Randomness at Infinity

Henry Towsner

December 10, 2018

Contents

Chapter 1

Introduction

The goal of this book is to present a proof of 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.

There are a rather large number of proofs of Szermerédi's Theorem ([2, 17, 18, 20, 21, 22, 24, 28, 38, 45, 53, 54, 55, 59], though these proofs are far from completely distinct), none of which require a book to present. Rather than presenting the shortest or most direct proof, I intend to meander through a number of detours. To preview what's ahead, let's consider some of the steps in the proof of the $k = 3$ case of Szemerédi's Theorem, also known as Roth's Theorem [46].

1.1 Converting to a Question about Graphs

Suppose n is large and we are given a set $A \subseteq \{1, 2, \ldots, n\}$ with $\frac{|A|}{n} \geq \epsilon$.

We are looking for arithmetic progressions $a, a + d, a + 2d \in A$. Because we are looking for triples, it turns out to be convenient to find a way of describing these arithmetic progressions using three numbers which can vary independently: we associate the triple (x, y, z) with an arithmetic

progression given by $a = x + 2y$ and $d = z - (x + y)$. Then we have $a = x + 2y$, $a + d = z + y$, and $a + 2d = 2z - x$. The important feature of this representation is that each of the three positions $a, a + d, a + 2d$ is described by a different pair of coordinates from (x, y, z), while every triple describes an arithmetic progression. In particular, we can vary one of the coordinates—say, switching from (x, y, z) to (x', y, z)—and change from one arithmetic progression to a different one. (Whereas if we changed from $(a, a + d, a + 2d)$ to $(a', a + d, a + 2d)$ we would no longer have an arithmetic progression.)

We define a "tripartite graph"—let the sets X, Y, and Z be three distinct copies of $\{1, 2, \ldots, n\}$, and we define sets of edges E, F, G by

- a pair $(x, y) \in X \times Y$ belongs to E if $x + 2y \in A$,
- a pair $(x, z) \in X \times Z$ belongs to F if $2z - x \in A$,
- a pair $(y, z) \in Y \times Z$ belongs to G if $z + y \in A$.

These definitions are set up so that when we have a *triangle*—a triple $(x, y, z) \in X \times Y \times Z$ such that when each of the pairs belongs to the appropriate set E, F, or G—then $x+2y, z+y, 2z-x$ describes an arithmetic progression in A.

So in order to prove the $k = 3$ case of Szemerédi's Theorem, we need to find a triangle in this graph. But not quite any triangle will do, because there are "trivial" triangles corresponding to progressions where $d = 0$: whenever $x + 2y \in A$, consider the triple $(x, y, x + y)$. Since $x + 2y \in A$ we have $(x, y) \in E$, but since $2(x + y) - x = x + 2y$ and $(x + y) + y = x + 2y$, we also have $(x, x + y) \in F, (y, x + y) \in G$.

So we can reduce our problem to the question of showing that this graph has a non-trivial triangle.

1.2 Probability on Graphs

We will want to think about the "density" of how many triangles there should be. Our assumption was that A was a "dense" set—its cardinality was on the order of ϵn where $\epsilon > 0$ was a fixed number and n is then very large relative to ϵ.

Similarly, with some more effort (and some tweaks to the definitions), we will ultimately be able to show that E, F, and G are also dense sets: $\frac{|E|}{n^2} \geq \epsilon'$ (where $\epsilon' > 0$ is a real number depending only on ϵ, though it may be a bit smaller than ϵ), and similarly for F and G.

It will be natural to think of these as probability measures: we will say $\mu_2(E) \geq \epsilon'$, where $\mu_2(S) = \frac{|S|}{n^2}$. This will allow us to think of counting triangles as a question about integrals: the number of triangles is

$$\left(\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z)\, d\mu_3\right) \cdot n^3$$

where $\iiint \cdots d\mu_3$ just means $\frac{1}{n^3}\sum_{(x,y,z)\in X\times Y\times Z} \cdots$ and $\chi_E(x,y)$ is the *characteristic function* which is 1 when $(x,y) \in E$ and 0 when $(x,y) \notin E$.

Indeed, what we will eventually show is that, for n sufficiently large, we can bound this integral

$$\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z)\, d\mu_3 \geq \delta > 0.$$

This will suffice, since there are at most n^2 trivial triangles—each pair (x,y) determines at most one trivial triangle $(x,y,x+y)$. So if we can find δn^3 triangles and choose $n > 1/\delta$, we have guaranteed there is at least one non-trivial triangle, and therefore an arithmetic progression of length 3.

1.3 Randomness and Structure

The remaining question is how we can hope to obtain this bound on the integral

$$\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z)\, d\mu_3.$$

We might hope, naïvely, that the sets E, F, and G are independent in a way that allows us to integrate over the three sets separately:

$$\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z)\, d\mu_3(x,y,z) \stackrel{?}{=} \iint \chi_E(x,y)d\mu_2 \iint \chi_F(x,z)d\mu_2 \iint \chi_G(y,z)d\mu_2.$$

This is implausible, not least because the sets E, F, and G are all derived from the same set A.

However our proof will depend on replacing the functions χ_E, χ_F, χ_G with functions for which this is true: we will have a decomposition

$$\chi_E(x,y) = e^{\top}(x,y) + e^{\perp}(x,y)$$

where $e^{\perp}$ is a "random" function with the property that

$$\iiint e^{\perp}(x,y)f(x,z)g(y,z)d\mu_3 = 0$$

for any f and g, while $e^\top$ is a "random-free" functions which has a "structure"—a particular kind of description in terms of simpler functions.

By repeating this decomposition for all three functions, we will be able to show that

$$\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z)\,d\mu_3 = \iiint e^\top(x,y)f^\top(x,z)g^\top(y,z)d\mu_3.$$

We will then—at last—be able to use the descriptions of $e^\top$, $f^\top$, and $g^\top$ by simpler functions to prove that this last integral is positive.

1.4 Our Approach

The many proofs of Szemerédi's Theorem vary along two main dimensions. Every proof of Szermerédi's Theorem depends on some sort of dichotomy where the set A is divided into a "structured" part and a "random" part, but different proofs use distinct notions of what it means for a set to be structured have appeared across the various proofs.

Every proof of Szemerédi's Theorem has a certain "analytic" character, with notions like density and averages playing a central role, but different proofs approach this in a different way. Some proofs entirely ignore this perspective entirely, preferring to focus on counting and cardinality (for example, speaking of $|A| \geq \epsilon n$ rather than $\frac{|A|}{n} \geq \epsilon$). Others use terminology that reflects this perspective, speaking of integrals and expected value, but remaining entirely in a finite setting. But other proofs use limiting techniques to pass to an infinitary setting—an ergodic theoretic or measure theoretic setting.

My real goal is to present, simultaneously, a particular approach to what randomness (and, dually, structure) means and a particular approach to working with infinite limits of finite graphs. This will lead us on a roundabout path which begins in finite combinatorics and ends up in the setting of probability and measure theory, with a decent helping of techniques from logic to mediate between the two.

Chapter 2

Random and Quasi-random Graphs

2.1 The Random Graph

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

Definition 2.1. 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: 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, 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.

Theorem 2.2. *For every $\epsilon > 0$ and every $\delta > 0$, whenever V is sufficiently large,*

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

It will usually 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 $\{v,w\} \notin \mathbf{R}_{1/2}$.

What this 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 $\mathbf{1}_{\{v,w\}} = 0$.

It's not hard to verify that the expected value of $\frac{|\mathbf{R}_{1/2}|}{\binom{|V|}{2}}$ is also $1/2$: by the the linearity of expected value

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

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—based on what we have shown so far—it might be 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*

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

Since $\mu(E)$ has exactly this form, 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 small. □

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 2.3. The *complete bipartite graph,* $\mathrm{K}_{n,n}$, 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 $\frac{1}{2}\binom{|V\cup W|}{2} = \frac{1}{2}\frac{2n(2n-1)}{2} = n^2 - \frac{n}{2}$.

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. (Of course, this error can be fixed by simply removing $n/2$ edges—say, choosing a single vertex in v and removing half its edges.)

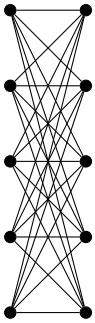


Figure 2.1: $K_{5,5}$

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.

2.2 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 probability-theoretic terms, we will ask what fraction of the possible copies of H are actually present.

Definition 2.4. 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 2.5. 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 $t_{C_3}(G)$: since $t_{C_3}(G)$ 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 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 2.6. When $H = (W, F)$ and $G = (V, E)$ are graphs, an *induced copy of H in V* is a possibly $\pi : W \to V$ 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:

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

*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.

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 $\mathbf{R}_{1/2}$ 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 represents a copy of H—that is, if, for each $\{w, w'\} \in F$, $\{\pi(w), \pi(w')\} \in \mathbf{R}_{1/2}$. For each pair $\{w, w'\} \in F$, there is a $1/2$ chance that $\{\pi(w), \pi(w')\}$ ends up being put into $\mathbf{R}_{1/2}$. $\mathbf{1}_\pi$ is 0 if π does not represent a copy of H. As long as π is injective, each edge is determined independently, so $\mathbb{E}(\mathbf{1}_\pi) = 2^{-|F|}$.

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

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

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

In particular,

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

(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 now 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 2.7. *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 2.2 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

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

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 very 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 2.8. *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 $K_{n,n}$: in a random graph, about $\frac{1}{8}$ of the potential triangles will almost certainly be actual triangles, while in $K_{n,n}$ 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 $t_H(G)$.

2.3 $t_H(G)$ as an Integral

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

Definition 2.9. 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 $t_{C_3}(G)$ is precisely the measure of this set under μ_3.

It will be convenient to identify $t_H(G)$ and $t_H^{\text{ind}}(G)$ with certain integrals.

Definition 2.10. When E is a graph on V, $\chi_E : V^2 \to \{0, 1\}$, the *characteristic function of* E, is the function given by:

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

This lets us write, for instance,

$$t_{C_3}(G) = \iiint \chi_E(x, y)\chi_E(x, z)\chi_E(y, z)\, d\mu_3.$$

More generally, we have

Theorem 2.11. *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:

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

In particular, the edge density $\frac{|E|}{\binom{|V|}{2}}$ can (almost) be expressed this way.

Definition 2.12. 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

$$t_{\mathrm{K}_2}(G) = \frac{2|E|}{|V|^2},$$

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.

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

Definition 2.13. 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 2.14. *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 in order for this quantity to be meaningful, it 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 2.15. *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

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

□

2.4 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.

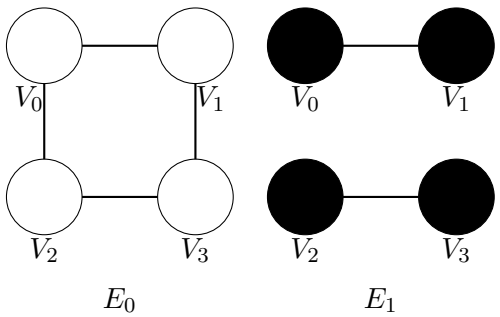
We've already seen a graph with too few triangles: $K_{n,n}$ has about half the edges, but no triangles at all. For a graph with too many triangles, the "complement" of $K_{n,n}$, which we will call $\overline{K_{n,n}}$ works: the graph $(V \cup W, \binom{V}{2} \cup \binom{W}{2})$.

$\overline{K_{n,n}}$ 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 $\overline{K_{n,n}}$ as the disjoint union of the complete graph on V with the complete graph on W.

$\overline{K_{n,n}}$ 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 an even-numbered part and one vertex in an odd-numbered part, 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 $V_2 \cup V_3$.



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, $\mathbf{G}_p$. $\mathbf{G}_p$ 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 if it's present in E_1. If the coin comes up tails, we place the edge in 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.

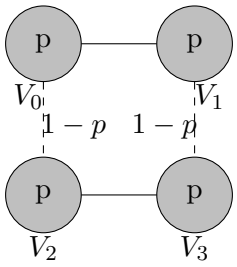
Analogous to the way we counted densities in the random graph using a function constantly equal to $1/2$, we can count densities in $\mathbf{G}_p$ using a function f_p which reflects the grid above.

Theorem 2.16. *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 2.17. *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 2.18. *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 $t_{\mathrm{C}_3}(f_p)$ 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:

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

□

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 about 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, so 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 2.19. *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. First, suppose we fix a particular choice of X and Y in advance. Then the only edges we care about are those between X and Y. The claimed property will hold for this particular choice of X and Y by the same arguments we used for the whole graph $\mathbf{R}_{1/2}$.

Observe that

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

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

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

Next we need to argue that since each choice of sets X and Y is individually very unlikely to have the wrong edge density, actually it's unlikely for any choice to have the wrong edge density. We use the *union bound*:

$$\mathbb{P}(X_1 \text{ or } X_2 \text{ or } \cdots \text{ or } X_k) \leq \sum_{i=1}^{k} \mathbb{P}(X_i).$$

Fix sizes $x \geq C \ln |V|$ and $y \geq C \ln |V|$, and consider all possible choices of X, Y with $|X| = x$ and $|Y| = y$. There are $\binom{|V|}{x}$ possible choices for X and $\binom{|V|}{y}$ possible choices for Y. Using *Stirling's Approximation*, we have the bound $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$, so

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

Therefore the probability that there exists any sets X and Y with $|X| = x$, $|Y| = y$, and $\left|\frac{|\mathbf{R}_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2}\right| \geq \epsilon$ is bounded by

$$2e^{-2\epsilon^2 xy} e^{x \ln |V| + y \ln |V| + x + y} = 2e^{x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 xy} \leq 2e^{x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V|}.$$

We can use the union bound again over all possible sizes x and y; there are at most n choices for x and n for y, so the probability that there are any sets X and Y with $|X| \geq C \ln |V|$, $|Y| \geq C \ln |V|$, and $\left|\frac{|\mathbf{R}_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2}\right| \geq \epsilon$ is bounded by

$$n^2 \cdot 2e^{x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V|} \leq 2e^{2 \ln n - x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V|}.$$

When $x, y \geq C \ln |V|$ for C sufficiently large,

$$2 \ln n + x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V| \leq -\ln^2 |V|.$$

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 $\mathbf{R}_{1/2}$.

2.5 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 2.20. 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 2.21. *For each $\epsilon > 0$ and $\delta > 0$, when n is sufficiently large, with probability $\geq 1 - \delta$, $|t_{\mathrm{C}_4}(\mathbf{G}_p) - \frac{1}{8}(p^4 - 2p^3 + 3p^2 - 2p + 1)| < \epsilon$.*

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 2.22. *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)**.**

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

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

□

This 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 2.23. 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 $t_{\mathrm{C}_4}(G)$ is close to $(t_{\mathrm{K}_2}(G))^4$; however, because of the previous lemma, it is equivalent just to have $t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4 < \epsilon$.

As the name suggests, what we plan to show is that when $t_{\mathrm{C}_4}(G) \approx (t_{\mathrm{K}_2}(G))^4$, G must resemble a random graph.

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 G is ϵ-quasirandom, $t_{K_2}(G) = p$, and 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 quasirandom.

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

Theorem 2.24. *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 2.7, when V is big enough, the probability that both

$$|t_{\mathrm{K}_2}(\mathbf{R}_p) - \mathbb{E}(t_{\mathrm{K}_2}(\mathbf{R}_p))| < \epsilon'$$

and

$$|t_{\mathrm{C}_4}(\mathbf{R}_p) - \mathbb{E}(t_{\mathrm{C}_4}(\mathbf{R}_p))| < \epsilon/2$$

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

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

$$\mathbb{E}(t_{\mathrm{K}_2}(\mathbf{R}_p)) = t_{\mathrm{K}_2}(f_p) = p$$

and

$$\mathbb{E}(t_{\mathrm{C}_4}(\mathbf{R}_p)) = t_{\mathrm{C}_4}(f_p) = \int f_p(x, y) f_p(y, x') f_p(x', y') f_p(y', x) d\mu_4 = p^4.$$

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

$$\left|t_{\mathrm{C}_4}(\mathbf{R}_p) - (t_{\mathrm{K}_2}(\mathbf{R}_p))^4\right| \leq \left|t_{\mathrm{C}_4}(\mathbf{R}_p) - p^4\right| + \left|(t_{\mathrm{K}_2}(\mathbf{R}_p))^4 - p^4\right| < \epsilon/2 + \epsilon/2 = \epsilon.$$

□

2.6 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 2.25. When $G = (V, E)$ is a graph and $x \in V$, $N_G(x) = \{y \in x \mid \{x, y\} \in E\}$ is the *neighborhood* of x.

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

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

Theorem 2.26. *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 converse. 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:

$$\begin{aligned}t_{\mathrm{C}_4}(G) &= \int \left(\int \chi_E(x,y)\chi_E(x,z)d\mu(x)\right)^2 d\mu_2(y,z)\\ &\geq \left(\int \chi_E(x,y)\chi_E(x,z)d\mu_3\right)^2\\ &\geq (p^2+\epsilon^3)^2\\ &\geq p^4 + \epsilon^6.\end{aligned}$$

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 2.19. 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 2.27. When $G = (V, E)$ is a graph and $X \subseteq V, Y \subseteq V$, the *edge density between X and Y*, $d_E(X,Y) = \frac{|E \cap (X \times Y)|}{|X| \cdot |Y|}$.

Theorem 2.28. *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 $t_{C_4}(G)$ 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

$$\begin{aligned} t_{C_4}(G) &= \int \chi_E(x,y)\chi_E(y,x')\chi_E(x',y')\chi_E(y',x)d\mu_4 \\ &= \int (f(x,y)-p)(f(y,x')-p)(f(x',y')-p)(f(y',x)-p)d\mu_4 \\ &= p^4 + 4p\int f(x,y)f(y,x')f(x',y')d\mu_4 + 4p^2\int f(x,y)f(y,x')d\mu_3 \\ &\quad + 2p^2\int f(x,y)f(x',y')d\mu_4 + 4p^3\int f(x,y)d\mu_2 + t_{C_4}(f). \end{aligned}$$

Again, p^4 accounts for all the copies of C_4 we should have. $t_{C_4}(f)$ 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 $t_{C_4}(f)$.

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

$$t_{C_4}(G) \geq p^4 + t_{C_4}(f) - \epsilon^{12}/2 \geq p^4 + \epsilon^8 d_E(X,Y)^4 - \epsilon^{12}/2.$$

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 2.29. $G = (V, E)$ 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 2.28 can be rephrased

Theorem 2.30. *For every $\epsilon > 0$ there is a $\delta > 0$ so that if G is δ-regular then G is ϵ-quasirandom.*

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 2.31. *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 $N_G(x) = \{y \mid \{x, y\} \in E\}$, so $\deg_G(x) = \mu(N_G(x))$. Then, for most x, $\mu(N_G(x)) \approx p$ by Theorem 2.26. Furthermore, by Theorem 2.28, $d_E(N_G(x), N_G(x)) \approx p$. But a triangle whose first vertex is x is exactly an edge between $N_G(x)$ 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 2.26, 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

$$\begin{aligned} t_{\mathrm{C}_3}(G) &= \int \chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3 \\ &= \int \chi_{V\setminus S}(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3 + \int \chi_S(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3 \\ &= \int \chi_{V\setminus S}(x)\mu(N_G(x))^2 d_E(N_G(x), N_G(x))d\mu_3 + \int \chi_S(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3. \end{aligned}$$

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

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

On the other hand, by Theorem 2.28 with $X = Y = N_G(x)$, we also have $|d_E(N_G(x), N_G(x)) - p| < \epsilon/6p$. So

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

So

$$\left|t_{\mathrm{C}_3}(G) - p^3\right| < \epsilon.$$

□

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

Theorem 2.32. *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 $t_{\mathrm{K}_2}(G) = p$ and V sufficiently large, $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| < \epsilon$.*

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 $t_{\mathrm{C}_r}(G) \approx p^4$, which means

Theorem 2.33. *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 2.32, so we defer it to later as well.

2.7 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 which will generalize better in later chapters.

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 2.34. 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 2.35. When $h : V^2 \to \mathbb{R}$ is a symmetric function, we associate a transformation T_h: given $f : V \to \mathbb{R}$, we define $T_E f : V \to \mathbb{R}$ 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 2.36. *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) $\lambda_1, \dots, \lambda_n$ and vectors $f_1, \dots, f_n$ 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 $f_1, \dots, f_n$ form an orthonormal basis.

For example, recall our bipartite graph $\mathrm{K}_{n,n}$, 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 2.37. *When $\lambda_1, \dots, \lambda_n$ and $f_1, \dots, f_n$ are the eigenvalues and eigenvectors associated to T_h, $h(x,y) = \sum_{i \leq n} \lambda_i f_i(x) f_i(y)$.*

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

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

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

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 a standard fact:

Theorem 2.38. *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

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

since each f_i has norm 1. □

When we think of h as a matrix, $\int h(x,x)d\mu$ is essentially the trace of the matrix, so this just says that the trace is equal to the sum of the eigenvalues.

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

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, $|\lambda_1| \geq \langle T_h g, g\rangle = t_{\mathrm{K}_2}(h)$.

If we do not wish to invoke this, we can show that $|\lambda_1| \geq \langle T_h g, g\rangle$ 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 2.40. • *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 $\lambda_1, \ldots, \lambda_n$ and $f_1, \ldots, f_n$ are the eigenvalues and eigenvectors associated with h then $\lambda_1^2, \ldots, \lambda_n^2$ and $f_1, \ldots, f_n$ 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 2.41. *$G = (V, E)$ is δ-quasirandom if and only if*

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

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 $(\lambda_1 - t_{\mathrm{K}_2}(G))^4 + \sum_{1<i\le n} \lambda_i^4$—the point is that G is δ-quasirandom exactly when two things happen: first, λ_1 is close to $t_{\mathrm{K}_2}(G)$, and second, the remaining eigenvalues are all small.

2.8 Dense Graphs

Ultimately, our interest will extend beyond random graphs to the more general behavior of dense graphs. By a dense graph, we mean a graph G where $t_{\mathrm{K}_2}(G) = \epsilon > 0$, where the number of vertices in G is generally much larger than $1/\epsilon$. How much larger will depend on the particular question; typically we'll be interested in properties which hold eventually, once the number of vertices is sufficiently large.

For example, the oldest result in extremal graph theory is *Mantel's Theorem*:

Theorem 2.42. *If G has no triangles then $t_{\mathrm{K}_2}(G) \le \frac{1}{2}$.*

The bipartite graph illustrates that this bound is optimal: the bipartite graph $K_{n,n}$ has $t_{\mathrm{K}_2}(K_{n,n}) = 1/2$.

Proof. Since the graph has no triangles, whenever $\{x, y\} \in E$, we must have $\deg_G(x) + \deg_G(y) \le 1$, so

$$\begin{aligned} t_{\mathrm{K}_2}(G) &= \int \chi_E(x,y) \cdot 1 \, dy \, dx \\ &\ge \int \chi_E(x,y) (\int \chi_E(x,z) dz + \int \chi_E(x,w) dw) dy \, dx \\ &= 2 \int (\chi_E(x,y) dy)^2 \, dx \\ &\ge 2 \left(\int \chi_E(x,y) dx dy \right)^2 \\ &= 2 t_{\mathrm{K}_2}(G)^2, \end{aligned}$$

so $\frac{1}{2} \ge t_{\mathrm{K}_2}(G)$. □

A natural generalization is to ask what happens when we replace the triangle with other small graphs H and ask about the density of graphs with no copies of H. If $G_t = (V, E)$ is a graph where V is partitioned into t parts $V = V_1 \cup V_2 \cup \cdots \cup V_t$ and E consists of all edges between distinct parts—that is, $E = \binom{V}{2} \setminus \bigcup_{i \leq t} \binom{V_i}{2}$—then G_t contains no copies of K_{t+1}. When $|V| = n$ and the V_i each have size roughly n/t, $\frac{|E|}{|\binom{V}{2}|} \approx 1 - \frac{1}{t}$.

Indeed, if $T_H(G_t) \neq \emptyset$ then $H = (W, F)$ is *t-colorable*: there is a function $c : W \to \{1, 2, \ldots, t\}$ such that whenever $\{x, y\} \in F$, $c(x) \neq c(y)$. (To find such a coloring, take any $\pi \in T_H(G_r)$ and set $c(x) = i$ where $\pi(x) \in V_i$.)

Definition 2.43. For any $H = (W, F)$, $\chi(H)$, the *coloring number* of H, is the smallest integer r such that there is a function $c : W \to \{1, 2, \ldots, r\}$ such that, for each $\{x, y\} \in F$, $c(x) \neq c(y)$.

Turán showed that graphs like G_t are the densest possible graphs containing no copies of K_{t+1} (and therefore, for any H with $\chi(H) = t + 1$, the densest graphs containing no copies of H):

Theorem 2.44. *If G contains no copies of the H then $t_{\mathrm{K}_2}(G) \leq 1 - \frac{1}{\chi(H)-1}$.*

Rather than prove these here, we will wait for versions of these results to follow from the general theory we develop in later chapters.

2.9 Related Topic: The Payley Graph

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

Definition 2.45. 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

$$t_{\mathrm{K}_2}(Q_q) = \int \chi_{Q_q}(x, y) d\mu = \frac{1}{2} - \frac{1}{2q},$$

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

Theorem 2.46. *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. If Q_q is going to be quasirandom, each pair x, y should have about $q/4$ neighbors in common.

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

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

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

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

□

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 [23]: every pair of distinct elements of X is an edge. That means that $d_E(X,X)$ is very close to 1, but Theorem 2.19 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.

2.10 What's Next

Our next step will be a detour to develop a framework for working with infinite graphs as the limits of the finite graphs we considered here. In this setting the analytic paradigm where we work with densities and integrals will be necessary, but we will no longer need to worry about approximations and estimates: in the limit, the error terms will fall to 0.

In this setting, quasirandom graphs will become the randomness side of a dichotomy, and we will be able to turn to understanding the "structured" side of that dichotomy: speaking very roughly, we will be able to write a graph as a sum of a quasirandom part and a structured part, where the structured part corresponds roughly to the eigenfunctions of the graph.

After studying this decomposition (and using it to prove the case of Szemerédi's Theorem for arithmetic progressions of length 3) we will turn to generalzing the notion of quasirandomness to hypergraphs.

2.11 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 [6] is a canonical reference, especially when supplemented by more recent books [1, 16, 31].

The study of quasirandom graphs was introduced by Chung, Graham, and Wilson in a paper [8], and most of this chapter is drawn from that paper. Since then, a number of additional characterizations of quasirandomness have been investigated [37, 40, 47, 48, 50, 51, 52, 58]. 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 have been studied as well like the notion of jumbledness introduced by Thomason [56, 57] and also well-studied since [34]. 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 well-studied [10, 43, 52]; 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 [49], which concerns a generalization of the Cauchy-Schwarz calculations we were using above to [25].

The quantity $\max_{X,Y \subseteq V} |d_E(X,Y)|X| \cdot |Y| - |E||$ (or the restriction to when $X = Y$) is called the *discrepency* of a graph. Erdős and Spencer [14] showed that there is a constant so that every graph with $|V|$ has discrepency at least $cn^{3/2}$, and that random graphs have close to this minimum discrepency. The behavior of the discrepency and related quantities in random, quasirandom, and other graphs has also been further studied; Chazelle's book [7] 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 [9].

Chapter 3

Ultraproducts

3.1 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 limiting 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 would need to pass to a subsequence—we would 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 3.1 (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

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

These partition $\mathbb{N}$, 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 $[\frac{a}{2^k}, \frac{a+1}{2^k}]$ (the endpoints might or might not be included) 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 3.2. 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 3.3. *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).

3.2 Ultralimits

We would like to go a step further: we would like to choose a 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 3.4. 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 3.5. *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 3.6. 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 3.7. 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 will often 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 3.8. *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 $K_i \in \mathcal{U}$.*

*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".

Proof. By induction on n. When $n = 1$, this is tautological.

Suppose the claim holds for n and $K_1 \cup K_2 \cup \cdots \cup K_n \cup K_{n+1} \in \mathcal{U}$. If $K_{n+1} \in \mathcal{U}$, we are done, so suppose not. By the "ultra" property, $\mathbb{N} \setminus K_{n+1} \in \mathcal{U}$, so also

$$(K_1 \cup K_2 \cup \cdots \cup K_n \cup K_{n+1}) \cap (\mathbb{N} \setminus K_{n+1}) \in \mathcal{U}.$$

But

$$(K_1 \cup K_2 \cup \cdots \cup K_n \cup K_{n+1}) \cap (\mathbb{N} \setminus K_{n+1}) \subseteq (K_1 \cup K_2 \cup \cdots \cup K_n),$$

so $K_1 \cup K_2 \cup \cdots \cup K_n \in \mathcal{U}$ as well. But then, by the inductive hypothesis, there is an $i \leq n$ so that $K_i \in \mathcal{U}$. □

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

Theorem 3.9. *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. 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

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

□

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.

Definition 3.10. 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.

3.3 Ultrafilters

Making every sequence converge is an impressive property, so we should prove that ultrafilters actually 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 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 $K \in \mathcal{F}_0$.

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 3.11. *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. Suppose there is some $J \in \mathcal{F}$ so that $K \cap J$ is finite. Let $I = \mathbb{N} \setminus (K \cap J)$, so $I \in \mathcal{F}$ as well. Then $I \cap J \in \mathcal{F}$. If $n \in I \cap J$ then $n \notin J \cap K$, but $n \in J$, so $n \notin K$, so $I \cap J \subseteq (\mathbb{N} \setminus K)$ and therefore $\mathbb{N} \setminus K \in \mathcal{F}$.

So if $\mathbb{N} \setminus K \notin \mathcal{F}$ then, for every $J \in \mathcal{F}$, $K \cap J$ must be infinite. Define $\mathcal{F}'$ to consist of all sets I such that, for some $J \in \mathcal{F}$, $K \cap J \subseteq I$. (This is the smallest possible choice for $\mathcal{F}'$—the closure properties of filters tell us that if $\mathcal{F}'$ extends $\mathcal{F}$ and also contains K, $\mathcal{F}'$ must also contain the intersection of K with any set from $\mathcal{F}$, and any set larger than such a set.)

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 $K \in \mathcal{F}'$.

We must check that $\mathcal{F}$ satisfies the properties of 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 $I_1 \in \mathcal{F}'$.

We cannot have $\emptyset \in \mathcal{F}'$, because if $\emptyset \in \mathcal{F}'$ then $\emptyset \supseteq K \cap J$, so $K \cap J$ is finite, contradicting our assumption.

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 3.11, 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 very 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 3.12. *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 $K \in \mathcal{F}$. □

We now have two choices for how to proceed. Both 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 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 3.13 (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: it considers some collection of objects, in our case filters, where we are looking for larger and larger objects. One way we might fail to have a maximal object is if we had a chain with no top—some sequence $P_0 \subseteq P_1 \subseteq \cdots$ (potentially infinitely or even uncountably infinitely long) of larger and larger objects 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.

All that remains is noticing that a maximal free filter is an ultrafilter.

Theorem 3.14. *There is an ultrafilter.*

Proof. Let $\mathcal{P}$ be the set of free filters, ordered by $\subseteq$. Using Lemma 3.12 and Zorn's Lemma, there must be some maximal free filter $\mathcal{F}$. We claim $\mathcal{F}$ is an ultrafilter: for any $K \subseteq \mathbb{N}$, if $\mathbb{N} \setminus K \notin \mathcal{F}$ then, by Lemma 3.11, there is a $\mathcal{F}' \supseteq \mathcal{F} \cup \{K\}$.

But $\mathcal{F}$ is maximal, so $\mathcal{F}' = \mathcal{F}$, so we already have $K \in \mathcal{F}$. □

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 [5, 15].*

*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 [29].

3.4 Products

We now return to our original question: 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.

Because we want it to somehow incorporate information from all the graphs G_n, we will take our limiting graph to be based on the *product* of 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 3.15. *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 on only one point—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 here: since most indices think these sequences are equal, we should decide they actually are equal.

Definition 3.16. 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 3.17. *$\sim_{\mathcal{U}}$ 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 3.18. 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 when sequences should be adjacent, 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 orther 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 $[V_n]_{\mathcal{U}}$.

3.5 Ultraproducts

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

Definition 3.19. 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.

We call the G_n "ground" graphs because they are the "grounded" or "concrete" objects on which the ultraproduct is based.

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 sizes of uncountable infinity, but we will not be concerned with the distinctions among them.

Theorem 3.20. *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, let $k = \min\{|V_n|, n\}$, and chose $w_n \in V_n \setminus \{v_n^1, \ldots, v_n^{k-1}\}$. Since we have chosen $k \leq |V_n|$ and we are only excluding $k-1$ values from V_n, we know some chose 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 $i < k$, so when both $i < |V_n|$ and $i < n$. Therefore

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

Since $\lim_{n\to\infty} |V_n| = \infty$, $\{n \mid |V_n| > i\}$ is cofinite and therefore in $\mathcal{U}$. The set $\{n \mid i < n\}$ is also 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 3.21. *Let H be a finite graph. There is a copy of H in $[G_n]_{\mathcal{U}}$ if and only if*

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

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 $\{n \mid$ there is an induced copy of H 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 $\{n \mid \mathfrak{P}$ 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. (Note that a graph with an isolated triangle could still have many other, non-isolated, triangles.)

Theorem 3.22. *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

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

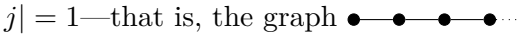
Since $\emptyset \notin \mathcal{U}$, there must be an n in all these sets. But this gives a contradiction: we chose u_n, v_n, w_n to be an isolated triangle in V_n, and x_n contradicts that isolation.

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 3.23. *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 3.24. When $X_n \subseteq V_n$ is a sequence of sets, and 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: 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 3.25. *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 $[v_n]_{\mathcal{U}} \in [X_n]_{\mathcal{U}} \setminus X$. □

3.6 Internal Sets and First-Order Logic

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

Definition 3.26. 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 $v \in [X_n]_{\mathcal{U}}$).

If a subset of $[V_n]_{\mathcal{U}}$ is not internal, it is *external*.

More generally, we will speak of internal subsets of $[V_n]^k_{\mathcal{U}}$ 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 3.27. *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 3.28. *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.

We will often want to prove that sets are internal, so it will be helpful to have some abstract techniques for proving that sets are internal. Indeed, we will ask for something more: that if we "describe" a set X_n in each G_n then $[X_n]_\mathcal{U}$ should be the set with the same "description" in $G_\mathcal{U}$.

Of course, not just any description will work. The examples above suggest, correctly, that if

$$X_n = \{(v, w) \in V_n^2 \mid w \text{ and } w \text{ are in the same connected component}\},$$

we should not generally expect $[X_n]_{\mathcal{U}}$ to be equal to

$$\{(v, w) \in V_{\mathcal{U}}^2 \mid v \text{ and } w \text{ are in the same connected component}\}$$

—indeed, the latter is not usually an internal set. So we need to pick a particular family of descriptions which do pass from the ground structures to the ultraproduct.

Descriptions which only involve equality have the property we want. For instance, if

$$X_n = \{(a^1, a^2) \in V_n^2 \mid a^1 = a^2\}$$

then

$$[X_n]_{\mathcal{U}} = \{(a^1, a^2) \in V_{\mathcal{U}}^2 \mid a^1 = a^2\}.$$

Similarly, if for each V_n we pick some $a_n \in V_n$ and

$$X_n = \{b \in V_n \mid b = a_n\}$$

then

$$[X_n]_{\mathcal{U}} = \{b \in [V_n]_{\mathcal{U}} \mid b = [a_n]_{\mathcal{U}}\}.$$

Stated more abstractly, equality is our first example of a *first-order formula*. Let us take $\phi(x_1, x_2)$ to be the formula $x_1 = x_2$. We customarily use letters like ϕ and ψ for formulas, and indicate the variables—in this case x_1 and x_2—in parentheses after the formula.* If $a^1, a^2 \in V$ then $\phi(a^1, a^2)$ means the statement $a^1 = a^2$: $\phi(a^1, a^2)$ is true if $a^1 = a^2$ and false if $a^1 \neq a^2$.

The point of this abstraction is that it lets us compare the notion of equality across different graphs: we can see that $b_n = a_n$ (where $a_n, b_n \in V_n$) and $b = [a_n]_{\mathcal{U}}$ (where $b, [a_n]_{\mathcal{U}} \in [V_n]_{\mathcal{U}}$) are instances of the same formula—one is $\phi(b_n, a_n)$ and the other is $\phi(b, [a_n]_{\mathcal{U}})$. Then we can rephrase what we said above as:

(1) if $X_n = \{(a^1, a^2) \in V_n^2 \mid \phi(a^1, a^2)\}$ then $[X_n]_{\mathcal{U}} = \{(a^1, a^2) \in V_{\mathcal{U}}^2 \mid \phi(a^1, a^2)\}$, and

(2) if $X_n = \{b \in V_n \mid \phi(b, a_n)\}$ then $[X_n]_{\mathcal{U}} = \{b \in [V_n]_{\mathcal{U}} \mid \phi(b, [a_n]_{\mathcal{U}})\}$.

This is precisely the property we want formulas to have. More generally, the theorem we will prove (once we have finished defining the first-order formulas) is:

*We will generally use letters like x and y for the variables and letters like a and b for elements of graphs. The distinction is subtle, and can mostly be ignored; the difference is that a is an element of some specific graph, while x is a purely abstract placeholder symbol. So $\phi(x)$ is an abstract formula while $\phi(a)$ is a statement about a specific vertex a in some specific graphs which is either true or false

Theorem 3.29. *If $\phi(x_1, \ldots, x_m)$ is a first-order formula, $k \leq m$, and $[a_n^{k+1}]_{\mathcal{U}}, \ldots, [a_n^m]_{\mathcal{U}} \in [V_n]_{\mathcal{U}}$, and*

$$X_n = \{(a^1, \ldots, a^k) \in V_n^k \mid \phi(a^1, \ldots, a^k, a_n^{k+1}, \ldots, a_n^m)\}$$

then

$$[X_n]_{\mathcal{U}} = \{(a^1, \ldots, a^k) \in V_{\mathcal{U}}^k \mid \phi(a^1, \ldots, a^k, [a_n^{k+1}]_{\mathcal{U}}, \ldots, [a_n^m]_{\mathcal{U}})\}.$$

This theorem is the only reason we need to worry about first-order formulas; since the definition is rather technical, what follows is closer to a review of first-order logic than a proper introduction. The reader may wish to skip to the next section, agreeing to take for granted that various sets we need to work with are defined by formulas, and therefore are subject to this theorem.

We need to identify the rest of the first-order formulas. First, we note that formulas do not need to use all their variables: we are perfectly comfortable taking $\phi(x_1, x_2, x_3, x_4)$ to be the formula $x_2 = x_4$, ignoring x_1 and x_3, so for any $b \in V$,

$$X = \{(a_1, a_2, a_3) \in V^3 \mid a_2 = b\}$$

is the set of triples of the form (a_1, b, a_3).

In addition to equality, we might have other *atomic formulas*—formulas not built from smaller formulas. We certainly have at least one: $\{x_1, x_2\} \in E$ is a formula (and, more generally, $\{x_i, x_j\} \in E$). More generally, if we have some related family of sets $R(V_n) \subseteq V_n^k$ for all n and $R([V_n]_{\mathcal{U}}) \subseteq [V_n]_{\mathcal{U}}$ so that $[R(V_n)]_{\mathcal{U}} = R([V_n]_{\mathcal{U}})$ then we may introduce an atomic formula $\{x_1, \ldots, x_n\} \in R$.

We will also have five ways of building more complicated formulas out of simpler ones. The first is negation: when $\phi(x_1, \ldots, x_k)$ is a formula, we write $\neg\phi(x_1, \ldots, x_k)$ for the *negation* of ϕ, which is true exactly when ϕ is false.

For instance, if $\phi(x_1, x_2)$ is $\neg(\{x_1, x_2\} \in E)$ then $\phi(a_1, a_2)$ is true when there is *not* an edge between a_1 and a_2.[†] (We will freely use parentheses in formulas whenever we think it makes the meaning of the formula clearer.)

In general,

$$\begin{aligned}&\{(a^1, \ldots, a^k) \in V^k \mid \neg\phi(a^1, \ldots, a^k, a^{k+1}, \ldots, a^m)\}\\ =&V^k \setminus \{(a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, a^{k+1}, \ldots, a^m)\} :\end{aligned}$$

[†]Though we are not belaboring this point, note that whether or not a formula is true depends on what graph we are considering. Suppose we consider two different graphs whose vertices are the integers 1 and 2; $V = (\{1, 2\}, \{\{1, 2\}\})$ is the graph with two vertices and an edge between then, while $W = (\{1, 2\}, \emptyset)$ is the graph with two vertice and no edges. Then $\phi(1, 2)$ is false in V and true in W.

the set defined by $\neg\phi$ is the *complement* of the set defined by ϕ.

The next two ways of combining formulas are *conjunction* and *disjunction*: when $\phi(x_1, \ldots, x_k)$ and $\psi(x_1, \ldots, x_k)$ are formulas,

- $(\phi \wedge \psi)(x_1, \ldots, x_k)$, read "$\phi$ and ψ", is the conjunction of ϕ with ψ, and is true when both ϕ and ψ are true,

- $(\phi \vee \psi)(x_1, \ldots, x_k)$, read "$\phi$ or ψ", is the disjunction of ϕ with ψ, and is true when at least one of ϕ and ψ is true (including when both are true).

For example, if $\phi(x_1, x_2, x_3)$ is the formula $\{x_1, x_2\} \in E \wedge \{x_1, x_3\} \in E \wedge \{x_2, x_3\} \in E$ then $\phi(a_1, a_2, a_3)$ is true exactly when a_1, a_2, a_3 form a triangle.

If $\phi(x_1, x_2, x_3, x_4)$ is the formula

$$\begin{aligned}&\{x_1, x_2\} \in E \wedge \{x_2, x_3\} \in E \wedge \{x_3, x_4\} \in E \wedge \{x_4, x_1\} \in E \\ &\wedge \neg\{x_1, x_2\} \in E \wedge \neg\{x_3, x_4\} \in E\end{aligned}$$

then $\phi(a_1, a_2, a_3, a_4)$ is true exactly when a_1, a_2, a_3, a_4 are an induced copy of C_4.

More generally, if $H = (W, F)$ is any finite graph with $W = \{w_1, \ldots, w_k\}$, we can write down a long formula

$$\phi(x_1, \ldots, x_k) = \{x_i, x_j\} \in E \wedge \cdots \wedge \{x_{i'}, x_{j'}\} \in E$$

so that $\phi(a_1, \ldots, a_k)$ is true exactly when $\pi(w_i) = a_i$ is a copy of H—the conjunction lists all edges in F. We can write down an even longer formula

$$\psi(x_1, \ldots, x_k) = \{x_i, x_j\} \in E \wedge \cdots \wedge \{x_{i'}, x_{j'}\} \in E \wedge \neg\{x_{i''}, x_{j''}\} \in E \wedge \cdots \wedge \neg\{x_{i'''}, x_{j'''}\} \in E$$

so that $\phi(a_1, \ldots, a_k)$ is true exactly when $\pi(w_i) = a_i$ is an induced copy of H—now the conjunction lists all edges in F and then all the non-edge in $\binom{W}{2} \setminus F$.

Finally we have the *quantifiers*: when $\phi(x_1, \ldots, x_k, y)$ is a formula,

- $(\exists y\phi)(x_1, \ldots, x_k)$, read "there is a y such that ϕ", is true when there is some vertex y so that ϕ is true,

- $(\forall y\phi)(x_1, \ldots, x_k)$, read "for every y, ϕ", is true when ϕ is true for every vertex y.

For example, consider the formula $\phi(x_1, x_2, x_3)$ given by

$$\begin{aligned}&\{x_1, x_2\} \in E \wedge \{x_1, x_3\} \in E \wedge \{x_2, x_3\} \in E \\ \wedge &\forall y \left[(x_1 = y) \vee (x_2 = y) \vee (x_3 = y)\right. \\ &\quad \left.\vee(\neg\{x_1, y\} \in E \wedge \neg\{x_2, y\} \in E \wedge \neg\{x_3, y\} \in E)\right].\end{aligned}$$

Then $\phi(a_1, a_2, a_3)$ is true exactly when a_1, a_2, a_3 are an isolated triangle: the first three clauses say that a_1, a_2, a_3 are a triangle, and the last clause says that, for any vertex y, either y is equal to one of a_1, a_2, a_3, or y is not adjacent to any of these vertices.

We are now ready to prove the main theorem about first-order formulas in ultraproducts. This theorem is sometimes called the Fundamental Theorem of Ultraproducts, because the property it establishes is the fundamental tool for working with ultraproducts.

Theorem 3.30 (Łoś's Theorem)**.** *If $\phi(x_1, \ldots, x_k)$ is a first-order formula and $[a_n^i]_{\mathcal{U}} \in [V_n]_{\mathcal{U}}$ for each $i \leq k$ then*

$$\{n \mid \phi(a_n^1, \ldots, a^k) \text{ is true}\} \in \mathcal{U}$$

if and only if

$$\phi([a_n^1]_{\mathcal{U}}, \ldots, [a_n^k]_{\mathcal{U}}) \text{ is true.}$$

Proof. The formulas were defined recursively: the atomic formulas are formulas, and then, given some formulas, one can build larger formulas using $\neg, \wedge, \vee, \exists, \forall$. So the proof naturally proceeds inductively, following the construction of formulas.

The $\neg, \vee$, and $\wedge$ cases are very similar, and follow from carefully working through all the definitions. The $\exists$ and $\forall$ cases require slightly more care because we need to construct an additional element in the ultraproduct from sequences in the ground graphs.

Throughout, we take $a^i = [a_n^i]_{\mathcal{U}}$ for all i.

First, we consider the atomic formulas. Suppose $\phi(x_1, \ldots, x_m)$ is $x_i = x_j$. We follow a chain of equivalences:

$$\begin{aligned}\phi(a^1, \ldots, a^k) \text{ is true if and only if } &a^i = a^j \\ &\text{if and only if } \{n \mid a_n^i = a_n^j\} \in \mathcal{U} \\ &\text{if and only if } \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U}.\end{aligned}$$

The case where $\phi(x_1, \ldots, x_m)$ is some other atomic formula is identical—the first "if and only if" are the definition of being an atomic formula, and

we only allow ourselves to choose atomic formulas satisfying the middle "if and only if".

Suppose $\phi(x_1, \ldots, x_m)$ is $(\neg\psi)(x_1, \ldots, x_m)$. By the inductive hypothesis, we have

$$\{n \mid \psi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \text{ if and only if } \psi(a^1, \ldots, a^k) \text{ is true.}$$

Then

$$\begin{aligned}\phi(a^1, \ldots, a^k) \text{ is true if and only if } & \psi(a^1, \ldots, a^k) \text{ is false} \\ \text{if and only if } & \{n \mid \psi(a_n^1, \ldots, a_n^k) \text{ is true}\} \notin \mathcal{U} \\ \text{if and only if } & \{n \mid \psi(a_n^1, \ldots, a_n^k) \text{ is false}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U}.\end{aligned}$$

The conjunction and disjunction cases are similar. Suppose ψ_0 and ψ_1 are two formulas for which, inductively, we have already proven the thoerem, so

$$\{n \mid \psi_0(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \text{ if and only if } \psi_0(a^1, \ldots, a^k) \text{ is true}$$

and

$$\{n \mid \psi_1(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \text{ if and only if } \psi_1(a^1, \ldots, a^k) \text{ is true.}$$

First, suppose ϕ is $\psi_0 \wedge \psi_1$. Then

$$\begin{aligned}\phi(a^1, \ldots, a^k) \text{ is true if and only if } & \psi_0(a^1, \ldots, a^k) \text{ and } \psi_1(a^1, \ldots, a^k) \text{ are both true} \\ \text{if and only if } & \{n \mid \psi_0(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \text{ and } \{n \mid \psi_1(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \psi_0(a_n^1, \ldots, a_n^k) \text{ and } \psi_1(a_n^1, \ldots, a_n^k) \text{ are both true}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U}.\end{aligned}$$

Next, suppose ϕ is $\psi_0 \vee \psi_1$. Then

$$\begin{aligned}\phi(a^1, \ldots, a^k) \text{ is true if and only if either } & \psi_0(a^1, \ldots, a^k) \text{ or } \psi_1(a^1, \ldots, a^k) \text{ is true} \\ \text{if and only if } & \{n \mid \psi_0(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \text{ or } \{n \mid \psi_1(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \psi_0(a_n^1, \ldots, a_n^k) \text{ is true}\} \cup \{n \mid \psi_1(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \text{either } \psi_0(a_n^1, \ldots, a_n^k) \text{ or } \psi_1(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U} \\ \text{if and only if } & \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U}.\end{aligned}$$

The remaining cases are for formulas built using quantifiers; these two cases are very similar to each other. Suppose we have a formula $\psi(x_1, \ldots, x_k, y)$ for which, inductively, we have already shown that

$$\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is true}\} \in \mathcal{U} \text{ if and only if } \psi(a^1, \ldots, a^k, b).$$

Suppose ϕ is $\exists z\psi$. If $\phi(a^1, \ldots, a^k)$ is true then there is some $b = [b_n]_\mathcal{U} \in [V_n]_\mathcal{U}$ so that $\psi(a^1, \ldots, a^k, b)$ is true, and therefore, by the inductive hypothesis, $\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is true}\} \in \mathcal{U}$, so $\{n \mid \phi(a_n^1, \ldots, a_n^k)\} \in \mathcal{U}$.

Conversely, if $K = \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \in \mathcal{U}$ then, fore each $n \in K$, there is a b_n so that $\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is true}\} \in \mathcal{U}$. Then we can take $b = [b_n]_\mathcal{U}$ (for $n \notin K$, b_n can be chosen arbitrarily). Then, by the inductive hypothesis, $\psi(a^1, \ldots, a^k, b)$ is true, so $\phi(a^1, \ldots, a^k)$.[‡]

Finally, suppose ϕ is $\forall z\psi$. We prove the contrapositive of the implications. Suppose $\phi(a^1, \ldots, a^k)$ is not true. Then it is not the case that, for every $b \in [V_n]_\mathcal{U}$, $\psi(a^1, \ldots, a^k, b)$ is true, so there must be some $b = [b_n]_\mathcal{U} \in [V_n]_\mathcal{U}$ such that $\psi(a^1, \ldots, a^k, b)$ is false. Then, by the inductive hypothesis, $\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is true}\} \notin \mathcal{U}$, so—because $\mathcal{U}$ is an ultrafilter—$\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is false}\} \in \mathcal{U}$, and therefore $\{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is false}\} \in \mathcal{U}$. That means that $\{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \notin \mathcal{U}$, which is what we needed to show.

For the converse, suppose $\{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is true}\} \notin \mathcal{U}$, so $K = \{n \mid \phi(a_n^1, \ldots, a_n^k) \text{ is false}\} \in \mathcal{U}$. Then, for each $n \in K$, there must be some $b_n \in V_n$ so that $\psi(a_n^1, \ldots, a_n^k, b_n)$ is false. Taking $b = [b_n]_\mathcal{U}$, we have $\{n \mid \psi(a_n^1, \ldots, a_n^k, b_n) \text{ is false}\} \supseteq K \in \mathcal{U}$, so by the inductive hypothesis, $\psi(a^1, \ldots, a^k, b)$ is false, and therefore $\phi(a^1, \ldots, a^k)$ is false as well. □

The main consequence we need is a quick corollary.

Theorem 3.29. *If $\phi(x_1, \ldots, x_m)$ is a first-order formula, $k \leq m$, and $[a_n^{k+1}]_\mathcal{U}, \ldots, [a_n^m]_\mathcal{U} \in [V_n]_\mathcal{U}$, and*

$$X_n = \{(a^1, \ldots, a^k) \in V_n^k \mid \phi(a^1, \ldots, a^k, a_n^{k+1}, \ldots, a_n^m)\}$$

then

$$[X_n]_\mathcal{U} = \{(a^1, \ldots, a^k) \in V_\mathcal{U}^k \mid \phi(a^1, \ldots, a^k, [a_n^{k+1}]_\mathcal{U}, \ldots, [a_n^m]_\mathcal{U})\}.$$

[‡]The ability to take the sequence of witnesses b_n and combine them into the single witness $b = [b_n]_\mathcal{U}$ is of course crucial here.

Proof. Let $X = \{(a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, [a_n^{k+1}]_{\mathcal{U}}, \ldots, [a_n^m]_{\mathcal{U}})\}$ and let $a^i = [a_n^i]_{\mathcal{U}}$. Then

$$\begin{aligned}(a^1, \ldots, a^k) \in X &\text{ if and only if } \phi(a^1, \ldots, a^k, a^{k+1}, \ldots, a^m)\\ &\text{if and only if } \{n \mid \phi(a_n^1, \ldots, a_n^k, a_n^{k+1}, \ldots, a_n^m) \text{ is true}\} \in \mathcal{U}\\ &\text{if and only if } \{n \mid (a_n^1, \ldots, a_n^k) \in X_n\} \in \mathcal{U}\\ &\text{if and only if } (a^1, \ldots, a^k) \in [X_n]_{\mathcal{U}}.\end{aligned}$$

Therefore $X = [X_n]_{\mathcal{U}}$. □

Definition 3.31. If $\phi(x_1, \ldots, x_m)$ is a first-order formula, $k \leq m$, and $a^{k+1}, \ldots, a^m \in V$ then $\{(a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, a^{k+1}, \ldots, a^m)\}$ is a *definable* set.[§]

Then the main fact we need can stated briefly as:

Corollary 3.32. *Definable subsets of an ultraproduct are internal.*

We have also given the beginning of answer to the question of when a property of the ground graphs passes to the ultraproduct.

Corollary 3.33. *Suppose σ is a first-order formula with no variables. Then σ is true in $[G_n]_{\mathcal{U}}$ if and only if $\{n \mid \sigma$ is true in $G_n\} \in \mathcal{U}$.*

3.7 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 3.34 (Saturation)**.** *Suppose that, for each $i \in \mathbb{N}$, $X^i = [X_n^i]_{\mathcal{U}}$ 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 such that X_n^1 is non-empty, let $k \leq n$ be least so that $\bigcap_{i \leq k} X_n^i$ is non-empty and take $v_n \in \bigcap_{i \leq k} X_n^i$. (Otherwise choose v_n arbitrarily.)

[§]Some would call this notion "definable with parameters", and restrict "definable" to mean that $k = m$—that is, that no *parameters* $a^{k+1}, \ldots, a^m$ are needed in the definition.

For each i,

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

so $[v_n]_{\mathcal{U}} \in [X_n^i]_{\mathcal{U}} = X^i$. 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 3.35. *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 the general theme underlying many uses of ultraproducts. We finish this section by working through an example of saturation which, while not central to what comes later, illustrate the way saturation is typically used.

Definition 3.36. 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 3.37. *$[G_n]_{\mathcal{U}}$ 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 defined by a formula: the distance between v and w is greater than k if there does not exist a path from v to w of length at most k, which can

be written:

$$\begin{aligned}\neg[&v = w \vee \{v,w\} \in E \\ &\vee \exists y_1(\{v,y_1\} \in E \wedge \{y_1,w\} \in E) \\ &\vee \exists y_1 \exists y_2(\{v,y_1\} \in E \wedge \{y_1,y_2\} \in E \wedge \{y_2,w\} \in E) \\ &\vee \cdots \\ &\vee \exists y_1 \exists y_2 \cdots \exists y_{k-1}(\{v,y_1\} \in E) \wedge \{y_1,y_2\} \in E \wedge \cdots \wedge \{y_k,w\} \in E)].\end{aligned}$$

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.

3.8 Extending Sequences

We will later need to use saturation to take countable sequences and extend them. 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.

Pick representations $A^i = [A^i_n]_{\mathcal{U}}$. Then there is a "diagonal" set $A^* = [\bigcup_{i \leq n} A^i_n]_{\mathcal{U}}$. (One might wonder why we use $\bigcup_{i \leq n} A^i_n$ rather than just A^n_n, since $\bigcup_{i \leq n} A^i = A^n$. However we only know that when $i \leq j$, $A^i_n \subseteq A^j_n$ for 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 $A^f = [\bigcup_{i \le f(n)} A_n^i]_{\mathcal{U}}$. As long as $\{n \mid i \le f(n)\} \in \mathcal{U}$, we will have $A^i \subseteq A^f$. (Indeed, more generally, if $\{n \mid g(n) \le f(n)\} \in \mathcal{U}$ 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 3.38. *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 $A^i = [A_n^i]_{\mathcal{U}}$ and $B^i = [B_n^i]_{\mathcal{U}}$. Let $C = A^f$ where $f(n)$ is the smallest $j \le n$ such that, for all $i, i' \le j$, $A_n^i \subseteq B_n^{i'}$.

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

Similarly, whenever $f(n) \ge j$, $\bigcup_{i \le f(n)} A_n^i \subseteq B_n^j$. Since $\{n \mid f(n) \ge j\} \in \mathcal{U}$, we also have $C \subseteq B^j$. □

3.9 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.

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: 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 3.39 (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 suble. 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 $K \setminus N$ 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 first two requirements then the victorious sets form an ultrafilter on N.

3.10 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 $\mathbb{N}$.

*Though not necessarily a free filter—we have not promised that every cofinite set is victorious, and indeed, N is typically finite, so there are no cofinite sets.

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

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 [4, 26], and it is an interesting open question whether all the uses of ultrafilters can be channeled through ultraproducts [12].

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 [13], 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 [11, 19]. Ultraproducts also pay an important role in set theory [32, 35], where one can even consider ultraproducts of the entire universe of sets. The ultraproduct construction, with a suitable modification, also applies to structures in continuous logic [3], including Banach spaces and C*-algebras. (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.)

*As the old joke goes, 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.

Chapter 4

Integrals in Ultraproducts

4.1 Density is not Cardinality

We now turn to making sure 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 property is expressed by a first-order formula,

$$\forall x \exists y_1 \exists y_2 \left[\{x, y_1\} \in E \wedge \{x, y_2\} \in E \wedge y_1 \neq y_2 \wedge \forall z \neg\{x, z\} \in E \vee z = y_1 \vee z = y_2\right],$$

so the ultraproduct $[G_n]_{\mathcal{U}}$ has the same property.

For any fixed finite cycle k, $t_{\mathrm{C}_k}(\mathrm{C}_n) = 0$ once $n > k$. Since the absence of any copies of C_k is also expressed by a first-order formula and, for each k, there is only one n so that G_n contains a copy of C_k, $[G_n]_{\mathcal{U}}$ 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 $[G_n]_{\mathcal{U}}$ in this case: uncountably many chains which stretch forever in both directions.

However $\lim_{n\to\mathcal{U}} t_{\mathrm{K}_2}(G_n) = 0$: the graph G_n has n vertices and n edges, so $t_{\mathrm{K}_2}(G_n) = \frac{n}{n^2} = \frac{1}{n}$. 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 $t_{\mathrm{K}_2}(G'_n) = 1$ for every n, the density of edges in $[G'_n]_{\mathcal{U}}$ should 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.

4.2 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 only one reasonable 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 2.3 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 4.1. *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 2, 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 4.2. *Whenever $X \subseteq [V_n]_{\mathcal{U}}$, 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 $\lim_{n\to\mathcal{U}}$.

μ inherits many of the rules we expect of probability from the ground models:

Lemma 4.3. *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

$$\begin{aligned}
\mu(X \cup Y) &= \lim_{n \to \mathcal{U}} \mu_n(X_n \cup Y_n) \\
&= \lim_{n \to \mathcal{U}} (\mu_n(X_n) + \mu_n(Y_n) - \mu_n(X_n \cap Y_n)) \\
&= \lim_{n \to \mathcal{U}} \mu_n(X_n) + \lim_{n \to \mathcal{U}} \mu_n(Y_n) - \lim_{n \to \mathcal{U}} \mu_n(X_n \cap Y_n) \\
&= \mu(X) + \mu(Y) - \mu(X \cap Y).
\end{aligned}$$

□

The conventional setting for probability theory is a *σ-algebra*, in which we have not only finite unions and intersections, but countable ones.

Definition 4.4. If $\mathcal{B} \subseteq \mathcal{P}(V)$ (the power set of V), we say $\mathcal{B}$ is a *σ-algebra* if:

- $\emptyset \in \mathcal{B}$ and $V \in \mathcal{B}$,
- whenever $B \in \mathcal{B}$, $V \setminus B \in \mathcal{B}$,
- whenever $B_i \in \mathcal{B}$ for every i, $\bigcap_{i \in \mathbb{N}} B_i \in \mathcal{B}$ and $\bigcup_{i \in \mathbb{N}} B_i \in \mathcal{B}$.

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.

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. However, 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 4.5. *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 3.35: 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 $i > k$.

So what remains is showing that $\mu(\bigcup_{i\leq k} B^i) = \sum_{i\leq k} \mu(B^i)$ where each B^i is internal, which uses a similar argument to the previous lemma. Pick a representation $B^i = [B^i_n]_{\mathcal{U}}$ for each $i \leq k$.

We claim that $[\bigcup_{i\leq k} B^i_n]_{\mathcal{U}}$ is a representation of $\bigcup_{i\leq k} B^i$. Let $K = \{n \mid v_n \in \bigcup_{i\leq k} B^i_n\}$ and, for each $i \leq k$, let $K_i = \{n \mid v_n \in B^i_n\}$, so $K = \bigcup_{i\leq k} K_i$. Then

$$\begin{aligned}
[v_n]_{\mathcal{U}} \in [\bigcup_{i\leq k} B^i_n]_{\mathcal{U}} &\text{ if and only if } K \in \mathcal{U} \\
&\text{if and only if there is some } i \leq k \text{ so that } K_i \in \mathcal{U} \\
&\text{if and only if there is some } i \leq k \text{ so that } [v_n]_{\mathcal{U}} \in B^i \\
&\text{if and only if } [v_n]_{\mathcal{U}} \in \bigcup_{i\leq k} B^i.
\end{aligned}$$

Since $B^i \cap B^j = \emptyset$ for $i \neq j$, it follows that $\{n \mid B^i_n \cap B^j_n = \emptyset\} \in \mathcal{U}$ for each $i \neq j$, and taking the intersection over the finitely many pairs i, j, also

$$\{n \mid \text{ for each } i \neq j,\ B^i_n \cap B^j_n = \emptyset\} \in \mathcal{U}.$$

Therefore

$$\mu(\bigcup_{i\leq k} B^i) = \mu([\bigcup_{i\leq k} B^i_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} \frac{|\bigcup_{i\leq k} B^i_n|}{|V_n|} = \sum_{i\leq k} \lim_{n\to\mathcal{U}} \frac{|B^i_n|}{|V_n|} = \sum_{i\leq k} \mu(B^i).$$

□

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 4.6. *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. Define 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 $m \leq n$ 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 i, $\mu(\bigcup_{j\leq i} A^j) \leq c < c + \epsilon$, and therefore $\{n \mid \frac{|\bigcup_{i\leq j} A^i_n|}{|V_n|} \leq c + \epsilon\} \in \mathcal{U}$, so $\{n \mid A^i_n \subseteq \bigcup_{i\leq f^\epsilon(n)} A^i_n\} \in \mathcal{U}$, and therefore $A^i \subseteq A^{+\epsilon}$.

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 3.38, 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 often apply this lemma when the sets A^i are pairwise disjoint, in which case $\lim_{m\to\infty} \mu(\bigcup_{i\leq m} A^i) = \sum_i \mu(A_i)$.

4.3 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 4.7. 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 4.8. *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 4.6 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 by to any set which is within a null set of an internal set:

Definition 4.9. 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 4.10. *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 4.6, 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 4.11. *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 has measure 0.

Lemma 4.12. *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 4.13. *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 $\mu(X) = 0$. □

4.4 Probability Spaces on $[V_n]^k_{\mathcal{U}}$

The sets we're most interested in the probability of are not subsets of $[V_n]_{\mathcal{U}}$ itself, but rather 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 $[V_n^k]_{\mathcal{U}}$. That is, we can forget about the fact that the elements of V_n^k happen to be tuples: taking $W_n = V_n^k$, we can apply the previous section to $[W_n]_{\mathcal{U}} = [V_n^k]_{\mathcal{U}} = [V_n]^k_{\mathcal{U}}$.

Definition 4.14. 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-approximation* 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 4.15. *$\mathcal{B}(\mu_k)$ 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 $[V_n]^2_{\mathcal{U}}$ should have 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, we nonetheless expect these approaches to double integrals to give the same value: for Lebesgue measure this is Fubini's Theorem, a familiar fact from multi-variable calculus.

We have to be careful here—these approaches *do* all give the same value, but not because of Fubini's Theorem. Fubini's Theorem applies to the *product* of probability measure spaces, and we will see later that $([V_n]_{\mathcal{U}}^k, \mathcal{B}(\mu_k), \mu_k)$ is *not* a product space.

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 $(V^k, \mathcal{B}(\mu_k), \mu_k)$ 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 $([V_n]_{\mathcal{U}}^k, \mathcal{B}(\mu_k), \mu_k)$ relate to each other. The core properties are summed up by the following definition:

Definition 4.16. 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

$$X_{x_1,\ldots,x_k} = \{(x_{k+1}, \ldots, x_r) \in V^{r-k} \mid (x_1, \ldots, x_k, x_{k+1}, \ldots, x_r) \in X\}.$$

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 and Cylinder Sets) let $r < k$ be given, and suppose that, for each $s \in \binom{[k]}{r}$, we have a set $B_s \in \mathcal{B}_r$; then

$$\mathcal{K}(\{B_s\}_{s\in\binom{[k]}{r}}) = \{(x_1,\dots,x_k) \mid \text{for every } \{s_1,\dots,s_r\} \in \binom{[k]}{r}, (x_{s_1},\dots,x_{s_r}) \in B_s\} \in \mathcal{B}_k,$$

- (Fubini Property) whenever $X \in \mathcal{B}_{k+r}$
  - the set of $x_1,\dots,x_k \in V$ such that $X_{x_1,\dots,x_k} \in \mathcal{B}_r$ belongs to $\mathcal{B}_k$ and has μ_k-measure 1, and
  - $\mu_{k+r}(X) = \int \mu_r(X_{x_1,\dots,x_k}) d\mu_k(x_1,\dots,x_k)$.

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.

Consider the product and cylinder set property when $r = 1$; then it just says that if $B_i \in \mathcal{B}_1$ for each $i \le k$, $\mathcal{K}(\{B_i\}_{i\in\binom{[k]}{1}}) = \prod_{i\le k} B_i \in \mathcal{B}_k$. The full property also calls for more general types of product-like-sets, called *cylinder sets*, like

$$\{(x_1,x_2,x_3) \mid (x_1,x_2) \in B_{1,2} \text{ and } (x_1,x_3) \in B_{1,3} \text{ and } (x_2,x_3) \in B_{2,3}\}.$$

However we are allowed to—as we typically will—have additional sets beyond those required by the cylinder set 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.

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.

We also note that having the Fubini property for measures lifts it to integrals:

Lemma 4.17. *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},*

$$\int f\, d\mu_{k+r} = \int \left(\int f(x_1,\dots,x_{k+r})\, d\mu_r(x_{k+1},\dots,x_{k+r}) \right) d\mu_k(x_1,\dots,x_k).$$

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

$$\begin{aligned}
\int f \, d\mu_{k+r} &= \int g \, d\mu_{k+r} + e \, |e| < \epsilon \\
&= \sum_{i \le d} c_i \mu_{k+r}(B_i) + e \text{ where } |e| < \epsilon \\
&= \sum_{i \le d} c_i \int \mu_r((B_i)_{x_1,\ldots,x_k}) \, d\mu_k + e \text{ where } |e| < \epsilon \\
&= \int \sum_{i \le d} c_i \mu_r((B_i)_{x_1,\ldots,x_k}) \, d\mu_k + e \text{ where } |e| < \epsilon \\
&= \int \left(\int f(x_1, \ldots, x_{k+r}) \, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right) d\mu_k(x_1, \ldots, x_k) + e' \text{ where } |e'| < 2\epsilon.
\end{aligned}$$

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.

Theorem 4.18. *The collection of probability measure spaces* $([V_n]_{\mathcal{U}}^k, \mathcal{B}(\mu_k), \mu_k)$ *is a Keisler graded probability space.*

Proof. The proof is routine, but long and technical. 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 $A^\pi = [A_n^\pi]_{\mathcal{U}}$ and $(\mu_k)_n(A_n^\pi) = (\mu_k)_n(A_n)$, 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 $\mu_k(A_\epsilon^\pi) = \mu_k(A_\epsilon) < \epsilon$ and $(A^\pi \triangle X^\pi) \subseteq A_\epsilon^\pi$, so A^π is a μ_k-approximation to X^π, so $X^\pi \in \mathcal{B}^k$ and $\mu_k(X^\pi) = \mu_k(A^\pi) = \mu_k(A) = \mu_k(X)$.

Cylinder sets of internal sets are certainly internal: if each $B_s = [B_{s,n}]_{\mathcal{U}} \subseteq$

$[V_n]^r_{\mathcal{U}}$ is internal then

$$\{(x_1,\dots,x_k) \mid \text{for all } \{s_1,\dots,s_r\} \in \binom{[k]}{r}, (x_{s_1},\dots,x_{s_r}) \in B_s\}$$
$$=[\{(x_1,\dots,x_k) \mid \text{for all } \{s_1,\dots,s_r\} \in \binom{[k]}{r}, (x_{s_1},\dots,x_{s_r}) \in B_{s,n}\}]_{\mathcal{U}}.$$

Suppose that, for each $s \in \binom{[k]}{r}$, we have $X_s \in \mathcal{B}_r$. Consider a μ_r-approximation A_s of X_s. Then

$$\mathcal{K}(\{A_s\}) \triangle \mathcal{K}(\{X_s\}) \subseteq \bigcup_{s_0 \in \binom{[k]}{r}} \mathcal{K}(\{A_{s_0} \triangle X_{s_0}\} \cup \{[V_n]^r\}_{s \in \binom{[k]}{r} \setminus \{s_0\}}).$$

Therefore, for any $\epsilon > 0$, we may choose internal sets $A_{s,\epsilon/\binom{k}{r}} \supseteq (A_s \triangle X_s)$. Then

$$\mathcal{K}(\{A_s\}) \triangle \mathcal{K}(\{X_s\}) \subseteq \bigcup_{s_0 \in \binom{[k]}{r}} \mathcal{K}(\{A_{s_0,\epsilon/\binom{k}{r}} \triangle X_{s_0}\} \cup \{[V_n]^r\}_{s \in \binom{[k]}{r} \setminus \{s_0\}})$$

and

$$\mu_k(\bigcup_{s_0 \in \binom{[k]}{r}} \mathcal{K}(\{A_{s_0,\epsilon/\binom{k}{r}} \triangle X_{s_0}\} \cup \{[V_n]^r\}_{s \in \binom{[k]}{r} \setminus \{s_0\}})) \leq \binom{k}{r} \frac{\epsilon}{\binom{k}{r}} = \epsilon.$$

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, $A_{x_1,\dots,x_k}$ is also internal for any choice of $x_1,\dots,x_k$, and therefore is in $\mathcal{B}^r$.

The ground graphs are finite, so the analogous statements hold by counting:

$$\begin{aligned}\int (\mu_r)_n(A_{x_1,\dots,x_k}) d(\mu_k)_n(x_1,\dots,x_k) &= \frac{1}{|V_n|^k} \sum_{(x_1,\dots,x_k) \in V_k^n} \frac{|(A_n)_{x_1,\dots,x_k}|}{|V_n|^r} \\ &= \frac{1}{|V_n|^{k+r}} \sum_{(x_1,\dots,x_k) \in V_k^n} |(A_n)_{x_1,\dots,x_k}| \\ &= \frac{|A_n|}{|V_n|^{k+r}} = \mu_n^{k+r}(A_n).\end{aligned}$$

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

$$\lim_{n \to \mathcal{U}} \int (\mu_r)_n(A_{x_1,\dots,x_k}) d(\mu_k)_n(x_1,\dots,x_k) = \int \mu_r(A_{x_1,\dots,x_k}) d\mu_k(x_1,\dots,x_k).$$

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

$$\begin{aligned}(\mu_{k+r})_n(A_n) &= \int (\mu_r)_n(A_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1, \ldots, x_k) \\ &= \sum_{j \leq J} \int_{C_n^j} (\mu_r)_n(A_{x_1,\ldots,x_k}) d(\mu_k)_n(x_1, \ldots, x_k) \\ &\approx \sum_{j \leq J} (\mu_k)_n(C_n^j)(j + 1/2)\epsilon.\end{aligned}$$

More precisely, since each interval I_j has radius $\epsilon/2$, we have

$$\left| \int (\mu_r)_n(A_{x_1,\ldots,x_n}) d(\mu_k)_n(x_1, \ldots, x_n) - \sum_{j \leq J} (\mu_k)_n(C_n^j)(j + 1/2)\epsilon \right| < \epsilon/2.$$

Consider the corresponding sets in the ultraproduct, $C^j = [C_n^j]_{\mathcal{U}}$. 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 $(x_1, \ldots, x_k) = ([x_n^1]_{\mathcal{U}}, \ldots, [x_n^k]_{\mathcal{U}}) \in C^j$ then

$$\mu_r(A_{x_1,\ldots,x_n}) = \lim_{n \to \mathcal{U}} (\mu_r)_n((A_n)_{x_n^1,\ldots,x_n^k}) \in \overline{I_j} = [j\epsilon, (j+1)\epsilon].$$

Moreover, $\mu_k(C^j) = \lim_{n \to \mathcal{U}} (\mu_k)_n(C_n^j)$. 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,

$$\begin{aligned}\mu_{k+r}(A) &= \lim_{n\to\mathcal{U}}(\mu_{k+r})_n(A_n)\\ &\approx \lim_{n\to\mathcal{U}}\sum_{j\leq J}(\mu_k)_n(C_n^j)(j+1/2)\epsilon\\ &= \sum_{j\leq J}\int_{C_j}\mu_r(A_{x_1,\ldots,x_k})d\mu_k(x_1,\ldots,x_k)\\ &= \int\mu_r(A_{x_1,\ldots,x_k})d\mu_k(x_1,\ldots,x_k)\end{aligned}$$

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 $D\subseteq V^k_{\mathcal{U}}$ 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 4.13, $\mu(D_d)=0$.

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

$$\begin{aligned}\mu_{k+r}(X) &= \mu_{k+r}(A)\\ &= \int\mu_r(A_{x_1,\ldots,x_k})d\mu_k(x_1,\ldots,x_k)\\ &= \int\mu_r(X_{x_1,\ldots,x_k})d\mu_k(x_1,\ldots,x_k).\end{aligned}$$

□

So the spaces $([V_n]^k_{\mathcal{U}}, \mathcal{B}(\mu_k), \mu_k)$ have the relationships we expect. In particular, we can calculate measures by integrating over one of the coordinates. For example, let $T_{C_3} = \{(x, y, z) \mid \{x, y\} \in E \text{ and } \{x, z\} \in E \text{ and } \{y, z\} \in E\}$ be the set of triangles. Then $(T_{C_3})_x = E \cap (E_x \times E_x)$ 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 $\{x, y\} \in E$, $(T_{C_3})_{(x,y)} = E_x \cap E_y$.

If we want to find the measure of the set of triangles, $\mu_3(T_{C_3})$, we can calculate either $\int \mu_2((T_{C_3})_x)d\mu(x)$ or $\int \mu((T_{C_3})_{(x,y)})d\mu_2(x, y)$ if one of these is easier to calculate. Indeed, we will routinely find ourselves switching between different integrals which calculate the same measure.

4.5 Subgraph Density

The work in the previous sections gives us the properties we mostly need to identify the sets we want to work with.

Definition 4.19. 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 *infinite* if, for all $v \in V$, $\mu_1(\{v\}) = 0$.

We often say "$G = (V, E, \mu_1)$ is a measureable 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 now, finally, ready to define subgraph density in an ultraproduct and, more generally, in a measurable graph.

Definition 4.20. 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 $(x_1, \ldots, x_k) \in V^k_{\mathcal{U}}$ 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 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 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, $T_H = [T_H(G_n)]_n$, where $T_H(G_n)$ is the set of copies of H in G_n.

The definition of a Keisler measure—specifically, the cylinder set property with $r = 2$—ensures that T_H is measurable.

Definition 4.21. We define $t_H(G)$, the subgraph density of H in G, to be $\mu(T_H)$.

When $G = [G_n]_{\mathcal{U}}$, since the T_H are internal, we have $t_H([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} \mu_n(T_H(G_n)) = \lim_{n\to\mathcal{U}} t_H(G_n)$.

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, which we can think of as weighted graphs:

Definition 4.22. If $H = (W, F)$ then

$$t_H(f) = \int \prod_{\{v_i,v_j\}\in F} f(x_i, x_j)\, d\mu_{|W|}.$$

Then $t_H(\chi_E) = t_H(E)$.

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 4.23. Let $H = (W, F)$ with $W = \{w_1, \ldots, w_k\}$. For any $d \leq k$ 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

$$t_{\mathrm{K}_3}(G, x, y) = \int \chi_E(x, z)\chi_E(y, z)\, d\mu_1(z),$$

the number of ways to extend a pair x, y to a triangle. 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.)

4.6 Quasirandomness

We can identify quasirandom (rather than just ϵ-quasirandom) measurable graphs.

Definition 4.24. $G = (V, E, \mu_1)$ is *quasirandom* if $t_{\mathrm{C}_4}(G) = (t_{\mathrm{K}_2}(G))^4$.

The theorem that $t_{\mathrm{C}_4}(G) \geq (t_{\mathrm{K}_2}(G))^4$ for any graph G still applies, so again quasirandomness means that $t_{\mathrm{C}_4}(G)$ is as small as possible: the value of $t_{\mathrm{K}_2}(G)$ 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 4.25. *$[G_n]_{\mathcal{U}}$ is quasirandom if and only if, for every $\epsilon > 0$, $\{n \mid G_n$ is ϵ-quasirandom$\} \in \mathcal{U}$.*

Proof. Suppose $[G_n]_{\mathcal{U}}$ is not quasirandom. Let $p = t_{\mathrm{K}_2}([G_n]_{\mathcal{U}})$. Then there is some $\epsilon \in (0, 1)$ so that $t_{\mathrm{C}_4}([G_n]_{\mathcal{U}}) > p^4 + \epsilon$. Let $J = \{n \mid |t_{\mathrm{C}_4}(G_n) - t_{\mathrm{C}_4}([G_n]_{\mathcal{U}})| < \epsilon/2\}$ and $K = \{n \mid |t_{\mathrm{K}_2}(G_n) - p| < \epsilon/60\}$. Then $J, K \in \mathcal{U}$, so $J \cap K \in \mathcal{U}$, and for any $n \in J \cap K$ we have

$$\begin{aligned} t_{\mathrm{C}_4}(G_n) &\geq t_{\mathrm{C}_4}([G_n]_{\mathcal{U}}) - \epsilon/2 \\ &> p^4 + \epsilon/2 \\ &> (t_{\mathrm{K}_2}(G_n) - \epsilon/60)^4 + \epsilon/2 \\ &\geq t_{\mathrm{K}_2}(G_n) + \epsilon/4. \end{aligned}$$

Conversely, if there is some $\epsilon > 0$ such that $\{n \mid G_n \text{ is } \epsilon\text{-quasirandom}\} \notin \mathcal{U}$ then

$$t_{\mathrm{C}_4}([G_n]_{\mathcal{U}}) = \lim_{n \to \mathcal{U}} t_{\mathrm{C}_4}(G_n) \geq \lim_{n \to \mathcal{U}} (t_{\mathrm{K}_2}(G_n))^4 + \epsilon = (t_{\mathrm{K}_2}([G_n]_{\mathcal{U}}))^4 + \epsilon.$$

□

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 4.26. *If $G = (V, E, \mu_1)$ is quasirandom then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $\mathcal{B}_1$,*

$$\mu_1(E \cap (X \times Y)) = \mu(E)\mu(X)\mu(Y).$$

The proof we gave of the analogous theorem in the finite setting, Theorem 2.28, goes through essentially unchanged. Rather than repeat that here, we will wait for Corollary **??**, where we will give another proof building on the machinery we develop below.

Using this, we can give our long delayed proof of Theorem 2.32, 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 4.27. *If $G = (V, E, \mu_1)$ is quasirandom then for every finite graph $H = (W, F)$, $t_H(G) = p^{|F|}$ where $p = t_{\mathrm{K}_2}(G) = \mu_2(E)$.*

Proof. First, consider the 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 4.26—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 4.26 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

$$t_H(G) = \mu_{|W|}(\{(x_1, \ldots, x_{|W|}) \mid \text{for each } \{v_i, v_j\} \in F,\ \{x_i, x_j\} \in E_{\mathcal{U}}\})$$
$$= \int \prod_{\{v_i, v_j\} \in F} \chi_E(x_i, x_j) d\mu_{|W|}(x_1, \ldots, x_{|W|}).$$

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 $t_H(G) = p \cdot p^{|F|-1} = p^{|F|}$. $\square$

4.7 Consequences for Finite Graphs

We can use Theorem 4.27 to give a proof of Theorem 2.32. The technique is typical of the way we obtain results about finite graphs using ultraproducts.

Theorem (2.32)**.** *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 $t_{K_2}(G) = p$ and V sufficiently large, $|t_H(G) - t_{K_2}(G)^{|F|}| < \epsilon$.*

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 4.27, $\lim_{n \to \mathcal{U}} t_H(G_n) = t_H([G_n]_{\mathcal{U}}) = p^{|F|}$. 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 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 4.27 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 service of the ultraproduct in this argument is hiding all those ϵ's and δ's by letting them reach the limit value of 0.

4.8 Sampling

Every ultraproduct of finite graphs is a measurable graph; we next verify a converse of sorts—that, at least for questions of subgraph density, every infinite measurable graph resembles an ultraproduct.

There is a canonical way to sample a finite graph from $G = (V, E, \mu_1)$: 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.

Lemma 4.28. *Let $G = (V, E)$ be a 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 $n \geq m$,*

$$\mathbb{P}[|t_H(G) - t_H(\mathbf{G}_n)| < \epsilon] > 1 - \epsilon.$$

Proof. Our arguments look much like they did for the random graph.

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 $\mathbf{1}_\pi$ be be the random variable which is 1 if $\pi(i) = \mathbf{v}_{\pi(i)}$ is an actual copy of H and 0 otherwise. When π is injective, observe that $\mathbf{1}_\pi$ is simply the probability that, when we select the points $\mathbf{v}_{\pi(1)}, \ldots, \mathbf{v}_{\pi(k)}$ from V, that we get an actual copy of H. Therefore

$$\mathbf{E}(\mathbf{1}_\pi) = \mu_{|W|}(T_H(G)) = t_H(G),$$

more or less by definition—these are actually the same notion written in different notation.

By the linearity of expectation,

$$\mathbf{E}(t_H(\mathbf{G}_n)) = \frac{1}{n^{|W|}} \sum_\pi \mathbb{E}(\mathbf{1}_\pi).$$

Since there are at most $Cn^{|W|-1}$ non-injective functions $\pi : W \to \{1, \ldots, n\}$ for some constant C, ths means that

$$\mathbb{E}(t_H(\mathbf{G}_n)) = t_H(G) + O(\frac{1}{n}).$$

As in the proof of Theorem 2.7, we use McDiarmid's inequality. This time the random variables are the $\mathbf{v}_i$ and $f(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 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 alue $\frac{2\epsilon^2}{|W|^2}$ is fixed, so by choosing n large enough, we can make this bound as small as we like. □

This immediately implies the same claim for finitely many graphs at once.

Corollary 4.29. *Let $G = (V, E)$ be a 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 4.30. *Let $G = (V, E)$ be a 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.

4.9 The Possible Subgraph Densities are Compact

One reason to consider infinite 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 4.31. *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 measurable graph G so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G)| < \epsilon$.*

Then there is a 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 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 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 measurabe graph G_k so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G_k)| < 1/k$. Then choose a sequence of graphs $\mathbf{G}_{k,n}$ by sampling so that $\lim_{n\to\infty} t_H(\mathbf{G}_{k,n}) = t_H(G_k)$ for all H. (If G_k is finite, we may just take $\mathbf{G}_{k,n} = G_k$ for all n.)

Let us interleave these sequences by setting $G^*_{2^k(n-1)+2^{k-1}} = \mathbf{G}_{k,n}$. (That is, the odd values $G^*_1, G^*_3, \ldots$ enumerate $\mathbf{G}_{1,n}$, then $G^*_2, G^*_6, G^*_{10}, \ldots$ enumerate $\mathbf{G}_{2,n}$, and so on.)

We would like to build a free filter $\mathcal{F}$ (which we will then extend to an ultrafilter) so that for each H, $\lim_{n\to\mathcal{F}} t_H(G^*_n) = \zeta(H)$. For each H and each $\epsilon > 0$, consider $J_{H,\epsilon} = \{n \mid |t_H(G^*_n) - \zeta(H)| < \epsilon\}$. Let $\mathcal{F}^-$ be the collection of all sets $J_{H,\epsilon}$. We would like to let $\mathcal{F}$ be the smallest free filter extending $\mathcal{F}^-$. More precisely, let $\mathcal{F}$ consist of all sets K such that there are finitely may sets $J_1, \ldots, J_d \in \mathcal{F}^-$ such that $(J_1 \cap \cdots \cap J_d) \setminus K$ is finite. (For formal reasons, we include the case where $k = 0$, in which case we require that $\mathbb{N} \setminus K$ is finite.) We must show that $\mathcal{F}$ is a free filter.

Certainly $\mathcal{F}$ is closed under finite intersections, and if $K \in \mathcal{F}$ and $K \subseteq K'$ then $K' \in \mathcal{F}$, by the definition of $\mathcal{F}$. If K is cofinite then $K \in \mathcal{F}$, again by definition. So it suffices to show that $\emptyset \notin \mathcal{F}$—that is, that whenever $J_{H_1,\epsilon_1}, \ldots, J_{H_d,\epsilon_d} \in \mathcal{F}^-$, $\bigcap_{i \leq d} J_{H_i,\epsilon_i}$ is infinite. But this follows by our construction of the $\mathbf{G}_{k,n}$: choosing k sufficiently large and then m sufficiently large, we have $\{2^k n + 2^{k-1} \mid n \geq m\} \subseteq \bigcap_{i \leq d} J_{H_i,\epsilon_i}$.

Therefore, for every H and every $\epsilon > 0$, we have $\{n \mid |t_H(G_n^*) - \zeta(H)| < \epsilon\} \in \mathcal{F}$, and therefore

$$\lim_{n \to \mathcal{F}} t_H(G_n^*) = \zeta(H).$$

Choosing $\mathcal{U}$ to be an ultrafilter extending $\mathcal{F}$, it follows that $t_H([G_n^*]_\mathcal{U}) = \zeta(H)$ for all finite graphs H. □

We should note that the same results apply if we restrict ourselves to graphs omitting certain subgraphs entirely.

Theorem 4.32. *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 (for instance, by Łos' Theorem). □

4.10 Shifting Measures

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 finite graph by making small changes—say, deleting a small fraction of the vertices, or changing a small fraction of the edges. Some of these modifications become natural analytic 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 4.33. 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 whenever $B \in \mathcal{B}$ and $\mu(B) = 0$, also $\nu(B) = 0$.

The key fact we will need is the Radon-Nikodym Theorem.

Theorem (Radon-Nikodym Inequality)**.** *If $\nu \ll \mu$ then there is a measurable function f, called the* Radon-Nikodym derivative *of ν, such that, for every measurable $B \in \mathcal{B}$,*

$$\nu(B) = \int_B f \, d\mu.$$

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

Lemma 4.34. *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 $B \in \mathcal{B}_k$, we define

$$\nu_k(B) = \int \chi_B(x_1, \ldots, x_k) \prod_{i\leq k} f(x_i) \, d\mu_k.$$

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

$$\begin{aligned}
\nu_{k+r}(X) &= \int \chi_X(x_1, \ldots, x_{k+r}) \prod_{i\leq k+r} f(x_i) \, d\mu_{k+r} \\
&= \int \left[\int \chi_X(x_1, \ldots, x_{k+r}) \prod_{k<i\leq k+r} f(x_i) \, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right] \prod_{i\leq k} f(x_i) \, d\mu_k(x_1, \ldots, x_k) \\
&= \int \nu_r(X_{x_1,\ldots,x_k}) \prod_{i\leq k} f(x_i) \, d\mu_k \\
&= \int \nu_r(X_{x_1,\ldots,x_k}) \, d\nu_k.
\end{aligned}$$

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 $B \in \mathcal{B}_{k+1}$, we have

$$\begin{aligned}\nu'_{k+1}(B) &= \int \nu'_k(B_{x_1})\, d\nu'_1(x_1) \\ &= \int \nu_k(B_{x_1})\, d\nu_1(x_1) \\ &= \nu_{k+1}(B).\end{aligned}$$

□

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 sets as G, but the new measure.

4.11 Turan's Theorem

The results of the previous sections will allow us to move back and forth freely between subgraph densities in measurable graphs and in finite graphs.

Theorem 4.35. *For any $t \geq 3$, if G is a measurable graph with $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{t-1}$ then $|T_{\mathrm{K}_t}(G)| > 0$.*

Proof. Suppose not, so there is a measurable graph $G = (V, E, \mu)$ with $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{t-1}$ and $T_{\mathrm{K}_t}(G) = \emptyset$. By Theorem 4.32, there is a measurable graph G with $T_{\mathrm{K}_t}(G) = \emptyset$ maximizing the value $t_{\mathrm{K}_2}(G)$ among all such measurable graphs.

Claim 1. For almost every x, $\deg_G(x) = t_{\mathrm{K}_2}(G)$.

Proof. Suppose not. The idea is that we will pick a set B of vertices whose degree is too high and tweak the measure μ_1 so that B has a slightly higher measure than it did before. Since the underlying edge relation is the same, this won't create copies of K_t where there were none before. We hope that this modified graph has higher edge density, because the high degree vertices count for more.

Let $c = t_{\mathrm{K}_2}(G)$. Since G is not evenly distributed, choose some $\epsilon > 0$ so that the set of x with $\deg_G(x) - t_{\mathrm{K}_2}(G) > \epsilon$ has positive measure, and let B be the set of such x.

For each $\delta > 0$, we will define a modified measure ν_δ: we define

$$\nu_\delta(S) = \frac{1}{1+\delta}\left[(1+\frac{\delta}{\mu_1(B)})\mu_1(S\cap B) + \mu_1(S\setminus B)\right].$$

Then

$$\nu_\delta(V) = \frac{1}{1+\delta}\left[(1+\frac{\delta}{\mu_1(B)})\mu_1(B) + \mu_1(S\setminus B)\right] = \frac{1}{1+\delta}[\mu_1(B)+\delta+(1-\mu_1(B))] = 1,$$

so ν_δ is still a probability measure and $\nu_\delta \ll \mu_1$.

Let $G_\delta = (V, E, \nu_\delta)$. Then

$$\begin{aligned}
t_{\mathrm{K}_2}(G_\delta) &= \int \chi_E(x,y)\, d(\nu_\delta)_2 \\
&= \frac{1}{(1+\delta)^2}\left[\int \chi_E(x,y)\, d\mu_2 + 2\frac{\delta}{d}\int_{B\times V} \chi_E(x,y)\, d\mu_1(x)\, d\nu_\delta(y)\right] \\
&= \frac{1}{(1+\delta)^2}\left[\int \chi_E(x,y)\, d\mu_2 + 2\frac{\delta}{d}\int_{B\times V} \chi_E(x,y)\, d\mu_2 + 2\frac{\delta^2}{d^2}\int_{B\times B} \chi_E(x,y)\, d\mu_2\right] \\
&= \frac{1}{(1+\delta)^2}\left[c + 2\delta(c+\epsilon) + 2\delta^2\frac{\int_{B\times B}\chi_E(x,y)\, d\mu_2}{d^2}\right].
\end{aligned}$$

In particular,

$$\begin{aligned}
\lim_{\delta\to 0}\frac{t_{\mathrm{K}_2}(G_\delta) - t_{\mathrm{K}_2}(G)}{\delta} &= \lim_{\delta\to 0}\frac{\left[c + 2\delta(c+\epsilon) + 2\delta^2\frac{\int_{B\times B}\chi_E(x,y)\, d\mu_2}{d^2}\right] - c(1+\delta)^2}{\delta(1+\delta)^2} \\
&= \lim_{\delta\to 0}\frac{2(c+\epsilon) + 2\delta\frac{\int_{B\times B}\chi_E(x,y)\, d\mu_2}{d^2} - 2c - c\delta}{\delta(1+\delta)^2} \\
&= 2\epsilon,
\end{aligned}$$

so when δ is sufficiently small, we must have $t_{\mathrm{K}_2}(G_\delta) - t_{\mathrm{K}_2}(G) > 0$. But $T_{\mathrm{K}_t}(G_\delta) = T_{\mathrm{K}_t}(G) = \emptyset$, contradicting the maximality of $t_{\mathrm{K}_2}(G)$ among such graphs ⊣

Since G is evenly distributed, let $V' \subseteq V$ be a set with $\mu(V') = 1$ and, for all $x \in V'$, $\deg_G(x) > 1 - \frac{1}{t-1}$. We obtain a copy of K_t in G by successively choosing elements: given $x_1, \ldots, x_m$ a copy of K_m in V' with $m < t$, for each $i \le m$, $\{y \in V' \mid \{x_i, y\} \notin E\}$ has measure $< \frac{1}{t-1}$, and therefore the set of $y \in V'$ such that $\{x_i, y\} \in E$ for each i has measure $> 1 - \frac{m}{t-1} \ge 0$, 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_t, showing that $|T_{\mathrm{K}_t}(G)| > 0$. □

This immediately gives a finite version.

Corollary 4.36. *For each $t \geq 3$ and any $\epsilon > 0$, there is an n so that if G is a graph on $\geq n$ vertices with $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{t-1} + \epsilon$ then $|T_{\mathrm{K}_t}(G)| > 0$.*

Proof. Suppose not. Then for some t and some $\epsilon > 0$, for each n there is a graph G_n with $\geq n$ vertices and $t_{\mathrm{K}_2}(G_n) > 1 - \frac{1}{t-1} + \epsilon$ and $|T_{\mathrm{K}_t}(G_n)| = 0$.

Then take any ultrafilter $\mathcal{U}$ and consider $G = [G_n]_{\mathcal{U}}$: $t_{\mathrm{K}_2}(G) = \lim_{n \to \mathcal{U}} t_{\mathrm{K}_2}(G_n) \geq 1 - \frac{1}{t-1} + \epsilon$ while $T_{\mathrm{K}_t}(G) = \emptyset$ 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, there are sharper results available by other methods: these techniques are not well suited to identifying the exact number of edges at which copies of K_t appear.

As noted in Section 2.8, the graph where we partition V into $t-1$ parts of equal measure and the edges are exactly those pairs in distinct parts achieves the maximum bound of $t_{\mathrm{K}_2}(G) = 1 - \frac{1}{t-1}$ while having no copies of K_t.

4.12 Minimizing Density

More generally, if $t_{\mathrm{K}_2}(G) > 1 - \frac{1}{t-1}$, we could ask what the smallest possible value of $t_{\mathrm{K}_t}(G)$ is.

For triangles, we have the following bound.

Lemma 4.37 (Goodman's Bound)**.** *When G is a measurable graph,*

$$t_{\mathrm{K}_3}(G) \geq t_{\mathrm{K}_2}(G) \cdot (2t_{\mathrm{K}_2}(G) - 1).$$

The proof here—like most of the results in this section—is due to Razborov [41]. Razborov introduced a formalism—*flag algebras*—for doing calculations with various subgraph densities. Like many proofs developed in that formalism, the proofs are quite short, but depend on finding exactly the right combinations of graphs to make certain calculations go through.

Proof. We notice that

$$\begin{aligned}
t_{\mathrm{K}_3}(G) + t_{\mathrm{K}_2}(G) &= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + \int \chi_E(x,y)\,d\mu_2 \\
&= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + \int \chi_E(x,y)(1-\chi_E(x,z))\,d\mu_3 \\
&\qquad + \int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + \int \chi_E(x,y)(1-\chi_E(x,z))\chi_E(y,z)\,d\mu_3 \\
&\qquad + \int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\,d\mu_3 + \int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)\chi_E(y,z)\,d\mu_3 + \int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\,d\mu_3 \\
&\qquad + \int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&= \int \chi_E(x,y)(1-\chi_E(x,z))(1-\chi_E(y,z))\,d\mu_3 + 2\int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 \\
&\geq 2\int (\int \chi_E(x,y)d\mu_1(y))^2\,d\mu_1(x) \\
&\geq 2(\int \chi_E(x,y)\,d\mu_2)^2 \\
&= 2(t_{\mathrm{K}_2}(G))^2.
\end{aligned}$$

Rearranging, we have

$$t_{\mathrm{K}_3}(G) = 2(t_{\mathrm{K}_2}(G))^2 - t_{\mathrm{K}_2}(G) = t_{\mathrm{K}_2}(G)(2t_{\mathrm{K}_2}(G) - 1).$$

□

This gives a lower bound: a graph with $t_{\mathrm{K}_2}(G)$ edges needs at least $t_{\mathrm{K}_2}(G)(2t_{\mathrm{K}_2}(G) - 1)$ triangles.

Consider the evenly t-partite graphs from the previous section: the graph P_t has $t_{\mathrm{K}_2}(P_t) = 1 - \frac{1}{t}$ and $t_{\mathrm{K}_3}(P_t) = (1 - \frac{1}{t})(1 - \frac{2}{t})$. Noticing that

$$(1 - \frac{1}{t})(2(1 - \frac{1}{t}) - 1) = (1 - \frac{1}{t})(1 - \frac{2}{t}),$$

this means that the even t-partite graphs have as few triangles as possible. Since the upper and lower bounds match in this case, this settles the question of how few triangles a graph can have, *if* the graph happens to have $1 - \frac{1}{t}$ edges for an integer t.

When $c = t_{\mathrm{K}_2}(G)$ does not have the form $1 - \frac{1}{t}$, there is a way to interpolate between the t-partite graphs: consider a $t+1$-partite graph where t of the parts have equal size and the last part is smaller. We can think of these graphs as coming from a process where we begin with a "one part" graph—the graph with no edges. We let the second part begin growing, so we get a graph with two parts, all edges between those parts, and one part is smaller than the other. As the smaller part grows, we approach the even bipartite graph. Once we reach two equal parts, a third part begins growing until it equals the first two, at which point a fourth part begins, and so on.

More precisely, given t and $a \leq \frac{1}{t+1}$, let us define $P_{t,a} = (V, E)$ to be the graph with $t+1$ parts—$V = V_0 \cup \cdots \cup V_t$ where $\mu(V_0) = a$ and, for $i > 0$, $\mu(V_i) = \frac{1-a}{t}$. Let us define:

Definition 4.38. $\mathfrak{g}_r(x) = t_{\mathrm{K}_r}(P_{t,a})$ where t, a are chosen so that $t_{\mathrm{K}_2}(P_{t,a}) = x$.

The graphs $P_{t,a}$ give us a candidate for the graph with as few copies of K_r as possible. Indeed, we have

Theorem 4.39 ([44])**.** *For every r and every x, the graph $P_{t,a}$ with t, a chosen so that $t_{\mathrm{K}_2}(P_{t,a}) = x$ minimizes $t_{\mathrm{K}_r}(G)$ among all graphs with $t_{\mathrm{K}_2}(G) = x$.*

The case for triangles, where $r = 3$ is shown in [42], and when $r = 4$ in [39].

The general proof is long and involved, so we will focus on a special case: when $r = 3$ and $x \in (1/2, 2/3)$ (that is, when $t = 2$). This will be enough to illustrate the main technique behind these arguments: the use of a *differentiable* structure in the measurable graphs.

First, let us carry out some calculations to figure out the value of $\mathfrak{g}_3$. Given t, a, we have

$$t_{\mathrm{K}_2}(P_{t,a}) = a(1-a) + (1-a)(a + \frac{(1-a)}{t}(t-1)) = -a^2(1+\frac{1}{t}) + 2a\frac{1}{t} + 1 - \frac{1}{t}.$$

In particular, if we want the edge density to be x and $x \in (1 - \frac{1}{t}, 1 - \frac{1}{t+1}]$, we need

$$\begin{aligned} a &= \frac{2\frac{1}{t} \pm \sqrt{\frac{4}{t^2} + 4(1 - \frac{1}{t} - x)(1 + \frac{1}{t})}}{2(1 + 1/t)} \\ &= \frac{2 \pm \sqrt{4 + 4(t - 1 - tx)(1 + t)}}{2t + 2} \\ &= \frac{1 \pm \sqrt{t(t - (t+1)x)}}{t+1}. \end{aligned}$$

Since we need $a \leq \frac{1}{t+1}$, notice that the larger root satisfies

$$\frac{1 \pm \sqrt{t(t-(t+1)x)}}{t+1} \geq \frac{1 \pm \sqrt{t(t-(t+1)(1-\frac{1}{t+1}))}}{t+1} = \frac{1}{t+1},$$

so we need $a = \frac{1-\sqrt{t(t-(t+1)x)}}{t+1}$. Therefore when $t_{\mathrm{K}_2}(P_{t,a}) = x$, we have

$$\begin{aligned}
t_{\mathrm{K}_3}(P_{t,a}) &= 3a(1-a)^2\frac{t-1}{t} + (1-a)^3\frac{t-1}{t}\frac{t-2}{t} \\
&= \left(\frac{t+\sqrt{t(t-(t+1)x)}}{t+1}\right)^2\left[3\frac{1-\sqrt{t(t-(t+1)x)}}{t+1}\frac{t-1}{t} + \frac{t+\sqrt{t(t-(t+1)x)}}{t+1}\frac{(t-1)(t-2)}{t^2}\right] \\
&= \frac{(t-1)(t+\sqrt{t(t-(t+1)x)})^2}{t^2(t+1)^2}\left[\frac{3t-3t\sqrt{t(t-(t+1)x)}+(t-2)(t+\sqrt{t(t-(t+1)x)})}{t+1}\right] \\
&= \frac{(t-1)(t+\sqrt{t(t-(t+1)x)})^2}{t^2(t+1)^2}\left[\frac{t+t^2-2t\sqrt{t(t-(t+1)x)}+\sqrt{t(t-(t+1)x)}}{t+1}\right] \\
&= \frac{(t-1)(t+\sqrt{t(t-(t+1)x)})^2(t-2\sqrt{t(t-(t+1)x)})}{t^2(t+1)^2}.
\end{aligned}$$

This formula—despite its complexity—gives us a bound. For any $x \in (0,1]$, let t be given such that $x \in (1-\frac{1}{t}, 1-\frac{1}{t+1}]$, and define

$$\mathfrak{g}_3(x) = \frac{(t-1)(t+\sqrt{t(t-(t+1)x)})^2(t-2\sqrt{t(t-(t+1)x)})}{t^2(t+1)^2}.$$

Then the example of the graph $P_{t,a}$ shows that it is possible to have $t_{\mathrm{K}_2}(G) = x$ and $t_{\mathrm{K}_3}(G) = \mathfrak{g}_3(x)$.

Lemma 4.40. *Let $G = (V, E, \mu_1)$ be a measurable graph with $t_{\mathrm{K}_t}(G) = c$, let $B \subseteq V$, and let $u = \int_B t_{\mathrm{K}_t}(G, x)\, d\mu_1$. Let ν_δ be the measure*

$$\nu_\delta(S) = \frac{1}{1+\delta}[\mu(S \setminus B) + (1 + \frac{\delta}{\mu(B)})\mu(S \cap B)]$$

and let $G_\delta = (V, E, \nu_\delta)$. Let h be the function with $h(\delta) = t_{\mathrm{K}_t}(G_\delta)$. Then $h'(0) = t(u - c)$.

Proof. Observe that

$$\begin{aligned}
t_{\mathrm{K}_t}(G_\delta) &= \int \prod_{1\le i<j\le t} \chi_E(x_i,x_j)\, d(\nu_\delta)_t \\
&= \frac{1}{(1+\delta)^2}\left[\int \prod_{1\le i<j\le t} \chi_E(x_i,x_j)\, d\mu_t + t\frac{\delta}{\mu(B)}\int \chi_B(x_1) \prod_{1\le i<j\le t} \chi_E(x_i,x_j)\, d\mu_t \right.\\
&\quad + \binom{t}{2}\frac{\delta^2}{\mu(B)}\int \chi_B(x_1)\chi_B(x_2) \prod_{1\le i<j\le t} \chi_E(x_i,x_j)\, d\mu_t + \cdots \\
&\quad \left. + \frac{\delta^t}{(\mu(B))^t}\int \prod_{i\le t}\chi_B(x_i) \prod_{1\le i<j\le t} \chi_E(x_i,x_j)\, d\mu_t\right].
\end{aligned}$$

In particular, since $h(0) = c = t_{\mathrm{K}_t}(G)$,

$$\begin{aligned}
h'(0) &= \lim_{\delta\to 0} \frac{t_{\mathrm{K}_t}(G_\delta) - t_{\mathrm{K}_t}(G)}{\delta} \\
&= \lim_{\delta\to 0} \frac{c + tu\delta + \delta^2(\cdots) - c(1+\delta)^t}{\delta(1+\delta)^t} \\
&= t(u-c).
\end{aligned}$$

□

Theorem 4.41. *For any $x \in (1/2, 2/3)$ and any infinite measurable graph G, if $t_{\mathrm{K}_2}(G) = x$ then $t_{\mathrm{K}_3}(G) \ge \mathfrak{g}_3(x)$.*

Proof. Suppose not, so there is an infinite measurable graph G with $t_{\mathrm{K}_2}(G) \in (1/2, 2/3)$ and $\mathfrak{g}_3(x)(t_{\mathrm{K}_2}(G) - t_{\mathrm{K}_3}(G) > 0$. By Theorem 4.31, we may choose a measurable graph G maximizing $\mathfrak{g}_3(t_{\mathrm{K}_2}(G)) - t_{\mathrm{K}_3}(G)$. Let $c = t_{\mathrm{K}_2}(G)$ and $d = t_{\mathrm{K}_3}(G)$.

This means that G has "too few" triangles (relative to its number of edges). We will argue that, because we have chosen G to maximize the defect in the number of triangles, it must be "locally maximal" as well, in two senses.

First, each vertex must contribute the right number of triangles relative to how many edges it contributes.

Claim 2. For a set of $x \in V$ of measure 1,

$$2\mathfrak{g}_3'(c)t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x) = 2\mathfrak{g}_3'(c)c - 3d.$$

Proof. Suppose not. Then there is some $\epsilon > 0$ so that either

$$B^+ = \{x \mid 2\mathfrak{g}_3'(c)t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x) - (2\mathfrak{g}_3'(c)c - 3d) \geq \epsilon\}$$

or

$$B^= \{x \mid 2\mathfrak{g}_3'(c)t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x) - (2\mathfrak{g}_3'(c)c - 3d) \leq -\epsilon\}$$

has positive measure. Let us assume the former, since the other case is symmetric.

We will apply Lemma 4.40. Let $u_2 = \int_{B^+} t_{\mathrm{K}_2}(G,x)\,d\mu_1$ and $u_3 = \int_{B^+} t_{\mathrm{K}_3}(G,x)\,d\mu_1$. By assumption,

$$2\mathfrak{g}_3'(c)u_2 - 3u_3 - (2\mathfrak{g}_3'(c)c - 3d) = \int_{B^+} 2\mathfrak{g}_3'(c)t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x) - (2\mathfrak{g}_3'(c)c - 3d)\,d\mu_1 \geq \epsilon\mu(B^+).$$

Letting ν_δ and G_δ be as in the statement of Lemma 4.40, we take $h_2(\delta) = t_{\mathrm{K}_2}(G_\delta)$ and $h_3(\delta) = t_{\mathrm{K}_3}(G_\delta)$. Consider the function $p(\delta) = \mathfrak{g}_3(t_{\mathrm{K}_2}(G_\delta)) - t_{\mathrm{K}_3}(G_\delta)$. Then

$$p'(0) = \mathfrak{g}_3'(c)h_2'(0) - h_3'(0) = 2\mathfrak{g}_3'(c)(u_2 - c) - 3(u_3 - d) = 2\mathfrak{g}_3'(c)u_2 - 3u_3 - (2\mathfrak{g}_3'(c)c - 3d) \geq \epsilon\mu(B^+) > 0.$$

In particular, this means that there is a small δ with $p(\delta) > p(0)$. But then G_δ would contradict the maximality of G. ⊣

Similarly, if an edge contributed too many triangles, we could delete it and thereby improve a more extreme choice of G. Unlike with vertices, we don't get a symmetric bound here—if an edge contributes too *few* triangles, we have no way to make more copies and thereby do better.

Claim 3. For a set of $(x,y) \in E$ of measure $\mu_2(E)$, $3t_{\mathrm{K}_3}(G,x,y) \leq \mathfrak{g}_3'(c)$.

Proof. Suppose not. Then there is an $\epsilon > 0$ so that, letting $B = \{(x,y) \in E \mid 3t_{\mathrm{K}_3}(G,x,y) > \mathfrak{g}_3'(c) + \epsilon\}$, $\mu_2(B) > 0$. For each $\delta \in (0, \mu_2(B))$, choose some $B_\delta \subseteq B$ and set $G_\delta = (V, E \setminus B_\delta, \mu_1)$. Then $t_{\mathrm{K}_2}(G_\delta) = c - \delta$ and

$$\begin{aligned} t_{\mathrm{K}_3}(G_\delta) &= \int \chi_{E\setminus B_\delta}(x,y)\chi_{E\setminus B_\delta}(x,z)\chi_{E\setminus B_\delta}(y,z)\,d\mu_3 \\ &= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 - 3\int \chi_{B_\delta}(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 \\ &\quad + 3\int \chi_{B_\delta}(x,y)\chi_{B_\delta}(x,z)\chi_E(y,z)\,d\mu_3 - \int \chi_{B_\delta}(x,y)\chi_{B_\delta}(x,z)\chi_{B_\delta}(y,z)\,d\mu_3. \end{aligned}$$

In particular,

$$\begin{aligned} \limsup_{\delta\to 0^+} \frac{t_{\mathrm{K}_3}(G_\delta) - d}{\delta} &= \frac{-3\int \chi_{B_\delta}(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + 3\int \chi_{B_\delta}(x,y)\chi_{B_\delta}(x,z)\chi_E(y,z)\,d\mu_3 - \int \chi_{B_\delta}(x,y)}{\delta} \\ &< -\mathfrak{g}_3'(c) - \epsilon. \end{aligned}$$

Letting $p(\delta) = \mathfrak{g}_3(t_{\mathrm{K}_2}(G_\delta)) - t_{\mathrm{K}_3}(G_\delta)$, we have

$$\limsup_{\delta \to 0^+} \frac{p(\delta) - p(0)}{\delta} > -\mathfrak{g}_3'(c) + (\mathfrak{g}_3'(c) + \epsilon) = \epsilon > 0.$$

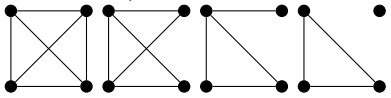
Then there is a small value of δ so that $\mathfrak{g}_3(t_{\mathrm{K}_3}(G_\delta)) - t_{\mathrm{K}_3}(G_\delta) > \mathfrak{g}_3(c) - d$, contradicting the maximality of G. ⊣

Next, we need an inequality:

Claim 4.

$$3\int \chi_E(x,y) t_{\mathrm{K}_3}(G,x) d\mu_2 + 3\int (1-\chi_E(x,z))(1-\chi_E(y,z)) t_{\mathrm{K}_3}(G,x,y)\, d\mu_3 \geq 2d.$$

Proof. The idea is to count over quadruples: any set of four vertices containing at least one triangle must be in one of four configurations (up to permutations):



Let $P_i \subseteq V^4$ for $i \in \{1,2,3,4\}$ be the set of quadruples (x,y,z,w) such that, up to some permutation, $\{x,y,z,w\}$ form a subgraph isomorphic to the i-th graph above. (For instance, $P_1 = T_{\mathrm{K}_4}(G)$.) Then $24d = 24\mu_4(P_1) + 12\mu_4(P_2) + 6\mu_4(P_3) + 6\mu_4(P_4)$.

Similarly,

$$24\int \chi_E(x,y) t_{\mathrm{K}_3}(G,x) d\mu_2 = 24\mu_4(P_1) + 16\mu_4(P_2) + 2\mu_4(P_3)$$

and

$$24\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_E(x,y) t_{\mathrm{K}_3}(G,x,y)\, d\mu_3 = 2\mu_4(P_3) + 6\mu_3(P_4),$$

so

$$\begin{aligned}
&3\int \chi_E(x,y) t_{\mathrm{K}_3}(G,x) d\mu_2 + 3\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_E(x,y) t_{\mathrm{K}_3}(G,x,y)\, d\mu_3 \\
&= 72\mu_4(P_1) + 48\mu_4(P_2) + 12\mu_4(P_3) + 18\mu_4(P_4) \\
&\geq 48\mu_4(P_1) + 24\mu_4(P_2) + 12\mu_4(P_3) + 12\mu_4(P_4) \\
&= d.
\end{aligned}$$

⊣

Observe that the first claim means that

$$\int \chi_E(x,y)(2\mathfrak{g}_3'(c)t_{\mathrm{K}_2}(G,x) - 3t_{\mathrm{K}_3}(G,x))\,d\mu_2 = c(2\mathfrak{g}_3'(c)c - 3d)$$

and the second claim means that

$$\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_E(x,y)t_{\mathrm{K}_3}(G,x,y)\,d\mu_3 \le \frac{\mathfrak{g}_3'(c)}{3}\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_E(x,y)\,d\mu_3.$$

Therefore

$$\begin{aligned}
2d &\le 3\int \chi_E(x,y)t_{\mathrm{K}_3}(G,x)d\mu_2 + 3\int (1-\chi_E(x,z))(1-\chi_E(y,z))t_{\mathrm{K}_3}(G,x,y)\,d\mu_3\\
&\le -c(2\mathfrak{g}_3'(c)c - 3d) + 2\mathfrak{g}_3'(c)\int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 + \mathfrak{g}_3'(c)\int (1-\chi_E(x,z))(1-\chi_E(y,z))\chi_E(x,y)\,d\mu_3\\
&= -c(2\mathfrak{g}_3'(c)c - 3d) + \mathfrak{g}_3'(c)\int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + \mathfrak{g}_3'(c)\int \chi_E(x,y)\chi_E(x,z)\,d\mu_3 + \mathfrak{g}_3'(c)\int (1-\\
&= -c(2\mathfrak{g}_3'(c)c - 3d) + \mathfrak{g}_3'(c)\int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z)\,d\mu_3 + \mathfrak{g}_3'(c)\int \chi_E(x,y)\,d\mu_2\\
&= -2\mathfrak{g}_3'(c)c^2 + \mathfrak{g}_3'(c)c + 3cd + \mathfrak{g}_3'(c)d.
\end{aligned}$$

Reorganizing gives us

$$\mathfrak{g}_3'(c)c(2c-1) \le d(\mathfrak{g}_3'(c) + 3c - 2).$$

When $x \in [1/2, 2/3]$, $\mathfrak{g}_3'(x) \ge 1$, so $\mathfrak{g}_3'(c) + 3c - 2 > 0$ and therefore

$$d \ge \frac{\mathfrak{g}_3'(c)c(2c-1)}{\mathfrak{g}_3'(c) + 3c - 2}.$$

Finally, we notice that $\mathfrak{g}_3$ satisfies the differential equation

$$x(2x-1)\mathfrak{g}_3'(x) = \mathfrak{g}_3(x)(\mathfrak{g}_3'(x) + 3x - 2),$$

so

$$d \ge \frac{\mathfrak{g}_3(c)(\mathfrak{g}_3'(c) + 3c - 2)}{\mathfrak{g}_3'(c) + 3c - 2} = \mathfrak{g}_3(c).$$

This gives us the lower bound we need: $d = t_{\mathrm{K}_2}(G) \ge \mathfrak{g}_3(t_{\mathrm{K}_3}(G))$. □

We have stated this for infinite measurable graphs but, as usual, we can quickly derive a consequence about large finite graphs.

Corollary 4.42. *For each $x \in (1/2, 2/3)$ and each $\epsilon > 0$, there is an n and a $\delta > 0$ so that whenever G has at least n vertices and $|t_{\mathrm{K}_2}(G) - x| < \delta$, $t_{\mathrm{K}_3}(G) \ge \mathfrak{g}_3(x) - \epsilon$.*

Proof. Suppose not. Let x and ϵ be a counterexample and, for each n, choose G_n with at least n vertices so that $|t_{K_2}(G_n) - x| < 1/n$ but $t_{K_3}(G_n) \geq \mathfrak{g}_3(x) - \epsilon$, and take any ultrafilter $\mathcal{U}$.

Then $t_{K_2}([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} t_{K_2}(G_n) = x$ and $t_{K_3}([G_n]_{\mathcal{U}}) = \lim_{n\to\mathcal{U}} t_{K_3}(G_n) \geq \mathfrak{g}_3(x) - \epsilon$. But this contradicts the theorem. □

4.13 Transfer

We keep proving results in the ultraproduct setting and then pulling them down to finite graphs. The proofs that the analogous results hold in finite graphs seem rather repetitive, so we'd like to give a general result containing all such arguments.

Theorem 4.43 (Transfer)**.** *Suppose that for every pair of natural numbers m and k, $\sigma_{m,k}$ is a first-order sentence. Then the following are equivalent:*

- *in $[G_n]_{\mathcal{U}}$, for every m there is a k so that $\sigma_{m,k}$ is true,*
- *for every m there is a k so that $\{n \mid \sigma_{m,k}$ is true in $G_n\} \in \mathcal{U}$.*

Proof. Suppose the first item fails: there is an m so that, for every k, $\sigma_{m,k}$ is false in $[G_n]_{\mathcal{U}}$. Then, by Łoś's Theorem, for each k, $\{n \mid \sigma_{m,k}$ is not true in $G_n\} \in \mathcal{U}$.

Similarly, if the second item fails then there is an m so that, for each k, $\{n \mid \sigma_{m,k}$ is not true in $G_n\} \in \mathcal{U}$, and therefore, by Łoś's Theorem again, $\sigma_{m,k}$ is false in $[G_n]_{\mathcal{U}}$. □

Corollary 4.44. *Suppose that for every pair of natural numbers m and k, $\sigma_{m,k}$ is a first-order sentence. Then the following are equivalent:*

- *in every infinite ultraproduct $[G_n]_{\mathcal{U}}$, for every m there is a k so that $\sigma_{m,k}$ is true,*
- *for every m there is a K and an n so that whenever G_n is a graph with $\geq n$ vertices, there is a $k \leq K$ so that $\sigma_{m,k}$ is true in G_n.*

Proof. Suppose the second item is true, let $[G_n]_{\mathcal{U}}$ be any infinite ultraproduct, and let m be given. There is a K and an N so that if G_n has $\geq N$ vertices, there is a $k \leq K$ so that $\sigma_{m,k}$ is true in G_n. $\{n \mid G_n$ has $\geq N\}$ vertices belongs to $\mathcal{U}$, so $\{n \mid$ for some $k \leq K$, $\sigma_{m,k}$ is true in $G_n\} \in \mathcal{U}$. But this is a finite union of the sets $\{n \mid \sigma_{m,k}$ is true in $G_n\}$, so there is a $k \leq K$ with $\{n \mid \sigma_{m,k}$ is true in $G_n\} \in \mathcal{U}$, and therefore $\sigma_{m,k}$ is true in $[G_n]_{\mathcal{U}}$.

Conversely, suppose that there is an m so that, for each K, there are arbitrarily large graphs G_n where each $\sigma_{m,k}$ with $k \leq K$ is false. Then, for each n, take G_n to be a graph with $\geq n$ vertices and such that, for each $k \leq n$, $\sigma_{m,k}$ is false. Let $[G_n]_{\mathcal{U}}$ be any ultraproduct. Since, for each k, $\{n \mid \sigma_{m,k} \text{ is true in } G_n\}$ is finite, the lemma says that each $\sigma_{n,k}$ is false in $[G_n]_{\mathcal{U}}$. □

Note that we are avoiding writing "$\forall m \exists k\, \sigma_{m,k}$" because it is important to remember that the quantifiers over m and k are *not* first-order quantifiers: they are quantifiers over the natural numbers.

When we want to relate statements about ultraproducts to statements in the finite setting, often all we need to do is observe that the statement we are interested in has the correct form. There is a slight complication, because most of the statements we are interested in involve statements about *measure*, which we noticed are not exactly first-order. We can address this with the following technical observation:

Lemma 4.45. *Let $f(x_1, \ldots, x_d)$ be a simple function—that is, f is a finite linear combination of functions involving internal sets. Then, for each $r < d$ and each ϵ, there is an internal set $M_{f,r,\epsilon}$ such that*

$$\{(a_1, \ldots, a_r) \mid \int f(a_1, \ldots, a_r, x_{r+1}, \ldots, x_d)\, d\mu_{d-r} < \epsilon\} \subseteq M_{f,r,\epsilon} \subseteq \{(a_1, \ldots, a_r) \mid \int f(a_1, \ldots, a_r, x_{r+1}, \ldots, x_d$$

Proof. We construct $M_{f,r,\epsilon}$ almost by definition: for each n, take $M_{f,e,\epsilon,n} \subseteq V_n^r$ to be those $(a_{1,n}, \ldots, a_{r,n})$ such that $\int f(a_1, \ldots, a_r, x_{r+1}, \ldots, x_d)\, d\mu_{d-r,n} < \epsilon$. □

This means that, while we cannot always use transfer with exact statements about measure, we can use it with the sort of inequalities we have been considering. For instance, our first result about finite graphs was to prove

> 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 $t_{\mathrm{K}_2}(G) = p$ and V sufficiently large, $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| < \epsilon$.

For a fixed value of H, if we replace ϵ with $E = \lceil 1/\epsilon \rceil$ and δ with $D = \lceil 1/\delta \rceil$, we can take $\sigma_{E,D}$ to be "either $|t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4| > 1/D$ then $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| < 1/E$". Understood in the ultraproduct, this corresponds to the weaker statement "either $|t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4| \geq 1/D$ or $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| \leq 1/E$". Since we proved that $|t_{\mathrm{C}_4}(G) - (t_{\mathrm{K}_2}(G))^4| = 0$ implies $|t_H(G) - t_{\mathrm{K}_2}(G)^{|F|}| = 0$, we showed that "for every E there is a D so that

$\sigma_{E,D}$" is true in any ultraproduct, and so we obtain the desired statement in sufficiently large finite models.

4.14 Remarks

The development of measures in ultraproducts (and nonstandard models more generally) goes back to Loeb [36], and measures of this kind are called *Loeb measures*. The notion of a graded probability space was introduced by Keisler [33] and used, in a context very close to the one we are concerned with, by Keisler's student Hoover [27].

The *Furstenberg correspondence*—a correspondence between finite sets and dynamical systems introduced by Furstenberg to give his proof of Szemerédi's Theorem [18]—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.)

First-order formulas have the nice property that they transfer easily the ground structures to the ultraproduct. Finding the right transference between statements about measure in the ground structure and the ultraproduct requires slightly more care. Many arguments in the literature handle this in an ad hoc way (as in [30]), but a general framework extending first-order logic by predicates for measures exists as well [20].

Bibliography

[1] 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. 45).

[2] 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. 5).

[3] 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).

[4] 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. 82).

[5] 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. 57).

[6] 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. 45).

[7] B. Chazelle. *The discrepancy method*. Randomness and complexity. Cambridge University Press, Cambridge, 2000, pp. xviii+463. ISBN: 0-521-77093-9 (cit. on p. 46).

[8] 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 p. 45).

[9] 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. 46).

[10] 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 p. 46).

[11] 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. 82).

[12] M. Di Nasso. "Hypernatural numbers as ultrafilters". In: *Nonstandard analysis for the working mathematician*. Springer, Dordrecht, 2015, pp. 443–474 (cit. on p. 82).

[13] 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. 82).

[14] P. Erdős and J. Spencer. "Imbalances in k-colorations". In: *Networks* 1 (1971/72), pp. 379–385. ISSN: 0028-3045 (cit. on p. 46).

[15] 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. 57).

[16] 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. 45).

[17] 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. 5).

[18] 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. 5, 122).

[19] 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 p. 82).

[20] 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. 5, 122).

[21] 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 p. 5).

[22] 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 p. 5).

[23] 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. 44).

[24] 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. 5).

[25] H. Hatami. "Graph norms and Sidorenko's conjecture". In: *Israel J. Math.* 175 (2010), pp. 125–150. ISSN: 0021-2172 (cit. on p. 46).

[26] 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. 82).

[27] D. Hoover. *Relations on Probability Spaces and Arrays of Random Variables*. Preprint. Institute for Advanced Study, Princeton, NJ, 1979 (cit. on p. 122).

[28] 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. 5).

[29] 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. 57).

[30] E. Hrushovski. “Stable group theory and approximate subgroups”. In: *J. Amer. Math. Soc.* 25.1 (2012), pp. 189–243. ISSN: 0894-0347 (cit. on p. 122).

[31] 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. 45).

[32] 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. 82).

[33] 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. 122).

[34] 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. 45).

[35] 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. 82).

[36] 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. 122).

[37] 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. 45).

[38] 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 p. 5).

[39] V. Nikiforov. “The number of cliques in graphs of given order and size”. In: *Trans. Amer. Math. Soc.* 363.3 (2011), pp. 1599–1618. ISSN: 0002-9947 (cit. on p. 114).

[40] 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. 45).

[41] A. A. Razborov. “Flag algebras”. In: *J. Symbolic Logic* 72.4 (2007), pp. 1239–1282. ISSN: 0022-4812 (cit. on p. 112).

[42] A. A. Razborov. "On the minimal density of triangles in graphs". In: *Combin. Probab. Comput.* 17.4 (2008), pp. 603–618. ISSN: 0963-5483 (cit. on p. 114).

[43] C. Reiher and M. Schacht. "Forcing quasirandomness with triangles". In: *ArXiv e-prints* (Nov. 2017). eprint: 1711.04754 (math.CO) (cit. on p. 46).

[44] C. Reiher. "The clique density theorem". In: *Ann. of Math. (2)* 184.3 (2016), pp. 683–707. ISSN: 0003-486X (cit. on p. 114).

[45] 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 p. 5).

[46] K. F. Roth. "On certain sets of integers". In: *J. London Math. Soc.* 28 (1953), pp. 104–109. ISSN: 0024-6107 (cit. on p. 5).

[47] 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. 45).

[48] 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. 45).

[49] A. F. Sidorenko. "Inequalities for functionals generated by bipartite graphs". In: *Diskret. Mat.* 3.3 (1991), pp. 50–65. ISSN: 0234-0860 (cit. on p. 46).

[50] 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. 45).

[51] 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. 45).

[52] J. Skokan and L. Thoma. "Bipartite subgraphs and quasi-randomness". In: *Graphs Combin.* 20.2 (2004), pp. 255–262. ISSN: 0911-0119 (cit. on pp. 45, 46).

[53] 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. 5).

[54] 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. 5).

[55] 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 p. 5).

[56] 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. 45).

[57] 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. 45).

[58] 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. 45).

[59] 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. 5).