Since σ-algebras are closed under complements and countable unions, they are also closed under countable intersections.

We saw in the previous chapter that the union or intersection of countably many internal sets is generally *not* an internal set, so if we want to do probability theory properly, we will need to extend μ beyond the internal sets. As a first step, we can show that μ satisfies the countable additivity it is supposed to have as long as the countable union happens to be internal.

Lemma 3.7. *Suppose that, for each i, $B^i \in \mathcal{B}$ is internal and $B^i \cap B^j = \emptyset$ whenver $i \neq j$. If $\bigcup_{i\in\mathbb{N}} B^i$ is internal then $\mu(\bigcup_{i\in\mathbb{N}} B^i) = \sum_{i=1}^{\infty} \mu(B^i)$.*

Proof. In the previous chapter, we proved Theorem 2.39: if $\bigcup_{i\in\mathbb{N}} B^i$ is internal then $\bigcup_{i\in\mathbb{N}} B^i = \bigcup_{i\leq k} B^i$ for some finite k. Since the B^i are pairwise disjoint, that means $B^i = \emptyset$ for $i > k$ and therefore $\mu(B^i) = 0$ for $i > k$. Therefore this follows from finite additivity. □

This makes μ a *pre-measure* on the internal sets: it has the properties a probability measure should have as long as the sets involved are internal.

Although the union of countably many internal sets is not internal, we can get close to a countable union: the union of countably many internal sets is contained in a set with the same measure as the sum.

Lemma 3.8. *If each A^i is internal then there is an internal set A^+ such that:*

- $\bigcup_{i\in\mathbb{N}} A^i \subseteq A^+$, *and*
- $\mu(A^+) = \lim_{m\to\infty} \mu(\bigcup_{i\leq m} A^i)$.

Proof. Name the target measure $c = \lim_{m\to\infty} \mu(\bigcup_{i\leq m} A^i)$

We first show that we can prove the statement up to ϵ: for each $\epsilon > 0$, we will define an internal set $A^{+\epsilon}$ so that $\bigcup_{i\in\mathbb{N}} A^i \subseteq A^{+\epsilon}$ and $\mu(A^{+\epsilon}) \leq c + \epsilon$. Define a function f^ϵ by setting $f^\epsilon(n)$ to be the largest m so that $\frac{|\bigcup_{i\leq m} A^i_n|}{|V_n|} \leq c+\epsilon$ and set $A^{+\epsilon} = A^{f^\epsilon}$—that is, $A^{+\epsilon} = [\bigcup_{i\leq f^\epsilon(n)} A^i_n]_{\mathcal{U}}$. Since $\frac{|\bigcup_{i\leq f^\epsilon(n)} A^i_n|}{|V_n|} \leq$

$c+\epsilon$ for each n, certainly $\mu(A^{+\epsilon}) \leq c+\epsilon$. Since $\lim_{m\to\infty} \mu(\bigcup_{i\leq m} A^i) = c$, for each m, $\mu(\bigcup_{i\leq m} A^i) \leq c < c+\epsilon$, and therefore $\{n \mid \frac{|\bigcup_{i\leq m} A^i_n|}{|V_n|} \leq c+\epsilon\} \in \mathcal{U}$, so $\{n \mid A^i_n \subseteq \bigcup_{i\leq f^\epsilon(n)} A^i_n\} \in \mathcal{U}$, and therefore $A^i \subseteq A^{+\epsilon}$. Similarly, when $\delta < \epsilon$, $A^{+\delta} \subseteq A^{+\epsilon}$.

Now consider the sequences

$$A^1 \subseteq \bigcup_{i\leq 2} A^i \subseteq \bigcup_{i\leq 3} A^i \subseteq \cdots \subseteq A^{+1/3} \subseteq A^{+1/2} \subseteq A^{+1}.$$

Then by Theorem 2.42, there is an internal A^+ with $\bigcup_{i\in\mathbb{N}} A^i \subseteq A^+$, and $A^+ \subseteq A^{+1/i}$ for all i. Therefore $\mu(A^+) \leq c + 1/i$ for all i, so $\mu(A^+) = c$. □

We will usually apply this lemma when the sets A^i are pairwise disjoint, in which case $\lim_{m\to\infty} \mu(\bigcup_{i\leq m} A^i) = \sum_i \mu(A_i)$.

There is one further property of finite sets that we will want μ to share. When A and B are finite sets with the same cardinality, we also expect A and B to have the same measure. In the ultraproduct setting, we have already seen that this is too much to expect.

Instead, we should re-examine what we mean by cardinality: two sets have the same cardinality if there is a bijection between them. In the ultraproduct, we should only expect the measure to respect *internal* bijections.

Lemma 3.9. *If A and B are internal sets and $f : A \to B$ is an internal bijection then $\mu(A) = \mu(B)$.*

It will be convenient later to have the more general form of this:

Lemma 3.10. *If $R \subseteq A \times B$ is an internal relation such that*

- *for each $a \in A$, $|\{b \mid (a,b) \in R\}| = k$,*
- *for each $b \in B$, $|\{a \mid (a,b) \in R\}| = m$,*

then $k\mu(A) = m\mu(B)$.

Then the graph of a bijection is exactly such a relation with $k = m = 1$. We do not need the full generality, but we will need, for instance, the case where $k = 2$ and $m = 1$, saying that each element of A is matched to exactly two elements of B. Then the measures behave the way we expect: if each element of A is matched to two elements of B then $\mu(B) = 2\mu(A)$.

Proof. We have $A = [A_n]_{\mathcal{U}}$, $B = [B_n]_{\mathcal{U}}$, and $R = [R_n]_{\mathcal{U}}$. Since the properties "for every a, $|\{b \mid (a,b) \in R\}| = k$" and "for every b, $|\{a \mid (a,b) \in R\}| = m$" are first-order, we know that the set of n so that $R_n \subseteq A_n \times B_n$ has these properties is in $\mathcal{U}$.

For any such n, we have $k|A_n| = m|B_n|$, so $k\mu(A_n) = m\mu(B_n)$. Since this holds for a set of n belonging to $\mathcal{U}$, we also have $k\mu(A_n) = m\mu(B_n)$. □

This makes sense of the difference between $\mathrm{K}_{\infty,\infty}$ and $\mathrm{K}_{\infty,\infty^2}$ we saw in Section 2.6: these structures both uncountable complete bipartite graphs, and are therefore isomorphic, but they have different internal sets. In $\mathrm{K}_{\infty,\infty}$, there is an internal bijection between the two parts, so for instance each part has measure $1/2$. On the other hand, in $\mathrm{K}_{\infty,\infty^2}$, for every m there is an m-to-1 relation between the two parts, so the left part has measure 0 while the right part has measure 1.

3.3 Probability Spaces

What about non-internal sets? We cannot hope to extend μ to make sense for all the subsets of $[V_n]_{\mathcal{U}}$: just as in the more familiar case of the Lebesgue measure on $[0,1]$, there is not a way to consistently define a probability on every single subset of $[V_n]_{\mathcal{U}}$. There will have to be some unmeasurable sets which are not in the domain of μ.

There is one class of sets we can immediately extend our measure to: if A is an internal set with $\mu(A) = 0$ then certainly we could say $\mu(B) = 0$ for every subset $B \subseteq A$, internal or not. Slightly more generally, a set might be contained in arbitrarily small internal sets:

Definition 3.11. We say $B \subseteq [V_n]_{\mathcal{U}}$ is *μ-null* if, for every $\epsilon > 0$, there is an internal set A_ϵ such that $B \subseteq A_\epsilon$ and $\mu(A_\epsilon) < \epsilon$.

It is almost immediate that a finite union of μ-null sets is also μ-null. Less trivally, even a countable union of μ-null sets is μ-null.

Lemma 3.12. *If each B_i is μ-null then so is $\bigcup_{i\in\mathbb{N}} B_i$.*

Proof. Let $\epsilon > 0$ be given. For each i, choose an internal set $A_i \supseteq B_i$ so that $\mu(A_i) < \epsilon \cdot 2^{-i}$. Let $A'_i = A_i \setminus \bigcup_{j<i} A'_j$, so the A'_i are pairwise disjoint and $\bigcup_{i\in\mathbb{N}} A'_i = \bigcup_{i\in\mathbb{N}} A_i$.

Then by Lemma 3.8 there is an internal set $A^+ \supseteq \bigcup_{i\in\mathbb{N}} A_i \supseteq \bigcup_{i\in\mathbb{N}} B_i$ with $\mu(A^+) = \sum_i \mu(A'_i) \leq \sum_i \mu(A_i) < \epsilon$. □

More generally, we could extend μ to any set which is within a null set of an internal set:

Definition 3.13. We say $B \subseteq [V_n]_{\mathcal{U}}$ is *μ-approximable* if there is an internal set A so that $A \triangle B$ is μ-null. We call A a *μ-approximation of B*.

We write $\mathcal{B}(\mu)$ for the set of μ-approximable sets. We define $\mu : \mathcal{B}(\mu) \to [0,1]$ by $\mu(B) = \mu(A)$ where A is any μ-approximation of B.

Note that this definition gives a unique value for the measure of each μ-approximable set: if A and A' are two μ-approximations of B then $A \triangle A' \subseteq (A \triangle B) \cup (A' \triangle B)$ is an internal μ-null set, and therefore has measure 0, so $\mu(A) = \mu(A')$.

The μ-approximable sets are a suitable family of sets to use as the basis for probability theory.

Theorem 3.14. *The μ-approximable sets are a σ-algebra.*

Proof. $\emptyset$ and V are internal, so certainly μ-approximable. If B is μ-approximable, so $A \triangle B$ is μ-null then $V \setminus A$ is internal as well, and $(V \setminus A) \triangle (V \setminus B) = A \triangle B$ is μ-null, so $V \setminus B$ is μ-approximable.

Suppose B_i is μ-approximable for $i \in \mathbb{N}$, so for each i we have an internal set A_i so $A_i \triangle B_i$ is μ-null. Let $A'_i = A_i \setminus \bigcup_{j<i} A_j$, so $\bigcup_{i\in\mathbb{N}} A'_i = \bigcup_{i\in\mathbb{N}} A_i$ and the A'_i are pairwise disjoint. Then by Lemma 3.8, we have a set $A^+ \supseteq \bigcup_{i\in\mathbb{N}} A_i$ with $\mu(A^+) = \sum_i \mu(A'_i)$.

Then the internal sets $A^+ \setminus \bigcup_{j<i} A'_i$ must have measures approaching 0, so $A^+ \setminus \bigcup_{i\in\mathbb{N}} A_i$ is μ-null.

Then

$$(A^+ \triangle \bigcup_{i\in\mathbb{N}} B_i) \subseteq (A^+ \setminus \bigcup_{i\in\mathbb{N}} A_i) \cup \bigcup_i (A_i \triangle B_i).$$

This is a countable union of μ-null sets, so is also μ-null.

Since $\bigcap_{i\in\mathbb{N}} = V \setminus (\bigcup_{i\in\mathbb{N}}(V \setminus B_i))$, $\bigcap_{i\in\mathbb{N}} B_i$ is μ-approximable as well. □

Theorem 3.15. *μ is a probability measure on $\mathcal{B}(\mu)$.*

Proof. The only thing to check is countable additivity. Suppose each B_i is μ-approximable and $B_i \cap B_j = \emptyset$ whenever $i \neq j$. For each B_i, fix a μ-approximation A_i. Observe that when $i \neq j$, $\mu(A_i \cap A_j) = 0$: since $B_i \cap B_j = \emptyset$, $A_i \cap A_j \subseteq (A_i \triangle B_i) \cup (A_j \triangle B_j)$, and is therefore μ-null.

So let $A'_i = A_i \setminus \bigcup_{j<i} A_j$, so $\mu(A'_i) = \mu(A_i)$. Choose $A^+ \supseteq \bigcup_{i\in\mathbb{N}} A_i$ with $\mu(A^+) = \sum_i \mu(A'_i) = \sum_i \mu(A_i)$. Since

$$A^+ \triangle \bigcup_i B_i \subseteq (A^+ \setminus \bigcup_{i\in\mathbb{N}} A_i) \cup \bigcup_{i\in\mathbb{N}} (A_i \triangle B_i)$$

is μ-null, A^+ is a μ-approximation of $\bigcup_{i\in\mathbb{N}} B_i$, so

$$\mu(\bigcup_{i\in\mathbb{N}} B_i) = \mu(A^+) = \sum_i \mu(A_i) = \sum_i \mu(B_i).$$

□

The μ-approximable sets are naturally *complete*: any subset of a measure 0 set is measurable.

Lemma 3.16. *Suppose $X \subseteq Y$, $Y \in \mathcal{B}(\mu)$, and $\mu(Y) = 0$. Then $X \in \mathcal{B}(\mu)$ and $\mu(X) = 0$.*

Proof. First, we show that the empty set is a μ-approximation of Y. Since $Y \in \mathcal{B}(\mu)$ with measure 0, there is an internal set A with $\mu(A) = 0$ so that $Y \triangle A$ is μ-null. Then, for each $\epsilon > 0$, there is a B_ϵ with $\mu(B_\epsilon) < \epsilon$ and $Y \triangle A \subseteq B_\epsilon$. But since $\mu(A) = 0$, $\mu(B_\epsilon \cup A) < \epsilon$ and $(Y \triangle \emptyset) = Y \subseteq B_\epsilon \cup A$.

But then $(X \triangle \emptyset) = X \subseteq B_\epsilon \cup A$ as well, so $\emptyset$ is a μ-approximation of X, so $\mu(X) = 0$. □

Corollary 3.17. *If $X \subseteq [V_n]_{\mathcal{U}}$ and, for every $\epsilon > 0$, there is a $Y_\epsilon \in \mathcal{B}(\mu)$ with $X \subseteq Y_\epsilon$ and $\mu(Y_\epsilon) < \epsilon$ then $X \in \mathcal{B}(\mu)$ and $\mu(X) = 0$.*

Proof. If the Y_ϵ were internal, this would be immediate because X would be μ-null. Instead, note that $X \subseteq \bigcap_d Y_{1/d} = Y$ and, since $\mathcal{B}(\mu)$ is a σ-algebra, $Y \in \mathcal{B}(\mu)$ and $\mu(Y) = 0$. Therefore $X \in \mathcal{B}(\mu)$ and $\mu(X) = 0$. □

3.4 Probability Spaces on $[V_n]^k_{\mathcal{U}}$

The sets whose densities we're interested in are not just subsets of $[V_n]_{\mathcal{U}}$ itself, but also subsets of $[V_n]^k_{\mathcal{U}}$ for various k. For instance, we are interested in the measure of $\{(x, y) \mid \{x, y\} \in [E_n]_{\mathcal{U}}\} \subseteq [V_n]^2_{\mathcal{U}}$, or the measure of $\{(x, y, z) \mid \{x, y\} \in [E_n]_{\mathcal{U}}$ and $\{x, z\} \in [E_n]_{\mathcal{U}}$ and $\{y, z\} \in [E_n]_{\mathcal{U}}\} \subseteq [V_n]^3_{\mathcal{U}}$ (that is, the density of the set of triangles).

The sets $[V_n]^k_{\mathcal{U}}$ are themselves ultraproducts—$[V_n]^k_{\mathcal{U}}$ is the same as $[V^k_n]_{\mathcal{U}}$. That is, we can forget about the fact that the elements of V^k_n happen to be tuples: taking $W_n = V^k_n$, we can apply the previous section to $[W_n]_{\mathcal{U}} = [V^k_n]_{\mathcal{U}} = [V_n]^k_{\mathcal{U}}$.

Definition 3.18. For any $k \in \mathbb{N}$, we define μ_k on internal subsets $X = [X_n]_{\mathcal{U}} \subseteq [V_n]^k_{\mathcal{U}}$ by

$$\mu_k(X) = \lim_{n\to\mathcal{U}} (\mu_k)_n(X_n) = \lim_{n\to\mathcal{U}} \frac{|X_n|}{|V_n|^k}.$$

A set $B \subseteq [V_n]^k_{\mathcal{U}}$ is μ_k-*null* if, for every $\epsilon > 0$, there is an internal set A_ϵ such that $B \subseteq A_\epsilon$ and $\mu_k(A_\epsilon) < \epsilon$.

A set $B \subseteq [V_n]^k_{\mathcal{U}}$ is μ_k-*approximable* if there is an internal set A such that $A \triangle B$ is μ-null, and we call A a μ_k-*approximable* of B.

We write $\mathcal{B}(\mu_k)$ for the set of μ_k-approximable sets and define $\mu_k : \mathcal{B}(\mu_k) \to [0,1]$ by $\mu_k(B) = \mu_k(A)$ where A is any μ_k-approximation of B.

As in the previous section, we have

Theorem 3.19. *$\mathcal{B}(\mu_k)$ is a σ-algebra and μ_k is a probability measure on $\mathcal{B}(\mu_k)$.*

However the spaces $[V_n]^k_{\mathcal{U}}$ are more interesting than just $[V_n]_{\mathcal{U}}$ by itself, because these spaces are related to each other—the measure on $[V_n]^2_{\mathcal{U}}$ has something to do with the measure on $[V_n]_{\mathcal{U}}$.

Indeed, there is a second way we could have tried to define μ_2: by an integral. For instance, when $X \subseteq [V_n]^2_{\mathcal{U}}$, we could also evaluate

$$\int \mu(\{y \mid (x,y) \in X\})\, d\mu(x).$$

Of course, when X is not symmetric, we have to worry about the possibility that integrating in the other order

$$\int \mu(\{x \mid (x,y) \in X\})\, d\mu(y)$$

might give a different value.

These integrals mean something different than the simple measure $\mu_2(X)$. The measure μ_2 is looking at the set of pairs X as simply an unstructured set, ignoring the fact that its elements are pairs. For instance, when $X = [X_n]_{\mathcal{U}}$, $\mu_2(X) = \lim_{n \to \mathcal{U}} \frac{|X_n|}{|V_n^2|}$. On the other hand, the integral is calculating the average, across all values of x, of the measure of the set of neighbors of x.

Another way to look at this is to think of measuring a set by randomly sampling points. μ_2 corresponds to the uniform measure on the set of pairs: we determine $\mu_2(X)$ by randomly choosing pairs of points, with each pair equally likely, and checking whether that pair belongs to X. $\int \mu(\{x \mid (x,y) \in X\})\, d\mu(x)$ corresponds to first choosing a value of x randomly, with each possible x equally likely, and only after choosing a point y and checking whether y belongs to the set $\{x \mid (x,y) \in X\}$.

Despite the different meanings, you may have a strong intuition that these should be the same, which we should be a bit careful about. The fact that these are the same for Lebesgue measure is Fubini's Theorem, a familiar

fact from multi-variable calculus. But Fubini's Theorem does not apply here: Fubini's Theorem only applies to the *product* of probability measure spaces, and we will see later that $([V_n]^k_{\mathcal{U}}, \mathcal{B}(\mu_k), \mu_k)$ is *not* a product space.

These different interpreations *are* still the same here, but for a different reason. This distinction is important, because we will be interested in the product spaces, which we will define precisely later. In fact, the distinction between the product space and $([V_n]^k_{\mathcal{U}}, \mathcal{B}(\mu_k), \mu_k)$ is a central concern: we will discover that the sets which are measurable in the sense of the product space are exactly the "non-random" sets, and that $\mathcal{B}(\mu_k)$ contains additional measurable sets which are random.

For the moment, however, we want to understand how the spaces $([V_n]^k_{\mathcal{U}}, \mathcal{B}(\mu_k), \mu_k)$ relate to each other. First, we need to consider several ways we could use the description of a set by coordinates to define other sets.

Definition 3.20. If $X \subseteq V^r$ and $\pi : [1, r] \to [1, r]$ is a bijection then

$$X^\pi = \{(x_1, \ldots, x_r) \mid (x_{\pi(1)}, \ldots, x_{\pi(r)}) \in X\} \in \mathcal{B}_r.$$

If $k \leq r$ and $x_1, \ldots, x_k \in V$, the *slice of* X corresponding to $x_1, \ldots, x_k$ is

$$X_{x_1,\ldots,x_k} = \{(x_{k+1}, \ldots, x_r) \in V^{r-k} \mid (x_1, \ldots, x_k, x_{k+1}, \ldots, x_r) \in X\}.$$

We can now write down an abstract description of the properties we would expect a family of probability measure spaces to have.

Definition 3.21. Let V be a set and suppose that, for each k, we have a probability measure space $(V^k, \mathcal{B}_k, \mu_k)$. The spaces $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ are a *Keisler graded probability space* if:

- (Symmetry) whenever $\pi : [1, k] \to [1, k]$ is a bijection and $X \in \mathcal{B}_k$,
  - $X^\pi \in \mathcal{B}_k$, and
  - $\mu_k(X^\pi) = \mu_k(X)$,
- (Products) whenever $B \in \mathcal{B}_k$ and $C \in \mathcal{B}_r$, $B \times C \in \mathcal{B}_{k+r}$,
- (Fubini Property) whenever $X \in \mathcal{B}_{k+r}$
  - the set of $x_1, \ldots, x_k \in V$ such that $X_{x_1,\ldots,x_k} \in \mathcal{B}_r$ belongs to $\mathcal{B}_k$ and has μ_k-measure 1, and
  - $\mu_{k+r}(X) = \int \mu_r(X_{x_1,\ldots,x_k}) d\mu_k(x_1, \ldots, x_k)$.

These properties seem technical, but they are exactly the properties we would expect a family of probability measures on V^k to have. The first two properties guarantee that there are enough measurable sets.

However we are allowed to—as we typically will—have additional sets beyond those required by the product property. The Fubini property puts a restriction on these new sets, requiring that these additional sets be assigned measures in a way that is consistent with the measures assigned to sets of lower arity: their slices have to exist in the lower arity σ-algebras, and the measure must be the one obtained by integrating over the slices.

In particular, we can calculate measures by integrating over one of the coordinates. For example, let $T_{\mathrm{C}_3} = \{(x, y, z) \mid \{x, y\} \in E \text{ and } \{x, z\} \in E \text{ and } \{y, z\} \in E\}$ be the set of triangles. Then $(T_{\mathrm{C}_3})_x = E \cap (E_x \times E_x)$ is the set of pairs (y, z) which are both neighbors of x and are also neighbors of each other (so that (x, y, z) is a triangle), and, when $\{x, y\} \in E$, $(T_{\mathrm{C}_3})_{(x,y)} = E_x \cap E_y$.

If we want to find the measure of the set of triangles, $\mu_3(T_{\mathrm{C}_3})$, we can calculate either $\int \mu_2((T_{\mathrm{C}_3})_x) d\mu(x)$ or $\int \mu((T_{\mathrm{C}_3})_{(x,y)}) d\mu_2(x, y)$ if one of these is easier to calculate. Indeed, we will routinely find ourselves switching between different integrals which calculate the same measure.

The Fubini property is only stated for a single order of integration, but symmetry allows us to rearrange the coordinates to consider other orders of integration. For instance, if X is a set of pairs (x_1, x_2) and π is the permutation swapping 1 and 2, we have

$$\mu_2(X) = \mu_2(X^\pi) = \int \mu_1(X^\pi_{x_2})\, d\mu_1(x_2) = \int \mu_1(X_{x_1})\, d\mu_1(x_2).$$

We also note that having the Fubini property for measures is enough to obtain the same for integrals:

Lemma 3.22. *If $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ is a Keisler graded probability space then for any measurable function on V^{k+r},*

$$\int f\, d\mu_{k+r} = \int \left(\int f(x_1, \ldots, x_{k+r})\, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right) d\mu_k(x_1, \ldots, x_k).$$

Proof. For any $\epsilon > 0$, we may choose a simple function $g = \sum_{i\le d} c_i \chi_{B_i}$ with

$|f(x_1,\ldots,x_{k+r}) - g(x_1,\ldots,x_{k+r})| < \epsilon$ for almost all $x_1,\ldots,x_{k+r}$. Therefore

$$\begin{aligned}
\int f\,d\mu_{k+r} &= \int g\,d\mu_{k+r} + e\ |e| < \epsilon \\
&= \sum_{i\leq d} c_i \mu_{k+r}(B_i) + e \text{ where } |e| < \epsilon \\
&= \sum_{i\leq d} c_i \int \mu_r((B_i)_{x_1,\ldots,x_k})\,d\mu_k + e \text{ where } |e| < \epsilon \\
&= \int \sum_{i\leq d} c_i \mu_r((B_i)_{x_1,\ldots,x_k})\,d\mu_k + e \text{ where } |e| < \epsilon \\
&= \int \left(\int f(x_1,\ldots,x_{k+r})\,d\mu_r(x_{k+1},\ldots,x_{k+r}) \right) d\mu_k(x_1,\ldots,x_k) + e' \text{ where } |e'| < 2\epsilon.
\end{aligned}$$

Since this holds for every $\epsilon > 0$,

$$\int f\,d\mu_{k+r} = \int \left(\int f(x_1,\ldots,x_{k+r})\,d\mu_r(x_{k+1},\ldots,x_{k+r}) \right) d\mu_k(x_1,\ldots,x_k).$$

□

The key fact we need is that the natural measure spaces on the ultraproduct form a Keisler graded probability space. The proof is routine, but long and technical, so we finish the section with it.

Theorem 3.23. *The collection of probability measure spaces $([V_n]^k_{\mathcal{U}}, \mathcal{B}(\mu_k), \mu_k)$ is a Keisler graded probability space.*

Proof. The approach is that, for each property, we first show that the property holds for internal sets, and then lift that to all of $\mathcal{B}(\mu_k)$ by looking at an approximation.

First we show symmetry for internal sets. Let $A = [A_n]_{\mathcal{U}}$ be internal. For any bijection $\pi : [1,k] \to [1,k]$, since $A^\pi = [A^\pi_n]_{\mathcal{U}}$ and $(\mu_k)_n(A^\pi_n) = (\mu_k)_n(A_n)$, we have $\mu(A^\pi) = \mu(A)$.

Next, consider an arbitrary $X \in \mathcal{B}^k$ and take a μ_k-approximation A. Then, for any $\epsilon > 0$, we have an internal set A_ϵ with $\mu_k(A_\epsilon) < \epsilon$ and $(A \triangle X) \subseteq A_\epsilon$. But also $\mu_k(A^\pi_\epsilon) = \mu_k(A_\epsilon) < \epsilon$ and $(A^\pi \triangle X^\pi) \subseteq A^\pi_\epsilon$, so A^π is a μ_k-approximation to X^π, so $X^\pi \in \mathcal{B}^k$ and $\mu_k(X^\pi) = \mu_k(A^\pi) = \mu_k(A) = \mu_k(X)$.

Products of internal sets are certainly internal: if $A = [A_n]_{\mathcal{U}} \subseteq [V_n]^k_{\mathcal{U}}$ and $B = [B_n]_{\mathcal{U}} \subseteq [V_n]^r_{\mathcal{U}}$ are internal then $A \times B = [A_n \times B_n]_{\mathcal{U}}$.

Suppose that, $X \in \mathcal{B}_k$ and $Y \in \mathcal{B}_r$. Consider a μ_k-approximation A of X and μ_r-approximation B of Y. Then for any $\epsilon > 0$, we may choose internal

sets $A_\epsilon \supseteq (A \triangle X)$ and $B_\epsilon \supseteq (B \triangle Y)$ with $\mu_k(A_\epsilon) < \epsilon/2$ and $\mu_r(B_\epsilon) < \epsilon/2$. Then

$$(A \times B) \triangle (X \times Y) \subseteq [(A \triangle X) \times [V_n]^r_{\mathcal{U}}] \cup \Big[[V_n]^k_{\mathcal{U}} \times (B \triangle Y)\Big]$$
$$(A_\epsilon \times [V_n]^r_{\mathcal{U}}) \cup ([V_n]^k_{\mathcal{U}} \times B_\epsilon),$$

and $\mu_{k+r}((A_\epsilon \times [V_n]^r_{\mathcal{U}}) \cup ([V_n]^k_{\mathcal{U}} \times B_\epsilon)) < \epsilon$.

Finally, we turn to the Fubini property. Once again, we start with internal sets. When $A = [A_n]_{\mathcal{U}} \in \mathcal{B}^{k+r}$ is internal, $A_{x_1,\ldots,x_k}$ is also internal for any choice of $x_1, \ldots, x_k$, and therefore is in $\mathcal{B}^r$.

The ground graphs are finite, so the analogous statements hold by counting:

$$\begin{aligned}\int (\mu_r)_n((A_n)_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1,\ldots,x_k) &= \frac{1}{|V_n|^k} \sum_{(x_1,\ldots,x_k)\in V_k^n} \frac{|(A_n)_{x_1,\ldots,x_k}|}{|V_n|^r} \\ &= \frac{1}{|V_n|^{k+r}} \sum_{(x_1,\ldots,x_k)\in V_k^n} |(A_n)_{x_1,\ldots,x_k}| \\ &= \frac{|A_n|}{|V_n|^{k+r}} = \mu_n^{k+r}(A_n).\end{aligned}$$

Moreover, we have $\lim_{n\to\mathcal{U}} \mu_n^{k+r}(A_n) = \mu_{k+r}(A_n)$, so what we need to do is show that

$$\lim_{n\to\mathcal{U}} \int (\mu_r)_n((A_n)_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1,\ldots,x_k) = \int \mu_r((A_n)_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k).$$

This does not immediately follow from our definitions—to prove this, we have to look carefully at how integrals are calculated. Integrals are approximated by finite sums: we can break the interval $[0,1]$ into small sub-intervals: let $J = \lfloor 1/\epsilon \rfloor$ and take $[0,1] = I_0 \cup I_1 \cup I_2 \cup \cdots \cup I_J$ where $I_j = [j\epsilon, (j+1)\epsilon)$. Then, for each $(x_1,\ldots,x_k)$, $(\mu_r)_n(A_{x_1,\ldots,x_k}) \in I_j$ for exactly one j; let $C_n^j = \{(x_1,\ldots,x_k) \mid \mu_r(A_{x_1,\ldots,x_k}) \in I_j\}$. Then—still working in the ground model—we notice that

$$\begin{aligned}(\mu_{k+r})_n(A_n) &= \int (\mu_r)_n(A_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1,\ldots,x_k) \\ &= \sum_{j\le J} \int_{C_n^j} (\mu_r)_n(A_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1,\ldots,x_k) \\ &\approx \sum_{j\le J} (\mu_k)_n(C_n^j)(j+1/2)\epsilon.\end{aligned}$$

More precisely, since each interval I_j has radius $\epsilon/2$, we have

$$\left| \int (\mu_r)_n(A_{x_1,\ldots,x_n}) d(\mu_k)_n(x_1,\ldots,x_n) - \sum_{j\leq J} (\mu_k)_n(C^j_n)(j+1/2)\epsilon \right| < \epsilon/2.$$

Consider the corresponding sets in the ultraproduct, $C^j = [C^j_n]_{\mathcal{U}}$. Note that C^j is not quite $\{(x_1,\ldots,x_k) \mid \mu_r(A_{x_1,\ldots,x_k}) \in I_j\}$—because of the behavior at limits, when $\mu_r(A_{x_1,\ldots,x_k}) = j\epsilon$, we could have either $(x_1,\ldots,x_k) \in I_j$ or $(x_1,\ldots,x_k) \in I_{j-1}$. Instead, what we have is that $\mu_r(A_{x_1,\ldots,x_k})$ is in the *closure* of I_j: if $(x_1,\ldots,x_k) = ([x^1_n]_{\mathcal{U}},\ldots,[x^k_n]_{\mathcal{U}}) \in C^j$ then

$$\mu_r(A_{x_1,\ldots,x_n}) = \lim_{n\to\mathcal{U}} (\mu_r)_n((A_n)_{x^1_n,\ldots,x^k_n}) \in \overline{I_j} = [j\epsilon,(j+1)\epsilon].$$

Moreover, $\mu_k(C^j) = \lim_{n\to\mathcal{U}} (\mu_k)_n(C^j_n)$. Therefore

$$\int \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k) = \sum_{j\leq J} \int_{C_j} \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k)$$

and so

$$\left| \int \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k) - \sum_{j\leq J} \mu_k(C^j)(j+1/2)\epsilon \right| < \epsilon/2.$$

Putting this together,

$$\begin{aligned}
\mu_{k+r}(A) &= \lim_{n\to\mathcal{U}} (\mu_{k+r})_n(A_n) \\
&\approx \lim_{n\to\mathcal{U}} \sum_{j\leq J} (\mu_k)_n(C^j_n)(j+1/2)\epsilon \\
&= \sum_{j\leq J} \int_{C_j} \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k) \\
&= \int \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k)
\end{aligned}$$

where the $\approx$ indicates an error of size $< \epsilon/2$. Since we can make ϵ as small as we like, we have

$$\mu_{k+r}(A) = \int \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1,\ldots,x_k).$$

Finally, we must show the Fubini property for an arbitrary internal set $X \in \mathcal{B}^{k+r}(\mu_{k+r})$. Let A be a μ_{k+r}-approximation of X. We will show that, for almost every $x_1, \ldots, x_k$, $A_{x_1,\ldots,x_k}$ is a μ_r-approximation of $X_{x_1,\ldots,x_k}$.

Let $D \subseteq V_{\mathcal{U}}^k$ consist of the "defective" points—those points for which $A_{x_1,\ldots,x_k}$ is not a μ_r-approximation of $X_{x_1,\ldots,x_k}$. Then $A_{x_1,\ldots,x_k} \triangle X_{x_1,\ldots,x_k}$ is *not* μ_r-null, so there must be some natural number $d > 0$ so that $A_{x_1,\ldots,x_k} \triangle X_{x_1,\ldots,x_k}$ is not contained in an internal set of measure $< 1/d$. Let D_d consist of those $x_1, \ldots, x_k$ such that $A_{x_1,\ldots,x_k} \triangle X_{x_1,\ldots,x_k}$ is not contained in an internal set of measure $< 1/d$, so $D = \bigcup_d D_d$.

We will show that each D_d is in $\mathcal{B}(\mu)$ and has measure 0, and therefore D does as well. For any $\epsilon > 0$, $A \triangle X$ is contained in some internal set B with $\mu_{k+r}(B) < \epsilon/2d$. Since $\epsilon/d > \mu_{k+r}(B) = \int \mu_r(B_{x_1,\ldots,x_k}) d\mu_k(x_1, \ldots, x_k)$, the set of $x_1, \ldots, x_k$ such that $\mu_r(B_{x_1,\ldots,x_k}) \geq 1/d$ must be have measure $< \epsilon$. This set contains D_d, so D_d is contained in a set of measure $< \epsilon$. Since D_d is contained in a set of measure $< \epsilon$ for every $\epsilon > 0$, by Corollary 3.17, $\mu(D_d) = 0$.

Then for almost every $x_1, \ldots, x_k$ we have $\mu_r(A_{x_1,\ldots,x_k}) = \mu_r(X_{x_1,\ldots,x_k})$, so

$$\begin{aligned}
\mu_{k+r}(X) &= \mu_{k+r}(A) \\
&= \int \mu_r(A_{x_1,\ldots,x_k}) d\mu_k(x_1, \ldots, x_k) \\
&= \int \mu_r(X_{x_1,\ldots,x_k}) d\mu_k(x_1, \ldots, x_k).
\end{aligned}$$

□

3.5 Subgraph Density

The work above justifies the definition of a *measurable graph*. Ultraproducts are our main example, and whenever we see a measurable graph, an ultraproduct will not be far behind, but is convenient to sometimes forget about the extra structure of an ultraproduct.

Definition 3.24. A Keisler graded probability space $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k \in \mathbb{N}}$ is *atomless* if, for all $v \in V$, $\mu_1(\{v\}) = 0$.

A *measurable graph* is a Keisler graded probability space $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k \in \mathbb{N}}$ together with a symmetric set $E \in \mathcal{B}_2$. We say a measurable graph is *atomless* if the underlying Keisler graded probability space is atomless.

We often say "$G = (V, E, \mu_1)$ is a measurable graph", leaving the spaces $\mathcal{B}_k$ and measures μ_k for $k > 1$ implicit. In the rare situation where the space

$\mathcal{B}_k$ is significant and not implicit from the context, we will spell out the measurable graph more carefully.

Finite graphs are always measurable graphs, taking $\mathcal{B}_k$ to be the set of all k-tuples and μ_k to be the counting measure. The work above shows that ultraproducts of finite graphs are also measurable graphs, and these will be our main examples of measurable graphs.

We are only interested in either finite or atomless measurable graphs. The notion includes other examples—for instance, imagine a measurable graph where there is a single vertex with half the measure, but the other half of the measure has no atoms—but we are generally not interested in those cases.

We are now, finally, ready to define subgraph density in an ultraproduct and, more generally, in a measurable graph.

Definition 3.25. If $H = (W, F)$ is a finite graph with $W = \{v_1, \ldots, v_k\}$ and $G = (V, E, \mu_1)$ is a measurable graph, for any $(x_1, \ldots, x_k) \in V_{\mathcal{U}}^k$ we can define a potential copy $\pi_{x_1,\ldots,x_k}$ of H in V by setting $\pi_{x_1,\ldots,x_k}(v_i) = x_i$. $T_H(G)$ is the set of $(x_1, \ldots, x_k)$ such that $\pi_{x_1,\ldots,x_k}$ is an actual copy.

This is the definition we would expect: $T_H(G)$ is precisely the set of tuples $(x_1, \ldots, x_k)$ which form a copy of H. When $G = [G_n]_{\mathcal{U}}$ is an ultraproduct, $T_H([G_n]_{\mathcal{U}})$ is certainly internal: it is the set $\bigcap_{\{v_i,v_j\}\in F}\{(x_1, \ldots, x_k) \mid \{x_i, x_j\} \in [E_n]_{\mathcal{U}}\}$. By the Lós Theorem, $T_H(G) = [T_H(G_n)]_n$, where $T_H(G_n)$ is the set of copies of H in G_n.

The definition of the Keisler graded probability space ensures that each of the sets $\{(x_1, \ldots, x_k) \mid \{x_i, x_j\} \in [E_n]_{\mathcal{U}}\}$ is measurable, and therefore their intersection is as well.

Definition 3.26. We define $t_H(G)$, the *subgraph density* of H in G, to be $\mu(T_H(G))$.

When $G = [G_n]_{\mathcal{U}}$, since the T_H are internal, we have $t_H([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} \mu_n(T_H(G_n)) = \lim_{n\to\mathcal{U}} t_H(G_n)$. This is the property we originally demanded our notion of a limit graphs should have, which we have finally achieved—not with a big final proof, but falling right out of our definitions and the work we did to show that those definitions made sense.

The quantity $t_{\mathrm{K}_2}(G)$ has a particular significance, since it is essentially the density of E itself. Indeed, by definition,

$$t_{\mathrm{K}_2}(E) = \mu_2(\{(x, y) \mid \{x, y\} \in E\}).$$

We can generalize t_H to symmetric functions:

Definition 3.27. If $H = (W, F)$ then

$$t_H(f) = \int \prod_{\{v_i,v_j\}\in F} f(x_i, x_j)\, d\mu_{|W|}$$

if the integral is defined.

Often we restrict to L^∞ functions (that is, functions f so that there is a c with $|f(x)| < c$ for almost all x), for which such integrals are always defined. We need f to be symmetric because the edge $\{v_i, v_j\} \in F$ is unordered, so if $f(x_i, x_j)$ and $f(x_j, x_i)$ were not the same, we wouldn't know which one to include in our product.

Then $t_H(\chi_E) = t_H(E)$. At least when f only takes on positive values, we can think of this as counting the density of "copies" of H in a "weighted graph" where an edge with $f(x, y) = 2$ counts as a double edge, an edge with $f(x, y) = 1/2$ counts as half an edge, and so on.

It is convenient to count not only the number of copies of a graph, but the number of extensions of a partial copy. That is, suppose $H = (W, F)$ is a graph with $W = \{w_1, w_2, \ldots, w_k\}$, and we have already picked, for instance, vertices x_1 and x_2 in V to correspond to w_1 and w_2. We would like to know how many choices of $x_3, \ldots, x_k$ correspond to copies of H.

Definition 3.28. Let $H = (W, F)$ with $W = \{w_1, \ldots, w_k\}$. For any $d \le k$ and any $x_1, \ldots, x_d \in V$, define

$$T_H(G, x_1, \ldots, x_d) = \{(x_{d+1}, \ldots, x_k) \in V^{k-d} \mid \pi_{x_1,\ldots,x_k} \text{ is an actual copy of } H\}$$

and

$$t_H(G, x_1, \ldots, x_d) = \mu_{k-d}(T_H(G, x_1, \ldots, x_d)).$$

The Fubini property assures us that when $H = (W, F)$ is a graph and $H' = (W_0, F \restriction \binom{W_0}{2})$ is a subgraph,

$$t_H(G) = \int_{T'_H(G)} t_H(G, x_1, \ldots, x_{|W_0|})\, d\mu_{|W_0|}.$$

This notation generalizes our earlier approach to degree: $t_{\mathrm{K}_2}(G, x) = \int \chi_E(x, y)\, d\mu_1(y) = \deg_G(x)$. More generally, this lets us count things like

$$t_{\mathrm{K}_3}(G, x) = \int \chi_E(x, y)\chi_E(x, z)\chi_E(y, z)\, d\mu_2(z),$$

the number of pairs $\{y, z\}$ which form a triangle with x, and

$$t_{\mathrm{K}_3}(G, x, y) = \chi_E(x, y) \int \chi_E(x, z)\chi_E(y, z)\, d\mu_1(z),$$

the number of ways to extend a pair x, y to a triangle. As this example illustrates, $t_H(G, x_1, \ldots, x_d)$ is 0 if it is impossible to extend $x_1, \ldots, x_d$ to a copy of H because an edge from H is already missing—if there is no edge between x and y, there are no choices of z which will make $\{x, y, z\}$ a triangle.

The Fubini property promises that

$$\int t_{\mathrm{K}_3}(G, x)\, d\mu_1 = t_{\mathrm{K}_3}(G)$$

—that is, if we add up, over all vertices, how many extensions there are to a triangle, we get the number of triangles—and

$$\int \chi_E(x, y) t_{\mathrm{K}_3}(G, x, y)\, d\mu_2 = t_{\mathrm{K}_3}(G)$$

—if we add up, over all edges, how many extensions there are to a triangle, we again get the average number of triangles. (A non-edge, of course, belongs to no triangles.)

Note that even though only H appears in the notation t_H, the definition depends not only on the graph (W, F), but on the ordering $\{w_1, \ldots, w_k\}$: if H is the graph  , the function $t_H(G, x, y)$ could be either

$$\int \chi_E(x, z)\chi_E(y, z)\chi_E(x, w)\chi_E(y, w)\, d\mu_2$$

or

$$\chi_E(x, y) \int \chi_E(x, z)\chi_E(y, w)\chi_E(z, w)\, d\mu_2,$$

depending on whether w_1, w_2 are chosen to be adjacent or non-adjacent vertices from H.

3.6 Sampling

Every ultraproduct of finite graphs is a measurable graph and we can now show that, at least for questions of subgraph density, every atomless measurable graph resembles an ultraproduct.

There is a canonical way to sample a random finite graph from a measurable graph $G = (V, E, \mu_1)$: choose finitely many vertices $\mathbf{v}_1, \ldots, \mathbf{v}_n$ according to the measure μ_1 on V, and let $\mathbf{G}_n = (\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, E \restriction \{\mathbf{v}_1, \ldots, \mathbf{v}_n\})$. Our requirement that $\mu_1(\{v\}) = 0$ for each $v \in V$ means that when $i \neq j$, $\mathbb{P}(\mathbf{v}_i = \mathbf{v}_j) = 0$, so, with probability 1, $\mathbf{G}_n$ is a well-defined finite graph whose properties we can consider.

Actually, we can say something slightly more general: whenever $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ is an atomless Keisler graded probability space and f is a symmetric, $\mathcal{B}_2$-measurable function with values in $[0,1]$, we can choose finitely many vertices $\mathbf{v}_1, \ldots, \mathbf{v}_n$ according to the measure μ_1 on V and then, for each pair $\mathbf{v}_i, \mathbf{v}_j$, let $\mathbf{p}_{\{i,j\}} \in \{0,1\}$ be a random value obtained by flipping a coin which comes up heads with probability $f(\mathbf{v}_i, \mathbf{v}_j)$. We can set $\mathbf{G}_n = (\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, \{\{\mathbf{v}_i, \mathbf{v}_j\} \mid \mathbf{p}_{\{i,j\}} = 1\})$.

Lemma 3.29. *Let $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ be an atomless Keisler graded probability space and let f be a symmetric $\mathcal{B}_2$-measurable function with values on $[0,1]$. For any finite graph H and any $\epsilon > 0$, there is an m so that whenever we sample an n vertex graph $\mathbf{G}_n$ from f with $n \geq m$,*

$$\mathbb{P}[|t_H(f) - t_H(\mathbf{G}_n)| < \epsilon] > 1 - \epsilon.$$

Proof. Our arguments look much like they did for $\mathbf{R}_p$.

Let $H = (W, F)$, write $W = \{w_1, \ldots, w_k\}$, and suppose n is a large finite number. First, we consider $\mathbb{E}(t_H(\mathbf{G}_n))$. For each $\pi : W \to \{1, \ldots, n\}$, let $\mathbf{c}_\pi$ be the random variable whose value is

$$\prod_{\{w_i,w_j\}\in F} f(\mathbf{v}_{\pi(i)}, \mathbf{v}_{\pi(j)}).$$

When π is injective, observe that $\mathbb{E}(\mathbf{c}_\pi)$ is simply $t_H(f)$.

By the linearity of expectation,

$$\mathbf{E}(t_H(\mathbf{G}_n)) = \frac{1}{n^{|W|}} \sum_\pi \mathbb{E}(\mathbf{c}_\pi).$$

Since there are at most $Cn^{|W|-1}$ non-injective functions $\pi : W \to \{1, \ldots, n\}$ for some constant C, ths means that

$$\mathbb{E}(t_H(\mathbf{G}_n)) = t_H(f) + O(\frac{1}{n}).$$

As in the proof of Theorem 1.7, we use McDiarmid's inequality. This time the random variables are the $\mathbf{v}_i$ and the function of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the subgraph density t_H in the graph induced by these vertices. A single vertex participates in at only $|W|n^{|W|-1}$ copies of π, and therefore changing a single vertex can change the subgraph density by at most $|W|/n$. Then McDiarmid's inequality says

$$\mathbb{P}(|\mathbb{E}(t_H(\mathbf{G}_n)) - t_H(\mathbf{G}_n)| \geq \epsilon) \leq 2e^{-\frac{2\epsilon^2}{n\frac{|W|^2}{n^2}}} = 2e^{-\frac{2\epsilon^2 n}{|W|^2}}.$$

Once we pick ϵ, the value $\frac{2\epsilon^2}{|W|^2}$ is fixed, so by choosing n large enough, we can make this bound as small as we like. □

Specializing to the case where $f = \chi_E$ gives the following.

Corollary 3.30. *Let $G = (V, E)$ be an atomless measurable graph. For any finite graph H and any $\epsilon > 0$, there is an m so that whenever we sample an n vertex graph $\mathbf{G}_n$ from G with $n \geq m$,*

$$\mathbb{P}[|t_H(G) - t_H(\mathbf{G}_n)| < \epsilon] > 1 - \epsilon.$$

This immediately implies the same claim for finitely many graphs at once.

Corollary 3.31. *Let $G = (V, E)$ be an atomless measurable graph. For any finite list of finite graphs $H_1, \ldots, H_k$ and any $\epsilon > 0$, there is an m so that whenever we sample an n vertex graph $\mathbf{G}_n$ from G with $n \geq m$, for each $i \leq k$,*

$$\mathbb{P}[|t_{H_i}(G) - t_{H_i}(\mathbf{G}_n)| < \epsilon] > 1 - \epsilon.$$

Proof. Apply the lemma with ϵ/k. □

Corollary 3.32. *Let $G = (V, E)$ be an atomless measurable graph. For each n, let $\mathbf{G}_n$ be an n vertex graph sampled from G (with the $\mathbf{G}_n$ sampled independently from each other). Then, with probability 1, for every finite graph H,*

$$\lim_{n\to\infty} t_H(\mathbf{G}_n) = t_H(G).$$

Proof. There are countably many finite graphs, so it suffices to show that, for each one individually, the probability that $\lim_{n\to\infty} t_H(\mathbf{G}_n) \neq t_H(G)$ is 0. So fix a finite graph H and fix $\epsilon > 0$. We will show that, there is an m so that, with probability $> 1 - \epsilon$, for every $n \geq m$, $|t_H(\mathbf{G}_n) - t_H(G)| < \epsilon$.

For a given m, the probability that there is any $n \geq m$ with $|t_H(\mathbf{G}_n) - t_H(G)| \geq \epsilon$ is at most

$$\sum_{n\geq m} \mathbb{P}(|t_H(\mathbf{G}_n) - t_H(G)| \geq \epsilon) < \sum_{n\geq m} 2e^{-\frac{\epsilon^2}{|W|^2}n} \leq \int_m^\infty 2e^{-\frac{\epsilon^2}{|W|^2}x}dx = \frac{2|W|^2}{\epsilon^2}e^{-\frac{\epsilon^2}{|W|^2}m}.$$

In particular, choosing m large enough, this probability is smaller than ϵ. □

Note that this corollary gives us true limits, not just ultralimits: *every* ultraproduct of the $\mathbf{G}_n$ has the same subgraph densities as G.

Consider the following: we begin with the function f which is constantly equal to $1/2$. By sampling, we obtain a sequence of graphs $\mathbf{G}_n$ whose ultraproduct G is a proper graph—it is either 1 or 0 everywhere. But this ultraproduct is clearly related to the original graph: for instance, $t_H(G) = t_H(f)$ for all finite graphs H. One of our goals later will be seeing how we recover the function f from the graph G; indeed, we will discover that f is the "non-random part" of the graph G.

3.7 The Possible Subgraph Densities are Compact

One reason to consider atomless measurable graphs, even if our main interest is finite graphs, is that they *compactify* the space of subgraph densities: rather than speaking of sequences of graphs with subgraph densities approaching some limiting values of interest, we can focus on a single measurable graph achieving these limiting values.

Theorem 3.33. *Let ζ be a function assigning, to each finite graph H, a density $\zeta(H)$. Suppose that for every finite list of finite graphs $H_1, \ldots, H_k$ and every $\epsilon > 0$, there is a finite or atomless measurable graph G so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G)| < \epsilon$.*

Then there is a finite or atomless measurable graph G such that $t_H(G) = \zeta(H)$ for all finite graphs H.

A more formal approach to this idea, which we do not investigate in detail because we will not otherwise need it, is to call ζ a *possible subgraph density* if there exists some finite or atomless measurable graph G so that, for every H, $\zeta(H) = t_G(H)$. There is a natural topology on the space of possible subgraph densities—take a basic open set to be $\{\zeta \mid \zeta(H) \in I\}$ for an open interval I. (This topology is even metrizable—if we put the finite graphs in an order $H_1, H_2, \ldots$, we can define $d(\zeta, \zeta) = \sum_i \frac{|\zeta(H_i) - \zeta(H_i)|}{2^i}$.) Then the theorem says that the space of possible subgraph densities are compact.

Proof. One way to prove this is to drop through the world of finite graphs. Fix an ordering of the finite graphs, $H_1, H_2, \ldots$. For each k, choose a finite or atomless measurable graph G'_k so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G'_k)| < 1/2k$. If G'_k is finite, let $G_k = G'_k$. If G'_k is atomless, by Corollary 3.31, we may sample a graph G_k from G'_k with $|t_{H_i}(G_k) - t_{H_i}(G'_k)| < 1/2k$ for each $i \leq k$, so $|\zeta(H_i) - t_{H_i}(G_k)| < 1/k$.

Then for each H_i, we have

$$t_{H_i}([G_k]_{\mathcal{U}}) = \lim_{k \to \mathcal{U}} t_{H_i}(G_k) = \zeta(H_i).$$

□

We will need to know that the same results apply if we restrict ourselves to graphs omitting certain subgraphs entirely.

Theorem 3.34. *Let ζ be a function assigning, to each finite graph H, a density $\zeta(H)$. Let $\mathcal{X}$ be a collection of forbidden finite graphs.*

Suppose that for every finite list of finite graphs $H_1, \ldots, H_k$ and every $\epsilon > 0$, there is a measurable graph G so that:

- *for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G)| < \epsilon$,*
- *for each $H \in \mathcal{X}$, $T_H(X) = \emptyset$.*

Then there is a measurable graph G such that $t_H(G) = \zeta(H)$ for all finite graphs H and $T_H(G) = \emptyset$ for all $H \in \mathcal{X}$.

Proof. This follows by the same proof, noting that when we sample $\mathbf{G}_n$ from G, if $T_H(G) = \emptyset$ then $T_H(\mathbf{G}_n) = \emptyset$ (because $\mathbf{G}_n$ is in fact a subgraph of G), and when $T_H(G_n) = \emptyset$ for all n then $T_H([G_n]_{\mathcal{U}}) = \emptyset$ as well (by Łos' Theorem). □

3.8 Remarks

The development of measures in ultraproducts (and nonstandard models more generally) goes back to Loeb [108], and measures of this kind are called *Loeb measures*. More precisely, the measure we have defined is the completion of the usual Loeb measure; the construction of Loeb measure is more robust than the construction we have used here (for instance, it can be carried out in settings of nonstandard analysis other than ultraproducts). See, for instance, [70] for details.

The notion of a graded probability space was introduced by Keisler [97] and used, in a context very close to the one we are concerned with, by Keisler's student Hoover [86].

The *Furstenberg correspondence*—a correspondence between finite sets and dynamical systems introduced by Furstenberg to give his proof of Szemerédi's Theorem [66]—can be understood as a special case of the ultraproduct construction, where the use of an ultrafilter can be replaced by a well-chosen countable filter. (In proofs using the Furstenberg correspondence, this takes the form of some sort of diagonalization argument.)

Chapter 4

Extrema

We have now developed enough machinery to begin applying it to prove some facts about finite graphs. One natural application is to questions about extrema—to questions about what the largest or smallest value of some $t_H(G)$ can be. Since the atomless measurable graphs compactify the possible subgraph densities, they are the natural setting to look for optimal (as opposed to "nearly optimal") graphs.

In this chapter we consider two related problems of this kind:

- What is the maximum possible value of $t_{\mathrm{K}_2}(G)$ given that G is large and does not contain any copies of some finite graph H?
- What is the smallest possible value of $t_{\mathrm{K}_r}(G)$ given that G is large and $t_{\mathrm{K}_2}(G) = c$?

4.1 Mantel's Theorem

The simplest case of these questions is also the first result of extremal graph theory, *Mantel's Theorem.*

Theorem 4.1. *If G has no triangles then $t_{\mathrm{K}_2}(G) \leq \frac{1}{2}$.*

The bipartite graph illustrates that this bound is optimal: the bipartite graph $K_{n,n}$ has no triangles and $t_{\mathrm{K}_2}(K_{n,n}) \approx 1/2$.

Proof. Since the graph has no triangles, whenever $\{x, y\} \in E$, x and y must

have no neighbors in common. In particular, $\deg_G(x) + \deg_G(y) \leq 1$, so

$$\begin{aligned} t_{\mathrm{K}_2}(G) &= \int \chi_E(x,y) \cdot 1 \, dy \, dx \\ &\geq \int \chi_E(x,y)(\int \chi_E(x,z)dz + \int \chi_E(y,w)dw)dy \, dx \\ &= 2\int (\chi_E(x,y)dy)^2 \, dx \\ &\geq 2\left(\int \chi_E(x,y)dxdy\right)^2 \\ &= 2t_{\mathrm{K}_2}(G)^2, \end{aligned}$$

so $\frac{1}{2} \geq t_{\mathrm{K}_2}(G)$. □

Our approach to generaling this will be to use the basic insight that makes calculus relevant to optimization: at an extremum, the derivative in every direction must be 0. In our context, that means that if G is a graph maximizing $t_{\mathrm{K}_2}(G)$ subject to $T_H(G) = \emptyset$ and we "change G a little bit" to some graph G_δ where we still have $T_H(G_\delta) = \emptyset$, we need $t_{\mathrm{K}_2}(G_\delta) \leq t_{\mathrm{K}_2}(G)$. In particular, this means that "the derivative in the direction of G_δ" must be ≤ 0.

This will allow us to take global information—like "$t_{\mathrm{K}_2}(G)$ is maximal"—and convert it into local information—like "almost every vertex in G has the same number of neighbors".

4.2 Shifting Measures

We already showed, in Section 3.7, that extremal graphs exist. We next need to make precise the idea of taking a graph and "changing it a little bit" in some direction.

We want to consider what happens if we take a measurable graph $G = (V, E, \mu_1)$ and modify μ_1 slightly. First, consider the analog in a finite graph. We could imagine modifying a large finite graph by making changes to a small number of vertices—say, by deleting a small fraction of the vertices, or taking one vertex and "duplicating" it (adding another vertex with the same neighbors). These modifications become natural operations in a measurable graph, because we can change the measure slightly to get a new measurable graph.

If this new measure could concentrate on a set of measure 0, it could end up having totally different behavior than the original one. So the case to

focus on is where the new measure is *absolutely continuous* with respect to the old one.

Definition 4.2. If μ and ν are both measures on $\mathcal{B}$, we say ν is *absolutely continuous with respect to* μ, written $\nu \ll \mu$, if there is a measurable function f such that, for every $S \in \mathcal{B}$, $\nu(S) = \int_S f \, d\mu$. In this case f is called the *Radon-Nikodym derivative* of ν.

We can sometimes think of f as a "weight"—when $f(x)$ is more than 1, it means x is "more imporant" to ν than it was to μ.

Lemma 4.3. *Suppose $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ is a Keisler graded probability space and $\nu \ll \mu_1$. Then there is a unique Keisler graded probability space $\{(V^k, \mathcal{B}_k, \nu_k)\}_{k\in\mathbb{N}}$ such that $\nu_1 = \nu$ and, for all k, $\nu_k \ll \mu_k$.*

Proof. Let f be the Radon-Nikodym derivative of ν. Then for $S \in \mathcal{B}_k$, we define

$$\nu_k(S) = \int \chi_S(x_1, \ldots, x_k) \prod_{i\leq k} f(x_i) \, d\mu_k.$$

Then $\nu_k \ll \mu_k$ and $\nu_1 = \nu$ by definition.

We must check that $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ is a Keisler graded probability space. Symmetry of ν_k follows by the definition and the symmetry of μ_k.

The Fubini Property holds since

$$\begin{aligned}
\nu_{k+r}(S) &= \int \chi_S(x_1, \ldots, x_{k+r}) \prod_{i\leq k+r} f(x_i) \, d\mu_{k+r} \\
&= \int \left[\int \chi_S(x_1, \ldots, x_{k+r}) \prod_{k<i\leq k+r} f(x_i) \, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right] \prod_{i\leq k} f(x_i) \, d\mu_k(x_1, \ldots, x_k) \\
&= \int \nu_r(S_{x_1,\ldots,x_k}) \prod_{i\leq k} f(x_i) \, d\mu_k \\
&= \int \nu_r(S_{x_1,\ldots,x_k}) \, d\nu_k.
\end{aligned}$$

To see uniqueness, suppose $\{(V^k, \mathcal{B}_k, \nu'_k)\}_{k\in\mathbb{N}}$ is a Keisler graded probability space with $\nu'_1 = \nu_1$ and each $\nu'_k \ll \mu_k$. We proceed by induction on k, showing that $\nu'_k = \nu_k$. Since $\nu'_1 = \nu_1$ by assumption, we assume that $\nu'_k = \nu_k$. Then for any $S \in \mathcal{B}_{k+1}$, we have

$$\begin{aligned}
\nu'_{k+1}(S) &= \int \nu'_k(S_{x_1}) \, d\nu'_1(x_1) \\
&= \int \nu_k(S_{x_1}) \, d\nu_1(x_1) \\
&= \nu_{k+1}(S).
\end{aligned}$$

□

In light of this lemma, when we have a measurable graph G with measure μ and a new measure $\nu \ll \mu$, we immediately have a new measurable graph with the same underlying measurable sets as G, but the new measure.

We will mostly be interested in a particular choice of ν: when we take a particular set $B \subseteq V$ and look at the measure in which B has slightly more (or less) measure and the rest of the graph has slightly less (or more).

Definition 4.4. For any $B \subseteq V$ with $\mu_1(B) > 0$, let

$$\nu_\delta^B(S) = \frac{1}{1+\delta}[\mu_1(S \setminus B) + (1 + \frac{\delta}{\mu_1(B)})\mu_1(S \cap B)].$$

The Radon-Nikodym derivative of ν_δ^B is the function whose value is $\frac{1}{1+\delta}(1 + \frac{\delta}{\mu_1(B)})$ on B and $\frac{1}{1+\delta}$ outside of B.

Then we have:

- ν_δ^B is a probability measure,
- $\nu_\delta^B \ll \mu_1$,
- $\nu_\delta^B(B) = \frac{\mu(B)+\delta}{1+\delta}$,
- $\nu_0^B = \mu_1$.

The measure ν_δ^B with $\delta > 0$ corresponds to the following idea: suppose we take a small number of points from the set B and duplicate them in such a way that we add a set of new points of measure δ; we then renormalize the measure so that the measure of the new, larger set is still 1 (by dividing by $1 + \delta$).

To see why we are interested in the measure ν_δ^B, consider the case where

$$B = \{x \in V \mid t_{K_2}(G, x) \geq \mu_2(E) + \epsilon\}$$

has positive measure. Recall that $t_{K_2}(G, x) = \int \chi_E(x, y)\, d\mu_1(y)$ is the normalized degree of x, so B is the vertices with more than their share of neighbors.

Consider what we expect ν_δ^B to do when $\delta > 0$: we are making the vertices with too many neighbors "count for more", so the edge density t_{K_2} should increase. But there is a countervailing trend: it might be that $t_{K_2}(G, x)$ is large because x has many neighbors in $V \setminus B$, and those neighbors count for less in ν_δ^B.

But the first trend is linear in δ while the second is quadratic. For instance, when we think of ν_δ^B as meaning that we duplicate a small number of points in B, when we select two points x, y randomly according to ν_δ^B, the probability that one of these is a new point is about $\frac{2\delta}{(1+\delta)^2}$. However the probability that *both* are new points is about $\frac{\delta^2}{(1+\delta)^2}$.

The linear part contributes about $\frac{2\delta(\mu_2(E)+\epsilon)}{(1+\delta)^2}$ edges—in particular, it increases the edge density by about $\frac{2\delta\epsilon}{(1+\delta)^2}$. Taking the derivative with respect to δ lets us focus on the linear part.

We can state this in the following general form, considering the density of K_r for any y.

Lemma 4.5. *Let $G = (V, E, \mu_1)$ be a measurable graph with $t_{\mathrm{K}_r}(G) = c$, let $B \subseteq V$ with positive measure, and let $u = \frac{1}{\mu(B)} \int_B t_{\mathrm{K}_r}(G, x)\, d\mu_1$. Let $G_\delta = (V, E, \nu_\delta^B)$. Let h be the function with $h(\delta) = t_{\mathrm{K}_r}(G_\delta)$. Then $h'(0) = r(u - c)$.*

In words, $u - c$ is the amount by which vertices in B belong to too many copies of K_r. (When $u - c$ is negative, vertices in B belong to too few copies instead.) This lemma says that when δ is very small, so we add a very small number of new copies of vertices in B then we add about $r(u - c)\delta$ new copies of K_r.

The change in copies of K_r comes from copies where at least one of the vertices in the copy is a "new" vertex. We get a coefficient of r because a copy of K_r has r vertices, each of which could be a new vertex. We look at what happens when δ is very small because in this case the number of copies of K_r containing multiple new vertices is negligible.

Proof. Observe that

$$\begin{aligned} t_{\mathrm{K}_r}(G_\delta) &= \int \prod_{1 \le i < j \le r} \chi_E(x_i, x_j)\, d(\nu_\delta)_r \\ &= \frac{1}{(1+\delta)^r} \left[\int \prod_{1 \le i < j \le r} \chi_E(x_i, x_j)\, d\mu_r + r \frac{\delta}{\mu(B)} \int \chi_B(x_1) \prod_{1 \le i < j \le r} \chi_E(x_i, x_j)\, d\mu_r \right. \\ &\quad + \binom{r}{2} \frac{\delta^2}{\mu(B)} \int \chi_B(x_1)\chi_B(x_2) \prod_{1 \le i < j \le r} \chi_E(x_i, x_j)\, d\mu_r + \cdots \\ &\quad \left. + \frac{\delta^r}{(\mu(B))^r} \int \prod_{i \le r} \chi_B(x_i) \prod_{1 \le i < j \le r} \chi_E(x_i, x_j)\, d\mu_r \right]. \end{aligned}$$

In particular, since $h_r(0) = c = t_{\mathrm{K}_r}(G)$,

$$\begin{aligned} h_r'(0) &= \lim_{\delta\to 0} \frac{t_{\mathrm{K}_r}(G_\delta) - t_{\mathrm{K}_r}(G)}{\delta} \\ &= \lim_{\delta\to 0} \frac{c + ru\delta + \delta^2(\cdots) - c(1+\delta)^r}{\delta(1+\delta)^r} \\ &= r(u-c). \end{aligned}$$

□

4.3 Turán's Theorem

We are now in a position to prove the generalization of Mantel's Theorem from triangles to K_r.

Theorem 4.6 (Turán's Theorem)**.** *For any $r \geq 3$, if $T_{\mathrm{K}_r}(G) = \emptyset$ then $t_{\mathrm{K}_2}(G) \leq 1 - \frac{1}{r-1}$.*

Proof. Suppose not, so there is a measurable graph $G = (V, E, \mu)$ with $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{r-1}$ and $T_{\mathrm{K}_r}(G) = \emptyset$. By Theorem 3.34, there is a measurable graph G with $T_{\mathrm{K}_r}(G) = \emptyset$ maximizing the value $t_{\mathrm{K}_2}(G)$ among all such measurable graphs.

We claim that, for almost every x, $t_{\mathrm{K}_2}(G, x) = t_{\mathrm{K}_2}(G)$. Suppose not. Let $c = t_{\mathrm{K}_2}(G)$ and, for any $\epsilon > 0$, let $B = \{x \mid t_{\mathrm{K}_2}(G, x) \geq c + \epsilon\}$. Suppose $\mu_1(B) > 0$. Consider the graphs $G_\delta = (V, E, \nu_\delta^B)$. By Lemma 4.5, when $h(\delta) = t_{\mathrm{K}_2}(G_\delta)$, $h'(\delta) \geq 2\epsilon > 0$. In particular, when δ is small, $t_{\mathrm{K}_2}(G_\delta) > c$. But $T_{\mathrm{K}_r}(G_\delta) = \emptyset$, so this contradicts the maximality of G. Therefore, for every $\epsilon > 0$, $\mu_1(B) = 0$.

Since G is evenly distributed, let $V' \subseteq V$ be a set with $\mu(V') = 1$ and, for all $x \in V'$, $t_{\mathrm{K}_2}(G, x) = c > 1 - \frac{1}{r-1}$. We obtain a copy of K_r in G by successively choosing elements: given $x_1, \ldots, x_m$ a copy of K_m in V' with $m < r$, for each $i \leq m$, $\{y \in V' \mid \{x_i, y\} \notin E\}$ has measure $< \frac{1}{r-1}$, and therefore the set of $y \in V'$ such that $\{x_i, y\} \in E$ for each i has measure $> 1 - \frac{m}{r-1} \geq 0$, and in particular is non-empty, so we may choose $x_{m+1} \in V'$ so that $x_1, \ldots, x_m, x_{m+1}$ is a copy of K_{m+1}. This gives us a copy of K_r, showing that $|T_{\mathrm{K}_r}(G)| > 0$. □

This immediately gives a finite version.

Corollary 4.7. *For each $r \geq 3$ and any $\epsilon > 0$, there is an n so that if G is a graph on $\geq n$ vertices with $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{r-1} + \epsilon$ then $|T_{\mathrm{K}_r}(G)| > 0$.*

Proof. Suppose not. Then for some r and some $\epsilon > 0$, for each n there is a graph G_n with $\geq n$ vertices and $t_{\mathrm{K}_2}(G_n) > 1 - \frac{1}{r-1} + \epsilon$ and $|T_{\mathrm{K}_r}(G_n)| = 0$.

Then take any ultrafilter $\mathcal{U}$ and consider $G = [G_n]_{\mathcal{U}}$: $t_{\mathrm{K}_2}(G) = \lim_{n \to \mathcal{U}} t_{\mathrm{K}_2}(G_n) \geq 1 - \frac{1}{r-1} + \epsilon$ while $T_{\mathrm{K}_r}(G) = \emptyset$ by Łoś's Theorem, contradicting the theorem. □

It is typical that our finite consequences have an asymptotic character—our ultraproduct arguments naturally lend themselves to working with "sufficiently large" graphs. In some cases, including this one [19, 34, 51], there are sharper results available by other methods: these ultraproduct techniques are not well suited to identifying the exact number of edges at which copies of K_r appear.

Turán's bound is optimal:

Lemma 4.8. *For every $r \geq 3$, there is a measurable graph G with $t_{K_2}(G) = 1 - \frac{1}{r-1}$ and $|T_{\mathrm{K}_r}(G)| = 0$.*

Proof. Take G to be the complete $r-1$-partite graph where the parts all have the same size. That is, let $G = (V, E)$ where V is a disjoint union of $r-1$ parts, $V = \bigcup_{i<r} X_i$, $\mu_1(X_i) = \frac{1}{r-1}$, and let E consist of all pairs $\{x, y\}$ such that $x \in X_i$ and $y \in X_j$ where $i \neq j$. Then $\mu_2(E) = 1 - \frac{1}{r-1}$, but if we have r vertices $x_1, \ldots, x_r$, two must belong to the same part and therefore do not have an edge between them. □

4.4 Counting K_t in the Equal s-partite Graphs

We will spend most of the rest of this chapter investigating the minimum number $t_{\mathrm{K}_r}(G)$ can be in large graphs given that $t_{\mathrm{K}_2}(G) = c$ for some value $c \in [0, 1]$. This turns out to be a difficult problem which took more than fifty years to resolve (from Goodman's work in 1959 [74] to Reiher's solution in 2012 [132]).

The first step will be showing that the complete s-partite graphs with all parts the same size have as few copies of K_r as possible; this will answer the question at the specific series of values $t_{\mathrm{K}_2}(G) = 1/2, 2/3, 3/4, \ldots$.

It will be convenient to give names to the complete s-partite graphs from Lemma 4.8.

Definition 4.9. Let P_s be the measurable graph $\mathrm{P}_s = (V, E)$ where:

- V is a disjoint union $V = \bigcup_{i \leq s} X_i$ where $\mu_1(X_i) = \frac{1}{s}$ for all i,
- $\{x, y\} \in E$ if and only if x and y are in different parts.

$t_{\mathrm{K}_r}(\mathrm{P}_s)$ is easy to count by considering what happens when we successively choose r vertices randomly: we get a copy of K_r if the vertices all end up in different parts. The first vertex can be in other part, the second vertex must in any of the other parts, which happens with probability $1 - \frac{1}{s}$, the third must be in any part other than the two already used, and so on.

Lemma 4.10. *When $s \geq r$, $t_{\mathrm{K}_r}(\mathrm{P}_s) = \prod_{i<r}(1 - \frac{i}{s})$.*

We will calculate a lower bound on $t_{\mathrm{K}_r}(G)$. This lower bound won't be sharp in general, but it happens to be sharp at exactly the values $1 - \frac{1}{s}$.

In order to do this, we will need a more sophisticated way to relate the values of t_G for different values of G. We noted earlier that our notion of counting subgraphs really counts "ordered" or "labeled" graphs—we defined t_G to be an integral involving *ordered* tuples $(x_1, \ldots, x_n)$. We now define a symmetric analog, which we abbreviate using the letter c (for "counting").

Definition 4.11. If H is a graph with n vertices,

$$c_H(x_1, \ldots, x_n) = \frac{1}{n!}|\{\pi \text{ a permutation of } \{1, \ldots, n\} \mid (x_{\pi(1)}, \ldots, x_{\pi(n)}) \in T_H(G)\}|.$$

That is, $c_H(x_1, \ldots, x_n)$ is the fraction of orderings of the vertices $x_1, \ldots, x_n$ which give us a copy of H. When G is a symmetric graph, $c_G(x_1, \ldots, x_n)$ is either 1 or 0, but when G is not symmetric, we could get intermediate values; for instance, if V is the graph with two edges on 3 vertices (the triangle minus one edge), and $\{x_1, x_2, x_3\}$ are vertices so that $\{x_1, x_2\}, \{x_1, x_3\} \in E$ but $\{x_2, x_3\} \notin E$, $c_{\mathrm{V}}(x_1, x_2, x_3) = \frac{1}{3}$—there are six possible orderings of the three vertices, and the only ones which give copies of V are the ones which match x_1 to the tip of the V.

Integrating over c_H just gives us t_H again.

Lemma 4.12. $\int c_H \, d\mu_n = t_H(G)$

Proof.

$$\begin{aligned}
\int c_H(x_1, \ldots, x_n) \, d\mu_n &= \int \frac{1}{n!} \sum_{\pi} \chi_{T_H}(x_{\pi(1)}, \ldots, x_{\pi(n)}) \, d\mu_n \\
&= \frac{1}{n!} \sum_{\pi} \int \chi_{T_H}(x_{\pi(1)}, \ldots, x_{\pi(n)}) \, d\mu_n \\
&= \frac{1}{n!} \sum_{\pi} t_H(G) \\
&= t_H(G).
\end{aligned}$$

□

The power of c_H comes from comparing different graphs, as in the following lemma.

Lemma 4.13. *Let K_r^- be the graph on $r \geq 2$ vertices with $\binom{r}{2} - 1$ edges. Then*

$$t_{K_r^-}(G) \leq \frac{r-2}{r-1} t_{K_r}(G) + \frac{1}{r-1} t_{K_{r-1}}(G).$$

Proof. Note that:

- if $\{x_1, \ldots, x_r\}$ is a copy of K_r then $c_{K_r^-}(x_1, \ldots, x_r) = 1$,
- if $\{x_1, \ldots, x_r\}$ is missing exactly one edge then $c_{K_r^-}(x_1, \ldots, x_r) = \frac{2}{(r-1)r}$,
- if $\{x_1, \ldots, x_r\}$ is missing more than one edge then $c_{K_r^-}(x_1, \ldots, x_r) = 0$.

Let $K_{r-1}(r)$ be the graph on r vertices where the first $r-1$ vertices are complete but the r-th vertex is isolated. (For instance, $K_3(4)$ is  .) Note that $t_{K_{r-1}(r)}(G) = t_{K_{r-1}}(G)$ (the extraneous vertex has no effect when calculating the integral since we are counting ordinary, not induced, subgraphs). Then

- if $\{x_1, \ldots, x_r\}$ is a copy of K_r then $c_{K_{r-1}(r)}(x_1, \ldots, x_r) = 1$,
- if $\{x_1, \ldots, x_r\}$ is missing exactly one edge then $c_{K_{r-1}(r)}(x_1, \ldots, x_r) = \frac{2}{r}$.

In particular, considering cases, we see that for any $(x_1, \ldots, x_r)$, we have

$$c_{K_r^-}(x_1, \ldots, x_r) \leq \frac{r-2}{r-1} c_{K_r}(x_1, \ldots, x_r) + \frac{1}{r-1} c_{K_{r-1}(r)}(x_1, \ldots, x_r).$$

Integrating, we get

$$t_{K_r^-}(G) \leq \frac{r-2}{r-1} t_{K_r}(G) + \frac{1}{r-1} t_{K_{r-1}}(G).$$

□

We should expect many inequalities of this kind to be true—we should expect that the presence of some subgraphs requires the presence of other subgraphs in various quantities. This lemma—saying that if we have copies of K_r^-, we ought to have either some copies of K_r or a large number of copies of K_{r-1}—is typical. Finding the right inequalities for a given proof is a matter of trial and error, though automated tools can now help in some situations [162].

Theorem 4.14. *If* $t_{\mathrm{K}_2}(G) = c \geq \frac{r-2}{r-1}$ *then*

$$t_{\mathrm{K}_r}(G) \geq c(2c-1)(3c-2)\cdot\cdots\cdot((r-1)c-(r-2)).$$

This inequality, for various values of r and in various forms, is from [74, 81, 109, 118, 124, 132]

Proof. Unsurprisingly, the argument is by induction on r. The main step is identifying an inductive relationship between the number of copies of K_r and the number of copies of K_{r-1}. We will show, by induction on r,

if $r \geq 2$ and $t_{\mathrm{K}_2}(G) = c \geq \frac{r-1}{r}$ then

$$t_{\mathrm{K}_{r+1}}(G) \geq (rc-(r-1))t_{\mathrm{K}_r}(G).$$

Applying this inequality inductively immediately gives the statement of the theorem, so all that remains is inductively proving this inequality.

When $r = 1$, this is immediate (noting that we always have $t_{\mathrm{K}_1}(G) = 1$). So suppose $r \geq 2$ and that the statement holds for $r-1$.

We get an upper bound on $t_{\mathrm{K}_r}(G)$ using Cauchy-Schwarz:

$$\begin{aligned}
(t_{\mathrm{K}_r}(G))^2 &= \left(\iint c_{\mathrm{K}_r}(x_1,\ldots,x_r)\,d\mu_1(x_r)\,d\mu_{r-1}(x_1,\ldots,x_{r-1})\right)^2 \\
&= \left(\int c_{\mathrm{K}_{r-1}}(x_1,\ldots,x_{r-1})\int\prod_{i<r}\chi_E(x_i,x_r)\,d\mu_1(x_r)\,d\mu_{r-1}(x_1,\ldots,x_{r-1})\right)^2 \\
&\leq t_{\mathrm{K}_{r-1}}(G)\int c_{\mathrm{K}_{r-1}}(x_1,\ldots,x_{r-1})\left(\int\prod_{i<r}\chi_E(x_i,x_r)\,d\mu_1(x_r)\right)^2 d\mu_{r-1}(x_1,\ldots,x_{r-1}) \\
&= t_{\mathrm{K}_{r-1}}(G)\int c_{\mathrm{K}_{r-1}}(x_1,\ldots,x_{r-1})\prod_{i<r}\chi_E(x_i,y)\chi_E(x_iz)\,d\mu_{r+1}(x_1,\ldots,x_{r-1},y,z) \\
&= t_{\mathrm{K}_{r-1}}(G)\int c_{\mathrm{K}_{r-1}}(x_1,\ldots,x_{r-1})\prod_{i<r}\chi_E(x_i,y)\chi_E(x_iz)\,d\mu_{r+1}(x_1,\ldots,x_{r-1},y,z) \\
&= t_{\mathrm{K}_{r-1}}(G)t_{\mathrm{K}^-_{r+1}}(G)
\end{aligned}$$

where K^-_{r+1} is the graph from the preceding lemma containing all but one of the edges on $r+1$ vertices.

Combining this with the inequality from the previous lemma gives:

$$\begin{aligned}
(t_{\mathrm{K}_r}(G))^2 &\leq t_{\mathrm{K}_{r-1}}(G)t_{\mathrm{K}^-_{r+1}}(G) \\
&\leq t_{\mathrm{K}_{r-1}}(G)(\frac{r-1}{r}t_{\mathrm{K}_{r+1}}(G)+\frac{1}{r}t_{\mathrm{K}_r}(G)).
\end{aligned}$$

But applying the inductive hypothesis to one copy of $t_{\mathrm{K}_r}(G)$ on the left side of this inquality,

$$((r-1)c-(r-2))t_{\mathrm{K}_r}(G)t_{\mathrm{K}_{r-1}}(G) \leq t_{\mathrm{K}_{r-1}}(G)(\frac{r-1}{r}t_{\mathrm{K}_{r+1}}(G)+\frac{1}{r}t_{\mathrm{K}_r}(G)),$$

so canceling $\frac{t_{\mathrm{K}_{r-1}}(G)}{r}$ (which we may do since $c \geq \frac{r-1}{r} > 1-\frac{1}{r-1}$, so $t_{\mathrm{K}_{r-1}}(G) > 0$) from both sides gives

$$(r(r-1)c-r(r-2))t_{\mathrm{K}_r}(G) \leq (r-1)t_{\mathrm{K}_{r+1}}(G)+t_{\mathrm{K}_r}(G),$$

or

$$t_{\mathrm{K}_{r+1}}(G) \geq \frac{r(r-1)c-r(r-2)-1}{r-1}t_{\mathrm{K}_r}(G) = (rc-(r-1))t_{\mathrm{K}_r}(G).$$

□

When $t_{\mathrm{K}_2}(G) = 1-\frac{1}{s} = \frac{s-1}{s}$, the bound on the right is $\frac{s-1}{s}\frac{s-2}{s}\cdots\frac{s-(r-1)}{s} = t_{\mathrm{K}_r}(\mathrm{P}_s)$, so we have:

Theorem 4.15. *If $t_{\mathrm{K}_2}(G) = 1-\frac{1}{s}$ then $t_{\mathrm{K}_r}(G) \geq t_{\mathrm{K}_r}(\mathrm{P}_s)$.*

That is, for the specific edge densities $1-\frac{1}{s}$, the graph with s equal parts has as few copies of K_r as possible.

4.5 Interpolating Between s-Partite Graphs

The remaining question is what happens in the gaps between the graph P_s—say, the interval $t_{\mathrm{K}_2}(G) \in (1/2, 2/3)$. We know that $t_{\mathrm{K}_3}(G) \geq c(2c-1)$ in this interval, but we don't yet have examples of graphs achieving this minimum, so there might be a better lower bound.

There is a natural way to interpolate between the graphs P_s and P_{s+1}:

Definition 4.16. When $a \in [0, \frac{1}{s}]$, $\mathrm{P}_{s,a} = (V, E)$ is the measurable graph where:

- V is a disjoint union $V = \bigcup_{i \leq s+1} X_i$ where $\mu_1(X_{s+1}) = \frac{1-sa}{s+1}$ and $\mu_1(X_i) = \frac{1+a}{s+1}$ for $i \leq s$,
- $\{x, y\} \in E$ if and only if x and y are in different parts.

So $\mathrm{P}_{s,1/s} = \mathrm{P}_s$ while $\mathrm{P}_{s,0} = \mathrm{P}_{s+1}$, and when a is strictly between 0 and $\frac{1}{s}$, $\mathrm{P}_{s,a}$ is a graph with $s+1$ parts where one part is smaller than the others. We can imagine the sequence of graphs growing, starting with $\mathrm{P}_{1,1}$—the empty graph—and growing to $\mathrm{P}_{1,a}$ (a bipartite graph where one part is small); as a approaches 0, $\mathrm{P}_{1,a}$ begins to resemble the equal bipartite graph, and when a reaches 0, $\mathrm{P}_{1,0} = \mathrm{P}_2 = \mathrm{P}_{2,1/2}$, and a third part begins growing.

Our choice to index the graphs $\mathrm{P}_{s,a}$ in precisely this way may seem odd—for instance, our choice to arrange that a decreases as the density increases. The main reason is that, however we paramaterize the intermediate graphs, we will be more interested in $t_{\mathrm{K}_2}(\mathrm{P}_{s,a})$ than in a, and with this choice, we get the comparatively simple formula

$$\begin{aligned} t_{\mathrm{K}_2}(\mathrm{P}_{s,a}) &= \frac{s(1+a)}{s+1}\left[\frac{(s-1)(1+a)}{s+1} + \frac{1-sa}{s+1}\right] + \frac{1-sa}{s+1}\frac{s(1+a)}{s+1} \\ &= \frac{s}{s+1}(1-a^2) \end{aligned}$$

Lemma 4.17. $t_{\mathrm{K}_r}(\mathrm{P}_{s,a}) = \frac{(1+a)^{r-1}}{(s+1)^r}\binom{s+1}{r}(1-a(r-1))$.

Proof. There are two ways to find a copy of K_r in $\mathrm{P}_{s,a}$: by choosing all r vertices from the s large pieces, or by choosing one vertex from the small piece. So

$$\begin{aligned} t_{\mathrm{K}_r}(\mathrm{P}_{s,a}) &= r!\binom{s}{r}\left(\frac{1+a}{s+1}\right)^r + r!\binom{s}{r-1}\left(\frac{1+a}{s+1}\right)^{r-1}\frac{1-sa}{s+1} \\ &= \frac{r!(1+a)^{r-1}}{(s+1)^r}\left[\binom{s}{r}(1+a) + \binom{s}{r-1}(1-sa)\right] \\ &= \frac{r!(1+a)^{r-1}}{(s+1)^r}\left[\binom{s+1}{r} + a\left(\binom{s}{r} - s\binom{s}{r-1}\right)\right] \\ &= \frac{r!(1+a)^{r-1}}{(s+1)^r}\binom{s+1}{r}(1-a(r-1)). \end{aligned}$$

□

It will be convenient to name these values. For the remainder of this chapter, for any $c \in [0,1)$, we define:

- $s(c)$ is the unique integer s so that $c \in [1-\frac{1}{s}, 1-\frac{1}{s+1})$, and
- $a(c) = \sqrt{1-\frac{s(c)+1}{s(c)}c}$.

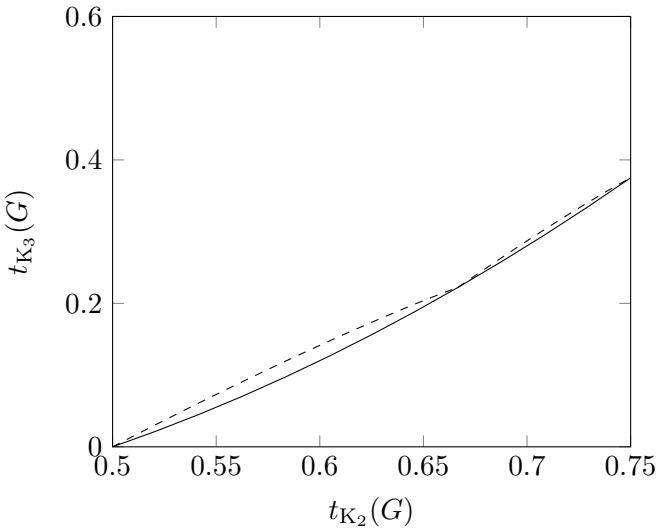
We have chosen these values precisely so that $t_{\mathrm{K}_2}(\mathrm{P}_{s(c),a(c)}) = c$.

We define

$$\mathcal{F}_r(c) = \frac{r!(1+a(c))^{r-1}}{(s(c)+1)^r}\binom{s(c)+1}{r}(1-a(c)(r-1))$$

so that $t_{\mathrm{K}_r}(\mathrm{P}_{s(c),a(c)}) = \mathcal{F}_r(c)$.

We can compare these densities to the lower bound we found in the previous section. For instance, in the case where $r = 3$,



The solid curve is the lower bound $t_{\mathrm{K}_2}(G)(2t_{\mathrm{K}_2}(G)-1)$ from the previous section. The dashed lines are $t_{\mathrm{K}_3}(\mathrm{P}_{s,a})$ when s, a are chosen to have the correct value of $t_{\mathrm{K}_2}(\mathrm{P}_{s,a})$. There is a gap, though a small one—that is, based on what we have seen so far, it is possible that there is a graph G with, say $t_{\mathrm{K}_2}(G) = 0.6$ but $t_{\mathrm{K}_3}(G) < t_{\mathrm{K}_3}(\mathrm{P}_{s(0.6),a(0.6)})$.

It was conjectured for decades that, in fact, there are no such graphs. This was settled in steps, for values of r and s, and shown in full generality by Reiher in 2012.

Theorem 4.18 ([132])**.** *For any G, $t_{\mathrm{K}_r}(G) \geq t_{\mathrm{K}_r}(\mathrm{P}_{s(t_{\mathrm{K}_2}(G)),a(t_{\mathrm{K}_2}(G))})$.*

Equivalently, for any G, $t_{\mathrm{K}_r}(G) \geq \mathcal{F}_r(t_{\mathrm{K}_2}(G))$. That is, $\mathrm{P}_{s,a}$ is actually the graph that has as few copies of $t_{\mathrm{K}_r}(G)$ as possible subject to $t_{\mathrm{K}_2}(G) = t_{\mathrm{K}_2}(\mathrm{P}_{s,a})$.

The proof involves looking at a graph which violates the conjecture by as much as possible—that is, a graph maximizing $\mathcal{F}_r(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_r}(G)$—and combining a great deal of careful work investigating the specific behavior

of the functions $\mathcal{F}_r(c)$ and $\mathcal{F}_r'(c)$ in various intervals with several different techniques for relating subgraph densities to limit the possibilities for what such an extremal graph can look like.

Rather than give the full proof, we will give the special case where $r = 3$ and $t_{K_2}(G) \in (1/2, 2/3)$ in the next section.

In order to do this, we need some specific elementary calculations involving $\mathcal{F}_3$, which we collect in the rest of this section. The calculations are purely algebra and calculus, and we gather them here only to avoid interrupting the flow of more substantive arguments to carry them out; the reader likely wishes to skip to the next section and refer back to these as needed.

Lemma 4.19. *$\mathcal{F}_3$ is concave down in each of the intervals $(1 - 1/s, 1 - 1/(s+1))$.*

Proof. From the formula for $\mathcal{F}_3(c)$, we can calculate $\mathcal{F}_3''$. Within each of these intervals, $s(c)$ is constant. Letting $s = s(c)$, we calculate:

$$a'(c) = -\frac{s+1}{2s\sqrt{1 - \frac{s+1}{s}c}} = -\frac{s+1}{2s \cdot a(c)}.$$

Then

$$\mathcal{F}_3'(c) = -\frac{36}{(s+1)^3}\binom{s+1}{3}a(c)(1+a(c))a'(c) = \frac{18}{s(s+1)^2}\binom{s+1}{3}(1+a(c))$$

and finally

$$\mathcal{F}_3''(c) = \frac{18}{s(s+1)^2}\binom{s+1}{3}a'(c)$$

which is negative since $a'(c)$ is always negative. □

Lemma 4.20. $\lim_{c \to 1-\frac{1}{s+1}^-} \mathcal{F}_3'(c) = \frac{3(s-1)}{s+1}$.

Proof. We can use the calculation from the previous proof: since $\lim_{c \to 1-\frac{1}{s+1}^-} a(c) = 0$, we have

$$\begin{aligned}
\lim_{c \to 1-\frac{1}{s+1}^-} \mathcal{F}_3'(c) &= \lim_{c \to 1-\frac{1}{s+1}^-} \frac{18}{s(s+1)^2}\binom{s+1}{3}(1+a(c)) \\
&= \frac{18}{s(s+1)^2}\binom{s+1}{3} \\
&= \frac{18}{s(s+1)^2}\frac{(s+1)s(s-1)}{6} \\
&= \frac{3(s-1)}{s+1}.
\end{aligned}$$

□

Lemma 4.21. *If $s > 2$ and $c \in (1-1/s, 1-1/(s+1))$ then $\mathcal{F}_3(c) < \frac{2}{3}\mathcal{F}_3'(c)c$.*

Proof. Since, on each interval $(1-1/s, 1-1/(s+1))$, $\mathcal{F}_3(c)$ is increasing while $\mathcal{F}_3'(c)$ is decreasing, it suffices to consider the limit as c approaches $1-\frac{1}{s+1}$ from the left. At this limit, $\mathcal{F}_3'(c) = \frac{3(s-1)}{s+1}$ by the previous lemma while $\mathcal{F}_3(c) = \frac{(s-1)(s-2)}{s^2}$. In this whole interval, $c > 1-\frac{1}{s}$, so

$$\begin{aligned}
\frac{2}{3}\mathcal{F}_3'(c)c - \mathcal{F}_3(c) &> 2c\frac{(s-1)}{s+1} - \frac{(s-1)(s-2)}{s^2} \\
&> 2\frac{(s-1)^2}{s(s+1)} - \frac{(s-1)(s-2)}{s^2} \\
&= \frac{s-1}{s}\left[2\frac{s-1}{s+1} - \frac{s-2}{s}\right] \\
&= \frac{s-1}{s}\frac{2s^2-2s-(s^2-s-2)}{s(s+1)} \\
&= \frac{s-1}{s}\frac{s^2-3s+2}{s(s+1)}.
\end{aligned}$$

The polynomial s^2-3s+2 has roots 1 and 2; in particular, this means that once $s > 2$, each of these terms is positive, so the expression is positive as well. □

4.6 Minimizing Triangles when the Density is in $(1/2, 2/3)$

We consider a graph maximizing $\mathcal{F}_3(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_3}(G)$ and spend most of this section deriving various properties such a graph must have. We will conclude the section by proving the special case of Theorem 4.18 where $r = 3$ and $t_{\mathrm{K}_2}(G) \in (1/2, 2/3)$. (This was shown in [130], and the proof here follows that paper closely.)

The main insight is the generalization of what we did in the proof of Turán's Theorem: an extremal graph (that is, a graph maximizing a function) must have some corresponding derivative equal to 0.

Lemma 4.22. *Suppose that $H(z_2, \ldots, z_d)$ is a function and G is a graph such that among all graphs, G maximizes*

$$H(t_{\mathrm{K}_2}(G), \ldots, t_{\mathrm{K}_d}(G))$$

and all partial derivatives of H exist at $(t_{\mathrm{K}_2}(G), \ldots, t_{\mathrm{K}_d}(G))$.

Then, for almost every vertex x,

$$\sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G,x) = \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G).$$

In our case, we will apply this with the function $H(z_2, z_r) = \mathcal{F}_r(z_2) - z_r$, so that

$$H(t_{\mathrm{K}_2}(G), t_{\mathrm{K}_r}(G)) = \mathcal{F}_r(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_r}(G).$$

This function measures how big the "deficiency" in G is—how many copies of K_r it is missing relative to $\mathrm{P}_{s(t_{\mathrm{K}_2}(G)),a(t_{\mathrm{K}_2}(G))}$.

Proof. Suppose not. Then there is some $\epsilon > 0$ so that there is a set A of positive measure such that, for all $x \in A$,

$$\left| \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G,x) - \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G) \right| \ge \epsilon.$$

Either those $x \in A$ such that

$$\sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G,x) \ge \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G) + \epsilon$$

or those x such that

$$\sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G) \ge \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G,x) + \epsilon$$

must have positive measure. Without loss of generality, let B be a set of positive measure so that, for all $x \in B$,

$$\sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G,x) \ge \sum_{2\le i\le d} i\frac{\partial H}{\partial z_i} t_{\mathrm{K}_i}(G) + \epsilon.$$

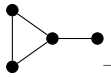
(This other case is symmetric.)

For $i \in [2,d]$, let $u_i = \frac{1}{\mu_1(B)} \int_B t_{\mathrm{K}_i}(G,x)\, d\mu_1$. Let $G_\delta = (V, E, \nu_\delta^B)$ as in Lemma 4.5 and let $h_i(\delta) = t_{\mathrm{K}_i}(G_\delta)$. Then, by Lemma 4.5, $h_i'(0) = i(u_i - t_{\mathrm{K}_i}(G))$.

Therefore

$$\begin{aligned}\frac{d}{d\delta}H(t_{\mathrm{K}_2}(G_\delta),\dots,t_{\mathrm{K}_d}(G_\delta))|_{\delta=0} &= \sum_{2\le i\le d} t'_{\mathrm{K}_i}(G_\delta)|_{\delta=0}\frac{\partial H}{\partial i}\\ &= \sum_{2\le i\le d} i(u_i - t_{\mathrm{K}_i}(G))\frac{\partial H}{\partial i}\\ &= \frac{1}{\mu_1(B)}\int_B \sum_{2\le i\le d} i(t_{\mathrm{K}_i}(G,x) - t_{\mathrm{K}_i}(G))\frac{\partial H}{\partial i}\\ &\ge \frac{\epsilon}{\mu_1(B)}.\end{aligned}$$

Since this derivative is positive, there must be some small enough δ so that $H(t_{\mathrm{K}_2}(G_\delta),\dots,t_{\mathrm{K}_d}(G_\delta)) > H(t_{\mathrm{K}_2}(G),\dots,t_{\mathrm{K}_d}(G))$, contradicting the maximality of G. This contradiction completes the proof. □

In this section, we just need the following consequence. Let us write K_3^+ for the graph  —that is, K_3 with one extra edge attached.

Lemma 4.23. *Suppose G is a graph maximizing $\mathcal{F}_3(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_3}(G)$ and $a(t_{\mathrm{K}_2}(G)) \neq \frac{1}{s(t_{\mathrm{K}_2}(G))}$. Then, taking $\lambda = \mathcal{F}_3'(t_{\mathrm{K}_2}(G))$,*

$$2\lambda t_{\mathrm{V}}(G) - 3t_{\mathrm{K}_3^+}(G) = (2\lambda t_{\mathrm{K}_2}(G) - 3t_{\mathrm{K}_3}(G))t_{\mathrm{K}_2}(G).$$

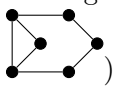
Proof. Since $a(t_{\mathrm{K}_2}(G)) \neq \frac{1}{s(t_{\mathrm{K}_2}(G))}$, we are not at one of the sharp corners in $\mathcal{F}_3$, so $\mathcal{F}_3'$ exists and by the previous lemma, for almost every x we have

$$2\mathcal{F}_3'(t_{\mathrm{K}_2}(G))t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x) = 2\lambda t_{\mathrm{K}_2}(G) - 3t_{\mathrm{K}_3}(G).$$

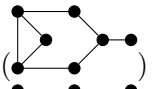
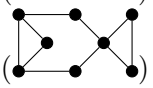
So we may calculate

$$\begin{aligned}2\lambda t_{\mathrm{V}}(G) - 3t_{\mathrm{K}_3^+}(G) &= 2\lambda\int \chi_E(x,y)t_{\mathrm{K}_2}(G,x)\,d\mu_2 - 3\int \chi_E(x,y)t_{\mathrm{K}_3}(G,x)\,d\mu_2\\ &= \int \chi_E(x,y)(2\lambda t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x))\,d\mu_2\\ &= \int \chi_E(x,y)\,d\mu_2(2\lambda t_{\mathrm{K}_2}(G) - 3t_{\mathrm{K}_3}(G))\\ &= t_{\mathrm{K}_2}(G)(2\lambda t_{\mathrm{K}_2}(G) - 3t_{\mathrm{K}_3}(G)).\end{aligned}$$

□

This is a typical application. By replacing χ_E with some other graph, we could prove a more general result: in a graph maximizing $\mathcal{F}_3(t_{\mathrm{K}_2}(G))-t_{\mathrm{K}_3}(G)$, for any graph H (for instance, ) we have

$$2\lambda t_{H+\mathrm{K}_2}(G) - 3t_{H+\mathrm{K}_3}(G) = (2\lambda t_{\mathrm{K}_2}(G) - 3t_{\mathrm{K}_3}(G))t_H(G)$$

where $H+\mathrm{K}_2$ is the result of attaching an edge to a vertex of H () and $H+\mathrm{K}_3$ is the result of attaching a triangle to the same vertex () In this notation, V is $\mathrm{K}_2+\mathrm{K}_2$ and K_3^+ is $\mathrm{K}_2+\mathrm{K}_3$.*

We can do something similar with edges.

Lemma 4.24. *Suppose that $H(z_2,\dots,z_d)$ is a function and G is a graph such that among all graphs, G maximizes*

$$H(t_{\mathrm{K}_2}(G),\dots,t_{\mathrm{K}_d}(G))$$

and all partial derivatives of H exist at $(t_{\mathrm{K}_2}(G),\dots,t_{\mathrm{K}_d}(G))$.

Then, for almost every edge $\{x,y\}\in E$,

$$\sum_{2\le i\le d}\binom{i}{2}\frac{\partial H}{\partial z_i}t_{\mathrm{K}_i}(G,x,y)\ge 0.$$

Note that with vertices we obtained an equality, but with edges we get an inequality. This is because when edges are "suboptimal", we could delete them and improve G, but when edges are "superoptimal", we cannot duplicate them. In the language of optimization, we are encountering a boundary condition—an edge cannot have value greater than 1—and so, at such the boundary, we can only expect a derivative to be non-positive. (We could find a dual lemma, showing an inequality in the other direction when $\{x,y\}\in V^2\setminus E$, but we will not need this here.)

Proof. Again, suppose not. When $\chi_E(x,y)=0$, also $t_{\mathrm{K}_t}(G,x,y)=0$, so for the claim to fail, there must be an $\epsilon>0$ so that, letting $B=\{(x,y)\in E\mid$

*Strictly speaking, we are attaching K_2 and K_3 not to just a graph H, but to a graph with a distinguished vertex. In the terminology of [130], we are "adding 1-flags"—taking two graphs, each with a distinguished vertex, and combining the distinguished vertices.

$\sum_{2\le i\le d}\binom{i}{2}\frac{\partial H}{\partial z_i}t_{\mathrm{K}_i}(G,x,y) < -\epsilon\}$, $\mu_2(B) > 0$. For each $\delta \in (0, \mu_2(B))$, choose some $B_\delta \subseteq B$ with $\mu_1(B_\delta) = \delta$. Set $G_\delta = (V, E \setminus B_\delta, \mu_1)$. Then

$$\begin{aligned} t_{\mathrm{K}_i}(G_\delta) &= \int \prod_{j,j'\le i} \chi_{E\setminus B_\delta}(x_j, x_{j'})\, d\mu_i \\ &= \int \prod_{j,j'\le i} \chi_E(x_j,x_{j'})(1-\chi_{B_\delta}(x_j,x_{j'}))\, d\mu_i \\ &= \int \sum_{S\subseteq\binom{i}{2}} (-1)^{|S|} \prod_{(j,j')\in\binom{t}{2}\setminus S} \chi_E(x_j,x_{j'}) \prod_{(j,j')\in S} \chi_{B_\delta}(x_j,x_{j'})\, d\mu_i. \end{aligned}$$

Once again, we can take the limit as δ goes to 0, so all terms where $|S| > 1$ will become negligible:

$$\begin{aligned} \lim_{\delta\to 0^+} \frac{t_{\mathrm{K}_i}(G_\delta) - t_{\mathrm{K}_i}(G)}{\delta} &= \frac{1}{\delta}\int \sum_{S\subseteq\binom{i}{2}, 0<|S|} (-1)^{|S|} \prod_{(j,j')\in\binom{t}{2}\setminus S} \chi_E(x_j,x_{j'}) \prod_{(j,j')\in S} \chi_{B_\delta}(x_j,x_{j'})\, d\mu_i \\ &= -\binom{i}{2}\int \chi_{B_\delta}(x_1,x_2) \prod_{(j,j')\in\binom{i}{2}\setminus\{(1,2)\}} \chi_E(x_j,x_{j'})\, d\mu_i \\ &= -\binom{i}{2}\int_{B_\delta} t_{\mathrm{K}_i}(G,x,y)\, d\mu_2. \end{aligned}$$

Therefore the right-hand derivative

$$\begin{aligned} \frac{d}{d_+\delta} H(t_{\mathrm{K}_2}(G_\delta),\dots,t_{\mathrm{K}_d}(G_\delta))|_{\delta=0} &= \sum_{2\le i\le d} -\binom{i}{2}\int_{B_\delta} t_{\mathrm{K}_i}(G,x,y)\, d\mu_2 \frac{\partial H}{\partial z_i} \\ &> \epsilon. \end{aligned}$$

But this means that, for some sufficiently small δ, $H(t_{\mathrm{K}_2}(G_\delta),\dots,t_{\mathrm{K}_d}(G_\delta)) > H(t_{\mathrm{K}_2}(G),\dots,t_{\mathrm{K}_d}(G))$, contradicting the maximality of G. □

Recall the graphs $\mathrm{K}_{r-1}(r)$ from Lemma 4.13; in particular, we need the graphs $\mathrm{K}_2(3)$—  —and $\mathrm{K}_3(4)$—  .

Lemma 4.25. *Suppose G is a graph maximizing $\mathcal{F}_3(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_3}(G)$ $a(t_{\mathrm{K}_2}(G)) \neq \frac{1}{s(t_{\mathrm{K}_2}(G))}$. Then, taking $\lambda = \mathcal{F}_3'(t_{\mathrm{K}_2}(G))$,*

$$3t^{ind}_{\mathrm{K}_3^+}(G) + 3t^{ind}_{\mathrm{K}_3(4)}(G) \le \mathcal{F}_3'(t_{\mathrm{K}_2}(G)) t^{ind}_{\mathrm{K}_2(3)}(G).$$

Proof. Since $a(t_{\mathrm{K}_2}(G)) \neq \frac{1}{s(t_{\mathrm{K}_2}(G))}$, we are not at one of the sharp corners in $\mathcal{F}_3$, so $\mathcal{F}_3'$ exists and by the previous lemma, for almost every x, y we have

$$0 \leq 2\mathcal{F}_3'(t_{\mathrm{K}_2}(G))\chi_E(x,y) - 6t_{\mathrm{K}_3}(G,x,y),$$

which is equivalent to

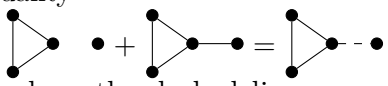
$$3t_{\mathrm{K}_3}(G,x,y) \leq \mathcal{F}_3'(t_{\mathrm{K}_2}(G))\chi_E(x,y).$$

Therefore

$$\begin{aligned}
3t^{ind}_{\mathrm{K}_3^+}(G) + 3t^{ind}_{\mathrm{K}_3(4)}(G) &= 3\int \chi_E(x,y)\chi_E(x,w)\chi_E(y,w) \\
&\qquad \cdot (1-\chi_E(x,z))(1-\chi_E(y,z))\left[\chi_E(z,w) + (1-\chi_E(z,w))\right] d\mu_4 \\
&= 3\int \chi_E(x,y)\chi_E(x,w)\chi_E(y,w)(1-\chi_E(x,z))(1-\chi_E(y,z))\, d\mu_4 \\
&= \int 3t_{\mathrm{K}_3}(G,x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\, d\mu_3 \\
&\leq \int \mathcal{F}_3'(t_{\mathrm{K}_2}(G))\chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\, d\mu_3 \\
&= \mathcal{F}_3'(t_{\mathrm{K}_2}(G))t^{ind}_{\mathrm{K}_2(3)}(G).
\end{aligned}$$

□

We could think of the first two lines of the calculation as expressing the "equality"



where the dashed line represents an edge which may or may not be present. In particular, the quantity

$$\int \chi_E(x,y)\chi_E(x,w)\chi_E(y,w)(1-\chi_E(x,z))(1-\chi_E(y,z))\, d\mu_4$$

counts the induced copies of this third "graph", consisting of a triangle x, y, w and an additional vertex which is definitely not adjacent to x or y but might or might not be adjacent to w.

Again, the quantity $(1-\chi_E(x,z))(1-\chi_E(y,z))$ could be replaced by any other function of x, y when we apply the inequality, giving relationships between other pairs of graphs, and the flag algebra framework [130] introduces notation for dealing with these combinations, as well as more general notation for graphs where some edges must be present, some must be absent, and some might be present or absent.

The reader might wonder where these graphs are coming from and why we are counting them. We have no satisfying answer to that, except that they will make the proof work: discovering the argument presumably involved a substantial amount of experimenting with the relationships between various graph densities to discover the essential ones.

We also need two general relationships, in the style of Lemma 4.13, among different subgraph densities in arbitrary graphs.

Lemma 4.26.

$$t_{\mathrm{K}_4}(G) + 2t_{\mathrm{K}_3}(G) \leq 3t_{\mathrm{K}_3^+}(G) + 3t^{ind}_{\mathrm{K}_3^+}(G) + 3t^{ind}_{\mathrm{K}_{3(4)}}(G).$$

Proof. As in the proof of Lemma 4.13, we do this by considering each possible set of four vertices and considering how many copies of each of these graphs it could contribute. For this we need the analog

$$c^{ind}_H(x_1, \ldots, x_n) = \frac{1}{n!}|\{\pi \text{ a permutation of } \{1, \ldots, n\} \mid (x_{\pi(1)}, \ldots, x_{\pi(n)}) \in T^{ind}_H(G)$$

of c_H. Then we have $\int c^{ind}_H \, d\mu_n = t^{ind}_H(G)$ as in Lemma 4.12. So it suffices to show that, for every (x, y, z, w),

$$c_{\mathrm{K}_4}(x, y, z, w) + 2c_{\mathrm{K}_{3(4)}}(x, y, z, w) \leq 3c_{\mathrm{K}_3^+}(x, y, z, w) + 3c^{ind}_{\mathrm{K}_3^+}(x, y, z, w) + 3c^{ind}_{\mathrm{K}_{3(4)}}(x, y, z, w)$$

Observe that:

- if $\{x, y, z, w\}$ contains no triangles, all four terms are 0,
- if $\{x, y, z, w\}$ contains one triangle and no extra edges,

$$c_{\mathrm{K}_4}(x, y, z, w) = 0, c_{\mathrm{K}_{3(4)}}(x, y, z, w) = 1/4,$$

$$c_{\mathrm{K}_3^+}(x, y, z, w) = 0, c^{ind}_{\mathrm{K}_3^+}(x, y, z, w) = 0, c^{ind}_{\mathrm{K}_{3(4)}}(x, y, z, w) = 1/4,$$

- if $\{x, y, z, w\}$ contains one triangle and one extra edge,

$$c_{\mathrm{K}_4}(x, y, z, w) = 0, c_{\mathrm{K}_{3(4)}}(x, y, z, w) = 1/4,$$

$$c_{\mathrm{K}_3^+}(x, y, z, w) = 1/4, c^{ind}_{\mathrm{K}_3^+}(x, y, z, w) = 1/4, c^{ind}_{\mathrm{K}_{3(4)}}(x, y, z, w) = 0,$$

- if $\{x, y, z, w\}$ contains one triangle and two extra edges,

$$c_{\mathrm{K}_4}(x, y, z, w) = 0, c_{\mathrm{K}_{3(4)}}(x, y, z, w) = 1/2,$$

$$c_{\mathrm{K}_3^+}(x, y, z, w) = 1/3, c^{ind}_{\mathrm{K}_3^+}(x, y, z, w) = 0, c^{ind}_{\mathrm{K}_{3(4)}}(x, y, z, w) = 0,$$

- if $\{x, y, z, w\}$ is a copy of K_4,

$$c_{\mathrm{K}_4}(x, y, z, w) = 1, c_{\mathrm{K}_3(4)}(x, y, z, w) = 1,$$

$$c_{\mathrm{K}_3^+}(x, y, z, w) = 1, c^{ind}_{\mathrm{K}_3^+}(x, y, z, w) = 0, c^{ind}_{\mathrm{K}_3(4)}(x, y, z, w) = 0.$$

□

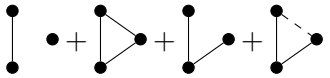
We need one more equality of this kind. We could show it the same way, by counting up what happens in each configuration, but we can also simply calculate the integrals.

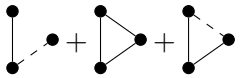
Lemma 4.27.

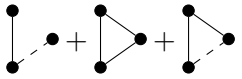
$$t^{ind}_{\mathrm{K}_2(3)} + 2t_{\mathrm{V}}(G) = t_{\mathrm{K}_2}(G) + t_{\mathrm{K}_3}(G).$$

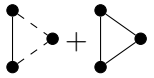
Proof. We can just combine the integrals in the correct way. For illustration, we draw the "graphs" whose counting is represented by the integral:

$$\begin{aligned}
&t^{ind}_{\mathrm{K}_2(3)} + 2t_{\mathrm{V}}(G) \\
&= \int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z)) + 2\chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z)) + \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \\
&\quad + \chi_E(x,y)\chi_E(x,z)(1-\chi_E(y,z)) + \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)(1-\chi_E(y,z)) + \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \\
&\quad + \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)(1-\chi_E(y,z)) + \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \\
&\quad + \chi_E(x,y)\chi_E(y,z)\,d\mu_3 \\
&= \int \chi_E(x,y) + \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 \\
&= t_{\mathrm{K}_2}(G) + t_{\mathrm{K}_3}(G).
\end{aligned}$$









□

With just a bit of calculating, we are prepared to prove the first case of Theorem 4.18.

Theorem 4.28. *For any G with $t_{K_2}(G) \in (1/2, 2/3)$, $t_{K_3}(G) \geq \mathcal{F}_3(t_{K_2}(G))$.*

Proof. By Theorem 3.33, there must be some G maximizing the gap $\mathcal{F}_3(t_{K_2}(G)) - t_{K_3}(G)$, and we will work in this G. Let $c = t_{K_2}(G)$, $b = t_{K_3}(G)$, $s = s(c) = 2$, and $a = a(c)$. The function $\mathcal{F}_3$ is differentiable in the interval $(1-1/2, 1-1/3)$, so let $\lambda = \mathcal{F}_3'(c)$.

Combining, in order, Lemmata 4.26, 4.25, 4.23, and 4.27 we have

$$\begin{aligned} t_{K_4}(G) + 2b &\leq 3t_{K_3^+}(G) + 3t^{ind}_{K_3^+}(G) + 3t^{ind}_{K_3(4)}(G) \\ &\leq 3t_{K_3^+}(G) + \lambda t^{ind}_{K_2(3)}(G) \\ &= 2\lambda t_V(G) - (2\lambda c - 3b)c + \lambda t^{ind}_{K_2(3)}(G) \\ &= 2\lambda t_V(G) - (2\lambda c - 3b)c + \lambda(c + b - 2t_V(G)) \end{aligned}$$

In particular, since $t_{K_4}(G) \geq 0$, we can isolate b:

$$b(2 - 3c - \lambda) \leq \lambda c(1 - 2c).$$

By Lemma 4.19, the function $\mathcal{F}_3(x)$ is concave down on the interval $(1/2, 2/3)$. In particular, for every $c \in (1/2, 2/3)$, $\mathcal{F}_3'(c) > \lim_{c \to 2/3} \mathcal{F}_3'(c) = 1$ by Lemma 4.20, so

$$b \geq \frac{\lambda c(1 - 2c)}{2 - 3c - \lambda}.$$

We are left with some tedious algebra to show that the right side of this equality is, in fact, equal to $\mathcal{F}_3(c)$. Since $s(c) = 2$, we have

$$\mathcal{F}_3(c) = \frac{2(1 + a)^2}{9}(1 - 2a) \text{ and } \mathcal{F}_3'(c) = 1 + a.$$

It is more convenient to work in terms of a, so noting that $c = \frac{2}{3}(1 - a^2)$, we

have

$$\begin{aligned}
\frac{\lambda c(1-2c)}{2-3c-\lambda} &= \frac{(1+a)\cdot c(1-2c)}{2-3c-(1+a)} \\
&= (1+a)\frac{\frac{2}{3}(1-a^2)(1-\frac{4}{3}(1-a^2))}{2-2(1-a^2)-(1+a)} \\
&= \frac{2}{9}(1+a)\frac{-1+5a^2-4a^4}{2a^2-a-1} \\
&= \frac{2}{9}(1+a)\frac{-(a-1)(a+1)(2a-1)(2a+1)}{(2a+1)(a-1)} \\
&= \frac{2}{9}(1+a)(-(a+1)(2a-1)) \\
&= \frac{2(1+a)^2}{9}(1-2a) \\
&= \mathcal{F}_3(c).
\end{aligned}$$

Therefore $b \geq \mathcal{F}_3(c)$.

This holds for the graph G maximizing $\mathcal{F}_3(c)-b$, so for all G, $\mathcal{F}_3(t_{\mathrm{K}_2}(G)) \leq t_{\mathrm{K}_3}(G)$. □

4.7 Remarks

We have cheated a bit with our definition of absolute continuity: normally one says that $\nu \ll \mu$ if, for every set S with $\mu(S) = 0$, also $\nu(S) = 0$. It is immediate that our definition—that $\nu(S) = \int_S f\, d\mu$ for some f—implies this, and the quite non-trivial Randon-Nikodym theorem shows that these two definitions are equivalent.

While we have focused in this chapter on proving exact inequalities for densities in measurable graphs, any of these theorems could also be viewed as a statement about approximate inequalities in sufficiently large finite graphs. Since statements about measures and integrals are not first-order, finding the right statement can sometimes be a bit trickier. Many arguments in the literature handle this in an ad hoc way (as in [90]), which usually suffices. A general framework extending first-order logic by predicates for measures exists as well [71].

Razborov's *flag algebra* framework [130] gives a sort of calculus for calculating the sorts of inequalities between graph densities we needed several of. For examples much beyond the sorts here, working by hand quickly becomes unwieldy, but examples have been calculated using the *Flagmatic* software [162], including a variety of proofs of new extremal results [52, 53].

Chapter 5

Structure and Randomness

5.1 Quasirandomness

We can identify quasirandom (rather than just ϵ-quasirandom) measurable graphs.

Definition 5.1. $G = (V, E, \mu_1)$ is *quasirandom* if $t_{C_4}(G) = (t_{K_2}(G))^4$.

By Lemma 1.22, $t_{C_4}(G) \geq (t_{K_2}(G))^4$ for any graph G. (We originally proved this in the setting of finite graphs, but our proof, written entirely in terms of calculations with integrals, was really always in the measurable graph setting.) So again, quasirandomness means that $t_{C_4}(G)$ is as small as possible: the value of $t_{K_2}(G)$ forces there to be a certain number of cycles of length 4, and a quasirandom graph has only as many as it has to.

Theorem 5.2. *$[G_n]_{\mathcal{U}}$ is quasirandom if and only if, for every $\epsilon > 0$, $\{n \mid G_n$ is ϵ-quasirandom$\} \in \mathcal{U}$.*

Proof. We have $t_{C_4}([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} t_{C_4}(G_n)$ and $(t_{K_2}([G_n]_{\mathcal{U}}))^4 = \lim_{n\to\mathcal{U}}(t_{K_2}(G_n))^4$.

If, for every $\epsilon > 0$, $\{n \mid G_n$ is ϵ-quasirandom$\} \in \mathcal{U}$ then, for every $\epsilon > 0$,

$$t_{C_4}([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} t_{C_4}(G_n) \leq \lim_{n\to\mathcal{U}}(t_{K_2}(G_n))^4 + \epsilon = (t_{K_2}([G_n]_{\mathcal{U}}))^4 + \epsilon,$$

so $t_{C_4}([G_n]_{\mathcal{U}}) = (t_{K_2}([G_n]_{\mathcal{U}}))^4$.

If there is an $\epsilon > 0$ such that $\{n \mid G_n$ is ϵ-quasirandom$\} \notin \mathcal{U}$ then

$$t_{C_4}([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} t_{C_4}(G_n) \geq \lim_{n\to\mathcal{U}}(t_{K_2}(G_n))^4 + \epsilon = (t_{K_2}([G_n]_{\mathcal{U}}))^4 + \epsilon.$$

□

As in the finite case, quasirandomness implies that edges are evenly distributed, in the sense that whenever X and Y are sets, the density of edges between X and Y is the same as the density of edges in total.

Theorem 5.3. *If $G = (V, E, \mu_1)$ is quasirandom then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $\mathcal{B}_1$,*

$$\mu_2(E \cap (X \times Y)) = \mu_2(E)\mu_1(X)\mu_1(Y).$$

The proof we gave of the analogous theorem in the finite setting, Theorem 1.28, goes through essentially unchanged. Rather than repeat that here, we will wait for Corollary 5.23, where we will give another proof building on the machinery we develop below.

Using this, we can give our long delayed proof of Theorem 1.31, that quasirandom graphs have the correct subgraph density for all finite graphs. We first show an analogous statement for the ultraproduct and then, in the next section, use properties of ultraproducts to prove the original version of the theorem.

Theorem 5.4. *If $G = (V, E, \mu_1)$ is quasirandom then for every finite graph $H = (W, F)$, we have $t_H(G) = p^{|F|}$ where $p = t_{\mathrm{K}_2}(G) = \mu_2(E)$.*

Proof. First, consider the case where $H = \mathrm{C}_3$, the triangle, to get the main idea. In this case

$$t_{\mathrm{C}_3}(G) = \int \chi_E(x, y)\chi_E(y, z)\chi_E(z, x)\, d\mu_3.$$

That is, $t_{\mathrm{C}_3}(G)$ is the probability that, when we select three random vertices x, y, and z, that they form a triangle. Using the Fubini property, this is equal to

$$\int \left(\int \chi_E(x, y)\chi_E(y, z)\chi_E(z, x) d\mu_2(x, y) \right) d\mu(z).$$

That means that we fix the vertex z and, for each z, ask what the probability that two random vertices x and y will form a triangle with z is.

So consider some fixed vertex z. It has a neighborhood $N_G(z) = \{x \in V \mid \{x, z\} \in E\}$. In order for (x, y) to form a triangle with z, we need $x \in N_G(z)$, $y \in N_G(z)$, and also $\{x, y\} \in E$.

That is, we are looking for

$$\mu_2([N_G(z) \times N_G(z)] \cap E).$$

Quasirandomness—specifically, Theorem 5.3—implies that this is equal to

$$p\mu(N_G(z))^2.$$

If we pick two random vertices, the probability that they have an edge between them is p. Quasirandomness tells us that if we pick two random neighbors of z, the probability that they have an edge between them is still p—the neighbors of z are neither more nor less likely to have edges between them.

In particular, that means

$$t_{C_3}(G) = p \int \chi_E(y,z)\chi_E(z,x)\, d\mu_3.$$

Notice that $\int \chi_E(y,z)\chi_E(z,x)\, d\mu_3$ is the same as $t_V(G)$ where V is the graph  , so we have reduced counting triangles to counting copies of a graph with one fewer edge.

To find $t_V(G)$, we can repeat this argument with a different edge:

$$t_{C_3}(G) = p \cdot t_V(G) = p \int (\chi_E(y,z)\chi_E(z,x) d\mu_2(y,z))\, d\mu(x).$$

That is, we are fixing x and looking for

$$\mu_2([V \times N_G(x)] \cap E).$$

Theorem 5.3 applies again: for each x, this is equal to

$$p\mu(V)\mu(N_G(x)) = \int \chi_E(z,x) d\mu(y),$$

so

$$t_{C_3}(G_{\mathcal{U}}) = p^2 \int \chi_E(z,x)\, d\mu_3 = p^3.$$

The general argument follows the same structure. We proceed by induction on the number of edges, and at each step, we choose a single edge $e \in F$ and look at the iterated integral where we fix all the vertices except the two vertices on that edge. Quasirandomness will show that $t_H(G) = p \cdot t_{H^-}(G)$ where $H^- = (W, F \setminus \{e\})$. This reduces us to finding $t_{H^-}(G)$, which is covered by the inductive hypothesis since H^- has one fewer edge.

Formally, note that

$$\begin{aligned} t_H(G) &= \mu_{|W|}(\{(x_1, \ldots, x_{|W|}) \mid \text{for each } \{v_i, v_j\} \in F,\ \{x_i, x_j\} \in E_{\mathcal{U}}\}) \\ &= \int \prod_{\{v_i,v_j\} \in F} \chi_E(x_i, x_j) d\mu_{|W|}(x_1, \ldots, x_{|W|}). \end{aligned}$$

We proceed by induction on $|F|$. If $|F| = 0$, so H is a graph with no edges, then $t_H(G) = 1$ and the claim is immediate.

So suppose $|F| > 0$. Pick some edge $\{v_{j_0}, v_{j_1}\} \in F$. Then we will calculate $t_H(G)$ using the integral

$$t_H(G) = \int \mu_2(\{(x_{j_0}, x_{j_1}) \mid \text{for each } \{v_i, v_j\} \in F,\ \{x_i, x_j\} \in E\})\mu_{|W|-2}.$$

This corresponds to first choosing $\bar{x} \in V^{|W|-2}$, representing all the vertices other than x_{j_0} and x_{j_1}, and then asking how many ways there are to choose x_{j_0} and x_{j_1} so that we get a copy of H.

In order for $\bar{x}, x_{j_0}, x_{j_1}$ to be a copy of H, we need four things to happen: $\bar{x}$ needs to contain all the edges it supposed to have, x_{j_0} has to be adjacent to certain vertices in $\bar{x}$, x_{j_1} has to be adjacent to certain vertices in $\bar{x}$, and x_{j_0} and x_{j_1} have to themselves be adjacent.

We can split up these requirements. First, let us take $P^- \subseteq V^{|W|-2}$ to consist of those $\bar{x} = (x_1, \ldots, x_k) \in V^{|W|-2}$ (omitting the indices x_{j_0} and x_{j_1}) such that whenever $\{v_i, v_j\} \in F$ with $i, j \notin \{j_0, j_1\}$, $\{x_i, x_j\} \in E$. (Phrased another way, if we let H' be the subgraph of H obtained by deleting v_{j_0} and v_{j_1}, and all edges incident on either, P^- is precisely $T_{H'}(G)$, the set of copies of H', indexed appropriately.)

Given $\bar{x} = (x_1, \ldots, x_k)$ (omitting the indices x_{j_0} and x_{j_1}), let $P^0_{\bar{x}} \subseteq V$ consist of those x which are suitable choices for x_{j_0}: $x \in P^0$ if, for each i such that $\{v_i, v_{j_0}\} \in F$, $\{x_i, x\} \in E$. Similarly, let $P^1_{\bar{x}} \subseteq V$ consist of those x which are suitable choices for x_{j_1}: $x \in P^1$ if, for each i such that $\{v_i, v_{j_1}\} \in F$, $\{x_i, x\} \in E$.

Then

$$t_H(G) = \int_{P^-} \mu_2(E \cap (P^0_{\bar{x}} \times P^1_{\bar{x}}))d\mu_{|W|-2}(\bar{x}).$$

Here we use quasirandomness: no matter what $\bar{x}$ we choose, $\mu_2(E \cap (P^0_{\bar{x}} \times P^1_{\bar{x}})) = p\mu(P^0_{\bar{x}})\mu(P^1_{\bar{x}})$, so

$$t_H(G) = p \int_{P^-} \mu(P^0_{\bar{x}})\mu(P^1_{\bar{x}})d\mu_{|W|-2}(\bar{x}).$$

But $\int_{P^-} \mu(P^0_{\bar{x}})\mu(P^1_{\bar{x}})d\mu_{|W|-2}(\bar{x})$ is precisely $t_{(W,F\setminus\{\{v_{j_0},v_{j_1}\})}}(G)$, which, by the inductive hypothesis, is equal to $p^{|F|-1}$, so $t_H(G) = p \cdot p^{|F|-1} = p^{|F|}$. □

5.2 Consequences for Finite Graphs

We can use Theorem 5.4 to give a proof of Theorem 1.31. The technique is typical of the way we obtain results about finite graphs using ultraproducts.

Theorem (1.31)**.** *For every finite graph $H = (W, F)$, each $\epsilon > 0$, there is a $\delta > 0$ so that if $G = (V, E)$ is δ-quasirandom, $|t_H(G) - t_{K_2}(G)^{|F|}| < \epsilon$.*

Proof. Suppose not. Then there is a finite graph $H = (W, F)$ and an $\epsilon > 0$ so that for every $n > 0$ there is a $G_n = (V_n, E_n)$ which is $1/n$-quasirandom, $|V_n| \geq n$, and

$$|t_H(G_n) - (t_{K_2}(G_n))^{|F|}| \geq \epsilon.$$

Let $[G_n]_{\mathcal{U}}$ be an ultraproduct of the sequence $\langle G_n \rangle_{n \in \mathbb{N}}$. Let $p = \lim_{n \to \mathcal{U}} t_{K_2}(G_n) = t_{K_2}([G_n]_{\mathcal{U}})$. Since G_n is $1/n$-quasirandom, $t_{C_4}([G_n]_{\mathcal{U}}) = \lim_{n \to \mathcal{U}} t_{C_4}(G_n) = p^4$, so $[G_n]_{\mathcal{U}}$ is quasirandom.

By Theorem 5.4, $\lim_{n \to \mathcal{U}} t_H(G_n) = t_H([G_n]_{\mathcal{U}}) = p^{|F|}$. Therefore, taking δ small enough,

$$\{n \mid |t_{K_2}(G_n) - p| < \delta\} \cap \{n \mid |t_H(G_n) - p^{|F|}| < \epsilon/2\} \in \mathcal{U}.$$

But consider some n in both of these sets: we have

$$|t_H(G_n) - (t_{K_2}(G_n))^{|F|}| \leq |p^{|F|} - (t_{K_2}(G_n))^{|F|}| + |t_H(G_n) - p^{|F|}| \leq \epsilon/2 + \epsilon/2 < \epsilon,$$

which is a contradiction. □

The basic structure here is almost ubiquitous in our proofs. When we wish to prove a statement about sufficiently large finite graphs, we begin by assuming the statement is false. This will lead to a sequence of counterexamples; we work with the ultraproduct of this sequence and get a property which contradicts our claim about the sequence.

We could have carried out this proof without ever mentioning ultraproducts. Indeed, we could take the arguments in the proof of Theorem 5.4 and translate them, step by step, into arguments in the finite setting. The difficulty is that the arguments would become peppered with additional ϵ's; for instance, each time we used quasirandomness to argue that $\mu_2([X \times Y] \cap E) = p\mu(X)\mu(Y)$, we would have to instead argue that ϵ-quasirandomness ensures $|\mu_2([X \times Y] \cap E) - p\mu(X)\mu(Y)| < \delta$ for some δ. The main value of the ultraproduct in this argument is hiding all those ϵ's and δ's by letting them reach the limit value of 0.

5.3 The L^2 Norm

We are going to develop tools for working with quasirandom graphs: we will develop a "norm" which will let us measure how random a graph, and more generally, a function is. This will give us an explanation for why $t_{C_4}(G)$ is

significant in ways that, say, $t_{C_3}(G)$ is not: t_{C_3} does not have an associated norm.

As a preview, we recall the properties of the L^2 norm, which will be a model for the seminorm we are interested in.

Definition 5.5. When $f : V^k \to \mathbb{R}$ is measurable, we define $||f||_{L^2(\mu_k)} = \sqrt{\int (f(x_1, \ldots, x_k))^2 \, d\mu_k}$.

We define $L^2(\mu_k)$ to be the set of measurable functions such that $||f||_{L^2(\mu_k)}$ is finite (i.e. the integral exists).

When considering the space $L^2(\mu_k)$, it is standard to identify functions "up to almost everywhere equality"—that is, when $||f - g||_{L^2(\mu_k)}$, we treat f and g as equivalent, and even write $f = g$. Note that this is weaker than pointwise equality—two functions could agree in this sense, but actually have $f(x) \neq g(x)$ for some values of x. However this is the same as requiring that $\{x \mid f(x) \neq g(x)\}$ have measure 0. So when we write $f = g$, we mean equality except possibly on a set of measure 0. This is consistent with our view that sets of measure 0 are negligible.

It is a standard fact that the L^2 norm *is* a norm; that is:

- for any f, $||f||_{L^2(\mu_k)} \geq 0$,
- if $||f||_{L^2(\mu_k)} = 0$ then $f = 0$ (the function constantly equal to 0, or at least equal to 0 except on a set of measure 0),
- for any real number c, $||c \cdot f||_{L^2(\mu_k)} = |c| \cdot ||f||_{L^2(\mu_k)}$, and
- $||f + g||_{L^2(\mu_k)} \leq ||f||_{L^2(\mu_k)} + ||g||_{L^2(\mu_k)}$.

The last property is the *triangle inequality*.

One might ask why we use the power 2—why not work with $(\int (f(x_1, \ldots, x_k))^p \, d\mu_k)^{1/p}$ for some other value of p? One good reason is that the L^2 norm associated to an *inner product*.

Definition 5.6. When $f, g \in L^2(\mu_k)$, $\langle f, g \rangle_{L^2(\mu_k)} = \int f(x_1, \ldots, x_k) \cdot g(x_1, \ldots, x_k) \, d\mu_k$.

With this definition, $||f||_{L^2(\mu_k)} = \sqrt{\langle f, f \rangle_{L^2(\mu_k)}}$.

This inner product is bilinear:

$$\langle f, ag + bh \rangle = a\langle f, g \rangle + b\langle f, h \rangle$$

and

$$\langle af + bh, g \rangle = a\langle f, g \rangle + b\langle h, g \rangle.$$

The Cauchy-Schwarz inequality can be stated in the form

$$\langle f, g \rangle_{L^2(\mu_k)} \leq ||f||_{L^2(\mu_k)} \cdot ||g||_{L^2(\mu_k)}.$$

5.4 The U^2 Seminorm

We will define a norm-like operation based on t_{C_4}:

Definition 5.7.

$$||f||_{U^2} = \sqrt[4]{\int f(x,y)f(x',y)f(x,y')f(x',y')d\mu_4}.$$

Of course, when $f = \chi_E$ is the characteristic function of a graph, $||f||_{U^2} = \sqrt[4]{t_{C_4}(E)}$. We will establish the basic properties of this operation and then use those to see how it relates to quasirandomness.

For an arbitrary measurable function f, this integral might not exist. It would be enough to restrict to functions f whose fourth power is integrable—the $L^4(\mu_2)$ functions, which are the functions for which $\int |f|^4 \, d\mu_2$ exists—but it will be sufficient for us, and simpler, to impose a stricter restriction and consider only the almost everywhere bounded functions.

Definition 5.8. $L^\infty(\mu_k) \subseteq L^2(\mu_k)$ is the space of functions f such that there is a real number $c > 0$ such that $\mu_k(\{(x_1, \ldots, x_k) \mid |f(x_1, \ldots, x_k)| > c\}) = 0$.

That is, $L^\infty(\mu_k)$ is the functions which are restricted to $[-c, c]$ for some c, except on a set of measure 0.

$||\cdot||_{U^2}$ is known as the *Gowers U^2 seminorm*. It is the first of the *Gowers uniformity seminorms*. (The others will appear later, when we generalize to hypergraphs.)

As the name suggests, $||\cdot||_{U^2}$ is a seminorm, not a norm: it is possible to have $||f||_{U^2} = 0$ even when f is non-zero function, but the other three properties of a norm (non-negativity, multiplication by scalars, and the triangle inequality) are satisfied, as we will prove shortly.

Definition 5.9. An operation $||\cdot||$ is a *seminorm* if it satisfies:

- for every f, $||f|| \geq 0$,
- for any real number c and any f, $||c \cdot f|| = |c| \cdot ||f||$,
- for any f and g, $||f + g|| \leq ||f|| + ||g||$.

We could define operations like $||\cdot||_{U^2}$ corresponding to other graphs, like $||f||_{C_3} = \sqrt[3]{\int f(x,y)f(y,z)f(z,x)\, d\mu_3}$. However for most graphs, this fails to be a seminorm. Even the form suggests immediately that the result will not always be non-negative: there is no squaring going on, and indeed, it is not hard to find f so that $||f||_{C_3}$ is negative.

Example 5.10. Let $A \subseteq V$ be a set with $\mu(A) = 1/2$ and define $f(x, y)$ by:

$$f(x,y) = \begin{cases} -1 & \text{if } x \in A \text{ and } y \in A \\ -1 & \text{if } x \notin A \text{ and } y \notin A \\ 0 & \text{if } x \in A \text{ and } y \notin A \\ 0 & \text{if } x \notin A \text{ and } y \in A \end{cases}$$

Then, for any x, y, z, $f(x,y)f(y,z)f(z,y)$ is 0 unless all three belong to A or all three do not belong to A, in which case $f(x,y)f(y,z)f(z,y)$ is -1. In particular, $||f||_{\mathrm{C}_3} = -1/4 < 0$.

However C_4 is not the only graph which gives us a seminorm, and it will be useful to consider a second example in parallel.

Definition 5.11. When $f \in L^\infty(\mu_2)$,

$$||f||_{\mathrm{V}} = \sqrt{\int f(x,y)f(x,z)\,d\mu_3}.$$

We call this the V seminorm because it corresponds to the graph  . The V seminorm is not as common (or important) as the U^2 seminorm, but belongs to a larger family of seminorms which extends the Gowers uniformity seminorms.

Before proving that $||\cdot||_{U^2}$ and $||\cdot||_{\mathrm{V}}$ are seminorms, it is helpful to introduce the inner product-like notions corresponding to them. Unlike an ordinary inner product, the U^2 version of an inner product has *four* arguments.

Definition 5.12.

$$\langle f_1, f_2\rangle_{\mathrm{V}} = \int f_1(x,y)f_2(x,z)\,d\mu_3$$

and

$$\langle f_1, f_2, f_3, f_4\rangle_{U_2} = \int f_1(x,y)f_2(x,y')f_3(x',y)f_4(x',y')\,d\mu_4.$$

Analogous to the equality $||f||^2_{L^2(\mu_k)} = \langle f, f\rangle_{L^2(\mu_k)}$, we have $||f||^2_{\mathrm{V}} = \langle f, f\rangle_{\mathrm{V}}$ and $||f||^4_{U^2} = \langle f, f, f, f\rangle_{U^2}$.

Linearity in each coordinate follows by the linearity of integrals; for instance,

$$\langle f, ag + bg\rangle_{\mathrm{V}} = a\langle f, g\rangle_{\mathrm{V}} + b\langle f, h\rangle_{\mathrm{V}}$$

and

$$\langle f_1, f_2, f_3, ag + bh\rangle_{U^2} = a\langle f_1, f_2, f_3, g\rangle_{U^2} + b\langle f_1, f_2, f_3, h\rangle_{U^2}.$$

Most importantly, the V and U^2 norms satisfies a version of Cauchy-Schwarz we will use over and over again.

Theorem 5.13 (Gowers–Cauchy–Schwarz)**.**

- $|\langle f_1, f_2\rangle_{\mathrm{V}}| \leq ||f_1||_{\mathrm{V}} \cdot ||f_2||_{\mathrm{V}}$*, and*
- $|\langle f_1, f_2, f_3, f_4\rangle_{U^2}| \leq ||f_1||_{U^2} \cdot ||f_2||_{U^2} \cdot ||f_3||_{U^2} \cdot ||f_4||_{U^2}$.

Proof. Both parts follow by the use of Cauchy-Schwarz.

For the first part,

$$\begin{aligned}
|\langle f_1, f_2\rangle_{\mathrm{V}}|^2 &= \left|\int f_1(x,y) f_2(x,z)\, d\mu_3\right|^2 \\
&= \left|\int \left(\int f_1(x,y) d\mu(y)\right)\left(\int f_2(x,z) d\mu(z)\right) d\mu(x)\right|^2 \\
&\leq \int \left(\int f_1(x,y) d\mu(y)\right)^2 d\mu \int \left(\int f_2(x,y) d\mu(y)\right)^2 d\mu \\
&= ||f_1||_{\mathrm{V}}^2 \cdot ||f_2||_{\mathrm{V}}^2,
\end{aligned}$$

so taking square roots of both sides,

$$|\langle f_1, f_2\rangle_{\mathrm{V}}| \leq ||f_1||_{\mathrm{V}} \cdot ||f_2||_{\mathrm{V}}.$$

The second part is similar, but we use Cauchy-Schwarz twice.

$$\begin{aligned}
&|\langle f_1, f_2, f_3, f_4\rangle|_{U^2}^4 \\
&= \left|\int f_1(x,y) f_2(x,y') f_3(x',y) f_4(x',y')\, d\mu_4\right|^4 \\
&= \left|\int \left(\int f_1(x,y) f_3(x',y)\, d\mu(y)\right)\left(\int f_2(x,y') f_4(x',y') d\mu(y')\right) d\mu_2(x,x')\right|^4 \\
&\leq \left|\int \left(\int f_1(x,y) f_3(x',y)\, d\mu(y)\right)^2 d\mu_2(x,x')\right|^2 \cdot \left|\int \left(\int f_2(x,y') f_4(x',y') d\mu(y')\right)^2 d\mu_2(x,x')\right|^2 \\
&\leq \left|\int f_1(x,y) f_1(x,y') f_3(x',y) f_3(x',y')\, d\mu_4\right|^2 \cdot \left|\int f_2(x,y) f_2(x,y') f_4(x',y) f_4(x',y')\, d\mu_4\right|^2.
\end{aligned}$$

Consider one of these pieces:

$$\begin{aligned}
&\left|\int f_1(x,y)f_1(x,y')f_3(x',y)f_3(x',y')\,d\mu_4\right|^2 \\
=&\left|\int\left(\int f_1(x,y)f_1(x,y')d\mu(x)\right)\cdot\left(\int f_3(x',y)f_3(x',y)d\mu(x')\right)d\mu_2(y,y')\right|^2 \\
\leq&\int\left(\int f_1(x,y)f_1(x,y')d\mu(x)\right)^2 d\mu_2(y,y')\cdot\int\left(\int f_3(x',y)f_3(x',y)d\mu(x')\right)^2 d\mu_2(y,y') \\
=&\int f_1(x,y)f_1(x,y')f_1(x',y)f_1(x',y')d\mu_4\cdot\int f_3(x,y)f_3(x,y')f_3(x',y)f_3(x',y')d\mu_4 \\
=&||f_1||_{U^2}^4\cdot||f_3||_{U^2}^4.
\end{aligned}$$

Similarly,

$$\left|\int f_2(x,y)f_2(x,y')f_4(x',y)f_4(x',y')\,d\mu_4\right|^2 \leq ||f_2||_{U^2}^4\cdot||f_4||_{U^2}^4.$$

Putting these together,

$$\langle f_1,f_2,f_3,f_4\rangle^4 \leq ||f_1||_{U^2}^4\cdot||f_3||_{U^2}^4\cdot||f_2||_{U^2}^4||f_4||_{U^2}^4,$$

and therefore

$$|\langle f_1,f_2,f_3,f_4\rangle| \leq ||f_1||_{U^2}\cdot||f_2||_{U^2}\cdot||f_3||_{U^2}\cdot||f_4||_{U^2}.$$

□

This argument gives some idea what makes C_4 and the V graphs special: they correspond to "doubling vertices". If we begin with a single edge $\{x,y\}$ and "double" the y vertex, we get the V graph—the single vertex x adjacent to two different vertices. If we take the V graph and now double the x vertex, we get the graph C_4: two copies of y and two copies of x, with both y vertices adjacent to both x vertices.

This doubling is exactly what lets us apply the Cauchy-Schwarz inequality, and helps explain why C_4 and V give us seminorms while most other graphs don't.

Theorem 5.14. *$||\cdot||_{\mathrm{V}}$ and $||\cdot||_{U^2}$ are seminorms.*

Proof. For any L^∞ function f, $||f||_{\mathrm{V}} = \sqrt{\int\left(\int f(x,y)\,d\mu(y)\right)^2 d\mu(x)}$; since the quantity $\left(\int f(x,y)\,d\mu(y)\right)^2$ is non-negative, $||f||_{\mathrm{V}}$ is always defined and non-negative. Similarly, $||f||_{U^2} = \sqrt[4]{\int\left(\int f(x,y)f(x',y)\,d\mu(y)\right)^2 d\mu_2(x,x')}$;

since the quantity $\left(\int f(x,y)f(x',y)\,d\mu(y)\right)^2$ is non-negative, we ensure that $||f||_{U^2}$ is always defined and non-negative.

For any real number c,

$$||c\cdot f||_{\mathrm{V}} = \sqrt{\int c^2 f(x,y)f(x,z)\,d\mu_3} = |c|\cdot||f||_V$$

and

$$||c\cdot f||_{U^2} = \sqrt[4]{\int c^4 f(x,y)f(x',y)f(x,y')f(x',y')d\mu_4} = |c|\cdot||f||_{U^2}.$$

To see the triangle inequality for $||\cdot||_V$, observe that

$$\begin{aligned} ||f+g||_{\mathrm{V}}^2 &= \langle f+g, f+g\rangle_{\mathrm{V}} \\ &= \langle f,f\rangle_{\mathrm{V}} + 2\langle f,g\rangle_{\mathrm{V}} + \langle g,g\rangle_{\mathrm{V}} \\ &\le ||f||_{\mathrm{V}}^2 + 2||f||_{\mathrm{V}}\cdot||g||_{\mathrm{V}} + ||g||_{\mathrm{V}}^2 \\ &= (||f||_{\mathrm{V}} + ||g||_{\mathrm{V}})^2, \end{aligned}$$

so $||f+g||_{\mathrm{V}} \le ||f||_{\mathrm{V}} + ||g||_{\mathrm{V}}$.

Similarly, to see the triangle inequality for $||\cdot||_{U^2}$, we work with the fourth power;

$$\begin{aligned} ||f+g||_{U^2}^4 &= \langle f+g, f+g, f+g, f+g\rangle_{U^2} \\ &= \langle f,f,f,f\rangle_{U^2} + \langle f,f,f,g\rangle_{U^2} + \cdot + \langle g,g,g,g\rangle_{U^2} \\ &\le ||f||_{U^2}^4 + ||f||_{U^2}^4||g||_{U^2} + \cdots + ||g||_{U^2}^4 \\ &= (||f||_{U^2} + ||g||_{U^2})^4. \end{aligned}$$

Therefore

$$||f+g||_{U^2} \le ||f||_{U^2} + ||g||_{U^2}.$$

□

It will be convenient to place the V and U^2 norms in an order with the L^2 norm and the absolute value.

Lemma 5.15. *For any f, $|\int f\,d\mu_2| \le ||f||_{\mathrm{V}} \le ||f||_{U^2} \le ||f||_{L^2(\mu_2)}$.*

Proof. Write 1 for the function constantly equal to 1. Then

$$\begin{aligned} \left|\int f(x,y)d\mu_2\right| &= \left|\int f(x,y)1(x,z)\,d\mu_3\right| \\ &= |\langle f,1\rangle_{\mathrm{V}}| \\ &\le ||f||_V\cdot||1||_{\mathrm{V}} \\ &= ||f||_{\mathrm{V}} \end{aligned}$$

since $||1||_{\mathrm{V}} = \int 1 \, d\mu_3 = 1$.

Similarly,

$$\begin{aligned} ||f||_{\mathrm{V}}^2 &= \int f(x,y)f(x,z) \, d\mu_3 \\ &= \langle f, f, 1, 1 \rangle_{U^2} \\ &\leq ||f||_{U^2}^2 \cdot ||1||_{U^2}^2 \\ &= ||f||_{U^2}^2 \end{aligned}$$

since $||1||_{U^2} = \int 1 \, d\mu_4 = 1$. Taking the square root of both sides gives $||f||_{\mathrm{V}} \leq ||f||_{U^2}$.

Finally,

$$\begin{aligned} ||f||_{U^2}^4 &= \int f(x,y)f(x,y')f(x',y)f(x',y') \, d\mu_4 \\ &\leq ||f||_{L^2}^4 \end{aligned}$$

by Cauchy-Schwarz. □

The inequality $|\int f \, d\mu_2| \leq ||f||_{U^2}$ is an extension of our old observation that $t_{\mathrm{K}_2}(G)^4 \leq t_{\mathrm{C}_4}(G)$. This suggests that we might think of an arbitrary function f as quasirandom exactly when $||f||_{U^2} = |\int f \, d\mu_2|$.

The V and U^2 norms satisfy an additional inequality which connects them directly to the structure of the graded probability space, which characterizes what makes them useful:

Theorem 5.16. *For any measurable set B and any $f \in L^\infty(\mu_2)$,*

$$||f(x,y)\chi_B(x)||_{\mathrm{V}} \leq ||f||_{\mathrm{V}}$$

and

$$||f(x,y)\chi_B(x)||_{U^2} \leq ||f||_{U^2} \text{ and } ||f(x,y)\chi_B(y)||_{U^2} \leq ||f||_{U^2}.$$

Applying the U^2 version of this lemma twice, once in the x coordinate and once in the y coordinate, shows that

$$||f(x,y)\chi_B(x)\chi_C(y)||_{U^2} \leq ||f||_{U^2}.$$

That is, when we restrict a function to a rectangle, the U^2 norm does not increase.

Proof. Let B be measurable and let $f \in L^\infty(\mu_2)$ be given. Letting $\overline{B} = V \setminus B$, we have

$$f(x,y) = f(x,y)\chi_B(x) + f(x,y)\chi_{\overline{B}}(x).$$

For the V norm, observe that

$$\begin{aligned}
||f||_{\mathrm{V}}^2 &= \langle f, f\rangle_{\mathrm{V}} \\
&= \langle f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}\rangle_{\mathrm{V}} \\
&= \langle f\chi_B, f\chi_B\rangle_{\mathrm{V}} + 2\langle f\chi_B, f\chi_{\overline{B}}\rangle_{\mathrm{V}} + \langle f\chi_{\overline{B}}, f\chi_{\overline{B}}\rangle_{\mathrm{V}} \\
&= ||f\chi_B||_{\mathrm{V}}^2 + 2\int f(x,y)\chi_B(x)f(x,z)\chi_{\overline{B}}(x)\,d\mu_3 + ||f\chi_{\overline{B}}||_{\mathrm{V}}^2 \\
&= ||f\chi_B||_{\mathrm{V}}^2 + ||f\chi_{\overline{B}}||_{\mathrm{V}}^2 \\
&\geq ||f\chi_B||_{\mathrm{V}}^2.
\end{aligned}$$

For the U^2 norm, the argument is similar except that there are more terms to worry about. We prove the first inequality for the U^2 norm since the last inequality follows symmetrically by the same argument.

$$\begin{aligned}
||f||_{U^2}^4 &= \langle f, f, f, f\rangle_{U^2} \\
&= \langle f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}\rangle_{U^2} \\
&= \langle f\chi_B, f\chi_B, f\chi_B, f\chi_B\rangle_{U^2} + \langle f\chi_{\overline{B}}, f\chi_B, f\chi_B, f\chi_B\rangle_{U^2} + \cdots + \langle f\chi_{\overline{B}}, f\chi_{\overline{B}}, f\chi_{\overline{B}}, f\chi_{\overline{B}}\rangle_{U^2}.
\end{aligned}$$

There are a total of 16 terms we need to consider in this sum. We will show that each term is non-negative.

Each term has the form

$$\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')f(x,y')\chi_{S_3}(x)f(x',y')\chi_{S_4}(x')\,d\mu_4$$

where S_1, S_2, S_3, S_4 are each either B or $\overline{B}$.

If $S_1 \neq S_3$ or $S_2 \neq S_4$ then this integral is 0, like the middle term in the V norm argument above. For instance,

$$\int f(x,y)\chi_B(x)f(x',y)\chi_B f(x,y')\chi_{\overline{B}}(x)f(x',y')\chi_B(x')\,d\mu_4 = 0$$

because for every x, one of $\chi_B(x)$ and $\chi_{\overline{B}}(x)$ must be 0, so the product is always 0.

However there are four terms,

$$\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')f(x,y')\chi_{S_1}(x)f(x',y')\chi_{S_2}(x')\,d\mu_4$$

which do not cancel out so simply. But

$$\begin{aligned}&\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')f(x,y')\chi_{S_1}(x)f(x',y')\chi_{S_2}(x')\,d\mu_4\\ =&\int\left(\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')\,d\mu(y)\right)^2 d\mu_2(x,x'),\end{aligned}$$

which is always non-negative because the integrand is squared.

Since all the terms are non-negative,

$$||f||_{U^2}^4 \geq \langle f\chi_B, f\chi_B, f\chi_B, f\chi_B\rangle_{U^2} = ||f\chi_B||_{U^2}^4.$$

□

This implies that when $||f||_{U^2} = 0$, f is orthogonal to any "rectangle" of the form $X \times Y$.

Lemma 5.17. *If $||f||_{U^2} = 0$ then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $\mathcal{B}(\mu)$,*

$$\int f(x,y)\chi_X(x)\chi_Y(y)\,d\mu_2 = 0.$$

Proof. We have

$$\left|\int f(x,y)\chi_X(x)\chi_Y(y)\,d\mu_2\right| \leq ||f(x,y)\chi_X(x)\chi_Y(y)||_{U^2} \leq ||f||_{U^2} = 0.$$

□

5.5 Seminorms, Even Distribution, and Quasirandomness

We would like to identify the relationship between the U^2 norm and quasirandomness. When $E \subseteq V^2$ is a symmetric measurable set, the definitions say that E is quasirandom exactly when $||\chi_E||_{U^2} = \mu_2(E)$.

However this turns out not to be the most important perspective on the relationship. We will ultimately want to be able to take any $L^\infty(\mu_2)$ function f and decompose it into a *structured part* and a *random part*—to write $f = f^+ + f^-$ where $||f^-||_{U^2} = 0$ and f^+ has some sort of nice description. This is similar to the spectral perspective, where we write f as a sum of eigenfunctions; indeed, one way to give a description of f^+ will be as the sum of the eigenfunctions.

When $f = \chi_E$ is quasirandom—and, more generally, when $||f||_{U^2} = |\int f \, d\mu_2|$—this decomposition will be particularly useful because f^+ will turn out to be a constant function.

Before proving this, it will be useful to develop the analogous theory for the V norm, both to preview the techniques and as a step in the proof.

The V norm corresponds to a weaker notion of randomness than quasirandomness.

Definition 5.18. f is *evenly distributed* if, for almost every x, $\int f(x,y)\, d\mu(y) = \int f(x,y)\, d\mu_2$.

Like the V norm itself, the definition of even distribution is asymmetric in the two variables, so we will most often apply it when f itself is symmetric.

In particular, when $f = \chi_E$ is the characteristic function of a graph, f is evenly distributed when (outside of a set of measure 0), every point x has the "right number" of neighbors.

For instance, the bipartite graph with two equal parts—the graph where there is an $A \subseteq V$ with $\mu(A) = 1/2$ and $E = A \times (V \setminus A) \cup (V \setminus A) \times A$—is evenly distributed, since $\mu_2(E) = 1/2$ and, for every x, $\mu(E_x) = 1/2$. (But this graph is very far from being quasirandom—for instance, it has no triangles.)

Theorem 5.19. *Let $f \in L^\infty(\mu_2)$ and take $p = |\int f \, d\mu_2|$. The following are equivalent:*

- *f is evenly distributed,*
- $||f||_{\mathrm{V}} = p$,
- $||f - p||_{\mathrm{V}} = 0$.

Proof. If f is evenly distributed then

$$\begin{aligned} ||f||_{\mathrm{V}} &= \sqrt{\int f(x,y) f(x,z)\, d\mu_3} \\ &= \sqrt{\int (\int f(x,y)\, d\mu(y))^2 d\mu} \\ &= \sqrt{\int p^2\, d\mu} \\ &= p. \end{aligned}$$

If $||f||_{\mathrm{V}} = p$ then

$$\begin{aligned} ||f-p||_{\mathrm{V}}^2 &= \langle f-p, f-p\rangle_{\mathrm{V}} \\ &= \langle f,f\rangle_{\mathrm{V}} - 2\langle f,p\rangle_{\mathrm{V}} + \langle p,p\rangle_{\mathrm{V}} \\ &= p^2 - 2\int f(x,y)p\,d\mu_3 + p^2 \\ &= p^2 - 2p\int f(x,y)\,d\mu_2 + p^2 \\ &= 0. \end{aligned}$$

If $||f-p||_V = 0$ then

$$\begin{aligned} 0 &= ||f-p||_V^2 \\ &= \int \left(\int f(x,y)\,d\mu(y) - p\right)^2 d\mu(x). \end{aligned}$$

Therefore $\{x \mid \int f(x,y)\,d\mu(y) = p\}$ must have measure 1, so f is evenly distributed. □

Knowing that χ_E is evenly distributed—that is, that $t_{\mathrm{V}}(E) = t_{\mathrm{K}_2}(E)^2$—is not enough to guarantee that E contains the right number of copies of all graphs, as the example of the bipartite graph above shows. However equidstributed graphs satisfy an analog of Theorem 5.4 for a limited family of graphs:

Theorem 5.20. *If $G = (V,E)$ is evenly distributed with $t_{\mathrm{K}_2}(E) = p$ then whenever $H = (W,F)$ is a* tree*—that is, H contains no cycles—$t_H(E) = p^{|F|}$.*

Problem 5.21. Adapt the proof of Theorem 5.4 to prove this. Adapt the proof of Theorem 1.31 to prove the finite analog: for every tree H and every $\epsilon > 0$, there is a $\delta > 0$ so that if $|t_{\mathrm{V}}(G) - t_{\mathrm{K}_2}(G)^2| < \delta$ and G is sufficiently large, then $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| < \epsilon$.

Theorem 5.22. *Let $f \in L^\infty(\mu_2)$ be symmetric and take $p = |\int f\,d\mu_2|$. The following are equivalent:*

- $||f||_{U^2} = p$,
- $||f-p||_{U^2} = 0$.

Note that when $f = \chi_E$, $||f||_{U^2} = |\int f\,d\mu_2|$ is saying that E is quasirandom.

Proof. First, suppose $||f||_{U^2} = p$. Since $p \leq ||f||_V \leq ||f||_{U^2}$, we also have $||f||_V = p$, so f is evenly distributed. We can calculate

$$\begin{aligned} ||f-p||_{U^2}^4 &= \langle f-p, f-p, f-p, f-p\rangle_{U^2} \\ &= \langle f,f,f,f\rangle_{U^2} - \langle f,f,f,p\rangle_{U^2} + \cdots + \langle p,p,p,p\rangle_{U^2}. \end{aligned}$$

There are sixteen total terms in this sum which we need to consider; we will show that all of them are equal to p^4. The first and last terms are certainly equal to p^4.

Consider the four terms with f three times and p once. For instance

$$\begin{aligned} \langle f,f,f,p\rangle_{U^2} &= \int f(x,y)f(x,y')f(x',y)p\,d\mu_4 \\ &= p\int f(x,y)f(x',y)\int f(x,y')\,d\mu(y')\,d\mu_3(x,x',y) \\ &= p^2\int f(x,y)f(x',y)\,d\mu_3 \\ &= p^2||f||_V^2 \\ &= p^4. \end{aligned}$$

The other three of these terms are equal to p^4 by symmetric arguments.

Consider the six terms with f twice and p twice. These are not all symmetric: four of them are like

$$\langle f,p,f,p\rangle_V = \int f(x,y)pf(x',y)p\,d\mu_4$$

where the two copies of f share a variable, while two are like

$$\langle f,p,p,f\rangle_V = \int f(x,y)p^2f(x',y')\,d\mu_4.$$

For the first group,

$$\int f(x,y)pf(x',y)p\,d\mu_4 = p^2t_V(f) = p^4.$$

For the second group,

$$\int f(x,y)p^2f(x',y')\,d\mu_4 = p^2(\int f(x,y)\,d\mu_2)^2 = p^4.$$

There are four terms where f only appears once, all of which are symmetric with

$$\langle f,p,p,p\rangle_V = \int f(x,y)p^3\,d\mu_4 = p^4.$$

Putting these sixteen terms together,

$$||f-p||_{U^2}^4 = p^4 - 4p^4 + 6p^4 - 4p^4 + p^4 = 0.$$

Conversely, suppose $||f-p||_{U^2} = 0$. Since

$$||f||_{U^2} = ||f-p+p||_{U^2} \leq ||f-p||_{U^2} + ||p||_{U^2} = 0 + p,$$

we have $||f||_{U^2} \leq p$. Since we always have $p \leq ||f||_{U^2}$, it follows that $||f||_{U^2} = p$. □

This gives us an almost trivial proof of a version of Theorem 1.28. (Carefully comparing the finite proof we gave with the work above will show that this is really the same argument, with all the work packaged into various lemmata.)

Corollary 5.23. *If E is quasirandom then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $\mathcal{B}(\mu)$,*

$$\mu_2(E \cap (X \times Y)) = \mu_2(E)\mu(X)\mu(Y).$$

Proof. Let $p = \mu_2(E)$. Since E is quasirandom, $||\chi_E - p||_{U^2} = 0$, so by Lemma 5.17 we have

$$0 = \int (\chi_E(x,y) - p)\chi_X(x)\chi_Y(y)\,d\mu_2 = \int \chi_E(x,y)\chi_X(x)\chi_Y(y)\,d\mu_2 - p\int \chi_X(x)\chi_Y(y)\,d\mu_2,$$

which is the same as saying that

$$\mu(E \cap (X \times Y)) = p\mu(X)\mu(Y).$$

□

5.6 Transfer

We would like a systematic way of drawing conclusions about finite graphs from theorems about measurable graphs. We have pointed out that "internal statements" have the property that they are true in an ultraproduct exactly when they are true of most of the ground graphs.

However the last two chapters have been mostly concerned with statements which are not internal. Consider Turán's Theorem, Theorem 4.6. The statement

$$\text{for each } r \geq 3, \text{ if } t_{\mathrm{K}_2}(G) > 1 - \tfrac{1}{r-1} \text{ then } |T_{\mathrm{K}_r}(G)| > 0$$

is not internal. First, the "for each r" is considering only natural numbers r—it is not the sort of "for all" contemplated by Lemma 2.34. Second, we know that expressions about measure like $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{r-1}$ are generally not internal.

However (recalling our internal comparison $\gtrsim$ from Definition 3.3) this is equivalent to the statement:

> for every $r \geq 3$ and every $e > 0$, if $t_{\mathrm{K}_2}(G) \gtrsim 1 - \frac{1}{r-1} + \frac{1}{e}$ then $|T_{\mathrm{K}_r}(G)| > 0$.

The last part—"$t_{\mathrm{K}_2}(G) \gtrsim 1 - \frac{1}{r-1} + \frac{1}{e}$ then $|T_{\mathrm{K}_r}(G)| > 0$"—is the sort of statement that passes between ultraproducts and ground models, so this has the form "for all natural numbers r, e, $\sigma_{r,e}$" where, for each r and e, $\sigma_{r,e}$ is an internal statement. This is the perspective taken in the proof of Corollary 4.7: we first fix values of the natural numbers r and e and then use the internality of $\sigma_{r,e}$.

This leads to a general principle, which we could formulate as follows. (We will restate it more formally and more generally later in this section.)

> Suppose that, for every natural number k, σ_k is an internal statement. If, for every k, σ_k is true in $[G_n]_{\mathcal{U}}$ then, for each k, $\{n \mid \sigma_k \text{ is true in } G_n\} \in \mathcal{U}\}$.

Note that it might not be the case that $\{n \mid \text{every } \sigma_k \text{ is true in } G_n\} \in \mathcal{U}$. For instance, consider the case where σ_k is the statement "there are at least k elements" and each G_n has n elements: then σ_k is true in G_n if and only if $k \leq n$, but $[G_n]_{\mathcal{U}}$ is infinite, so every σ_k is true in $[G_n]_{\mathcal{U}}$. That is, we are comparing the situation where every σ_k is true, simultaneously, in the ultraproduct to the case where *each* σ_k, individually, is true in "most" of the G_n.

Next, let us consider a more complicated statement. Theorem 5.4 says that

> whenever G is a measurable graph, for every finite graph $H = (W, F)$, if $t_{\mathrm{C}_4}(G) = (t_{\mathrm{K}_2}(G))^4$ then $t_H(G) = (t_{\mathrm{K}_2}(G))^{|F|}$

while the corresponding statement about finite graphs is Theorem 1.31:

> for every finite graph $H = (W, F)$ and every $\epsilon > 0$, there is a $\delta > 0$ so that if $|t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4| < \delta$ and $|V| > 1/\delta$, $|t_H(G) - (t_{\mathrm{K}_2}(G))^{|F|}| < \epsilon$.

These are clearly related claims, but there has been some transformation, from the exact equalities in the measurable graph to approximate equalities in the finite graph.

The "for every finite graph" quantifies over the countable set of finite graphs, and we can replace exact equalities between integrals with $\succsim$ comparisons: the statement of Theorem 5.4 is equivalent to

> whenever G is a measurable graph, for every finite graph $H = (W, F)$, if, for every $\delta > 0$, $|t_{C_4}(G) - (t_{K_2}(G))^4| \precsim \delta$, then for every $\epsilon > 0$, $|t_H(G) - (t_{K_2}(G))^{|F|}| \precsim \epsilon$.

Then we can reorganize the quantifiers to get the also equivalent statement:

> whenever G is a measurable graph, for every finite graph $H = (W, F)$ and every $e > 0$, there is a $d > 0$ such that if $|t_{C_4}(G) - (t_{K_2}(G))^4| \precsim 1/e$ then $|t_H(G) - (t_{K_2}(G))^{|F|}| \precsim 1/d$.

The statement "if $|t_{C_4}(G) - (t_{K_2}(G))^4| \precsim 1/e$ then $|t_H(G) - (t_{K_2}(G))^{|F|}| \precsim 1/d$" is internal. So Theorem 5.4 is ultimately equivalent to a statement of the form "for all H, e there exists a d such that $\sigma_{H,e,d}$ holds" where $\sigma_{H,e,d}$ is internal.

This is more complicated than the structure of Turán's Theorem—there are now two "external quantifiers"—first the "for all H and e" and then "there exists a d" followed by an internal statement.*

Theorem 5.24 (Transfer)**.** *Suppose that for every pair of natural numbers y and z, $\sigma_{y,z}$ is an internal statement†. Then the following are equivalent:*

- *in $[G_n]_{\mathcal{U}}$, for every y there is a z so that $\sigma_{y,z}$ is true,*
- *for every y, there is a z so that $\{n \mid \sigma_{y,z} \text{ is true in } G_n\} \in \mathcal{U}$.*

Note that the second clause has a built-in uniformity: it requires that, for each y, there is a *single* choice of a value z which works for most G_n.

Proof. Let $[G_n]_{\mathcal{U}}$ be given. First, suppose that, for every y there is a z so that $\sigma_{y,z}$ is true in $[G_n]_{\mathcal{U}}$. Then, since $\sigma_{y,z}$ is internal, $\{n \mid \sigma_{y,z} \text{ is true in } G_n\} \in \mathcal{U}$.

Conversely, suppose that, for each y, there is a z so that $\{n \mid \sigma_{y,z} \text{ is true in } G_n\} \in \mathcal{U}$. Then for this z, again since $\sigma_{y,z}$ is internal, $\sigma_{y,z}$ is true in $[G_n]_{\mathcal{U}}$. □

*As is common, we count "for all H and for all e" as a single quantifier over the pair (H, e).

†Here—and everywhere we use the term "internal statement"—one can replace it with "formula of first-order logic", which means essentially the same thing.

There is an analogous statement for internal sets: suppose we want to consider sets of the form $S = \bigcap_{y\in\mathbb{N}} \bigcup_{z\in\mathbb{N}} S^{y,z}$ where the $S^{y,z}$ are internal. The set S is not internal (unless the unions and intersections trivialize).

Theorem 5.25. *Suppose that, for every pair of natural numbers y and z and each n, $S_n^{y,z} \subseteq G_n^k$. Then, for any $[\vec{b}_n]_{\mathcal{U}} \in [G_n]_{\mathcal{U}}^k$, the following are equivalent:*

- *for every natural number y there is a natural number z so that $[\vec{b}_n]_{\mathcal{U}} \in [S_n^{y,z}]_{\mathcal{U}}$,*
- *for every natural number y there is a natural number z so that $\{n \mid \vec{b}_n \in S_n^{y,z}\} \in \mathcal{U}$.*

That is, $\bigcap_{y\in\mathbb{N}} \bigcup_{z\in\mathbb{N}} [S_n^{y,z}]_{\mathcal{U}}$ consists of those elements $[\vec{b}_n]_{\mathcal{U}}$ which belong "uniformly" to $\bigcap_{y\in\mathbb{N}} \bigcup_{z\in\mathbb{N}} S_n^{y,z}$. The proof is exactly the same as the proof of the Transfer Theorem. We have stated this for single natural number parameters y and z, but it would be the same if we had several, $y_1, \ldots, y_m$ and $z_1, \ldots, z_n$.

5.7 Remarks

The U^2 norm was well-known prior to Gowers work, at least as early as the original papers on graph quasirandomness [30]. Gowers' name and the notation U^2 have become attached to it since [75] gave the right generalization to hypergraphs. The name "V norm" and notation $||\cdot||_V$ are nonstandard; the perspective here emphasizing this norm as a building block to the U^2 norm comes from [158], where a generalization of this approach is used for hypergraphs.

The question of when a graph $G = (V, E)$ has, like V and C_4, the propety that $t_G(\cdot)^{1/|E|}$ is a norm, has been studied ([37, 82]) and is closely related to Sidorenko's conjecture [140].

The transfer theorem, as we have stated it, is more commonly stated in the setting of nonstandard analysis [10, 70, 125, 126], where it is a central property. In axiomatic approaches (as opposed to the semantic approach our ultraproducts represent), it is generally included as one of the characteristic axioms [15, 55, 122].

Chapter 6

Regularity

6.1 Rectangles

$L^2(\mu_2)$ is a vector space, and the functions which have U^2 norm 0 form a subspace, so we expect to be able identify a dual space. That will let us decompose functions, writing

$$f = f^+ + f^-$$

where $||f^-||_{U^2} = 0$ and f^+ is orthogonal to all functions with U^2 norm 0.

Definition 6.1. We say f is *dual to* U^2 if whenever $||r||_{U^2} = 0$, $\int fr\, d\mu_2 = 0$.

One of our goals this chapter is to characterize the functions which are dual to U^2. By Lemma 5.17, whenever $X, Y \subseteq V$, $\chi_X(x)\chi_Y(y)$ is dual to U^2. On the other hand, if f is dual to U^2 and non-zero then, in particular, $\int ff\, d\mu^2 = ||f||^2_{L^2(\mu_2)} > 0$, so we must have $||f||_{U^2} \neq 0$.

The following result then tells us that when f is dual to U^2, at least f correlates with a rectangle.

Theorem 6.2. *When $f \in L^\infty(\mu_2)$, $||f||_{U^2} > 0$ if and only if there is a rectangle $X \times Y$ such that $\left|\int_{X\times Y} f\, d\mu_2\right| > 0$.*

Proof. The right to left direction essentially follows from what we have already done. Suppose there is a rectangle $X \times Y$ such that $\left|\int_{X\times Y} f\, d\mu_2\right| > 0$. Then using Lemma 5.15 and Theorem 5.16 we have

$$0 < \left|\int f(x,y)\chi_X(x)\chi_Y(y)\, d\mu_2\right| \leq ||f(x,y)\chi_X(x)\chi_Y(y)||_{U^2} \leq ||f||_{U^2}.$$

Conversely, suppose $||f||_{U^2} > 0$, so

$$\begin{aligned}0 &< ||f||_{U^2}^4 \\ &= \int f(x,y)f(x',y)f(x,y')f(x',y')\,d\mu_4 \\ &= \int \left[\int f(x,y)f(x',y)f(x,y')f(x',y')\,d\mu_2(x,y)\right] d\mu_2(x',y').\end{aligned}$$

Let us assume that $\mu_2(\{(x,y) \mid |f(x,y)| > 1\}) = 0$ (the case where this holds with 1 replaced by some value d follows the same argument after scaling).

There must be some x', y' with $|f(x',y')| \le 1$ so that

$$\left|\int f(x,y)f(x',y)f(x,y')f(x',y')\,d\mu_2\right| \ge \left|\int f(x,y)f(x',y)f(x,y')\,d\mu_2\right| = \epsilon > 0.$$

(And also $\mu(\{x \mid |f(x,y')| > 1\}) = \mu(\{y \mid |f(x',y)| > 1\}) = 0$.)

For any interval $[a,b]$, let $X_{a,b} = \{x \mid f(x,y') \in [a,b]\}$ and $Y_{[a,b]} = \{y \mid f(x',y) \in [a,b]\}$. If we choose $[a,b]$ and $[c,d]$ so that $|b-a|$ and $|d-c|$ are small, the value of this integral on $X_{[a,b]} \times Y_{[c,d]}$ does not depend much on $f(x',y)$ or $f(x,y')$:

$$\begin{aligned}\left|\int_{X_{[a,b]}\times Y_{[c,d]}} f(x,y)f(x',y)f(x,y')\,d\mu_2\right| &\le \left|ac\int_{X_{[a,b]}\times Y_{[c,d]}} f(x,y)\,d\mu_2\right| \\ &\quad + \left|\int_{X_{[a,b]}} f(x,y') - a\,d\mu\right| \cdot \left|\int_{Y_{[c,d]}} f(x',y) - c\,d\mu\right| \\ &\le \left|ac\int_{X_{[a,b]}\times Y_{[c,d]}} f(x,y)\,d\mu_2\right| \\ &\quad + |b-a|\cdot|c-d|\mu(X_{[a,b]})\mu(Y_{[c,d]}).\end{aligned}$$

In particular, if we take $K \ge \sqrt{\frac{2}{\epsilon}}$,

$$\begin{aligned}\epsilon &\le \left|\int f(x,y)f(x',y)f(x,y')\,d\mu_2\right| \\ &\le \left|\sum_{i\le K}\sum_{j\le K}\int_{X_{[i\sqrt{\epsilon/2},(i+1)\sqrt{\epsilon/2}]}\times Y_{[j\sqrt{\epsilon/2},(j+1)\sqrt{\epsilon/2}]}} f(x,y)f(x,y')f(x',y)\,d\mu_2\right| \\ &\le \sum_{i\le K}\sum_{j\le K} ij\frac{\epsilon}{2}\left|\int_{X_{[i\sqrt{\epsilon/2},(i+1)\sqrt{\epsilon/2}]}\times Y_{[j\sqrt{\epsilon/2},(j+1)\sqrt{\epsilon/2}]}} f(x,y)\,d\mu_2\right| + \frac{\epsilon}{2}.\end{aligned}$$

In particular, this means there must be some i, j so that

$$\left| \int_{X_{[i\sqrt{\epsilon/2},(i+1)\sqrt{\epsilon/2}]} \times Y_{[j\sqrt{\epsilon/2},(j+1)\sqrt{\epsilon/2}]}} f(x,y)\, d\mu_2 \right| > 0.$$

□

This suggests that the rectangles will be the building blocks for the functions which are dual to U^2.

Definition 6.3. $\mathcal{B}^0_{2,1} \subseteq \mathcal{B}_2$ consists of sets of the form $\bigcup_{i\leq k} X_i \times Y_i$ where each $X_i, Y_i \in \mathcal{B}_1$.

The 2 in the notation $2,1$ reminds us that $\mathcal{B}^0_{2,1}$ consists of sets of pairs, while the 1 tells us that these are the sets generated by unary sets—that is, by rectangles.

What we really need is for $\mathcal{B}^0_{2,1}$ to be an algebra—that is, to be closed under complements, finite unions, and finite intersections. Normally this requires more than just closure under finite unions, but since the complement of a rectangle is a finite union of rectangles, our definition of $\mathcal{B}^0_{2,1}$ is sufficient.

Lemma 6.4. *$\mathcal{B}^0_{2,1}$ is an algebra—that is:*

- *whenever $B \in \mathcal{B}_{2,1}$, also $V \setminus B \in \mathcal{B}_{2,1}$,*
- *whenever $B_1, \ldots, B_k \in \mathcal{B}_{2,1}$, also $\bigcup_{i\leq k} B_i \in \mathcal{B}_{2,1}$,*
- *whenever $B_1, \ldots, B_k \in \mathcal{B}_{2,1}$, also $\bigcap_{i\leq k} B_i \in \mathcal{B}_{2,1}$.*

Proof. It is helpful to note that when $B = \bigcup_{i\leq k} X_i \times Y_i$, we can rewrite B as a union $\bigcup_{i\leq k'} X'_i \times Y'_i$ so that for any $i, j \leq k'$, either $X'_i = X'_j$ or $X'_i \cap X'_j = \emptyset$, and either $Y'_i = Y'_j$ or $Y'_i \cap Y'_j = \emptyset$. To see this, we identify all the intersections of the X_i and the Y_i—all sets of the form $X_s = \bigcap_{i\in s} X_i \cap \bigcap_{i\notin s}(V \setminus X_i)$ for some $s \subseteq [1,k]$, and all $Y_s = \bigcap_{i\in s} Y_i \cap \bigcap_{i\notin s}(V \setminus Y_i)$ for some $s \subseteq [1,k]$. (The sets X_s are precisely the atoms of the finite algebra of sets generated by the sets $\{X_i\}$, and similarly for the Y_s.)

Suppose $s \neq s'$. Then there is an $i \in s \mathbin{\triangle} s'$; without loss of generality, assume $i \in s$. Then $X_s \subseteq X_i$ but $X_{s'} \cap X_i = \emptyset$. Therefore $X_s \cap X_{s'} = \emptyset$.

So we can take B to be the union of those $X_s \times Y_s$ such that there is some i with $X_s \subseteq X_i$ and $Y_s \subseteq Y_i$.

Without loss of generality, let us write $B = \bigcup_{i\leq k} X_i \times Y_i$ with the additional condition that for any $i, j \leq k$, either $X_i = X_j$ or $X_i \cap X_j = \emptyset$,

and either $Y_i = Y_j$ or $Y_i \cap Y_j = \emptyset$. Then we can see that $V^2 \setminus B$ is also a union of rectangles: for each X_i, let $Y_i' = \bigcup_{j \leq k, X_j = X_i} Y_j$, and set

$$V^2 \setminus B = \bigcup_{i \leq k} X_i \times (V \setminus Y_i').$$

The second part holds since a union of unions is a union, and the third part follows from the first two since

$$\bigcap_{i \leq k} B_i = V^2 \setminus \left[\bigcup_{i \leq k} (V^2 \setminus B_i) \right].$$

□

We really want to work with σ-algebras, so we have to extend $\mathcal{B}^0_{2,1}$ slightly, to those sets which can be approximated using $\mathcal{B}^0_{2,1}$:

Definition 6.5. $\mathcal{B}_{2,1} \subseteq \mathcal{B}_2$ consists of those sets $B \in \mathcal{B}_2$ such that, for every $\epsilon > 0$, there is a $B_\epsilon \in \mathcal{B}^0_{2,1}$ such that $\mu_2(B \triangle B_\epsilon) < \epsilon$.

Theorem 6.6. *$\mathcal{B}_{2,1}$ is a σ-algebra.*

Proof. Clearly $\emptyset$ and $V^2 \in \mathcal{B}_{2,1}$. If $B \in \mathcal{B}_{2,1}$ then, for every $\epsilon > 0$, $\mu_2((V^2 \setminus B) \triangle (V^2 \setminus B_\epsilon)) = \mu_2(B \triangle B_\epsilon) < \epsilon$.

We can show that $\mathcal{B}_{2,1}$ is closed under unions of two elements (and therefore under finite unions and intersections): if $B, B' \in \mathcal{B}_{2,1}$ then, for any $\epsilon > 0$, we may find B_ϵ, B'_ϵ with $\mu_2(B \triangle B_\epsilon) < \epsilon/2$ and $\mu_2(B' \triangle B'_\epsilon) < \epsilon/2$. Then

$$\mu_2((B \cup B') \triangle (B_\epsilon \cup B'_\epsilon)) \leq \mu_2(B \triangle B_\epsilon) + \mu_2(B' \triangle B'_\epsilon) < \epsilon.$$

Suppose we have a sequence with $B_i \in \mathcal{B}_{2,1}$ for all $i \in \mathbb{N}$. By replacing each B_i with $B_i \setminus \bigcup_{j<i} B_j$, we may assume the B_i are pairwise disjoint. For each i, choose $B_{2^{-i}} \in \mathcal{B}^0_{2,1}$ with $\mu_2(B_i \triangle B_{2^{-i}}) < \epsilon 2^{-i-1}$. Since the B_i are pairwise disjoint, $\mu_2(\bigcup_{i \geq N} B_i) = \sum_{i \geq N} \mu_2(B_i) \leq 1$, so we may choose N large enough that $\mu_2(\bigcup_{i \geq N} B_i) < \epsilon/2$. Then

$$\mu_2(\bigcup_i B_i \triangle \bigcup_{i<N} B_{2^{-i}}) \leq \sum_{i<N} \mu_2(B_i \triangle B_{2^{-i}}) + \mu_2(\bigcup_{i \geq N} B_i) \leq \epsilon \sum_{i<N} 2^{-i-1} + \epsilon/2 < \epsilon.$$

□

6.2 Grids

We can now define a subspace of $L^\infty(\mu_2)$ which, we will eventually show, exactly captures the notion of being dual to U^2.

Definition 6.7. $L^\infty(\mathcal{B}_{2,1})$ is the subspace of $L^\infty(\mu_2)$ consisting of functions which are measurable with respect to $\mathcal{B}_{2,1}$.

Suppose $V = \bigcup_{i\leq k} B_i$ is a finite partition (so when $i \neq j$, $B_i \cap B_j = \emptyset$). Then the corresponding rectangles form a partition of V^2. We will—temporarily—define a "grid" on $\{B_i\}_{i\leq k}$ to be a step function on these rectangles—that is, a function of the form

$$\sum_{i,j\leq k} \gamma_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)$$

for some choices of constants $\gamma_{i,j}$. (This definition is temporary because we will give it its proper name in the next section.)

Lemma 6.8. *For any function f and any finite partition $V = \bigcup_{i\leq k} B_i$, the grid*

$$\sum_{i,j\leq k} \alpha_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)$$

with $\alpha_{i,j} = \frac{\int_{B_i\times B_j} f(x,y)\,d\mu_2}{\mu_2(B_i\times B_j)}$ is the grid minimizing the $L^2(\mu_2)$ distance from f.

Proof. Consider an arbitrary grid $\sum_{i,j\leq k} \gamma_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)$ and let $e_{i,j} = \gamma_{i,j} - \alpha_{i,j}$. We will show that choosing $e_{i,j} = 0$ minimizes the $L^2(\mu_2)$ distance. The main point is that $\alpha_{i,j}$ is chosen so that $\int_{B_i\times B_j}(f - \alpha_{i,j})\,d\mu_2 = 0$. Then

$$\begin{aligned}
||f - \sum_{i,j\leq k} \gamma_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)||^2_{L^2(\mu_2)} &= \int (f - \sum_{i,j\leq k} \gamma_{i,j}\chi_{B_i}(x)\chi_{B_j}(y))^2\,d\mu_2 \\
&= \sum_{i,j\leq k} \int (f - \gamma_{i,j})^2\chi_{B_i}(x)\chi_{B_j}(y)\,d\mu_2 \\
&= \sum_{i,j\leq k} \int (f - \alpha_{i,j} - e_{i,j})^2\chi_{B_i}(x)\chi_{B_j}(y)\,d\mu_2 \\
&= \sum_{i,j\leq k} \int ((f - \alpha_{i,j})^2 - 2(f - \alpha_{i,j})e_{i,j} + e_{i,j}^2)\chi_{B_i}(x)\chi_{B_j}(y)\,d\mu_2 \\
&= \sum_{i,j\leq k} \int ((f - \alpha_{i,j})^2 + e_{i,j}^2)\chi_{B_i}(x)\chi_{B_j}(y)\,d\mu_2.
\end{aligned}$$

In particular, choosing all values of $e_{i,j}$ to be equal to 0—so $\gamma_{i,j} = \alpha_{i,j}$—minimizes the $L^2(\mu_2)$ distance. □

Theorem 6.9. *If $f \in L^\infty(\mathcal{B}_{2,1})$ then, for every $\epsilon > 0$, there is a partition $V = \bigcup_{i \leq k} B_i$ such that, taking $\alpha_{i,j} = \frac{\int_{B_i \times B_j} f(x,y)\, d\mu_2}{\mu_2(B_i \times B_j)}$, we have*

$$||f - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon.$$

Proof. Let f and $\epsilon > 0$ be given. It is a standard result about measurable functions that f is approximated by simple functions:

$$||f - \sum_{i \leq d} \gamma_i \chi_{D_i}(x,y)||_{L^2(\mu_2)} < \epsilon/2$$

with $D_i \in \mathcal{B}_{2,1}$. Each D_i can be approximated by finite unions of rectangles: for each D_i, we have

$$\mu(D_i \triangle \bigcup_{j \leq d_i} X_{i,j} \times Y_{i,j}) < \frac{\epsilon}{2\gamma_i},$$

so

$$||f - \sum_{i \leq d, j \leq d_i} \gamma_i \chi_{X_{i,j}}(x) \chi_{Y_{i,j}}(y)||_{L^2(\mu_2)} < \epsilon.$$

We rearrange the $X_{i,j}, Y_{i,j}$ into a partition $\{B_i\}_{i \leq k}$ as in the proof of Lemma 6.4. Since $\sum_{i \leq d, j \leq d_i} \gamma_i \chi_{X_{i,j}}(x) \chi_{Y_{i,j}}(y)$ is a grid, we have

$$||f - \sum_{i,j} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} \leq ||f - \mathbb{E}(f \mid \{B_i \times B_j\}_{i,j \leq k})||_{L^2(\mu_2)} < \epsilon.$$

□

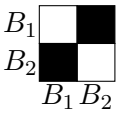
We actually want a small refinement of this result which is more complicated to state, but more useful. In particular, we want to insist that the sets B_i in the partition be internal. The property of the internal sets we need is that they are a dense algebra in $\mathcal{B}_1$—that is, the internal sets are closed under complement, finite union, and finite intersection, and every set in $\mathcal{B}_1$ can be approximated by internal sets. In practice, the case where are always interested in is when $\mathcal{B}$ is the internal sets.

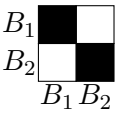
Corollary 6.10. *If $f \in L^\infty(\mathcal{B}_{2,1})$ and $\mathcal{B} \subseteq \mathcal{B}_1$ is a dense algebra then, for every $\epsilon > 0$, there is a partition $V = \bigcup_{i \leq k} B_i$ with each $B_i \in \mathcal{B}$ and $\mu_1(B_i) > 0$ such that, taking $\alpha_{i,j} = \frac{\int_{B_i \times B_j} f(x,y)\, d\mu_2}{\mu_2(B_i \times B_j)}$, we have*

$$||f - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon.$$

Proof. Follow the same proof, using the fact that each D_i can be can be approximated by rectangles from $\mathcal{B}$, and taking sets of measure 0 and combining them with a set of positive measure. □

The simplest examples of such functions are things like the characteristic function of a bipartite graph: we have $V = B_1 \cup B_2$ and $f(x,y) = \chi_{B_1}(x)\chi_{B_2}(y) + \chi_{B_2}(x)\chi_{B_1}(y)$.

When $0 \leq f \leq 1$, we can represent these functions by drawing grids. For instance, the bipartite function can be drawn as 

where we think of this grid as representing V^2, with one copy of V on each axis. The filled in boxes are where the function is equal to 1 and the empty boxes are where the function is equal to 0. On the other hand, 

is the characteristic function of the complement—two complete graphs, one on B_1 and one on B_2, with no edges between them.

In Section 1.4 we constructed the graphs $\mathbf{G}_p$ which interpolated between these two examples. The graph $\mathbf{G}_p$ itself is not represented by a grid, but there is a related grid indicating the densities of edges between various regions. To define the graph $\mathbf{G}_p$ we partitioned V into four regions and (when $p < 1/2$) gave them densities like  where black regions are ones where edges are always present, white regions are where no edges are present, edges in the light grey region are present with probability p, and edges in the dark gray regions are present with probability $1 - p$.

More generally, a grid can have different regions of varying sizes. In the grid  we have $V = B_1 \cup B_2 \cup B_3 \cup B_4$ where $\mu_1(B_2)$ is smaller than the measure of the other regions and the $\alpha_{i,j}$ are various values in $[0,1]$ (for

instance, we can see that $\alpha_{3,3}$ is a small but non-zero value, while $\alpha_{1,1}$ is a dark grey corresponding to a value close to but not equal to 1).

Whenever $f \in L^\infty(\mathcal{B}_{2,1})$ and $0 \leq f \leq 1$, Theorem 6.9 says that f can always be approximated by grids like these. In particular, *sets* in $\mathcal{B}_{2,1}$ can be approximated by grids which are mostly "black or white"—that is, we can arrange for most of the measure to be on rectangles where the constant is either close to 0 or close to 1.

Lemma 6.11. *Suppose $E \in \mathcal{B}_{2,1}$, $V = \bigcup_{i \leq k} B_i$ is a partition, and there is a grid so that*

$$||\chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon^{3/2}.$$

Then

$$\mu_2(\bigcup_{i,j \leq k, \alpha_{i,j} \in (\epsilon, 1-\epsilon)} B_i \times B_j) < \epsilon.$$

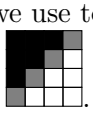
Proof. Since $\chi_E(x,y) \in \{0,1\}$ for all x, y, whenever $x \in B_i$ and $y \in B_j$ so that $\alpha_{i,j} \in (\epsilon, 1-\epsilon)$ we have

$$|\chi_E(x,y) - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)| = |\chi_E(x,y) - \alpha_{i,j}| \geq \epsilon.$$

Therefore, letting $U = \bigcup_{i,j \leq k, \alpha_{i,j} \in (\epsilon, 1-\epsilon)} B_i \times B_j$,

$$\begin{aligned} \epsilon^3 &> ||\chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||^2_{L^2(\mu_2)} \\ &= \int (\chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y))^2 \, d\mu_2 \\ &\geq \int_U \epsilon^2 \, d\mu_2 \\ &= \mu_2(U) \epsilon^2. \end{aligned}$$

So $\mu_2(U) < \epsilon$. □

The existence of *some* rectangles where the constant $\alpha_{i,j}$ is in the interval $(\epsilon, 1-\epsilon)$ is sometimes unavoidable, as the following example shows. The idea is to use the diagonal:  No matter how many rectangles we use to approximate this, the part along the diagonal will still be "fuzzy":  .

Theorem 6.12. *There is a Keisler graded probability space $\{(V, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ and an $E \in \mathcal{B}_{2,1}$ so that whenever $V = \bigcup_{i\leq k} B_i$ is a partition there are $i, j \leq k$ so that $1/4096 \leq \frac{\mu_2((B_i\times B_j)\cap E)}{\mu_2(B_i\times B_j)} < 4095/4096$.*

Proof. Naturally, we will construct G as an ultraproduct. Let $V_n = \{1, 2, \ldots, n\}\times \{0, 1\}$ consisting of two copies of the interval $\{1, 2, \ldots, n\}$. We let E_n consist of those pairs $((x, 0), (y, 1))$ with $x < y$.

Let $V = [V_n]_\mathcal{U}$ and $E = [E_n]_\mathcal{U}$. First we must show that $E \in \mathcal{B}_{2,1}$—that is, for every k, E can be approximated up to $1/k$ by unions of rectangles. Given $k > 0$, whenever n is big enough we can partition $\{1, 2, \ldots, n\}$ into intervals I^j_k of size roughly $1/k$ by taking $I^1_n = \{1, \ldots, \lfloor n/k\rfloor\}$, $I^2_n = \{\lfloor n/k\rfloor + 1, \ldots, \lfloor 2n/k\rfloor\}$, and so on. Then the sets $I^{j,b}_n = I^j_n \times \{b\} \subseteq V_n$ where $b \in \{0, 1\}$ form a partition of V_n. Take $I^{j,b} = [I^{j,b}_n]_\mathcal{U}$, so $\mu_1(I^{j,b}) = 1/2k$.

Observe that

$$\bigcup_{j<j'\leq k} I^{j,0} \times I^{j',1} \subseteq E \subseteq \bigcup_{j\leq j'\leq k} I^{j,0} \times I^{j,1}.$$

Therefore

$$\begin{aligned}
\mu_2(E \triangle \bigcup_{j<j'\leq k} I^{j,0} \times I^{j',1}) &= \mu_2(E \setminus \bigcup_{j<j'\leq k} I^{j,0} \times I^{j',1}) \\
&\leq \mu_2((\bigcup_{j<j'\leq k} I^{j,0} \times I^{j',1}) \setminus (\bigcup_{j\leq j'\leq k} I^{j,0} \times I^{j',1})) \\
&= \mu_2(\bigcup_{j\leq k} I^{j,0} \times I^{j,1}) \\
&= 1/4k^2.
\end{aligned}$$

Next we must show that any partition contains some "fuzzy" rectangles. The idea is not so complicated, though the details involve some technical computations. We need to distinguish between the two parts of V: for each n and each $b \in \{0, 1\}$, let $V_{n,b} = \{1, 2, \ldots, n\} \times \{b\} \subseteq V_n$, and set $V_b = [V_{n,b}]_\mathcal{U}$. (Technically we are writing V_0 and V_1 to mean both the first two elements of the sequence V_n and the sets $V_b \subseteq V$, but from here on we always mean the subsets of V.) Although these are separate sets, we know that they are really two copies of the same set: there is an internal measure-preserving bijection $\beta : V_0 \to V_1$ defined by taking $\beta = [\beta_n]_\mathcal{U}$ where $\beta_n(x, 0) = (x, 1)$.

The simplest sort of partition would be to break the sets V_0 and V_1 into intervals. If we could find two intervals which overlap, in the sense that $\mu_0(I_1 \cap \beta(I_0)) > 0$, then E would have intermediate density in $I_0 \times I_1$.

A more general partition might have sets B_i which contain parts of V_0 and parts of V_1; we can fix this by partitioning B_i into the sets $B_{i,0} = B_i \cap V_0$ and $B_{i,1} \cap V_1$. We will then look for i_0, i_1 so that E has intermediate density in $B_{i_0,0} \times B_{i_1,1}$. To lift this to intermediate density in $B_{i_0} \times B_{i_1}$, we will have to discard the case where the set $B_{i,b}$ is only a tiny part of B_i; fortunately, there are few points belonging to $B_{i,b}$ where B_i is almost entirely contained in V_{1-b}, so we can discard these points.

The sets $B_{i,b}$ might not be intervals. However what we really need from an interval is that it contains both many small points and many large points. So we will look at an arbitrary set $B_{i,b}$ and identify a "middle interval"—that is, we partition $B_{i,b}$ into three parts, the small elements (below some c_i), the large elements (above some d_i), and the middle elements (between c_i and d_i). If we can find sets $B_{i_0,0}$ and $B_{i_1,1}$ whose middle parts overlap—that is, so that c_{i_0} is below d_{i_1} and c_{i_1} is below d_{i_0}—then $B_{i_0,0} \times B_{i_1,1}$ contains both many pairs in E and many pairs not in E. To ensure that we can find a pair with overlapping middle intervals, we require that the middle intervals be large, so that most points belong to a middle interval.

To make all this precise, let us note that our example has a lot of extra structure we will want to use.

We can make sense of the ordering $<$ on the sets V_0 and V_1: for $x, y \in V_0$, we choose representatives $x = [(x_n, 0)]_{\mathcal{U}}$ and $y = [(y_n, 0)]_{\mathcal{U}}$ and set $x < y$ if and only if $\{n \mid x_n < y_n\} \in \mathcal{U}$. As usual, we can check that this is well-defined, in the sense that it does not depend on the choice of representatives. Furthermore, by Łoś's Theorem, $<$ is a linear order on V_0. We may similarly define a linear ordering on V_1, which we also call $<$.

We can also make sense of the ordering between the two parts—indeed, this is precisely what E does, so if $x \in V_0$ and $y \in V_1$, we can say $x < y$ exactly when $(x, y) \in E$. Note that this is monotonic in the ways we would expect; for instance, if $(x, y) \in E$ and $x' < x$ then $(x', y) \in E$ as well.

So now suppose we are given an arbitrary partition $V = \bigcup_{i \leq k} B_i$. Let us define a further partition $B_{i,b} = B_i \cap V_b$.

For each i, b where $\mu_1(B_{i,b}) > 0$, we can divide each $B_{i,b}$ into a "lowest part", a "highest part", and a "middle". More precisely, we choose elements $c_{i,b} < d_{i,b}$ so that $\mu_1(\{x \in B_{i,b} \mid x < c_{i,b}\}) = \frac{1}{8}\mu_1(B_{i,b})$ and $\mu_1(\{x \in B_{i,b} \mid x > d_{i,b}\}) = \frac{1}{8}\mu_1(B_{i,b})$. To see that these exist, observe that we can divide V_b into finitely many intervals I_b^j with $\mu_1(I_b^j) < \mu_1(B_{i,b})/k$ very small and then take the least j with $\mu_1(B_{i,b} \cap \bigcup_{i \leq j} I_b^i) > r$, and then by choosing $c_k \in I_b^j$, we have $\mu_1(\{x \in B_{i,b} \mid x < c_k\}) \in [r - \mu_1(B_{i,b})/k, r + \mu_1(B_{i,b})/k]$. Choosing such a sequence with $k \to \infty$ and using sequential comprehensiveness gives

us the desired c.

Suppose we find i_0, i_1 so that the middle intervals overlap—that is, $(c_{i_0,0}, d_{i_1,1}), (c_{i_1,1}, d_{i_0,0}) \in E$. So

$$\mu_2((B_{i_0,0}\times B_{i_1,1})\cap E) \geq \mu_2(\{x \in B_{i_0,0} \mid x < c_{i_0,0}\}\times\{y \in B_{i_1,1} \mid y > d_{i_1,1}\}) \geq \frac{1}{64}\mu_1(B_{i_0,0})\mu_1(B_{i_1,1}),$$

and also

$$\mu_2((B_{i_0,0}\times B_{i_1,1})\setminus E) \geq \mu_2(\{x \in B_{i_0,0} \mid x > d_{i_0,0}\times\{y \in B_{i_1,1} \mid y < c_{i_1,1}\}) \geq \frac{1}{64}\mu_1(B_{i_0,0})\mu_1(B_{i_1,1}).$$

So it remains to find elements i_0, i_1 so that the middle intervals overlap. Observe that the middle intervals, collectively, contain $3/4$ of the measure of each $V_{i,b}$—that is, setting $M_b = \bigcup_i \{x \in B_{i,b} \mid c_{i,b} \leq x \leq d_{i,b}\}$, we have $\mu_1(M_b) \geq 3\mu_1(V_b)/4$. Since β is measure-preserving, $M_0 \cap \beta^{-1}(M_1)$ has measure at least $3\mu(V_0)/4$, so we may choose an $x \in M_0 \cap \beta^{-1}(M_1)$, and then choose i_0, i_1 with $c_{i_0,0} < x < d_{i_0,0}$ and $c_{i_1,1} < \beta(x) < d_{i_1,1}$.

There is one remaining technical difficulty: we have arranged for E to have positive density in $B_{i_0,0} \times B_{i_1,1}$, but our original promise was to consider $B_{i_0} \times B_{i_1}$. To deal with this, we refine the sets M_b. For each b, let S_b consist of those i such that $\mu_1(B_{i,b}) < \mu_1(B_i)/8$. Observe that $\mu_1(\bigcup_{i\in S_b} B_{i,b}) \leq \mu(V_b)/8$ (any measure in $\bigcup_{i\in S_b} B_{i,b}$ must be supported by twenty times as much measure in V_{1-b}). So let $M'_b = \bigcup_{i\notin S_b}\{x \in B_{i,b} \mid c_{i,b} \leq x \leq d_{i,b}\}$, so $\mu_1(M'_b) \geq 5\mu(V_b)/8$, so $M'_0 \cap \beta^{-1}(M'_1)$ still has positive measure. Then we select i_0, i_1 as before and we have

$$\frac{\mu_2((B_{i_0} \times B_{i_1}) \cap E)}{\mu_1(B_{i_0})\mu_1(B_{i_1})} \geq \frac{\mu_2((B_{i_0,0} \times B_{i_1,1}) \cap E)}{64\mu_1(B_{i_0,0})\mu_1(B_{i_1,1})} \geq \frac{1}{4096},$$

and, similarly, $\frac{\mu_2((B_{i_0}\times B_{i_1})\setminus E)}{\mu_1(B_{i_0})\mu_1(B_{i_1})} \geq \frac{1}{4096}$. □

6.3 Conditional Expectation

The σ-algebra $\mathcal{B}_{2,1}$ is a *sub-σ-algebra* of $\mathcal{B}_2$. Whenever we have two σ-algebras like this, we can talk about the *conditional expectation* of functions from the larger σ-algebra with respect to the smaller one.

For the purposes of this section, we do not need to worry about the full abstraction of a Keisler graded probability space. All we need is a probability measure space $(V, \mathcal{B}, \mu)$ and a sub-σ-algebra $\mathcal{D} \subseteq \mathcal{B}$.

We write $L^2(\mathcal{B})$ for the set of functions f which are measurable with respect to $\mathcal{B}$ and such that $\int f^2\, d\mu$ is finite. We write $L^2(\mathcal{D})$ for the subset of L^2 consisting only of functions which are measurable with respect to $\mathcal{D}$.

In the presence of a measure, we can define the *conditional expectation*.

Theorem 6.13. *For any $f \in L^2(\mathcal{B})$:*

- *There is a function $g \in L^2(\mathcal{D})$ such that, for all $h \in L^2(\mathcal{D})$,*

$$||f - g||_{L^2} \leq ||f - h||_{L^2},$$

and

- *If $g_0, g_1 \in L^2(\mathcal{D})$ both have the property that for all $h \in L^2(\mathcal{B})$, $||f - g_i||_{L^2} \leq ||f - h||_{L^2}$, then $||g_0 - g_1||_{L^2}$.*

Proof. The proofs of both parts follow from a "quantitative" version of the second part. Let f be given and let $\alpha = \inf_{h \in L^2(\mathcal{D})} ||f - h||_{L^2}$.

What we will show is that if $g_0, g_1 \in L^2(\mathcal{D})$ are "almost as good as α" at approximating f then g_0 and g_1 must be near each other. More precisely, if $||f - g_i||_{L^2} \leq \alpha + \epsilon$ for each i then $||g_0 - g_1||_{L^2} \leq \sqrt{8\alpha\epsilon + 4\epsilon^2}$.

For observe that

$$\begin{aligned}
||f - \frac{g_0 + g_1}{2}||_{L^2}^2 &= \langle f - \frac{g_0 + g_1}{2}, f - \frac{g_0 + g_1}{2} \rangle_{L^2} \\
&= \langle f, f \rangle_{L^2(\mu_2)} - \langle f, g_0 \rangle_{L^2} - \langle f, g_1 \rangle_{L^2} + \frac{1}{4}\langle g_0, g_0 \rangle_{L^2} + \frac{1}{4}\langle g_1, g_1 \rangle_{L^2} + \frac{1}{2}\langle g_0, g_1 \rangle_{L^2} \\
&= \frac{1}{2}||f - g_0||_{L^2}^2 + \frac{1}{2}||f - g_1||_{L^2}^2 - \frac{1}{4}||g_0 - g_1||_{L^2}^2 \\
&\leq (\alpha + \epsilon)^2 - \frac{1}{4}||g_0 - g_1||_{L^2}^2.
\end{aligned}$$

Since we must have $||f - \frac{g_0+g_1}{2}||_{L^2} \geq \alpha$, it follows that $||g_0 - g_1||_{L^2} \leq \sqrt{8\alpha\epsilon + 4\epsilon^2}$.

This immediately gives the second part: if we had such a g_0, g_1, we would have $||g_0 - g_1||_{L^2} < \delta$ for all δ.

To prove the first part, for each n we may choose g_n with $||f - g_n||_{L^2} < \alpha + 1/n$. This is a Cauchy sequence, since for each $\epsilon > 0$, when n is small enough, $||g_n - g_m||_{L^2} < \epsilon$ for all $m \geq n$. Then by the completeness of L^2, we may choose g to be the limit of this sequence. □

Definition 6.14. When $f \in L^2$, we write

$$\mathbb{E}(f \mid \mathcal{D})$$

for the function in $L^2(\mathcal{D})$ given by the preceding lemma.

When $B \in \mathcal{B}$, we write $\mathbb{E}(B \mid \mathcal{D})$ as an abbreviation of $\mathbb{E}(\chi_D \mid \mathcal{D})$.

We refer to the conditional expectation interchangeably as the *projection onto* $\mathcal{D}$ (or, more properly, onto $L^2(\mathcal{D})$), since the conditional expectation operation is indeed a projection of a vector space onto a subspace.

$\mathbb{E}(f \mid \mathcal{D})$ represents the "best approximation" to f using only the sets from $\mathcal{D}$.

Lemma 6.15. *For any $f \in L^2(\mathcal{B})$ and $g \in L^2(\mathcal{D})$,*

- $\int fg\, d\mu = \int \mathbb{E}(f \mid \mathcal{D})g\, d\mu$,
- $\int (f - \mathbb{E}(f \mid \mathcal{D}))g\, d\mu = 0$.

Proof. The two parts are equivalent (using the linearity of the integral and moving an integral to the other side of the equality), so it suffices to prove the second part.

Suppose not—suppose $|\int (f - \mathbb{E}(f \mid \mathcal{D}))g\, d\mu| = c > 0$. Let $h = \mathbb{E}(f \mid \mathcal{D}) + \frac{c}{||g||_{L^2}} g$. h is also in $L^2(\mathcal{D})$, and we will show it is a better approximation of f:

$$\begin{aligned}
||f - h||_{L^2}^2 &= \int (f - \mathbb{E}(f \mid \mathcal{D}) - \frac{c}{||g||_{L^2}} g)^2\, d\mu \\
&= ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 - 2\frac{c}{||g||_{L^2}} \int (f - \mathbb{E}(f \mid \mathcal{D}))g\, d\mu + \frac{c^2}{||g||_{L^2}^2} ||g||_{L^2} \\
&= ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 - \frac{c^2}{||g||_{L^2}} \\
&< ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2,
\end{aligned}$$

which contradicts the definition of $\mathbb{E}(f \mid \mathcal{D})$. □

We can think of $||\mathbb{E}(f \mid \mathcal{D})||_{L^2}$ as a measurement of what portion of f has been explained by $\mathcal{D}$.

Lemma 6.16.

$$||f||_{L^2}^2 = ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2.$$

Proof.

$$\begin{aligned}
||f||_{L^2}^2 &= \int (f - \mathbb{E}(f \mid \mathcal{D}) + \mathbb{E}(f \mid \mathcal{D}))^2\, d\mu \\
&= ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + 2\int (f - \mathbb{E}(f \mid \mathcal{D}))\mathbb{E}(f \mid \mathcal{D})\, d\mu + ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 \\
&= ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2.
\end{aligned}$$

□

The "grids" of the previous section were examples of conditional expectation: when $V = \bigcup_{i \leq k} B_i$ is a partition of V then the finite collection $\mathcal{D} = \{B_i \times B_j\}_{i,j \leq k}$ is a sub-σ-algebra of $\mathcal{B}_{2,1}$ (because the collection is finite, there are no countable unions or intersections to consider). Because this collection is finite, $L^2(\mathcal{D})$ is exactly the functions we called grids, and Lemma 6.8 shows exactly that the choice of coefficients $\alpha_{i,} = \frac{\int_{B_i \times B_j} f(x,y)\, d\mu_2}{\mu_2(B_i \times B_j)}$ gives the function in $L^2(\mathcal{D})$ minimizes the $L^2(\mu_2)$ distance to f, and therefore in this case

$$\mathbb{E}(f \mid \mathcal{D}) = \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y).$$

It is sometimes convenient to note that when we have bounds on f, these bounds pass over to $\mathbb{E}(f \mid \mathcal{D})$.

Lemma 6.17. *If $f \leq a$ almost everywhere then $\mathbb{E}(f \mid \mathcal{D}) \leq a$ almost everywhere.*

Dually, if $a \leq f$ almost everywhere then $a \leq \mathbb{E}(f \mid \mathcal{D})$ almost everywhere.

Proof. We prove the first part since the second is symmetric. Let $g = \mathbb{E}(f \mid \mathcal{D})$. Suppose that for some $\epsilon > 0$, $\mu(\{x \mid g(x) > a\}) = \delta > 0$. Then let $g'(x) = \max\{g(x), a\}$. Since g is measurable with respect to $\mathcal{D}$, the level set $\{x \mid g(x) > a\}$ is in $\mathcal{D}$, so g' is measurable with respect to $\mathcal{D}$ as well. But

$$||f - g'||_{L^2} \leq ||f - g||_{L^2} - \epsilon^2 \delta < ||f - g||_{L^2},$$

contradicting the minimality of $||f - g||_{L^2}$. □

In particular, if $f = \chi_B$ then $\mathbb{E}(f \mid \mathcal{D})$ is bounded between 0 and 1. Slightly more generally, if $f \in L^\infty(\mathcal{B})$ then $\mathbb{E}(f \mid \mathcal{D}) \in L^\infty(\mathcal{D})$.

Orthogonality to $L^2(\mathcal{D})$ is equivalent to having trivial projection.

Theorem 6.18. *For any $f \in L^2$, $\mathbb{E}(f \mid \mathcal{D})$ is the function which is equal to 0 almost everywhere exactly when, for any $g \in L^2(\mathcal{D})$, $\langle f, g \rangle_{L^2} = 0$.*

Proof. If $\mathbb{E}(f \mid \mathcal{D}) = 0$ then $f - \mathbb{E}(f \mid \mathcal{D}) = f$ and, by the previous lemma, $\langle f, g \rangle_{L^2} = 0$ for any $g \in L^2(\mathcal{D})$.

Conversely, if $\mathbb{E}(f \mid \mathcal{D})$ is not 0 almost everywhere then $||\mathbb{E}(f \mid \mathcal{D})||_{L^2} = \epsilon > 0$. Then

$$\begin{aligned} ||f||_{L^2}^2 &= ||f - \mathbb{E}(f \mid \mathcal{D}) + \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 \\ &= ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + 2\langle f - \mathbb{E}(f \mid \mathcal{D}) \rangle_{L^2} + ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 \\ &= ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2. \end{aligned}$$

In particular, $||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 < ||f||_{L^2}^2$. Since

$$||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 = ||f||_{L^2}^2 - 2\langle f, \mathbb{E}(f \mid \mathcal{D})\rangle_{L^2} + ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2,$$

we must have $\langle f, \mathbb{E}(f \mid \mathcal{D})\rangle_{L^2} > 0$. □

Lemma 6.19. *$\mathbb{E}$ is linear:*

$$\mathbb{E}(af + bg \mid \mathcal{D}) = a\mathbb{E}(f \mid \mathcal{D}) + b\mathbb{E}(g \mid \mathcal{D}).$$

Proof. Consider any $h \in L^2(\mathcal{D})$. Let $z = a\mathbb{E}(f \mid \mathcal{D}) + b\mathbb{E}(g \mid \mathcal{D})$ and $h' = h - z$. Note that $h - z \in L^2(\mathcal{D})$, so $\langle af + bg - z, h - z\rangle_{L^2} = 0$. Then

$$\begin{aligned} ||af + bg - h||_{L^2}^2 &= ||(af + bg - z) - (h - z)||_{L^2}^2 \\ &= ||af + bg - z||_{L^2}^2 - 2\langle af + bg - z, h - z\rangle_{L^2} + ||h - z||_{L^2}^2 \\ &= ||af + bg - z||_{L^2}^2 + ||h - z||_{L^2}^2 \\ &\geq ||af + bg - z||_{L^2}^2, \end{aligned}$$

so $z = \mathbb{E}(af + bg \mid \mathcal{D})$. □

Lemma 6.20. $||\mathbb{E}(f \mid \mathcal{D})||_{L^2} \leq ||f||_{L^2}$

Proof.

$$\begin{aligned} ||f||_{L^2}^2 &= \langle \mathbb{E}(f \mid \mathcal{D}) + (f - \mathbb{E}(f \mid \mathcal{D})), \mathbb{E}(f \mid \mathcal{D}) + (f - \mathbb{E}(f \mid \mathcal{D}))\rangle_{L^2}^2 \\ &= ||\mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 + ||f - \mathbb{E}(f \mid \mathcal{D})||_{L^2}^2 \end{aligned}$$

since $\langle \mathbb{E}(f \mid \mathcal{D}), f - \mathbb{E}(f \mid \mathcal{D})\rangle_{L^2} = 0$. □

6.4 Conditional Expectation on Rectangles

Naturally, we want to apply the work of the previous section with the σ-algebras $\mathcal{B}_{2,1} \subseteq \mathcal{B}_2$.

In addition to the other properties preserved by projections, in this setting there is a notion of symmetry which is also preserved by projections.

Lemma 6.21. *If $f \in L^\infty(\mu_2)$ is symmetric, so is $\mathbb{E}(f \mid \mathcal{B}_{2,1})$.*

Proof. Let $f^+ = \mathbb{E}(f \mid \mathcal{B}_{2,1})$. Suppose not, so $\{(x, y) \mid f^+(x, y) \neq f^+(y, x)\}$ has positive measure. Then there is some $\delta > 0$ so that $B = \{(x, y) \mid$

$f^+(x,y) - f^+(y,x) > \delta\}$ has positive measure. This set belongs to $\mathcal{B}_{2,1}$, as does $B^{op} = \{(y,x) \mid (x,y) \in B\}$, so

$$\begin{aligned}\int_B f^+(x,y) - f^+(y,x)\, d\mu^2 &= \int_B f^+(x,y)\, d\mu^2 - \int_B f^+(y,x)\, d\mu^2 \\ &= \int_B f^+(x,y)\, d\mu^2 - \int_{B^{op}} f^+(x,y)\, d\mu^2 \\ &= \int_B f(x,y)\, d\mu^2 - \int_{B^{op}} f(x,y)\, d\mu^2 \\ &= \int_B f(x,y)\, d\mu^2 - \int_B f(x,y)\, d\mu^2 \\ &= 0.\end{aligned}$$

But this is a contradiction, since $\int_B f^+(x,y) - f^+(y,x)\, d\mu^2 > \delta\mu(B)$. □

We can compare this projection to the U^2 norm.

Theorem 6.22. *$||f||_{U^2} = 0$ if and only if $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} = 0$.*

Note that $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} = 0$ is the same as $\mathbb{E}(f \mid \mathcal{B}_{2,1}) = 0$—that is, except on a set of measure 0, $\mathbb{E}(f \mid \mathcal{B}_{2,1})(x,y) = 0$. Therefore the U^2 norm is 0 exactly if the projection is the trivial function which is 0 almost everywhere.

Proof. By Theorem 6.18, $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2} = 0$ exactly when, for every $g \in \mathcal{L}^2(\mathcal{B}_{2,1})$, $\langle f, g\rangle_{L^2(\mu_2)} = 0$.

Every rectangle $\chi_X(x)\chi_Y(y)$ is in $L^2(\mathcal{B}_{2,1})$, so if $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} = 0$ then for every rectangle $X \times Y$, $\langle f, \chi_X(x)\chi_Y(y)\rangle_{L^2(\mu_2)} = 0$, so by Lemma 6.2, $||f||_{U^2} = 0$.

Conversely, suppose $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} > 0$, so there is a $g \in L^2(\mathcal{B}_{2,1})$ such that $\langle f, g\rangle_{L^2(\mu_2)} > 0$. We want to work with sets, not the function g, so we have to stratify g into level sets: sets of the form $B_{a,b} = \{(x,y) \mid a < g(x,y) \leq b\}$. These all belong to $\mathcal{B}_{2,1}$ (because g is $\mathcal{B}_{2,1}$-measurable, which, by definition, means all its level sets belong to $\mathcal{B}_{2,1}$). Also, for any partition $(-\infty, \infty) = \bigcup_{i<k}(a_i, a_{i+1}]$

$$\int fg\, d\mu_2 = \sum_i \int fg\chi_{B_{a_i,a_{i+1}}}\, d\mu_2 \approx \sum_i \frac{a_i + a_{i+1}}{2} \int f\chi_{B_{a_i,a_{i+1}}}\, d\mu_2,$$

there must be some a, b so that $\left|\int f\chi_{B_{a,b}}\, d\mu_2\right| = \epsilon > 0$.

Since $B_{a,b} \in \mathcal{B}_{2,1}$, there must be some $B \in \mathcal{B}^0_{2,1}$ so that $\mu_2(B_{a,b} \triangle B) < \epsilon/||f||_{L^2(\mu_2)}$, and therefore

$$\left| \int f\chi_B \, d\mu_2 \right| > 0$$

as well.

Since B is a finite union of rectangles, there must be some rectangle with $|\int f\chi_{X\times Y} \, d\mu_2| > 0$. Therefore, by Lemma 6.2, $||f||_{U^2} > 0$. □

It is worth noting that one half of this proof is much harder than the other: when the U^2 norm is non-zero, this quickly implies that the $L^2(\mu_2)$ norm of the projection is non-zero. In the other direction, it requires a lot more work to get from knowing that the $L^2(\mu_2)$ norm of the projection is large to conclude that the U^2 norm is large. Later we will see that this reflects a quantitative fact: knowing that the U^2 norm is larger than some $\epsilon > 0$ will tell us that the $L^2(\mu_2)$ norm of the projection must also be larger than some $\delta > 0$ which we can calculate from ϵ. But in the reverse direction, we will see examples where even though the $L^2(\mu_2)$ norm of the projection is as large as we want—say, 1—the U^2 norm can be arbitrarily small.

Lemma 6.23. *f belongs to $L^2(\mathcal{B}_{2,1})$ if and only if f is dual to U^2.*

Proof. Suppose $||r||_{U^2} = 0$ and $|\int fr \, d\mu_2| = \epsilon > 0$. We may choose $g = \sum_{i\leq k} \alpha_i \chi_{X_i}(x)\chi_{Y_i}(y)$ so that $||f - g||_{L^2(\mu^2)} < \frac{\epsilon}{2||r||_{L^2(\mu^2)}}$. Then

$$\begin{aligned} |\int gr \, d\mu_2 &= \int (f + (g - f))r \, d\mu_2 \\ &= \int fr \, d\mu_2 + \int (g - f)r \, d\mu_2 \\ &> \epsilon - \frac{\epsilon}{2||r||_{L^2(\mu_2)}} ||r||_{L^2(\mu_2)} \\ &> \epsilon/2. \end{aligned}$$

But

$$\int gr \, d\mu_2 = \sum_{i\leq k} \alpha_i \int \chi_{X_i}(x)\chi_{Y_i}(y) d\mu_2 = 0,$$

a contradiction.

Suppose f is dual to U^2. Let $f^+ = \mathbb{E}(f \mid \mathcal{B}_{2,1})$ and $f^- = f - f^+$. Then

$$0 = \int ff^- \, d\mu_2 = \int f^+ f^- \, d\mu_2 + \int f^- f^- \, d\mu_2 = 0 + ||f^-||^2_{L^2(\mu^2)}.$$

So $||f^-||_{L^2(\mu^2)} = 0$, so $f = f^+$, so f is measurable with respect to $\mathcal{B}_{2,1}$. □

So, for our purposes, a function is "structured"—totally non-random—if it belongs to $L^2(\mathcal{B}_{2,1})$. These are exactly the functions approximated by grids we discussed above.

The analogous statement for graphs has to take into account that $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is never 0 (unless E is nearly empty) because $\mu_2(E) = \int \chi_E \, d\mu_2 = \int \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) \, d\mu_2$. Instead, the analogous notion of $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ being trivial is for it to be a constant function (up to measure 0).

Theorem 6.24. *$||\chi_E||_{U^2} = \mu_2(E)$ if and only if $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is the function constantly equal to $\mu_2(E)$.*

Proof. Let $p = \mu_2(E)$ and let f be the "balanced" version of E, $f = \chi_E - p$, so $\int f \, d\mu_2 = 0$.

By Theorem 5.22, $||\chi_E||_{U^2} = \mu_2(E)$ if and only if $||f||_{U^2} = 0$ which, by the previous theorem, holds if and only if $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} = 0$. We finish by noting that $\mathbb{E}(f \mid \mathcal{B}_{2,1}) = \mathbb{E}(\chi_E - p \mid \mathcal{B}_{2,1}) = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) - p$ (since conditional expectation distributions over sums and p is $\mathcal{B}_{2,1}$-measurable), so $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} = 0$ if and only if $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) = p$. □

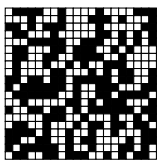
The U^2 norm $||f||_{U^2}$ is positive exactly when f has some correlation with $\mathcal{B}_{2,1}$. However it is not accurate to say that the U^2 norm measures how much f correlates with $\mathcal{B}_{2,1}$; in fact, even if $f \in L^2(\mathcal{B}_{2,1})$, so f is entirely described by $\mathcal{B}_{2,1}$, the U^2 norm can take on any value between $|\int f \, d\mu_2|$ and $||f||_{L^2(\mu_2)}$.

Instead, the U^2 norm reflects something about the "complexity" of f. For example, consider graphs E with $\mu_2(E) = 1/2$. We always have $||\chi_E||_{U^2} \geq 1/2$ in this case, so it makes sense to focus on the U^2 norm of $f(x, y) = \chi_E(x, y) - 1/2$, which counts how many "extra" rectangles the graph E has, beyond the ones it must have. When $E \in \mathcal{B}_{2,1}$, $f \in L^2(\mathcal{B})$, so we will have $||f||_{U^2} > 0$.

When E can be described using a small number of rectangles, $||f||_{U^2}$ will be larger. For example, consider the case where E is the bipartite graph $B_1 \times B_2 \cup B_2 \times B_1$ with $\mu(B_1) = \mu(B_2) = 1/2$. This can be represented by the grid

which only has four "boxes". The U^2 norm of f is 1/2.

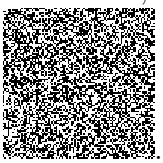
On the other hand, suppose E comes from a grid like



where the boxes are filled in randomly. Again, let $f(x,y) = \chi_E(x,y) - 1/2$. Then $||f||_{U^2}$ is smaller—roughly 0.279 (depending on the exact distribution of squares). And, as the number of squares increases, the U^2 norm continues to decrease. (If we fill in a $k \times k$ grid randomly with black and white boxes to get the graph E, the U^2 norm of $\chi_E - 1/2$ will be roughly $\frac{1}{2^{3/4}k^{1/4}}$.)

Note that $||f||_{L^2(\mu_2)} = ||\chi_E - 1/2||_{L^2(\mu_2)} = 1/2$, so χ_E is always "far" from the constantly $1/2$ function in the $L^2(\mu_2)$ norm. However, when E is built out of many rectangles—when E is "complicated"—χ_E is close to $1/2$ in the U^2 norm.

To see this, consider the 100×100 grid:



Visually, parts of the grid are beginning to look like a grey blur. A finer grid, say 10000×10000, would be indistinguishable to the eye from a constant grey. There is some sense that a graph built from a very large number of small, randomly chosen squares is close to the function which is constantly $1/2$. The norm which reflects that these are close is the U^2 norm.

More generally, suppose we begin with a graph E which may or may not belong to $\mathcal{B}_{2,1}$, but where $||\chi_E||_{U^2}$ is just a bit larger than $\mu_2(E)$. Since $||\chi_E||_{U^2} > \mu_2(E)$, this means that E is not quasirandom. We can decompose $\chi_E = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) + (\chi_E - \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}))$. It could be that $||\chi_E||_{U^2}$ is "not too big" because the structured part $||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)}$ is itself not too big. It could also be that E is entirely, or almost entirely, structured—that $||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)}$ is close to $\mu_2(E)$—but that $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is very complicated.

6.5 Graph Removal

We can now prove the main technical step in our proof of Roth's Theorem (that is, Szemedédi's Theorem for arithmetic progressions of length 3), which is an interesting theorem in its own right.

The *triangle removal lemma* says that if we have a graph with $t_{K_3}(E) = 0$ then there is a set Z whose measure can be a small as we want such that $T_{K_3}(E \setminus Z) = \emptyset$. That is, if a graph has very few triangles, we can remove all of them by removing a small set of edges.

First, we note that subgraph density depends only on $\mathcal{B}_{2,1}$—that is, if we want to examine how dense a subgraph is, we can focus only on the

projection to $\mathcal{B}_{2,1}$.

Lemma 6.25. *For any symmetric $f \in L^\infty(\mu_2)$, $t_H(f) = t_H(\mathbb{E}(f \mid \mathcal{B}_{2,1}))$.*

While the proof involves some fiddling around with Fubini's Theorem, the idea is simple. For simplicity, let us assume $f = \chi_E$. Suppose $H = (W, F)$ and I pick a potential copy $\pi : W \to V$ and want to know what the probabiltiy is that it's an actual copy. That is, I want to evaluate $\int \prod_{\{v_i,v_j\}\in F} \chi_E(x_i, x_j)\, d\mu_{|W|}$. The key observation is that if pick some particular term in this product, all the other terms can only interact with at most one coordinate—that is, I can write this in the form

$$\int \prod_{\{v_i,v_j\}\in F} \chi_E(x_i, x_j)\, d\mu_{|W|} = \int \chi_E(x_1, x_2) f(x_1, x_3, \ldots) g(x_2, x_3, \ldots)\, d\mu_{|W|}$$

where f does not depend on x_2 and g does not depend on x_1. This is precisely a function which depends only on one coordinate at a time (this is what we need to use Fubini's Theorem to show)—that is, the function $(x_1, x_2) \mapsto \int f(x_1, x_3, \ldots) g(x_2, x_3, \ldots)\, d\mu_{|W|-2}$ is $\mathcal{B}_{2,1}$-measurable, which means

$$\int \chi_E(x_1, x_2) f(x_1, x_3, \ldots) g(x_2, x_3, \ldots)\, d\mu_{|W|} = \int \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})(x_1, x_2) f(x_1, x_3, \ldots) g(x_2, x_3, \ldots)\, d\mu_{|}$$

Proof. Let $H = (W, F)$. Let $f^+ = \mathbb{E}(f \mid \mathcal{B}_{2,1})$ and $f^- = f - f^+$. Then

$$t_H(f) = \int \prod_{\{v_i,v_j\}\in F} f(x_i, x_j)\, d\mu_{|W|}.$$

We successively replace each f in this product with f^+: for $S \subseteq F$, we show by induction on $|S|$ that

$$t_H(f) = \int \prod_{\{v_i,v_j\}\in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\}\in F\setminus S} f(x_i, x_j)\, d\mu_{|W|}.$$

When $S = \emptyset$, this is the definition of $t_H(f)$. Suppose $S = S' \cup \{(v, v')\}$. Then

$$\begin{aligned} t_H(f) &= \int \prod_{\{v_i,v_j\}\in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i, x_j) f(v, v')\, d\mu_{|W|} \\ &= \int \prod_{\{v_i,v_j\}\in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i, x_j) f^+(v, v')\, d\mu_{|W|} \\ &\quad + \int \prod_{\{v_i,v_j\}\in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i, x_j) f^-(v, v')\, d\mu_{|W|}, \end{aligned}$$

so it suffices to show that

$$\int \prod_{\{v_i,v_j\}\in S} f^+(x_i,x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i,x_j) f^-(v,v')\, d\mu_{|W|} = 0.$$

Here we use Fubini's Theorem:

$$\int \prod_{\{v_i,v_j\}\in S} f^+(x_i,x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i,x_j) f^-(v,v')\, d\mu_{|W|}$$
$$= \int\int \prod_{\{v_i,v_j\}\in S} f^+(x_i,x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i,x_j) f^-(v,v')\, d\mu(x_i,x_j) d\mu_{|W|-2}.$$

For any fixed $|W|-2$ tuple $(x_1,\dots,x_{i-1},x_{i+1},\dots,x_{j-1},x_{j+1},\dots,x_{|W|})$, $\prod_{\{v_i,v_j\}\in S} f^+(x_i,x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i,x_j)$ belongs to $\mathcal{B}_{2,1}$—in fact, we can decompose this function into the form

$$c g_0(x) g_1(x').$$

Therefore

$$\int \prod_{\{v_i,v_j\}\in S} f^+(x_i,x_j) \prod_{\{v_i,v_j\}\in F\setminus S'} f(x_i,x_j) f^-(v,v')\, d\mu(x_i,x_j) = 0.$$

□

The same argument would apply to induced subgraph density.

Definition 6.26. When $f \in L^\infty(\mu_2)$ is symmetric and $H = (W,F)$ with $W = \{w_1,\dots,w_k\}$,

$$t_H^{ind}(f) = \int \prod_{\{w_i,w_j\}\in F} f(x_i,x_j) \prod_{\{w_i,w_j\}\in \binom{W}{2}\setminus F} (1-f(x_i,x_j))\, d\mu_k.$$

Lemma 6.27. *For any symmetric $f \in L^\infty(\mu_2)$, $t_H^{ind}(f) = t_H^{ind}(\mathbb{E}(f \mid \mathcal{B}_{2,1}))$.*

Theorem 6.28 (Triangle Removal)**.** *Suppose $t_{C_3}(E) = 0$. Let $\mathcal{B} \subseteq \mathcal{B}_2$ be a dense algebra. Then for every $\epsilon > 0$ there is a $B \in \mathcal{B}$ so that $\mu_2(E \cap B) < \epsilon$ and*

$$T_{C_3}(E \setminus B) = \emptyset.$$

That is, if there are very few triangles, we can remove a set that is as small as we like and thereby remove *all* triangles. Moreover, we don't want to remove just any set—we want to be able to pick a big enough sub-algebra

of sets $\mathcal{B}$ and promise that the set E belongs to $\mathcal{B}$. The only case we really care about is when $\mathcal{B}$ is the internal sets.

Specifically, following the ideas of the previous sections, we can take a sufficiently fine grid, $\chi_E \approx \sum_{i\leq n} \alpha_{i,j}\chi_{B_i}\chi_{B_j}$ and remove all the sets $E \cap (B_i \times B_j)$ where $\alpha_{i,j}$ is small. Since we only remove edges where the density is small, we remove few edges this way. On the other hand, if the grid is a good enough approximation of E, any time we have $(x, y, z) \in B_i \times B_j \times B_k$ where $\alpha_{i,j}, \alpha_{i,k}$, and $\alpha_{j,k}$ are all not too small, we should get about $\alpha_{i,j}\alpha_{i,k}\alpha_{j,k}\mu(B_i)\mu(B_j)\mu(B_k)$ triangles inside $B_i \times B_j \times B_k$.

Proof. Let $f^+ = \mathbb{E}(E \mid \mathcal{B}_{2,1})$. We have $t_{C_3}(f^+) = t_{C_3}(E) = 0$ by the previous lemma. Let $E^+ = \{(x, y) \mid f^+(x, y) > 0\}$; then we must also have $t_{C_3}(E^+) = 0$. Also

$$\begin{aligned}
\mu_2(E \setminus E^+) &= \int \chi_E(x, y)(1 - \chi_{E^+}(x, y))\, d\mu_2 \\
&= \mu_2(E) - \int \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})\chi_{E^+}\, d\mu_2 \\
&= \mu_2(E) - \int \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})\, d\mu_2 \\
&= \mu_2(E) - \int \chi_E\, d\mu_2 \\
&= 0.
\end{aligned}$$

By Corollary 6.10, we may choose positive measure sets $\{B_i\}_{i\leq d}$ in $\mathcal{B}$ so that $V = \bigcup_{i\leq d} B_i$ and, for suitable constants,

$$||\chi_{E^+} - \sum_{i,j\leq d} \alpha_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)||_{L^2(\mu_2)} < \frac{\sqrt{\epsilon}}{4}.$$

Indeed, we may assume that $\alpha_{i,j} = \frac{\mu_2((B_i\times B_j)\cap E^+)}{\mu_2(B_i\times B_j)}$. Let $E' = \bigcup_{i,j\leq d,\alpha_{i,j}>3/4} B_i \times B_j$.

If $(x, y) \in E^+ \setminus E'$ then $\chi_{E^+}(x, y) = 1$ while $\alpha_{i,j} \leq 3/4$, and therefore $\chi_{E^+} - \sum_{i,j\leq d} \alpha_{i,j}\chi_{B_i}(x)\chi_{B_j}(y) \geq 1/4$. Therefore

$$\begin{aligned}
\mu_2(E^+ \setminus E') &\leq 16 \int (\chi_{E^+} - \sum_{i,j\leq d} \alpha_{i,j}\chi_{B_i}(x)\chi_{B_j}(y))^2\, d\mu_2 \\
&= 16||\chi_{E^+} - \sum_{i,j\leq d} \alpha_{i,j}\chi_{B_i}(x)\chi_{B_j}(y)||^2_{L^2(\mu_2)} \\
&< \epsilon,
\end{aligned}$$

so $\mu_2(E^+ \setminus E') < \epsilon$, so $\mu_2(E \setminus E') < \epsilon$ as well.

We will let B be the complement of E', so $E \cap B = E \setminus E'$ and $E \setminus B \subseteq E'$. Towards a contradiction, suppose $T_{C_3}(E') \neq \emptyset$. E' is a union of rectangles from the grid $\{B_i\}_{i \leq d}$, so there must be some triangle $(i,j),(j,k),(i,k)$ so that $B_i \times B_j \times B_k \subseteq T_{C_3}(E')$, and therefore $t_{C_3}(E') \geq \mu(B_i)\mu(B_j)\mu(B_k) > 0$.

Since $\mu_2(E^+ \cap (B_i \times B_j)) > 3/4$, at most $\frac{1}{4}\mu(B_i)\mu(B_j)\mu(B_k)$ of the triangles in $B_i \times B_j \times B_k$ do not have $(x,y) \in E^+$. Similarly, at most $\frac{1}{4}\mu(B_i)\mu(B_j)\mu(B_k)$ fail to have $(x,z) \in E^+$ and at most $\frac{1}{4}\mu(B_i)\mu(B_j)\mu(B_k)$ fail to have $(y,z) \in E^+$.

So $t_{C_3}(E^+) \geq \frac{1}{4}\mu(B_i)\mu(B_j)\mu(B_k) > 0$, which contradicts the assumption. So $T_{C_3}(E') = \emptyset$, and therefore $B = E \setminus E'$ is the promised set. □

Theorem 6.28 generalizes to arbitrary graphs with only notational changes.

Theorem 6.29. *Let $H = (W,F)$ be a finite graph and suppose $t_H(E) = 0$. Let $\mathcal{B} \subseteq \mathcal{B}_2$ be a dense algebra. Then for every $\epsilon > 0$ there is a $B \in \mathcal{B}$ so that $\mu_2(B) < \epsilon$ and*

$$T_H(E \setminus B) = \emptyset.$$

Proof. Let $f^+ = \mathbb{E}(E \mid \mathcal{B}_{2,1})$. By Lemma 6.25, $t_H(f^+) = t_H(E) = 0$, so letting $E^+ = \{(x,y) \mid f^+(x,y) > 0\}$, also $t_H(E^+) = 0$ and $\mu_2(E \setminus E^+) = 0$.

By Corollary 6.10 we may choose a partition $V = \bigcup_{i \leq d} B_i$ so that each $B_i \in \mathcal{B}$, $\mu_1(B_i) > 0$, and, taking $\alpha_{i,j} = \frac{\mu_2((B_i \times B_j) \cap E^+)}{\mu_2(B_i \times B_j)}$,

$$||\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \frac{\sqrt{\epsilon}}{\sqrt{|F|+1}}.$$

Set $E' = \bigcup_{i,j \leq d, \alpha_{i,j} > 1 - \frac{1}{|F|+1}} B_i \times B_j$. If $(x,y) \in E^+ \setminus E'$ then $\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) > \frac{1}{|F|+1}$, so

$$\begin{aligned} \mu_2(E^+ \setminus E') &\leq (|F|+1) \int \chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) \, d\mu_2 \\ &= (|F|+1) ||\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||^2_{L^2(\mu_2)} \\ &< \epsilon. \end{aligned}$$

So we take $B = E \setminus E'$, and therefore $E \setminus B = E \cap E' \subseteq E'$. Suppose $T_H(E') \neq \emptyset$, so there is a copy $\pi : W \to V$ of H in (V, E'). For each $w \in W$, there is an atom $B_w \in \mathcal{B}'$ so that $w \in B_w$. Then for *every* $\pi' : W \to V$ such that $\pi'(w) \in B_w$, π' must be a copy of H, so $t_H(E') \geq \prod_{w \in W} \mu_1(B_w) > 0$.

We now show that $t_H(E^+) \geq \frac{1}{|F|+1} t_H(E')$, which gives the desired contradiction. Fix an edge $e = (w, w') \in F$ and consider the copies π of H in (V, E') such that $(\pi(w), \pi(w')) \notin E^+$. This is

$$\int \chi_{E'}(1 - \chi_{E^+}) \prod_{f \in F \setminus \{e\}} \chi_{E'} \, d\mu_{|W|}.$$

Since all the copies of $\chi_{E'}$ are measurable with respect to $\mathcal{B}'$, this is equal to

$$\int \chi_{E'}(1 - \mathbb{E}(\chi_{E^+} \mid (\mathcal{B}')^2)) \prod_{f \in F \setminus \{e\}} \chi_{E'} \, d\mu_{|W|}.$$

By the definition of E', this is

$$\leq \frac{1}{|F|+1} t_H(E').$$

If π is a copy of H in (V, E') but not (V, E^+) then there must be some such edge, so the set of such π has measure

$$\leq \sum_{e \in F} \frac{1}{|F|+1} t_H(E') = \frac{|F|}{|F|+1} t_H(E').$$

That leaves at least $\frac{1}{|F|+1} t_H(E')$ copies which must also be copies in (V, E^+), so $t_H(E^+) \geq \frac{1}{|F|+1} t_H(E') > 0$. This gives the desired contradiction. □

There is a finite analog of this theorem, known as the Graph Removal Theorem.

Corollary 6.30 (Graph Removal)**.** *For every $\epsilon > 0$ and every finite graph $H = (W, F)$, there is a $\delta > 0$ so that whenever $G = (V, E)$ is a graph with $t_H(G) < \delta$, there is a $B \subseteq \binom{V}{2}$ so that $\frac{|B|}{\binom{|V|}{2}} < \epsilon$ and $T_H(E \setminus B) = \emptyset$.*

Proof. Suppose not, so for some $\epsilon > 0$ and some finite graph $H = (W, F)$, for each n there is a $G_n = (V_n, E_n)$ with $t_H(G_n) < 1/n$ but whenever $B \subseteq \binom{V_n}{2}$ with $\frac{|B|}{\binom{|V_n|}{2}} < \epsilon$, $T_H((V_n, E_n \setminus B)) \neq \emptyset$.

Note that $t_H(G) < 1/n$ and $T_H(G) \neq \emptyset$ implies that $|V_n|^{|W|} > n$, so $\lim_{n \to \infty} |V_n| = \infty$.

Consider any ultraproduct $[G_n]_{\mathcal{U}} = G = (V, E)$ and the corresponding Keisler graded probability space, so $t_H(G) = 0$. Let $\mathcal{B} \subseteq \mathcal{B}_2$ be the internal sets. Then there is a $B \in \mathcal{B}$ with $\mu_2(B) < \epsilon$ and $T_H(E \setminus B) = \emptyset$. Since B is internal, $B = [B_n]_{\mathcal{U}}$. In particular, we may choose an n with $\frac{|B_n|}{|V_n|} < \epsilon$ and such that $T_H(E_n \setminus B_n) = \emptyset$, contradicting the assumption. □

6.6 Roth's Theorem

We can now prove Roth's Theorem—that is, Szemerédi's Theorem for arithmetic progressions of length 3, which we state as Theorem 6.31 below. The idea is to construct a graph whose triangles represent arithmetic progressions of length 3; the "trivial progressions"—progressions of the form a, a, a—will be a collection of triangles which, despite having measure 0, cannot be removed by removing a small number of edges, and then triangle removal will imply that there must be many triangles, some of which are non-trivial.

Theorem 6.31 (Roth's Theorem)**.** *For every $\epsilon > 0$, there is an N so that whenever $n \geq N$ and $A \subseteq \{1, 2, \ldots, n\}$ is a set with $\frac{|A|}{n} \geq \epsilon$, there is an $a \in A$ and a $d > 0$ such that $a, a + d, a + 2d \in A$.*

Proof. Suppose not. Then there is some $\epsilon > 0$ which is a counterexample: for every N there is an $n \geq N$ and an $A_n \subseteq \{1, 2, \ldots, n\}$ with $\frac{|A_n|}{n} \geq \epsilon$ but A contains no arithmetic progression $a, a + d, a + 2d$.

We define graphs $G_n = (V_n, E_n)$ as follows:

- V_n is the disjoint union of X_n, Y_n, and Z_n where $X_n = Y_n = Z_n = \{1, 2, \ldots, 3n\}$,
- $(x, y) \in X_n \times Y_n$ belongs to E_n if $x + 2y \mod 3n \in A_n$,
- $(x, z) \in X_n \times Z_n$ belongs to E_n if $2z - x \mod 3n \in A_n$,
- $(y, z) \in Y_n \times Z_n$ belongs to E_n if $z + y \mod 3n \in A_n$.

We go up to $3n$ rather than n to avoid some minor technical issues where the modulus would cause fake progressions from overlowing past n.

Suppose $(x, y, z) \in T_{C_3}(E_n)$. Let $a = x + 2y \mod 3n$ and $d = z - (x + y) \mod 3n$. Then $a \in A_n$; in particular, $a \leq n$. Also $a + d = x + 2y + z - x - y = y + z \mod 3n \in A_n$, so $a + d \mod 3n \leq n$ as well. Similarly, $a + 2d \mod 3n \in A_n$. If $d > 0$ then $d \mod n$ gives an arithmetic progression.

However if $d = 0$ then $z = x + y \mod 3n$. Let $T_n = \{(x, y, x + y \mod 3n) \mid x + 2y \in A_n\}$ (the "trivial" triangles).

For any fixed $y \in Y_n$, $|\{x \mid x - 2y \mod 3n \in A_n\}| = |A_n| \geq \epsilon n^2$, so

$$\frac{|E_n \cap (X_n \times Y_n)|}{9n^2} \geq \frac{\epsilon}{9}.$$

Similarly, $\frac{|E_n \cap (Y_n \times Z_n)|}{9n^2} \geq \frac{\epsilon}{9}$ and $\frac{|E_n \cap (X_n \times Z_n)|}{9n^2} \geq \frac{\epsilon}{9}$, so

$$\frac{|E_n|}{9n^2} \geq \frac{\epsilon}{3}.$$

Relatedly, we need one additional structure on the finite sets: consider the functions $\rho_n^x : X_n \times Y_n \to X_n \times Z_n$ and $\rho_n^y : X_n \times Y_n \to Y_n \times Z_n$ given by $\rho_n^x(x,y) = (x, x+y \mod 3n)$ and $\rho_n^y(x,y) = (y, x+y \mod 3n)$. Note that both these maps are one-to-one, and so in particular measure-preserving.

Pick any nonprincipal ultrafilter $\mathcal{U}$ and let $G = [G_n]_\mathcal{U}$ and $E = [E_n]_\mathcal{U}$. We also set $X = [X_n]_\mathcal{U}$, $Y = [Y_n]_\mathcal{U}$, $Z = [Z_n]_\mathcal{U}$, $\rho^x = [\rho_n^x]_\mathcal{U}$, and $\rho^y = [\rho_n^y]_\mathcal{U}$.

Suppose $t_{C_3}(E) = 0$. Then there is an internal set $B = [B_n]_\mathcal{U}$ with $\mu_2(E \setminus B) < \epsilon/12$ such that $T_{C_3}(E \setminus B) = \emptyset$.

It will be more convenient to divide B into the three pieces corresponding to the three pieces of E: set $B_{XY} = B \cap (X \times Y)$, $B_{XZ} = B \cap (X \times Z)$, and $B_{YZ} = B \cap (Y \times Z)$. Consider those $(x,y) \in X \times Y$ such that

- $(x,y) \in E$,
- $(x,y) \notin E \cap B_{XY}$,
- $(x, x+y) \notin E \cap B_{XZ}$,
- $(y, x+y) \notin E \cap B_{YZ}$.

More formally, this is the set

$$E^- = (E \cap (X \times Y)) \setminus [B_{XY} \cup (\rho^x)^{-1}(B_{XZ}) \cup (\rho^y)^{-1}(B_{YZ})].$$

Note that when $(x,y) \in E \cap (X \times Y)$, also $\rho^x(x,y) \in E$ and $\rho^y(x,y) \in E$ (by the definition of E above), so if (x,y) is in $E \setminus E^-$ then either $(x,y) \in E \setminus B_{XY}$, $\rho^x(x,y) \in E \setminus B_{XZ}$, or $\rho^y(x,y) \in E \setminus B_{YZ}$. Therefore

$$\mu_2(E^-) \geq \mu^2(E) - \mu^2(B) > \epsilon/3 - 3\epsilon/4 > \epsilon/12.$$

Therefore E^- is non-empty, so there is an $(x,y) \in E^-$, so $(x, y, x+y) \in T_{C_3}(E \setminus B)$, which is a contradiction.

So $t_{C_3}(E) > 0$. Since $\mu_3(T) = \lim_{n\to\infty} \frac{|T_n|}{27n^3} < \lim_{n\to\infty} \frac{n^2}{27n^3} = 0$, there is an $([x_n]_\mathcal{U}, [y_n]_\mathcal{U}, [z_n]_\mathcal{U}) \in T_{C_3}(E) \setminus T$. Therefore for some (indeed, almost every) n, there is an $(x_n, y_n, z_n) \in E_n \setminus T_n$, giving an arithmetic progression in A_n, which is a contradiction. □

6.7 Szemerédi Regularity

Frequently Corollary 6.10 is combined with the observation that most pairs in the resulting approximation are "regular". Recall the notion of ϵ-regularity, as specialized to a bipartite graph:

If $E \subseteq X \times Y$, the triple (X, Y, E) is *ϵ-regular* if whenever $X' \subseteq X$ and $Y' \subseteq Y$ with $\mu_1(X') \geq \epsilon\mu_1(X)$ and $\mu_1(Y') \geq \epsilon\mu_1(Y)$, $|d_E(X', Y') - d_E(X, Y)| < \epsilon$.

Lemma 6.32. *If $f \in L^\infty(\mathcal{B}_{2,1})$ and $\mathcal{B} \subseteq \mathcal{B}_1$ is a dense algebra then, for every $\epsilon > 0$, there is a partition $V = \bigcup_{i \leq k} B_i$ with each $B_i \in \mathcal{B}$, $\mu_1(B_i) > 0$, and such that $\mu_2(\bigcup_{i,j \leq k | (B_i, B_j, E) \text{ is not } \epsilon\text{-regular}} B_i \times B_j) < \epsilon$.*

What this lemma is telling us is that we have a finite partition of V^2 into rectangles $B_i \times B_j$ so that E is "well-behaved"—that is, ϵ-regular—in most of the rectangles.

Proof. By Corollary 6.10, choose a a $V = \bigcup_{i \leq k} B_i$ so that each $B_i \in \mathcal{B}$, $\mu_1(B_i) > 0$, and, taking $\alpha_{i,j} = \frac{\mu_2((B_i \times B_j) \cap E)}{\mu_2(B_i \times B_j)}$,

$$||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon^5.$$

Let $\mathcal{S}_0$ be the algebra of sets generated by $\{B_i\}_{i,j \leq k}$ and let $\mathcal{B}_0$ be the algebra of sets generated by rectangles from $\mathcal{S}_0$, so $\mathbb{E}(\chi_E \mid \mathcal{B}_0) = \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)$.

Consider some pair i, j so that (B_i, B_j, E) is not ϵ-regular, and choose $X_{i,j} \subseteq B_i$ and $Y_{i,j} \subseteq B_j$ so that $|d_E(X_{i,j}, Y_{i,j}) - \alpha_{i,j}| \geq \epsilon$. Let $\mathcal{S}_{i,j}$ be the algebra of sets generated by $\mathcal{S}_0$ together with $\{X_{i,j}, Y_{i,j}\}$ and let $\mathcal{B}_{i,j}$ be the algebra of sets generated by rectangles from $\mathcal{S}_{i,j}$. In particular, the atoms of $\mathcal{B}_{i,j}$ are the same as those of $\mathcal{B}_0$ except that $B_i \times B_j$ has been split into four pieces—$X_{i,j} \times Y_{i,j}$, $(B_i \setminus X_{i,j}) \times Y_{i,j}$, $X_{i,j} \times (B_j \setminus Y_{i,j})$, and $(B_i \setminus X_{i,j}) \times (B_j \setminus Y_{i,j})$. Therefore

$$\begin{aligned} ||\mathbb{E}(\chi_E \mid \mathcal{B}_{i,j}) - \mathbb{E}(\chi_E \mid \mathcal{B}_0)||^2_{L^2(\mu_2)} &\geq (d_E(X_{i,j}, Y_{i,j}) - \alpha_{i,j})^2 \mu_1(X_{i,j}) \mu_1(Y_{i,j}) \\ &\geq \epsilon^4 \mu(B_i) \mu(B_j). \end{aligned}$$

(In the first line, we are discarding the other three sub-rectangles since we only care about the inequality.)

Let $Z = \{(i, j) \mid (B_i, B_j, E) \text{ is not } \epsilon\text{-regular}\}$. Let $\mathcal{S}_1$ be the algebra generated by $\bigcup_{(i,j) \in Z} \mathcal{S}_{i,j}$ and let $\mathcal{B}_1$ be the algebra of sets generated by

rectangles from $\mathcal{S}_1$. Then, by our choice of $\{B_i\}_{i\leq k}$,

$$\begin{aligned}\epsilon^5 &> ||\mathbb{E}(\chi_E \mid \mathcal{S}_1) - \mathbb{E}(\chi_E \mid \mathcal{S}_0)||^2_{L^2(\mu_2)} \\ &= \sum_{i,j\leq k} \int_{B_i\times B_j} (\mathbb{E}(\chi_E \mid \mathcal{S}_1) - \mathbb{E}(\chi_E \mid \mathcal{S}_0))^2 \, d\mu_2 \\ &\geq \sum_{(i,j)\in Z} \int_{B_i\times B_j} (\mathbb{E}(\chi_E \mid \mathcal{S}_{i,j}) - \mathbb{E}(\chi_E \mid \mathcal{S}_0))^2 \, d\mu_2 \\ &\geq \epsilon^4 \sum_{(i,j)\in Z} \mu(B_i)\mu(B_j).\end{aligned}$$

□

Lemma 6.32 also has a finitary version. Before proving it, we should note that the property of being an ϵ-regular bipartite graph is preserved by the ultraproduct.

Lemma 6.33. *If $([B_n]_\mathcal{U}, [C_n]_\mathcal{U}, [E_n]_\mathcal{U})$ is ϵ-regular then $\{n \mid (B_n, C_n, E_n)$ is $\epsilon-$ regular$\} \in \mathcal{U}$.*

Proof. The contrapositive is easier: suppose $\{n \mid (B_n, C_n, E_n)$ is not $\epsilon -$ regular$\} \in \mathcal{U}$. Then for each such n, there is $X_n \supseteq B_n$ and $Y_n \subseteq C_n$ with $\frac{|X_n|}{|B_n|} \geq \epsilon$, $\frac{|Y_n|}{|C_n|} \geq \epsilon$, and $|d_{E_n}(X_n, Y_n) - d_{E_n}(B_n, C_n)| \geq \epsilon$.

Therefore, taking $X = [X_n]_\mathcal{U}$ and $Y = [Y_n]_\mathcal{U}$, $|d_E(X, Y) - d_E(B, C)| \geq \epsilon$, so (B, C, E) is not ϵ-regular. □

Theorem 6.34 (Szemerédi's Regularity Lemma)**.** *For every $\epsilon > 0$ and every k_0, there is an N so that whenever $G = (V, E)$ is a finite graph with $|V| \geq N$, there is a partition $V = \bigcup_{i\leq k} B_i$ such that:*

- *$k \leq N$,*
- *for each i, $|B_i| \leq |V|/k_0$,*
- *there is a set $R \subseteq [1, k] \times [1, k]$ of* regular pairs *so that:*
  - *if $(i, j) \in R$ then (B_i, B_j, E) is ϵ-regular,*
  - *$|\bigcup_{(i,j)\notin R} B_i \times B_j| < \epsilon|V|^2$.*

The first condition says that the size of the partition is not too big—in particular, V can be chosen to be much, much larger than N. It is common to replace the second requirement with the requirement that there be at least k_0 parts and that the partition be an "equipartition"—that the pieces B_i differ in size by at most 1. Such a version can be obtained from this one by a certain amount of additional fiddling with the pieces.

Proof. Suppose the thorem were false, so there is an $\epsilon > 0$ and an n and, for every N, a graph $G_N = (V_N, E_N)$ with $|V_N| \geq N$ so that there is no such partition of size $\leq N$.

Let $G = [G_N]_{\mathcal{U}}$, $V = [V_N]_{\mathcal{U}}$, and $E = [E_N]_{\mathcal{U}}$. By Corollary 6.10, choose a partition $V = \bigcup_{i \leq k} B_i$ so that each B_i is internal, each $\mu_1(B_i) > 0$, and

$$||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) - \mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i,j \leq k})||_{L^2(\mu_2)} < \epsilon^5.$$

If any of the B_i have measure $> 1/k_0$, we split them into into smaller pieces. Then, as in the proof of Lemma 6.32, $\mu_2(\bigcup_{0 \leq i,j \leq k | (B_i, B_j, E) \text{ is not } \epsilon-\text{regular}} B_i \times B_j) < \epsilon$.

Each $B_i = [B_{i,n}]_{\mathcal{U}}$. There are only finitely many i, so we may find an n so that whenever (B_i, B_j, E) is ϵ-regular, $(B_{i,n}, B_{j,n}, E_n)$ is ϵ-regular, and $\mu_1(B_{i,n})$ is $\leq 1/k_0$. □

This does not fully capture the strength of Corollary 6.10. In the ultraproduct, we get a partition which is within ϵ of being optimal among *all* partitions. But if we examine the proof of Lemma 6.32, we only compared the partition we got to one other partition. (Indeed, a partition whose size can be calculated—it has size at most $k2^k$.)

In the finite world, we can't hope for a uniform (independent of $|V|$) bound on an almost optimal partition. For suppose we had some bound N; then consider some number M much larger than N, and a graph $|V|$ with many more than M vertices where we partition $V = \bigcup_{i \leq M} C_i$ where the C_i all have the same size, and then for each rectangle $C_i \times C_j \cup C_j \cup C_i$, we flip a coin and either place the whole rectangle in E (if the coin is heads) or none of it (if the coin is tails). (This is basically the example from the end of Section 6.4.) Then the partition into only N parts does not provide any meaningful information—on each rectangle we will have $\mu_2((B_i \times B_j) \cap E) \approx \frac{1}{2}\mu_2(B_i \times B_j)$ (because the rectangle is subdivided into many rectangles $C_i \times C_j$ which are randomly assigned to be in or out of the graph). Yet there is *some* excellent partition, namely $V = \bigcup_{i \leq M} C_i$.

So we can't hope to find a bound on optimal partitions. But we can hope to find a bound on partitions which are optimal among partitions which are not too much bigger.

Theorem 6.35. *For every $\epsilon > 0$, every n, and every function $F : \mathbb{N} \to \mathbb{N}$ there is an N so that whenever $G = (V, E)$ is a finite graph with $|V| \geq N$, there is a partition $V = \bigcup_{i \leq k} B_i$ such that:*

- $k \leq N$,

- *if $k' \leq F(k)$ and $V = \bigcup_{j \leq k'} D_j$ is a partition so that each $D_j \subseteq B_i$ for some $i \leq k$,*

$$||\mathbb{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j,j' \leq k'}) - \mathbb{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} < \epsilon.$$

Proof. Suppose not. Let $\epsilon > 0$ and F be a counterexample and, for each N, let $G_N = (V_n, E_n)$ with $|V_N| \geq N$ be a graph with no such partition of size $\leq N$.

Let $G = [G_N]_{\mathcal{U}}$, $V = [V_N]_{\mathcal{U}}$, and $E = [E_N]_{\mathcal{U}}$. By Corollary 6.10, choose a partition $V = \bigcup_{i \leq k} B_i$ so that each B_i is internal, each $\mu_1(B_i) > 0$, and

$$||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) - \mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i,j \leq k})||_{L^2(\mu_2)} < \epsilon.$$

Let each $B_i = [B_{i,N}]_{\mathcal{U}}$. In almost every G_N with $N \geq k$, the $B_{i,N}$ form a partition, so by assumption there must be some $k'_N \leq F(k)$ and some partition $V_N = \bigcup_{j \leq k'_N} \{D_{j,N}\}_{j \leq k'_N}$ so that each $D_{j,N} \subseteq B_{i,N}$ for some $i \leq k$ and

$$||\mathcal{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j,j' \leq k'_N}) - \mathcal{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} \geq \epsilon.$$

$F(k)$ is finite, so there must be some k' so that $\{N \mid k'_N = k'\} \in \mathcal{U}$. For each $j \leq k'$, let $D_j = [D_{j,N}]_{\mathcal{U}}$. For each j and each N, there is an $i_{j,N} \leq k$ so that $D_{j,N} \subseteq B_{i_{j,N},N}$, so there is an i_j so that $\{N \mid i_{j,N} = i_j\} \in \mathcal{U}$, so $D_j \subseteq B_{i_j}$. So the partition $V = \{D_j\}_{j \leq k'}$ refines $\{B_i\}_{i \leq k}$, and therefore

$$||\mathcal{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} + \epsilon \leq ||\mathcal{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j,j' \leq k'_N})||_{L^2(\mu_2)} \leq ||\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)},$$

contradicting the choice of $\{B_i\}_{i \leq k}$. □

This version was introduced by Alon, Fischer, Krivelevich, and Szegedy in [8] and has gone by a variety of names, like "strong", "robust", and "metastable".

6.8 Which Rectangles are Needed

Given $f \in L^2(\mu_2)$, the question of whether $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0$ seems like it ought to depend on the particular choice of σ-algebra $\mathcal{B}_1$. For instance, if we replace $\mathcal{B}_1$ with some $\mathcal{B}'_1 \subsetneq \mathcal{B}_1$, fewer functions are measurable with respect to $\mathcal{B}'_{2,1} = \mathcal{B}'_1 \times \mathcal{B}'_1$. We might think that a function f could be measurable with respect to $\mathcal{B}_{2,1}$ but not $\mathcal{B}'_{2,1}$.

Yet our work above shows that this is not the case—$||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0$ if and only if $||f||_{U^2} \neq 0$, and the integral $\int f(x,y)f(x',y)f(x,y')f(x',y')\,d\mu_4$

does not depend on the particular σ-algebra $\mathcal{B}_1$. The explanation is that the definition of a Keisler graded probability space forces certain sets to belong to $\mathcal{B}_1$, and therefore these must be sufficient to express the projection of f onto $\mathcal{B}_{2,1}$. In particular, this tells us that when $||\mathbb{E}(f \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0$, f not only correlates with some rectangle, but correlates with a rectangle which can iteself be described using f. We can look more closely at the proof to identify which sets are actually used to construct the rectangles we need.

Lemma 6.36. *Let $f \in L^\infty(\mu_2)$ with $||f||_{U^2} > 0$ and let $\mathcal{B} \subseteq \mathcal{B}_1$ be any σ-algebra large enough that, for almost every $a \in V$, we have the property*

> *for every interval $I \subseteq \mathbb{R}$, $\{x \mid f(a,x) \in I\}$ and $\{x \mid f(x,a) \in I\}$ are both in $\mathcal{B}$.*

Then $||\mathbb{E}(f \mid \mathcal{B} \times \mathcal{B})||_{L^2(\mu_2)} > 0$.

Proof. Without loss of generality, we assume that $\{(x,y) \mid |f(x,y)| \leq 1\}$ has measure 1. (If not, there is some $c > 0$ so that f/c has this property, and we can work with f/c instead.)

Since

$$0 < ||f||_{U^2}^4 = \int f(x,y) \left[\int f(x,y')f(x',y)f(x',y') \, d\mu_2(x',y') \right] d\mu(x,y),$$

there must be a set of x', y' of positive measure so that

$$\left| \int f(x,y)f(x,y')f(x',y) \, d\mu_2(x,y) \right| = \epsilon > 0.$$

We can choose some x', y' whose level sets are in $\mathcal{B}$.

Choosing $n > 1/\epsilon$, we may divide $[-1,1]$ into intervals, $[-1,1] \subseteq \bigcup_{i \leq n} [-1 + i\epsilon/2, -1 + (i+1)\epsilon/2)$. For each $i \leq n$, let $A_i = \{x \mid f(x,y') \in [-1+i\epsilon/2, -1+(i+1)\epsilon/2)$ and $B_i = \{y \mid f(x',y) \in [-1+i\epsilon/2, -1+(i+1)\epsilon.2)\}$. Then

$$\begin{aligned}
\epsilon &= \left| \int f(x,y)f(x,y')f(x',y) \, d\mu_2(x,y) \right| \\
&= \left| \sum_{i,j \leq n} \int_{A_i \times B_j} f(x,y)f(x,y')f(x',y) \, d\mu_2(x,y) \right| \\
&\leq \left| \sum_{i,j \leq n} \int_{A_i \times B_j} f(x,y)(-1+i\epsilon/2)(-1+j\epsilon/2) \, d\mu_2(x,y) \right| + \epsilon/2.
\end{aligned}$$

In particular, there must be some rectangle $A_i \times B_j \in \mathcal{B} \times \mathcal{B}$ such that $|\int_{A_i \times B_j} f(x,y) \, d\mu_2| > 0$. □

Theorem 6.37. *Let $f \in L^\infty(\mu_2)$ be given and let $\mathcal{B} \subseteq \mathcal{B}_1$ be any σ-algebra containing such that, for almost every $a \in V$,*

> *for every interval $I \subseteq \mathbb{R}$, the sets $\{x \mid f(a,x) \in I\}$ and $\{x \mid f(x,a) \in I\}$.*

Then $\mathbb{E}(f \mid \mathcal{B} \times \mathcal{B}) = \mathbb{E}(f \mid \mathcal{B}_{2,1})$.

Proof. Let $f^- = f - \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B})$.

Since $\mathbb{E}(f \mid \mathcal{B} \times \mathcal{B})$ is $\mathcal{B} \times \mathcal{B}$-measurable, it must be the case (by Fubini's Theorem) that for every $a \in V$ and every interval I, the sets $\{x \mid f(a,x) \in I\}$ and $\{x \mid f(x,a) \in I\}$ are in $\mathcal{B}$. So for almost every a, it is also the case that $\{x \mid f^-(a,x) \in I\} = \{x \mid f(a,x) - \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B})(a,x) \in I\}$ is in $\mathcal{B}$, and similarly for $\{x \mid f^-(x,a) \in I\}$.

If $\mathbb{E}(f \mid \mathcal{B} \times \mathcal{B}) \neq \mathbb{E}(f \mid \mathcal{B}_{2,1})$ then $||\mathbb{E}(f^- \mid \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0$, so by the previous lemma, $||\mathbb{E}(f^- \mid \mathcal{B} \times \mathcal{B})||_{L^2(\mu_2)} \neq 0$, which is a contradiction. □

6.9 Points of Density

There is another perspective on the proofs of Traingle and Graph Removal (Theorems 6.28 and 6.29) which is perhaps more consistent with our focus on infinitary and limit behavior. Instead of looking at a fixed finite partition, we will take a limit of finer and finer partitions.

We can think of the sets in our partitions as behaving like neighborhoods in a topological sense. (Indeed, we could think of the sets in our partition as being a clopen basis for some topology.) Around any point, we have smaller and smaller neighborhoods—that is, as we look at finer and finer partitions, we look at the component from the partitions containing our point.

We can then ask which points are "typical". As motivation, recall *Lebesgue's density theorem*: if $A \subseteq \mathbb{R}^n$ is a Lebesgue measurable set then for almost every point $x \in A$, $\lim_{\epsilon \to 0} \frac{\mu(A \cap B_\epsilon(x))}{\mu(B_\epsilon(x))} = 1$. (Here $B_\epsilon(x)$ is the ball of radius ϵ around x.) The same applies to the complement of A, so for almost every $x \in \mathbb{R}^n \setminus A$, $\lim_{\epsilon \to 0} \frac{\mu(A \cap B_\epsilon(x))}{\mu(B_\epsilon(x))} = 0$.

Our setting is slightly different—$E \in \mathcal{B}_2$ definitely need not be Lebesgue measurable, and our neighborhoods are clopen partition components instead of balls—but we can still find similar behavior saying that, for most points in E, E is "locally dense" in a certain sense.

A countable sequence of finite partitions will only involve countably many sets. Our space $\mathcal{B}_1$ is not necessarily separable, so we cannot expect to generate all of it using a countable sequence of partitions. We need to

establish a sub-σ-algebra which is big enough. For our uses here, we only need to be big enough to make sense of E, but it will be useful later to consider the more general case where we might have countably sets.

Lemma 6.38. *If $\{E_1, \ldots, E_i, \ldots\}$ is a countable subset of $\mathcal{B}_2$ then there is a separable $\mathcal{B}_1^- \subseteq \mathcal{B}_1$ such that, for all i, $\mathbb{E}(\chi_{E_i} \mid \mathcal{B}_{2,1})$ is $(\mathcal{B}_1^-)^2$-measurable.*

Proof. For each i, take the countably many level sets of the form $\{(x,y) \mid \mathbb{E}(\chi_{E_i} \mid \mathcal{B}_{2,1})(x,y) > q\}$. For each of these level sets and each k, there is a finite set $\mathcal{S}_{i,q,k} \subseteq \mathcal{B}_1$ such that $\{(x,y) \mid \mathbb{E}(\chi_{E_i} \mid \mathcal{B}_{2,1})(x,y) > q\}$ is approximated to within $1/k$ by rectangles from $\mathcal{S}_{i,q,k}$. Let $\mathcal{B}_1^-$ be the σ-algebra generated by the union of the countably many finite sets $\mathcal{S}_{i,q,k}$. $\square$

Let us fix, for this section, a sequence of neighborhoods. Fix $\mathcal{B}_1^-$ given by the lemma so that $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is $(\mathcal{B}_1^-)^2$-measurable. For each $j \in \mathbb{N}$, let $\mathcal{N}^j = \{N_1^j, \ldots, N_{k_j}^j\}$ be a partition of V such that:

- when $i < j$, $\mathcal{N}^j$ refines $\mathcal{N}^i$ (that is, for each $u < k_j$ there is a $v < k_i$ with $N_u^j \subseteq N_v^i$),
- $\lim_{j\to\infty} \max_{u \le k_j} \mu(N_u^j) = 0$,
- every N_u^j has positive measure,
- $\bigcup_j \mathcal{N}^j$ generates $\mathcal{B}_1^-$.

Each $\mathcal{N}^j$ gives us an approximation $\mathbb{E}(\chi_E \mid \mathcal{N}^j)$ to $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$. Taking the pointwise limit, define

$$\widetilde{E}(x,y) = \lim_{j\to\infty} \mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y)$$

wherever this limit exists. This is the same as setting

$$\widetilde{E}(x,y) = \lim_{j\to\infty} \frac{\mu(E \cap (N_u^j \times N_v^j))}{\mu(N_u^j \times N_v^j)}$$

where u, v are chosen so that $(x,y) \in N_u^j \times N_v^j$.

Usually we only think of $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ as being defined up to the L^2-norm. But, as we will see, $\widetilde{E}$ gives a particular pointwise representation of this function (albeit one that depends on the particular choice of partitions).

Note that, even though E is a set, $\widetilde{E}$ is a function. (We write $\widetilde{E}$ to indicate that E has been "smoothed out"—instead of worrying about the precise value of χ_E at (x,y), $\widetilde{E}$ is concerned with the average value of χ_E

near (x, y).) Unlike in the Lebesgue measurable case, we cannot expect that $\widetilde{E}$ will almost everywhere be 0 or 1—for instance, if E is quasirandom then $\widetilde{E}$ will be the function constantly equal to $1/2$.

However this is the only additional complication, so we can define a point of density to be one where all the points "near" (x, y) (in the sense of our neighborhoods) behave similarly to (x, y).

Definition 6.39. For any $(x, y) \in V^2$, write $\mathcal{N}^j(x, y)$ for the unique set $N^j_u \times N^j_v$ such that $x \in N^j_u$ and $y \in N^j_v$.

(x, y) is a *point of density for* E if $\widetilde{E}(x, y)$ exists and

$$\lim_{j\to\infty} \frac{1}{\mu(\mathcal{N}^j(x,y))} \int_{\mathcal{N}^j(x,y)} |\widetilde{E}(u, v) - \widetilde{E}(x, y)|\, d\mu(u, v) = 0.$$

Lemma 6.40. *$\widetilde{E}$ is defined almost everywhere and is a representation of $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$. Additionally, almost every (x, y) is a point of density for E, for almost every $(x, y) \in E$, $\widetilde{E}(x, y) > 0$, and for almost every $(x, y) \notin E$, $\widetilde{E}(x, y) < 1$.*

Proof. First, note that the approximations $\mathbb{E}(\chi_E \mid \mathcal{N}^j)$ converge to $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ in the L^2 norm: for any $\epsilon > 0$, we may choose j_0 large enough that $||\mathbb{E}(\chi_E \mid \mathcal{N}^{j_0}) - \mathbb{E}(\chi_E \mid \bigcup_j \mathcal{N}^j)|| < \epsilon$, and since the $\mathcal{N}^j$ with $j \geq j_0$ refine $\mathcal{N}^{j_0}$, the same holds with any $j \geq j_0$. But $\mathbb{E}(\chi_E \mid \bigcup_j \mathcal{N}^j) = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$.

To see that the pointwise limit is defined almost everywhere and that almost every point is a point of density, consider any $\epsilon > 0$ and $\alpha < \beta$. Let $g = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ be any representative. Choose j_0 large enough that there is a union of rectangles from $\mathcal{N}^{j_0}$, S, so that $\mu(S \triangle \{(x, y) \mid g(x, y) \leq \alpha\}) < \frac{\beta-\alpha}{1-\alpha}\epsilon$.

We will argue that on most of S, the average of g remains $< \beta$ for all $j \geq j_0$. Consider the exceptions: for each $j \geq j_0$, we can list those rectangles in $\mathcal{N}^j$ so that $R \subseteq S$ and $\frac{1}{\mu(R)} \int_R \chi_E\, d\mu \geq \beta$; call this $\mathcal{D}_j$. Let $\mathcal{D} \subseteq \bigcup_j \mathcal{D}_j$ consist of those $R \in \mathcal{D}_j$ such that, for all $j' \in [j_0, j)$, no element of $\mathcal{D}_{j'}$ contains R. That is, for each j we identify the *new* rectangles where our average is too high and throw them into $\mathcal{D}$, and once we have done so we never throw a subset of that rectangle into $\mathcal{D}$. Then the union of the rectangles in $\mathcal{D}$ is the same as the union of the rectangles in $\bigcup_j \mathcal{D}_j$, but the elements of $\mathcal{D}$ are pairwise disjoint.

Let $D = \bigcup \mathcal{D}$. Because D is a disjoint union of rectangles where the average is above β, we also have $\frac{1}{\mu(D)} \int_D \chi_E\, d\mu \geq \beta$. Since $\chi_E \leq 1$, we must have $\{(x, y) \in D \mid \chi_E(x, y) > \alpha\} \geq \frac{\beta-\alpha}{1-\alpha}\mu(D)$, and therefore $\mu(D) < \epsilon$. Therefore, once $j \geq j_0$, except for a subset of S of measure $\epsilon + \frac{\beta-\alpha}{1-\alpha}\epsilon$, if

$g(x,y) \le \alpha$ then for all $j \ge j_0$, $\mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) \le \beta$ as well. So the set of points with $g(x,y) \le \alpha$ but $\limsup \mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) > \beta$ has measure $< \epsilon + \frac{\beta-\alpha}{1-\alpha}\epsilon$. Dually, we can show that the set of points with $g(x,y) \ge \beta$ but $\liminf \mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) < \alpha$ has measure $< \epsilon$. Since this holds for all ϵ and all $\alpha < \beta$, for almost all (x,y) we have $\widetilde{E}(x,y) = \lim_{j\to\infty} \mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) = g(x,y)$.

Choose j_0 large enough that, except on a set S of measure ϵ^2, for all $j \ge j_0$ we have $|\mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) - \widetilde{E}(x,y)| < \epsilon$. Consider any $j \ge j_0$ and any $R \in \mathcal{N}^j$ so that $\mu(R \cap S) < \epsilon\mu(R)$. Then for any $(x,y) \in R \setminus S$, we have

$$\begin{aligned}
\frac{1}{\mu(R)} \int_R |\widetilde{E}(u,v) - \widetilde{E}(x,y)|\, d\mu(u,v) &\le \frac{1}{\mu(R)} \Bigg[\int_{R\setminus S} |\widetilde{E}(u,v) - \mathbb{E}(\chi_E \mid \mathcal{N}^j)(u,v)| \\
&\quad + |\mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y) - \widetilde{E}(x,y)| \\
&\quad + |\mathbb{E}(\chi_E \mid \mathcal{N}^j)(u,v) - \mathbb{E}(\chi_E \mid \mathcal{N}^j)(x,y)|\, d\mu(u,v) \\
&\quad + \int_S |\widetilde{E}(u,v) - \widetilde{E}(x,y)|\, d\mu(u,v) \Bigg] \\
&\le \frac{1}{\mu(R)} \left[\int_{R\setminus S} \epsilon + \epsilon + 0\, d\mu + \mu(S) \right] \\
&= 3\epsilon.
\end{aligned}$$

Since this holds for every ϵ, almost every (x,y) is a point of density.

Let Z be the set of points with $\widetilde{E}(x,y) = 0$, so Z is $\mathcal{B}^-_{2,1}$-measurable. Therefore $0 = \int_Z \widetilde{E}(x,y)\, d\mu = \mu(E \cap Z)$. Similarly, letting P be the set of points with $\widetilde{E}(x,y) = 1$, P is $\mathcal{B}^-_{2,1}$-measurable, so $0 = \int_P 1 - \widetilde{E}(x,y)\, d\mu = \mu(P \setminus E)$. □

The main result we need is an argument that says that when we can "blow up" graphs on points of density to get positive measure copies.

Theorem 6.41. *Let (W,F) be a finite graph and $\pi : W \to V$ be such that, for each $\{x,y\} \in F$, $\widetilde{E}(\pi(x),\pi(y)) > 0$ and $(\pi(x),\pi(y))$ is a point of density of E. Then $t_{(W,F)}(E) > 0$.*

Proof. The idea is that we can replace each of the points $\pi(x)$ with a neighborhood $\mathcal{N}^j(\pi(x))$ and then use the fact that each E appears with positive density in the necessary neighborhoods to find many copies.

Let $\epsilon \le \min_{\{x,y\}\in F} \widetilde{E}(\pi(x),\pi(y))$. Since each $\{\pi(x),\pi(y)\}$ is a point of density, we may choose j large enough that, for each $\{x,y\} \in F$,

$$\frac{1}{\mu(\mathcal{N}^j(\pi(x),\pi(y)))} \mu(\{(u,v) \in \mathcal{N}^j(\pi(x),\pi(y)) \mid \widetilde{E}(u,v) \ge \epsilon/2\}) > 1 - \frac{1}{|F|}.$$

Consider the set $N = \prod_{w \in W} \mathcal{N}^j(\pi(w))$. Taking the product, we also have

$$\frac{1}{\mu(N)}\mu(\{\{x_w\}_{w\in W} \in N \mid \widetilde{E}(x_u, x_v) \geq \epsilon/2\}) > 1 - \frac{1}{|F|}$$

for each $\{u, v\} \in F$. Therefore

$$\begin{aligned} t_{(W,F)}(\widetilde{E}) &\geq \left(\frac{\epsilon}{2}\right)^{|F|} \mu(\{x_w\}_{w\in W} \in N \mid \text{for every } \{u,v\} \in F,\ \widetilde{E}(x_u, x_v) \geq \epsilon/2) \\ &> \left(\frac{\epsilon}{2}\right)^{|F|} (1 - |F|\frac{1}{|F|})\mu(N) \\ &= 0. \end{aligned}$$

Then by Lemma 6.25, we have $t_{(W,F)}(E) = t_{(W,F)}(\widetilde{E}) > 0$. □

The proof of graph removal, Theorem 6.29, is then quite short. Indeed, it makes clear that we can prove a slightly stronger statement.

Theorem 6.42. *There is a $Z \in \mathcal{B}_2$ so that $\mu_2(Z) = 0$ and for every finite graph H, either $T_H(E \setminus Z) = \emptyset$ or $t_H(E) > 0$.*

Proof. Let Z consist of those points in E which are either not points of density of E, or such that $\widetilde{E}(x, y) = 0$. Then $\mu(Z) = 0$. If $T_H(Z) \neq \emptyset$ for some $H = (W, F)$ then any $\pi \in T_H(Z)$ has the property that $\widetilde{E}(\pi(x), \pi(y)) > 0$ and $(\pi(x), \pi(y))$ is a point of density of E. Therefore, by the previous lemma, $t_H(E) > 0$. □

6.10 Separable Realizations and Graphons

The measurable graph (V, E, μ_1) we get out of an ultraproduct is a useful representation, but as a formal object, it is not as simple as it could be: the accompanying Keisler graded probability space $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ takes some work to define, and is less familiar than a conventional probability space.

We can take some steps to rearrange this object in terms of more familiar probability spaces—specifically, we can instead think of E as coming from a product of separable measure spaces.

The idea is this: we will obtain two sub-σ-algebras of $\mathcal{B}_2$. One will be $\mathcal{B}_{2,1}$ (more precisely, a separable sub-σ-algebra of $\mathcal{B}_{2,1}$), and the other, $\mathcal{R}$ (a "random complement"), will consist entirely of elements of $\mathcal{B}_2$ which are orthogonal to $\mathcal{B}_{2,1}$. A typical element of $\mathcal{B}_2$ will a union of sets of the form $B \cap R$ where $B \in \mathcal{B}_{2,1}$ and $R \in \mathcal{R}$. Since the elements of R are random, we

always have $\mu(B \cap R) = \mu(B)\mu(R)$, so instead of thinking of B and R as being subsets of the same copy of V^2, we will think of them as subsets of *distinct* copies of V^2.

That is, we will replace $(V^2, \mathcal{B}_2)$ with $(V \times V \times V^2, \mathcal{B}_1 \times \mathcal{B}_1 \times \mathcal{R})$, where the latter is a traditional product measure.

Formally, what relationship can we have between these two spaces so that E can be "the same" as some as some set from this product measure space? We can't mean a isomorphism of measure spaces, because an isomorphism would preserve being a product measure. We need a weaker notion: that the σ-algebras are equivalent as algebras with a measure.

Definition 6.43. When $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ are probability measure spaces, a *measurable equivalence of measure algebras from X to Y* is a function $\rho : X \to Y$ such that:

- ρ is measurable and measure-preserving (that is, for every $C \in \mathcal{C}$, $\rho^{-1}(C) \in \mathcal{B}$ and $\mu(\rho^{-1}(C)) = \nu(C)$), and
- for every $B \in \mathcal{B}$, there is a $C \in \mathcal{C}$ with $\mu(B \triangle \rho^{-1}(C)) = 0$.

This does not obviously give an equivalence relation on probability measure spaces (for instance, a measurable equivalence need not have an inverse). However it does give us a natural way to see elements of $\mathcal{C}$ as giving an alternate representation of the elements of $\mathcal{B}$.

Constructing the measurable equvialence requires a certain amount of measure-theoretic formality.

The main theorem we need is to show that we can actually obtain the sub-σ-algebra $\mathcal{R}$. (This is essentially Maharam's Theorem [112]. We are following the proof in [48].)

Theorem 6.44. *Let $\mathcal{B}$ be a separable σ-algebra and let $\mathcal{C} \subseteq \mathcal{B}$ be a sub-σ-algebra. Suppose that for every k there is a partition $\{S_{1,k}, \ldots, S_{k,k}\}$ in $\mathcal{C}$ so that, for each i, $\mathbb{E}(S_{i,k} \mid \mathcal{C})$ is the function constantly equal to $1/k$.*

Then there is a $\mathcal{D} \subseteq \mathcal{B}$ such that every element of $\mathcal{D}$ is orthogonal to $\mathcal{C}$ and every element of $\mathcal{B}$ is generated by $\mathcal{C}$ together with $\mathcal{D}$.

Proof. Fix a countable sequence $\{E_1, E_2, \ldots\}$ generating $\mathcal{B}$. Let $\mathcal{P}_k$ be the finite algebra generated by $\{E_1, \ldots, E_k\} \cup \{S_{i,j} \mid i \leq j \leq k\}$. Let $\mathcal{P}_k^*$ be the atoms of $\mathcal{P}_k$.

Fix an ordering on each $\mathcal{P}_k^*$ so that the orderings are compatible—if $A_1, A_2 \in \mathcal{P}_k^*$ with $A_1 < A_2$ and $A_1', A_2' \in \mathcal{P}_{k+1}^*$ with $A_1' \subseteq A_1$ and $A_2' \subseteq A_2$ then $A_1' < A_2'$.

For each atom in $\mathcal{P}_k^*$, we can fix a version of the projection onto $\mathcal{C}$, and we may assume that for all x, $\sum_{A \in \mathcal{P}_k^*} \mathbb{E}(\chi_A \mid \mathcal{C})(x) = 1$.

We should imagine the atoms, as ordered in $\mathcal{P}_k^*$, as being an order in which points are listed—the points in the first atom appear before the points in the second atom, and so on. We also consider the $\mathcal{C}$-measurable approximation to this ordering, in which we look at the partial sums $\sum_{A' \leq A} \mathbb{E}(\chi_{A'} \mid \mathcal{C})(x)$. This partial sum tells us "how likely is it that x has been list by the time we reach A".

Then for each x and $\lambda \in [0, 1)$, we can ask when the sum crosses λ: we define $A(x, \lambda, k)$ to be the least $A \in \mathcal{P}_k^*$ such that $\sum_{A' \leq A} \mathbb{E}(\chi_{A'} \mid \mathcal{C})(x) > \lambda$. That is, $A(x, \lambda, k)$ is "the position at which the probability that x has been listed exceeds λ". For each $A \in \mathcal{P}_k^*$, we may set $T(A, \lambda, k)$ to be the set of x such that $A(x, \lambda, k) = A$. Note that $T(A, \lambda, k)$ is $\mathcal{C}$-measurable (it is determined by the $\mathcal{C}$-measurable functions $\mathbb{E}(\chi_A \mid \mathcal{C})$).

Then we define two sets

$$S(\lambda, k) = \bigcup_{A \in \mathcal{P}_k^*} (T(A, \lambda, k) \cap \bigcup_{A' < A} A')$$

and

$$S'(\lambda, k) = \bigcup_{A \in \mathcal{P}_k^*} (T(A, \lambda, k) \cap \bigcup_{A' \leq A} A').$$

That is, $S(\lambda, k)$ consists of those points which $\mathcal{C}$ thinks have probability more than λ of appearing by A, and which actually appear before A. $S'(\lambda, k)$ is those points which $\mathcal{C}$ thinks have probability more than λ of appearing by A, and which actually appear by A (either before A, or in A).

The idea is that if $\mathcal{C}$ thinks some points have probability λ of appearing by A, the chance that a point has actually appeared should be just about λ. So $S(\lambda, k)$ is going to be an approximation of a set whose projection onto $\mathcal{C}$ is pretty close to constantly equal to λ.

Note that $S(\lambda, k) \subseteq S(\lambda, k')$, since we arranged the orders on the atoms to be compatable. So we will take $S(\lambda) = \bigcup_k S(\lambda, k)$. The sets $S(\lambda)$ are increasing: if $\lambda < \lambda'$ then $S(\lambda) \subseteq S(\lambda')$.

When $A \in \mathcal{P}_k^*$, $\mathbb{E}(\chi_A \mid \mathcal{C}) \leq 1/k$, since, as an atom, each A is contained in one of the sets $S_{i,k}$. This means that $\sum_{A < A(x,\lambda,k)} \mathbb{E}(\chi_A \mid \mathcal{C})(x) \in (\lambda - 1/k, \lambda]$ (because we know $\mathbb{E}(\chi_{A(x,\lambda,k)} \mid \mathcal{C})(x) \leq 1/k$, and that adding this extra term will bring us over λ).

Then since

$$\mathbb{E}(S(\lambda, k) \mid \mathcal{C}) = \sum_{A, A' \in \mathcal{P}_k^*, A < A'} \mathbb{E}(\chi_A \mid \mathcal{C}) \chi_{T(A', \lambda, k)} = \sum_{A' < A(x,\lambda,k)} \mathbb{E}(\chi_{A'} \mid \mathcal{C})(x),$$

we also have $\mathbb{E}(S(\lambda,k) \mid \mathcal{C}) \in (\lambda - 1/k, \lambda]$.

Therefore $\mathbb{E}(S(\lambda) \mid \mathcal{C}) = \lambda$, so we can take $\mathcal{D}$ to be generated by the sets $S(\lambda)$.

It remains to show that every element of $\mathcal{B}$ is generated by $\mathcal{C}$ and $\mathcal{D}$. It suffices to show that initial segments of $\mathcal{P}_k^*$—that is, sets of the form $\bigcup_{A' \leq A} A'$—are generated by $\mathcal{C}$ and $\mathcal{D}$, since this suffices to generate every atom, and therefore every element of $\mathcal{P}_k$, and the $\mathcal{P}_k$ generate all of $\mathcal{B}$.

So consider some initial segment $I = \bigcup_{A \leq A_0} A$ in $\mathcal{P}_k$. I is also an initial segment in each $\mathcal{P}_t$ with $t > k$. For each $d < t$, we can let $F_d = \{x \mid \mathbb{E}(I \mid \mathcal{C})(x) \in [d/t, (d+t)/t)\}$. Then we will take $I_t = \bigcup_{d<t}(F_d \cap S(d/t))$. The set I_t is generated by $\mathcal{C}$ and $\mathcal{D}$, so it suffices to show that the I_t approximate I.

First, observe that

$$\mathbb{E}(\chi_{I_t} \mid \mathcal{C})(x) = \sum_{d<t} \chi_{F_d}(x)\mathbb{E}(\chi_{S(d/t)} \mid \mathcal{C})(x).$$

Consider some x. There is a $d < t$ so that $x \in F_d$. Then $\mathbb{E}(\chi_{I_t} \mid \mathcal{C})(x)$ is within $1/t$ of $\sum_{A<A(x,d/t,t)} \mathbb{E}(\chi_A \mid \mathcal{C})(x)$. Since $\mathbb{E}(\chi_I \mid \mathcal{C})(x) = \sum_{A \leq A_0} \mathbb{E}(\chi_A \mid \mathcal{C})(x)$ is between d/t and $(d+1)/t$, we have

$$|\mathbb{E}(\chi_I \mid \mathcal{C})(x) - \mathbb{E}(\chi_{I_t} \mid \mathcal{C})(x)| \leq 3/t.$$

So, for all x, $|\mathbb{E}(\chi_I \mid \mathcal{C})(x) - \mathbb{E}(\chi_{I_t} \mid \mathcal{C})(x)| \leq 3/t$. But $\mu(I \triangle I_t) = \int |\chi_I - \chi_{I_t}|\, d\mu \leq \int \mathbb{E}(|\chi_I - \chi_{I_t}| \mid \mathcal{C})\, d\mu$, so the I_t approximate I. □

Corollary 6.45. *There is a σ-algebra $\mathcal{B}_2^- \subseteq \mathcal{B}_2$ containing E, σ-algebra $\mathcal{B}_1^-$ and $\mathcal{R}$, and a measurable equivalence $\rho : V^2 \to V \times V \times V^2$ from $(V^2, \mathcal{B}_2^-, \mu_2)$ to $(V \times V \times V^2, \mathcal{B}_1^- \times \mathcal{B}_1^- \times \mathcal{R}, \mu)$.*

Proof. Applying Lemma 6.38, we obtain $\mathcal{B}_1^- \subseteq \mathcal{B}_1$ so that $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is $(\mathcal{B}_1^-)^2$-measurable. Let $\mathcal{B}_2^-$ be the separable sub-σ-algebra of $\mathcal{B}_2$ generated by $(\mathcal{B}_1^-)^2$ and E.

By Theorem 6.44, we find $\mathcal{R} \subseteq \mathcal{B}_{2,1}^-$ so that every element of $\mathcal{R}$ is orthogonal to $(\mathcal{B}_1^-)^2$, and together $(\mathcal{B}_1^-)^2$ and $\mathcal{R}$ generate $\mathcal{B}_{2,1}^-$. Then the function $\rho : V^2 \to V \times V \times V^2$ given by $\rho(w,v) = (w,v,(w,v))$ is a measurable equivalence: when we have $C \in \mathcal{B}_{2,1}^-$ and $R \in \mathcal{R}$, $C \cap R = \rho^{-1}(C \times R)$, and more generally, given $B \in \mathcal{B}_2^-$, it is approximated by sets of the form $\bigcup_{i \leq n} C_i \cap R_i$, so B is approximated by sets of the form $\rho^{-1}(\bigcup_{i \leq n} C_i \times R_i)$.) □

So instead of working with measurable graphs on a Keisler graded probability space, we could work with subsets of a product space. One disadvantage of this is that it becomes less clear what the elements mean: it is somewhat

harder to view an element $(v, w, (v', w'))$ as a pair of points in a graph in a direct way.

With graphs, one solution is to simply integrate away the random part.

Definition 6.46. A *graphon* is a symmetric measurable function $W : V^2 \to [0, 1]$ on an atomless probability measure space $(V, \mathcal{B}, \mu)$.

Note that this definition requires W to be measurable in the usual sense of the product measure on V^2, not a more general Keisler graded probability space.

Given a measurable graph E, we can obtain a graphon by setting $W = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$. Equivalently, if $E \in \mathcal{B}_1^- \times \mathcal{B}_1^- \times \mathcal{R}$ then

$$W(v, w) = \int \chi_E(v, w, (v', w'))\, d\mu_2(v', w').$$

When $W : V^2 \to [0, 1]$ is a graphon, the elements of V represent vertices of a graph-like object and the value $W(x, y)$ represents a probability that there is an edge between x and y. Like with a measurable graph, we can make sense of notions like subgraph densities by taking integrals—for instance, the triangle density of W is $\iiint W(x, y)W(x, z)W(y, z)\, dx\, dy\, dz$.

We can sample a finite graph from W in two steps: first we select n vertices $\mathbf{v}_n$ randomly from V, and then for each pair $(\mathbf{v}_i, \mathbf{v}_j)$, we independently determine if there is an edge with probability $W(\mathbf{v}_i, \mathbf{v}_j)$.

We have already seen how to get a graphon from a sequence of finite graphs: if G_n is a sequence of finite graphs, we may take $(V, E) = [G_n]_{\mathcal{U}}$ and then let the corresponding graphon be $W = \mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$.

This suggests why considering only $\{0, 1\}$-valued graphons is not enough if we want graphons to be product measurable. We want every sequence of finite graphs to have a subsequence which converges to a graphon (in the sense that $\lim_{n\to\infty} t_H(G_n) = t_H(W)$ for each finite graph H). If we take $G_n = \mathbf{R}_{1/2}$, the random graph on n vertices, we need $t_{\mathrm{K}_2}(W) = 1/2$ and $t_{\mathrm{C}_4}(W) = 1/16$. As we have seen in Theorem 6.24, this can only happen if W is the function constantly equal to $1/2$.

It is common to add an additional requirement to the definition of a graphon, that Ω be the interval $[0, 1]$ with the usual Lebesgue measure. However this is only a notational change: all separable atomless measure spaces are equivalent to Lebesgue measure on the interval (see, for instance, [61, 331P]).

6.11 Higher Order Transfer

The relationship between Corollary 6.10 and Theorem 6.35 is an example of a more general equivalence.

Note that Corollary 6.10 has the wrong form to apply the transfer theorem (Theorem 5.24): transfer applies to statements of the form "for every natural number y there is a natural number z so that $\sigma_{y,z}$ is true", while Corollary 6.10 can be interpreted as saying

> for every $\epsilon > 0$ there is a k so that for every d
>
> > there is a partition $V = \bigcup_{i \leq k} B_i$ so that whenever $V = \bigcup_{i \leq d} C_i$, $||f - \mathbb{E}(f \mid \{B_i \times B_j\})|| < ||f - \mathbb{E}(f \mid \{C_i \times C_j\})|| + \epsilon$.

The inner part ("there is a partition...") is internal, but the outer quantifier structure involves three quantifers: "for every ϵ there is a k so that for every d". A generalization of the transfer theorem holds for such statements, and resembles the strong regularity lemma.

Theorem 6.47. *Suppose that, for every triple of natural numbers y, z, w, $\phi_{y,z,w}(x_1, \ldots, x_m)$ is a first-order formula. Chose parameters $[b_n^1]_{\mathcal{U}}, \ldots, [b^{m-k}]_{\mathcal{U}}$, and let*

$$X = \{([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \mid \forall y \exists z \forall w \, \phi_{y,z,w}([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}, [b_n^1]_{\mathcal{U}}, \ldots, [b_n^{m-k}]_{\mathcal{U}})\}.$$

Then $([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \in X$ if and only if for every y and every $W : \mathbb{N} \to \mathbb{N}$, there is a z so that

$$\{n \mid \phi_{y,z,W(z)}(a_n^1, \ldots, a_n^k, b_n^1, \ldots, b_n^{m-k})\} \in \mathcal{U}.$$

Proof. Consider some $([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \in [V_n^k]_{\mathcal{U}}$. Suppose $([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \notin X$, so there is some y so that, for every z there is a w_z with $\phi_{y,z,w_z}([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}, [b_n^1]_{\mathcal{U}}, \ldots, [b_n^{m-k}]_{\mathcal{U}})$ false. Let W be the function with $W(z) = w_z$. Suppose that, for this y and W, we had some z so that

$$\{n \mid \phi_{y,z,W(z)}(a_n^1, \ldots, a_n^k, b_n^1, \ldots, b_n^{m-k})\} \in \mathcal{U}.$$

But this would mean that $\phi_{y,z,w_z}([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}, [b_n^1]_{\mathcal{U}}, \ldots, [b_n^{m-k}]_{\mathcal{U}})$, which contradicts our choice of w_z.

Conversely, suppose that $([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \in X$. Then for any y and W, we can ignore W—there is some z so that, for every w, $\phi_{y,z,w}([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}, [b_n^1]_{\mathcal{U}}, \ldots, [b_n^{m-k}]_{\mathcal{U}})$ holds. Then, for every w,

$$\{n \mid \phi_{y,z,w}(a_n^1, \ldots, a_n^k, b_n^1, \ldots, b_n^{m-k})\} \in \mathcal{U}.$$

So, in particular,

$$\{n \mid \phi_{y,z,W(z)}(a_n^1, \ldots, a_n^k, b_n^1, \ldots, b_n^{m-k})\} \in \mathcal{U}.$$

□

Corollary 6.48 (Metastable Transfer)**.** *Suppose that, for every triple of natural numbers y, z, w, $\sigma_{y,z,w}(x_1, \ldots, x_m)$ is a first-order sentence. Then the following are equivalent:*

- *in $[G_n]_{\mathcal{U}}$, for every y there is a z so that for every w $\sigma_{y,z,w}$ is true,*
- *for every y and every W, there is a z so that $\{n \mid \sigma_{y,z,W(z)}$ is true in $G_n\} \in \mathcal{U}$.*

Corollary 6.49. *Suppose that, for every triple of natural numbers y, z, w, $\sigma_{y,z,w}(x_1, \ldots, x_m)$ is a first-order sentence. Then the following are equivalent:*

- *in every infinite ultraproduct $[G_n]_{\mathcal{U}}$, for every y there is a z so that for every w, $\sigma_{y,z,w}$ is true,*
- *for every y and W there is a Z and an n so that whenever G_n is a graph with $\geq n$ vertices, there is a $z \leq Z$ so that $\sigma_{y,z,W(z)}$ is true in G_n.*

6.12 Remarks

Szemerédi's regularity lemma first appeared as a step in his proof of Szemerédi's Theorem, and was isolated in [146]. It immediately became a central tool in extremal graph theory, and there are a number of excellent surveys on the uses of the theorem [102, 103, 133].

The σ-algebra perspective on the regularity lemma developed over a period of time, especially in [150, 153], and was developed into the form we use here in [158, 160].

The regularity lemma has notoriously poor bounds: the bound on the number of pieces is known to be bounded by a function on the order of a tower of exponents of size $1/\epsilon$, and this cannot be improved [57, 77, 119]. As a result, alternative proofs have been sought for many of its consequences.

Graph removal in particular has been extensively studied. Graph removal and its variants turn out to have applications in computer science, where they are related to questions of "property testing" [73]—roughly speaking, the question of whether a property of a graph can be tested by a computer

program in a reasonable way. Many variants of graph removal have been proven—for instance, where one modifies a graph with few *induced* copies of a graph to contain no copies [8], where one deals with infinite families of graphs [5], and directed graphs [6]. Proofs of some of these results have been given which do not use the regularity lemma [56] and give somewhat better bounds. Some of these and other results are surveyed in [36].

The Lebesgue density theorem is used to prove graph (and hypergraph) removal in [48]. The canonical book on graphons in general is [110].

The generalization of Szemerédi regularity, Theorem 6.35, was introduced in [8] in order to prove graph removal with induced subgraphs. (Another variant, which corresponds to choosing the function $F(k) = k$, is sometimes called the "weak" regularity lemma [62].) The approach here was identified by Tao [153], who introduced the term "metastable" for such results in a slightly different context [152]. The more general notion of higher order transfer between "metastable" formulations—that is, formulations of the form "for every ϵ and every function F"—and Π_3 statements was first noticed by Kohlenbach [100] and has subsequently been extensively studied [11, 12, 46, 99, 101].

Chapter 7

Combinatorial Structure

In this chapter we will consider graphs which are "combinatorially simple". We'll see that these combinatorial properties imply that the graphs have nice measurable structures—for instance, that they are $\mathcal{B}_{2,1}$-measurable.

7.1 VC Dimension

If we have a graph $G = (V, E)$ (finite or infinite) and a finite set of vertices $\{x_1, \ldots, x_n\} \subseteq V$, each vertex $y \in V$ picks out a subset, namely $\{x_i \mid \{x_i, y\} \in E\} = \{x_1, \ldots, x_n\} \cap N_G(y)$. As we consider different vertices in V, we identify different subsets of $\{x_1, \ldots, x_n\}$ in this way. In a random graph (for example, $\mathbf{R}_{1/2}(V)$ where $|V|$ is much larger than n), we expect to obtain all 2^n subsets of $\{x_1, \ldots, x_n\}$ as we vary y across different values.

Let us consider graphs which are very non-random: graphs where we do not find all the subsets in this way.

Recall that when (V, E) is a graph, we write E_x for the neighborhood of x, $\{y \mid \{x, y\} \in E\}$.

Definition 7.1. If $X \subseteq V$ (typically X is finite), we say E *shatters* X if, for every $S \subseteq X$, there is a $y \in V$ so that $X \cap E_y = S$.

The *VC dimension* of G is the largest n such that there exists some $\{x_1, \ldots, x_n\} \subseteq V$ which is shattered by E, or ∞ if there is no such n.

VC stands for Vapnik-Chervonenkis, the names of the two computational learning theorists who introduced the notion.

Haing having high VC dimension, particularly infinite VC dimension, is a notion of complexity. Note that VC dimension is defined by a worst case scenario: we have high VC dimension if we find a single choice of set

$\{x_1, \ldots, x_n\}$ which is shattered. This reflects the idea that if we combine a simple graph and a complex one (say, by taking the disjoint union of their vertices), we ought to get a complex graph.

Very structured graphs tend to have low VC dimension. For instance, the complete bipartite graph $K_{n,n}$ (with $n \geq 3$) has VC dimension 1: if we take a set $\{x_1, x_2\}$, if these vertices are from different parts then no y has $\{x_1, x_2\} \cap E_y = \{x_1, x_2\}$, and if they are from the same part then no y has $\{x_1, x_2\} \cap E_y = \{x_1\}$.

Similarly, consider the bipartite graph where $V_0 = V_1 = \{1, 2, \ldots, n\}$ and V is the disjoint union of V_0 and V_1, and E is the set of pairs $\{x, y\}$ with $x \in V_0$, $y \in V_1$, and $x < y$. This has VC dimension 1: suppose we take two vertices $\{x_1, x_2\}$. If they are in different parts, we cannot shatter the set because there would be no y with $\{x_1, x_2\} \cap E_y = \{x_1, x_2\}$, and if they are in the same part—without loss of generality, assume V_0—they have an order—say, $x_1 < x_2$—and then we cannot have $\{x_1, x_2\} \cap E_y = \{x_2\}$.

At the other extreme, $\mathbf{R}_{1/2}(V)$ tends to have large VC dimension; in particular, with probability 1, $[\mathbf{R}_{1/2}(V_n)]_{\mathcal{U}}$ will have VC dimension ∞ (assuming $|V_n| \to \infty$).

The notion of VC dimension has been invented several separate times in the literature. In particular, Shelah introduced it in model theory [139], where it goes by the name NIP ("not the independence property"). More precisely, a theory is NIP if every definable graph has finite VC dimension. Although beyond our scope here, the point is that a wide variety of theories have the NIP property, and therefore any graph defined in a model of one of those theories will have finite VC dimension.

An equivalent property is that a graph has finite VC dimension exactly if it omits some bipartite graph.

Theorem 7.2. *(V, E) has infinite VC dimension if and only if whenever $(W_0 \cup W_1, F)$ is a finite bipartite graph, there is a copy $\pi : W_0 \cup W_1 \to V$ of $(W_0 \cup W_1, F)$ such that whenever $w_0 \in W_0$ and $w_1 \in W_1$, $\{\pi(w_0), \pi(w_1)\} \in E$ if and only if $\{w_0, w_1\} \in F$.*

Note that the requirement on π is somewhere between our definition of a copy and our definition of an induced copy: what we are asking for is that π embeds the two parts W_0 and W_1 so that the edges between $\pi(W_0)$ and $\pi(W_1)$ exactly correspond to F, but we don't care what happens inside W_0 and W_1.

Proof. Suppose (V, E) has infinite VC dimension and let $(W_0 \cup W_1, F)$ be a finite bipartite graph. Since the VC dimension of (V, E) is infinite, choose

$X \subseteq V$ with $|X| = |W_0|$ so that X is shattered by E. Choose any bijection $\pi : W_0 \to X$.

For each $w_1 \in W_1$, consider $S_{w_1} = \{\pi(w_0) \mid \{w_0, w_1\} \in F\} \subseteq X$. Since E shatters X, we may find some v_1 with $E_{v_1} \cap X = S_{w_1}$ and define $\pi(w_1) = v_1$. Then π is a induced copy of $(W_0 \cup W_1, F)$ as a bipartite graph as in the sense of the statement of the theorem.

Conversely, suppose (V, E) contains every bipartite graph. In particular, for any n, consider the bipartite graph with $W_0 = \{1, 2, \ldots, n\}$ and $W_1 = \mathcal{P}(\{1, 2, \ldots, n\})$ with $\{w_0, w_1\} \in F$ if and only if $w_0 \in w_1$. If π is a copy of $(W_0 \cup W_1, F)$ then $\pi(W_0)$ is a set of size n and $\pi(W_1)$ witnesses that E shatters $\pi(W_0)$. □

VC dimension is usually considered in a slightly more abstract setting.

Definition 7.3. A *set system* is a set V and a collection $\mathcal{F}$ of subsets of V. When $X \subseteq V$, we say $\mathcal{F}$ *shatters* X if, for every $S \subseteq X$, there is an $F \in \mathcal{F}$ with $X \cap F = S$.

The *VC dimension* of $(V, \mathcal{F})$ is the largest n such that there exists some $X \subseteq V$ with $|X| = n$ so that X is shattered by $\mathcal{F}$.

A graph (V, E) corresponds to the set system on V with $\mathcal{F} = \{E_y \mid y \in V\}$. (Conversely, given a set system $(V, \mathcal{F})$, we can obtain a bipartite graph whose vertices are $V \cup \mathcal{F}$ with an edge between x and F exactly when $x \in F$.)

Unlike the definition in a graph, this definition looks asymmetric. If we are only concerned with whether a set system has finite or infinite VC dimension, however, there is a symmetry.

Lemma 7.4. *Suppose $(V, \mathcal{F})$ has infinite VC dimension. Then for every d, there exist $F_0, \ldots, F_{d-1} \in \mathcal{F}$ so that, for every $S \subseteq \{F_0, \ldots, F_{d-1}\}$, there is an $x \in V$ so that $\{F \in \{F_0, \ldots, F_{d-1}\} \mid x \in F\} = S$.*

Proof. Since $(V, \mathcal{F})$ has infinite VC dimension, we may choose a set $\{x_1, \ldots, x_{2^d}\}$ which is shattered by $\mathcal{F}$. For each $j < d$, we choose F_j containing exactly those x_i such that, when we write i in base 2, the j-th digit is a 1. Then for any $\{F_{j_1}, \ldots, F_{j_s}\} \subseteq \{F_0, \ldots, F_{d-1}\}$, we may let $i = \sum_{k \leq s} 2^{j_k}$, and we have $\{F_{j_1}, \ldots, F_{j_s}\} = \{F \in \{F_0, \ldots, F_{d-1}\} \mid x_i \in F\}$. □

VC dimension gives a striking dividing line. For any finite $X \subseteq V$, we can define $\Pi_{\mathcal{F}}(X) = \{S \mid \exists F \in \mathcal{F}\, X \cap F = S\}$ and $\pi_{\mathcal{F}}(n) = \max\{|\Pi_{\mathcal{F}}(X)| \mid X \subseteq V,\ |X| = n\}$. That is, $\Pi_{\mathcal{F}}(X)$ is the collection of subsets of X we can obtain by looking at neighborhoods and $\pi_{\mathcal{F}}(n)$ is the largest possible size of $\Pi_{\mathcal{F}}(X)$ among sets of size n. Since $\Pi_{\mathcal{F}}(X)$ is a set of subsets of X,

$|\Pi_{\mathcal{F}}(X)| \leq 2^{|X|}$, so $\pi_{\mathcal{F}}(n) \leq 2^n$. The VC dimension of $(V, \mathcal{F})$ is the smallest n with $\pi_{\mathcal{F}}(n+1) < 2^{n+1}$, so when the VC dimension is infinite, $\pi_{\mathcal{F}}(n) = 2^n$ for all n.

It turns out that there are only two possible behaviors for the function $\pi_{\mathcal{F}}$: either the function is the exponential 2^n or it is bounded by a polynomial.

Theorem 7.5 (Sauer–Shelah)**.** *If the VC dimension of $(V, \mathcal{F})$ is d then, for all n, $\pi_{\mathcal{F}}(n) \leq \sum_{i=0}^{d} \binom{n}{i}$.*

There are a very large number of proofs of this theorem, of which we only give one, which uses the following crucial lemma.

Lemma 7.6. *Whenever $(V, \mathcal{F})$ is a set system and $X \subseteq V$ is finite, there are at least $|\Pi_{\mathcal{F}}(X)|$ subsets of X which are shattered by $\mathcal{F}$.*

Proof. By induction on $|X|$. If $X = \emptyset$ then either $\mathcal{F} = \emptyset$, in which case $|\Pi_{\mathcal{F}}(X)| = 0$, which suffices because at least 0 subsets of X are certainly shattered, or $\mathcal{F} \neq \emptyset$, in which case $|\Pi_{\mathcal{F}}(X)| = 1$ and $\emptyset$ is shattered (by definition, as long as $\mathcal{F}$ is non-empty), so the 1 needed subset of X is shattered.

Suppose that $|X| = m + 1$ and choose some $x \in X$. Let $\mathcal{F}_x = \{v \setminus \{x\} \in \mathcal{F} \mid x \in v\}$ and let $\mathcal{F}_- = \mathcal{F} \setminus \mathcal{F}_x$. Then, by the inductive hypothesis, $\mathcal{F}_x$ shatters at least $|\Pi_{\mathcal{F}_x}(X \setminus \{x\})|$ subsets of $X \setminus \{x\}$ while $\mathcal{F}_-$ shatters at least $|\Pi_{\mathcal{F}_-}(X \setminus \{x\})|$ subsets of $X \setminus \{x\}$.

We now count subsets of X shattered by $\mathcal{F}$. First, if $S \subseteq X$ is shattered by $\mathcal{F}_-$, X is shattered by $\mathcal{F}$. Suppose $S \subseteq X$ is shattered by $\mathcal{F}_x$. If S is *also* shattered by $\mathcal{F}_-$ then $S \cup \{x\}$ is shattered by $\mathcal{F}$. Otherwise S is *not* shattered by $\mathcal{F}_-$ but is shattered by $\mathcal{F}$. In particular, for each set shattered by $\mathcal{F}_x$, we obtain an additional set shattered by $\mathcal{F}$, so the number of sets shattered by $\mathcal{F}$ is at least $|\Pi_{\mathcal{F}_-}(X)| + |\Pi_{\mathcal{F}_x}(X)|$.

Finally, observe that

$$\begin{aligned}
|\Pi_{\mathcal{F}}(X)| &= |\{S \mid \exists F \in \mathcal{F}\ X \cap F = S\}| \\
&= |\{S \mid \exists F \in \mathcal{F}\ X \cap F = S \text{ and } x \in S\}| + |\{S \mid \exists F \in \mathcal{F}\ X \cap F = S \text{ and } x \notin S\}| \\
&= |\Pi_{\mathcal{F}_x}(X \setminus \{x\})| + |\Pi_{\mathcal{F}_-}(X \setminus \{x\})|,
\end{aligned}$$

so we have shown that $\mathcal{F}$ shatters enough subsets of X. □

Proof of Sauer-Shelah. Suppose not: then there is an $X \subseteq V$ with $|X| = n$ and $|\Pi_{\mathcal{F}}(X)| > \sum_{i=0}^{d} \binom{n}{i}$. There are only $\sum_{i=0}^{d} \binom{n}{i}$ subsets of X of size $\leq d$, so there must be a subset of X of size $> d$ shattered by $\mathcal{F}$, so the VC dimension of $(V, \mathcal{F})$ is $> d$. □

7.2 Closure of VC Dimension

It is convenient to note that having finite VC dimension is preserved under various ways of combining set systems.

Lemma 7.7. *If $(V, \mathcal{F})$ is a set system with finite VC dimension, the set system consisting of complements, $(V, \{V \setminus F \mid F \in \mathcal{F}\})$, also has finite VC dimension.*

Proof. Let $X \subseteq V$ be a finite set not shattered by $\mathcal{F}$. Then there is an $S \subseteq X$ so that $X \cap F \neq S$ for any $F \in \mathcal{F}$. Let $T = X \setminus S$; then $X \cap (V \setminus F) \neq T$ for any $F \in \mathcal{F}$. □

Lemma 7.8. *Let $(V, \mathcal{F})$ and $(V, \mathcal{G})$ be two set systems on the same sets of points, both having finite VC dimension. Then the intersection set system, $(V, \{F \cap G \mid F \in \mathcal{F} \text{ and } G \in \mathcal{G}\})$ also has finite VC dimension.*

Proof. A slick proof is to make use of the Sauer-Shelah theorem. Let d bound the VC dimension of both set systems, and choose n much larger than d. Consider a subset $X \subseteq V$ of size n and consider how many subsets of X have the form $X \cap (F \cap G)$ with $F \in \mathcal{F}$ and $G \in \mathcal{G}$.

First, consider the subsets of X of the form $X \cap F$ with $F \in \mathcal{F}$: by Sauer-Shelah, there are at most $\sum_{i \leq d} \binom{n}{i}$ of these. For each set $X \cap F$, consider the number of subsets of the form $X \cap F \cap G$ with $G \in \mathcal{G}$: there are, again, at most $\sum_{i \leq d} \binom{|X \cap F|}{i} \leq \sum_{i \leq d} \binom{n}{i}$ of these.

So, in total, there can be at most $(\sum_{i \leq d} \binom{n}{i})^2$ subsets of X of the form $X \cap (F \cap G)$. When n is large enough, this is less than 2^n, so in particular X cannot be shattered. □

Lemma 7.9. *Let $(V, \mathcal{F})$ and $(V, \mathcal{G})$ be two set systems on the same sets of points, both having finite VC dimension. Then the union set system, $(V, \{F \cup G \mid F \in \mathcal{F} \text{ and } G \in \mathcal{G}\})$ also has finite VC dimension.*

Proof. We could give an argument similar to the previous lemma, but we can also note that the union is the complement of the intersection of the complements, so this follows from the previous two lemmata. □

The main application we will need is actually the symmetric difference.

Corollary 7.10. *Let $(V, \mathcal{F})$ and $(V, \mathcal{G})$ be two set systems on the same sets of points, both having finite VC dimension. Then the symmetric difference set system, $(V, \{F \triangle G \mid F \in \mathcal{F} \text{ and } G \in \mathcal{G}\})$ also has finite VC dimension.*

This again follows from the lemmata above because the symmetric difference can be formed using union, intersection, and complement.

7.3 ϵ-Nets

Suppose G is a measurable graph and y is a vertex with $\mu(E_y) \geq \epsilon > 0$. If we select a large number of points $x_1, \ldots, x_n$ at random (with n large relative to $1/\epsilon$), we expect that at least one of these points will probably belong to E_y. If we have several vertices, $y_1, \ldots, y_m$ with each $\mu(E_{y_i}) \geq \epsilon$, we would expect to need to make n larger if we want to find a point in *every* set E_{y_i}—even though we're likely to get a point in E_{y_1}, and a point in E_{y_2}, and so on, the likelihood that we miss at least one of the sets E_{y_i} increases with m. In particular, when there are uncountably many values of y, it could become very likely that we miss at least one of them. For instance, in $[\mathbf{R}_{1/2}]_{\mathcal{U}}$, almost every y has $\mu(E_y) = 1/2$, but we cannot pick an x belonging to all these E_y simultaneously.

If the graph has finite VC dimension, however, we can hope to select a reasonable number of x's so that we have found at least one member of every single E_{y_i}. We call such a set of x's an *ϵ-net*—like a net, it "catches" every set E_y which is not too small.

Definition 7.11. A set $X \subseteq V$ is an *ϵ-net* for $G = (V, E, \mu)$ if, for every $y \in V$ with $\mu(E_y) \geq \epsilon$, $X \cap E_y \neq \emptyset$.

Theorem 7.12. *If $G = (V, E, \mu)$ has finite VC dimension then there is an ϵ-net.*

The proof here is based on the one given in [116].

Proof. Let d be the VC dimension of G. We will show that there is a large enough size r (depending on d and ϵ) so that if X is a set of size r chosen randomly according to μ, there is positive probability that X is an ϵ-net.

Suppose we choose a set X_0 of size r, and let p be the probability that X_0 is *not* an ϵ-net. We ultimately wish to show that p is small, so consider what happens when X_0 is not an ϵ-net: there must be some y so that $\mu(E_y) \geq \epsilon$ but $X_0 \cap E_y = \emptyset$. That means that if we choose r additional elements—an additional set X_1—then (by the Hoeffding inequality), when r is large enough, the probability that we choose X_1 so that $|X_1 \cap E_y| \geq \epsilon r/2$ is at least $3/4$.

So, when we choose *two* sets of size r, X_0 and X_1, the probability is at least $3p/4$ that there is some y with $X_0 \cap E_y = \emptyset$ but $|X_1 \cap E_y| \geq \epsilon r/2$. It looks a bit unlikely that we should be able to totally avoid E_y with the first set, but then find lots of elements in it with the second set, and will indeed show that this situation is unlikely, and therefore that p must be small.

We can find the probability of this situation a different way. Suppose that we first pick a set Z with $2r$ elements, and then choose the r elements

$X_0 \subseteq Z$ randomly from Z. Once we picked the set Z, for any given y with $|E_y \cap Z| \geq \epsilon r/2$, we can ask for the possibility that, when we choose $X_0 \subseteq Z$ so that $E_y \cap X_0 = \emptyset$.

For any fixed y with $|E_y \cap Z| \geq \epsilon r/2$, the probability that we choose r elements from z while avoiding all $\epsilon r/2$ elements of $E_y \cap Z$ is at most $(1 - \epsilon/4)^r$.

Of course, there are many values of y. However—after we've chosen Z—this outcome depends only on the set $E_y \cap Z$. The Sauer-Shelah Theorem tells us that there are only $\Pi_G(2r) \leq C(2r)^d$ possibilities (for some C) for the set $E_y \cap Z$. Therefore the probability that we have $E_y \cap X_0 = \emptyset$ for at least one value of y is bounded by $C(2r)^d(1 - \epsilon/4)^r$. In particular, this decreases exponentially in r, so when r is large enough we can ensure that this probability is at most $1/2$.

Therefore $3p/4 \leq 1/2$, so $p \leq 2/3 < 1$, so in particular there is a positive probability that X_0 is an ϵ-net. □

7.4 VC Dimension and Rectangles

The existence of ϵ-nets gives us a quick tool for constructing approximations using rectangles.

Theorem 7.13. *If $G = (V, E, \mu)$ is a measurable graph with finite VC dimension then $E \in \mathcal{B}_{2,1}$.*

Proof. It suffices to show that, for every $\epsilon > 0$, there is an approximation of E to within ϵ using rectangles.

Consider the set system for symmetric differences, $(V, \{E_x \triangle E_y \mid x, y \in V\})$. This set system also has finite VC dimension, and therefore we may choose an ϵ-net $Z = \{z_1, \ldots, z_m\}$. Since this is an ϵ-net, whenever $\mu_1(E_x \triangle E_y) \geq \epsilon$, there is an i with $z_i \in E_x \triangle E_y$.

For each $s \subseteq Z$, let $V_s = \{x \in V \mid E_x \cap Z = s\}$. For each $s \subseteq Z$ with $V_s \neq \emptyset$, choose some $y_s \in V_s$. Observe that if $x \in V_s$ then $\mu_1(E_x \triangle E_{y_s}) < \epsilon$.

Then let $E' = \bigcup_{s \subseteq Z, V_s \neq \emptyset} V_s \times E_{y_s}$. We claim that $\mu_2(E \triangle E') < \epsilon$:

$$\begin{aligned}
\mu_2(E \triangle E') &= \sum_{s \subseteq Z} \int_{V_s} \mu_1(E_x \triangle E_{y_s})\, d\mu_1 \\
&< \sum_{s \subseteq Z} \mu_1(V_s)\epsilon \\
&= \epsilon.
\end{aligned}$$

□

This gives us a corresponding improvement of Szemerédi's Regularity Lemma (Theorem 6.34) under the assumption of VC dimension. Recall that the regularity lemma promises a uniform bound N so that every finite graph (V, E) has a partition into pieces $V = \bigcup_{i \leq k} B_i$ with $k \leq N$ so that most pairs (B_i, B_j, E) are ϵ-regular.

In applications of the regularity lemma, there are a number of ways in which one could hope it might be improved. First, one would like to improve the bounds—the bound on N is a tower of exponents in $1/\epsilon$. Second, one could hope that most of the pairs are not only ϵ-regular, but have density $d_E(B_i, B_j)$ close to either 0 or 1. Third, one would like to eliminate the "irregular pairs"—one would like to have *every* pair (B_i, B_j, E) be ϵ-regular, instead of most. And finally, one could hope to strengthen regularity to *homogeneity*: we could hope to have either $B_i \times B_j \subseteq E$ or $(B_i \times B_j) \cap E = \emptyset$.

VC dimension gives us the first two of these improvements. In keeping with our focus on what happens in the limit, we do not worry about the bounds, and only prove that graphs with bounded VC dimension satisfy a version of the strong regularity lemma (Theorem 6.35).

Theorem 7.14. *For every d and $\epsilon > 0$ there is an N so that whenever (V, E) is a graph with $|V| \geq N$, there is a partition $V = \bigcup_{i \leq k} B_i$ such that:*

- $k \leq N$,
- $||\chi_E - \mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i,j \leq k})||_{L^2(\mu_2)} < \epsilon$.

Note that this means that, except on rectangles with combined measure at most $\sqrt{\epsilon}$, the average $\mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i,j \leq k})$ must be in $[0, \sqrt{\epsilon}) \cup (1 - \sqrt{\epsilon}, 1]$—that is, on most rectangles, the density of χ_E is either close to 0 or close to 1.

Proof. Suppose not. Let d and $\epsilon > 0$ be given so that this fails, so for each N we have a counterexample $G_N = (V_N, E_N)$ with $|V_N| \geq N$. By the previous theorem, $[E_n]_{\mathcal{U}}$ belongs to $\mathcal{B}_{2,1}$, so there is a partition $[V_N]_{\mathcal{U}} = \bigcup_{i \leq k} B_i$ with each B_i internal so that

$$||\chi_E - \mathbb{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} < \epsilon/2.$$

Then we may pick representatives $B_i = [B_{i,N}]_{\mathcal{U}}$ and find an N so that

$$||\chi_{E_N} - \mathbb{E}(\chi_{E_N} \mid \{B_{i,N} \times B_{i',N}\}_{i,i' \leq k})||_{L^2(\mu_2)} < \epsilon,$$

giving a contradiction. □

Note that (as in Lemma 6.32), this implies that most of the (B_i, B_j, E) are ϵ-regular.

7.5 A Converse of Sorts

Neither the existence of ϵ-nets nor belonging to $\mathcal{B}_{2,1}$ can plausibly be equivalent to having finite VC dimension. These consequences are measure-theoretic, while VC dimension is determined by finite sets, so we can always take a graph with finite VC dimension and add an additional graph with infinite VC dimension, but put no measure on the additional part.

To obtain an equivalence, we have to ask about what happens under *uniformly* under *arbitrary* measures.

Theorem 7.15. *$G = (V, E)$ has finite VC dimension if and only if, for every $\epsilon > 0$, there is an n so that, for any μ making (V, E, μ) a measurable graph, the probability that an element of V^n is an ϵ-net is at least $1 - \epsilon$.*

Proof. We have already shown that graphs with finite VC dimension have this property.

For the converse, suppose $G = (V, E)$ has infinite VC dimension; we will show that, for $\epsilon = 1/2$, for each n there is a measure μ where (V, E, μ) is a measurable graph and the probability that a subset of V^n is an ϵ-net is 0. Let $d \geq 2n$ and take a subset S of V of size d which is shattered, and let μ be the uniform distribution on S (with measure 0 on $V \setminus S$). Then whenever $X \subseteq S$ with $|X| = n$, we may choose, for instance, a $y \in V$ with $E_y \cap S = S \setminus X$, so $\mu(E_y) = \frac{|E_y \cap S|}{|S|} \geq 1/2$, but $E_y \cap X = \emptyset$, so X is not a $1/2$-net. □

That is, VC dimension is equivalent to the existence of ϵ-nets in a uniform way under all possible measures. Analogously, it is equivalent to E being uniformly $\mathcal{B}_{2,1}$-measurable.

Theorem 7.16. *$G = (V, E)$ has finite VC dimension if and only if, for every $\epsilon > 0$, there is an N so that, for any μ making (V, E, μ) a measurable graph, there is a partition of V into $n \leq N$ parts, $V = \bigcup_{i \leq n} B_i$ so that*

$$||\chi_E - \mathbb{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq n})||_{L^2(\mu_2)} < \epsilon.$$

Proof. We have already shown that graphs with finite VC dimension have this property.

For the converse, suppose $G = (V, E)$ has infinite VC dimension. We must pick an $\epsilon > 0$ and, for each N, a measure on (V, E) so that there is no partition of V into $n \leq N$ parts with the desired property.

We let $\epsilon = 1/2$. For each m, choose a random bipartite graph on $m \times m$ vertices—that is, fix sets U_m, V_m with $|U_m| = |V_m| = m$ and, for each

$(u, v) \in U_m \times V_m$, flip an independent fair coin, and let $E_m \subseteq U_m \times V_m$ be the set of pairs for which this coin came up heads. Let G_m be the graph $(U_m \times V_m, E_m)$.

Note that, with probability 1, the sequence of graphs $\langle G_m \rangle$ is quasirandom as a bipartite graph; in particular, we have $\lim_{m\to\infty} \frac{|E_m|}{|U_m \times V_m|} = 1/2$ and $\lim_{m\to\infty} t_{\mathrm{C}_4}(G_m) = 1/16$.

Since G is bipartite, G contains a copy of each G_m, so for each m, we may consider the measure μ_m on G which places its measure uniformly on some copy of G_m. That is, we fix some $\pi_m : G_m \to G$ and define

$$\mu_m(\{x\}) = \begin{cases} \frac{1}{2m} & \text{if } x \in \mathrm{rng}(\pi_m) \\ 0 & \text{otherwise} \end{cases}$$

The ultraproduct of these structures is going to be a quasirandom measurable graph, so we need to notice that if we had a uniform partition of G, we could lift it to the ultraproduct. So consider some ultrafilter $\mathcal{U}$ and the graph $G^* = (V^*, E^*, \mu^*) = [(V, E, \mu_m)]_{\mathcal{U}}$. Note that, up to measure 0, (V, E, μ_m) is isomorphic to G_m with the uniform measure; in particular, $t_{\mathrm{K}_2}(G^*) = 1/2$ and $t_{\mathrm{C}_4}(G^*) = 1/16$, so G^* is quasirandom.

Towards a contradiction, suppose there were an N satisfying the second half of the statement of the theorem: for each m, we have $n_M \leq N$ and a partition $V = \bigcup_{i \leq n} B_i^m$. Let $n = \lim_{n\to\mathcal{U}} n_M$, and for each $i \leq n$, let $B_i = [B_i^m]_{\mathcal{U}}$. For each $i, i' \leq n$, we have $\mathbb{E}(\chi_{E^*} \mid B_i \times B_{i'}) = \lim_{m\to\mathcal{U}} \mathbb{E}_{\mu_m}(\chi_E \mid B_i^m \times B_{i'}^m)$, so we also have

$$||chi_{E^*} - \mathbb{E}(\chi_{E^*} \mid \{B_i \times B_{i'}\}_{i,i' \leq n})||_{L^2(\mu^*)} = \lim_{m\to\mathcal{U}} ||\chi_E - \mathbb{E}_{\mu_m}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq n_m})||_{L^2(\mu_m)} < 1/2.$$

But E^* is quasirandom, so

$$||\chi_{E^*} - \mathbb{E}(\chi_{E^*} \mid \{B_i \times B_{i'}\}_{i,i' \leq n})||_{L^2(\mu^*)}|| \geq ||\chi_{E^*} - \mathbb{E}(\chi_{E^*} \mid \mathcal{B}_{2,1})||_{L^2(\mu^*)} = 1/2,$$

which gives the desired contradiction. □

7.6 Stability

We showed in Theorem 6.12 that there are graphs E so that every finite partition includes some rectangles where E has intermediate density. The example we used was the *half graph*—a graph roughly of the form $\{(x, y) \mid x < y\}$.

We will now show that this is, in a precise sense, the only obstacle.

Definition 7.17. If $G = (V, E)$ is a graph, a *strict ladder of length* d in G is a pair of sequences $x_1, \ldots, x_d, y_1, \ldots, y_d \in V$ so that $\{x_i, y_j\} \in E$ if and only if $i < j$.

We say G is *stable* if there is a d so that there does not exist a ladder of length d.

The word "strict" refers to the fact that we have a requirement when $i = j$ as well; a simple "ladder" would allow either $\{x_i, y_i\} \in E$ or $\{x_i, y_i\} \notin E$.

Stability is a strengthening of having finite VC dimension: finite VC dimension says that, given $\{x_1, \ldots, x_d\}$ with d sufficiently large, there is *some* subset we cannot obtain by looking at the intersection with E_y. Stability prohibits specific combinations of subsets from appearing.

As the example of the half graph shows, stability is strictly stronger—the half graph has finite VC dimension but is unstable.

Like VC dimension, stability gives a bound of sorts on the number of subsets of a set we can hope to obtain, though it has a different character: stability implies that there are only countably many subsets of any countable set.

Theorem 7.18. *Suppose (V, E) is stable. Then whenever $B \subseteq V$ is countable, there are only countably many subsets of B of the form $E_y \cap B$.*

Stated in this form, this follows directly from the Erdős–Makkai Theorem:

Theorem 7.19. *Let $(B, \mathcal{F})$ be a set system with $|\mathcal{F}| > |B|$ and B infinite. Then there exist $x_1, \ldots, x_d \in B$ and $F_0, \ldots, F_d \in \mathcal{F}$ so that $x_i \in F_j$ iff $i \leq j$.*

Proof. By induction on d. For $d = 0$ this is immediate: take any $F_0 \in \mathcal{F}$.

Suppose the claim holds for d. Choose any $F_{d+1} \in \mathcal{F}$ with $F_{d+1} \neq B$. Consider the set system $\mathcal{F}' = \{F \cap F_{d+1} \mid F \in \mathcal{F}\}$.

First, suppose $|\mathcal{F}'| > |B|$. For each $b \in F_{d+1}$, consider $\mathcal{F}_b = \{F \in \mathcal{F}' \mid b \notin F\}$. Observe that, other than F_{d+1} itself, every $F \in \mathcal{F}'$ belongs to at least one $\mathcal{F}_b$, so $\mathcal{F}' = \{F_{d+1}\} \cup \bigcup_{b \in B} \mathcal{F}_b$. Since $|\mathcal{F}'| > |B|$, this means there is some b with $|\mathcal{F}_b| > |B|$. We apply the inductive hypothesis to $(F_{d+1}, \mathcal{F}_b)$, obtaining $x_1, \ldots, x_d, F_0, \ldots, F_d$ with each $x_1, \ldots, x_d \in F_{d+1}$ and $x_i \in F_j$ if and only $i \leq j$. Taking $x_{d+1} = b$ completes the construction.

Otherwise, $|\mathcal{F}'| = |B|$. Then consider instead the family $\mathcal{F}^c = \{B \setminus F \mid F \in \mathcal{F}\}$. Then $|\mathcal{F}^c| > |B|$ as well, and we may apply the same argument as the first case with $B \setminus F_{d+1}$ to obtain $x_1, \ldots, x_{d+1}$ and $B \setminus F_1, \ldots, B \setminus F_{d+1}$ so that $x_i \in B \setminus F_{j+1}$ if and only if $i \leq j$. Then the sequence $x_{d+1}, \ldots, x_1, F_{d+1}, \ldots, F_0$ witnesses the claim. □

In an ultraproduct (and, more generally, structures which satisfy the right saturation properties) this is equivalent to stability.

Theorem 7.20. *Suppose $([V_n]_{\mathcal{U}}, [E_n]_{\mathcal{U}})$ is unstable. Then there is a countable $B \subseteq [V_n]_{\mathcal{U}}$ so that there are uncountably many subsets of B of the form $([E_n]_{\mathcal{U}})_y$.*

Proof. The idea is that instability together with the compactness of ultraproducts allows us to construct a copy of the half graph. For each d we have sequences $x^{d,1}, \ldots, x^{d,d}, y^{d,1}, \ldots, y^{d,d}$ so that $\{x^{d,i}, y^{d,j}\} \in E$ if and only if $i < j$.

Choosing representatives, take $x^{d,i} = [x_n^{d,i}]_{\mathcal{U}}$ and $y^{d,i} = [y_n^{d,i}]_{\mathcal{U}}$. Choose a sequence of sets $I_0 \supseteq I_1 \supseteq \cdots$ in $\mathcal{U}$ such that if $n \in I_d$ then $\{x_n^{d,i}, y_n^{d,j}\} \in E_n$ if and only if $i < j$. For each real number $r \in [0,1]$, define x_n^r, y_n^r for $n \in I_{d+1} \setminus I_d$ by taking $x_n^r = x_n^{d,\lceil dr \rceil}$ and $y_n^r = y_n^{d,\lceil dr \rceil}$. For any $r < s$, observe that $\{n \mid \{x_n^r, y_n^s\} \in E_n\} \in I_{\lceil 1/(r-s) \rceil} \in \mathcal{U}$, so $[x_n^r]_{\mathcal{U}} < [y_n^s]_{\mathcal{U}}$.

Let $B = \{[x_n^q]_{\mathcal{U}} \mid q \in \mathbb{Q} \cap [0,1]\}$. Then for each real number r, $E_{[y_n^r]_{\mathcal{U}}} \cap B = \{[x_n^q]_{\mathcal{U}} \mid q < r\}$. This gives uncountably many distinct subsets of B. □

It would be tempting, by analogy to VC dimension, to define d-stability to mean that d is the largest size for which a strict ladder exists. In fact, this term is reserved for a slightly different configuration.

Definition 7.21. We write $\{0,1\}^d$ for the set of binary sequences of length d—that is, functions $\sigma : [0, d-1] \to \{0,1\}$ and $\{0,1\}^{<d} = \bigcup_{k<d} \{0,1\}^k$. When $\sigma \in \{0,1\}^d$, we set $|\sigma| = d$ and call $|\sigma|$ the *length* of d. We write $\sigma \sqsubseteq \tau$ if σ is an initial segment of τ—that is, if $\tau \restriction [0, |\sigma| - 1]$.

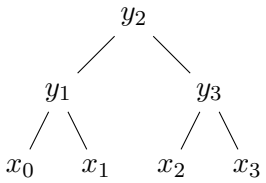
We say $G = (V, E)$ is *d-stable* if there do not exist collections of points $\{x_\sigma\}_{\sigma \in \{0,1\}^d}$ and $\{y_\eta\}_{\eta \in \{0,1\}^{<d}}$ such that whenever $\eta \sqsubset \sigma$, $x_\sigma \in E_{y_\eta}$ if and only if $\eta^\frown\langle 1 \rangle \sqsubseteq \sigma$.

We can think of the y_η as labeling the nodes of a binary tree, indicating that elements in E_{y_η} should go to the left and elements not in E_{y_η} should go to the right. Starting at the root, every element x follows a path: left or right depending on whether it belongs to $E_{y_{\langle\rangle}}$, then left or right again depending on $E_{y_{\langle 0 \rangle}}$ or $E_{y_{\langle 1 \rangle}}$, and so on. d-stability means that, however we label the internal nodes of the tree of height d, there is a path which no element b will follow.

The d in stability and the d in d-stability don't necessarily match, but up to that numeric difference, the notions are equivalent.

Lemma 7.22. *If G is d-stable then G is stable.*

Proof. Suppose G is not stable. For any d, we will show that G is not d-stable. The idea is that if we have a ladder of length 2^d, we can use the x's to label the leaves of a tree of height d, and the y's to label the internal nodes; for instance, when $d = 2$:



Making this precise for a general d is just a matter of working through the indexing. Start with a ladder $x_0, \ldots, x_{2^d-1}, y_0, \ldots, y_{2^d-1} \in V$ so that $\{x_i, y_j\} \in E$ if and only if $j \leq i$. (We have shifted the indexing by one and swapped the names of the x's and y's in order to match the definition of the tree better.) We can associate each $i < 2^d$ with a sequence of legth d by writing it in binary with $\sigma(0)$ being the most significant bit—that is, let σ_i be the sequence such that $i = \sum_{j<d} 2^{d-1-j}\sigma_i(j)$, and take $x_{\sigma_i} = x_i$.

Then for each η, we take $y_\eta = y_{\sum_{j<|\eta|} 2^{d-1-j}\eta(j)+2^{d-1-|\eta|}}$. Observe that if $\eta^\frown\langle 0\rangle \sqsubseteq \sigma$ then $x_\sigma = x_i$ for some $i \leq \sum_{j<|\eta|} 2^{d-1-j}\eta(j) + \sum_{|\eta|<j<d} 2^{d-1-j} < \sum_{j<|\eta|} 2^{d-1-j}\eta(j) + 2^{d-1-|\eta|}$, and therefore $\{x_\sigma, y_\eta\} \in E$, while if $\eta^\frown\langle 1\rangle \sqsubseteq \sigma$ then $x_\sigma = x_i$ for some $i \geq \sum_{j<|\eta|} 2^{d-1-j}\eta(j) + 2^{d-1-|\eta|}$, and therefore $\{x_\sigma, y_\eta\} \notin E$. □

The converse also holds, but is more difficult to prove. We follow the proof in [85], which needs a Ramsey-like theorem for trees.

Definition 7.23. The *binary tree of height* n, written $\{0,1\}^{<n}$, is the collection of all sequences of 0's and 1's of length less than n—that is, $\{0,1\}^{<n} = \bigcup_{0\leq m<n}\{0,1\}^{[1,\ldots,m]}$.

We write $\sigma \sqsubseteq \tau$ if $\tau \restriction \mathrm{dom}(\sigma) = \sigma$.

When $m = 0$, the notation "$\{0,1\}^{[1,\ldots,0]}$" means the function from the empty set to $\{0,1\}$—that is, the empty function, which we write $\langle\rangle$; we call this the *root* of the tree. $\sigma \sqsubseteq \tau$ means that τ extends σ.

Definition 7.24. A *subtree of* $\{0,1\}^{<n}$ *of height* m is a function $f : \{0,1\}^{<m} \to \{0,1\}^{<n}$ such that $f(\sigma) \sqsubseteq f(\tau)$ if and only if $\sigma \sqsubseteq \tau$.

For instance, $f(\sigma)$ must extend $f(\langle\rangle)$ for all $\sigma \in \{0,1\}^{<m}$, and the two nodes $f(\sigma^\frown\langle 0\rangle)$ and $f(\sigma^\frown\langle 1\rangle)$ must be two distinct, incomparable extensions of $f(\sigma)$.

Lemma 7.25. *Suppose that $c : \{0,1\}^{<n_0+n_1} \to \{0,1\}$. Then for some $i \in \{0,1\}$, there is a subtree $f : \{0,1\}^{<n_i} \to \{0,1\}^{<n_0+n_1}$ such that $c(f(\sigma)) = i$ for all $\sigma \in \{0,1\}^{<n_i}$.*

That is, we can find a "monochromatic subtree".

Proof. We proceed by induction on $n_0 + n_1$. When $n_0 + n_1 = 0$ all these trees are empty, so this is trivial.

Suppose $c(\langle\rangle) = 0$. (The case where $c(\langle\rangle) = 1$ is symmetric.) There are two natural subtrees of height $n_0 + n_1 - 1$—the nodes extending $\langle 0\rangle$ and the nodes extending $\langle 1\rangle$.

Formally, consider two colorings $c_j : \{0,1\}^{<n_0+n_1-1} \to \{0,1\}$ given by $c_j(\sigma) = c(\langle j\rangle^\frown\sigma)$. We may apply the inductive hypothesis to each of c_0 and c_1 with $n_0 - 1, n_1$; if, for either j, we obtain $f_j : \{0,1\}^{<n_1} \to \{0,1\}^{<n_0+n_1-1}$ so that $c_j(f_j(\sigma)) = 1$ for all $\sigma \in \{0,1\}^{<n_1}$ then we may take $f(\sigma) = \langle j\rangle^\frown f_j(\sigma)$.

Suppose not. Then we obtain two functions $f_j : \{0,1\}^{<n_0-1} \to \{0,1\}^{<n_0+n_1-1}$ so that $c_j(f_j(\sigma)) = 0$ for all $\sigma \in \{0,1\}^{<n_0-1}$. Then we define $f : \{0,1\}^{<n_0} \to \{0,1\}^{<n_0+n_1}$ by $f(\langle\rangle) = \langle\rangle$ and $f(\langle j\rangle^\frown\sigma) = \langle j\rangle^\frown f_j(\sigma)$. □

Theorem 7.26. *If G is stable then there is a d so that G is d-stable.*

Proof. We will show that if G fails to be $3 \cdot 2^{d-1} - 2$-stable then we can find a strict ladder of length d. Specifically, we will show the following by induction on d:

> Let $h = 3\cdot 2^{d-1}-2$. Suppose we have a tree of vertices $\{x_\sigma\}_{\sigma\in\{0,1\}^h}$ and $\{y_\eta\}_{\eta\in\{0,1\}^{<h}}$ such that when $\eta \sqsubset \sigma$, $\{x_\sigma, y_\eta\} \in E$ if and only if $\eta^\frown\langle 1\rangle \sqsubseteq \sigma$. Then there exist $x_1, \ldots, x_d, y_1, \ldots, y_d$ so that $\{x_i, y_j\} \in E$ if and only if $i < j$, for each i there is a σ so that $x_i = x_\sigma$, and for each j there is an η so that $y_j = y_\eta$.

Let such a tree be given. When $d = 1$, $h = 1$, so we may take $y_1 = y_{\langle\rangle}$ and $x_1 = x_{\langle 0\rangle}$.

Suppose $d > 1$. Let $h' = 3 \cdot 2^{d-2} - 2$, so that $1 + h' + (h' + 1) = h$. Choose some $\sigma \sqsupseteq \langle 0\rangle$ and define $c : \{0,1\}^{<h-1} \to \{0,1\}$ by setting $c(\eta) = 1$ if $\{x_\sigma, y_{\langle 0\rangle^\frown\eta}\} \in E$ and 0 otherwise.

Suppose that there is a function $f : \{0,1\}^{<h'} \to \{0,1\}^{<h-1}$ such that, for each $\eta \in \{0,1\}^{<h'}$, $c(f(\eta)) = 1$. Then for each $\eta \in \{0,1\}^{<h'}$, set $y'_\eta = y_{\langle 0\rangle^\frown f(\eta)}$. For each $\tau \in \{0,1\}^{h'}$, we have $\tau = \eta^\frown\langle j\rangle$ for some $j \in \{0,1\}$; choose any leaf $\tau' \sqsupseteq \langle 0\rangle^\frown f(\eta)^\frown\langle j\rangle$ and set $x'_\tau = x_{\tau'}$. Applying the inductive hypothesis to $\{x'_\tau\}_{\tau\in\{0,1\}^{h'}}, \{y'_\eta\}_{\eta\in\{0,1\}^{<h'}}$, we obtain $x_1, \ldots, x_{d-1}, y_1, \ldots, y_{d-1}$ with $\{x_i, y_j\} \in E$ if and only if $i < j$. We may

set $x_0 = x_\sigma$ and $y_0 = y_{\langle\rangle}$; then we also have $\{x_i, y_0\} \notin E$ for all i, and $\{x_0, y_i\} \in E$ for all $i > 0$.

Suppose there is no such f. Then the previous lemma gives us a function $f : \{0,1\}^{<h'+1} \to \{0,1\}^{<h+1}$ such that, for each $\eta \in \{0,1\}^{<h'+1}$, $c(f(\eta)) = 0$. For each $\eta \in \{0,1\}^{<h'}$, set $y'_\eta = y_{\langle 0\rangle^\frown f(\langle 1\rangle^\frown \eta)}$. For each $\tau \in \{0,1\}^{h'}$, we have $\tau = \eta^\frown\langle j\rangle$ for some $j \in \{0,1\}$; choose any leaf $\tau' \sqsupseteq \langle 0\rangle^\frown f(\langle 1\rangle^\frown\eta)^\frown\langle j\rangle$. Then, by the inductive hypothesis again, we find $x_1, \ldots, x_{d-1}, y_1, \ldots, y_{d-1}$ with $\{x_i, y_j\} \in E$ if and only if $i < j$. We may set $x_d = x_\sigma$ and $y_d = y_{f(\langle\rangle)}$. Then we have $\{x_d, y_i\} \notin E$ for all i and $\{x_i, y_d\} \in E$ for all $i < d$. □

7.7 Stable Regularity

Since stable graphs have finite VC dimension, we know they belong to $\mathcal{B}_{2,1}$. However stable graphs belong to $\mathcal{B}_{2,1}$ in an especially clean way: we can partition V^2 into rectangles so that, in *every* rectangle, E is either almost the whole rectangle or almost none of it.

Definition 7.27. We say A is 0-*good* for E if, for every $b \in V$, either $\frac{\mu(A\cap E_b)}{\mu(A)} \in \{0,1\}$.

Lemma 7.28. *If E is stable and $\mu(A) > 0$ then there is a 0-good $A' \subseteq A$ with $\mu(A') > 0$.*

Proof. We construct a tree of subsets of A inductively so that when η is a sequence of length n, $\mu(A_\eta) > 0$. Let $A_{\langle\rangle} = A$. Given A_η, if A_η is not 0-good then there is some y_η so that $0 < \frac{\mu(A_\eta\cap E_{y_\eta})}{\mu(A_\eta)} < 1$. Then we may take $A_{\eta^\frown\langle 0\rangle} = A_\eta \cap E_{y_\eta}$ and $A_{\eta^\frown\langle 1\rangle} = A_\eta \setminus E_{y_\eta}$.

Then there must be some η of length $< d$ so that A_η is 0-good. For suppose not: then for each $\sigma \in \{0,1\}^d$ we can pick an $x_\sigma \in A_\sigma$ since A_σ has positive measure. Then when $\eta \sqsubset \sigma$, we have $x_\sigma \in E_{y_\eta}$ if and only if $\eta^\frown\langle 1\rangle \sqsubseteq \sigma$, contradicting d-stability. □

Lemma 7.29. *If A and B are both 0-good then $\frac{\mu((A\times B)\cap E)}{\mu(A\times B)} \in \{0,1\}$.*

Proof. Since B is 0-good, for everty $a \in A$ we have $\frac{\mu(B\cap E_a)}{\mu(B)} \in \{0,1\}$. Partition $A = A^0 \cup A^1$ where $A^0 = \{a \in A \mid \frac{\mu(B\cap E_a)}{\mu(B)} = 0\}$ and $A^1 = \{a \in A \mid \frac{\mu(B\cap E_a)}{\mu(B)} = 1\}$.

Analogously, we can define $B^0 = \{b \in B \mid \frac{\mu(A\cap E_b)}{\mu(A)} = 0\}$ and $B^1 = \{b \in B \mid \frac{\mu(A\cap E_b)}{\mu(A)} = 1\}$, and since A is 0-good, $B = B^0 \cup B^1$.

Suppose $\mu(A^0) > 0$; then we have $\frac{\mu((A^0 \times B) \cap E)}{\mu(A^0 \times B)} = 0$. In particular, this implies $\mu(B^1) = 0$, so $\frac{\mu((A \times B) \cap E)}{\mu(A \times B)} = \frac{\mu((A \times B^0) \cap E)}{\mu(A \times B^0)} = 0$.

On the other hand, if $\mu(A^0) = 0$ then $\frac{\mu((A \times B) \cap E)}{\mu(A \times B)} = \frac{\mu((A^1 \times B) \cap E)}{\mu(A^1 \times B)} = 1$. □

Theorem 7.30. *If E is stable then there is a countable partition $V = \bigcup_i A_i$ so that for for every pair A_i, A_j, $\frac{\mu((A_i \times A_j) \cap E)}{\mu(A_i \times A_j)} \in \{0, 1\}$.*

Proof. By the previous lemma, it suffices to choose the A_i so that each A_i is 0-good.

We choose the A_i inductively. Suppose that, for $j < i$, we have chosen A_i. If $\mu(V \setminus \bigcup_{j<i} A_j) = 0$ then we are done. (Technically, $\bigcup_{j<i} A_j$ might not be a partition, because there is an extra bit of measure 0; but we may add the extra bit onto A_1 and, since the extra bit has measure 0, A_1 continues to be 0-good.)

Otherwise, by Lemma 7.28, there exists a subset of $V \setminus \bigcup_{j<i} A_j$ which is 0-good. Let

$$\epsilon = \sup_{B \subseteq V \setminus \bigcup_{j<i} A_j,\ B \text{ is 0-good}} \mu(B).$$

There may be no subset of $V \setminus \bigcup_{j<i} A_j$ achieving this supremum, so we choose a 0-good $A_i \subseteq V \setminus \bigcup_{j<i} A_j$ with $\mu(A_i) > \epsilon/2$.

We continue this until we have A_i for all $i \in \mathbb{N}$ (unless the process finishes at a finite step). We claim that $\mu(V \setminus \bigcup_{i \in \mathbb{N}} A_i) = 0$. Suppose not. Note that since the A_i are disjoint subsets of V, $\lim_{i \to \infty} \mu(A_i) = 0$. By Lemma 7.28, there is a 0-good $B \subseteq V \setminus \bigcup_{i \in \mathbb{N}} A_i$. But there must be some i large enough that $\mu(B) \geq 2\mu(A_i)$, contradicting the choice of A_i.

So $\mu(V \setminus \bigcup_{i \in \mathbb{N}} A_i) = 0$. We replace A_1 with $A_1' = A_1 \cup (V \setminus \bigcup_{i \in \mathbb{N}} A_i)$; since A_1' differs from A_1 on a set of measure 0, A_1' is also 0-good, so $A_1', A_2, A_3, \ldots$ is the desired partition. □

Corollary 7.31. *If E is stable then for any $\epsilon > 0$ there is a partition $V = \bigcup_{i \leq k} B_i$ so that for every pair B_i, B_j, $\frac{\mu((B_i \times B_j) \cap E)}{\mu(B_i \times B_j)} \in [0, \epsilon) \cup (1 - \epsilon, 1]$.*

Proof. Let $V = \bigcup_i A_i$ be the countable partition into 0-good sets given by the theorem.

Choose k large enough that $\mu(\bigcup_{i>k} A_i) = \delta < \epsilon\mu(A_1)/3$. We define $B_1 = A_1 \cup \bigcup_{i>k} A_i$ and, for $1 < i \leq k$, $B_i = A_i$. Then for any pair i, j, if $1 < \min\{i, j\}$ we have $\frac{\mu((B_i \times B_j) \cap E)}{\mu(B_i \times B_j)} \in \{0, 1\}$. When one (or both) of i, j is 1, the extra part $B_1 \setminus A_1$ contributes an error of at most $(2\delta + \delta^2)\mu(A_1) < \epsilon$. □

7.8 Internal Cardinality and Pseudofinite Dimension

So far we have focused on the density structure of finite graphs and their ultraproducts—that is, on sets with positive measure. However when we have a large finite set, we can also consider behavior at different orders of magnitudes.

Definition 7.32. When $\mathcal{U}$ is an ultrafilter, the *nonstandard real numbers* are the ultraproduct $[\mathbb{R}]_{\mathcal{U}}$.

That is, a nonstandard natural number is a sequence $\langle r_n \rangle_{n\in\mathbb{N}}$ where we say $\langle r_n \rangle_{n\in\mathbb{N}}$ represents the same sequence as $\langle s_n \rangle_{n\in\mathbb{N}}$ if $\{n \mid r_n = s_n\} \in \mathcal{U}$. As always, we write $[r_n]_{\mathcal{U}}$ for the equivalence class of $\langle r_n \rangle_{n\in\mathbb{N}}$.

It is common to abbreivate the nonstandard reals as ${}^*\mathbb{R}$; this notation can be misleading, because it implies that there is a single clearly defined object called ${}^*\mathbb{R}$.

Just like finite sets have sizes which are a natural number, internal sets have "internal cardinality", which is a nonstandard natural number:

Definition 7.33. The *nonstandard natural numbers* are the ultraproduct $[\mathbb{N}]_{\mathcal{U}}$—that is, those $[r_n]_{\mathcal{U}}$ such that $\{n \mid r_n \in \mathbb{N}\} \in \mathcal{U}$.

If $S \subseteq [V_n]^k$ is internal, so $S = [S_n]_{\mathcal{U}}$, the *internal cardinality* is $|S| = [|S_n|]_{\mathcal{U}}$.

Internal cardinality interacts with internal functions in the way we expect:

Lemma 7.34. *If A and B are internal sets,*

- *there is an internal injection $f : A \to B$ if and only if $|A| \leq |B|$,*
- *there is an internal surjection $f : A \to B$ if and only if $|B| \leq |A|$,*
- *there is an internal bijection $f : A \to B$ if and only if $|A| = |B|$.*

The nonstandard real numbers have many of the properties we expect; we can add, multiply, and divide them, for instance. We can also compare them: we say $[r_n]_{\mathcal{U}} < [s_n]_{\mathcal{U}}$ if $\{n \mid r_n < s_n\} \in \mathcal{U}$.

Definition 7.35. $[r_n]_{\mathcal{U}}$ is *bounded* if there is a standard real number so that $\{n \mid |r_n| < r\} \in \mathcal{U}$.

If $[r_n]_{\mathcal{U}}$ is bounded then $st([r_n]_{\mathcal{U}})$, the *standard part of* $[r_n]_{\mathcal{U}}$ is $\lim_{n\to\mathcal{U}} r_n$.

For instance, we have $\mu_k(S) = st(\frac{|S|}{|[V_n]^k_{\mathcal{U}}|})$: from this perspective, the measure μ_k *is* the counting measure, suitably interpreted.

The nonstandard natural numbers satisfy a sort of induction principle for internal sets.

Theorem 7.36. *Let S be a set of nonstandard natural numbers such that $[0]_{n\in\mathcal{U}} \in S$ and, whenever $[k_n]_{\mathcal{U}} \in S$, $[k_n + 1]_{\mathcal{U}} \in S$. If S is internal then $S = [\mathbb{N}]_{\mathcal{U}}$.*

We certainly need to demand that S be internal; for instance, if S consisted of the constant sequences $[k]_{\mathcal{U}}$ for $k \in \mathbb{N}$, S would contain 0 and be closed under successor, but would not contain $[n]_{\mathcal{U}}$.

Proof. Since S is internal, $S = [S_n]_{\mathcal{U}}$. Suppose $\{n \mid 0 \in S_n \text{ and } S_n \neq \mathbb{N}\} \in \mathcal{U}$. Then, for each such n, there is a least $k_n \in \mathbb{N} \setminus S_n$. Since $[0]_{n\in\mathcal{U}} \in S$, $\{n \mid k_n > 0\} \in \mathcal{U}$, so we may consider $[k_n - 1]_{\mathcal{U}}$ and we have $[k_n - 1]_{\mathcal{U}} \in S$. But $[k_n]_{\mathcal{U}} \notin S$. □

The following variant is often more useful:

Corollary 7.37. *If $S \subseteq [\mathbb{N}]_{\mathcal{U}}$ is internal and non-empty then S has a least element.*

Rather than focusing on sets of positive density, we can use the internal cardinality to characterize the "dimension" of a set, relative to $[V_n]_{\mathcal{U}}$. For instance, if each V_n is a set with $|V_n| = n$, we could consider subsets $X_n \subseteq V_n$ with $|X_n| \approx \sqrt{n}$. We can say that X_n has "dimension 1/2". This makes sense when we consider that a subset of V_n with size $\approx n$ is a 1 dimensional set, a subset of V_n^2 with size $\approx n^2$ is a 2 dimensional set, and analogously, a subset of V_n^k with size roughly n^r should be an r dimensional set—even when r is a real number.

More precisely, to calculate the dimension we look at the logarithm: if $|X_n| \approx n^r$ then $\log |X_n| \approx r \log n$, so $r \approx \frac{\log |X_n|}{\log |V_n|}$.

Definition 7.38. Let $X = [X_n]_{\mathcal{U}}$ be an internal subset of $V^k = [V_n^k]_{\mathcal{U}}$. The *(coarse) pseudofinite dimension* of X is defined to be

$$st(\frac{\log |X|}{\log |V|}) = \lim_{n\to\mathcal{U}} \frac{\log |X_n|}{\log |V_n|}.$$

We write $\delta(X)$ for the coarse pseudofinite dimension of X.

This is the only pseudofinite dimension we will consider, but we note that there is a broader family of pseudofinite dimensions; within that family, this is known as the "coarse" pseudofinite dimension.

Lemma 7.39. *When $V = [V_n]_{\mathcal{U}}$,*

(1) $\delta(V^k) = k$,

(2) *if $A, B \subseteq V$ are internal then $\delta(A \cup B) = \max\{\delta(A), \delta(B)\}$,*

(3) *if $Y \subseteq V^{k+m}$ and $Z \subseteq V^k$ are internal, $\delta(Z) = r$, and for all $\vec{v} \in Z$, $\delta(Y_{\vec{v}}) \leq s$ then $\delta(\bigcup_{\vec{v}\in Z} Y_{\vec{v}}) \leq r + s$,*

(4) *if $Y \subseteq V^{k+m}$ and $Z \subseteq V^k$ are internal, $\delta(Z) = r$, and for all $\vec{v} \in Z$, $\delta(Y_{\vec{v}}) \geq s$ then $\delta(\bigcup_{\vec{v}\in Z} Y_{\vec{v}}) \geq r + s$,*

The last property is sometimes called the "fiber property": it says that if we have a collection of sets $\{Y_{\vec{v}}\}_{\vec{v}\in Z}$, the size of the union is bounded by adding the number of sets in the collection and the size of each set. (Consider that the *cardinality* of the union should be bounded by the product of the cardinalities, and the logarithmic nature of dimension converts that to a sum.)

Proof. The first part follows directly from the definition. For the second part, if $A = [A_n]_{\mathcal{U}}$ and $B = [B_n]_{\mathcal{U}}$, note that $\delta(A \cup B) = \lim_{n\to\mathcal{U}} \frac{\log|A_n \cup B_n|}{\log|V_n|}$. We certainly have $\delta(A \cup B) \geq \max\{\delta(A), \delta(B)\}$ since $\frac{\log|A_n\cup B_n|}{\log|V_n|} \geq \frac{\log|A_n|}{\log|V_n|}$ and similarly for B_n. We also have $\log|A_n \cup B_n| \leq \log 2 \max\{|A_n|, |B_n|\} = (\log\max\{|A_n|, |B_n|\}) + \log 2$, and since $\frac{\log 2}{\log|V_n|} \to 0$, we have $\delta(A \cup B) \leq \max\{\delta(A), \delta(B)\}$ as well.

For the third part, we have

$$\begin{aligned}\delta(\bigcup_{\vec{v}\in Z} Y_{\vec{v}}) &= \lim_{n\to\mathcal{U}} \frac{\log|\bigcup_{\vec{v}\in Z_n}(Y_n)_{\vec{v}}|}{\log|V_n|} \\ &\leq \lim_{n\to\mathcal{U}} \frac{\log|Z_n| \cdot \max_{\vec{v}\in Z_n}|(Y_n)_{\vec{v}}|}{\log|V_n|} \\ &\leq r + s.\end{aligned}$$

The fourth part is analogous. □

7.9 Stable Erdős-Hajnal

Coarse pseudofinite dimension is the right setting for considering the Erdős-Hajnal conjecture, which says that, for every finite graph H, any graph with no induced copies of H either has a large clique or an anti-clique.

Definition 7.40. If $G = (V, E)$ is a graph, we say $S \subseteq V$ is *E-homogeneous* if either $\binom{S}{2} \subseteq E$ or $\binom{S}{2} \cap E = \emptyset$.

That is, S is either a clique in E (every pair is an edge) or an anti-clique (no pair is an edge).

Conjecture 7.41 (Erdő–Hajnal Conjecture)**.** For every finite graph H there is a constant $d(H) > 0$ so that whenever $G = (V, E)$ is a finite graph with $T_H^{ind}(G) = \emptyset$, there is an E-homogeneous set $S \subseteq V$ with $\delta(S) \geq d(H)$.

While this is open, it is known that stable graphs—that is, graphs omitting the ladder of size d for all d—do have large homogeneous sets. Before proving this, we need the following lemma, which says that if we have a set in a graph where every vertex has very few neighbors, we can color the vertices in the set with a small number of colors so that neighbors get different colors.

Lemma 7.42. *Let $B \subseteq [V_n]_{\mathcal{U}}$ and $E \subseteq [V_n]^2_{\mathcal{U}}$ be internal sets with E symmetric, and suppose that for each $x \in B$, $\delta(\{y \in B \mid (x, y) \in E\}) = 0$. Then there is an internal set I and internal function $c : B \to I$ such that:*

- *$\delta(I) = 0$,*
- *if $b, b' \in B$ and $c(b) = c(b')$ then $(b, b') \notin E$.*

The finite version of this is not surprising: suppose that B is finite and for each $x \in B$ we have $|\{y \in B \mid (x, y) \in E\}| \leq m$. Then we expect to be able to partition B into sets $B_1, \ldots, B_{m+1}$ so that if $x, x' \in B_i$ then $(x, x') \notin E$. Indeed, we could construct such a partition greedily: place B in an order and when we decide which partition component to put the j-th element of B in, there are at most m of them which have been ruled out, so we place this element in whichever one remains. The infinite version follows by doing this in the ultraproduct.

Proof. There are really two parts to this. The first is noting that, since B is internal and $\delta(E_x) = 0$ for every $x \in B$, there is actually a supremum on the sizes of the E_x which itself has dimension 0.

Indeed, let $S \subseteq [\mathbb{N}]_{\mathcal{U}}$ be the set of nonstandard natural numbers k such that, for all $x \in B$, $|E_x| \leq k$. This is an internal set (for instance, it can be

expressed by a first-order formula using the predicate which holds of x, k when $|E_x| \leq k$) and contains every k with $\delta(k) > 0$, so has a least element k. If $\delta(k) > 0$ then $\delta(k-1) > 0$, so $k \in S$ as well; so for the least $k \in S$, we must have $\delta(k) = 0$.

Choose an internal set I with $|I| = k+1$. Now consider the set $T \subseteq [\mathbb{N}]_{\mathcal{U}}$ of natural numbers m such that there is an internal $U \subseteq B$ and $c : U \to I$ with $|U| = m$ so that if $b, b' \in U$ and $c(b) = c(b')$ then $(b, b') \notin E$. Clearly $[0] \in S$ since $|\emptyset| = [0]$ the existence of such a function for the empty set is trivial.

Consider the least $m \notin T$. If $m = |B|$ then we are done. Suppose $m < |B|$. Then $m - 1 \in T$, so we may choose $U \subseteq B$ with $|U| = m - 1$ and $c : U \to I$. Choose any $b \in B \setminus U$. Since $|E_b| < |I|$, there must be some $i \in I$ so that there is no $b' \in U$ with $(b, b') \in E$ and $c(b') = i$, so we may extend U to $U \cup \{b\}$ by adding $c(b) = i$. This contradicts the assumption that $m \notin T$. $\square$

The ultraproduct version of stable Erdős-Hajnal is the following.

Theorem 7.43. *If $[G_n]_{\mathcal{U}} = (V, E)$ is stable then there is an internal set S with $\delta(S) > 0$ so that S is E-homogeneous.*

Proof. Let $G = [G_n]_{\mathcal{U}} = (V, E)$ be d-stable. We will first show that there is a set B with $\delta(B) > 0$ such that either for every $y \in B$, $\delta(\{x \in B \mid (x, y) \in E\}) = 0$ or for every $y \in B$, $\delta(\{x \in B \mid (x, y) \notin E\}) = 0$.

We construct a tree of sets: let $A_{\langle\rangle} = V$. Suppose we have constructed A_η with $\delta(A_\eta) > 0$. If there is a $y \in A_\eta$ such that both $\delta(\{x \in A_\eta \mid (x, y) \in E\}) > 0$ and $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) > 0$ then let $y_\eta = y$, $A_{\eta^\frown\langle 0\rangle} = \{x \in A_\eta \mid (x, y) \in E\}$ and $A_{\eta^\frown\langle 1\rangle} = \{x \in A_\eta \mid (x, y) \notin E\}$.

If we can construct such a tree of size d, for each leaf σ we have $\delta(A_\sigma) > 0$, so we may choose any $x_\sigma \in A_\sigma$ and we have a tree contradicting d-stability.

So we eventually obtain an A_η so that, for every $y \in A_\eta$, either $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) = 0$ or $\delta(\{x \in A_\eta \mid (x, y) \in E\}) = 0$. Let A^-_η be those y such that $\delta(\{x \in A_\eta \mid (x, y) \in E\}) = 0$ and A^+_η be those y such that $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) = 0$, so $A_\eta = A^+_\eta \cup A^-_\eta$.

If $\delta(A^-_\eta) = 0$, we may take $B = A_\eta \setminus A^-_\eta$. Otherwise, for each $y \in A^+_\eta$, since $\delta(\{x \in A^-_\eta \mid (x, y) \in E\}) = 0$, $\delta(\{x \in A^-_\eta \mid (x, y) \in E\}) = \delta(A^-_\eta)$. Therefore $\delta(\{(x, y) \in E \mid x \in A^-_\eta \text{ and } y \in A^+_\eta\}) = \delta(A^-_\eta) + \delta(A^+_\eta) > 0$.

But for each $x \in A^-_\eta$, $\delta(\{y \in A^+_\eta \mid (x, y) \in E\}) = 0$, so $\delta(\{(x, y) \in E \mid x \in A^-_\eta \text{ and } y \in A^+_\eta\}) \leq \delta(A^-_\eta) + 0 = \delta(A^-_\eta)$. Therefore $\delta(A^+_\eta) = 0$, so we may take $B = A_\eta \setminus A^+_\eta$.

Without loss of generality, let us assume we have a B with $\delta(B) > 0$ such that for every $y \in B$, $\delta(\{x \in B \mid (x, y) \in E\}) = 0$. Then by Lemma 7.42, we

have an I with $\delta(I) = 0$ and a $c : B \to I$ so that if $c(b) = c(b')$ then $(b, b') \notin E$. Consider $Y = \{(i, b) \mid c(b) = i\}$. Since $\delta(B) = \delta(\bigcup_{i \in I} Y_i) \leq 0 + \sup_{i \in I} \delta(Y_i)$, there must be some $i \in I$ so that $\delta(Y_i) = \delta(B)$. Then Y_i is the desired anti-clique. □

As always, we can extract a finitary conclusion:

Theorem 7.44. *For every d there is a $\delta > 0$ so that whenever $G = (V, E)$ is d-stable and sufficiently large, there is an $S \subseteq V$ with $|S| \geq |V|^\delta$ so that S is E-homogeneous.*

Proof. Suppose not. Then there is a d so that for every n there is a $G_n = (V_n, E_n)$ with $\geq n$ vertices which is d-stable but every clique or anti-clique has size $< |V|^{1/n}$.

Let $G = [G_n]_{\mathcal{U}}$. Since being d-stable is a first-order property and all the G_n are d-stable, G is d-stable. Then there is an internal homogeneous set $S = [S_n]_{\mathcal{U}}$ with $\delta(S) > 0$. Therefore we may choose an n so that S_n is homogeneous and $|S_n| \geq |V|^{1/n}$, contradicting our assumption. □

7.10 Distal Cell Decompositions

We can think of stability as one way for graphs with finite VC dimension to be particularly nice. We now consider a different way that a graph of finite VC dimension might be particularly nice.

Definition 7.45. Let $E \subseteq V \times W$ be a graph. We say E has a *distal cell decomposition* if there is a k and a $C \subseteq V \times W^k$ such that whenever $B \subseteq W$ is finite with $|B| \geq 2$ and $a \in V$, there are $(b_1, \ldots, b_k) \in B^k$ so that $(a, b_1, \ldots, b_k) \in C$ and for every $b \in B$, either

- $C_{b_1,\ldots,b_k} \subseteq E_b$, or
- $C_{b_1,\ldots,b_k} \cap E_b = \emptyset$.

The sets $C_{b_1,\ldots,b_k}$ are the "cells" in the name. (This notion comes from [22], where the authors require some additional definable conditions on a distal cell decomposition.)

The idea is that $(b_1, \ldots, b_k)$ is a bounded amount of information which encodes the set $E_a \cap B$. To see that this happens, observe that if we have chosen $b_1, \ldots, b_k$ so that $(a, b_1, \ldots, b_k) \in C$, the second condition requires that, for all $b \in B$, $C_{b_1,\ldots,b_k} \subseteq E_b$ if and only if $b \in E_a$. (If $C_{b_1,\ldots,b_k} \subseteq E_b$

then, since $a \in C_{b_1,\ldots,b_k}$, we must have $a \in E_b$, so $b \in E_a$. If $C_{b_1,\ldots,b_k} \cap E_b = \emptyset$ then, since $a \in C_{b_1,\ldots,b_k}$, we have $a \notin E_b$, so $b \notin E_a$.)

For instance, suppose $V = W = [0,1]$ and $E = \{(x,y) \mid x < y\}$. The idea is that when B is a finite set, if we want to know about $E_a \cap B$, all we really need to know is which interval inside B contains a (or if a is either strictly above or strictly below all of b). Then we can take $k = 2$ and

$$C = \{(a,b_1,b_2) \mid b_1 \leq a < b_2\} \cup \{(a,b_1,b_2) \mid b_2 < b_1 \leq a\} \cup \{(a,b_1,b_1) \mid a < b_1\}.$$

The first set is the main case; the second two are encoding that a is either larger than or smaller than everything in B. (This is why we need the requirement $|B| \geq 2$—it makes sure we can do things like selectively repeat elements to encode different cases.)

Whenever $B \subseteq W$ is a finite set and $a \in V$, if $a \leq \min B$ then we take $b_1 = \min B$ and have $(a,b_1,b_1) \in C$, so $C_{b_1,b_1} = [0,b_1) \subseteq [0,b)$ for all $b \in B$. If $a \geq \max B$ then we take $b_1 = \max B$ and $b_2 \in B$ to be any other element; then $(a,b_2,b_1) \in C$ and $C_{b_2,b_1} = [b_1,1]$, which is disjoint from E_b for all $b \in B$. Otherwise, we choose $b_1 = \max B \cap [0,a]$ and $b_2 = \min B \cap (a,1]$, so $(a,b_1,b_2) \in C$ and $C_{b_1,b_2} = [b_1,b_2)$, which is contained in E_b for every $b \in B$ with $b_2 \leq b$ and disjoint from E_b for every $b \in B$ with $b \leq b_1$.

On the other hand, any graph with infinite VC dimension fails to have a distal cell decomposition.

Lemma 7.46. *If $E \subseteq V \times W$ has a distal cell decomposition then E has finite VC dimension.*

Proof. Suppose $C \subseteq V \times W^k$ is a distal cell decomposition for E. Let $B \subseteq W$ with $2^{|B|} > |B|^k$.

Observe that each choice of $b_1,\ldots,b_k \in B$ defines a subset of B, namely $S_{b_1,\ldots,b_k} = \{b \mid E_b \supseteq C_{b_1,\ldots,b_k}\}$. We will show that, for every a, $E_a \cap B$ is one of these sets; since there are not enough such sets to shatter B, it follows that E has finite VC dimension.

Take any $a \in V$. Since C is a distal cell decomposition, there must be $b_1,\ldots,b_k$ with $(a,b_1,\ldots,b_k) \in C$. If $b \in E_a \cap B$ then we have $a \in E_b \cap C_{b_1,\ldots,b_k}$, so $E_b \cap C_{b_1,\ldots,b_k}$, so we must have $E_b \supseteq C_{b_1,\ldots,b_k}$, so $b \in S_{b_1,\ldots,b_k}$. Conversely, if $b \in S_{b_1,\ldots,b_k}$ then $a \in C_{b_1,\ldots,b_k} \subseteq E_b$. Therefore $E_a \cap B = S_{b_1,\ldots,b_k}$. □

Most sets with finite VC dimension have distal cell decompositions. For a while, only one essential example was known.

Theorem 7.47 ([27])**.** *Let p be a prime and let $\mathbb{F}$ be the algebraic closure of $\mathbb{F}_p$. Let $P \subseteq \mathbb{F}^2$ be the set of points and let L be the set of lines in $\mathbb{F}^2$. Then*

the incidence relation $I \subseteq P \times L$ consisting of pairs (p, ℓ) where the point p is on line ℓ is not distally compressible.

This example is known to be stable. Two more examples have been constructed using more [20, 94] using graph theoretic methods.

The motivation for distal cell decompositions come from from the model theoretic notion of *distality* [141]. A characterization of distality given in [24] says that a theory is distal if every graph defined in the theory has a distal cell decomposition which is itself definable in the theory. The original purpose of distality was to capture the notion of NIP theories (that is, theories where every definable set has finite VC dimension) which are "purely unstable"; indeed, non-trivial theories cannot be both distal and stable. However distality is fundamentally different in character from stablity; in particular, it is not monotone in the theory—we could have a theory which is not distal because the compression schemes are not definable, but by adding additional formulas, the theory becomes distal. (And in the process, we might expand a stable theory to an unstable one because the new formulas are unstable.)

Graphs with distal cell decompositions satisfy a strong form of the Erdős–Hajnal conjecture in which we find a homogeneous set of positive *measure*—not just dimension—in a bipartite graph. We will prove this under an extra assumption, that a related set system has finite VC dimension.

Definition 7.48. When $C \subseteq V \times W^k$ is a distal cell decomposition for $E \subseteq V \times W$, we say $d \in W$ *crosses* $\vec{b}$ if both $C_{\vec{b}} \cap E_d \neq \emptyset$ and $C_{\vec{b}} \setminus E_d \neq \emptyset$.

For each $\vec{b} \in W^k$, we define $Cr(C, \vec{b})$ to be the set of $d \in W$ which cross $\vec{b}$.

The next lemma is strong version of Erdős-Hajnal for graphs with a nice enough distal cell decomposion.

Theorem 7.49. *If $G = (V, E)$ is a measurable graph with a distal cell decomposition such that the set system $(V, \{Cr(C, \vec{b})\}_{\vec{b} \in W^k})$ has finite VC dimension then there is a $\epsilon > 0$ and sets X, Y with $\mu_1(X) \geq \epsilon$ and $\mu_1(Y) \geq 1/4$ which are E-homogeneous—that is, either $X \times Y \subseteq E$ or $(X \times Y) \cap E = \emptyset$.*

Indeed, the proof will show that we can even ensure that Y has measure close to $1/2$.

Proof. Fix a chambered compression scheme $C \subseteq V \times V^k$ so that $(V, \{Cr(C, \vec{b})\}_{\vec{b} \in V^k})$ has finite VC dimension. Let ϵ be $1/c^k$ where c is large enough that there is an $1/2$-net of size $\leq c$.

Let $B \subseteq V$ be a 1/2-net with $|B| \leq c$, so for every $\vec{b} \in V^k$ (note that $\vec{b}$ need not be from B^k), if $\mu_1(Cr(C, \vec{b})) \geq 1/2$ then $B \cap Cr(C, \vec{b}) \neq \emptyset$.

For every $a \in V$, there must be some $\vec{b} \in B^k$ so that $a \in C_{\vec{b}}$ and $Cr(C, \vec{b}) \cap B = \emptyset$. So $V = \bigcup_{\vec{b} \in B^k | Cr(C,\vec{b}) \cap B = \emptyset} C_{\vec{b}}$, so there must be some $\vec{b} \in B^k$ so that $Cr(C, \vec{b}) \cap B = \emptyset$ and $\mu_1(C_{\vec{b}}) \geq 1/|B|^k \geq \epsilon$.

Fixing this $\vec{b}$, let $D_0 = \{d \mid C_{\vec{b}} \subseteq E_d\}$ and $D_1 = \{d \mid C_{\vec{b}} \cap E_d = \emptyset\}$. We have $V = D_0 \cup D_1 \cup Cr(C, \vec{b})$. Since we chose B to be an 1/2-net, we must have $\mu_1(Cr(C, \vec{b})) < 1/2$, so there is an $i \in \{0, 1\}$ with $\mu_1(D_i) > \frac{1}{2}(1 - \epsilon)$. Since $C_{\vec{b}} \times D_i$ is homogeneous regardless of the value of i, are homogeneous, we have found a homogeneous rectangle. □

As the proof suggests, by replacing the 1/2-net with a δ-net for some small δ, we could make $\mu_1(Y)$ as close to 1/2 as we like, at the cost of making ϵ smaller.

Finally we give a strengthening of regularity for distal graphs. When a graph is stable, we can arrange that E has density close to 0 or 1 on *every* rectangle. On the other hand, for distal graphs we still have to allow a small number of rectangles to have intermediate density, but we can guarantee that the good rectangles don't merely have density *close* to 0 or 1, but are actually homogeneous.

Theorem 7.50 (Regularity for Distal Graphs)**.** *For each $\epsilon > 0$ there is an n so that whenever $G = (V, E)$ is a graph with a distal cell decomposition so that $(V, \{Cr(C, \vec{b})\}_{\vec{b} \in W^k})$ has finite VC dimension, there is a partition $V = \bigcup_{i \leq k} A_i$ so that either $A_i \times A_j \subseteq E$ or $A_i \times A_j \cap E = \emptyset$ except for a set R of (i, j) so that $\mu_2(\bigcup_{(i,j) \in R} A_i \times A_j) < \epsilon$.*

Proof. We begin with the trivial partition $V = V$. Given a partition $V = \bigcup_{i \leq k} A_i$, let $R(\{A_i\})$ be the set of (i, j) so that A_i, A_j is not E-homogeneous and let $m(\{A_i\}) = \mu_2(\bigcup_{(i,j) \in R(\{A_i\})} A_i \times A_j)$.

Consider some partition $V = \bigcup_{i \leq k} A_i$. Take any pair $A_i \times A_j$ which is not E-homogeneous. The previous theorem gives us subsets $B_i \subseteq A_i$ and $B_j \subseteq A_j$ with $\mu_1(B_i) \geq \delta \mu_i(A_i)$ and $\mu_1(B_j) \geq \frac{1}{4}\mu_1(A_j)$. Consider the partition $V = \bigcup_{i \leq k+2} A'_i$ where we replace A_i with $B_i, A_i \setminus B_i$ and $A_j, A_j \setminus B_j$. Then

$$m(\{A'_i\}) \leq m(\{A_i\}) - \frac{\delta}{4}\mu_1(A_i)\mu_1(A_j).$$

Applying this repeatedly, we eventually obtain a partition with $m(\{A_i\}) < \epsilon$. □

7.11 Remarks

VC dimension is a notion which has been reinvented a very large number of times in a variety of contexts, including computational learning theory [161], model theory [139], and combinatorics [135]. In model theory, the notion of "finite VC dimension" is often called "NIP" (standing for "not the independence property"). (More precisely, one typically calls a *theory* NIP if all its formulas define sets of finite VC dimension; the development of NIP theories and VC dimension were completely separate for almost twenty years before Laskowski identified the connection [106].)

One might ask whether the VC dimension *characterizes* the rate of growth of the function of $\pi_G(n)$. In fact, it does not: there are examples where G has VC dimension d but the function $\pi_G(n)$ grows at a much slower than n^d. The infimum of those r such that $\pi_G(n) \in O(n^r)$ is called the *VC density* of G [44, 64, 117], and can be much lower than the VC dimension (and need not be an integer).

The proof we give of the existence of an ϵ-net is essentially taken from [116], which is an excellent reference for the geometric consequences of finite VC dimension. The existence of ϵ-approximations is known as the *Glivenko-Cantelli property*; the relationship between various forms and strengthenings of the Glivenko-Cantelli property and combinatorial characterizations has been extensively studied [9, 45, 147, 148].

The stronger form of regularity for graphs with finite VC dimension was first noted in [111]. Special cases of regularity for graphs defined in (usually algebraic) structures known to have finite VC dimension were further studied in [58, 59] and a more general investigation was begun in [26].

Stability was the first model theoretic property of its type to be discovered, and has been a central topic in model theory for several decades [128, 129]. The stronger regularity for stable graphs was shown in [114], and the connection between properties like stability, bounded VC dimension, and distality on the one hand, and regularity on the other, has been extensively investigated in the setting of particular (usually algebraic) structures having these properties [58, 59, 151] as well as from the analytic and model theoretic perspective [2, 25, 26, 27, 35, 114, 115]. The specific form of the result here is essentially from [39]. Many of these results include quantitative bounds of the sort we have omitted here.

The notion of pseudofinite dimension was introduced in [90] and developed in [67, 68, 72, 89].

Chapter 8

Hypergraphs

8.1 k-Graphs

We would like to generalize the results of the previous chapters to hypergraphs in which an "edge" is a k-tuple instead of a pair.

Definition 8.1. A *k-graph* on V is a set $E \subseteq \binom{V}{k}$.

So a 2-graph is exactly a graph. As with graphs, our definition of k-graph implies irreflexivity and symmetry—a k-graph is a collection of subsets of V, each of size exactly k. Naturally, we call a set $e \in E$ a "k-edge".

The definition of an ultraproduct is essentially unchanged: if, for each n, (V_n, E_n) is a k-graph then $([V_n]_{\mathcal{U}}, [E_n]_{\mathcal{U}})$ is a k-graph as well.

Definition 8.2. A measurable k-graph is a Keisler graded probability space $\{(V^k, \mathcal{B}_k, \mu_k)\}_{k\in\mathbb{N}}$ together with a symmetric set $E \in \mathcal{B}_k$. A measurable k-graph is atomless if, for every $v \in V$, $\mu_1(\{v\}) = 0$.

Again our main examples are when V is finite and when V is an ultraproduct. As before, we will often simply say "(V, E, μ_1) is a measurable k-graph".

We can define sub-k-graphs and induced sub-k-graphs analogously to graphs.

Definition 8.3. If (V, E, μ_1) is a measurable k-graph and $H = (\{w_1, \ldots, w_d\}, F)$ is a finite k-graph, we define:

- $T_H(E)$ is the set of d-tuples $(v_1, \ldots, v_d) \in V^d$ such that whenever $\{w_{i_1}, \ldots, w_{i_k}\} \in F$, $\{v_{i_1}, \ldots, v_{i_k}\} \in E$,

- $T_H^{ind}(E)$ is the set of d-tuples $(v_1, \ldots, v_d) \in V^d$ such that $\{w_{i_1}, \ldots, w_{i_k}\} \in F$ if and only if $\{v_{i_1}, \ldots, v_{i_k}\} \in E$,
- $t_H(E) = \mu_d(T_H(E))$,
- $t_H^{ind}(E) = \mu_d(T_H^{ind}(E))$.

We mostly focus on 3-graphs as the natural next step after graphs, though the ideas immediately generalize to $k > 3$.

We can consider an assortment of examples. Analogous to the complete bipartite graph, there is a complete tripartite 3-graph.

Example 8.4. A complete tripartite 3-graph is a 3-graph (V, E) where $V = V_0 \cup V_1 \cup V_2$, the three parts are pairwise disjoint, and E contains those triples with one vertex in each V_i.

There are several ways to define bipartite 3-graphs.

Example 8.5. A symmetric bipartite 3-graph is a 3-graph (V, E) where $V = V_0 \cup V_1$ and E consists of those triples with at least one vertex in each V_i.

An asymmetric bipartite 3-graph is a 3-graph (V, E) where $V = V_0 \cup V_1$ and E consists of those triples with at exactly one vertex in V_1.

We can also define a random 3-graph analogous to our definition of $\mathbf{R}_p(V)$.

Example 8.6. Define $\mathbf{R}_p^3(V)$ to be the 3-graph where, for each triple $\{x, y, z\} \in \binom{V}{3}$ we flip a coin which is heads with probability p and place $\{x, y, z\} \in E$ if the coin comes up heads.

We can verify results analogous to those we found for the random graph: in particular, for every finite $H = (W, F)$, $t_H(\mathbf{R}_p^3(V)) \approx p^{|F|}$.

One of the new features of 3-graphs that we did not see with ordinary graphs is that there are other "random-like" 3-graphs.

Definition 8.7. If E is a graph on V, define $K^{\text{odd}}(E)$ to be the 3-graph consisting of those triples $\{x, y, z\}$ so that either one or three of the pairs $\{x, y\}, \{x, z\}, \{y, z\}$ belong to E.

In particular, consider $K^{\text{odd}}(\mathbf{R}_{1/2})$. We can verify that, with high probability, this 3-graph has density $1/2$, but it cannot be fully random. Indeed, $K^{\text{odd}}(\mathbf{R}_{1/2})$ omits some induced sub-3-graphs entirely.

Definition 8.8. Let $K_{3,4}^-$ be the 3-graph on 4-vertices omitting exactly one of the 4 possible 3-edges.

Observe that, for any E, $T^{\text{ind}}_{K_{3,4}^-}(K^{\text{odd}}(E)) = \emptyset$. To see this, consider any four vertices x, y, z, w, and observe that $\{x, y, z\} \in K^{\text{odd}}(E)$ exactly when $\chi_E(x, y) + \chi_E(x, z) + \chi_E(y, z)$ is odd. For any four vertices $\{x, y, z, w\}$, if the three edges $\{x, y, z\}$, $\{x, y, w\}$, and $\{x, z, w\}$ are all are present in $K^{\text{odd}}(E)$ then can add up these three triples and we see that

$$2\chi_E(x, y) + 2\chi_E(x, z) + 2\chi_E(x, w) + \chi_E(y, z) + \chi_E(y, w) + \chi_E(z, w),$$

as the sum of three odd numbers, is also odd. But this means $\chi_E(y, z) + \chi_E(y, w) + \chi_E(z, w)$ is odd, and therefore, the fourth edge $\{y, z, w\}$ is present as well.

On the other hand, we will see that there is a class of 3-graphs that $K^{\text{odd}}(\mathbf{R}_{1/2})$ does have the "correct" number of copies of, as if it were random: indeed, we will identify $K^{\text{odd}}(\mathbf{R}_{1/2})$ as having a "partial randomness" property (analogous, in some sense, to the way that equidistribution is a weak notion of randomness for graphs).

Example 8.9. Let $\to$ be a *tournament* on V—that is, $\to \subseteq V^2$ is a relation such that for every pair $\{x, y\} \in \binom{V}{2}$, exactly one of $x \to y$ or $y \to x$ holds. Then we can define a 3-graph $\mathrm{C}(\to)$ to consist of those triples $\{x, y, z\}$ such that $\{x, y, z\}$ forms a cycle in $\to$ (in either direction—that is, either $x \to y, y \to z, z \to x$, or $x \to z, z \to y, y \to x$).

In particular, consider the case where $\to$ is the random tournament on V—that is, for each pair $\{x, y\}$, we flip a coin to determine whether $x \to y$ or $y \to x$. Then $\mathrm{C}(\to)$ has density roughly $1/4$; one way to see this is to observe that if we fix the order $x \to y$, the other two components $y \to z$ and $z \to y$ each independently have a $1/2$ chance to point the right direction to make a cycle, so the probability is $1/4$ that both do.

Again, this 3-graph will turn out to satisfy a weak notion of randomness, but cannot be fully random since, for any tournament $\to$, $T_{K_{3,4}^-}(\mathrm{C}(\to)) = \emptyset$. To see this, suppose $\{x, y, z, w\}$ are vertices and both $\{x, y, z\} \in H$ and $\{x, y, w\} \in \mathrm{C}(\to)$. Without loss of generality, assume $x \to y$, so also $y \to z$ and $y \to w$. Therefore $\{y, z, w\} \notin \mathrm{C}(\to)$. Similarly, $z \to x$ and $w \to x$, so $\{x, z, w\} \notin \mathrm{C}(\to)$.

Example 8.10. Let V be linearly ordered by $<$ and take a random 3-coloring of $\binom{V}{2}$ into red, blue, and yellow, and if $x < y < z$ then put $\{x, y, z\} \in E$ if $\{x, y\}$ is red, $\{x, z\}$ is blue, and $\{y, z\}$ is yellow.

This 3-graph has density roughly 1/27—given $x < y < z$, this is the probability that all three pairs get the right colors. Once again, we will see that this has a weak randomness property. In an important sense, this will turn out to be a "minimal" 3-graph—roughly speaking, every large 3-graph contains (not necessarily induced) sub-3-graphs which look like this.

8.2 Notation for Integrals Indexed by Graphs

In order to consider notions like $t_H(f)$ when H is a k-graph and f is a symmetric function on V^k, it will be helpful to introduce some notation for indexing integrals by tuples.

Definition 8.11. Suppose (W, F) is a finite k-graph. We write V^W for the set of functions from W to V. When $\vec{x} \in V^W$, we call $\vec{x}$ a *W-tuple*. For any $w \in W$, we write $x_w = \vec{x}(w)$. If $e \in F$ is a k-edge, we write $\vec{x}_e$ for the set $\{x_w \mid w \in e\}$.

This notation naturally extends the usual notation for tuples: we are identifying $(x_1, \ldots, x_d) \in V^d$ with the function $\vec{x}(i) = x_i$. Note that V^W is isomorphic to $V^{|W|}$—the only difference is the labeling: we think of $V^{|W|}$ containing ordered sequences $(x_1, \ldots, x_d)$, while we think of the elements of V^W as sequences whose elements are named by the vertices of W.

When V is atomless and $f : V^k \to \mathbb{R}$ is symmetric, $\vec{x} \in V^W$, and $e \in \binom{W}{k}$, the notation $f(\vec{x}_e)$ makes sense everywhere except for the set of measure 0 where $x_w = x_{w'}$ for two distinct $w, w' \in e$—taking $e = \{w_1, \ldots, w_k\}$, we identify $f(x_e)$ with $f(x_{w_1}, \ldots, x_{w_k})$, and since f is symmetric, the value of $f(x_e)$ is independent of the choice of ordering of the vertices in e.

This notation will give us a useful, compact way to express various integrals; for instance, we can write

$$t_H(f) = \int \prod_{e \in F} f(x_e)\, d\mu_{|W|}.$$

8.3 Weak Quasirandomness

For graphs, we ended up with three equivalent formulations of quasirandomness for atomless measurable graphs—$G = (V, E)$ being quasirandom was equivalent to each of the following:

- for every $H = (W, F)$, $t_H(G) = (t_{\mathrm{K}_2}(G))^{|F|}$,

- $t_{C_4}(G) = (t_{K_2}(G))^4$,
- $\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1})$ is the function constantly equal to $t_{K_2}(G)$.

We would like to find three analogous characterizations for quasirandom k-graphs.

There is a sub-σ-algebra analogous to $\mathcal{B}_{2,1}$:

Definition 8.12. Let $\mathcal{B}^0_{k,1}$ consist of sets of the form $\bigcup_{i\le r}\prod_{j\le k} X_{i,j}$ where each $X_{i,j} \in \mathcal{B}_1$

That is, $\mathcal{B}^0_{k,1}$ consists of "boxes"—subsets of V^k which are formed by products of a set on each side—and their unions. As before, $\mathcal{B}^0_{k,1}$ is an algebra.

Definition 8.13. $\mathcal{B}_{k,1} \subseteq \mathcal{B}_k$ consists of those sets $B \in \mathcal{B}_k$ such that, for every $\epsilon > 0$, there is a $B_\epsilon \in \mathcal{B}^0_{k,1}$ such that $\mu_k(B \triangle B_\epsilon) < \epsilon$.

As in Theorem 6.6, we get:

Theorem 8.14. *$\mathcal{B}_{k,1}$ is a σ-algebra.*

Definition 8.15. We say E is $\binom{k}{1}$*-quasirandom* if $\mathbb{E}(\chi_E \mid \mathcal{B}_{k,1})$ is a constant function.

Theorem 8.16. *The following 3-graphs are $\binom{3}{1}$-quasirandom:*

- $K^{\text{odd}}(\mathbf{R}_{1/2})$,
- $\mathrm{C}(\to)$ *where $\to$ is the random tournament.*

Proof. Consider a box $A \times B \times C$.

For $K^{\text{odd}}(\mathbf{R}_{1/2})$, let $E = \mathbf{R}_{1/2}$ and observe that $\mu_3(K^{\text{odd}}(E) \cap (A\times B\times C))$ is a sum of four terms like

$$\int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3$$

over the various ways to have either one or three edges present. For each of these terms, we can successively consider integrals over two variables at a

time, as in the proof of Theorem 5.4. For instance

$$\begin{aligned}
&\int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3 \\
&= \int \chi_C(z)\left[\int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\chi_A(x)\chi_B(y)\,d\mu_2(x,y)\right] d\mu_1(z) \\
&= \int \chi_C(z)\left[\frac{1}{2}\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_A(x)\chi_B(y)\,d\mu_2(x,y)\right] d\mu_1(z) \\
&= \int \chi_B(y)\left[\frac{1}{2}\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_A(x)\chi_C(z)\,d\mu_2(x,z)\right] d\mu_1(y) \\
&= \int \chi_B(y)\left[\frac{1}{4}\int (1-\chi_E(y,z))\chi_A(x)\chi_C(z)\,d\mu_2(x,z)\right] d\mu_1(y) \\
&= \int \chi_A(x)\left[\frac{1}{4}\int (1-\chi_E(y,z))\chi_B(y)\chi_C(z)\,d\mu_2(y,z)\right] d\mu_1(x) \\
&= \int \chi_A(x)\left[\frac{1}{8}\int \chi_B(y)\chi_C(z)\,d\mu_2(y,z)\right] d\mu_1(x) \\
&= \frac{1}{8}\mu_3(A\times B\times C).
\end{aligned}$$

Since this holds for each term, $\mu_3(K^{\mathrm{odd}}(E)\cap(A\times B\times C)) = \frac{1}{2}\mu_3(A\times B\times C)$. Since this holds for every rectangle, $\mathbb{E}(K^{\mathrm{odd}}(E) \mid \mathcal{B}_{3,1}) = 1/2$.

The argument for $\mathrm{C}(\rightarrow)$ is similar: $\mu_3(\mathrm{C}(\rightarrow)\cap(A\times B\times C))$ is

$$\begin{aligned}
&\int \chi_\rightarrow(x,y)\chi_\rightarrow(y,z)\chi_\rightarrow(z,y)\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3 \\
+&\int \chi_\rightarrow(y,x)\chi_\rightarrow(x,z)\chi_\rightarrow(z,x)\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3
\end{aligned}$$

and the symmetric term, and the same argument, using the fact that $\mathbb{E}(\chi_\rightarrow \mid \mathcal{B}_{2,1})$ is constantly $1/2$, shows that

$$\int \chi_\rightarrow(x,y)\chi_\rightarrow(y,z)\chi_\rightarrow(z,y)\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3 = \frac{1}{8}\mu_3(A\times B\times C)$$

and

$$\int \chi_\rightarrow(y,x)\chi_\rightarrow(x,z)\chi_\rightarrow(z,x)\chi_A(x)\chi_B(y)\chi_C(z)\,d\mu_3 = \frac{1}{8}\mu_3(A\times B\times C),$$

so

$$\mu_3(\mathrm{C}(\rightarrow)\cap(A\times B\times C)) = \frac{1}{4}\mu_3(A\times B\times C) = \mu_3(C(\rightarrow))\mu_3(A\times B\times C).$$

□

Since these 3-graphs omit some small 3-graphs entirely, they certainly do not have the correct density for all 3-graphs. However $\binom{k}{1}$-quasirandomness does imply the correct density for *linear* k-graphs.

Definition 8.17. A k-graph (W, F) is *linear* if whenever $e, e' \in F$ with $e \neq e'$, $|e \cap e'| \leq 1$.

That is, a k-graph is linear if two k-edges never share more than a single vertex in common.

Theorem 8.18. *If $G = (V, E)$ is $\binom{k}{1}$-quasirandom and $H = (W, F)$ is linear then $t_H(G) = (\mu_k(E))^{|F|}$.*

Proof. By induction on $|F|$. When $|F| = 0$, H is the empty graph on $|W|$ vertices, so $t_H(G) = 1$ by definition.

Suppose $|F| > 0$. Choose some edge $e_0 \in F$ and let $H^- = (W, F \setminus \{e_0\})$. Numbering vertices in W appropriately, we may assume $e_0 = \{w_1, \ldots, w_k\}$. For each $i \leq k$, let $F_i \subseteq F$ consist of those edges $e \in F$ with $w_i \in e$. Since H is linear, the F_i are pairwise disjoint. Let $F^- = F \setminus \bigcup_{i \leq k} F_i$, so F^- consists of those e with $e \cap e_0 = \emptyset$. Then

$$\begin{aligned} t_H(G) &= \int \prod_{e \in F} \chi_E(x_e)\, d\mu_{|W|} \\ &= \int \prod_{e \in F^-} \chi_E(x_e) \left[\int \chi_E(x_{e_0}) \prod_{i \leq k} \prod_{e \in F_i} \chi_E(x_e)\, d\mu_k(x_{e_0}) \right] d\mu_{|W|-k}. \end{aligned}$$

Observe that the inner integral looks at the product of $\chi_E(x_{e_0})$ with the function $f_{x_{W \setminus e}}(x_e) = \prod_{i \leq k} \prod_{e \in F_i} \chi_E(x_e)$ (for some fixed value of the coordinates $x_{W \setminus e}$). The function f is measurable with respect to $\mathcal{B}_{k,1}$, since it is a product of terms each of which depends on only one of the coordinates x_i. Therefore

$$\int \chi_E(x_{e_0}) \prod_{i \leq k} \prod_{e \in F_i} \chi_E(x_e)\, d\mu_k(x_e) = \mu_k(E) \int \prod_{i \leq k} \prod_{e \in F_i} \chi_E(x_e)\, d\mu_k(x_e),$$

and therefore

$$t_H(G) = \mu_k(E) t_{H^-}(G) = \mu_k(E)(\mu_k(E))^{|F|-1} = (\mu_k(E))^{|F|}$$

by the inductive hypothesis. □

Note that we cannot simply extend this to induced sub-k-graphs. (If we were to try, we would end up looking at edges *not* in the linear k-graph, which would overlap in more than one vertex.)

As we might hope, there is a single linear k-graph analogous to C_4. It is easiest to understand this as a k-partite graph. The definition is chosen exactly to make the right applications of Cauchy-Schwarz work. We can obtain the relevant k-graph from a certain sequence.

Definition 8.19. Let (V, E) be a k-partite k-graph with $V = V_1 \cup V_2 \cup \cdots \cup V_k$—that is, $E \subseteq \prod_{i \leq k} V_i$, so each k-edge contains exactly one vertex from each V_i.

We define the *i-doubling* of (V, E), written $\mathcal{D}_i(V, E)$, to be the k-partite k-graph where:

- $\mathcal{D}_i(V, E)$ is k-partite with parts V'_j where:
  - if $j = i$ then $V'_j = V_i$,
  - if $j \neq i$ then $V'_j = \{0, 1\} \times V_i$,
- if $\{v_1, \ldots, v_k\} \in E$ with each $v_j \in V_j$ and $b \in \{0, 1\}$ then

$$\{(b, v_1), (b, v_2), \ldots, v_i, \ldots, (b, v_k)\}$$

  is an edge in $\mathcal{D}_i(V, E)$.

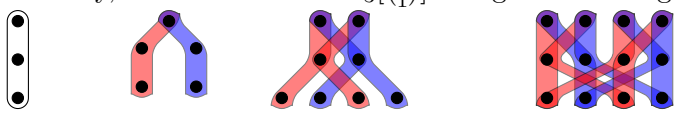
That is, the i-doubling makes two copies of each vertex *not* in the i-th part.

For example, consider beginning with the simple 2-graph consisting of a single edge, which we view as a bipartite graph with $V_1 = \{x\}$ and $V_2 = \{y\}$. The 1-doubling is the V graph: we double the vertex y, so $V'_2 = \{(0, y), (1, y)\}$, with both vertices adjacent to x. If we 2-double V, we get C_4: we now make two copies of x, so $V''_1 = \{(0, x), (1, x)\}$, $V''_2 = \{(0, y), (1, y)\}$, and all four edges between the two parts are present.

Definition 8.20. $M_k[\binom{k}{1}]$ is the k-graph $\mathcal{D}_k(\mathcal{D}_{k-1}(\cdots(\mathcal{D}_1(G))))$ where G is the k-partite k-graph consisting of a single k-edge.

In particular, $M_2[\binom{2}{1}]$ is exactly C_4.

Similarly, we can consider $M_3[\binom{3}{1}]$ being built in stages:



G $\mathcal{D}_1(G)$ $\mathcal{D}_2(\mathcal{D}_1(G))$ $\mathcal{D}_3(\mathcal{D}_2(\mathcal{D}_1(G)))$

Lemma 8.21. *If G is a linear k-partite k-graph then $\mathcal{D}_i(G)$ is linear.*

Proof. Suppose e, e' are distinct edges of $\mathcal{D}_i(G)$. Then they have the form $\{(b, v_1), \dots, v_i, \dots, (b, v_k)\}$ and $\{(b', v_1'), \dots, v_i', (b', v_k')\}$ for some edges $\{v_1, \dots, v_k\}$ and $\{v_1', \dots, v_k'\}$ of G. If e and e' share more than one vertex then we have $v_i' = v_i$ and $v_j' = v_j$ for two distinct indices $i \neq j$, contradicting the linearity of G. □

Corollary 8.22. *$M_k[\binom{k}{1}]$ is linear.*

An important consequence of the doubling construction is the following Cauchy-Schwarz-like inequality.

Lemma 8.23. *For any graph G, $|t_G(f)|^2 \leq t_{D_i(G)}(f)$.*

Proof. Let $G = (\bigcup_{i \leq k} V_i, E)$. We unfold the definitions and apply Cauchy-Schwarz.

$$\begin{aligned}
|t_G(f)|^2 &= \left(\int \prod_{\{v_1,\dots,v_k\} \in E} f(x_{v_1}, \dots, x_{v_k})\, d\mu_{\bigcup_{j \leq k} V_i} \right)^2 \\
&= \left(\iint \prod_{\{v_1,\dots,v_k\} \in E} f(x_{v_1}, \dots, x_{v_k})\, d\mu_{\bigcup_{j \neq i} V_j} d\mu_{V_i} \right)^2 \\
&= \int \left(\int \prod_{\{v_1,\dots,v_k\} \in E} f(x_{v_1}, \dots, x_{v_k})\, d\mu_{\bigcup_{j \neq i} V_j} \right)^2 d\mu_{V_i} \\
&= \int \int \prod_{\{v_1,\dots,v_k\} \in E} f(x_{v_1}, \dots, x_{v_k})\, d\mu_{\bigcup_{j \neq i} V_j \times \{0\}} \\
&\quad \cdot \int \prod_{\{v_1,\dots,v_k\} \in E} f(x_{v_1}, \dots, x_{v_k})\, d\mu_{\bigcup_{j \neq i} V_j \times \{1\}} d\mu_{V_i} \\
&= t_{D_i(G)}(f).
\end{aligned}$$

□

Corollary 8.24. $|\int f\, d\mu|^{2^k} \leq t_{M_k[\binom{k}{1}]}(f)$.

Proof. Recalling that when G is the k-partite k-graph consisting of a single k-edge, $\int f\, d\mu = t_G(f)$, this follows from the previous lemma applied to the construction of $M_k[\binom{k}{1}]$. □

This suggests, correctly, that we could think of $t_{M_k[\binom{k}{1}]}(f)$ as representing a norm, similar to the way $||f||_{U^2} = (t_{C_r}(f))^{1/4}$ represents a norm. We can even define a corresponding "inner product" with 2^k positions so that $t_{M_k[\binom{k}{1}]}(f) = \langle \{f_\sigma\}_{\sigma \in \{0,1\}^{\{1,\dots,k\}}} \rangle_{M_k[\binom{k}{1}]}$.

We can give a direct construction of $M_k[\binom{k}{1}]$ as well.

Lemma 8.25. *$M_k[\binom{k}{1}]$ is isomorphic to the k-graph (V, E) where:*

- *$V = V_1 \cup \cdots \cup V_k$ where each V_i contains 2^{k-1} vertices labeled by functions $\upsilon : (\{1, \dots, k\} \setminus \{i\}) \to \{0, 1\}$,*
- *E consists of 2^k edges, one for each function $\sigma : \{1, \dots, k\} \to \{0, 1\}$, where the edge e_σ consists of exactly the vertex labeled $\sigma \restriction (\{1, \dots, k\} \setminus \{i\})$ from each part.*

Theorem 8.26. *For a k-graph $G = (V, E)$, the following are equivalent:*

(1) *G is $\binom{k}{1}$-quasirandom,*

(2) *for every linear k-graph $H = (W, F)$, $t_H(G) = (\mu_k(E))^{|F|}$,*

(3) *$t_{M_k[\binom{k}{1}]}(E) = (\mu_k(E))^{2^k}$.*

$1 \Rightarrow 2$ is Theorem 8.18 and $2 \Rightarrow 3$ follows since $M_k[\binom{k}{1}]$ is linear, so all that remains is $3 \Rightarrow 2$. We will prove something slightly weaker, namely that for *functions* f, $\mathbb{E}(f \mid \mathcal{B}_{k,1}) = 0$ is equivalent to $t_{M_k[\binom{k}{1}]}(f) = 0$. In particular, if we apply this to the function $f = \chi_E - \mu(E)$, this means that showing $t_{M_k[\binom{k}{1}]}(\chi_E - \mu(E)) = 0$ suffices to show that $G = (V, E)$ is $\binom{k}{1}$-quasirandom. (Going further would require a bit more development of the seminorm perspective on $t_{M_k[\binom{k}{1}]}$ than we need.)

The key step is showing that for any function f, the number of weighted copies of $M_k[\binom{k}{1}]$ is monotonic if we restrict one coordinate. This is a generalization of Lemma 5.16, which showed the analogous statement for C_4 (in terms of the U_2 norm).

Lemma 8.27. *For any $f \in L^\infty(\mu_k)$ and any $B \subseteq V$,*

$$t_{M_k[\binom{k}{1}]}(f(x_1, \dots, x_k)\chi_B(x_i)) \leq t_{M_k[\binom{k}{1}]}(f).$$

This is trivial if f is non-negative (for instance, $f = \chi_E$), but the main case we will want to apply it to is $f = \chi_E - \mu(E)$, which can have "negative" copies of $M_k[\binom{k}{1}]$, so the monotonicity is meaningful.

Proof. We will abbreviate $f(x_1,\dots,x_k)\chi_B(x_i)$ by $f\chi_B$, remembering that χ_B is applied to the i-th coordinate. We have $f = (f\chi_B) + (f\chi_{\overline{B}})$ (where $\overline{B}$ is the complement of B, so $f\chi_{\overline{B}} = f(x_1,\dots,x_k)(1-\chi_B(x_i)))$.

We view $t_{M_k[\binom{k}{1}]}(f)$ is an integral of a big product indexed by the edges of $M_k[\binom{k}{1}]$. Given $\sigma : \{1,\dots,k\} \to \{0,1\}$, let us abbreviate $\sigma_i = \sigma \restriction (\{1,\dots,k\} \setminus \{i\})$. Then the edges of $M_k[\binom{k}{1}]$ have the form $(\sigma_1,\dots,\sigma_k)$ for a single σ, so we can write the integral as

$$\begin{aligned} t_{M_k[\binom{k}{1}]}(f) &= \int \prod_{\sigma\in\{0,1\}^{\{1,\dots,k\}}} f(x_{\sigma_1},\dots,x_{\sigma_k})\,d\mu \\ &= \int \prod_{\sigma\in\{0,1\}^{\{1,\dots,k\}}} (f(x_{\sigma_1},\dots,x_{\sigma_k})\chi_B(x_{\sigma_i}) + f(x_{\sigma_1},\dots,x_{\sigma_k})\chi_{\overline{B}}(x_{\sigma_i}))\,d\mu. \end{aligned}$$

We can expand out the product so that this is equal to a sum of 2^{2^k} integrals indexed by functions $\tau : \{0,1\}^{\{1,\dots,k\}} \to \{B,\overline{B}\}$: we have

$$t_{M_k[\binom{k}{1}]}(f) = \sum_{\tau\in\{B,\overline{B}\}^{\{0,1\}^{\{1,\dots,k\}}}} \int \prod_{\sigma\in\{0,1\}^{\{1,\dots,k\}}} f(x_{\sigma_1},\dots,x_{\sigma_k})\chi_{\tau(\sigma)}(x_{\sigma_i})\,d\mu.$$

Note that the term where $\tau(\sigma) = B$ for all B is precisely $t_{M_k[\binom{k}{1}]}(f\chi_B)$, so it suffices to prove that all other terms are non-negative.

Let us consider the term corresponding to some particular function τ. For any $\sigma \in \{0,1\}^{\{1,\dots,k\}}$, let $\overline{\sigma} \in \{0,1\}^{\{1,\dots,k\}}$ be given by

$$\overline{\sigma}(j) = \begin{cases} \sigma(j) & \text{if } j \neq i \\ 1-\sigma(i) & \text{if } j = i \end{cases}.$$

That is, $\overline{\sigma}$ flips the i-th coordinate and leaves the remaining coordinate alone. Suppose there is a σ so that $\tau(\sigma) \neq \tau(\overline{\sigma})$. Then the product is constantly 0, because $x_{\sigma_i} = x_{\overline{\sigma}_i}$, so for every x_{σ_i}, either $\chi_{\tau(\sigma)}(x_{\sigma_i}) = 0$ or $\chi_{\tau(\overline{\sigma})}(x_{\overline{\sigma}_i}) = 0$.

So consider the remaining terms, where τ has the property that $\tau(\sigma) = \tau(\overline{\sigma})$ for all σ. Then we may view the product as a square. Given $\sigma : (\{1,\dots,k\} \setminus \{i\}) \to \{0,1\}$ and $b \in \{0,1\}$, let us write σb for the element of $\{0,1\}^{\{1,\dots,k\}}$ with $\sigma b(j) = j$ for $j \neq i$ and $\sigma b(i) = b$. Note that our

assumption is exactly that $\tau(\sigma 0) = \tau(\sigma 1)$ for each σ. Then

$$\begin{aligned}
&\int \prod_{\sigma \in \{0,1\}^{\{1,\dots,k\}}} f(x_{\sigma_1}, \dots, x_{\sigma_k}) \chi_{\tau(\sigma)}(x_{\sigma_i})\, d\mu \\
&= \int \prod_{\sigma \in \{0,1\}^{\{1,\dots,k\}\setminus\{i\}}} f(x_{\sigma 0_1}, \dots, x_{\sigma 0_k}) \chi_{\tau(\sigma 0)}(x_\sigma) f(x_{\sigma 1_1}, \dots, x_{\sigma 1_k}) \chi_{\tau(\sigma 1)}(x_\sigma)\, d\mu \\
&= \int \left(\prod_{\sigma \in \{0,1\}^{\{1,\dots,k\}\setminus\{i\}}} f(x_{\sigma 0_1}, \dots, x_{\sigma 1_k}) \chi_{\tau(\sigma 0)}(x_\sigma) \right)^2 d\mu.
\end{aligned}$$

In particular, this is an integral of a square, and therefore non-negative.

So we have $t_{M_k[\binom{k}{1}]}(f) = t_{M_k[\binom{k}{1}]}(f\chi_B) + \cdots$, and since all the terms in $\cdots$ are non-negative, $t_{M_k[\binom{k}{1}]}(f) \geq t_{M_k[\binom{k}{1}]}(f\chi_B)$. □

Theorem 8.28. *If* $t_{M_k[\binom{k}{1}]}(f) = 0$ *then* $\mathbb{E}(f \mid \mathcal{B}_{k,1}) = 0$.

Proof. If $\mathbb{E}(f \mid \mathcal{B}_{k,1}) \neq 0$ then there must exist sets $A_i \subseteq V$ so that $|\int f(x_1, \dots, x_l) \prod_i \chi_{A_i}(x_i)\, d\mu_k| > 0$.

By Corollary 8.24, we have $t_{M_k[\binom{k}{1}]}(f \prod_i A_i) > 0$, and then the previous lemma applied k times gives $t_{M_k[\binom{k}{1}]}(f) > 0$. □

This is actually an equivalence: the converse follows by the argument of Theorem 8.18, applied to a general function instead of a graph, but we will not spell out the details because we do not need it.

8.4 Cylinder Intersection Sets

In order to have the property that $t_H(E) = (\mu_k(E))^{|F|}$ for all k-graphs $H = (W, F)$ we need a stronger notion of quasirandomness, which corresponds to having $\mathbb{E}(\chi_E \mid \mathcal{B})$ be constant for a larger σ-algebra than $\mathcal{B}_{k,1}$.

The 3-graph $K^{\text{odd}}(\mathbf{R}_{1/2})$ provides some motivation: this 3-graph is $\binom{3}{1}$-quasirandom, because it is based on $\mathbf{R}_{1/2}$, which is itself quasirandom. But the properties of $K^{\text{odd}}(\mathbf{R}_{1/2})$ are entirely determined by $\mathbf{R}_{1/2}$ itself—that is, there is a set of *pairs* which entirely determines which *triples* belong to $K^{\text{odd}}(\mathbf{R}_{1/2})$.

This suggests a stronger notion of randomness: that a quasirandom 3-graph should be orthogonal, not only to boxes, but to sets "formed from sets of pairs"—that is, sets of the form

$$\{(x, y, z) \mid (x, y) \in A \text{ and } (x, z) \in B \text{ and } (y, z) \in C\}.$$

We will call this a $\binom{3}{2}$-cylinder intersection set: it is a set of triples described as an intersection of sets defined by pairs:

$$\{(x,y,z) \mid (x,y) \in A\} \cap \{(x,y,z) \mid (x,z) \in B\} \cap \{(x,y,z) \mid (y,z)\}.$$

Definition 8.29. Let $m < k$ and suppose that, for each $s \in \binom{k}{m}$, $B_s \in \mathcal{B}_m$. We define the $\binom{k}{m}$-*cylinder intersection set*

$$K(\{B_s\}_{s\in\binom{k}{m}}) = \{\vec{x} \in V^{[1,\dots,k]} \mid \forall s \in \binom{k}{m} x_s \in B_s\}.$$

Note that

$$K(\{B_s\}_{s\in\binom{k}{m}}) = \bigcap_{s\in\binom{k}{m}} \{\vec{x} \in V^{[1,\dots,m]} \mid x_s \in B_s\}.$$

The individual sets $\{\vec{x} \in V^{[1,\dots,k]} \mid x_s \in B_s\}$ are *cylinder sets* (they have the form $B_s \times V^{[1,\dots,k]\setminus s}$), and a cylinder intersection set is precisely an intersection of cylinder sets.

A $\binom{k}{1}$-cylinder intersection set is exactly a box, but a $\binom{3}{2}$-cylinder intersection set like

$$\{(x,y,z) \mid (x,y) \in A \text{ and } (x,z) \in B \text{ and } (y,z) \in C\}$$

cannot usually be represented as a product.

The cylinder intersection sets give us a natural family of sub-σ-algebras of $\mathcal{B}_k$ generalizing $\mathcal{B}_{k,1}$.

Definition 8.30. $\mathcal{B}^0_{k,m}$ consists of all finite unions of $\binom{k}{m}$-cylinder intersection sets where each $B_s \in \mathcal{B}_s$.

$\mathcal{B}_{k,m} \subseteq \mathcal{B}_k$ consists of those $B \in \mathcal{B}_k$ such that, for every $\epsilon > 0$, there is a $B_\epsilon \in \mathcal{B}^0_{k,m}$ such that $\mu_k(B \mathbin{\triangle} B_\epsilon) < \epsilon$.

Once again we have

Theorem 8.31. *$\mathcal{B}_{k,m}$ is a σ-algebra.*

Definition 8.32. We say E is $\binom{k}{m}$-quasirandom if $\mathbb{E}(\chi_E \mid \mathcal{B}_{k,m})$ is a constant function.

We notice that $K^{odd}(\mathbf{R}_{1/2})$ and $C(\to)$ are *not* $\binom{3}{2}$-quasirandom—indeed, quite the opposite, they belong to $\mathcal{B}_{3,2}$. $K^{odd}(\mathbf{R}_{1/2})$ is a union of cylinder

intersection sets formed from $\mathbf{R}_{1,2}$ and $V^2 \setminus \mathbf{R}_{1,2}$ and $C(\to)$ is a union of cylinder intersection sets formed from $\to$ and $V^2 \setminus \to$.

On the other hand, $\mathbf{R}^3_p(V)$ *is* $\binom{3}{2}$-quasirandom: by the same Hoeffding inequality arguments we have used before, for any cylinder intersection set C, $\mu_3(\mathbf{R}^3_p(V) \cap C) = p\mu_3(C)$.

For each k, we now have a hierarchy of randomness notions: $\binom{k}{1}$-quasirandomness is weaker than $\binom{k}{2}$-quasirandomness, which (when $k > 3$) is weaker than $\binom{k}{3}$-quasirandomness, and so on. Although we will not prove this, the hierarchy is strict—for each $m < k-1$, there are k-graphs which are $\binom{k}{m}$-quasirandom but not $\binom{k}{m+1}$-quasirandom. The idea of the proof has been previewed by $K^{\text{odd}}(\mathbf{R}_{1/2})$: choose a random set of $m+1$-tuples and construct a k-graph from it.

The natural generalization of linear is to consider:

Definition 8.33. A k-graph $H = (W, F)$ has *m-bounded intersections* if whenever $e, e' \in F$ with $e \neq e'$, $|e \cap e'| \leq m$.

So 1-bounded intersections is the same as linear, while all k-graphs have $(k-1)$-bounded intersections.

Theorem 8.34. *Let $H = (W, F)$ have m-bounded intersections, and suppose that, for each $e \in F$, we have a measurable function $f_e : V^e \to [-1, 1]$. Then*

$$\int \prod_{e \in F} f_e(\vec{x}_e)\, d\mu_k(\vec{x}) = \int \prod_{e \in F} \mathbb{E}(f_e \mid \mathcal{B}_{k,m})(\vec{x}_e)\, d\mu_k(\vec{x}).$$

We are mostly interested in this when $f_e = \chi_E$ for all e, which gives the following corollary.

Corollary 8.35. *If $G = (V, E)$ is $\binom{k}{m}$-quasirandom and $H = (W, F)$ has m-bounded intersections then $t_H(G) = (\mu_k(E))^{|F|}$.*

Proof. Take $f_e = \chi_E$ for all $e \in F$. Since E is $\binom{k}{m}$-quasirandom, $\mathbb{E}(f_e \mid \mathcal{B}_{k,m})$ is the function constantly equal to $\mu_k(E)$, so $\int \prod_{e \in F} \mathbb{E}(f_e \mid \mathcal{B}_{k,m})(\vec{x}_e)\, d\mu_k(\vec{x})$ is exactly $(\mu_k(E))^{|F|}$. □

This implies that $\binom{k}{k-1}$-quasirandomness is the strongest notion of quasirandomness we should expect: a $\binom{k}{k-1}$-quasirandom k-graph correctly counts all possible k-subgraphs.

We can now prove Theorem 8.34.

Proof. Let F_0 be those $e \in F$ such that f_e is not $\mathcal{B}_{k,m}$-measurable. We proceed by induction on $|F_0|$. If $|F_0| = 0$ then the claim is trivial because $B_e = \mathbb{E}(\chi_{B_e} \mid \mathcal{B}_{k,m})$ for all $e \in F$.

Suppose $|F_0| > 0$. Choose some edge $e_0 \in F_0$. Numbering vertices in W appropriately, we may assume $e_0 = \{w_1, \dots, w_k\}$. For each $s \in \binom{k}{m}$, let $F'_s \subseteq F$ consist of those edges $e \in F$ with $e_0 \cap e \subseteq s$. Since $|e_0 \cap e|$ could be less than m, some e might belong to multiple F'_s. So choose $F_s \subseteq F'_s$ arbitrarily so that the F_s form a partition of F. Then

$$\int \prod_{e \in F} f_e(x_e)\, d\mu_{|W|} = \int \int f_{e_0}(x_{e_0}) \prod_{s \in \binom{k}{m}} \prod_{e \in F_s} f_e(x_e)\, d\mu_k(x_{e_0})\, d\mu_{|W|-k}.$$

Observe that the inner integral looks at the product of $f_{e_0}(x_{e_0})$ with the function $g_{x_{W \setminus e_0}}(x_{e_0}) = \prod_{s \in \binom{k}{m}} \prod_{e \in F_s} f_e(x_e)$ (for some fixed value of the coordinates $x_{W \setminus e_0}$). Observe that f is measurable with respect to $\mathcal{B}_{k,m}$, since it is a product of terms each of which depends only on the coordinates x_s. Therefore

$$\begin{aligned} &\int \int f_{e_0}(x_{e_0}) \prod_{s \in \binom{k}{m}} \prod_{e \in F_s} f_e(x_e)\, d\mu_k(x_{e_0})\, d\mu_{|W|-k} \\ &= \int \mathbb{E}(f_{e_0} \mid \mathcal{B}_{k,m}) \prod_{s \in \binom{k}{m}} \prod_{e \in F_s} f_e(x_e)\, d\mu_k(x_e), \end{aligned}$$

and then we may apply the inductive hypothesis to this to see that it is equal to $\int \prod_{e \in F} \mathbb{E}(f_e \mid \mathcal{B}_{k,m})(\vec{x}_e)\, d\mu_k(\vec{x})$. □

We could go on from here to develop a fuller characterization of $\binom{k}{m}$-quasirandomness, including identifying canonical k-graphs $M_k[\binom{k}{m}]$. The ideas are largely the same as for the linear case: in place of doubling operations $\mathcal{D}_i$, which double all coordinates other than i, we need the doubling operation $\mathcal{D}_I$, which double all coordinates in $\{1, \dots, k\} \setminus I$. The arguments are then quite similar to those in Section 8.3.

8.5 Hypergraph Removal

To prove Szemerédi's Theorem, we need to prove an extension of graph removal called, naturally enough, hypergraph removal. (For Szemerédi's Theorem, we only really need to remove one specific hypergraph for each k, the clique with all k-edges on $k+1$ vertices, but it is no harder to prove the general result.)

We will build on the topological argument we gave before: we will identify a notion of a "point of density", show that almost all k-tuples are points of density, and that if we have a single copy of H whose k-edges are points of density then we actually have many copies. Then we can prove removal simply by deleting all k-edges which are not points of density.

Our definition of a point of density, however, is more complicated. Instead of thinking about a partition of V, we want to think about partitions of V^m for various choices of m.

Again, we want to pass to separable sub-algebras of the $\mathcal{B}_m$ which are still big enough to make everything we care about measurable.

Lemma 8.36. *If $\{E_1, \ldots, E_j, \ldots\}$ is a countable subset of $\mathcal{B}_k$ then there is a sequence of separable $\mathcal{B}_i^- \subseteq \mathcal{B}_i$ for each $i \leq k$ such that:*

- *each $E_j \in \mathcal{B}_k^-$,*
- *whenever $1 < i \leq k$ and $B \in \mathcal{B}_i^-$, $\mathbb{E}(\chi_B \mid \mathcal{B}_{i,i-1})$ is $(\mathcal{B}_{i-1}^-)^{\binom{i}{i-1}}$-measurable.*

Proof. By induction on k. When $k = 1$, we take $\mathcal{B}_1^-$ to be the σ-algebra generated by the E_j.

When $k > 1$, we use the argument of Lemma 6.38. For each j, take the countably many level sets of the form $\{(x, y) \mid \mathbb{E}(\chi_{E_j} \mid \mathcal{B}_{k,k-1})(x, y) > q\}$. For each of these level sets and each n, there is a finite set $\mathcal{S}_{j,q,n} \subseteq \mathcal{B}_{k-1}$ such that $\{(x, y) \mid \mathbb{E}(\chi_{E_j} \mid \mathcal{B}_{k,k-1})(x, y) > q\}$ is approximated to within $1/n$ by cylinder intersection sets from $\mathcal{S}_{j,q,n}$. Then we may take $\mathcal{B}_k^-$ to be the σ-algebra generated by the E_j and, for $i < k$, we apply the inductive hypothesis to $\bigcup \mathcal{S}_{j,q,n}$ to obtain $\mathcal{B}_i^-$ for $i < k$. □

Next, we need to choose our neighborhoods—that is, a collection of finer and finer partitions. The new complication is that it is not enough to choose neighborhoods around individual points; we also need neighborhoods around pairs, and triples, and in general m-tuples for all $m < k$.

So, for every $m < k$, we fix a sequence of neighborhoods $\mathcal{N}_m^j = \{N_{m,1}^j, \ldots, N_{m,k_{m,j}}^j\}$ so that each $\mathcal{N}_m^j$ is a partition of V^m such that:

- when $i < j$, $\mathcal{N}_m^j$ refines $\mathcal{N}_m^i$,
- $\lim_{j\to\infty} \max_{u \leq k_{m,j}} \mu(N_{m,u}^j) = 0$,
- $\bigcup_j \mathcal{N}_m^j$ generates $\mathcal{B}_m^-$.

When $\vec{x} \in V^m$, we can talk about the j-th neighborhood of $\vec{x}$: $\mathcal{N}_m^j(\vec{x})$ is the unique $N_{m,u}^j$ such that $\vec{x} \in N_{m,u}^j$. The properties above say that these neighborhoods have the properties we might expect: as j increases, the neighborhood around $\vec{x}$ gets smaller and decreases to have measure 0. The final property promises that our partitions are fine enough, because the sets in our partitions collectively generate a big enough σ-algebra.

More generally, if $\vec{x} \in V^d$ with $m \leq d$, we could lift the notion of a neighborhood by asking which cylinder intersection set contains all the m-tuples from $\vec{x}$.

Definition 8.37. For any $\vec{x} \in V^d$ and $m \leq d$, we write $\mathcal{N}_m^j(\vec{x})$ for the unique $\binom{d}{m}$-cylinder intersection set $K(\{N_{m,u_s}^j\}_{s \in \binom{d}{m}})$ containing $\vec{x}$.

When $m = 1$, this cylinder intersection set is simply a box, and this matches the notation we used for graphs. Note that it could be that some of these cylinder intersection sets are non-empty but have measure 0. (This doesn't happen when $m = 1$, because a product of positive measure sets has positive measure; but one of the new complications with cylinder intersection sets is that positive measure cylinders could miss each other entirely, or could intersect in a set of measure 0.) We could define our way around this, but it will be easier to just remember that it might occur. This will not matter much, because there are only countably many cylinder intersection sets coming from our neighborhoods, so throwing away the points which belong to any of these bad cylinder intersection sets still only loses a set of measure 0.

Definition 8.38. $\vec{x} \in V^d$ *has proper neighborhoods* if, letting $m = \min\{d, k-1\}$, for any $j_1, \ldots, j_m$, $\bigcap_{i \leq m} \mathcal{N}_i^{j_i}(\vec{x})$ has positive measure.

Lemma 8.39. *The set of points with proper neighborhoods has measure* 1.

Proof. If $\vec{x}$ does not have a proper neighborhood, there are $j_1, \ldots, j_m$ so that $\vec{x} \in \bigcap_{i \leq m} \mathcal{N}_i^{j_i}(\vec{x})$ and $\mu_d(\bigcap_{i \leq m} \mathcal{N}_i^{j_i}(\vec{x})) = 0$. Therefore the set of points without proper neighborhoods are contained in a countable union of sets of measure 0. □

We again want to "smooth out" what happens at a point by taking the limit of averages over its neighborhood. The main new complication is that we have neighborhoods at different levels, and we need to take the limits in the right order. For instance, when $k = 3$, so we have both the partitions $\mathcal{N}_1^j$ of V and the partitions $\mathcal{N}_2^j$ of V^2, the correct notion of a "small neighborhood around $\vec{x}$" is

$$\mathcal{N}_2^{j_2}(\vec{x}) \cap \mathcal{N}_1^{j_1}(\vec{x})$$

where $j_2 \ll j_1$. That is, we want to consider a small neighborhood of the pair, and then a much smaller neighborhood made of the product of the individual vertices.

That idea becomes the limit in the following definition.

Definition 8.40. When $1 \leq d < r$, $B \in \mathcal{B}_r$, and any $\vec{x} \in V^r$ with proper neighborhoods, we define

$$\widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x}) = \mathbb{E}(\chi_B \mid \bigcup_{m \leq d} \mathcal{N}_m^{j_m})(\vec{x})$$

and set

$$\widetilde{B}_d = \lim_{j_d \to \infty} \cdots \lim_{j_1 \to \infty} \tilde{B}_d^{j_d,\ldots,j_1}.$$

We could equivalently define $\widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x})$ by

$$\widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x}) = \frac{\mu(B \cap \bigcap_{m \leq d} \mathcal{N}_m^{j_m}(\vec{x}))}{\mu(\bigcap_{m \leq d} \mathcal{N}_m^{j_m}(\vec{x}))}.$$

Definition 8.41. $\vec{x} \in V^r$ is a *point of density* of $B \in \mathcal{B}_r$ if $\vec{x}$ has proper neighborhoods and:

- $\widetilde{B}_d(\vec{x})$ exists,
- for each $m \leq d$, each $j \in \mathbb{N}$, and each $s \in \binom{r}{m}$, letting $N = \mathcal{N}_m^j(x_s)$, x_s is a point of density of N with $\widetilde{N}_{m-1}(x_s) > 0$,
- $$\lim_{j_d \to \infty} \cdots \lim_{j_1 \to \infty} \frac{\int_{\bigcap_{m \leq d} \mathcal{N}_m^{j_m}(\vec{x})} \left| \widetilde{B}_d(\vec{x}) - \widetilde{B}_d(\vec{y}) \right| \, d\mu_d(\vec{y})}{\mu_d(\bigcap_{m \leq d} \mathcal{N}_m^{j_m}(\vec{x}))} = 0.$$

The last condition is the main one, and the analog of what we needed for graphs: it says that $\vec{x}$ is surrounded by points which behave similarly to $\vec{x}$. We also need the second clause, which promise that our point doesn't lie at a badly behaved boundary of one of the lower arity neighborhoods.

Theorem 8.42. *$\widetilde{B}_d$ is a representative of $\mathbb{E}(\chi_B \mid \mathcal{B}_{r,d})$, almost every $\vec{x}$ is a point of density, and almost every $\vec{x} \in B$ has $\widetilde{B}_d(\vec{x}) > 0$.*

Proof. The proof is much like the analogous argument for graphs, where we only needed $d = 1$. We proceed by induction on d.

For each j, let us write $\widetilde{B}_d^j = \lim_{j_{d-1} \to \infty} \cdots \lim_{j_1 \to \infty} \widetilde{B}_d^{j,j_{d-1},\ldots,j_1}$. We first show that $\widetilde{B}_d^j$ converges to $\mathbb{E}(\chi_B \mid \mathcal{B}_{r,d}^-)$ in the L^2 norm. Let $\mathcal{K}_j$ be the

σ-algebra generated by $\binom{r}{m}$-cylinder intersection sets with $m \leq d$ where the sets come from $\mathcal{N}_d^j \cup \bigcup_{m<d} \mathcal{B}_m^-$. Then, using the inductive hypothesis, $\widetilde{B}_d^j$ is exactly $\mathbb{E}(\chi_B \mid \mathcal{K}_j)$. Since $\bigcup_j \mathcal{K}_j$ generates $\mathcal{B}_{r,d}^-$, we immediately have that $\mathcal{B}_d^j$ converges to $\mathbb{E}(\chi_B \mid \mathcal{B}_{r,d}^-)$ in the L^2 norm.

Next we show that $\widetilde{B}_d$ is a representative of $\mathbb{E}(\chi_B \mid \mathcal{B}_{r,d}^-)$. Let g be any representative of $\mathbb{E}(\chi_B \mid \mathcal{B}_{r,d}^-)$. For any $\epsilon > 0$, choose j_0 large enough that there is a set S in $\mathcal{K}_{j_0}$ so that $\mu(S \triangle \{\vec{x} \mid g(\vec{x}) \leq \alpha\}) < \frac{\beta-\alpha}{1-\alpha}\epsilon$.

For each $j > j_0$, let D_j be the subset of S where $\widetilde{B}_d^j$ is $\geq \beta$. For $j > j_0$, let $D_j' = D_j \backslash \bigcup_{j' \in (j_0, j)} D_j$. Observe that $\frac{1}{\mu(D_j')} \int_{D_j'} \chi_B \, d\mu \geq \beta$ for all j (since each $D_{j'}$ with $j' < j$ is also $\mathcal{K}_j$-measurable). Letting $D = \bigcup_{j>j_0} D_j = \bigcup_{j>j_0} D_j'$, we have $\frac{1}{\mu(D)} \int_D \chi_B \, d\mu \geq \beta$. Since $\chi_B \leq 1$, this means $\mu(D) < \epsilon$.

Therefore the set of $\vec{x}$ with $g(\vec{x}) \leq \alpha$ and $\limsup_{j\to\infty} \widetilde{B}_d^j(\vec{x}) > \beta$ has measure $< \epsilon + \frac{\beta-\alpha}{1-\alpha}\epsilon$. Since this holds for all ϵ and all $\alpha < \beta$, it follows that $\widetilde{B}_d^j$ converges to $g = \mathbb{E}(\chi_B \mid \mathcal{B}_{r,d}^-)$ pointwise almost everywhere.

This also shows that $\widetilde{B}_d(\vec{x})$ exists almost everywhere, and the inductive hypothesis (and the fact that a countable union of sets of measure 0 has measure 0) ensures that almost every point satisfies the second condition in being a point of density.

We finish showing that almost every point is a point of density by showing the third condition holds almost everywhere. Choose $j_d \ll j_{d-1} \ll \cdots \ll j_1$ so that, except on a set S of measure $< \epsilon^2$, we have $|\widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x}) - \widetilde{B}_d(\vec{x})| < \epsilon$.

Consider some neighborhood $N = \bigcap_{m\leq d} \mathcal{N}_m^{j_m}(\vec{x})$ with $\vec{x} \notin S$. Except for a set of neighborhoods of size at most ϵ, $\frac{\mu(N\cap S)}{\mu(N)} < \epsilon$. On such a neighborhood, we have

$$\frac{\int_{\bigcap_{m\leq d} \mathcal{N}_m^{j_m}(\vec{x})} \left|\widetilde{B}_d(\vec{x}) - \widetilde{B}_d(\vec{y})\right| \, d\mu_d(\vec{y})}{\mu_d(\bigcap_{m\leq d} \mathcal{N}_m^{j_m}(\vec{x}))} \leq 3\epsilon,$$

since for $\vec{y} \notin R$ we have

$$|\widetilde{B}_d(\vec{x}) - \widetilde{B}_d(\vec{y})| \leq |\widetilde{B}_d(\vec{x}) - \widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x})| + |\widetilde{B}_d^{j_d,\ldots,j_1}(\vec{x}) - \widetilde{B}_d^{j_d,\ldots,j_1}(\vec{y})| + |\widetilde{B}_d(\vec{y}) - \widetilde{B}_d^{j_d,\ldots,j_1}(\vec{y})| \leq \epsilon + 0 + \epsilon,$$

and the points with $\vec{y} \in R$ contribute measure at most ϵ.

Since this holds for every ϵ, we have

$$\lim_{j_d\to\infty} \cdots \lim_{j_1\to\infty} \frac{\int_{\bigcap_{m\leq d} \mathcal{N}_m^{j_m}(\vec{x})} \left|\widetilde{B}_d(\vec{x}) - \widetilde{B}_d(\vec{y})\right| \, d\mu_d(\vec{y})}{\mu_d(\bigcap_{m\leq d} \mathcal{N}_m^{j_m}(\vec{x}))} = 0$$

for almost every $\vec{x}$.

Let Z be the set of points with $\widetilde{B}_d(\vec{x}) = 0$. Since $\widetilde{B}_d$ is $\mathcal{B}^-_{r,d}$-measurable, so is Z. Therefore

$$\mu(B \cap Z) = \int \chi_Z \chi_B)\, d\mu_r = \int \chi_Z \mathbb{E}(\chi_B \mid \mathcal{B}^-_{r,d})\, d\mu_r = \int_Z \widetilde{B}_d(\vec{x})\, d\mu_r = 0.$$

Therefore for almost every $\vec{x} \in B$, $\widetilde{B}_d(\vec{X}) > 0$.

□

We are now ready to state a version of the *counting lemma*. In our case, it says that if we can find a single copy of a k-graph which is at a "typical" point then we have a positive measure collection of copies.

Slightly more generally, we allow the case that the edges of our graph get mapped to distinct sets. We need this in the induction, where we need to know that having each x_e be a point of density implies that $\vec{x}_W$ has proper neighborhoods.

Theorem 8.43. *Let W be a finite set and, for each $e \in \binom{W}{k}$, let $B_e \subseteq V^k$ be a measurable set. Let $\vec{x}_W$ be a point such that, for every $e \in \binom{W}{k}$, x_e is a point of density of B_e with $(\widetilde{B_e})_{k-1}(x_e) > 0$.*

Then $\mu(\widetilde{K(\{B_e\})}_{k-1}) > 0$, and so in particular $\mu(K(\{B_e\})) > 0$.

An element of $K(\{B_e\})$ is a "copy" of W in the sense that when $\vec{y}_W \in K(\{B_e\})$, for each e we have $\vec{y}_e \in B_e$. For instance, if we want to show $t_H(E) > 0$ with $H = (W, F)$, we can take $B_e = E$ whenever $e \in F$ and $B_e = V^k$ for $e \notin F$. We can also apply this to show $t^{ind}_H(E) > 0$, taking $B_e = E$ when $e \in F$ and $B_e = V^k \setminus E$ when $e \notin F$.

Proof. We proceed by induction on k. By 8.34, $\mu(K(\{B_e\})) = \int \prod_e \widetilde{B_e}_{k-1}(\vec{x}_e)\, d\mu_k(\vec{x})$, so it suffices to show that $\int \prod_e \widetilde{B_e}_{k-1}(\vec{x}_e)\, d\mu_k(\vec{x}) > 0$.

Since each x_e is a point of density of $N^e = \mathcal{N}^j_m(x_e)$, we have $\widetilde{N^e}_{m-1}(x_e) > 0$, and therefore by the inductive hypothesis, for all $j_{k-1}, \ldots, j_1$, we have $\mu(\bigcap_{m<k} \mathcal{N}^{j_m}_m(\vec{x}_W))$ has positive measure, so $\vec{x}_W$ has proper neighborhoods.

Choose $\epsilon = \min_e \chi_{(\widetilde{B_e})_{k-1}}(x_e) > 0$. Choose $j_{k-1} \ll j_{k-2} \ll \cdots j_1$ sufficiently large. Since $\vec{x}_W$ has proper neighborhoods, the set $N = \bigcap_{m<k} \mathcal{N}^{j_m}_m(\vec{x}_W)$ has positive measure.

For each $e \in \binom{W}{k}$, the set of points in N with $(\widetilde{B_e})_{k-1}(x_e) < \epsilon/2$ is a small fraction of $\bigcap_{m<k} \mathcal{N}^{j_m}_m(\vec{x}_W)$, so after removing all of these, we have most of N left, so $\int_N \prod_e \widetilde{B_e}_{k-1}(\vec{x}_e)\, d\mu_k(\vec{x}) \geq \frac{1}{2}\mu(N) > 0$. □

Corollary 8.44. *Let W be a finite set and suppose that for each $s \in \binom{W}{k}$ we have a set $B_s \subseteq V^k$. If $\mu(K(\{B_s\})) = 0$ then for each s there is a Z_s so that:*

- $\mu(Z_s) = 0$,
- $\bigcap_{j\in[r,k]} K(\{B_s \setminus Z_s\}_{s\in\binom{W}{k}}) = \emptyset$.

Proof. Let Z_s contain those x_s which are not points of density for B_s. □

8.6 Szemerédi's Theorem

We can, at last, prove Szemerédi's Theorem.

Theorem 8.45 (Szemerédi's Theorem)**.** *For every $\epsilon > 0$ and every k, there is an N so that whenever $n \geq N$ and $A \subseteq \{1, 2, \ldots, n\}$ is a set with $\frac{|A|}{n} \geq \epsilon$, there is an $a \in A$ and a $d > 0$ such that*

$$a, a + d, a + 2d, \ldots, a + (k - 1)d \in A.$$

The proof has the same basic structure as the proof of Roth's Theorem. Fix k and let $\mathrm{K}_{k,k-1}$ be the complete $k-1$-graph on k-vertices—that is, the $k-1$-graph $([1, \ldots, k], \binom{[1,\ldots,k]}{k-1})$. (So $\mathrm{K}_{3,2} = \mathrm{C}_3$.) Given a set A with density $\geq \epsilon$, we will identify a $k-1$-graph E of positive density with the property that we cannot "remove" $\mathrm{K}_{k,k-1}$—that is, so that whenever $\mu_{k-1}(Z) = 0$, $T_{\mathrm{K}_{k,k-1}}(E \setminus Z) \neq \emptyset$.

Proof. Towards a contradiction, suppose this fails: there is an $\epsilon > 0$ and a k and, for every N, an $n \geq N$ and an $S_N \subseteq \{1, 2, \ldots, n\}$ with $\frac{|S_N|}{|n|} \geq \epsilon$ so that S_N does not contain an arithmetic progression of length k.

We construct a corresponding $k-1$-graph. It will be convenient to make the graph partite. We will have a set of vertices $V_N = \bigcup_{i\leq k} X_{i,N}$ where each part $X_{i,N} = \{1, 2, \ldots, kn\}$.

For each $I \in \binom{\{1,2,\ldots,k\}}{k-1}$ we define a set $A_{I,N} \subseteq \prod_{i\in I} X_{i,N}$. When $I = \{1, 2, \ldots, k-1\}$, we set $A_{I,N}$ to consist of those $(x_1, \ldots, x_{k-1})$ where $\sum_{i<k} i \cdot x_i \in S_N$. When $I = \{1, 2, \ldots, k\} \setminus \{i\}$ with $i < k$, we take $A_{I,N}$ to consist of those $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$ where $\sum_{j<k, j\neq i} j \cdot x_j + i(x_k - \sum_{j<k, j\neq i} x_j) \in S_N$.

Suppose we have values $(x_1, \ldots, x_k)$ and take $a = \sum_{i<k} i \cdot x_i$ and $d = x_k - \sum_{i<k} x_i$. Observe that $a \in S_N$ exactly when $(x_1, \ldots, x_{k-1}) \in A_{\{1,2,\ldots,k-1\},N}$

and

$$a + id = \sum_{j<k} j \cdot x_j + ix_k - i\sum_{j<k} x_j = \sum_{j<k, j\neq i} j \cdot x_j + i(x_k - \sum_{j<k, j\neq i} x_j) \in S_N$$

exactly when $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \in A_{\{1,2,\ldots,i-1,i+1,\ldots,k\},N}$. In particular, if $(x_1, \ldots, x_k) \in K(\{A_{I,N}\}_{I\in\binom{k}{k-1}})$ then we have an a, d so that $a, a+d, \ldots, a+(k-1)d \in S_N$. (However this includes the possibility that $d = 0$.) On the other hand, whenever $a \in S_N$, we can take any sum $a = \sum_{i<k} i \cdot x_i$ and the sequence $(x_1, \ldots, x_{k-1}, \sum_i x_i) \in K(\{A_{I,N}\}_{I\in\binom{k}{k-1}})$. Let T_N be the set of these trivial k-tuples—that is,

$$T_N = \{(x_1, \ldots, x_{k-1}, \sum_{i<k} x_i) \mid (x_1, \ldots, x_{k-1}, \sum_{i<k} x_i) \in K(\{A_{I,N}\}_{I\in\binom{k}{k-1}}).$$

Note that, for each $(x_1, \ldots, x_{k-1})$, there is at most one b so that $(x_1, \ldots, x_{k-1}, b) \in T_N$, so $|T_N| \leq |V_N|^{k-1}$, so $\mu_k(T_N)$ is small when N is large.

Let $\rho_N : \prod_{j<k} X_{j,N} \to X_{k,N}$ be given by

$$\rho_N(x_1, \ldots, x_{k-1}) = \sum_{i<k} x_i,$$

so the function $\rho_{i,N}(x_1, \ldots, x_{k-1}) \mapsto (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k, \rho_N(x_1, \ldots, x_{k-1}))$ is one-to-one for each $i < k$.

Let $V = [V_n]_{\mathcal{U}}$ and, for each $I \in \binom{k}{k-1}$, $A_I = [A_{I,N}]_{\mathcal{U}}$. Let $T = [T_N]_{\mathcal{U}}$. Let $\rho = [\rho_N]_{\mathcal{U}}$ and $\rho_i = [\rho_{i,N}]_{\mathcal{U}}$.

Let $G = [G_N]_{\mathcal{U}}$. If $t_{\mathrm{K}_{k,k-1}}(G) > 0$ then we may find a complete $(x_1, \ldots, x_k) \in K(\{A_I\}_{I\in\binom{k}{k-1}}) \setminus T$. But this would mean that, for many N, we can find $(x_1, \ldots, x_k) \in K(\{A_{I,N}\}_{I\in\binom{k}{k-1}}) \setminus T_N$, and therefore an arithmetic progression in A_N with $d \neq 0$.

So suppose, towards a contradiction, that $t_{\mathrm{K}_{k,k-1}}(G) = 0$. Then by Theorem 8.44 (with $r = k$), we find sets Z_I with $\mu_{k-1}(Z_I) = 0$. Let $Z \subseteq \prod_{i<k} X_i$ be given by $Z = Z_{\{1,2,\ldots,k-1\}} \cup \bigcup_i \rho_i^{-1}(Z_{[k]\setminus\{i\}})$. Since the ρ_i are internal one-to-one functions, and therefore measure-preserving, $\mu_{k-1}(Z) = 0$.

In particular, $\mu(A_{\{1,2,\ldots,k-1\}} \setminus Z) > 0$. For any $(x_1, \ldots, x_{k-1}) \in A_{\{1,2,\ldots,k-1\}} \setminus Z$, for each i we have $\rho_i(x_1, \ldots, x_{k-1}) \in A_{[k]\setminus\{i\}} \setminus Z_{[k]\setminus\{i\}}$, and therefore $(x_1, \ldots, x_{k-1}, \rho(x_1, \ldots, x_{k-1})) \in K(\{A_I \setminus Z_I\}_{I\in\binom{k}{k-1}})$. This gives the desired contradiction. □

8.7 Hypergraphons

We can represent measurable k-graphs as measurable sets in a product measure space rather than a Keisler graded probability space, just like we did for graphs. Again, the catch is that the product will need to be over more than k sets.

We need to separate out all the σ-algebras $\mathcal{B}_{k,1}$, $\mathcal{B}_{k,2}$, and so on. We need k coordinates in our new representation just to capture the information in $\mathcal{B}_{k,1}$. We also need the information in $\mathcal{B}_{k,2}$; there are $\binom{k}{2}$ pairs of coordinates, each of which will need its own coordinate containing the binary information between that pair. In total, the new representation will be a product over $2^k - 1$ sets, one for each non-empty subset of $\{1, 2, \ldots, k\}$.

We will need to keep track of this correspondance between non-empty subsets of our k old coordinates and our $2^k - 1$ new coordinates. So we will proceed as follows. For each arity $i \leq k$, we will have a measure space $(\Omega_i, \mathcal{D}_i, \mu)$ representing i-ary information. We will then work in the space $\Omega_{(\leq k)} = \prod_{s \subseteq \{1,2,\ldots,k\}, s \neq \emptyset} \Omega_{|s|}$ where the measurable sets are the corresponding products of the $\mathcal{D}_i$. In particular, we will be able to identify a measurable k-graph E with a corresponding $E^* \subseteq \Omega_{(\leq k)}$—that is, we will have a measurable equivalence of measure algebras $\rho : V^k \to \Omega_{(\leq k)} = \prod_{s \subseteq \{1,2,\ldots,k\}, s \neq \emptyset} \Omega_i$.

It is common to use the fact that all separable atomless measure spaces are isomorphic to the Lebesgue measure to take all the Ω_i to be equal to $[0, 1]$. Then we can simply consider measurable sets and functions on $[0, 1]^{2^k - 1}$, at the price of suggesting the existence of additional structure (like bijections between coordinates representing different sizes, or the topology on $[0, 1]$) which is not meaningful.

One challenge with our new representation is that understanding what the coordinates mean is a bit subtle. With a measurable k-graph, V *is* a set of vertices, and E is a literal k-graph. When we instead consider $E^* \subseteq \Omega_{(\leq k)}$, we do not exactly have well-defined "vertices". (That perspective has not been lost—as we will see, we can recover a literal k-graph by sampling randomly in the right way.) Instead, the information about the k coordinates of V has been separated out among various coordinates (that is, it has been separated into pieces with different randomness properties). In particular, the meanings of the coordinates are not independent—the meaning of the $\{1, 2\}$ coordinate is related to the meanings of the $\{1\}$ and $\{2\}$ coordinates, but not the $\{3\}$ coordinate, and so on.

One way to see this relationship is to think about how the symmetry of $E \subseteq V^k$ is reflected in E^*. The symmetry of E meant that for any permutation π of $\{1, 2, \ldots, k\}$, $E^\pi = \{(x_{\pi(1)}, \ldots, x_{\pi(k)}) \mid (x_1, \ldots, x_k) \in E\}$

is equal to E. We should only consider these same $k!$ permutations of $\Omega_{(\leq k)}$: whenever π is a permutation of $\{1, 2, \ldots, k\}$, we have the corresponding permutation (which, by abuse of notation, we also call π) of $\mathcal{P}(\{1, 2, \ldots, k\})\backslash\emptyset$ taking a set s to $\pi(s) = \{\pi(i) \mid i \in s\}$.

Definition 8.46. When π is a permutation of $\{1, 2, \ldots, k\}$, the induced permutation $\tilde{\pi} : \Omega_{(\leq k)} \to \Omega_{(\leq k)}$ is defined by

$$\tilde{\pi}(\{x_s\}_{s\in\mathcal{P}(\{1,2,\ldots,k\})\backslash\emptyset}) = \{x_{\pi(s)}\}_{s\in\mathcal{P}(\{1,2,\ldots,k\})\backslash\emptyset}.$$

We have allowed the spaces Ω_i to be distinct for different sizes i; note that we always map a coordinate $x_s \in \Omega_{|s|}$ to a coordinate $x_{\pi(s)} \in \Omega_{|\pi(s)|} = \Omega_{|s|}$; we are never tempted to map, say Ω_2 to Ω_1 because these sets contain different sorts of data (even if they both happen to be equal to $[0, 1]$).

Our arguments will build on the work we did in Section 6.10. Indeed, we essentially carried out the $k = 2$ case of the construction of E^*. Recall that we began with a measurable graph (V, E, μ_1), and constructed a corresponding $E^* \subseteq V \times V \times V^2$ which was measurable with respect to the product measure $\mathcal{B}_1^- \times \mathcal{B}_1^- \times \mathcal{R}$, where $\mathcal{R}$ was an orthogonal complement to $\mathcal{B}_1^-$. (We did not verify at the time that $t_H(E^*) = t_H(E)$, though we implied it would be true.)

There is a technical point about symmetry we did not need to consider when dealing with graphs. There is one non-trivial permutation of the coordinates of V^2, namely when we swap the two coordinates. The natural analog for $V \times V \times V^2$ is $\pi(v, w, (v', w')) = (w, v, (w', v'))$—that is, we would want to swap, not only the first two coordinates, but we would want to carry out a swap "inside" the V^2 coordinate.

This will not suffice. When we take $\Omega_2 = V^2$, we do not want to have to keep track of internal structure for Ω_2. We didn't need to worry about this in the $k = 2$ case because E was symmetric, which means E^* was actually measurable with respect to the smaller σ-algebra $\mathcal{B}_1^- \times \mathcal{B}_1^- \times \mathcal{R}^{sym}$, where $\mathcal{R}^{sym}$ consists of the *symmetric* elements of $\mathcal{R}$. Once we restrict our measurable sets on $\Omega_2 = V^2$ to be only the symmetric sets, we can disregard swapping points within Ω_2 because (v', w') and (w', v') are indistinguishable points.

We might hope that if all we want to consider is k-graphs, the same idea would work in general. However it turns out that when E is a k-graph, we can ignore the symmery concerns for Ω_k in this way, but that even when E is symmetric, when we start writing it in terms of lower arity information, we may see asymmetric sets arise. For instance, recall our example $\mathrm{C}(\rightarrow)$, the set of $\{x, y, z\}$ where a random tournament $\rightarrow$ forms a cycle. $\mathrm{C}(\rightarrow)$ is a symmetric 3-graph, but the binary data we use to describe it is asymmetric.

(One might imagine that even if *this* description seems asymmetric, some other way to describe it might not be, but it can be shown that there is no way to represent this example using only symmetric binary data; see [40].)

Fortunately, there is another way around this issue, by pushing all the asymmetry down to the unary level. Before giving the general argument, let us consider what we do in the $k = 2$ case. First, observe that we can always have a measurable ordering on V. This is true in general—we have mentioned that we can always find a measurable equivalence from V to $[0, 1]$—but a more direct way to see it in our ultraproduct setting is to observe that when $V = [V_n]_{\mathcal{U}}$ where the V_n are finite, we can take orderings $\prec_n$ on each V_n, and then $\prec = [\prec_n]_{\mathcal{U}}$ is an internal ordering on V, so automatically measurable.

The graph of $\prec$, $G_\prec = \{(x, y) \in V^2 \mid x \prec y\}$ is always $\mathcal{B}_1^2$-measurable. For any m, we may choose a finite sequence $a_1 \prec a_2 \prec \cdots \prec a_m$ so that, the interval $I_i = \{x \mid a_i \prec x \preceq a_{i+1}\}$ has measure $< 2/m$. Then $(G_\prec \triangle \bigcup_{i<j}(I_i \times I_j)) \subseteq \bigcup_i I_i^2$ and therefore has measure $< 4/m$. Since we can take m as small as we want, we can approximate $G_\prec$ arbitrarily well by rectangles.

So now suppose we are given $f : V^2 \to \mathbb{R}$ and wish to construct an analogous function $f^* : V \times V \times V^2 \to \mathbb{R}$ which is measurable with respect to $\mathcal{B}_1 \times \mathcal{B}_1 \times \mathcal{R}^{sym}$. (And so in particular, the σ-algebra on the V^2 component is symmetric.)

We define two new functions based on f, but which ignore the given order of the inputs. $f_\prec(x, y)$ will always act like the inputs are ordered by $\prec$:

$$f_\prec(x, y) = \begin{cases} f(x, y) & \text{if } x \prec y \\ f(y, x) & \text{otherwise} \end{cases}$$

while $f_\succ$ will always act like the inputs are in the reverse order from $\prec$:

$$f_\succ(x, y) = \begin{cases} f(y, x) & \text{if } x \prec y \\ f(x, y) & \text{otherwise} \end{cases}.$$

Note that $f_\prec$ and $f_\succ$ are both symmetric: they don't depend on the order of the inputs because they ignore the input order, replacing it with an order derived from $\prec$.

Because $f_\prec$ and $f_\succ$ are symmetric, the corresponding functions $f^*_\prec$ and f^*_{succ} are both $\mathcal{B}_1 \times \mathcal{B}_1 \times \mathcal{R}^{sym}$-measurable. But then we can define

$$f^*(x, y) = \begin{cases} f^*_\prec(x, y) & \text{if } x \prec y \\ f^*_\succ(x, y) & \text{otherwise} \end{cases}$$

Since $G_\prec$ is $\mathcal{B}_1^2$-measurable, f^* is $\mathcal{B}_1^2 \times \mathcal{R}^{sym}$-measurable, which is exactly what we want.

Let us now prove our general result. First, we need to pass to a separable sub-σ-algebra.

Definition 8.47. When $\mathcal{B} \subseteq \mathcal{B}_{k-1}$, $\mathcal{B}^{\binom{k}{k-1}}$ is the σ-algebra of $\binom{k}{k-1}$-cylinder intersection sets whose elements belong to $\mathcal{B}$.

Theorem 8.48. *Let E be a measurable k-graph and suppose there is a measurable order $\prec$ on V. There is a σ-algebra $\mathcal{B}_k^- \subseteq \mathcal{B}_k$ containing E, probability measure spaces $(\Omega_i, \mathcal{D}_i, \mu)$, and a measurable equivalence of measure algebras $\rho : V^k \to \Omega_{(\leq k)} = \prod_{s \subseteq \{1,2,\ldots,k\}, s \neq \emptyset} \Omega_i$ such that whenever π is a permutation of $\{1, 2, \ldots, k\}$, $\rho \circ \pi = \tilde{\pi} \circ \rho$.*

Proof. Take the σ-algebras $\mathcal{B}_i^-$ given by applying Lemma 8.36 to $\{E\}$. For each $\mathcal{B}_i^-$, there are $\binom{i}{i-1}$ distinct natural sub-σ-algebras of cylinder sets isomorphic to $\mathcal{B}_{i-1}^-$, corresponding to the $\binom{i}{i-1}$ choices of $i-1$ coordinates, and together these generate a σ-algebra $\mathcal{B}_{i,i-1}^-$ generated by cylinder intersection sets from $\mathcal{B}_{i-1}^-$. By Lemma 6.44, we choose a σ-algebra $\mathcal{R}_i$ orthogonal to $\mathcal{B}_{i,i-1}^-$ so that $\mathcal{R}_i$ together with $\mathcal{B}_{i,i-1}^-$ generates $\mathcal{B}_i^-$.

For each i, we will let $\Omega_i = \binom{V}{i}$, the set of *unordered* i-tuples from V. We will take $(\Omega_i, \mathcal{R}_i^{sym}, \mu_i)$ to be our i-th probability measure space, where $\mathcal{R}_i^{sym}$ is the set of symmetric elements of $\mathcal{R}_i$.

For any $d \leq k$ we have the product space $\Omega_{(\leq d)} = \prod_{s \in \mathcal{P}(\{1,2,\ldots,d\}) \setminus \emptyset} \Omega_{|s|}$. We define $\rho_d : V^d \to \Omega_{(\leq d)}$ in the natural way, setting $\rho_d(x_1, \ldots, x_d)$ to be the element $\{y_s\}_{s \in \mathcal{P}(\{1,2,\ldots,d\}) \setminus \emptyset}$ with $y_s = \{x_i\}_{i \in s}$. ρ_d is measure-preserving because $\mathcal{R}_i^{sym}$ is orthogonal to $\mathcal{B}_{i,i-1}^-$.

It remains to show that for any $B \in \mathcal{B}_d^-$, there is a $D \in \prod_{s \in \mathcal{P}(\{1,2,\ldots,d\}) \setminus \emptyset} \mathcal{R}_{|s|}^{sym}$ such that $\mu(B \triangle \rho_d^{-1}(D)) = 0$. We show this by induction on d. When $d = 1$, ρ_1 is the identity on V, so the claim is trivial. Suppose the claim holds for $d < k$.

If B is symmetric, we B is approximated by a union of intersections of sets from $\mathcal{B}_{d+1,d}^-$ and $\mathcal{R}_d^{sym}$. We can choose sets so that $\mu(B \triangle \bigcup_{j \leq n} \bigcap_{s \in \binom{d+1}{d}} B_{j,s} \cap R_j) < \epsilon$, and then we have sets $D_{j,s} \subseteq \prod_{t \in \mathcal{P}(s) \setminus \{\emptyset, s\}} \Omega_{|t|}$ and we already have $R_j \subseteq \Omega_{d+1}$, so for each j we have $\mu(\bigcap_{s \in \binom{d+1}{d}} B_{j,s} \cap R_j \triangle \rho_{d+1}^{-1}(\bigcap_{s \in \binom{d+1}{d}} D_{j,s} \cap R_j)) = 0$.

If B is not symmetric, we rewrite it as a combination of symmetric sets: for each permutation π of $\{1, 2, \ldots, d+1\}$, define B_π to consist of those $(x_1, \ldots, x_{d+1})$ such that, taking the unique ordering of $\{x_1, \ldots, x_{d+1}\}$ to be $\prec$-increasing and applying π to this ordering, this tuple is in B. Then B_π is symmetric, so in the inverse image of ρ_d.

But then B is an intersection of $\bigcup_\pi B_\pi$ with a set $\rho_{d+1}^{-1}(U)$ where U is a measurable subset of Ω_1^{d+1}. □

We are only one step from defining a k-hypergraphon.

Definition 8.49. A *k-hypergraphon* is a measurable function $W : [0,1]^{\mathcal{P}(\{1,2,\ldots,k\})\setminus\{\emptyset,\{1,2,\ldots,k\}\}} \to [0,1]$ which is symmetric under all the permutations $\tilde{\pi}$.

Given E, we obtain a k-hypergraphon by taking E^* with $\rho^{-1}(E^*) = E$ and letting $W = \int \chi_{E^*}\, d\mu(x_{\{1,2,\ldots,k\}})$. The reason we integrate away the k-ary component $x_{\{1,2,\ldots,k\}}$ is that it is the "purely random" component: instead of a set with a purely random component, we get a *function* W telling us how likely $x_{\{1,2,\ldots,k\}}$ was to put us into the set.

8.8 k-ary Combinatorics

We saw a combinatorial property which almost characterized which graphs belonged to $\mathcal{B}_{2,1}$—E has finite VC dimension if it is $\mathcal{B}_{2,1}$-measurable under any measure.

Analogously, we can ask when a graph be $\mathcal{B}_{k,k-1}$-measurable.

Theorem 8.50. *Suppose a k-graph (V,E) is not $\mathcal{B}_{k,k-1}$-measurable. Then for every k-partite k-graph $H = (\bigcup_{i\le k} W_i, F)$ there is a $\pi : \bigcup_{i\le k} W_i \to V$ such that for any $w_1,\ldots,w_k$ with each $w_i \in W_i$, $\{\pi(w_1),\ldots,\pi(w_k)\} \in E$ if and only if $\{w_1,\ldots,w_k\} \in F$.*

What we are really describing is "containing an induced copy of H". But this is not quite an induced copy the way we defined it, because there could be extraneous edges inside some $\pi(W_i)$. All we are requiring is that the *partite* k-edges of $\bigcup_{i\le k}\pi(W_i)$ are exactly the k-edges in F, and we ignore potential k-edges which contain more than one vertex in the same part. (This complication is because (V,E) is not itself k-partite.)

Proof. Except for a set of points of measure 0, every point in E has $\widetilde{E}_{k-1}(\vec{x}) > 0$. We also know that for almost every point of $V^k \setminus E$, $\widetilde{E}_{k-1}(\vec{x}) < 1$. So suppose that almost every point of E has $\widetilde{E}_{k-1}(\vec{x}) = 1$; then, up to measure 0, E is precisely $\{\vec{x} \mid \widetilde{E}_{k-1}(\vec{x}) = 1\}$, which is $\mathcal{B}_{k,k-1}$-measurable.

So if E is not $\mathcal{B}_{k,k-1}$-measurable, there is a positive set of points with $0 < \widetilde{E}_{k-1}(\vec{x}) < 1$. Since almost every point is a point of density for E and for $V^k \setminus E$, there must be a point of density $\vec{x}$ for both with $0 < \widetilde{E}_{k-1}(\vec{x}) < 1$.

Now let $(\prod_{i\leq k} W_i, F)$ be any k-partite k-graph (that is, each $e \in F$ has $|e \cap W_i| = 1$ for each i). We can appy Theorem 8.43: take $\vec{a} \in V^{\prod_{i\leq k} W_i}$ to be the point with $a_w = x_i$ for each $w \in W_i$, and take $B_s = E$ for $e \in F$ and $B_e = V^k \setminus E$ for each $e \in \prod_{i\leq k} W_i \setminus e$. For all e which are not partite, take $B_e = V^k$ (so every point is trivially a point of density). Then, by Theorem 8.43, we have $t_{\prod_{i\leq k} W_i, F}(E) > 0$. □

In the case of ordinary graphs, this is an alternate proof of Theorem 7.13 using Theorem 7.2 which shows that containing all bipartite graphs is equivalent to saying that E has infinite VC dimension.

We could define a higher-arity notion of VC dimension analogously.

Definition 8.51. A k-graph (V, E) has infinite VC_{k-1} dimension if for every k-partite k-graph $H = (\bigcup_{i\leq k} W_i, F)$ there is a $\pi : \bigcup_{i\leq k} W_i \to V$ such that for any $w_1, \ldots, w_k$ with each $w_i \in W_i$, $\{\pi(w_1), \ldots, \pi(w_k)\} \in E$ if and only if $\{w_1, \ldots, w_k\} \in F$. Otherwise E has finite VC_{k-1} dimension.

This is indexed so that VC_1 dimension is simply ordinary VC dimension.

VC_k dimension is usually defined by an equivalent property which is more closely analogous to the standard definition of VC dimension.

Theorem 8.52. *E has infinite VC_k dimension if and only if, for every d, there exists a k-ary box $A = \{a_{1,1}, \ldots, a_{1,d}\} \times \cdots \times \{a_{k,1}, \ldots, a_{k,d}\}$ such that, for every $B \subseteq A$, there is a b with $E_b \cap A = B$.*

Proof. For the left to right direction, consider the $k+1$-partite graph $(\prod_{i\leq k+1} W_i, F)$ where $|W_i| = d$ for each $i \leq k$, $|W_{k+1}| = 2^{d^k}$, and for each $B \subseteq \prod_{i\leq k} W_i$, there is a $b \in W_k$ such that $\{a \in \prod_{i\leq k} W_i \mid (a, b) \in F\} = B$.

Conversely, let a $k+1$-paritite graph $(\bigcup_{i\leq k+1} W_i, F)$ be given. Choose $d = \max_{i\leq k} |W_i|$ and take the k-ary box $A = \{a_{1,1}, \ldots, a_{1,d}\} \times \cdots \times \{a_{k,1}, \ldots, a_{k,d}\}$. Define π on $\bigcup_{i\leq k} W_i$ so that, for each $i \leq k$, $\pi : W_i \to \{a_{i,1}, \ldots, a_{i,d}\}$ is an injection.

For each $w \in W_{k+1}$, there is a set $B_w = \{\{\pi(w_1), \ldots, \pi(w_k)\} \mid \{w_1, \ldots, w_k\} \in F_w\} \subseteq A$. By assumption, there is a b_w with $E_{b_w} \cap A = B_w$, so we define $\pi(w) = b_w$. Then π gives the desired copy of $(\bigcup_{i\leq k+1} W_i, F)$. □

As with VC dimension, we can rephrase VC_k dimension in terms of set systems—sets V and collections $\mathcal{F}$ of subsets of V.

Definition 8.53. When $(V, \mathcal{F})$ is a set system, the VC_k dimension of $(V, \mathcal{F})$ is the largest n such that there are sets $A_1, \ldots, A_k$ with $\prod_{i\leq k} A_i \subseteq V$ so that $\mathcal{F}$ shatters A.

For VC_k dimension, this makes the most sense when we think of V as itself being a product, $V = \prod_{i \leq k} V_i$, though this is no formal need to include this in the definition.

When $X \subseteq V$, we can ask how many subsets of X get picked out by $\mathcal{F}$:

$$\pi_{\mathcal{F},k}(n) = \max\{|\Pi_{\mathcal{F}}(X)| \mid X \subseteq V, X = \prod_{i \leq k} X_i, |X_i| = n \text{ for all i}\}.$$

Recall that $\Pi_{\mathcal{F}}(X)$ to be $\{X \cap F \mid F \in \mathcal{F}\}$, so $\pi_{\mathcal{F},k}$ asks how many subsets $\mathcal{F}$ can pick out of a box with sides of size n. (Again, the $k = 1$ case is precisely $\pi_{\mathcal{F}}$ as we defined it before.) Certainly $\pi_{\mathcal{F},k}(n) \leq 2^{n^k}$, since the sets X we consider in the definition have size n^k. Moreover, if $\pi_{\mathcal{F},k}(n) < 2^{n^k}$ then the VC_k dimension of $\mathcal{F}$ is $\leq n$ by definition.

We can prove a generalization of the Sauer–Shelah Theorem, Theorem 7.5, showing that when V has finite VC_k dimension, $\pi_{\mathcal{F},k}(n)$ is bounded strictly away from 2^{n^k}.

In order to characterize this definition, we need the Zarankiewicz numbers.

Definition 8.54. The Zarankiewicz number $z_k(n, d)$ is the smallest z such that if $(\prod_{i \leq k} V_i, E)$ is a k-partite k-graph with $|V_i| = m$ for all m and $|E| \geq z$ then there are $A_i \subseteq V_i$ with $|A_i| = d$ and $\prod_{i \leq k} A_i \subseteq E$.

That is, if we have a k-partite k-graph on m vertices then once it contains z edges, there must be a box with sides of size d which is entirely contained in E.

Theorem 8.55. *For and k, d, for sufficiently large n, $z_k(n, d) \leq 2^{k/d} n^{k - \frac{1}{d^{k-1}}}$.*

Proof. By induction on k.

Suppose we have $(\prod_{i \leq k} V_i, E)$ with $|V_i| = n$ for all i and $|E| \geq (kn)^{k - \frac{1}{d^{k-1}}}$. For each $v \in V_k$, let $E_v = \{\vec{y} \in \prod_{i<k} V_i \mid \vec{y} \cup \{v\} \in E\}$. Then $\sum_{v \in V_k} |E_v| \geq 2^{k/d} n^{k - \frac{1}{d^{k-1}}} = n^{k-1} n \frac{2^{k/d}}{n^{1/d^{k-1}}}$.

For each $\vec{w} \in \prod_{i<k} V_i$, let $c_{\vec{w}} = |\{v \mid \vec{w} \in E_v\}|$, so $\sum_{\vec{w}} c_{\vec{w}} = n^{k-1} n \frac{1}{n^{1/d^{k-1}}}$

as well. Observe that

$$\begin{aligned}
\sum_{v_1,\dots,v_d} |\bigcap_{j\le d} E_{v_j}| &= \sum_{v_1,\dots,v_d}\sum_{\vec{w}\in\prod_{i<k}V_i}\prod_{j\le d}\chi_{E_{v_j}}(\vec{w})\\
&= \sum_{\vec{w}\in\prod_{i<k}V_i}\sum_{v_1,\dots,v_d}\prod_{j\le d}\chi_{E_{v_j}}(\vec{w})\\
&= \sum_{\vec{w}\in\prod_{i<k}V_i}(\sum_{v\in V_k}\chi_{E_v}(\vec{w}))^d\\
&= \sum_{\vec{w}\in\prod_{i<k}V_i} c_{\vec{w}}^d\\
&\ge \frac{1}{n^{(k-1)(d-1)}}(\sum_{\vec{w}\in\prod_{i<k}V_i} c_{\vec{w}})^d\\
&> \frac{1}{n^{(k-1)(d-1)}}(n^{k-1}n\frac{2^{k/d}}{n^{1/d^{k-1}}})^d\\
&= n^{k-1}(n\frac{2^{k/d}}{n^{1/d^{k-1}}})^d.
\end{aligned}$$

There are fewer than d^2n^{d-1} choices of vertices from V_k which are *not* all distinct; each such choice contributes at most n^{k-1} to the sum. The remaining $\binom{n}{d} \le n^d$ choices must contribute the remainder, so there must be some choice of distinct $v_1,\dots,v_d$ with

$$|\bigcap_{j\le d} E_{v_j}| \ge \frac{n^{k-1}(n\frac{2^{k/d}}{n^{1/d^{k-1}}})^d - d^2n^{d-1}n^{k-1}}{n^d} = n^{k-1-1/d^{k-2}}2^k - d^2n^{k-2}.$$

When $n \ge (d^2/2^{k-1})^{\frac{1}{1-1/d^{k-2}}}$, $d^2n^{k-2} \le n^{k-1-1/d^{k-2}}2^{k-1}$, so there is some $v_1,\dots,v_d$ with $|\bigcap_{j\le d} E_{v_j}| \ge n^{k-1-1/d^{k-2}}2^{k-1}$.

So $(\prod_{i<k}V_i, \bigcap_{j\le d}E_{v_j})$ is a $k-1$-partite $k-1$-graph with $|V_i| = m$ for all m and $|\bigcap_{j\le d}E_{v_j}| \ge 2^{k-1}n^{k-1-1/d^{k-2}}$, so by the inductive hypothesis, we can find $A_i \subseteq V_i$ for $i < k$ so that $\prod_{i<k}A_i \subseteq \bigcap_{j\le d}E_{v_j}$. Therefore $\prod_{i\le k}A_i \subseteq E$. □

Theorem 8.56. *If the VC_k dimension of $(V,\mathcal{F})$ is d then, for all n, $\pi_{\mathcal{F},k}(n) \le \sum_{i=0}^{z_k(n,d+1)-1}\binom{n}{i}$.*

Proof. Suppose there is an $X \subseteq V$ with $X = \prod_{i\le k}X_i$ and $|X_i| = n$ for all i, and $|\Pi_{\mathcal{F}}(X)| > \sum_{i=0}^{z_k(n,d+1)}\binom{n}{i}$. Recall that, by Lemma 7.6, there are at least $|\Pi_{\mathcal{F}}(X)|$ subsets of X which are shattered by $\mathcal{F}$.

In particular, there must be a $B \subseteq X$ with $|B| \geq z_k(n, d+1)$ which is shattered by $\mathcal{F}$. Consider the k-partite k-graph $(\prod_{i\leq k} X_i, B)$. Since $|B| \geq z_k(n, d+1)$, we can find $A_i \subseteq X_i$ with $|A_i| = d+1$ so that $\prod_{i\leq k} A_i \subseteq B$. But $\mathcal{F}$ shatters B, so in particular $\mathcal{F}$ shatters $\prod_{i\leq k} A_i$. Therefore the VC_k dimension of $(V, \mathcal{F})$ is $\geq d+1$. □

Corollary 8.57. *For each k, d there is an $\epsilon > 0$ so that, for n sufficiently large, if the VC_k dimension of $(V, \mathcal{F})$ is d then $\pi_{\mathcal{F},k}(n) \leq 2^{n^{k-\epsilon}}$.*

Proof. We have $\pi_{\mathcal{F},k}(n) \leq \sum_{i=0}^{z_k(n,d+1)-1} \binom{n}{i} \leq \sum_{i=0}^{2^{k/d} n^{k-\frac{1}{d^{k-1}}}} \binom{n}{i} \leq n^{2^{k/d} n^{k-\frac{1}{d^{k-1}}}} + 1 = 2^{2^{k/d} n^{k-\frac{1}{d^{k-1}}} \ln n} + 1$. The exponent here has the form $cn^{k-\frac{1}{d^k-1}} \ln n$ for a constant c (depending on k and d, but not n), so when n is large enough, the exponent is at most $n^{k-\epsilon}$. □

8.9 Remarks

Hypergraph quasirandomness is substantially more complicated than graph quasirandomness, and the treatment here only touches on some of the complications. The study of hypergraph quasirandomness and hypergraph analogs of regularity begins right after the modern theory of graph quasirandomness [29, 31, 32], but finding the right notion of hypergraph regularity needed to formulate the removal lemma took substantially longer [47, 76, 78, 98, 121, 134].

While it was clear from early on that there were multiple notions of hypergraph quasirandomness [31], the structure became more complicated as a variety of incomparable notions were introduced [60, 107]. The σ-algebra based approach we used here was introduced in [158] and given a quantitative version in [3]. Yet more notions of hypergraph quasirandomness were introduced in [40].

The idea of viewing hypergraph regularity in terms of points of density was introduced in [47] in the setting of what are now known as hypergraphons; the version here is based on [159]. The hypergraphon approach to regularity and limits of hypergraphs has been further studied and generalized [38, 164]. These sorts of representations can be seen as examples of the Aldous–Hoover Theorem [4, 86] which we discuss later.

The study of k-ary combinatorics and its relationship with hypergraph regularity is quite new. VC_k dimension is introduced in [138] and further studied in [23]. The result that bounded VC_k-dimension implies $\mathcal{B}_{k,k-1}$-measurability was shown in [28]; a quantitative proof in the $k = 3$ case was

given in [155] and the bounds on the number of $k-1$-ary cylinder sets needed was shown to be polynomial in [154]. [155] also introduces two distinct candidates for "2-stability".

Chapter 9

Random Countable Graphs

There are other perspectives on the measurable k-graphs that have formed the central topic in this book. In this final chapter, we give the basics of two other perspectives. One, coming from probability theory, is to look at exchangeable arrays of random variables; we prove the foundational theorem of that perspective, the Aldous–Hoover Theorem, which is in some sense a restatement, in quite different language, of the decomposition into σ-algebras we have already seen. Another perspective, connected to both model theory and descriptive set theory, is to look at automorphism invariant probability measures; we finish the book with one of the key theorems from this approach, the Ackerman–Freer–Patel Theorem.

We have already considered, in Section 3.6, what happens when we begin with an atomless measurable graph (V, E, μ_1) and sample finitely many vertices $\mathbf{v}_1, \dots, \mathbf{v}_n$.

The same idea extends to sampling countably many vertices: we choose vertices $\mathbf{v}_1, \mathbf{v}_2, \dots \in V$, distributed according to μ_1, and let $\mathbf{G}_\omega = (\{\mathbf{v}_n\}_{n\in\mathbb{N}}, E \restriction \{\mathbf{v}_n\}_{n\in\mathbb{N}})$. We may as well state this for k-graphs, as well.

Definition 9.1. If (V, E, μ_1) is an atomless measurable k-graph, a *random countable substructure* is a k-graph of the form $(\{\mathbf{v}_n\}_{n\in\mathbb{N}}, E \restriction \{\mathbf{v}_n\}_{n\in\mathbb{N}})$ where the $\mathbf{v}_n$ are chosen independently from V according to μ_1.

Our use of the term "structure" here gestures at the fact that most of the ideas in this chapter make sense, not only for graphs and k-graphs, but for a much broader family of first-order structures, and most of the terminology comes from that setting. Nonetheless, here we will focus on the case of k-graphs.

As the bold font reminds us, we should really think of $\mathbf{G}_\omega$ as a random structure, not as a single fixed structure. For instance, the question "is there

an edge between $\mathbf{v}_1$ and $\mathbf{v}_2$?" is not especially interesting, since it will depend on the specific random choices we made. More typically we are interested in questions like "what is the probability that there's an edge between $\mathbf{v}_1$ and $\mathbf{v}_2$?".

In this chapter, we will consider two frameworks for thinking about $\mathbf{G}_\omega$ as a probabilistic structure.

9.1 Exchangeable Sequences

One perspective is to view $\mathbf{G}_\omega$ as a collection of related random variables: for each $\{i_1, \ldots, i_k\} \in \binom{\mathbb{N}}{k}$, we have a random variable

$$\mathbf{X}^{(V,E,\mu_1)}_{i_1,\ldots,i_k} = \begin{cases} 1 & \text{if } (\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_k}) \in E \\ 0 & \text{otherwise} \end{cases}$$

This makes sense even in the seemingly degenerate case where $k = 1$. In this section we will state some basic definitions and (without proof) some results about sampling random countable 1-graphs, since the same ideas appear in a slightly more complicated form when $k > 1$.

A "1-graph" is just a pair (V, P) where $P \subseteq V$—that is, a set with a distinguished subset. Similarly, an atomless measurable 1-graph is a triple (V, P, μ_1) where $P \subseteq V$ is a measurable set. If we sample a random countable substructure, the corresponding random variables are just $\mathbf{X}^{(V,P,\mu_1)}_i$ for $i \in \mathbb{N}$, which is 1 or 0 depending on whether $\mathbf{v}_i \in P$.

These random variables are very simple—they're i.i.d. (independent and identically distributed) with the propery that $\mathbb{P}(\mathbf{X}^{(V,P,\mu_1)}_i = 1) = \mu_1(P)$ for each i. Independence turns out to be an artifact of the $k = 1$ case. We will see that in general the array of random variables $\{\mathbf{X}^{(V,P,\mu_1)}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\} \in \binom{\mathbb{N}}{k}}$ has two properties we will define below—it is a *dissociated exchangeable* array, and when $k = 1$ it happens that dissociated exchangeable sequences are precisely the i.i.d. sequences.

Definition 9.2. A sequence of random variables $\{\mathbf{X}_i\}_{i \in \mathbb{N}}$ is *exchangeable* if for all finite sequences $i_1, \ldots, i_n$ and $j_1, \ldots, j_n$, the sequences $(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_n})$ and $(\mathbf{X}_{j_1}, \ldots, \mathbf{X}_{j_n})$ have the same distribution.

That is, we require that $\mathbb{P}(\mathbf{X}_i = 1)$ be the same for all i, but we also require that whenever $i \neq j$, that $\mathbb{P}(\mathbf{X}_i = 1 \text{ and } \mathbf{X}_j = 1)$ be the same regardless of i and j, and so on.

There is a good intuitive reason that the variables we get by sampling a random countable substructure are exchangeable, which will generalize to $k > 1$: when we sample the vertices $\mathbf{v}_i$, there's nothing special about which index we give to which vertex—whatever we chose for vertices $\mathbf{v}_1$ and $\mathbf{v}_2$, it was just as likely that we would have chosen those to be $\mathbf{v}_3$ and $\mathbf{v}_4$ instead. This is precisely what exchangeability says: that the choice of indices for our random variables was arbitrary.

If the $\mathbf{X}_i$ are independent and identically distributed then they are certainly exchangeable. So what makes this notion weaker?

Example 9.3. We define a sequence $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ as follows. We first flip a single fair coin. If it comes up heads then each $\mathbf{X}_i$ will be an independent Bernoulli random variable which is 1 with probability 2/3. If the original coin comes up tails, each $\mathbf{X}_i$ will be an independent Bernoulli random variable which is 1 with probability 1/3.

This example is not independent. The probability that $\mathbf{X}_4$ is 1 is $1/2 \cdot 1/3 + 1/2 \cdot 2/3 = 1/2$. But if we know that $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3 = 1$, we would strongly suspect that the original coin came up heads, and therefore $\mathbb{P}(\mathbf{X}_4 = 1 \mid \mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3 = 1)$ would be close to 2/3.

But this distribution is still exchangeable; the probability

$$\mathbb{P}(\mathbf{X}_{i_1} = b_1 \text{ and } \mathbf{X}_{i_2} = b_2 \text{ and } \cdots \mathbf{X}_{i_n} = b_n)$$

does not depend on the particular indices $i_1, \ldots, i_n$.

More generally, we could start with any distribution on values in $[0, 1]$, and then let the $\mathbf{X}_i$ be Bernoulli random variables with the given probability.

Example 9.4. Let ρ be a probability distribution on $[0, 1]$. Let $\{\mathbf{X}_i^\rho\}_{i\in\mathbb{N}}$ be the sequence of random variables given by selecting $p \in [0, 1]$ according to ρ and then letting the $\mathbf{X}_i^\rho$ be independent Bernoulli random variables which are 1 with probability p. Then $\{\mathbf{X}_i^\rho\}_{i\in\mathbb{N}}$ is exchangeable.

It turns out that all exchangeable sequences of random variables have this form.

Theorem 9.5 (di Finetti's Theorem)**.** *If $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ is an exchangeable sequence of random variables, there is a distribution ρ on $[0, 1]$ such that $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ has the same distribution as the sequence $\{\mathbf{X}_i^\rho\}_{i\in\mathbb{N}}$.*

We will not prove this here since it will follow from our more general proof of the Aldous–Hoover Theorem—the generalization to $k \geq 1$—below.

We can recover the independent identically distributed sequences as those sequences where the distribution ρ chooses a single value with probability 1. This turns out to be a useful notion later, so we will give it a name.

Definition 9.6. $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ is *dissociated* if, for all n, $\{\mathbf{X}_i\}_{i\leq n}$ is independent of $\{\mathbf{X}_i\}_{i>n}$.

Together with exchangeability, being dissociated implies that for any disjoint sets S, T, $\{\mathbf{X}_i\}_{i\in S}$ is independent of $\{\mathbf{X}_i\}_{i\in T}$.

Di Finetti's Theorem implies that every exchangeable sequence is a mixture of dissociated exchangeable sequences, and that dissociated exchangeable sequences are i.i.d.. (This is not really the natural way to prove either of these facts, though—usually one would need to prove both of these directly in order to then prove di Finetti's Theorem.)

9.2 Exchangeable Arrays

Now let us consider the case of graphs, where $k = 2$. We now have an *array* of random variables $\{\mathbf{X}_{i,j}^{(V,E,\mu_1)}\}_{i,j\in\mathbb{N}}$. This array will not typically be independent—for instance, $\mathbf{X}_{1,2}$ and $\mathbf{X}_{1,3}$ both depend on our choice of $\mathbf{v}_1$. For instance, if some vertices in V have higher degree, knowing that $\mathbf{X}_{1,2} = 1$ makes it more likely that $\mathbf{v}_1$ had high degree, and therefore makes $\mathbf{X}_{1,3}$ more likely.

The array is, however, still exchangeable and dissociated. Let us now define these notions for arbitrary k.

Definition 9.7. Let $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\mathbb{N}}{k}}$ be an array of random variables. We say the array is *exchangeable* if whenever $S, T \subseteq \mathbb{N}$ are finite sets and $\pi : S \to T$ is a bijection, the collections $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{S}{k}}$ and $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\pi(S)}{k}}$ have the same joint distributions.

We say the array is *dissociated* if whenever $S, T \subseteq \mathbb{N}$ are disjoint, $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{S}{k}}$ and $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{T}{k}}$ are independent.

Exchangability seems technical, but it just states the natural property that when we change indices, we don't change probabilities; for instance,

$$\mathbb{P}(\mathbf{X}_{1,2} = 1 \text{ and } \mathbf{X}_{1,3} = 1 \text{ and } \mathbf{X}_{2,3} = 0) = \mathbb{P}(\mathbf{X}_{4,3} = 1 \text{ and } \mathbf{X}_{4,7} = 1 \text{ and } \mathbf{X}_{3,7} = 0).$$

We have to state it carefully in terms of a bijection because we do have to change indices in the same way—we typically have $\mathbb{P}(\mathbf{X}_{1,2} = 1 \text{ and } \mathbf{X}_{1,3} = 1) \neq \mathbb{P}(\mathbf{X}_{1,2} = 1 \text{ and } \mathbf{X}_{4,3} = 1)$, because exchangeability only tells us what

happens if we change all the 1's to 4's in the indices, and makes no promises if we change only some.

Theorem 9.8. *The array* $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ *coming from a countable random substructure is exchangeable and dissociated.*

Proof. To see that $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is dissociated, observe that if S and T are disjoint sets, $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{S}{k}}$ is determined by the choice of $\{\mathbf{v}_i\}_{i\in S}$ while $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{T}{k}}$ is determined by the choice of $\{\mathbf{v}_i\}_{i\in T}$. Since the $\mathbf{v}_i$ are chosen independently, the sets $\{\mathbf{v}_i\}_{i\in S}$ and $\{\mathbf{v}_i\}_{i\in T}$ are independent, so $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{S}{k}}$ and $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{T}{k}}$ are independent as well.

To see that $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is exchangeable, fix any finite sets S and T with a bijection $\pi : S \to T$. We need to show that whenever we have an array $\{b_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{S}{k}}$ of values in $\{0,1\}$, we have

$$\mathbb{P}(\text{for each } i_1,\dots,i_k \in \binom{S}{k},\ \mathbf{X}_{i_1,\dots,i_k} = b_{i_1,\dots,i_k}) = \mathbb{P}(\text{for each } i_1,\dots,i_k \in \binom{S}{k},\ \mathbf{X}_{\pi(i_1),\dots,\pi(i_k)} = b_{i_1,\dots,i_k}).$$

The probability that, for each $i_1,\dots,i_k \in \binom{S}{k}$, we have $\mathbf{X}_{i_1,\dots,i_k} = b_{i_1,\dots,i_k}$ is really a question about the vertices we chose: this probability is precisely

$$\mathbb{P}(\text{for each } i_1,\dots,i_k \in \binom{S}{k},\ \{\mathbf{v}_{i_1},\dots,\mathbf{v}_{i_k}\} \in E \text{ if and only if } b_{i_1,\dots,i_k} = 1).$$

But we can say the same thing about the other probability we care about: $\mathbb{P}(\text{for each } i_1,\dots,i_k \in \binom{S}{k},\ \mathbf{X}_{\pi(i_1),\dots,\pi(i_k)} = b_{i_1,\dots,i_k})$ is, by definition,

$$\mathbb{P}(\text{for each } i_1,\dots,i_k \in \binom{S}{k},\ \{\mathbf{v}_{\pi(i_1)},\dots,\mathbf{v}_{\pi(i_k)}\} \in E \text{ if and only if } b_{i_1,\dots,i_k} = 1).$$

And since the $\mathbf{v}_i$ are idependent and identically distributed, these probabilities are equal. □

Every dissociated exchangeable array can be obtained in this way.

Theorem 9.9. *Suppose* $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ *is exchangeable and dissociated. Then there is a measurable* k*-graph* (V,E,μ_1) *such that* $\{\mathbf{X}^{(V,E,\mu_1)}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ *has the same joint distribution as* $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$.

Proof. Let $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\mathbb{N}}{k}}$ be an exchangeable, dissociated array of random variables. We can obtain finite structures by sampling from it: let $V_n = \{1, 2, \ldots, n\}$ and let $\mathbf{E}_n \subseteq \binom{V_n}{k}$ be the set of $(i_1, \ldots, i_k)$ such that $\mathbf{X}_{i_1,\ldots,i_k} = 1$. That is, $(V_n, \mathbf{E}_n)$ is a random k-graph chosen according to $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\{1,2,\ldots,n\}}{k}}$.

We will obtain (V, E, μ_1) from an ultraproduct (of course), letting $G = (V, E) = [(V_n, \mathbf{E}_n)]_{\mathcal{U}}$. (Note that, formally, the choice of E we get depends not only on $\mathcal{U}$, but also on the specific random choices we make for each $\mathbf{E}_n$.) We then get the new exchangeable array $\{\mathbf{X}^{(V,E,\mu_1)}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\mathbb{N}}{k}}$

To show that the joint distributions are the same, it suffices to show that for every n and every $F \subseteq \binom{\{1,2,\ldots,m\}}{k}$, we have

$$\mathbb{P}(\text{for all } \vec{i} \in \binom{\{1,2,\ldots,m\}}{k},\ \mathbf{X}_{\vec{i}} = \chi_F(\vec{i})) = \mathbb{P}(\text{for all } \vec{i} \in \binom{\{1,2,\ldots,m\}}{k},\ \mathbf{X}^{(V,E,\mu_1)}_{\vec{i}} = \chi_F(\vec{i})).$$

We have

$$\mathbb{P}(\text{for all } \vec{i} \in \binom{\{1,2,\ldots,n\}}{k},\ \mathbf{X}^{(V,E,\mu_1)}_{\vec{i}} = \chi_F(\vec{i})) = t_F(E) = \lim_{\mathcal{U}} t_F(\mathbf{E}_n),$$

so it suffices to consider $t_F(\mathbf{E}_n)$ for large n.

Let $c = \mathbb{P}(\text{for all } \vec{i} \in \binom{\{1,2,\ldots,m\}}{k},\ \mathbf{X}_{\vec{i}} = \chi_F(\vec{i}))$. We follow a familiar strategy: we first will show that $\mathbb{E}(t_F(\mathbf{E}_n))$ has the correct value c, and then use a concentration inequality to show that, with high probability, $t_F(\mathbf{E}_n)$ is close to c.

For each $\pi : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$, we can consider the indicator random variable $\mathbf{1}_\pi$ which is 1 if

$$\text{for all } \vec{i} \in \binom{\{1,2,\ldots,m\}}{k},\ \mathbf{X}_{\pi(\vec{i})} = \chi_F(\vec{i})$$

and 0 otherwise. Whenever π is injective, exchangeability ensures that $\mathbb{E}(\mathbf{1}_\pi) = c$. When n is large, most of the π are injective, so $\mathbb{E}(t_F(\mathbf{E}_n))$ is close to c.

McDiarmid's inequality will not quite suffice: the variables $\mathbf{1}_\pi$ are not necessarily independent. Instead we have a slightly weaker property: when n is much larger than m, for each π we have that for *most* π', the variables $\mathbf{1}_\pi$ and $\mathbf{1}_{\pi'}$ are independent. Specifically, whenever the range of π and the range of π' are disjoint, $\mathbf{1}_\pi$ and $\mathbf{1}_{\pi'}$ are independent because $\{\mathbf{X}_{\vec{i}}\}$ is dissociated.

Although this is similar in spirit to the other concentration inequalities we have seen, there does not seem to be a standard named inequality that covers precisely this case. One can derive what we need as an application of

a very general concentration inequality, the Azuma-Hoeffding inequality, but it would require some effort for us to even state this inequality (we would have to define a martingale). Fortunately, the result we need is a special case of the one given in [91].

Theorem. *If $\mathbf{X} = \frac{1}{k}\sum_{i\leq k}\mathbf{X}_i$ where the $\mathbf{X}_i$ are random variables such such that $0 \leq \mathbf{X}_i \leq 1$ always holds and, for each i, there is a set $U_i \subseteq k$ with $|U_i| > k - d$ such that $\mathbf{X}_i$ is independent of $\{\mathbf{X}_j\}_{j\in U_i}$. Then*

$$\mathbb{P}(|\mathbf{X} - \mathbb{E}(\mathbf{X})| \geq \epsilon) \leq 2e^{-2k\epsilon^2/d}.$$

We are interested in $t_F(\mathbf{E}_n) = \frac{1}{\binom{n}{m}}\sum_\pi \mathbf{1}_\pi$, so we apply this inequality when $k = n^m$. For a given π, there are at most mn^{m-1} choices of π' so that the range of π' overlaps the range of π, so

$$\mathbb{P}(|t_F(\mathbf{E}_n) - c| \geq \epsilon) \leq 2e^{-2n\epsilon^2/m}.$$

When n is much larger than ϵ^2/m, $t_F(\mathbf{E}_n)$ is close to c with high probability, so in particular, with probability 1, $\lim_{n\to\infty} t_F(\mathbf{E}_n) = c$.

This completes the proof: we have shown that $\{\mathbf{X}^{(V,E,\mu_1)}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\mathbb{N}}{k}}$ has the same joint distribution as the original array $\{\mathbf{X}_{i_1,\ldots,i_k}\}_{\{i_1,\ldots,i_k\}\in\binom{\mathbb{N}}{k}}$. □

We finish our discussion of exchangeable arrays with the result which is really the beginning of the subject, the Aldous–Hoover Theorem, which shows that every exchangeable array can be represented by a Borel measurable function applied to independent random variables with a particular structures.

Let us first consider what Aldous–Hoover says when $k = 2$ and the array is dissociated.

Given a measurable function $f : [0,1]^3 \to \{0,1\}$, we define an array as follows. For each $i \in \mathbb{N}$ we choose $\xi_i \in [0,1]$, and for each $\{i,j\} \in \binom{\mathbb{N}}{2}$ we choose $\xi_{\{i,j\}} \in [0,1]$, with all these choices uniform and independent. Then we set

$$\mathbf{Y}_{i,j} = f(\xi_i, \xi_j, \xi_{i,j}).$$

This says that the random variable has a "unary component"—the variable ξ_i which is shared across all variables $\{\mathbf{Y}_{i,j}\}_{j\in\mathbb{N}}$—and a "binary component" unique to $\mathbf{Y}_{i,j}$.

It's important here that f is measurable in the usual sense of the product measure on $[0,1]^3$, *not* an arbitrary Keisler measure space. In a precise way,

as we will see in the proof, the ξ_i, ξ_j part corresponds to the projection onto $\mathcal{B}_{2,1}$ while the $\xi_{i,j}$ part corresponds to the quasirandom part of f.

If we drop the requirement that $\{\mathbf{X}_{i,j}\}_{\{i,j\}\in\binom{\mathbb{N}}{2}}$ be dissociated, we only need to modify the representation slightly. We take a measurable function $f : [0,1]^4 \to \{0,1\}$ and, in addition to the ξ_i and the $\xi_{i,j}$, we take a $\xi_\emptyset$ which is also chosen uniformly and independently from $[0,1]$, and we set

$$\mathbf{Y}_{i,j} = f(\xi_\emptyset, \xi_i, \xi_j, \xi_{i,j}).$$

That is, $\xi_\emptyset$ is a "global" random variable shared by all the $\mathbf{Y}_{i,j}$. A general exchangeable array is a mixture of dissociated exchangeable arrays: we can first choose $\xi_\emptyset$, which tells us which exchangeable array to use, and then we choose $\xi_i, \xi_{i,j}$ to generate the exchangeable array.

For a general k, the point is that the random variables range over the subsets of $[1,k]$.

Definition 9.10. When $f : [0,1]^{\mathcal{P}([1,k])} \to \{0,1\}$ is measurable, define an array of random variables $\{\mathbf{Y}^f_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ by choosing random variables $\{\xi_R\}_{R\in\bigcup_{i\le k}\binom{\mathbb{N}}{i}}$ uniformly and independently from $[0,1]$, and then setting

$$\mathbf{Y}^f_{i_1,\dots,i_k} = f(\{\xi_R\}_{R\subseteq\{i_1,\dots,i_k\}}).$$

Lemma 9.11. *$\{\mathbf{Y}^f_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is exchangeable.*

Proof. Essentially by definition: the finite array $\{\mathbf{Y}^f_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{S}{k}}$ is determined by $\{\xi_R\}_{R\in\bigcup_{i\le k}\binom{S}{i}}$. Since the ξ_R are choosen independently, the variables $\{\xi_{\pi(R)}\}_{R\in\bigcup_{i\le k}\binom{\pi(S)}{i}}$ have the same distribution, so $\{\mathbf{Y}^f_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\pi(S)}{k}}$ has the same distribution as well. □

Theorem 9.12. *If $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is a dissociated exchangeable array then there is a measurable function $f : [0,1]^{\mathcal{P}([1,k])\setminus\{\emptyset\}} \to \{0,1\}$ such that, taking $f' : [0,1]^{\mathcal{P}([1,k])} \to \{0,1\}$ to be $f'(\{\xi_S\}_{S\subseteq[1,k]}) = f(\{\xi_S\}_{S\subseteq[1,k],S\neq\emptyset})$, the random variable $\mathbf{Y}^{f'}_{i_1,\dots,i_k}$ has the same distribution as $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$.*

Proof. By Theorem 9.9, we have a measurable k-graph (V,E,μ_1) so that $\{\mathbf{X}^{(V,E,\mu_1)}_{i_1,\dots,I_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ has the same distribution as $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$.

(V, E, μ_1) was created using an ultraproduct, so there is a measurable ordering on V. Applying Theorem 8.48 gives us separable probability measure spaces $(\Omega_i, \mathcal{D}_i, \mu)$ and a measurable equivalence of measure algebras $\rho : V^k \to \prod_{s\subseteq\{1,2,\dots,k\},s\neq\emptyset} \Omega_i$. Since separable atomless probability measure spaces are equivalent to a subspace of the Lebesgue measurable interval $[0,1]$ [61, 331P], so we may take $\rho : V^k \to [0,1]^{2^k-1}$.

Choose $E^* \subseteq [0,1]^{2^k-1}$ with $\mu_k(\rho^{-1}(E^*) \triangle E) = 0$. Then we may take $f = \chi_{E^*}$. For any finite $H = (W, F)$ with $W \subseteq \mathbb{N}$, the probability that $\{\mathbf{Y}^{f'}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{W}{k}}$ is a copy of H (that is, that $\mathbf{Y}^{f'}_{i_1,\dots,i_k} = 1$ if and only if $\{i_1,\dots,i_k\} \in W$) is $t_H(f) = t_H(E)$, which is equal to the probability that $\{\mathbf{X}^{(V,E,\mu_1)}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{W}{k}}$ is a copy of H, which is equal to the probability that $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{W}{k}}$, so $\mathbf{Y}^{f'}_{i_1,\dots,i_k}$ has the same distribution as $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$. □

Corollary 9.13 (Aldous–Hoover Theorem)**.** *If $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is an exchangeable array then there is a measurable function $g : [0,1]^{\mathcal{P}([1,k])} \to \{0,1\}$ such that the random variable $\mathbf{Y}^g_{i_1,\dots,i_k}$ have the same distribution.*

Proof. The idea is this: if we took lots of independent samples from $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$, we could take the ultraproduct of each, and each one would gives us a function f as in the theorem. Because the array is not dissocated, these functions f will typically be distinct, so we get a distribution over functions f. Indeed, we could essentially take this to be our function g—let the $\xi_\emptyset$ variable represent our choice of a sample from $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$, and then use the function f from the previous theorem. The problem is that it's not obvious that f depends in a nice enough way on $\xi_\emptyset$.

To capture this idea, we define a related array which captures the information in our original way: define $\{\mathbf{X}'_{j,i_1,\dots,i_k}\}_{j\in\mathbb{N},\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ so that, for each j, $\{\mathbf{X}'_{j,i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is an independent copy of $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$. Note that we are not asking for this array to be exchangeable—we treat the first coordinate as an index, distinct from the other coordinates.

We will think of this as a subset of $V \times V^k$—that is, we take a sample of this array and let $V_n = \{1, 2, \dots, n\}$ and $\mathbf{E}_n \subseteq V_n \times \binom{V_n}{k}$ consist of those $(j, i_1, \dots, i_k)$ with $\mathbf{X}'_{j,i_1,\dots,i_k} = 1$. We take $(V, E) = [(V_n, \mathbf{E}_n)]_{\mathcal{U}}$, so we may view E as a subset of $V \times \binom{V}{k}$.

Since $\mathbf{X}'_{j,i_1,\dots,i_k}$ and $\mathbf{X}'_{j',i'_1,\dots,i'_k}$ are independent whenever $j \neq j'$, it follows that E is $\mathcal{B}_1 \times \mathcal{B}_k$-measurable. We can now proceed as in the preceding theorem: take $\rho : V^{k+1} \to [0,1]^{2^k}$ (where the first coordinate comes

from $\mathcal{B}_1$ and the remaining $2^k - 1$ represent $\mathcal{B}_k$), choose $E^* \subseteq [0,1]^{2^k}$ with $\mu_{k+1}(\rho^{-1}(E^*) \triangle E) = 0$, and let $f = \chi_{E^*}$.

□

9.3 Invariant Measures

Another perspective on the random array $\{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ is that we can define a related random variable $\mathbf{E}$ by setting $\{i_1,\dots,i_k\} \in \mathbf{E}$ exactly when $\mathbf{X}_{i_1,\dots,i_k} = 1$. Then $\mathbf{E}$ is a random k-graph on $\mathbb{N}$.

Formally, let us consider the space where $\mathcal{G}_k$ consisting of all k-graphs on $\mathbb{N}$. There is a natural choice of σ-algebra on $\mathcal{G}_k$: we would like each of the sets $\{E \in \mathcal{G}_k \mid \{i_1,\dots,i_k\} \in E\}$ to be measurable, so let us take $\mathcal{D}_k$ to be the smallest σ-algebra containing all these sets. To define a measure on $\mathcal{D}_k$, it would be good enough to define it on each set which a finite intersection of basic sets and complements of basic sets—that is, on sets of the form

$$\{E \in \mathcal{G}_k \mid F_1 \subseteq E \text{ and } F_0 \cap E = \emptyset\}$$

for disjoint finite sets $F_0, F_1 \subseteq \binom{\mathbb{N}}{k}$.

Definition 9.14. Let $\mathbf{X} = \{\mathbf{X}_{i_1,\dots,i_k}\}_{\{i_1,\dots,i_k\}\in\binom{\mathbb{N}}{k}}$ be an array. There is a probability measure $\mu_{\mathbf{X}}$ on $(\mathcal{G}_k, \mathcal{D}_k)$ defined by

$$\begin{aligned}&\mu_{\mathbf{X}}(\{E \in \mathcal{G}_k \mid F_1 \subseteq E \text{ and } F_0 \cap E = \emptyset)\\ =&\mathbb{P}(\text{for every } \{i_1,\dots,i_k\} \in F_1,\ \mathbf{X}_{i_1,\dots,i_k} = 1\\ &\qquad \text{and for every } \{i_1,\dots,i_k\} \in F_0,\ \mathbf{X}_{i_1,\dots,i_k} = 0\}).\end{aligned}$$

When $\mathbf{X}$ is exchangeable, this is invariant under permutations.

Theorem 9.15. *When $\mathbf{X}$ is exchangeable, $\mu_{\mathbf{X}}$ is invariant under permutations of $\mathbb{N}$.*

Proof. We need to show that when $\pi : \mathbb{N} \to \mathbb{N}$ is a permutation and B is a measurable set, that $\mu_{\mathbf{X}}(B) = \mu_{\mathbf{X}}(\pi(B))$.

It suffices to show that this holds for an algebra which generates the measurable sets, so we can restrict ourselves to finite Boolean combinations of sets of the form $\{(\mathbb{N}, E) \mid \{i_1,\dots,i_k\} \in E\}$, and to account for all these, it suffices to account for sets of the form $B = \{(\mathbb{N}, E) \mid (W, E \restriction S) = K\}$ for some finite $W \subseteq \mathbb{N}$ and some $K = (W, F)$. The claim that $\mu_{\mathbf{X}}(B) = \mu_{\mathbf{X}}(\pi(B))$ for sets of this form is a case of exchangeability. □

Note that the converse holds as well: if μ is a probability measure on $(\mathcal{G}_k, \mathcal{D}_k)$ then then we can define an array of random variables by choosing $\mathbf{E}$ according to μ and setting $\mathbf{X}_{i_1,\ldots,i_k} = 1$ when $\{i_1, \ldots, i_k\} \in \mathbf{E}$, and when μ is invariant under permutations, this array is exchangeable.

Although we did not describe them this way, we have seen examples of these kinds of measures. For instance, the random graph $\mathbf{R}_p(\mathbb{N})$ describes a measure $\mu_{\mathbf{R}_p}$ in which each of the sets $\{E \mid \{i, j\} \in E\}$ has measure p and they are all independent.

One distinctive feature of this particular measure it that it concentrates on a single graph up to isomorphism: there is a specific graph on $\mathbb{N}$ such that, with probability 1, a graph chosen according to $\mu_{\mathbf{R}_p}$ is isomorphic to it.

This graph R, called the Rado graph or, confusingly, the random graph, is characterized (up to isomorphism) by the following *extension property*: whenever H and H' are finite graphs and H is an induced sub-graph of H', every copy of H in R can be extended to a copy of H'. To make this precise, suppose $H = (W, F)$ and $H' = (W', F')$ are finite graphs, $\pi : W \to W'$ is an induced copy of H in H', and $\pi_0 : W \to \mathbb{N}$ is an induced copy of H in R; then the extension properties promises us an induced copy $\pi_1 : W' \to \mathbb{N}$ of H' in R so that $\pi_1 \circ \pi = \pi_0$. (This says not only that π_1 is a copy of H' extending π_0, but that the copy of H in the image of π_0 lies inside the image of π_1 in the specific way specified by π.)

This means that the Rado graph is the *Fraïssé limit* of the set of finite graphs: it is a countable structure which contains all finite graphs in a universal, homogeneous way. Graphs generated by exchangeable arrays are often (though not always) Fraïssé limits, and it will be useful to consider some of the properties of these limits.

9.4 Fraïssé Limits

Given a collection of finite structures with some suitable properties, we can ask whether it admits a limit analogous to the random graph.

Definition 9.16. Let K be a collection of finite k-graphs. Then K has the *hereditary property* if whenever $H \in K$ and H' is an induced sub-k-graph of H then $H' \in K$.

K has the *amalgamation property* if whenever we have $H, H_0, H_1 \in K$ and induced copies π_0, π_1 of H in H_0 and H_1, respectively, then we have an $H' \in K$ and induced copies π_0', π_1' of H_0 and H_1 in H', respectively, so that $\pi_0' \circ \pi_0 = \pi_1' \circ \pi_1$.

If K has both the hereditary property and the amalgamation property then K is a *Fraïssé class.*

The notation obscures the idea of amalgamation: we have two finite structures H_0 and H_1 which overlap on some set of elements H. Amalgamation promises that we can find a single structure H' containing both H_0 and H_1.

The collection of all graphs, for instance, has this property: take the union of H_0 and H_1. For another example, consider the collection of all *matchings*: all graphs where each vertex has degree at most 1. This time we can't just take the union—it might be that a vertex in H has a unique neighbor in H_0 and also a unique neighbor in H_1; in order to find the amalgam H', we need to identify those vertices—that is, if $h \in H$, $h_0 \in H_0$, $h_1 \in H_1$, and h_0 and h_1 are both neighbors of h, then we need $\pi_0'(h_0) = \pi_1'(h_1)$. (There is a stronger notion of amalgamation, *strong amalgamation*, where we add the requirement that $\mathrm{rng}(\pi_0') \cap \mathrm{rng}(\pi_1') = \mathrm{rng}(\pi_0' \circ \pi_0)$—that is, that the only vertices of H_0 and H_1 which get identified in H' are those in H.)

For a non-example, consider the collection of all graphs which do not contain a copy of C_5. Let H consist of two vertices and no edges, let H_0 connect those two vertices by a path of length 2, and let H_1 connect those two vertices by a path of length 3. Then none of H, H_0, or H_1 contain a cycle of length 5, but if we try to find an H' containing both H_0 and H_1, it must contain a cycle of length 5.

Definition 9.17. Let K be a Fraïssé class. We say M is a *Fraïssé limit* of K if M is a countable structure such that:

- every finite induced sub-k-graph of M is isomorphic to an element of K,
- whenever $K_0, K_1 \in K$ with K_0 an induced sub-k-graph of K_1, every induced copy of K_0 in M can be extended to an induced copy of K_1 in M.

Theorem 9.18 (Fraïssé's Theorem)**.** *When K is a Fraïssé class, a Fraïssé limit exists.*

Proof. Enumerate $K = \{K_0, K_1, \dots, \}$ in such a way that every element of K appears infinitely often. We will define a structure M on $\mathbb{N}$ as the union of a sequence of finite structures $(\{0, 1, \dots, n_i\}, E_i)$ so that $E_{i+1} \restriction \{0, 1, \dots, n_i\} = E_i$, and so that each $(\{0, 1, \dots, n_i\}, E_i)$ is an element of K.

We can let n_0 be the size of the domain of K_0 and choose E_0 so that $(\{0, 1, \dots, n_-\}, E_0)$ is a copy of K_0. Suppose we have constructed

$(\{0,1,\ldots,n_i\},E_i)$. We will construct n_{i+1},E_{i+1} as the union of a finite sequence of extensions. Let $(u_0,\pi_0,\rho_0),\ldots,(u_d,\pi_d,\rho_d)$ be the list of all triples such that

(1) $u_j \leq i$,

(2) π_j is an induced copy of K_{u_j} in K_{i+1},

(3) ρ_j is an induced copy of K_{u_j} in $(\{0,1,\ldots,n_i\},E_i)$.

We want to arrange that by the time we build $(\{0,1,\ldots,n_{i+1}\},E_{i+1})$, each ρ_j has been extended to a copy of K_{i+1}.

Let $n'_0 = n_i$ and $E'_0 = E_i$. Given $(\{0,1,\ldots,n'_j\},E'_j)$, the induced copies π_j,ρ_j are an instance of the amalgamation property, so there is a $K_v \in K$, an induced copy π'_k of K_{i+1} in K_v, and an induced copy ρ'_j of $(\{0,1,\ldots,n'_j\},E'_j)$ in K_v. In particular, we may choose n'_{j+1} to be the size of the domain of K_v and choose E'_{j+1} so that $E'_{j+1} \restriction \{0,1,\ldots,n'_j\} = E'_j$ and $(\{0,1,\ldots,n'_{j+1}\},E'_{j+1})$ is isomorphic to K_v.

Finally we take $n_{i+1} = n'_{d+1}$ and $E_{i+1} = E'_{d+1}$.

We let M be the union $(\mathbb{N},E)$ we get after infinitely many sets. If H is a finite induced sub-k-graph of M then H is an induced sub-k-graph of $(\{0,1,\ldots,n_i\},E_i)$ for some large enough i, so by the hereditary property, H is in K.

If $H,H' \in K$, ρ is an induced copy of H in M, and π is an induced copy of H in H', we have some large enough i so that $\mathrm{rng}(\rho) \subseteq \{0,1,\ldots,n_i\}$ and also $K_{i+1} = H'$ (since the former happens cofinitely and the latter happens infinitely often). We have constructed E_{i+1} precisely so that there is some induced copy π' of H' in $(\{0,1,\ldots,n_{i+1}\},E_{i+1})$ with $\pi \circ \pi = \rho$. □

Theorem 9.19. *Any two Fraïssé limits of K are isomorphic.*

Proof. Let M,N be Fraïssé limits of K. We use a classic back-and-forth argument, constructing a sequence of finite bijections. It suffices to show that if π is a partial isomorphism from a finite subset of the domain of M to a finite subset of the domain of N and a is an element of M, there is a partial isomorphism $\pi' \supseteq \pi$ with $\mathrm{dom}(\pi') = \mathrm{dom}(\pi) \cup \{a\}$. (To add some $b \in N$ to the range, we can use the fact that π^{-1} is also a partial isomorphism, use the same result to get $\pi' \supseteq \pi^{-1}$ and then $\mathrm{rng}(\pi')^{-1} = \mathrm{rng}(\pi) \cup \{b\}$.)

So suppose the partial isomorphism π is given. Then there are H,H_0 in K so that H is isomorphic to $M \restriction \mathrm{dom}(\pi)$ and H_0 is isomorphic to $M \restriction \mathrm{dom}(\pi) \cup \{a\}$. Then $\mathrm{rng}(\pi)$ is a copy of H in N, and since N is a Fraïssé limit, $\mathrm{rng}(\pi)$ can be extended to a copy of H_0 in N; we define π' by mapping a to the new element in this copy of H_0. □

We have already mentioned that the set of finite graphs is a Fraïssé class, and its Fraïssé limit is the Rado graph. More generally, it is not hard to see that the set of finite k-graphs, for any fixed k, is a Fraïssé class (for the amalgamation property, we may just take the union); there is an analogous Fraïssé limit, usually called the random k-graph. There are also variations on these notions, like the random bipartite graph (the Fraïssé limit of the set of finite bipartite graphs), and similar variants for k-graphs.

For a slightly different looking class of structures, consider the collection of triangle-free graphs—graphs with the property that, among any set of three vertices, at least one edge is missing—forms a Fraïssé class. The hereditary property follows from the definition (if K has no triangles, removing vertices will certainly not create a triangle). For the amalgamation property, if we have $H \subseteq H_0, H_1$, we can take the union of H_0 and H_1 which adds no new edges; then any set of three vertices is either totally contained in H_0 (so triangle-free because H_0 is), totally contained in H_1 (so triangle-free because H_1 is), or contains a vertex in $H_0 \setminus H_1$ and a vertex in $H_1 \setminus H_0$, and these two have no edge between them.

Note that the same argument would apply in general to the collection of k-graphs which contain no clique of m vertices, for any $k < m$.

9.5 Measures Concentrating on a Structure

The Fraïssé limit of the triangle-free graphs is called the *Henson graph*. For many years, it was an open question whether the Henson graph could be generated randomly in a way analogous to the Rado graph—that is, is there a probability measure on graphs which, with probability 1, gives a copy of the Henson graph?

More generally, we can ask which k-graphs can be randomly generated with probability 1—that is, when is there a probability measure which is invariant under permutations which concentrates on k-graphs isomorphic to a particular k-graph. The main constraint is the observation that "anything which happens must happen infinitely often". For instance, if there is a postive probability of finding a triangle then, with probability 1, the graph must contain infinitely many triangles: we can take infinitely many pairwise disjoint sets of three vertices, and since each set of three vertices has the same probability of being a triangle, independently of the others, infinitely many must be one.

We can push this slightly further: if I take any set of vertices and ask whether an extension to a particular graph happens, either the extension

never happens or it happens infinitely often. For instance, I can pick three vertices which happen to be a triangle and ask whether they extend to a copy of K_4; I can take infinitely many vertices, and each has the same probability, independently of the others, of completing the triangle to a K_4, so either this has probability 0 and so never happens, or positive probability, and therefore happens infinitely often.

Making this precise gives us the following property.

Definition 9.20. A structure $\mathcal{M}$ has *trivial group-theoretic definable closure* if whenever $S \subseteq \mathbb{N}$ is finite and $a \notin S$, there is an automorphism of $\mathcal{M}$ which fixes S and moves a.

Not all Fraïssé limits have this property. For example, let K be the collection of *matchings*: graphs in which each vertex has degree at most 1. This certainly has the hereditary property. To see the amalgamation property, note that H might have a vertex a with degree 0 but that $\pi_0(a)$ and $\pi_1(a)$ might both have degree 1. In this case H' must identify the neighbors of a with the same vertex—we have $\pi_0'(b_0) = \pi_1'(b_1)$, even though b_0, b_1 are not in the range of π_0 and π_1.

The Fraïssé limit is a graph with countably many isolated vertices and countably many pairs of matched vertices. But if we hold one end of an edge fixed, an automorphism cannot move the other end of the same edge: knowing one end of an edge is enough to "define" the other end of the same edge.

Conversely, not every structure with trivial group-theoretic definable closure is a Fraïssé limit. Consider the graph on $\mathbb{N}$ where there's an edge between any pair of distinct even vertices, and no others. That is, every odd vertex is isolated while the even vertices form a single infinite clique. This graph cannot be a Fraïssé limit: consider the *age* of this graph, the set of all finite induced sub-graphs; it consists of vertices with a single finite clique and finitely many isolated vertices. This set does not have the amalgamation property: take H to consist of a single vertex, let H_0 consist of two adjacent vertice, and let H_1 consist of two adjacent vertices together with a third isolated vertex. Let $\pi_0 : H \to H_0$ map the vertex of H to any vertex of H_0, let $\pi_1 : H \to H_1$ map the vertex of H to the isolated vertex of H_1. Then any amalgam would need two edges in distinct connected components. On the other hand, this graph does have trivial group-theoretic definable closure: we get an isomorphism by permuting the even vertices and, separately, permuting the odd vertices; given any finite set S and any a, we can find an automorphism holding S fixed and mapping a to any number of the same parity not in S.

Theorem 9.21. *If $\mathcal{M}$ is a structure and μ is a permutation invariant probability measure which gives a copy of $\mathcal{M}$ with probability 1 then $\mathcal{M}$ has trivial group-theoretic definable closure.*

Proof. Consider some fixed finite $S \subseteq \mathbb{N}$ and some $a \notin S$. For a given tuple $(m_1, \ldots, m_{|S|}, n) \in \mathbb{N}^{|S|+1}$, let us write $A_{m_1,\ldots,m_{|S|},n}$ for the event "there exists an isomorphism of $\mathbf{X}$ to $\mathcal{M}$ mapping $m_1, \ldots, m_{|S|}$ to S and n to a". Since, with probability 1, $\mathbf{X}$ is isomorphic to $\mathcal{M}$, we have $\mu(\bigcup_{m_1,\ldots,m_{|S|},n} A_{m_1,\ldots,m_{|S|},n}) = 1$, so by countable additivity, there is some $m_1, \ldots, m_{|S|}, n$ with $\mu(A_{m_1,\ldots,m_{|S|},n}) > 0$, and by permutation invariance, this holds for all $(m_1, \ldots, m_{|S|}, n)$.

For fixed $m_1, \ldots, m_{|S|}$, we have $0 < \mu(\bigcup_n A_{m_1,\ldots,m_{|S|},n}) \leq 1$. This means we cannot have $\mu(\bigcup_n A_{m_1,\ldots,m_{|S|},n}) = \sum_n A_{m_1,\ldots,m_{|S|},n}$, so there must be n, n' with $\mu(A_{m_1,\ldots,m_{|S|},n} \cap A_{m_1,\ldots,m_{|S|},n'}) > 0$.

So consider some choice of $\mathbf{X}$ in this event, and let π, π' be isomorphisms of $\mathbf{X}$ taking $m_1, \ldots, m_{|S|}$ to S and n or n' to a, respectively. Then $\pi' \circ \pi^{-1}$ is an automorphism of $\mathcal{M}$ fixing S and mapping a to something other than a.

Since this holds for any S and any $a \notin S$, $\mathcal{M}$ does not have trivial group-theoretic definable closure. □

In order to prove a converse, it is helpful to work directly with the way automorphisms move elements around.

Definition 9.22. Given a structure $\mathcal{M}$ with universe $\mathbb{N}$, write $[\mathcal{M}]_d$ for the collection of equivalence classes of $\mathbb{N}^d$ under the automorphisms of $\mathcal{M}$.

We can think of $[\mathcal{M}]_d$ as representing a notion of "type" of a d-tuple, where two tuples have the same type exactly when one can be mapped to the other by an automorphism. (This is a bit stronger than the usual model-theoretic notion of a type—more precisely, it's roughly a type in the logic $\mathcal{L}_{\omega_1,\omega}$.) Note that each element of $[\mathcal{M}]_k$ determines E—that is, each element of $[\mathcal{M}]_k$ is either an equivalence class of edges or an equivalence class of non-edges—since an automorphism certainly preserves whether there's an edge between a given k-tuple.

We are skirting the bounds of what it is reasonable to do while working only with k-graphs. We should really think of this as the passage to a new structure, $\mathcal{M}'$, in an expanded language of first-order logic containing a relation symbol for each element of each $[\mathcal{M}]_d$. This new structure continues to have trivial group-theoretic definable closure, and additionally is *ultrahomogeneous*—whenever we have a partial automorphism between finite subsets, it can be extended to an automorphism of the whole structure.

Still, we can state this fact without introducing more abstraction.

Lemma 9.23. *Let $d < d_0, d_1$ be given, and suppose we have $(n_1, \ldots, n_d) \in \mathbb{N}^d$, and two tuples $(n_1, \ldots, n_d, m^0_1, \ldots, m^0_{d_0-d}) \in \mathbb{N}^{d_0}$ and $(n_1, \ldots, n_d, m^1_1, \ldots, m^1_{d_1-d}) \in \mathbb{N}^{d_1}$.*

There exists a tuple $(n_1, \ldots, n_d, m^2_1, \ldots, m^2_{d_1-d}) \in \mathbb{N}^{d_1}$ which has the same d_1-type as $(n_1, \ldots, n_d, m^1_1, \ldots, m^1_{d_1-d})$ and so that $\{m^0_1, \ldots, m^0_{d_0-d}\} \cap \{m^2_1, \ldots, m^2_{d_1-d}\} = \emptyset$.

This is essentially an almagamation theorem for types: it says that if p is a sub-type of both p_0 and p_1 then there is a type p_2 which amalgamates p_0 and $p_1$without identifying any elements of $p_0 \setminus p$ with elements of $p_1 \setminus p$. In this guise, it is a standard fact from model theory—that a when a structure has both trivial group-theoretic definable closure and is ultrahomogeneous, it has the *strong amalgamation* property, which is the strengthening of the amalgamation property in which the copies of H_0 and H_1 in H' do not share any elements other than those in H.

Stating this directly as amalgamation about types, however, would require some sort of notation for writing embeddings of types. Rather than introduce this notation for one application, we have stated this more bare bones version which suffices for us.

Proof. Let $N = \{n_1, \ldots, n_d\}$, $M^0 = \{m^0_1, \ldots, m^0_{d_1-d}\}$, and $M^1 = \{m^1_1, \ldots, m^1_{d_1-d}\}$. We show the claim by induction on $|M^0 \cap M^1|$. If $M^0 \cap M^1 = \emptyset$, we may simply take $M^2 = M^1$ and be done.

Otherwise, choose some $m \in M^0 \cap M^1$. Since $\mathcal{M}$ has trivial group-theoretic definable closure, we may find an automorphism fixing N and $M^0 \cup M^1 \setminus \{m\}$ and mapping m to some m'. Then $m' \notin M^0$, so $N \cup (M^1 \setminus \{m\}) \cup \{m'\}$ has the same d_1-type as $(n_1, \ldots, n_d, m^1_1, \ldots, m^1_{d_1-d})$, but intersects M^0 in one fewer point. The inductive hypothesis then gives the desired M^2. □

We can now prove a converse to Theorem 9.21, giving us an exact characterization of those structures which can be generated with probability 1 by a permutation invariant probability measure.

Theorem 9.24 (Ackerman–Freer–Patel)**.** *Suppose $\mathcal{M}$ has trivial definable group-theoretic closure. Then there is a permutation invariant measure μ which, with probability* 1*, concentrates on structures isomorphic to $\mathcal{M}$.*

The idea is that we'll build our measure by deciding how to assign d-tuples to elements of $\mathcal{M}$: we'll build our measure in stages where, at a given stage, we've divided $[0, 1]$ into intervals $I_1, \ldots, I_d$, and whenever $(x_1, \ldots, x_d)$ is a

tuple with $x_i \in I_i$ for all i, we will have chosen an element of $\mathcal{M}_d$ to assign $(x_1, \ldots, x_d)$.

Proof. We fix a list of all the "one-point extensions"—that is, all pairs (p, p') where $p \in [\mathcal{M}]_d$, $p' \in [\mathcal{M}]_{d+1}$—so that all such pairs appear infinitely often.

At each stage, we will have a disjoint list of intervals $I_1 < \ldots < I_s$ of $[0, 1]$ so that $\max I_s < 1$, and a type p in $[\mathcal{M}]_d$, and we have committed that a tuple $(x_1, \ldots, x_d)$ with $x_i \in I_i$ for all i will have type p.

At infinitely many stages, we refine an interval. Let j be such that $|I_j|$ is largest and, if there are multiple such j, so j is least among such intervals. We divide I_j into two sub-intervals (say, by dividing in half). Consider any representative $(n_1, \ldots, n_s)$ of the chosen type p. Fixing all elements other than n_j, there is an automorphism of $\mathcal{M}$ moving n_j to some n'_j. Let p' be the type of $(n_1, \ldots, n_{j-1}, n_j, n'_j, n_{j+1}, \ldots, n_s)$; we decide that a tuple from from $I_1, \ldots, I_j/2, I_j/2, \ldots, I_s$ will be assigned to p'.

At infinitely many stages we arrange an extension. We choose the next tuple (p, p') our list. Let v be the number of sub-tuples of p_s which are p. We take the interval $[\max I_s, 1]$ and divide it into v sub-intervals $I_{s+1}, \ldots, I_{s+v}$. For each I_{s+j}, we want to ensure that the j-th interval extends the j-th tuple to a copy of p'. By strong amalgamation, we can then amalgamate these all to find a p_{s+1}.

The way we have defined our measure, it is only defined up to measure 0 (a countable set of points are boundaries of intervals); we can define the measure arbitrarily on tuples containing a boundary.

Now suppose we choose some structure $\mathcal{N}$ according to μ. We must show that, with probability 1, it is isomorphic to $\mathcal{M}$. We use a back-and-forth argument. Suppose we have constructed a finite automorphism $\pi : \mathcal{M} \to \mathcal{N}$. Let $(x_1, \ldots, x_d)$ be the range of $\mathcal{N}$, in increasing order.

Given any $a \in \mathbb{N} \setminus \mathrm{dom}(\pi)$, consider the type of $\mathrm{dom}(\pi) \cup \{a\}$. At some later stage, we considered $\mathrm{dom}(\pi), \mathrm{dom}(\pi) \cup \{a\}$, and at that stage, we added an interval so that any element from that interval would give the desired type; with probability 1, $\mathcal{N}$ contains that element as an interval, so we may choose that for $\pi(a)$.

Given any $b \in |\mathcal{N}| \setminus \{\mathrm{rng}(\pi)\}$, we chose $(x_1, \ldots, x_d, b)$ from some intervals, which were assigned to some type in $[\mathcal{M}]_{d+1}$. Since this type extends the type of $\mathrm{dom}(\pi)$, there must be an element of $\mathcal{M}$ realizing the type, so we set $\pi(a) = b$. □

9.6 Remarks

The study of exchangeable array, in their current generality begins with [4], which proves the $k = 2$ case of the Aldous–Hoover Theorem. The general case was settled in [86], using model theoretic methods similar to those we have used here. Many subsequent results on exchangeable arrays and variations on the idea are described in [95]. The connection with graph limits was noted in [43].

The construction of a probability measure concentrating on the Henson graph is from [127], which was then generalized to the construction given here in [1].

There is a connection between the properties of Fraïssé classes and the properties of measures which are invariant under the automorphisms of the Fraïssé limit. If G is the group automorphisms (or, indeed, any group), a G-invariant measure is a G-flow. The study of these is a large topic we cannot begin to survey here, but one fundamental result is [96], which connects the Ramsey properties of the Fraïssé class (that is, the question of whether one can "color" substructures of the Fraïssé limit and then find a copy of the Fraïssé limit which is monochromatic) to the extreme amenability of the G-flow.

Index of Notation

Index

Bibliography

[1] N. Ackerman, C. Freer, and R. Patel. "Invariant measures concentrated on countable structures". In: *Forum Math. Sigma* 4 (2016), e17, 59. ISSN: 2050-5094 (cit. on p. 273).

[2] N. Ackerman, C. Freer, and R. Patel. "Stable regularity for relational structures". In: *arXiv preprint arXiv:1712.09305* (2017) (cit. on p. 222).

[3] E. Aigner-Horev et al. "Quasirandomness in hypergraphs". In: *Electron. J. Combin.* 25.3 (2018), Paper No. 3.34, 22 (cit. on p. 253).

[4] D. J. Aldous. "Representations for partially exchangeable arrays of random variables". In: *J. Multivariate Anal.* 11.4 (1981), pp. 581–598. ISSN: 0047-259X (cit. on pp. 253, 273).

[5] N. Alon and A. Shapira. "A characterization of the (natural) graph properties testable with one-sided error". In: *SIAM J. Comput.* 37.6 (2008), pp. 1703–1727. ISSN: 0097-5397 (cit. on p. 195).

[6] N. Alon and A. Shapira. "Testing subgraphs in directed graphs". In: *J. Comput. System Sci.* 69.3 (2004), pp. 353–382. ISSN: 0022-0000 (cit. on p. 195).

[7] N. Alon and J. H. Spencer. *The probabilistic method*. Fourth. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2016, pp. xiv+375. ISBN: 978-1-119-06195-3 (cit. on p. 46).

[8] N. Alon et al. "Efficient testing of large graphs". In: *Combinatorica* 20.4 (2000), pp. 451–476. ISSN: 0209-9683 (cit. on pp. 182, 195).

[9] N. Alon et al. "Scale-sensitive dimensions, uniform convergence, and learnability". In: *J. ACM* 44.4 (1997), pp. 615–631. ISSN: 0004-5411 (cit. on p. 222).

[10] J. Avigad and J. Helzner. "Transfer principles in nonstandard intuitionistic arithmetic". In: *Arch. Math. Logic* 41.6 (2002), pp. 581–602. ISSN: 0933-5846 (cit. on p. 151).

[11] J. Avigad, P. Gerhardy, and H. Towsner. "Local stability of ergodic averages". In: *Trans. Amer. Math. Soc.* 362.1 (2010), pp. 261–288. ISSN: 0002-9947 (cit. on p. 195).

[12] J. Avigad and J. Iovino. "Ultraproducts and metastability". In: *New York J. Math.* 19 (2013), pp. 713–727. ISSN: 1076-9803 (cit. on p. 195).

[13] J. Avigad and H. Towsner. "Metastability in the Furstenberg-Zimmer tower". In: *Fund. Math.* 210.3 (2010), pp. 243–268. ISSN: 0016-2736 (cit. on p. 7).

[14] I. Ben Yaacov et al. "Model theory for metric structures". In: *Model theory with applications to algebra and analysis. Vol. 2*. Ed. by Z. Chatzidakis et al. Vol. 350. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2008, pp. 315–427. ISBN: 978-0-521-70908-8 (cit. on p. 82).

[15] B. van den Berg and S. Sanders. "Reverse mathematics and parameter-free transfer". In: *Ann. Pure Appl. Logic* 170.3 (2019), pp. 273–296. ISSN: 0168-0072 (cit. on p. 151).

[16] V. Bergelson et al., eds. *Ultrafilters across mathematics*. Vol. 530. Contemporary Mathematics. Papers from the International Congress on Applications of Ultrafilters and Ultraproducts in Mathematics (ULTRAMATH 2008) held in Pisa, June 1–7, 2008. American Mathematical Society, Providence, RI, 2010, pp. x+200. ISBN: 978-0-8218-4833-3 (cit. on p. 81).

[17] A. Blass. "A model without ultrafilters". In: *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* 25.4 (1977), pp. 329–331. ISSN: 0001-4117 (cit. on p. 59).

[18] B. Bollobás. *Random graphs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1985, pp. xvi+447. ISBN: 0-12-111755-3; 0-12-111756-1 (cit. on p. 46).

[19] B. Bollobás and P. Erdős. "On the structure of edge graphs". In: *Bull. London Math. Soc.* 5 (1973), pp. 317–321. ISSN: 0024-6093 (cit. on p. 113).

[20] M. Briański, P. Micek, and M. Pilipczuk Michałand Seweryn. "Erdös-Hajnal properties for powers of sparse graphs". In: *SIAM J. Discrete Math.* 35.1 (2021), pp. 447–464. ISSN: 0895-4801 (cit. on p. 220).

[21] B. Chazelle. *The discrepancy method*. Randomness and complexity. Cambridge University Press, Cambridge, 2000, pp. xviii+463. ISBN: 0-521-77093-9 (cit. on p. 47).

[22] A. Chernikov, D. Galvin, and S. Starchenko. "Cutting lemma and Zarankiewicz's problem in distal structures". In: *Selecta Math. (N.S.)* 26.2 (2020), Paper No. 25, 27. ISSN: 1022-1824 (cit. on p. 218).

[23] A. Chernikov, D. Palacin, and K. Takeuchi. "On n-dependence". In: *Notre Dame J. Form. Log.* 60.2 (2019), pp. 195–214. ISSN: 0029-4527 (cit. on p. 253).

[24] A. Chernikov and P. Simon. "Externally definable sets and dependent pairs II". In: *Trans. Amer. Math. Soc.* 367.7 (2015), pp. 5217–5235. ISSN: 0002-9947 (cit. on p. 220).

[25] A. Chernikov and S. Starchenko. "A note on the Erdős-Hajnal property for stable graphs". In: *Proc. Amer. Math. Soc.* 146.2 (2018), pp. 785–790. ISSN: 0002-9939 (cit. on p. 222).

[26] A. Chernikov and S. Starchenko. "Definable regularity lemmas for NIP hypergraphs". In: *arXiv preprint arXiv:1607.07701* (2016) (cit. on p. 222).

[27] A. Chernikov and S. Starchenko. "Regularity lemma for distal structures". In: *J. Eur. Math. Soc. (JEMS)* 20.10 (2018), pp. 2437–2466. ISSN: 1435-9855 (cit. on pp. 219, 222).

[28] A. Chernikov and H. Towsner. "Hypergraph regularity with forbidden patterns". in preparation (cit. on p. 253).

[29] F. R. K. Chung and R. L. Graham. "Quasi-random hypergraphs". In: *Random Structures Algorithms* 1.1 (1990), pp. 105–124. ISSN: 1042-9832 (cit. on p. 253).

[30] F. R. K. Chung, R. L. Graham, and R. M. Wilson. "Quasi-random graphs". In: *Combinatorica* 9.4 (1989), pp. 345–362. ISSN: 0209-9683 (cit. on pp. 47, 151).

[31] F. R. K. Chung. "Quasi-random classes of hypergraphs". In: *Random Structures Algorithms* 1.4 (1990), pp. 363–382. ISSN: 1042-9832 (cit. on p. 253).

[32] F. R. K. Chung. "Regularity lemmas for hypergraphs and quasi-randomness". In: *Random Structures Algorithms* 2.2 (1991), pp. 241–252. ISSN: 1042-9832 (cit. on p. 253).

[33] F. R. K. Chung. *Spectral graph theory*. Vol. 92. CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997, pp. xii+207. ISBN: 0-8218-0315-8 (cit. on p. 47).

[34] V. Chvátal and E. Szemerédi. "Notes on the Erdős-Stone theorem". In: *Combinatorial mathematics (Marseille-Luminy, 1981)*. Vol. 75. North-Holland Math. Stud. North-Holland, Amsterdam, 1983, pp. 183–190 (cit. on p. 113).

[35] G. Conant, A. Pillay, and C. Terry. "A group version of stable regularity". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press. 2018, pp. 1–9 (cit. on p. 222).

[36] D. Conlon and J. Fox. "Graph removal lemmas". In: *Surveys in combinatorics 2013*. Vol. 409. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2013, pp. 1–49 (cit. on p. 195).

[37] D. Conlon, J. Fox, and B. Sudakov. "An approximate version of Sidorenko's conjecture". In: *Geom. Funct. Anal.* 20.6 (2010), pp. 1354–1366. ISSN: 1016-443X (cit. on pp. 47, 151).

[38] L. N. Coregliano and A. A. Razborov. "Semantic limits of dense combinatorial objects". In: *Uspekhi Mat. Nauk* 75.4(454) (2020), pp. 45–152. ISSN: 0042-1316 (cit. on p. 253).

[39] L. N. Coregliano and M. Malliaris. *Countable Ramsey*. 2022 (cit. on p. 222).

[40] L. N. Coregliano and A. A. Razborov. *Natural quasirandomness properties*. 2020. eprint: `arXiv:2012.11773` (cit. on pp. 247, 253).

[41] M. Di Nasso, I. Goldbring, and M. Lupini. "Nonstandard Methods in Ramsey Theory and Combinatorial Number Theory". In: *ArXiv e-prints* (Sept. 2017). eprint: `1709.04076` (math.CO) (cit. on p. 81).

[42] M. Di Nasso. "Hypernatural numbers as ultrafilters". In: *Nonstandard analysis for the working mathematician*. Springer, Dordrecht, 2015, pp. 443–474 (cit. on p. 81).

[43] P. Diaconis and S. Janson. "Graph limits and exchangeable random graphs". In: *Rend. Mat. Appl. (7)* 28.1 (2008), pp. 33–61. ISSN: 1120-7183 (cit. on p. 273).

[44] R. M. Dudley. "A course on empirical processes". In: *École d'été de probabilités de Saint-Flour, XII—1982*. Vol. 1097. Lecture Notes in Math. Papers from the summer school held in Saint-Flour, August 22–September 8, 1982, Edited by P. L Hennequin. Berlin: Springer, 1984, pp. 1–142. ISBN: 3-540-13897-8 (cit. on p. 222).

[45] R. M. Dudley, E. Giné, and J. Zinn. "Uniform and universal Glivenko-Cantelli classes". In: *J. Theoret. Probab.* 4.3 (1991), pp. 485–510. ISSN: 0894-9840 (cit. on p. 222).

[46] E. Dueñez and J. Iovino. "Model theory and metric convergence I: Metastability and dominated convergence". In: *Beyond First Order Model Theory*. Chapman and Hall/CRC, 2017 (cit. on p. 195).

[47] G. Elek and B. Szegedy. "Limits of Hypergraphs, Removal and Regularity Lemmas. A Non-standard Approach". unpublished. 2007 (cit. on p. 253).

[48] G. Elek and B. Szegedy. "A measure-theoretic approach to the theory of dense hypergraphs". In: *Adv. Math.* 231.3-4 (2012), pp. 1731–1772. ISSN: 0001-8708 (cit. on pp. 189, 195).

[49] H. B. Enderton. *A mathematical introduction to logic*. Second. Harcourt/Academic Press, Burlington, MA, 2001, pp. xii+317. ISBN: 0-12-238452-0 (cit. on p. 81).

[50] P. Erdős and J. Spencer. "Imbalances in k-colorations". In: *Networks* 1 (1971/72), pp. 379–385. ISSN: 0028-3045 (cit. on p. 47).

[51] P. Erdös and A. H. Stone. "On the structure of linear graphs". In: *Bull. Amer. Math. Soc.* 52 (1946), pp. 1087–1091. ISSN: 0002-9904 (cit. on p. 113).

[52] V. Falgas-Ravry and E. R. Vaughan. "On applications of Razborov's flag algebra calculus to extremal 3-graph theory". In: *ArXiv e-prints* (Oct. 2011). arXiv: `1110.1623 [math.CO]` (cit. on p. 130).

[53] V. Falgas-Ravry and E. R. Vaughan. "Turán H-densities for 3-graphs". In: *Electron. J. Combin.* 19.3 (2012), Paper 40, 26. ISSN: 1077-8926 (cit. on p. 130).

[54] S. Feferman. "Some applications of the notions of forcing and generic sets". In: *Fund. Math.* 56 (1964/1965), pp. 325–345. ISSN: 0016-2736 (cit. on p. 59).

[55] P. Fletcher. "Nonstandard set theory". In: *J. Symbolic Logic* 54.3 (1989), pp. 1000–1008. ISSN: 0022-4812 (cit. on p. 151).

[56] J. Fox. "A new proof of the graph removal lemma". In: *Ann. of Math. (2)* 174.1 (2011), pp. 561–579. ISSN: 0003-486X (cit. on p. 195).

[57] J. Fox and L. M. Lovász. "A tight lower bound for Szemerédi's regularity lemma". In: *Combinatorica* 37.5 (2017), pp. 911–951. ISSN: 0209-9683 (cit. on p. 194).

[58] J. Fox, J. Pach, and A. Suk. "A polynomial regularity lemma for semialgebraic hypergraphs and its applications in geometry and property testing". In: *SIAM J. Comput.* 45.6 (2016), pp. 2199–2223. ISSN: 0097-5397 (cit. on p. 222).

[59] J. Fox et al. "Overlap properties of geometric expanders". In: *J. Reine Angew. Math.* 671 (2012), pp. 49–83. ISSN: 0075-4102 (cit. on p. 222).

[60] P. Frankl and V. Rödl. "Extremal problems on set systems". In: *Random Structures Algorithms* 20.2 (2002), pp. 131–164. ISSN: 1042-9832 (cit. on p. 253).

[61] D. H. Fremlin. *Measure theory. Vol. 3*. Measure algebras, Corrected second printing of the 2002 original. Torres Fremlin, Colchester, 2004, pp. 693+13. ISBN: 0-9538129-3-6 (cit. on pp. 192, 263).

[62] A. Frieze and R. Kannan. "Quick approximation to matrices and applications". In: *Combinatorica* 19.2 (1999), pp. 175–220. ISSN: 0209-9683 (cit. on p. 195).

[63] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016, pp. xvii+464. ISBN: 978-1-107-11850-8 (cit. on p. 46).

[64] Z. Füredi and J. Pach. "Traces of finite sets: extremal problems and geometric applications". In: *Extremal problems for finite sets (Visegrád, 1991)*. Vol. 3. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, 1994, pp. 251–282 (cit. on p. 222).

[65] H. Furstenberg, Y. Katznelson, and D. Ornstein. "The ergodic theoretical proof of Szemerédi's theorem". In: *Bull. Amer. Math. Soc. (N.S.)* 7.3 (1982), pp. 527–552. ISSN: 0273-0979 (cit. on p. 7).

[66] H. Furstenberg. "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions". In: *J. Analyse Math.* 31 (1 1977). 10.1007/BF02813304, pp. 204–256. ISSN: 0021-7670 (cit. on pp. 7, 105).

[67] D. García. "A note on pseudofinite dimension and forking". In: *arXiv preprint arXiv:1402.5212* (2014) (cit. on p. 222).

[68] D. García, D. Macpherson, and C. Steinhorn. "Pseudofinite structures and simplicity". In: *J. Math. Log.* 15.1 (2015), pp. 1550002, 41. ISSN: 0219-0613 (cit. on p. 222).

[69] A. Gelman and D. Nolan. "You can load a die, but you can't bias a coin". In: *The American Statistician* 56.4 (2002), pp. 308–311 (cit. on p. 31).

[70] R. Goldblatt. *Lectures on the hyperreals*. Vol. 188. Graduate Texts in Mathematics. An introduction to nonstandard analysis. Springer-Verlag, New York, 1998, pp. xiv+289. ISBN: 0-387-98464-X (cit. on pp. 76, 81, 105, 151).

[71] I. Goldbring and H. Towsner. "An approximate logic for measures". English. In: *Israel Journal of Mathematics* 199.2 (2014), pp. 867–913. ISSN: 0021-2172 (cit. on pp. 7, 130).

[72] I. Goldbring and H. Towsner. "Dividing and weak quasi-dimensions in arbitrary theories". In: *Arch. Math. Logic* 54.7-8 (2015), pp. 915–920. ISSN: 0933-5846 (cit. on p. 222).

[73] O. Goldreich. "Introduction to testing graph properties". In: *Studies in complexity and cryptography*. Vol. 6650. Lecture Notes in Comput. Sci. Springer, Heidelberg, 2011, pp. 470–506 (cit. on p. 194).

[74] A. W. Goodman. "On sets of acquaintances and strangers at any party". In: *Amer. Math. Monthly* 66 (1959), pp. 778–783. ISSN: 0002-9890 (cit. on pp. 113, 116).

[75] W. T. Gowers. "A new proof of Szemerédi's theorem". In: *Geom. Funct. Anal.* 11.3 (2001), pp. 465–588. ISSN: 1016-443X (cit. on pp. 7, 151).

[76] W. T. Gowers. "Hypergraph regularity and the multidimensional Szemerédi theorem". In: *Ann. of Math. (2)* 166.3 (2007), pp. 897–946. ISSN: 0003-486X (cit. on pp. 7, 253).

[77] W. T. Gowers. "Lower bounds of tower type for Szemerédi's uniformity lemma". In: *Geom. Funct. Anal.* 7.2 (1997), pp. 322–337. ISSN: 1016-443X (cit. on p. 194).

[78] W. T. Gowers. "Quasirandomness, counting and regularity for 3-uniform hypergraphs". In: *Combin. Probab. Comput.* 15.1-2 (2006), pp. 143–184. ISSN: 0963-5483 (cit. on p. 253).

[79] S. W. Graham and C. J. Ringrose. "Lower bounds for least quadratic nonresidues". In: *Analytic number theory (Allerton Park, IL, 1989)*. Vol. 85. Progr. Math. Birkhäuser Boston, Boston, MA, 1990, pp. 269–309 (cit. on p. 46).

[80] B. Green and T. Tao. "Yet another proof of Szemerédi's theorem". In: *An irregular mind*. Vol. 21. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, 2010, pp. 335–342 (cit. on p. 7).

[81] N. G. Hadžiivanov and V. S. Nikiforov. "The Nordhaus-Stewart-Moon-Moser inequality". In: *Serdica* 4.4 (1978), pp. 344–350. ISSN: 0204-4110 (cit. on p. 116).

[82] H. Hatami. "Graph norms and Sidorenko's conjecture". In: *Israel J. Math.* 175 (2010), pp. 125–150. ISSN: 0021-2172 (cit. on pp. 47, 151).

[83] C. W. Henson and J. Iovino. "Ultraproducts in analysis". In: *Analysis and logic (Mons, 1997)*. Vol. 262. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2002, pp. 1–110 (cit. on p. 82).

[84] N. Hindman and D. Strauss. *Algebra in the Stone-Čech compactification*. De Gruyter Textbook. Theory and applications, Second revised and extended edition [of MR1642231]. Walter de Gruyter & Co., Berlin, 2012, pp. xviii+591. ISBN: 978-3-11-025623-9 (cit. on p. 81).

[85] W. Hodges. *A shorter model theory*. Cambridge University Press, Cambridge, 1997, pp. x+310. ISBN: 0-521-58713-1 (cit. on p. 209).

[86] D. Hoover. *Relations on Probability Spaces and Arrays of Random Variables*. Preprint. Institute for Advanced Study, Princeton, NJ, 1979 (cit. on pp. 105, 253, 273).

[87] B. Host and B. Kra. "Nonconventional ergodic averages and nilmanifolds". In: *Ann. of Math. (2)* 161.1 (2005). 10.4007/annals.2005.161.397, pp. 397–488. ISSN: 0003-486X (cit. on p. 7).

[88] P. Howard and J. E. Rubin. *Consequences of the axiom of choice*. Vol. 59. Mathematical Surveys and Monographs. With 1 IBM-PC floppy disk (3.5 inch; WD). American Mathematical Society, Providence, RI, 1998, pp. viii+432. ISBN: 0-8218-0977-6 (cit. on p. 59).

[89] E. Hrushovski. "On pseudo-finite dimensions". In: *Notre Dame J. Form. Log.* 54.3-4 (2013), pp. 463–495. ISSN: 0029-4527 (cit. on p. 222).

[90] E. Hrushovski. "Stable group theory and approximate subgroups". In: *J. Amer. Math. Soc.* 25.1 (2012), pp. 189–243. ISSN: 0894-0347 (cit. on pp. 130, 222).

[91] S. Janson. "Large deviations for sums of partly dependent random variables". In: *Random Structures Algorithms* 24.3 (2004), pp. 234–248. ISSN: 1042-9832 (cit. on p. 261).

[92] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000, pp. xii+333. ISBN: 0-471-17541-2 (cit. on p. 46).

[93] T. Jech. *Set theory*. Springer Monographs in Mathematics. The third millennium edition, revised and expanded. Springer-Verlag, Berlin, 2003, pp. xiv+769. ISBN: 3-540-44085-2 (cit. on p. 81).

[94] Y. Jiang et al. "Regular partitions of gentle graphs". In: *Acta Math. Hungar.* 161.2 (2020), pp. 719–755. ISSN: 0236-5294 (cit. on p. 220).

[95] O. Kallenberg. *Probabilistic symmetries and invariance principles*. Probability and its Applications (New York). New York: Springer, 2005, pp. xii+510. ISBN: 978-0387-25115-8; 0-387-25115-4 (cit. on p. 273).

[96] A. S. Kechris, V. G. Pestov, and S. Todorcevic. "Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups". In: *Geom. Funct. Anal.* 15.1 (2005), pp. 106–189. ISSN: 1016-443X (cit. on p. 273).

[97] H. J. Keisler. "Probability quantifiers". In: *Model-theoretic logics*. Ed. by J. Barwise and S. Feferman. Perspect. Math. Logic. Springer, New York, 1985, pp. 509–556. ISBN: 0-387-90936-2 (cit. on p. 105).

[98] Y. Kohayakawa, V. Rödl, and J. Skokan. "Hypergraphs, quasi-randomness, and conditions for regularity". In: *J. Combin. Theory Ser. A* 97.2 (2002), pp. 307–352. ISSN: 0097-3165 (cit. on p. 253).

[99] U. Kohlenbach and L. Leuştean. "A quantitative mean ergodic theorem for uniformly convex Banach spaces". In: *Ergodic Theory Dynam. Systems* 29.6 (2009), pp. 1907–1915. ISSN: 0143-3857 (cit. on p. 195).

[100] U. Kohlenbach. "Some logical metatheorems with applications in functional analysis". In: *Trans. Amer. Math. Soc.* 357.1 (2005), 89–128 (electronic). ISSN: 0002-9947 (cit. on p. 195).

[101] U. Kohlenbach and P. Safarik. "Fluctuations, effective learnability and metastability in analysis". In: *Ann. Pure Appl. Logic* 165.1 (2014), pp. 266–304. ISSN: 0168-0072 (cit. on p. 195).

[102] J. Komlós and M. Simonovits. "Szemerédi's Regularity Lemma and its applications in graph theory". In: *Combinatorics, Paul Erdős is Eighty, Vol 2*. Ed. by D. Miklos et al. János Bolyai Mathematical Society, 1996, pp. 295–352 (cit. on p. 194).

[103] J. Komlós et al. "The regularity lemma and its applications in graph theory". In: *Theoretical aspects of computer science (Tehran, 2000)*. Vol. 2292. Lecture Notes in Comput. Sci. Springer, Berlin, 2002, pp. 84–112 (cit. on p. 194).

[104] M. Krivelevich and B. Sudakov. "Pseudo-random graphs". In: *More sets, graphs and numbers*. Vol. 15. Bolyai Soc. Math. Stud. Springer, Berlin, 2006, pp. 199–262 (cit. on p. 47).

[105] K. Kunen. *Set theory*. Vol. 34. Studies in Logic (London). College Publications, London, 2011, pp. viii+401. ISBN: 978-1-84890-050-9 (cit. on p. 81).

[106] M. C. Laskowski. "Vapnik-Chervonenkis classes of definable sets". In: *J. London Math. Soc. (2)* 45.2 (1992), pp. 377–384. ISSN: 0024-6107 (cit. on p. 222).

[107] J. Lenz and D. Mubayi. "The poset of hypergraph quasirandomness". In: *Random Structures Algorithms* 46.4 (2015), pp. 762–800. ISSN: 1042-9832 (cit. on p. 253).

[108] P. A. Loeb. "Conversion from nonstandard to standard measure spaces and applications in probability theory". In: *Trans. Amer. Math. Soc.* 211 (1975), pp. 113–122. ISSN: 0002-9947 (cit. on p. 105).

[109] L. Lovász and M. Simonovits. "On the number of complete subgraphs of a graph. II". In: *Studies in pure mathematics*. Birkhäuser, Basel, 1983, pp. 459–495 (cit. on p. 116).

[110] L. Lovász. *Large networks and graph limits*. Vol. 60. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2012, pp. xiv+475. ISBN: 978-0-8218-9085-1 (cit. on p. 195).

[111] L. Lovász and B. Szegedy. "Regularity partitions and the topology of graphons". In: *An irregular mind*. Ed. by I. Bárány and J. Solymosi. Vol. 21. Bolyai Soc. Math. Stud. Szemerédi is 70. Budapest: János Bolyai Math. Soc., 2010, pp. 415–446. ISBN: 978-963-9453-14-2; 978-3-642-14443-1 (cit. on p. 222).

[112] D. Maharam. "On homogeneous measure algebras". In: *Proc. Nat. Acad. Sci. U.S.A.* 28 (1942), pp. 108–111. ISSN: 0027-8424 (cit. on p. 189).

[113] M. Makkai. "Stone duality for first order logic". In: *Adv. in Math.* 65.2 (1987), pp. 97–170. ISSN: 0001-8708 (cit. on p. 74).

[114] M. Malliaris and S. Shelah. "Regularity lemmas for stable graphs". In: *Trans. Amer. Math. Soc.* 366.3 (2014), pp. 1551–1585. ISSN: 0002-9947 (cit. on p. 222).

[115] M. Malliaris and A. Pillay. "The stable regularity lemma revisited". In: *Proc. Amer. Math. Soc.* 144.4 (2016), pp. 1761–1765. ISSN: 0002-9939 (cit. on p. 222).

[116] J. Matoušek. *Lectures on discrete geometry*. Vol. 212. Graduate Texts in Mathematics. New York: Springer-Verlag, 2002, pp. xvi+481. ISBN: 0-387-95373-6 (cit. on pp. 202, 222).

[117] J. Matoušek. "Geometric set systems". In: *European Congress of Mathematics, Vol. II (Budapest, 1996)*. Vol. 169. Progr. Math. Birkhäuser, Basel, 1998, pp. 1–27 (cit. on p. 222).

[118] J. W. Moon and L. Moser. "On a problem of Turán". In: *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 7 (1962), pp. 283–286 (cit. on p. 116).

[119] G. Moshkovitz and A. Shapira. "A short proof of Gowers' lower bound for the regularity lemma". In: *Combinatorica* 36.2 (2016), pp. 187–194. ISSN: 0209-9683 (cit. on p. 194).

[120] J. S. Myers. "Graphs without large complete minors are quasi-random". In: *Combin. Probab. Comput.* 11.6 (2002), pp. 571–585. ISSN: 0963-5483 (cit. on p. 47).

[121] B. Nagle, V. Rödl, and M. Schacht. "The counting lemma for regular k-uniform hypergraphs". In: *Random Structures Algorithms* 28.2 (2006), pp. 113–179. ISSN: 1042-9832 (cit. on pp. 7, 253).

[122] E. Nelson. "Internal set theory: a new approach to nonstandard analysis". In: *Bull. Amer. Math. Soc.* 83.6 (1977), pp. 1165–1198. ISSN: 0002-9904 (cit. on p. 151).

[123] V. Nikiforov. "Eigenvalues and extremal degrees of graphs". In: *Linear Algebra Appl.* 419.2-3 (2006), pp. 735–738. ISSN: 0024-3795 (cit. on p. 47).

[124] E. A. Nordhaus and B. M. Stewart. “Triangles in an ordinary graph”. In: *Canad. J. Math.* 15 (1963), pp. 33–41. ISSN: 0008-414X (cit. on p. 116).

[125] E. Palmgren. “A constructive approach to nonstandard analysis”. In: *Ann. Pure Appl. Logic* 73.3 (1995), pp. 297–325. ISSN: 0168-0072 (cit. on p. 151).

[126] E. Palmgren. “A note on Brouwer’s weak continuity principle and the transfer principle in nonstandard analysis”. In: *J. Log. Anal.* 4 (2012), Paper 2, 7 (cit. on p. 151).

[127] F. Petrov and A. Vershik. “Uncountable graphs and invariant measures on the set of universal countable graphs”. In: *Random Structures Algorithms* 37.3 (2010), pp. 389–406. ISSN: 1042-9832 (cit. on p. 273).

[128] A. Pillay. *Geometric stability theory*. Vol. 32. Oxford Logic Guides. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1996, pp. x+361. ISBN: 0-19-853437-X (cit. on p. 222).

[129] B. Poizat. *Stable groups*. Vol. 87. Mathematical Surveys and Monographs. Translated from the 1987 French original by Moses Gabriel Klein. American Mathematical Society, Providence, RI, 2001, pp. xiv+129. ISBN: 0-8218-2685-9 (cit. on p. 222).

[130] A. A. Razborov. “Flag algebras”. In: *J. Symbolic Logic* 72.4 (2007), pp. 1239–1282. ISSN: 0022-4812 (cit. on pp. 121, 124, 126, 130).

[131] C. Reiher and M. Schacht. “Forcing quasirandomness with triangles”. In: *ArXiv e-prints* (Nov. 2017). eprint: `1711.04754` (math.CO) (cit. on p. 47).

[132] C. Reiher. “The clique density theorem”. In: *Ann. of Math. (2)* 184.3 (2016), pp. 683–707. ISSN: 0003-486X (cit. on pp. 113, 116, 119).

[133] V. Rödl and M. Schacht. “Regularity lemmas for graphs”. In: *Fete of combinatorics and computer science*. Ed. by G. O. H. Katona, A. Schrijver, and T. Szőnyi. Vol. 20. Bolyai Soc. Math. Stud. Papers from the conference to commemorate the 60th birthday of László Lovász held in Keszthely, August 11–15, 2008. Budapest: János Bolyai Math. Soc., 2010, pp. 287–325. ISBN: 978-963-9453-12-8; 978-3-642-13579-8 (cit. on p. 194).

[134] V. Rödl and J. Skokan. “Regularity lemma for k-uniform hypergraphs”. In: *Random Structures Algorithms* 25.1 (2004), pp. 1–42. ISSN: 1042-9832 (cit. on pp. 7, 253).

[135] N. Sauer. "On the density of families of sets". In: *J. Combinatorial Theory Ser. A* 13 (1972), pp. 145–147 (cit. on p. 222).

[136] A. Shapira. "Quasi-randomness and the distribution of copies of a fixed graph". In: *Combinatorica* 28.6 (2008), pp. 735–745. ISSN: 0209-9683 (cit. on p. 47).

[137] A. Shapira and R. Yuster. "The effect of induced subgraphs on quasi-randomness". In: *Random Structures Algorithms* 36.1 (2010), pp. 90–109. ISSN: 1042-9832 (cit. on p. 47).

[138] S. Shelah. "Definable groups for dependent and 2-dependent theories". In: *Sarajevo J. Math.* 13(25).1 (2017), pp. 3–25. ISSN: 1840-0655 (cit. on p. 253).

[139] S. Shelah. "Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory". In: *Ann. Math. Logic* 3.3 (1971), pp. 271–362. ISSN: 0168-0072 (cit. on pp. 198, 222).

[140] A. F. Sidorenko. "Inequalities for functionals generated by bipartite graphs". In: *Diskret. Mat.* 3.3 (1991), pp. 50–65. ISSN: 0234-0860 (cit. on pp. 47, 151).

[141] P. Simon. "Distal and non-distal NIP theories". In: *Ann. Pure Appl. Logic* 164.3 (2013), pp. 294–318. ISSN: 0168-0072 (cit. on p. 220).

[142] M. Simonovits and V. T. Sós. "Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs". In: *Combinatorica* 17.4 (1997), pp. 577–596. ISSN: 0209-9683 (cit. on p. 47).

[143] M. Simonovits and V. T. Sós. "Szemerédi's partition and quasirandomness". In: *Random Structures Algorithms* 2.1 (1991), pp. 1–10. ISSN: 1042-9832 (cit. on p. 47).

[144] J. Skokan and L. Thoma. "Bipartite subgraphs and quasi-randomness". In: *Graphs Combin.* 20.2 (2004), pp. 255–262. ISSN: 0911-0119 (cit. on p. 47).

[145] E. Szemerédi. "On sets of integers containing no k elements in arithmetic progression". In: *Acta Arith.* 27 (1975). Collection of articles in memory of Juriĭ Vladimirovič Linnik, pp. 199–245. ISSN: 0065-1036 (cit. on p. 7).

[146] E. Szemerédi. "Regular partitions of graphs". In: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. Vol. 260. Colloq. Internat. CNRS. Colloque International CNRS held at the Université d'Orsay, Orsay, July 9–13, 1976. Paris: CNRS, 1978, pp. 399–401. ISBN: 2-222-02070-0 (cit. on p. 194).

[147] M. Talagrand. "The Glivenko-Cantelli problem". In: *Ann. Probab.* 15.3 (1987), pp. 837–870. ISSN: 0091-1798 (cit. on p. 222).

[148] M. Talagrand. "The Glivenko-Cantelli problem, ten years later". In: *J. Theoret. Probab.* 9.2 (1996), pp. 371–384. ISSN: 0894-9840 (cit. on p. 222).

[149] T. Tao. "A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma". In: *J. Analyse Math.* 103 (1 2007). 10.1007/s11854-008-0001-0, pp. 1–45. ISSN: 0021-7670 (cit. on p. 7).

[150] T. Tao. "A quantitative ergodic theory proof of Szemerédi's theorem". In: *Electron. J. Combin.* 13.1 (2006), Research Paper 99, 49 pp. (electronic). ISSN: 1077-8926 (cit. on pp. 7, 194).

[151] T. Tao. "Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets". In: *Contrib. Discrete Math.* 10.1 (2015), pp. 22–98. ISSN: 1715-0868 (cit. on p. 222).

[152] T. Tao. "Norm convergence of multiple ergodic averages for commuting transformations". In: *Ergodic Theory Dynam. Systems* 28.2 (2008), pp. 657–688. ISSN: 0143-3857 (cit. on p. 195).

[153] T. Tao. "Szemerédi's regularity lemma revisited". In: *Contrib. Discrete Math.* 1.1 (2006), pp. 8–28. ISSN: 1715-0868 (cit. on pp. 194, 195).

[154] C. Terry. *An improved bound for regular decompositions of* 3*-uniform hypergraphs of bounded* VC_2*-dimension*. 2022 (cit. on p. 254).

[155] C. Terry and J. Wolf. *Irregular triads in 3-uniform hypergraphs*. 2021 (cit. on p. 254).

[156] A. Thomason. "Pseudorandom graphs". In: *Random graphs '85 (Poznań, 1985)*. Ed. by Z. Karoński Michałand Palka. Vol. 144. North-Holland Math. Stud. Papers from the second international seminar on random graphs and probabilistic methods in combinatorics held at Adam Mickiewicz University, Poznań, August 5–9, 1985, Annals of Discrete Mathematics, 33. Amsterdam: North-Holland, 1987, pp. 307–331. ISBN: 0-444-70265-2 (cit. on p. 47).

[157] A. Thomason. "Random graphs, strongly regular graphs and pseudorandom graphs". In: *Surveys in combinatorics 1987 (New Cross, 1987)*. Vol. 123. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 1987, pp. 173–195 (cit. on p. 47).

[158] H. Towsner. "σ-algebras for quasirandom hypergraphs". In: *Random Structures Algorithms* 50.1 (2017), pp. 114–139. ISSN: 1042-9832 (cit. on pp. 151, 194, 253).

[159] H. Towsner. *A Removal Lemma for Ordered Hypergraphs*. 2021. eprint: `arXiv:2101.09769` (cit. on p. 253).

[160] H. Towsner. "An analytic approach to sparse hypergraphs: hypergraph removal". In: *Discrete Analysis* (2018) (cit. on p. 194).

[161] V. N. Vapnik and A. J. Červonenkis. "The uniform convergence of frequencies of the appearance of events to their probabilities". In: *Teor. Verojatnost. i Primenen.* 16 (1971), pp. 264–279. ISSN: 0040-361x (cit. on p. 222).

[162] E. R. Vaughn. *Flagmatic User's Guide* (cit. on pp. 115, 130).

[163] R. Yuster. "Quasi-randomness is determined by the distribution of copies of a fixed graph in equicardinal large sets". In: *Combinatorica* 30.2 (2010), pp. 239–246. ISSN: 0209-9683 (cit. on p. 47).

[164] Y. Zhao. "Hypergraph limits: a regularity approach". In: *Random Structures Algorithms* 47.2 (2015), pp. 205–226. ISSN: 1042-9832 (cit. on p. 253).

[165] T. Ziegler. "A non-conventional ergodic theorem for a nilsystem". In: *Ergodic Theory Dynam. Systems* 25.4 (2005), pp. 1357–1370. ISSN: 0143-3857 (cit. on p. 7).