Randomness at Infinity

Henry Towsner

April 7, 2019

NOTE: This is an INCOMPLETE DRAFT. Many citations and attributions are missing from this version.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6 Stability</td>
<td>182</td>
</tr>
<tr>
<td>6.7 Stable Regularity</td>
<td>187</td>
</tr>
<tr>
<td>6.8 Internal Cardinality and Pseudofinite Dimension</td>
<td>189</td>
</tr>
<tr>
<td>6.9 Stable Erdős-Hajnal</td>
<td>193</td>
</tr>
<tr>
<td>6.10 Compression Schemes</td>
<td>195</td>
</tr>
<tr>
<td>6.11 Remarks</td>
<td>197</td>
</tr>
</tbody>
</table>
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 $\epsilon$ 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 Szemerédi’s Theorem ([9], [37], [38], [42], [45], [46], [49], [54], [77], [87], [98], [103], [104], [112], 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 [88].

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 $\epsilon n$ where $\epsilon > 0$ was a fixed number and $n$ is then very large relative to $\epsilon$.

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: $|E| \geq \epsilon'$ (where $\epsilon' > 0$ is a real number depending only on $\epsilon$, though it may be a bit smaller than $\epsilon$), and similarly for $F$ and $G$. 
It will be natural to think of these as coming from 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 $\delta n^3$ triangles then once $n > 1/\delta$, 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, na"ively, that the sets $E$, $F$, and $G$ are independent in a way that allows us to integrate over the three sets separately:

$$\iiint \chi_E(x,y)\chi_F(x,z)\chi_G(y,z) d\mu_3(x,y,z) \approx \iiint \chi_E(x,y) d\mu_2 \iiint \chi_F(x,z) d\mu_2 \iiint \chi_G(y,z) d\mu_2.$$ 

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 $\chi_E, \chi_F, \chi_G$ 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” function 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 in two main ways. Every proof of Szemerédi’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 ignore this perspective entirely, preferring to focus on counting and cardinality (for example, speaking of $|A| \geq cn$ 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.

The true goal of this book is to use Szemerédi’s Theorem as a guide to presenting, 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 round-about 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” (edges between a vertex and itself): by definition, an edge is a pair of distinct vertices. Similarly, this definition excludes directed graphs—$E$ is a set of unordered pairs, so there is no difference between saying $\{v, w\} \in E$ and $\{w, v\} \in E$. We also prohibit a graph with no vertices at all (though we allow graphs with no edges).

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

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

Since we will deal with this graph repeatedly, we give it a name: we will call this random graph $R_{1/2}$, or $R_{1/2}(V)$ if we wish to be explicit about the set of vertices $V$. We will abuse notation to write $R_{1/2}$ for both the
CHAPTER 2. RANDOM AND QUASI-RANDOM GRAPHS

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 $\binom{|V|^2}{2}$ 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,

$$\Pr\left(\left|\frac{|R_{1/2}|}{\binom{|V|}{2}} - \frac{1}{2}\right| < \epsilon \left(\frac{|V|}{2}\right) \right) \geq 1 - \delta.$$ 

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

$$\Pr\left(\left|\frac{|R_{1/2}|}{\binom{|V|}{2}} - \frac{1}{2}\right| < \epsilon \right) \geq 1 - \delta.$$ 

**Proof.** The idea is that each edge is placed in $R_{1/2}$ independently, so $|R_{1/2}|$ is a sum of independent random variables: for each $\{v, w\}$, let $1_{\{v, w\}}$ be the random variable which is 1 if $\{v, w\} \in R_{1/2}$ and 0 if $\{v, w\} \not\in R_{1/2}$. Then $|R_{1/2}| = \sum_{\{v, w\} \in \binom{V}{2}} 1_{\{v, w\}}$. Since each of these random variables $1_{\{v, w\}}$ is chosen independently, it is very likely that close to half of them are 1 while the other half are 0.

Slightly more formally, what $1_{\{v, w\}}$ means is that we first pick a particular pair of vertices, $\{v, w\}$, and then generate the set of edges $R_{1/2}$ by flipping coins, and $1_{\{v, w\}}$ is 1 in the event that the potential edge we picked in advance actually turns up in our final graph. In particular, $1_{\{v, w\}}$ is simply the result of a single coin flip, so its expected value, $E(1_{\{v, w\}})$, is 1/2, since half the time the coin comes up heads and $1_{\{v, w\}} = 1$, and the other half of the time the coin comes up tails and $1_{\{v, w\}} = 0$.

The expected value of $\frac{|R_{1/2}|}{\binom{|V|}{2}}$ is also 1/2: by the the linearity of expected
value
\[ E\left(\frac{|R_{1/2}|}{|V|^2}\right) = E\left(\sum_{\{v,w\}\in(V)^2} \frac{1_{\{v,w\}}}{|V|^2}\right) = \frac{1}{|V|} \sum_{\{v,w\}\in(V)^2} E(1_{\{v,w\}}) = \frac{1}{|V|} \sum_{\{v,w\}\in(V)^2} \frac{1}{2} = \frac{1}{2}. \]

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

This should happen because \( \frac{|R_{1/2}|}{|V|^2} \) is the sum of a large number of independent random variables. The Hoeffding inequality covers precisely this situation.

**Theorem (Hoeffding Inequality).** If \( X = \frac{1}{k} \sum_{i \leq k} X_i \) where the \( X_i \) are independent random variables such that \( 0 \leq X_i \leq 1 \) always holds then
\[ \mathbb{P}(|X - E(X)| \geq \epsilon) \leq 2e^{-2k\epsilon^2}. \]

We take \( X \) to be the random variable \( \frac{|R_{1/2}|}{|V|^2} \), so the Hoeffding inequality says that
\[ \mathbb{P}\left(|\frac{|R_{1/2}|}{|V|^2} - \frac{1}{2}| \geq \epsilon\right) \leq 2e^{-2(|V|^2)\epsilon^2}. \]

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

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, \( K_{n,n} \), is the graph \( (V \cup W, E) \) where \( |V| = |W| = n \), \( V \) and \( W \) are disjoint sets, and \( E \) consists of all pairs of vertices between \( V \) and \( W \). Then \( |E| = n^2 \) while \( \frac{1}{2} \binom{|V \cup W|}{2} = \frac{1}{2} \frac{2n(2n-1)}{2} = n^2 - \frac{n}{2} \).
Figure 2.1: $K_{5,5}$

In this example, $|E|$ is not quite identical to half the possible edges, but the error—$n/2$—is small relative to $\binom{|V \cup W|}{2}$. 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.)

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 $G$. 
2.2. SUBGRAPH DENSITY

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 \( (\{0, 1, 2\}, (\{0, 1, 2\})_2) \)—that is, the triangle with three vertices and all three edges: \( \bullet \triangle \bullet \).

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 \( \pi \) 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 \( \square \). If \( H \) is the cycle on four vertices, \( \bullet \longrightarrow \bullet \), then by our definition, \( t_H(G) \) is positive: we count the copies where \( \pi \) 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 possible copy \( \pi : W \to V \) such that, for each pair \( \{w, w'\} \in \binom{W}{2} \),

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

*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 define \( t^\text{ind}_H(G) \) to be the fraction of potential copies of \( H \) which are induced copies:

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

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

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

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 \( t_H(R_{1/2}) \) is.

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

When \( \pi \) 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 we divide by \( |V||W| \), these terms will contribute at most \( C \cdot \frac{1}{|V|} \) to \( t_H(G) \)—that is, an error term on the order \( O(\frac{1}{|V|}) \).

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

\[
\mathbb{E}(t_H(R_{1/2})) = \frac{1}{|V||W|} \sum_\pi \mathbb{E}(1_\pi) = 2^{-|F|} + O\left(\frac{1}{|V|}\right).
\]

In particular,

\[
\lim_{|V| \to \infty} \mathbb{E}(t_H(R_{1/2})) = 2^{-|F|}.
\]

(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 $R_{1/2}(V)$. We only expect $R_{1/2}(V)$ 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,  

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

**Proof.** The idea is similar to the proof of Theorem 2.2 above: we want to argue that $t_H(R_{1/2})$ 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 $1_\pi$ are no longer independent: if $\pi(H)$ and $\pi'(H)$ share an edge, $1_\pi$ and $1_{\pi'}$ 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 $t_H(R_{1/2})$ is no longer a sum of many independent random variables, $t_H(R_{1/2})$ 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 $X_1, \ldots, X_k$ 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 $x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k$ and any $x_i, x'_i$,

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

Then

$$
\mathbb{P} \left( |\mathbb{E}(f(X_1, \ldots, X_k)) - f(X_1, \ldots, X_k)| \geq \epsilon \right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{k} c_i^2}}.
$$

The condition

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

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

$$|f(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_k) - f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_k)| < c_i$$

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 $|V|^{\lvert W \rvert - 2}$ ways to assign the remaining $|W| - 2$ vertices, so there are $\leq |W|^2 |V|^{\lvert W \rvert - 2}$ 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 $|W|^2 |V|^{\lvert W \rvert - 2}$ copies of $H$. This means that the density $t_H$ changes by at most $rac{|W|^2 |V|^{\lvert W \rvert - 2}}{|V|^2} = \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 $\sum_{i=1}^{k} c_i^2 = \binom{|V|}{2} \frac{|W|^4}{|V|^2} \leq \frac{|W|^4}{|V|^2}$.

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

$$\mathbb{P}\left( \left| 2^{-\lvert F \rvert} - t_H(\mathbf{R}_{1/2}) \right| \geq \epsilon \right) \leq \mathbb{P}\left( \left| \mathbb{E}(t_H(\mathbf{R}_{1/2})) - t_H(\mathbf{R}_{1/2}) \right| \geq \epsilon/2 \right) \leq 2e^{-\epsilon^2/|W|^2|V|^2}.$$ 

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

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 $\pi$ is injective then the probability that $\pi$ becomes an induced copy is $2^{-\binom{|W|}{2}}$, irrespective of how many edges $H$ has: in order for $\pi$ to be an induced copy, each pair $\{\pi(w), \pi(w')\}$ must do exactly the right thing—become an edge if $\{w, w'\} \in E$ 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:
2.3. \( t_H(G) \) AS AN INTEGRAL

**Theorem 2.8.** For every \( \epsilon > 0 \), every \( \delta > 0 \), and every finite graph \( H \), whenever \( V \) is sufficiently large,

\[
P\left( \left| t_H^{\text{ind}}(R_{1/2}) - 2^{-\binom{|V|}{2}} \right| < \epsilon \right) \geq (1 - \delta).
\]

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 \( \frac{1}{2} \) 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 \( \mu_k \) 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 \( \mu_3 \).

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

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

**Definition 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:

\[
\chi_E(v, w) = \begin{cases} 
1 & \text{if } (v, w) \in E \\
0 & \text{otherwise}
\end{cases}
\]
We will sometimes abuse notation and write \( t_H(E) \) or \( t_H^{\text{ind}}(E) \) if the vertexset is clear.

This lets us abbreviate, for instance,
\[
t_{C_3}(G) = \int \int \int \chi_E(x, y) \chi_E(x, z) \chi_E(y, z) \, d\mu_3.
\]

**Definition 2.11.** \( K_2 \) is the graph with two elements and an edge between them.

That is, \( K_2 \) is the graph consisting of a single edge. Then
\[
t_{K_2}(G) = \mu_2(\{(v, w) \mid \{v, w\} \in E\}) = \frac{2|E|}{|V|^2},
\]
which differs from \( \frac{|E|}{\binom{|V|}{2}} \) by an amount on the order of \( 1/|V| \), which we can treat as negligible when \( V \) is large enough. Going forward, we will often focus on the quantity \( t_{K_2}(G) \) rather than \( \frac{|E|}{\binom{|V|}{2}} \), and call this the edge density of \( G \).

More generally, we have

**Theorem 2.12.** For any graph \( H = (W, F) \) with \( W = \{w_1, \ldots, w_k\} \) and any graph \( G = (V, E) \),
\[
t_H(G) = \int \prod_{1 \leq i < j \leq k, \{w_i, w_j\} \in F} \chi_E(v_i, v_j) \, d\mu_k.
\]

Note that, since these are finite spaces, integrals are really averages:
\[
\int \prod_{1 \leq i < j \leq k, \{w_i, w_j\} \in F} f(v_i, v_j) \, d\mu_{|W|} = \frac{1}{|V||W|} \sum_{\{v_1, \ldots, v_{|W|}\} \in V^{|W|}} \prod_{1 \leq i < j \leq k, \{w_i, w_j\} \in F} f(v_i, v_j).
\]

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
\[
t_H(f) = \int \prod_{1 \leq i < j \leq k, \{w_i, w_j\} \in F} f(v_i, v_j) \, d\mu_{|W|}.
\]
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,
\[ t_{C_3}(f) = \int f(x,y)f(x,z)f(y,z) \, d\mu_3. \]
But, since we do not attach any significance to the order of the vertices, this
should be equal to
\[ \int f(x,y)f(z,x)f(y,z) \, d\mu_3. \]

One advantage of this notation is that it gives us a quick way to calculate
things like \( E(t_H(R_{1/2})) \): 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
\[ \left| E(t_H(R_{1/2})) - t_H(f) \right| < \frac{C}{|V|}. \]

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

When \( \pi \) is injective,
\[ E(1_\pi) = 1 \cdot P(\pi \text{ is a copy of } H) + 0 \cdot P(\pi \text{ is not a copy of } H) \]
\[ = 2^{-|F|} \]
\[ = \prod_{1 \leq i < j \leq k, \{w_i, w_j\} \in F} f(v_i, v_j). \]
CHAPTER 2. RANDOM AND QUASI-RANDOM GRAPHS

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

$$\left| E(t_H(R_{1/2})) - t_H(f) \right| < C/|V|.$$ 


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 $(V \cup W, (V)_2 \cup (W)_2)$.

$\overline{K_{n,n}}$ 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 $V_0 \cup V_3$ and one vertex in $V_1 \cup V_2$, and
- $E_1 = (V_0 \cup V_1)_2 \cup (V_2 \cup V_3)_2$—all pairs with both vertices in $V_0 \cup V_1$, or all pairs with both vertices in $V_2 \cup V_3$. 
2.4. COUNTING TRIANGLES

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

Next we define a family of partially random graphs interpolating between these. For the purposes of this section, we will call these graphs $G_p$ for $p \in [0, 1]$. $G_p$ 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 $G_p$ with a grid

<table>
<thead>
<tr>
<th></th>
<th>$V_0$</th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_0$</td>
<td>1-p</td>
<td>0</td>
<td>1-p</td>
<td>0</td>
</tr>
<tr>
<td>$V_1$</td>
<td>p</td>
<td>1</td>
<td>0</td>
<td>1-p</td>
</tr>
<tr>
<td>$V_2$</td>
<td>1-p</td>
<td>0</td>
<td>p</td>
<td>1</td>
</tr>
<tr>
<td>$V_3$</td>
<td>0</td>
<td>1-p</td>
<td>1</td>
<td>p</td>
</tr>
</tbody>
</table>

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 $G_p$ using a function $f_p$ which reflects the grid above.

**Theorem 2.16.** Let $G_p$ be the randomly generated graph defined above. Let
$f_p : V^2 \to [0,1]$ be the function given by

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

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

The function $f_p$ looks like this:

```
\begin{align*}
&\begin{tikzpicture}[scale=0.8]
&\node (a) at (0,0) [shape=circle,draw] {$V_0$};
&\node (b) at (1,0) [shape=circle,draw] {$V_1$};
&\node (c) at (0,1) [shape=circle,draw] {$V_2$};
&\node (d) at (1,1) [shape=circle,draw] {$V_3$};
&\draw [dashed] (a) -- (b);
&\draw [dashed] (c) -- (d);
&\draw (a) -- (c);
&\draw (b) -- (d);
&\end{tikzpicture}
\end{align*}
```

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:

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

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

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

$$
\mathbb{E}(t_H(G_p)) = \frac{1}{|V||W|} \sum_{\pi : W \to V, \pi \text{ injective}} \prod_{(w, w') \in F} f_p(\pi(w), \pi(w')) + O\left(\frac{1}{|V|}\right)
$$

$$
= t_H(f_p) + O\left(\frac{1}{|V|}\right).
$$

**Theorem 2.17.** For every $\epsilon > 0$, every $\delta > 0$, and every finite graph $H = (W, F)$, whenever $n$ is sufficiently large,

$$
\mathbb{P}\left(|t_H(G_p) - \mathbb{E}(t_H(G_p))| < \epsilon\right) \geq (1 - \delta).
$$
Proof. This is the same argument using McDiarmid’s inequality as for the random graph: the quantity \( t_H(G_p) \) is a function of the random variables \( 1_{\{v,w\}} \) with the property that changing any single edge can only change \( t_H(G_p) \) by at most \( \frac{|W|^2|V|(|V|-2)}{|V|^2} = |W|^2 \).

So, by McDiarmid’s inequality, the probability that \( t_H(G_p) \) differs from \( \mathbb{E}(t_H(G_p)) \) by more than \( \epsilon \) is at most
\[
2e^{-\frac{\epsilon^2|V|^2}{|W|^2}}.
\]
In particular, when \( n = |V|/4 \) is sufficiently large, the probability that \( |t_H(G_p) - \mathbb{E}(t_H(G_p))| \geq \epsilon \) is \( < \delta \).

\[\square\]

\textbf{Theorem 2.18.} For any \( \epsilon > 0 \), when \( n \) is sufficiently large, with probability \( \geq 1 - \epsilon \):

- \( |t_{K_2}(G_p) - 1/2| < \epsilon \), and
- \( |t_{C_3}(G_p) - \frac{1}{8}(p^3 - 2p^2 + 3p)| < \epsilon \).

\textbf{Proof.} Combining the previous two theorems, it suffices to show that \( t_{K_2}(f_p) = 1/2 \) and \( t_{C_3}(f_p) = \frac{1}{8}(p^3 - 2p^2 + 3p) \).

For the first claim, \( t_{K_2}(f_p) = \int f_p(v,w)d\mu_2 \). The four components \( V_0, V_1 \), and so on are symmetric, so
\[
\int f_p(v,w)d\mu_2 = 4\int_{V_0 \times V} f_p(v,w)d\mu_2
= 4(\int_{V_0 \times V_0} f_p(v,w)d\mu_2 + \int_{V_0 \times V_1} f_p(v,w)d\mu_2 + \int_{V_0 \times V_2} f_p(v,w)d\mu_2)
= 4(p\mu(V_0 \times V_0) + \mu(V_0 \times V_1) + (1-p)\mu(V_0 \times V_2) + 0\mu(V_0 \times V_3))
= 4(\frac{p}{4} + \frac{1}{4} + \frac{1}{4} - \frac{p}{4})
= 1/2.
\]

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

\[ t_{C_3}(f_p) = \int f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \]

\[ = 4 \int_{V_0 \times V \times V} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \]

\[ = 4 \left( \int_{V_0 \times V_0 \times V_0} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 + 2 \int_{V_0 \times V_0 \times V_1} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \right. \]

\[ + 2 \int_{V_0 \times V_0 \times V_2} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 + \int_{V_0 \times V_1 \times V_1} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \]

\[ + 2 \int_{V_0 \times V_1 \times V_2} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 + \int_{V_0 \times V_1 \times V_3} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \]

\[ + 2 \int_{V_0 \times V_2 \times V_3} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 + \int_{V_1 \times V_2 \times V_3} f_p(u, v)f_p(u, w)f_p(v, w)d\mu_3 \right) \]

\[ = 4 \frac{1}{3^3} (p^3 + 2p + 2p(1-p)^2 + 0 + 0 + 0 + p(1-p)^2 + 0 + 0) \]

\[ = \frac{1}{8}(p^3 - 2p^2 + 3p). \]

\[ \square \]

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

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

So the graph \( G_{p^*} \) 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 \( |Y| \geq C\ln |V| \),

\[ \left| \frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2} \right| < \epsilon. \]
2.4. COUNTING TRIANGLES

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

**Proof.** The idea is that if we pick a single pair of sets $X$ and $Y$ in advance, we expect close to half the edges in $X \times Y$ will belong to $R_{1/2}$. If we then look at many pairs of sets, each pair of sets has a small probability of “being defective”—of having either too many or too few edges between them. Exactly how small that probability is depends on the sizes of $X$ and $Y$—it is some value $p(|X|, |Y|)$.

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

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

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

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

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 $R_{1/2}$. 
Observe that
\[
\mathbb{E}\left( \frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} \right) = \frac{1}{|X| \cdot |Y|} \sum_{(x,y) \in X \times Y} 1_{\{x,y\}} = 1/2.
\]

The quantity \(\frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|}\) is a function of at least \(\frac{|X| \cdot |(Y|-1)}{2} \geq \frac{|X| \cdot |Y|}{4}\) 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,
\[
P\left( \left| \frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2} \right| \geq \epsilon \right) \leq 2e^{-2\epsilon^2|X| \cdot |Y|}.
\]

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 \textit{union bound}:
\[
P(X_1 \text{ or } X_2 \text{ or } \cdots \text{ or } X_k) \leq \sum_{i=1}^{k} P(X_i).
\]

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 \(\binom{|V|}{x} \binom{|V|}{y}\) possible choices for \(X\) and \(\binom{|V|}{y}\) possible choices for \(Y\). Using \textit{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 \(\left| \frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2} \right| \geq \epsilon\) 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 \(\left| \frac{|R_{1/2} \cap (X \times Y)|}{|X| \cdot |Y|} - \frac{1}{2} \right| \geq \epsilon\) is bounded by
\[
n^2 \cdot 2e^{x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V|} \leq 2e^{x(\ln |V| + 1) + y(\ln |V| + 1) - 2\epsilon^2 C^2 \ln^2 |V|}.
\]
When \( x, y \geq C \ln |V| \) for \( C \) sufficiently large,

\[
2 \ln n + x(\ln |V| + 1) + y(\ln |V| + 1) - 2e^2C^2 \ln^2 |V| \leq -\ln^2 |V|.
\]

In particular, when \( V \) is large enough, the probability is less than \( \delta \).

So the graph \( G_{p^*} \) still does not resemble \( R_{1/2} \):

\[
\frac{|G_{p^*} \cap (V_0 \times V_3)|}{|V_0||V_3|} = 0,
\]
which would be very unlikely in \( R_{1/2} \).

### 2.5 Counting Cycles of Length 4

Another way to try to distinguish \( G_{p^*} \) from \( R_{1/2} \) 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: \[
\square
\]

We could also call this graph \( K_{2,2} \), 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 \( \geq 1 - \delta \), \( |t_{C_4}(G_p) - \frac{1}{8}(p^4 - 2p^3 + 3p^2 - 2p + 1)| < \epsilon \) (where \( G_p \) is the graph from the previous section).

**Sketch.** Again, it suffices to show that \( t_{C_4}(f_p) = \frac{1}{8}(p^4 - 2p^3 + 3p^2 - 2p + 1) \).

The calculation is still a little tedious: since

\[
t_{C_4}(f_p) = \int f_p(x, y) f_p(y, x') f_p(x', y') f_p(y', x) d\mu_4,
\]

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 \( u \in V_0 \):

\[
t_{C_4}(f_p) = 4 \int_{V_0 \times V_3} f_p(x, y) f_p(y, x') f_p(x', y') f_p(y', x) d\mu_4,
\]

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

\[\square\]
As it happens, that means that $t_{C_4}(G_p) \approx 0.07123$, which is not the
1/16 that we would expect in a random graph. Indeed, there is no value of $p$
which gives $\mathbb{E}(t_{C_4}(G_p)) = 1/16$: 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, $G_p$ 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 $G$,

$$t_{C_4}(G) \geq (t_{K_2}(G))^4.$$ 

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

**Theorem** (Cauchy-Schwarz).

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

$$(t_{K_2}(G))^4 = \left(\int \chi_E(x, y) d\mu_2\right)^4$$

$$= \left(\int \left(\int \chi_E(x, y) d\mu(y)\right) d\mu(x)\right)^4$$

$$\leq \left(\int \left(\int \chi_E(x, y) d\mu(y)\right)^2 d\mu(x)\right)^2$$

$$= \left(\int \chi_E(x, y) \chi_E(x', y') d\mu_3\right)^2$$

$$= \left(\int \left(\int \chi_E(x, y) \chi_E(x, y') d\mu(x)\right) d\mu_2(y, y')\right)^2$$

$$\leq \int \left(\int \chi_E(x, y) \chi_E(x, y') d\mu(x)\right)^2 d\mu_2(y, y')$$

$$= \int \chi_E(x, y) \chi_E(x', y) \chi_E(x', y') d\mu_4$$

$$= t_{C_4}(G).$$
In graphs with very few edges (for instance, a “perfect matching” with $2n$ vertices and $n$ edges where each vertex has exactly one neighbor), this is a little counterintuitive—it depends on the fact that our definition of $t_{C_4}(G)$ includes “degenerate cycles” in which $x = x'$ and $y = y'$. However this only matters in very sparse graphs (the perfect matching has edge density $t_{K_2}(G) = \frac{1}{4n}$). In the case we are mostly interested in, where $t_{K_2}(G)$ is some fixed real number $p \in (0, 1)$ and $n$ is quite large, there are at least $p^4n^4$ copies of $C_4$, of which at most around $n^3$ are degenerate, so the degenerate copies contribute only a small error term to $t_{C_4}(G)$.

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

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

**Definition 2.23.** A graph $G = (V,E)$ is $\epsilon$-quasirandom if

$$|t_{C_4}(G) - (t_{K_2}(G))^4| < \epsilon.$$ 

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

As the name suggests, what we plan to show is that when $t_{C_4}(G) \approx (t_{K_2}(G))^4$, $G$ must resemble a random graph.

Although we have only discussed $R_{1/2}$ so far, it is not a big jump to generalize to the graph $R_p$, 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 $\epsilon$-quasirandom, $t_{K_2}(G) = p$, and has a large number of vertices, $G$ is supposed to resemble $R_p$, a claim we will justify below.

Note that being quasirandom is a very different sort of property than being random. $R_{1/2}$ 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 $R_{1/2}$?”; 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 $R_{1/2}$.
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, $R_p$ is $\epsilon$-quasirandom with probability $\geq 1 - \delta$.

**Proof.** We can choose $\epsilon'$ 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

$$|t_{K_2}(R_p) - \mathbb{E}(t_{K_2}(R_p))| < \epsilon'$$

and

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

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

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

$$\mathbb{E}(t_{K_2}(R_p)) = t_{K_2}(f_p) = p$$

and

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

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

$$|t_{C_4}(R_p) - (t_{K_2}(R_p))^4| \leq |t_{C_4}(R_p) - p^4| + |(t_{K_2}(R_p))^4 - p^4| < \epsilon/2 + \epsilon/2 = \epsilon.$$ 

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 $\deg_G(x) = \frac{|N_G(x)|}{|V|}$. 
2.6. QUASIRANDOM GRAPHS

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 $\delta$ so that if $G$ is $\delta$-quasirandom and $t_{K_4}(G) = p \in [0, 1]$ 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 $t_{K_4}(G) = p$ 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 $\chi_E$, 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

$$\int \chi_E(x, y) \chi_E(x, z) d\mu_3 = \int (p + f(x, y))(p + f(x, z)) d\mu_3$$

$$= p^2 + 2p \int f(x, y) d\mu_2 + \int f(x, y) f(x, z) d\mu_3.$$

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.
CHAPTER 2. RANDOM AND QUASI-RANDOM GRAPHS

We can give a bound
\[
\int f(x, y)f(x, z)d\mu_3 = \int \left( \int f(x, y)d\mu(y) \right)^2 d\mu(x) \\
= \int \left( \int \chi_{E}(x, y) - pd\mu(y) \right)^2 d\mu(x) \\
= \int (\deg_G(x) - p)^2 d\mu \\
\geq \int_S (\deg_G(x) - p)^2 d\mu \\
\geq \mu(S)\epsilon^2 \\
\geq \epsilon^3.
\]

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:
\[
t_{C_4}(G) = \int \left( \int \chi_{E}(x, y)\chi_{E}(x, z)d\mu(x) \right)^2 d\mu_2(y, z) \\
\geq \left( \int \chi_{E}(x, y)\chi_{E}(x, z)d\mu_3 \right)^2 \\
\geq (p^2 + \epsilon^3)^2 \\
\geq p^4 + \epsilon^6.
\]

So we see that \( G \) cannot be \( \epsilon^6 \)-quasirandom. \( \square \)

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 \( C \), 
\[
\frac{|R_{1/2}(X \times Y)|}{|X||Y|}
\]
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 \( Y \)**, \( d_E(X, Y) = \frac{|E \cap (X \times Y)|}{|X||Y|} \).

**Theorem 2.28.** For every \( \epsilon > 0 \) there is a \( \delta > 0 \) so that whenever \( G = (V, E) \) is \( \delta \)-quasirandom with \( t_{K_2}(G) = p \), \( X \subseteq V, Y \subseteq V \), and \( \frac{|X|}{|V|} \geq \epsilon \) and \( \frac{|Y|}{|V|} \geq \epsilon \),
\[
|d_E(X, Y) - p| < \epsilon.
\]
2.6. QUASIRANDOM GRAPHS

Proof. Suppose $G$ is $\delta$-quasirandom for a sufficiently small $\delta$, 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 $\chi_E$ from the function which is constantly equal to $p$. Let $f(x,y) = \chi_E(x,y) - p$.

Then

$$t_{C_4}(G) = \int \chi_E(x,y)\chi_E(y,x')\chi_E(x',y')\chi_E(y',x)d\mu_4$$

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

$$= p^4 + 4p \int f(x,y)f(y,x')f(x',y')d\mu_4 + 4p^2 \int f(x,y)f(y,x')d\mu_3$$

$$+ 2p^2 \int f(x,y)f(x',y')d\mu_4 + 4p^3 \int f(x,y)d\mu_2 + t_{C_4}(f).$$

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 $\delta$-quasirandom for some small enough $\delta$, 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:

\[ t_{C_4}(f) = \int F(x, x', y, y')d\mu_4 \]
\[ = \int F(x, x', y, y')\chi_X(x)\chi_X(x')d\mu_4 \]
\[ + 2\int F(x, x', y, y')\chi_X(x)\chi_{V\setminus X}(x')d\mu_4 \]
\[ + \int F(x, x', y, y')\chi_{V\setminus X}(x)\chi_{V\setminus X}(x')d\mu_4. \]

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:

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

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 \( \delta \)-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 \(\epsilon\)-**regular** if whenever \(X \subseteq V, Y \subseteq V\) with \(|X| \geq \epsilon \frac{|V|}{|V|} \) and \(|Y| \geq \epsilon \frac{|V|}{|V|} \),

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

where \(p = t_{K_2}(G)\).

So Theorem 2.28 can be rephrased

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

\(\delta\)-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 \(\delta\)-quasirandom with \(t_{K_2}(G) = p\) and \(V\) is sufficiently large, \(|t_{C_3}(G) - p^3| < \epsilon\).

**Proof.** Suppose \(G\) is \(\delta\)-quasirandom for some \(\delta\) small enough (based on the calculations to follow), and we assume that \(\epsilon\) is much smaller than \(p\) (otherwise we can replace it with a smaller \(\epsilon\) 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, \(x \not\in S, |\deg_G(x) - p| < \epsilon/6p\). We can divide \(t_{C_3}(G)\) into those triangles whose first vertex is in \(S\) and those whose first vertex isn’t: So

\[t_{C_3}(G) = \int \chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3\]

\[= \int \chi_{V\setminus S}(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3 + \int \chi_S(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3\]

\[= \int \chi_{V\setminus S}(x)\mu(N_G(x))^2d_E(N_G(x), N_G(x))d\mu_3 + \int \chi_S(x)\chi_E(x, y)\chi_E(x, z)\chi_E(y, z)d\mu_3.\]
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

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

$$\leq 3\epsilon/6p$$

$$\leq \epsilon/2.$$ 

So

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

\[\square\]

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

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

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

While we used quasirandomness in the proof, we could actually have used regularity instead. In particular, $\delta$-regularity for sufficiently small $\delta$ implies $t_{C_3}(G) \approx p^4$, which means

**Theorem 2.33.** For every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $G = (V, E)$ is $\delta$-regular and $V$ is sufficiently large, $G$ is $\epsilon$-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.

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 $\|f\|_{L^2} = \sqrt{\langle f, f \rangle} = \int (f(x))^2d\mu$.

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 $T_{\chi_E}$.

$T_h f$ 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) $\lambda_1, \ldots, \lambda_n$ and vectors $f_1, \ldots, f_n$ such that:

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

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

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

For example, recall our bipartite graph $K_{n,n}$, consisting of two pieces, say $X$ and $Y$, with $n$ vertices each, and exactly the edges between those two parts. Then $\lambda_1$ 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

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

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 $\chi_E$ as a weighted sum of the eigenvectors is not a coincidence.

Theorem 2.37. When $\lambda_1, \ldots, \lambda_n$ and $f_1, \ldots, f_n$ are the eigenvalues and eigenvectors associated to $T_h$, $h(x,y) = \sum_{i \leq n} \lambda_i f_i(x) f_i(y)$.

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

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

\[
\int h(x, y)g(y)d\mu(y) = \int h(x, y) \sum_{i \leq n} c_i f_i(x)d\mu(y) = \sum_{i \leq n} c_i \int h(x, y)f_i(x)d\mu(y) = \sum_{i \leq n} c_i \lambda_i f_i(x)
\]

while, similarly,

\[
\int \left( \sum_{i \leq n} \lambda_i f_i(x)f_i(y) \right) g(y)d\mu(y) = \int \left( \sum_{i \leq n} \lambda_i f_i(x)f_i(y) \right) \left( \sum_{i \leq n} c_i f_i(y) \right)d\mu(y) = \sum_{i \leq n} \sum_{j \leq n} \lambda_i f_i(x) \int f_i(y) f_j(y)d\mu(y) = \sum_{i \leq n} \sum_{j \leq n} c_j \lambda_i f_i(x) \int f_i(y) f_j(y)d\mu(y).
\]

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 \( y_0 \),

\[
\sum_{i \leq n} \lambda_i f_i(x)f_i(y) = h(x, y).
\]
CHAPTER 2. RANDOM AND QUASI-RANDOM GRAPHS

This gives a convenient proof of a standard fact:

**Theorem 2.38.** If \( h : V^2 \rightarrow \mathbb{R} \) 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
\[
\int h(x, x) d\mu = \int \sum_{i \leq n} \lambda_i f_i(x) f_i(x) d\mu = \sum_{i \leq n} \lambda_i \int (f_i(x))^2 d\mu = \sum_{i \leq n} \lambda_i
\]
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 \( K_{n,n} \) 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 \( \lambda_1 \) is the largest eigenvalue (in absolute value) of \( T_h \),
\[
|\lambda_1| \geq t_{K_2}(h).
\]

**Proof.** Consider the function \( g(x) \) which is constantly equal to 1. Then
\[
\langle T_h g, g \rangle = \int h(x, y) g(x) g(y) d\mu_2 = \int h(x, y) d\mu_2 = t_{K_2}(h).
\]

We can stop here if we invoke a standard fact about eigenvalues, that the largest eigenvalue \( \lambda_1 \) satisfies \( |\lambda_1| = \max_u \langle T_h u, u \rangle \) where \( u \) ranges over the unit vectors: since \( g \) is a unit vector, \( |\lambda_1| \geq \langle T_h g, g \rangle = t_{K_2}(h) \).

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 \( c_i, g(x) = \sum_{i \leq n} c_if_i(x) \), and we have

\[
1 = \int (g(x))^2 d\mu \\
= \int \left( \sum_{i \leq n} c_if_i(x) \right) \left( \sum_{j \leq n} c_jf_j(x) \right) d\mu \\
= \sum \sum_{i \leq n, j \leq n} c_ic_j \int f_i(x)f_j(x) d\mu \\
= \sum_{i \leq n} c_i^2.
\]

Then

\[
t_{K_2}(h) = \int h(x,y)g(x)g(y)d\mu_2 \\
= \int \left( \sum_{i \leq n} \lambda_if_i(x)f_i(y) \right) \left( \sum_{j \leq n} c_jf_j(x) \right) \left( \sum_{k \leq n} c_kf_k(y) \right) d\mu_2 \\
= \sum_{i \leq n} \lambda_i \left( \int \sum_{j \leq n} c_jf_i(x)f_j(x) d\mu \right) \left( \int \sum_{k \leq n} c_kf_i(x)f_k(y) d\mu \right) \\
= \sum_{i \leq n} \lambda_i c_i^2 \\
\leq |\lambda_1| \sum_{i \leq n} c_i^2 \\
= |\lambda_1|.
\]

The connection to quasirandomness comes from the observation that

\[
t_{C_4}(h) = \int h(x,y)h(y,x')h(x',y')h(y',x)d\mu_4 = \int h^4(x,x) d\mu
\]

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 \( t_{C_4}(h) \) 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 $f: V \rightarrow \mathbb{R}$,

$$((T_h)^2f)(x) = \int h(x, y) [(T_h f)(y)] d\mu(y)$$

$$= \int h(x, y) \left[ \int h(y, z) f(z) d\mu(z) \right] d\mu(y)$$

$$= \int \left( \int h(x, y) h(y, z) d\mu(y) \right) f(z) d\mu(z)$$

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

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

$$((T_h f_i)(x) = \int h^2(x, y) f_i(y) d\mu(y)$$

$$= \int \left( \int h(x, z) h(z, y) d\mu(z) \right) f_i(y) d\mu(y)$$

$$= \int h(x, z) h(z, y) f_i(y) d\mu_2(y, z)$$

$$= \int h(x, z) (T_h f_i)(z) d\mu(z)$$

$$= \lambda_i \int h(x, z) f_i(z) d\mu(z)$$

$$= \lambda_i (T_h f_i)(x)$$

$$= \lambda_i^2 f_i(x).$$

This allows us to prove that a graph is quasirandom exactly when all the eigenvalues other than the largest eigenvalue are small. (Where $\lambda_1$ is always the largest eigenvalue, and we always mean largest in the sense of absolute value.)

**Theorem 2.41.** $G = (V, E)$ is $\delta$-quasirandom if and only if

$$(\lambda_1 - \kappa_2(G))^4 + \sum_{1 < i \leq n} \lambda_i^4 < \delta.$$
2.8. DENSE GRAPHS

Proof. Observe that

\[ t_{C_4}(G) - t_{K_2}(G)^4 = \sum_{i \leq n} \lambda_i^4 - t_{K_2}(G)^4 = (\lambda_1^4 - t_{K_2}(G)^4) + \sum_{1 < i \leq n} \lambda_i^4. \]

We are usually not concerned with the exact quantity \((\lambda_1 - t_{K_2}(G))^4 + \sum_{1 < i \leq n} \lambda_i^4\)—the point is that \(G\) is \(\delta\)-quasirandom exactly when two things happen: first, \(\lambda_1\) is close to \(t_{K_2}(G)\), 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 \(t_{K_2}(G) = \epsilon > 0\), 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 \(t_{K_2}(G) \leq \frac{1}{2}\).

The bipartite graph illustrates that this bound is optimal: the bipartite graph \(K_{n,n}\) has \(t_{K_2}(K_{n,n}) = 1/2\).

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

\[ t_{K_2}(G) = \int \chi_E(x, y) \cdot 1 \, dy \, dx \]

\[ \geq \int \chi_E(x, y) (\int \chi_E(x, z) \, dz + \int \chi_E(y, w) \, dw) \, dy \, dx \]

\[ = 2 \int (\int \chi_E(x, y) \, dy) \, dx \]

\[ \geq 2 \left( \int \chi_E(x, y) \, dx \right)^2 \]

\[ = 2t_{K_2}(G)^2, \]

so \(\frac{1}{2} \geq t_{K_2}(G)\). \qed
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 $n/t$, $|E| \approx \frac{1 - 1}{t}$. 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 $\pi \in T_H(G)\mu$ 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 $t_{K_2}(G) \leq 1 - \frac{1}{\chi(H)-1}$. 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 *Payley graph $Q_q$* is the graph with vertices $\{0, 1, \ldots, q-1\}$ where $\{x, y\}$ is an edge precisely when $x - y$ is a quadratic residue modulo $q$ (that is, $x - y \neq 0$ and there is a $k$ so that $k^2 \equiv x - y \mod q$).

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

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

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

which approaches $1/2$ when $q$ is large.
Theorem 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 $\delta$-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 quadratic residue exactly when $z \in C(x, y)$.

Assume $x$ and $y$ are distinct. There are $\frac{q}{2} (q-1)$ 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

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

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 $|X| \geq \log q \log \log q$ which is a clique in $Q_q$[48]: 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 $R_{1/2}$, 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 generalizing 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 $\epsilon n^2$ 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 [13] is a canonical reference, especially when supplemented by more recent books [4, 36, 58].

The study of quasirandom graphs was introduced by Chung, Graham, and Wilson in a paper [19], and most of this chapter is drawn from that paper. Since then, a number of additional characterizations of quasirandomness have been investigated [76, 79, 90, 91, 95, 96, 97, 111]. 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 [108, 109] and also well-studied since [66]. 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 [23, 84, 97]; 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 [93], which concerns a generalization of the Cauchy-Schwarz calculations we were using above to [50].

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 discrepancy of a graph. Erdős and Spencer [29] showed that there is a constant so that every graph with $|V|$ has discrepancy at least $cn^{3/2}$, and that random graphs have close to this minimum discrepancy. The behavior of the discrepancy and related quantities in random, quasirandom, and other graphs has also been further studied; Chazelle’s book [14] 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 [20].
Chapter 3

Ultraproducts

3.1 Convergent Subsequences

Keeping track of the $\epsilon$’s and $\delta$’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 $K_{n/2,n/2}$, but when $n$ is odd, $G_n$ is the complete graph $\langle \{1, \ldots, n\}, \langle 1, \ldots, n \rangle \rangle$. Then the sequence $t_{C_3}(G_n)$ 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—

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

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.

\( \square \)

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
3.1. CONVERGENT SUBSEQUENCES

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 \( \mathbb{N} \setminus K \). 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 \( \langle 0, 2, 1, 3, 0, 2, 1, 3, \ldots \rangle \), 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 $K$. \qed

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 Ultraproducts

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

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

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

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 \( \lim ? r_n = \lim \frac{r_n}{n \to K} \) 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 \( K \in \mathcal{F} \),
- \( \emptyset \not\in \mathcal{F} \), 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 $(K_1 \cap \cdots \cap K_n) \setminus K_{n+1} \in \mathcal{F}$.

**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 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”.*
3.2. ULTRALIMITS

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

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

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

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

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

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

\[
\{n \mid |r_n - r| \leq 1/2\} \subseteq \{n \mid r_n \in [1/2, 1]\} \in \mathcal{U}.
\]
Next we split the interval in half again. For instance, if \( s_1 = 1/4 \) then we split \([0, 1/2] = [0, 1/4] \cup (1/4, 1/2] \). If \( \{ n \mid r_n \in [0, 1/4] \} \in \mathcal{U} \) then we take \( s_2 = 1/8 \) and promise that, for all \( n > 2, s_n \in [0, 1/4] \), which will mean that \( r \in [0, 1/4] \), and therefore that
\[
\{ n \mid |r_n - r| \leq 1/4 \} \setminus \{ n \mid r_n \in [0, 1/4] \} \in \mathcal{U}.
\]
Otherwise \( \{ n \mid r_n \in (1/4, 1] \} \in \mathcal{U} \), so also
\[
\{ n \mid r_n \in (1/4, 1] \} \cap \{ n \mid r_n \in [0, 1/2] \} = \{ n \mid r_n \in (1/4, 1/2] \} \in \mathcal{U}.
\]
In this case we take \( s_2 = 3/8 \) and promise that, for all \( n > 2, s_n \in (1/4, 1/2] \). This means that \( r \in [1/4, 1/2] \), and therefore that
\[
\{ n \mid |r_n - r| \leq 1/4 \} \subseteq \{ n \mid r_n \in (1/4, 1/2] \} \in \mathcal{U}.
\]
In general, after \( k \) stages we have chosen \( s_1, \ldots, s_k \) and an interval \( I = [\frac{a}{2^k}, \frac{a+1}{2^k}] \) so that:
- when \( i < j \), \( |s_i - s_j| \leq \frac{1}{2^k} \),
- \( s_k \in I \),
- \( \{ n \mid r_n \in I \} \in \mathcal{U} \).

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 \( |s_i - s_j| \leq \frac{1}{2^j} \). 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
\[
\{ n \mid |r_n - r| < \epsilon \} \subseteq \{ n \mid |r_n - r| \leq \frac{1}{2^{k+1}} \} \subseteq \{ n \mid |r_n - s_k| \leq \frac{1}{2^{k+1}} \} \in \mathcal{U}.
\] 

\( \square \)

**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 \( F_0 \) 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 \( J \cap K \in F_0 \).

- If \( J \) is cofinite and \( J \subseteq K \) then \( \mathbb{N} \setminus K \subseteq \mathbb{N} \setminus J \) is also finite, so \( K \in F_0 \).

We will try to expand \( F_0 \) into an ultrafilter. If we have any free filter \( F \) which is not an ultrafilter, so there is some set with both \( K \notin F \) and \( \mathbb{N} \setminus K \notin F \), then we can at least extend the free filter to contain one of these sets.

**Lemma 3.11.** Let \( F \) be a free filter such that \( \mathbb{N} \setminus K \notin F \). Then there is a free filter \( F' \supseteq F \cup \{K\} \).

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

Suppose \( I_0, I_1 \in F' \), so \( I_0 \supseteq K \cap J_0 \) an \( I_1 \supseteq K \cap J_1 \) with \( J_0, J_1 \in F \). 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 \( J_0 \cap J_1 \in F \), also \( I_0 \cap I_1 \in F' \).

If \( I_0 \in F' \) and \( I_0 \subseteq I_1 \) then \( I_1 \supseteq I_0 \supseteq K \cap J_0 \) so \( I_1 \in F' \).

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

Since \( F' \supseteq F \), it is the desired free filter. \( \square \)

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 \( F_0 \), 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 \( F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \) 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 \( i \in L \), \( F_i \) is a free filter so that when \( i \leq j \), \( F_i \subseteq F_j \). Then there is a free filter \( F \) such that, for all \( i \in L \), \( F_i \subseteq F \).

**Proof.** We simply take \( F = \bigcup_{i \in L} F_i \).

Suppose \( J, K \in F \). Then there are \( i, j \) so that \( J \in F_i \) and \( K \in F_j \). Either \( i \leq j \) or \( j \leq i \); without loss of generality, assume \( j \leq i \), so \( F_j \subseteq F_i \), so \( K \in F_i \). Then \( J \cap K \in F_i \), so also \( J \cap K \in F_j \).

If \( J \in F \) and \( J \subseteq K \) then \( J \in F_i \), so \( K \in F_i \), so \( K \in F \).

Since \( \emptyset \notin F_i \) for all \( i \), \( \emptyset \notin F \).

If \( K \) is cofinite then \( K \in F_i \) for any \( i \), so \( K \in F \).

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 \( F_0 \subseteq F_1 \subseteq \cdots \) so that, at stage \( i \), we ensure that either \( K_i \in F_{i+1} \) or \( \mathbb{N} \setminus K_i \in F_{i+1} \). Doing this requires some use of Axiom of Choice (the version usually called the “well-ordering principle”) to obtain an ordering of the subsets of \( \mathbb{N} \) on which transfinite recursion works.

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

**Theorem 3.13 (Zorn’s Lemma).** Let \( P \) be a set partially ordered by \( \subseteq \) so that whenever \((L, \leq)\) is a totally ordered set and \( f : I \to P \) is an order preserving function (so \( i \leq j \) implies \( f(i) \subseteq f(j) \)), there is a
3.4. PRODUCTS

\[ P \in \mathcal{P} \text{ such that, for every } i \in I, f(i) \subseteq P. \text{ Then there is a maximal element of } \mathcal{P} \text{—a } P \in \mathcal{P} \text{ such that, for any } Q \in \mathcal{P} \text{ such that } P \leq Q, \]

\[ P = Q. \]

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 [12, 30].

### 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

*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 [55].
be the bipartite graph $K_{n,n}$. 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 $(v_n)_{n \in \mathbb{N}}$ 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 $(v_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$, we need to decide whether to put an edge between them. The sequence $(v_n)_{n \in \mathbb{N}}$ could be a mix of even and odd values, as could $(w_n)_{n \in \mathbb{N}}$—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 $(v_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$ 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 think 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:

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

Note that it is very important that our choice of “most” $n$ is “coherent”. Suppose we take three elements of $\prod_{n \in \mathbb{N}} V_n$, $(u_n)_{n \in \mathbb{N}}$, $(v_n)_{n \in \mathbb{N}}$, and $(w_n)_{n \in \mathbb{N}}$. 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 $\langle (u_n)_{n \in \mathbb{N}}, (v_n)_{n \in \mathbb{N}} \rangle \in [E_n]_{\mathcal{U}}$ if $\{n \mid \{u_n, v_n\} \in E_n\} \in \mathcal{U}$, and, separately, we place $\langle (v_n)_{n \in \mathbb{N}}, (w_n)_{n \in \mathbb{N}} \rangle \in [E_n]_{\mathcal{U}}$ if $\{n \mid \{v_n, w_n\} \in E_n\} \in \mathcal{U}$,
and similarly for \( \{u_n\}_{n \in \mathbb{N}}, \{w_n\}_{n \in \mathbb{N}} \in [E_n]_U \). Yet these three decisions cannot be independent, because we cannot place all three pairs in \([E_n]_U\).

Coordinating all the different choices is the job of the ultrafilter \(U\). (This is the simultaneous part of asking that all sequences converge simultaneously.) Suppose all three pairs did somehow end up in \([E_n]_U\); that would mean

\[
\{n \mid \{u_n, v_n\} \in E_n\}, \{n \mid \{u_n, w_n\} \in E_n\}, \{n \mid \{v_n, w_n\} \in E_n\} \in U.
\]

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

\[
\{n \mid \{u_n, v_n\} \in E_n\} \cap \{n \mid \{u_n, w_n\} \in E_n\} \cap \{n \mid \{v_n, w_n\} \in E_n\} \in U.
\]

That is impossible, because there are no such \(n\) and \(\emptyset \notin U\), 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 \(v, w \in V\), \(\{v, w\} \in [E_n]_U\) 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 \(\{n \mid v_n\text{ is even}\} \in U\), 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 \(\{n \mid v_n\text{ is odd}\} \in U\). 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

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

so \(\langle\langle v_n \rangle_{n \in \mathbb{N}}, \langle w_n \rangle_{n \in \mathbb{N}\rangle} \in [E_n]_U\).

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

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

so \(\langle\langle v_n \rangle_{n \in \mathbb{N}}, \langle w_n \rangle_{n \in \mathbb{N}\rangle} \notin [E_n]_U\). \(\square\)

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 $[E_n]_U$ consists of all pairs $\{ \langle v_n \rangle_{n \in \mathbb{N}}, \langle w_n \rangle_{n \in \mathbb{N}} \}$ such that $\{ n \mid \{ v_n, w_n \} \in E_n \} \in U$.

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

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 $\langle v_n \rangle_{n \in \mathbb{N}} \sim_U \langle w_n \rangle_{n \in \mathbb{N}}$ if $\{ n \mid v_n = w_n \} \in U$.

**Theorem 3.17.** $\sim_U$ is an equivalence relation.

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

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

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

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

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

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

Formally, a vertex of $[V_n]_U$ is an equivalence class of sequences—that is, it is a set of sequences which are all $\sim_U$ equivalent to each other.

**Definition 3.18.** When $\langle v_n \rangle_{n \in \mathbb{N}}$, we will write $[v_n]_U$ for the equivalence class of $\langle v_n \rangle_{n \in \mathbb{N}}$ in $[V_n]_U$.

We call $\langle v_n \rangle_{n \in \mathbb{N}}$ a representative of the equivalence class $[v_n]_U$. 
3.5. ULTRAPRODUCTS

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 \([v_n] = v\). 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 other words, that we could have equivalent sequences which disagree about whether or not the pair belongs to \([E_n]_{\mathcal{U}}\).

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

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

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

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

This fixes the example that gave us trouble. If \( G_n \) is the complete graph \( \langle \{ 1, 2, \ldots, n \}, \langle \{ 1, 2, \ldots, n \} \rangle \rangle \) 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.
CHAPTER 3. ULTRAPRODUCTS

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.

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.

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 \(n\), choose \(w_n \in V_n \setminus \{v^n_1, \ldots, v^n_{|V_n|-1}\}\). Since we are only excluding at most \(|V_n| - 1\) values from \(V_n\), we know some choice of \(w_n\) is possible. (Note that, when we work index-wise, considering each index \(n\) individually, we

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.
also only consider the \( n \)-th terms of the representatives—we are considering \( v_n^i \) because this is an element of \( V_n \).

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

Therefore
\[
\{ n \mid w_n \neq v_n^i \} \supseteq \{ n \mid |V_n| > i \}.
\]

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

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

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

We said at the beginning of the chapter that we want the subgraph density of the limit to be the limit of the subgraph densities; in our new terminology, we can say that we want
\[
t_H(\prod_{n \to \mathcal{U}} G_n) = \lim_{n \to \mathcal{U}} t_H(G_n).
\]

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

**Theorem 3.21.** Let \( H \) be a finite graph. There is a copy of \( H \) in \([G_n]_\mathcal{U}\) if and only if
\[
\{ n \mid \text{there is a copy of } H \text{ in } G_n \} \in \mathcal{U}.
\]

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

Then, for each \( \{w, w'\} \in F \), we must have a set
\[
K_{w, w'} = \{ n \mid \{v^w_n, v^{w'}_n\} \in E_n \} \in \mathcal{U}
\]
in the ultraproduct which witnesses the presence of the edge between \( \pi(w) = [v^w_n]_\mathcal{U} \) and \( \pi(w') = [v^{w'}_n]_\mathcal{U} \).

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 \( \pi_n(w) = v^w_n \). Then for any \( \{w, w'\} \in F \), \( \{\pi_n(w), \pi_n(w')\} = \{v^w_n, v^{w'}_n\} \in E_n \) because \( n \in K_{w, w'} \).
Theorem 3.22. The ultraproduct $[G_n]_U$ has an isolated triangle if and only if \{n \mid G_n has an isolated triangle\} $\in U$.

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

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 $U$, so also $K = K_1 \cap K_2 \cap K_3$ belongs to $U$.

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 $U$ then we could take a vertex $x = [x_n]_U$ (where $x_n$ is any element of $V_n$ for $n \notin J_1$). But then we would have $\{x, u\} \in [E_n]_U, x \neq v,$ and $x \neq w,$ contradicting the fact that we started with an isolated triangle.
3.5. ULTRAPRODUCTS

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 $[G_n]_\mathcal{U}$. \hfill $\square$

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 $i \lt n$, $\{v_i, v_{i+1}\} \in E$.

**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 $\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$ 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 \( K_i = \{ n \mid \{ u_n^i, u_n^{i+1} \} \in E_n \} \), 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 \( u_n^1 = v_n = 1 \) adjacent to \( u_n^2 \), so \( u_n^2 = 2 \). Since \( u_n^2 \) is adjacent to \( u_n^3 \), we must have \( u_n^3 = 3 \) or \( u_n^3 = 1 \). Continuing in this way, we must have \( u_n^i \leq i \) for each \( i \), which would mean that \( k \geq u_n^k = w_n = n \), contradicting the choice that \( k < n \).

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

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, 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

\[
\{ n \mid v_n \in X_n \} = \{ n \mid X_n \neq \emptyset \} \supseteq \{ n \mid 1 \in X_n \}
\]
3.6. A FEW EXAMPLES

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 \([v_n]_\mathcal{U} \in [X_n]_\mathcal{U} \setminus X \). \( \square \)

3.6 A Few Examples

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

First, consider the ultraproduct \( K_{\infty,\infty} = [K_n]_\mathcal{U} \).

**Lemma 3.26.** \( K_{\infty,\infty} = (V,E) \) is a complete bipartite graph with two infinite parts.

**Proof.** For each \( n \), we have \( K_{n,n} = V_0^n \cup V_1^n \) with \( |V_0^n| = |V_1^n| \) and \( E_n = (V_0^n \times V_0^n) \cup (V_1^n \times V_1^n) \).

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

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

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

**Lemma 3.27.** \( K_{\infty,\infty^2} = (V,E) \) is a complete bipartite graph with two infinite parts.

It will be slightly harder to make sense of the idea that, in \( K_{\infty,\infty} \), the two parts have the same size, while in \( K_{\infty,\infty^2} \) one part is much larger than the other. In particular, it turns out that, in both graphs, the two parts have the same cardinality—there is a bijection between them. However we will show later that only in \( K_{\infty,\infty} \) can we find a bijection between the two parts which “respects the ultraproduct” in a certain sense.

If, for each \( n \), we take \( G_n = \mathbb{R}_p(\{1, \ldots, n\}) \), a random graph on \( n \) vertices, we get an ultraproduct we will call \( \mathbb{R}_{p,\mathcal{U}} = [G_n]_\mathcal{U} \). Instead of talking about what happens “with high probability”, the interesting properties of \( \mathbb{R}_{p,\mathcal{U}} \) will turn out to be most settled up to probability 1.
3.7 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 3.28.** We say a set \( X \subseteq [V_n]_U \) is *internal* if there is some sequence \( X_n \subseteq V_n \) such that \( X = [X_n]_U \) (as sets—that is, \( v \in X \) if and only if \( v \in [X_n]_U \)).

If a subset of \([V_n]_U\) is not internal, it is *external*.

More generally, we will speak of internal subsets of the product set \([V_n]^k]_U\) for \( k > 1 \) in the same way—\( X \subseteq [V_n]^k]_U \) is internal if there are sets \( X_n \subseteq V_n^k \) so that \( X \) is precisely the set of tuples \((v_1^n, \ldots, v_k^n)\) such that \( \{ n | (v_1^n, \ldots, v_k^n) \in X_n \} \in U \).

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

An internal set \( X \) has many representations in the form \([X_n]_U\), but these representations are as similar as we could hope for—they are equal almost everywhere.

**Lemma 3.29.** If \([X_n]_U = [Y_n]_U \) (as sets, any element of one set is also an element of the other) then \( \{ n | X_n = Y_n \} \in U \).

**Proof.** We prove the contrapositive. Suppose \( \{ n | X_n = Y_n \} \notin U \). Then
\[
\{ n | X_n \neq Y_n \} = \{ n | X_n \setminus Y_n \neq \emptyset \} \cup \{ n | Y_n \setminus X_n \neq \emptyset \} \in U.
\]
So one of these sets must in \( U \); without loss of generality, assume \( K = \{ n | X_n \setminus Y_n \neq \emptyset \} \in U \). 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 \( \{ n | a_n \in X_n \} \supseteq K \in U \) and \( \{ n | a_n \notin Y_n \} \supseteq K \in U \), so \([a_n]_U \in [X_n]_U \setminus [Y_n]_U \).

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 3.30. Suppose that \( \{ n \mid G_n \text{ is connected}\} \in \mathcal{U} \). 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 \( \{ n \mid G_n \text{ is connected}\} \), 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 \( \{[v_n]_\mathcal{U}, [w_n]_\mathcal{U}\} \in [E_n]_\mathcal{U} \).}

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 be able to work with abstract descriptions of sets, independent of the particular graph we are considering. In particular, we would like to have the property 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. We can say that “$x_1 = x_2$” is a formula which defines this set. When we emphasize that $x_1 = x_2$ describes a set, we are emphasizing that “$x_1 = x_2$” is an abstract mathematical object—a string of symbols—and that we can look at any graph $(V, E)$ and pick out a subset of $V^2$, in such a way that there is some relationship between the set we get in $(V, E)$ and the set we get in some other graph $(V', E')$.

We customarily use letters like $\phi$ and $\psi$ for formulas, and indicate the variables in parentheses after the formula. If $\phi(x_1, x_2)$ is our formula $x_1 = x_2$ and $a^1, a^2 \in V$ 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

$$\begin{align*}
\text{if } & X_n = \{ (a^1, a^2) \in V^2_n \mid \phi(a^1, a^2) \} \text{ then } [X_n]_\mathcal{U} = \{ (a^1, a^2) \in V^2_\mathcal{U} \mid \\
& \phi(a^1, a^2) \}, \text{ and} \\
\text{if } & X_n = \{ b \in V_n \mid \phi(b, a_n) \} \text{ then } [X_n]_\mathcal{U} = \{ b \in [V_n]_\mathcal{U} \mid \phi(b, [a_n]_\mathcal{U}) \}.
\end{align*}$$

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:

**Theorem 3.31.** If $\phi(x_1, \ldots, x_m)$ is a first-order formula, $k \leq m$, and $[b^1_\mathcal{U}, \ldots, b^{m-k}_\mathcal{U}] \in [V_n]_\mathcal{U}$, and

$$X_n = \{ (a^1, \ldots, a^k) \in V^k_n \mid \phi(a^1, \ldots, a^k, b^1_n, \ldots, b^{m-k}_n) \}$$

*In fact, one can define first-order logic in roughly this way, with the formulas precisely the functors from structures to sets which respect ultraproducts.

*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.
then

\[ [X_n]_\mathcal{U} = \{(a^1, \ldots, a^k) \in [V_n^k]_\mathcal{U} \mid \phi(a^1, \ldots, a^k, b^1, \ldots, b^{m-k})\}. \]

That is, the set a formula “carves out” of each \( V_n \) should correspond to the set it “carves out” of \([V_n]_\mathcal{U}\)—and this should remain true if we allow the formula to reference parameters \( b^1, \ldots, b^{m-k} \).

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, whenever 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 \( \phi \), which is true exactly when \( \phi \) 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,

\[ \{(a^1, \ldots, a^k) \in V^k \mid \neg \phi(a^1, \ldots, a^k, b^1, \ldots, b^{m-k})\} = V^k \setminus \{(a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, b^1, \ldots, b^{m-k})\} \]

the set defined by \( \neg \phi \) is the complement of the set defined by \( \phi \).

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,

\[ (\phi \land \psi)(x_1, \ldots, x_k) \]

(read “\( \phi \) and \( \psi \)”, is the conjunction of \( \phi \) with \( \psi \), and is true when both \( \phi \) and \( \psi \) are true).

\(^1\)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 them, while \( W = \{(1, 2), \emptyset\} \) is the graph with two vertices and no edges. Then \( \phi(1, 2) \) is false in \( V \) and true in \( W \).
• \((\phi \lor \psi)(x_1, \ldots, x_k)\), read “\(\phi\) or \(\psi\)”, is the disjunction of \(\phi\) with \(\psi\), and is true when at least one of \(\phi\) and \(\psi\) is true (including when both are true).

For example, if \(\phi(x_1, x_2, x_3)\) is the formula \(\{x_1, x_2\} \in E \land \{x_1, x_3\} \in E \land \{x_2, x_3\} \in E\) 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
\[
\{x_1, x_2\} \in E \land \{x_2, x_3\} \in E \land \{x_3, x_4\} \in E \land \{x_4, x_1\} \in E
\land \neg\{x_1, x_2\} \in E \land \neg\{x_3, x_4\} \in E
\]
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
\[
\phi(x_1, \ldots, x_k) = \{x_i, x_j\} \in E \land \cdots \land \{x_i', x_j'\} \in E
\]
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
\[
\psi(x_1, \ldots, x_k) = \{x_i, x_j\} \in E \land \cdots \land \{x_i', x_j'\} \in E \land \neg\{x_i'', x_j''\} \in E \land \cdots \land \neg\{x_i''', x_j'''\} \in E
\]
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 \(\phi\)”, is true when there is some vertex \(y\) so that \(\phi\) is true,

• \((\forall y \phi)(x_1, \ldots, x_k)\), read “for every \(y\), \(\phi\)”, is true when \(\phi\) is true for every vertex \(y\).

For example, consider the formula \(\phi(x_1, x_2, x_3)\) given by
\[
\{x_1, x_2\} \in E \land \{x_1, x_3\} \in E \land \{x_2, x_3\} \in E
\land \forall y [(x_1 = y) \lor (x_2 = y) \lor (x_3 = y)
\land \neg\{x_1, y\} \in E \land \neg\{x_2, y\} \in E \land \neg\{x_3, y\} \in E].
\]
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
3.7. INTERNAL SETS AND FIRST-ORDER LOGIC

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 3.32 (Łoś’s Theorem).** If \( \phi(x_1, \ldots, x_k) \) is a first-order formula and \([a_i^i]_U \in [V_n]_U \) for each \( i \leq k \) then

\[
\{ n \mid \phi(a_1^n, \ldots, a_k^n) \text{ is true} \} \in U
\]

if and only if

\[
\phi([a_1^i]_U, \ldots, [a_k^i]_U) \text{ is true.}
\]

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

The \( \neg, \lor, \land \) 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 \( a^i = [a_i^i]_U \) 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:

\[
\phi(a^1, \ldots, a^k) \text{ is true if and only if } a^i = a^j \quad \text{if and only if } \{ n \mid a_i^n = a_j^n \} \in U \quad \text{if and only if } \{ n \mid \phi(a_1^n, \ldots, a_k^n) \text{ is true} \} \in U.
\]

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

\[
\{ n \mid \psi(a_1^n, \ldots, a_k^n) \text{ is true} \} \in U \text{ if and only if } \psi(a^1, \ldots, a^k) \text{ is true.}
\]
Then
\[ \phi(a^1, \ldots, a^k) \text{ is true if and only if } \psi(a^1, \ldots, a^k) \text{ is false} \]
if and only if \( \{ n \mid \psi(a^1_n, \ldots, a^k_n) \text{ is true} \} \not\in U \)
if and only if \( \{ n \mid \psi(a^1_n, \ldots, a^k_n) \text{ is false} \} \in U \)
if and only if \( \{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \).

The conjunction and disjunction cases are similar. Suppose \( \psi_0 \) and \( \psi_1 \) are two formulas for which, inductively, we have already proven the theorem, so

\[ \{ n \mid \psi_0(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \text{ if and only if } \psi_0(a^1, \ldots, a^k) \text{ is true} \]
and

\[ \{ n \mid \psi_1(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \text{ if and only if } \psi_1(a^1, \ldots, a^k) \text{ is true} \]

First, suppose \( \phi \) is \( \psi_0 \land \psi_1 \). Then
\[ \phi(a^1, \ldots, a^k) \text{ is true if and only if } \psi_0(a^1, \ldots, a^k) \text{ and } \psi_1(a^1, \ldots, a^k) \text{ are both true} \]
if and only if \( \{ n \mid \psi_0(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \text{ and } \{ n \mid \psi_1(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \)
if and only if \( \{ n \mid \psi_0(a^1_n, \ldots, a^k_n) \text{ and } \psi_1(a^1_n, \ldots, a^k_n) \text{ are both true} \} \in U \)
if and only if \( \{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \).

Next, suppose \( \phi \) is \( \psi_0 \lor \psi_1 \). Then
\[ \phi(a^1, \ldots, a^k) \text{ is true if and only if either } \psi_0(a^1, \ldots, a^k) \text{ or } \psi_1(a^1, \ldots, a^k) \text{ is true} \]
if and only if \( \{ n \mid \psi_0(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \text{ or } \{ n \mid \psi_1(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \)
if and only if \( \{ n \mid \psi_0(a^1_n, \ldots, a^k_n) \text{ is true} \} \cup \{ n \mid \psi_1(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \)
if and only if \( \{ n \mid \text{ either } \psi_0(a^1_n, \ldots, a^k_n) \text{ or } \psi_1(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \)
if and only if \( \{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U \).

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

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

Conversely, suppose $K = \{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is true} \} \in U$ then, for each $n \in K$, there is a $b_n$ so that $\{ n \mid \psi(a^1_n, \ldots, a^k_n, b_n) \text{ is true} \} \in U$. Then we can take $b = [b_n]_U$ (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 $\phi$ 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 $b \in [V_n]_U$, $\psi(a^1, \ldots, a^k, b)$ is true, so there must be some $b = [b_n]_U \in [V_n]_U$ such that $\psi(a^1, \ldots, a^k, b)$ is false. Then, by the inductive hypothesis, $\{ n \mid \psi(a^1_n, \ldots, a^k_n, b_n) \text{ is true} \} \notin U$, so—because $U$ is an ultrafilter—$\{ n \mid \psi(a^1_n, \ldots, a^k_n, b_n) \text{ is false} \} \in U$, and therefore $\{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is false} \} \in U$. That means that $\{ n \mid \phi(a^1_n, \ldots, a^k_n) \} \notin U$, which is what we needed to show.

For the converse, suppose $\{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is true} \} \notin U$, so $K = \{ n \mid \phi(a^1_n, \ldots, a^k_n) \text{ is false} \} \in U$. Then, for each $n \in K$, there must be some $b_n \in V_n$ so that $\phi(a^1_n, \ldots, a^k_n, b_n)$ is false. Taking $b = [b_n]_U$, we have $\{ n \mid \psi(a^1_n, \ldots, a^k_n, b_n) \text{ is false} \} \supseteq K \in U$, 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 3.31.** If $\phi(x_1, \ldots, x_m)$ is a first-order formula, $k \leq m$, and $[a^k_{n+1}]_U, \ldots, [a^n_m]_U \in [V_n]_U$, and

$$X_n = \{ (a^1, \ldots, a^k) \in V_n \mid \phi(a^1, \ldots, a^k, a^k_{n+1}, \ldots, a^n_m) \}$$

then

$$[X_n]_U = \{ (a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, [a^k_{n+1}]_U, \ldots, [a^n_m]_U) \}.$$ 

**Proof.** Let $X = \{ (a^1, \ldots, a^k) \in V^k \mid \phi(a^1, \ldots, a^k, [a^k_{n+1}]_U, \ldots, [a^n_m]_U) \}$ and let $a^i = [a^i_n]_U$. Then

$$(a^1, \ldots, a^k) \in X$$

if and only if $\phi(a^1, \ldots, a^k, a^k_{n+1}, \ldots, a^n_m)$

if and only if $\{ n \mid \phi(a^1, \ldots, a^k, a^k_{n+1}, \ldots, a^n_m) \text{ is true} \} \in U$

if and only if $\{ n \mid (a^1_n, \ldots, a^k_n) \in X_n \} \in U$

if and only if $(a^1, \ldots, a^k) \in [X_n]_U$.

Therefore $X = [X_n]_U$. 

---

The ability to take the sequence of witnesses $b_n$ and combine them into the single witness $b = [b_n]_U$ is of course crucial here.
**Definition 3.33.** 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 be stated briefly as:

**Corollary 3.34.** Definable subsets of an ultraproduct are internal.

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

Notice that if a formula has no variables (other than those bound by quantifiers) then it defines a subset of the “empty product” $V^0$. For instance, consider a formula like $\forall x \exists y \{x,y\} \in E$. Formally, $V^0$ should be a set with a single element, say $\{\ast\}$, so $\forall x \exists y \{x,y\} \in E$ either defines the set $\{\ast\}$ or the set $\emptyset$—that is, $\forall x \exists y \{x,y\} \in E$ is either true in the graph $(V,E)$ or it is false in the graph $(V,E)$.

**Definition 3.35.** A formula in which all variables are bound by quantifiers is called a sentence.

We often write $\sigma$ or $\tau$ for a sentence.

**Corollary 3.36.** Suppose $\sigma$ is a sentence. Then $\sigma$ is true in $[G_n]_U$ if and only if $\{n \mid \sigma \text{ is true in } G_n\} \in U$.

We can now see what makes $K_{\infty,\infty}$ different from $K_{\infty,\infty^2}$:

**Theorem 3.37.** In the graph $K_{\infty,\infty}$, there is an internal bijection between the two parts. In the graph $K_{\infty,\infty^2}$, there is no internal bijection between the two parts.

**Proof.** Note that the property of being a bijection is first-order, and therefore internal: to say that $\pi : V_0 \rightarrow V_1$ is a bijection is to say

$$(\forall x \in V_0 \forall y \in V_0 \pi(x) = \pi(y) \rightarrow x = y) \land (\forall y \in V_1 \exists x \in V_0 \pi(x) = y).$$

So saying that $\pi = [\pi_n]_U$ is an internal bijection is the same as saying that $\{n \mid \pi_n \text{ is a bijection}\} \in U$. So to see that there is an internal bijection between the two parts in $K_{\infty,\infty}$, we need only observe that such a bijection $\pi_n$ exists for each $n$ and let $\pi = [\pi_n]_U$.

On the other hand, if there were an internal bijection $\pi = [\pi_n]_U$ between the two parts in $K_{\infty,\infty^2}$ then $\pi_n$ would have to be a bijection for almost every $n$. But there is no such bijection in $K_{n,n^2}$, so no such bijection in $K_{\infty,\infty^2}$. \[\square\]

---

*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.*
3.8 Saturation

Every finite set is internal (the set \( \{a^1, \ldots, a^k\} \) can be defined as \( \{x \in V \mid x = a^1 \vee \ldots \vee x = a^k\} \) 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 3.38 (Saturation).** Suppose that, for each \( i \in \mathbb{N} \), \( X^i = [X^i_n]_\mathcal{U} \) is an internal set, and that, for every \( k \), \( \bigcap_{i \leq k} X^i \) is non-empty. Then \( \bigcap_{i \in \mathbb{N}} X^i \) is non-empty.

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

For each \( i \),
\[
\{n \mid v_n \in X^i_n\} \supseteq \{n \mid i \leq n \text{ and } \bigcap_{i \leq n} X^i \text{ is non-empty}\} \in \mathcal{U},
\]
so \( [v_n]_\mathcal{U} \in [X^i_n]_\mathcal{U} = X^i \). Since this holds for every \( i \), \( [v_n]_\mathcal{U} \in \bigcap_i X^i \).

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

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

**Theorem 3.39.** Suppose that, for each \( i \), \( X^i \) is internal. If \( \bigcup_i X^i \) is internal then there is a \( k \) so that \( \bigcup_i X^i = \bigcup_{i \leq k} X^i \).

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

Saturation is 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 3.40.** If \( G = (V, E) \) is a graph and \( v, w \in V \), the distance between \( v \) and \( w \) is the smallest \( k \) such that there is a path \( v = v_0, v_1, \ldots, v_k = w \) such that \( \{v_i, v_{i+1}\} \in E \) for all \( i < k \), or \( \infty \) 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 3.41. \([G_n]_U\) is connected if and only if \([G_n]_U\) has finite diameter.

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

\(X^k\) 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:

\[
\neg[v = w \lor \{v, w\} \in E \lor \exists y_1(\{v, y_1\} \in E \land \{y_1, w\} \in E) \lor \exists y_1 \exists y_2(\{v, y_1\} \in E \land \{y_1, y_2\} \in E \land \{y_2, w\} \in E) \lor \ldots \lor \exists y_1 \exists y_2 \ldots \exists y_{k-1}(\{v, y_1\} \in E \land \{y_1, y_2\} \in E \land \ldots \land \{y_k, w\} \in E)].
\]

Since \([G_n]_U\) 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 \([G_n]_U\) 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.

3.9 Extending Sequences

We will later need to use saturation to take countable sequences and extend them. Suppose we have internal sets \(A^1 \subseteq A^2 \subseteq A^3 \subseteq \ldots\). We have already seen that, unless the sequence stops growing at some finite \(A^n\), 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 \(A^i = [A^i_n]_U\). Then there is a “diagonal” set \(A^* = [\bigcup_{i \leq n} A^n_i]_U\). (One might wonder why we use \(\bigcup_{i \leq n} A^n_i\) rather than just \(A^n\), 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 $[v_n]_U \in \bigcup_{i \in \mathbb{N}} A_i$ then there is some $i$ so that $\{n \mid v_n \in A^i_n\} \in U$, and therefore $\{n \mid i \leq n \text{ and } v_n \in A^i_n\} \in U$ as well, so $[v_n]_U \in A^*$.

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

More generally, whenever $f : \mathbb{N} \to \mathbb{N}$ is a function, we can define an internal set $A^f = \bigcup_{i \leq f(n)} A^i_n |_U$. As long as $\{n \mid i \leq f(n)\} \in U$, we will have $A^i \subseteq A^f$. (Indeed, more generally, if $\{n \mid g(n) \leq f(n)\} \in U$ 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 3.42.** Suppose that

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

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

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

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

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

---

### 3.10 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, $\{[a_n]_U, [b_n]_U\} \in E_U$ when $\{n \mid \{a_n, b_n\} \in E_n\} \in U$, 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, \( C \), 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 \( C \). (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 \( C \) depending on the voter preferences. (Note that we are requiring that \( F \) produce a full, linear ranking on \( C \), 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 \( |C| \) points to the top candidate in \( \prec_n \), \(|C| - 1\) 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 3.43** (Arrow’s Impossibility Theorem). If \( N \) is finite and there are at least three candidates, there is no preference aggregation rule such that:

1. If \( a \succ_n b \) for every \( n \in N \) then \( a \succ b \).
2. Suppose that \( \{\prec_n\}_{n \in N} \) and \( \{\prec'_n\}_{n \in N} \) are two different lists of preferences, \( a, b \in C \), and for every \( n \in N \),

   \[
   a \prec_n b \iff a \prec'_n b.
   \]

   Then also

   \[
   a \prec b \iff a \prec' b.
   \]

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 $n \in K \setminus J$, 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.\footnote{Though not necessarily a free filter—we have not promised that every cofinite set is victorious, and indeed, $N$ is finite, so we do not have a free filter.}

Next, suppose $K \setminus N$ is not victorious. That means there is some list of preferences $\{\prec_n\}_{n \in N}$ so that $a \succ \prec_n b$ 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.\footnote{In the conventional definition, where we work over filters rather than free filters.} But, since $N$ is finite, $N = \{n_1\} \cup \cdots \cup \{n_k\}$ for some finite $k$, and therefore there is some $i$ so that $\{n_i\} \subseteq N$. But then $n_i$ is a dictator: whenever $a \succ_{n_i} b$, we have $a \succ b$.\end{proof}

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 $N$.\end{proof}
3.11 Remarks

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

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

The space of ultrafilters has a structure of its own, as a topological semi-group—indeed, with a discrete space like \( \mathbb{N} \), the collection of ultrafilters coincides with the maximal compactification of the space, the Stone-Čech compactification. Ultrafilters have a number of direct (that is, without constructing an ultraproduct) applications in mathematics \([11, 51]\), and it is an interesting open question whether all the uses of ultrafilters can be channeled through ultraproducts \([25]\).

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 \([28]\), 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 \([24, 11]\). Ultraproducts also pay an important role in set theory \([59, 67]\), 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 \([10]\), 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

---

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

\[
\forall x \exists y_1 \exists y_2 \{ \{ x, y_1 \} \in E \land \{ x, y_2 \} \in E \land y_1 \neq y_2 \land \forall z \neg \{ x, z \} \in E \lor z = y_1 \lor z = y_2 \},
\]

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

For any fixed finite cycle \( k \), \( t_{C_k}(C_n) = 0 \) 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 \( [G_n]_\mathcal{U} \) as a graph: uncountably many chains which stretch forever in both directions.
However \( \lim_{n \to \mathcal{U}} t_{K_2}(G_n) = 0 \): the graph \( G_n \) has \( n \) vertices and \( n \) edges, so \( t_{K_2}(G_n) = \frac{n}{n^2} = \frac{1}{n} \). So even though \([G_n]_{\mathcal{U}}\) has uncountably many edges, the “density” of these edges should still be 0.

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

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

\( \square \)
4.2. PROBABILITY ON INTERNAL SETS

For example, suppose that each $G_n$ is a complete bipartite graph where the sides have \textit{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, $[E_n]_U$ will be a complete bipartite graph on $[V_n]_U$. $[X_n]_U$ and $[Y_n]_U$ are subsets of $[V_n]_U$ and

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

In the finite graphs $K_{n,n+1}$, $X_n$ is \textit{approximately} half the vertices (at least when $n$ is large), but in the ultraproduct, the measure of $[X_n]_U$ is \textit{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, errors 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, $\{n \mid \mu_n(X_n) > 1/2\} = \mathbb{N} \in U$, but $\mu([X_n]_U) \leq 1/2$.

However this is the most that $\mu([X_n]_U)$ can deviate from the behavior of the $\mu_n(X_n)$:

\textbf{Lemma 4.2.} Whenever $X \subseteq [V_n]_U$ is internal, taking $X = [X_n]_U$ we have:

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

All parts of the lemma follow from the definition of $\lim_{n \to U}$.

$\mu$ inherits many of the rules we expect of probability from the ground models:

\textbf{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 \( X \cup Y = [X_n \cup Y_n]_\mathcal{U} \):

\[
[v_n]_\mathcal{U} \in X \cup Y \text{ if and only if } [v_n]_\mathcal{U} \in X \text{ or } [v_n]_\mathcal{U} \in Y \\
\text{ if and only if } \{ n \mid v_n \in X_n \} \in \mathcal{U} \text{ or } \{ n \mid v_n \in Y_n \} \in \mathcal{U} \\
\text{ if and only if } \{ n \mid v_n \in X_n \text{ or } v_n \in Y_n \} \in \mathcal{U} \\
\text{ if and only if } [v_n]_\mathcal{U} \in [X_n \cup Y_n]_\mathcal{U}.
\]

Similarly, \( X \cap Y = [X_n \cap Y_n]_\mathcal{U} \):

\[
[v_n]_\mathcal{U} \in X \cup Y \text{ if and only if } [v_n]_\mathcal{U} \in X \text{ and } [v_n]_\mathcal{U} \in Y \\
\text{ if and only if } \{ n \mid v_n \in X_n \} \in \mathcal{U} \text{ and } \{ n \mid v_n \in Y_n \} \in \mathcal{U} \\
\text{ if and only if } \{ n \mid v_n \in X_n \text{ and } v_n \in Y_n \} \in \mathcal{U} \\
\text{ if and only if } [v_n]_\mathcal{U} \in [X_n \cap Y_n]_\mathcal{U}.
\]

Therefore

\[
\mu(X \cup Y) = \lim_{n \to \mathcal{U}} \mu_n(X_n \cup Y_n) \\
= \lim_{n \to \mathcal{U}} (\mu_n(X_n) + \mu_n(Y_n) - \mu_n(X_n \cap Y_n)) \\
= \lim_{n \to \mathcal{U}} \mu_n(X_n) + \lim_{n \to \mathcal{U}} \mu_n(Y_n) - \lim_{n \to \mathcal{U}} \mu_n(X_n \cap Y_n) \\
= \mu(X) + \mu(Y) - \mu(X \cap Y).
\]

More generally, we have finite additivity.

**Corollary 4.4.** Suppose \( B^1, \ldots, B^k \) are pairwise disjoint internal sets. Then \( \mu(\bigcup_{i \leq k} B^i) = \sum_{i \leq k} \mu(B^i) \).

**Proof.** By induction on \( k \), using the previous lemma. \( \square \)

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

**Definition 4.5.** If \( \mathcal{B} \subseteq \mathcal{P}(V) \) (the power set of \( V \)), we say \( \mathcal{B} \) is an algebra if:

- \( \emptyset \in \mathcal{B} \) and \( V \in \mathcal{B} \),
- whenever \( B \in \mathcal{B} \), \( V \setminus B \in \mathcal{B} \),
- whenever \( B_0, B_1 \in \mathcal{B} \), also \( B_0 \cup B_1 \in \mathcal{B} \).
We say $B$ is a $\sigma$-algebra if, additionally, whenever $B_i \in B$ for every $i \in \mathbb{N}$, $\bigcup_{i \in \mathbb{N}} B_i \in B$.

When $B \subseteq \mathcal{P}(V)$ is a $\sigma$-algebra, a probability measure on $(V, B)$ is a function $\mu : B \to [0, 1]$ such that:

- $\mu(V) = 1$,
- if $B_i \in B$ 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 $(V, B, \mu)$ where $B \subseteq \mathcal{P}(V)$ is a $\sigma$-algebra and $\mu$ is a probability measure.

Since algebras and $\sigma$-algebras are closed under unions and complements, they are also closed under finite and countable intersections, respectively.

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 $\mu$ beyond the internal sets. However, as a first step, we can show that $\mu$ satisfies the countable additivity it is supposed to have as long as the countable union happens to be internal.

Lemma 4.6. Suppose that, for each $i$, $B^i \in B$ is internal and $B^i \cap B^j = \emptyset$ whenever $i \neq j$. 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 3.39 if $\bigcup_{i \in \mathbb{N}} B^i$ is internal then $\bigcup_{i \in \mathbb{N}} B^i = \bigcup_{i \leq k} B^i$ for some finite $k$. Since the $B^i$ are pairwise disjoint, that means $B^i = \emptyset$ for $i > k$ and therefore $\mu(B^i) = 0$ for $i > k$. Therefore this follows from finite additivity.

This makes $\mu$ a pre-measure on the internal sets: it has the properties a probability measure should have as long as the sets involved are internal.

We should note that $\mu$ respects “internal cardinality”, in the following sense:

Lemma 4.7.

- If $A$ and $B$ are internal sets and $f : A \to B$ is an internal one-to-one function then $\mu(A) = \mu(B)$.
- More generally, if $R \subseteq A \times B$ is an internal relation such that
  - for each $a \in A$, $|\{b \mid (a, b) \in R\}| = k$,
  - for each $b \in B$, $|\{a \mid (a, b) \in R\}| = m$,
then $k \mu(A) = m \mu(B)$.

**Proof.** We prove the more general second part. We have $A = [A_n]_\mathcal{U}$, $B = [B_n]_\mathcal{U}$, and $R = [R_n]_\mathcal{U}$. Since the properties “for every $a$, $|\{b \mid (a, b) \in R\}| = k$” and “for every $b$, $|\{a \mid (a, b) \in R\}| = m$” are first-order, we know that the set of $n$ so that $R_n \subseteq A_n \times B_n$ has these properties is in $\mathcal{U}$.

For any such $n$, we have $k|A_n| = m|B_n|$, so $k \mu(A_n) = m \mu(B_n)$. Since this holds for a set of $n$ belonging to $\mathcal{U}$, we also have $k \mu(A_n) = m \mu(B_n)$.

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 4.8.** If each $A^i$ is internal then there is an internal set $A^+$ such that:

- $\bigcup_{i \in \mathbb{N}} A^i \subseteq A^+$, and
- $\mu(A^+) = \lim_{m \to \infty} \mu(\bigcup_{i \leq m} A^i)$.

**Proof.** 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 $\epsilon$: 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 $f^\epsilon$ by setting $f^\epsilon(n)$ to be the largest $m$ so that $\frac{|\bigcup_{i \leq m} A^i_n|}{|A_n^i|} \leq c + \epsilon$ and set $A^{+\epsilon} = A^{f^\epsilon}$—that is, $A^{+\epsilon} = [\bigcup_{i \leq f^\epsilon(n)} A^i_n]_\mathcal{U}$. Since $\frac{|\bigcup_{i \leq f^\epsilon(n)} A^i_n|}{|A_n^i|} \leq c + \epsilon$ 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 $\{n \mid \frac{|\bigcup_{i \leq i} A^i_n|}{|A_n^i|} \leq c + \epsilon\} \in \mathcal{U}$, so $\{n \mid A^i_n \subseteq \bigcup_{j \leq f^\epsilon(n)} A^j_n\} \in \mathcal{U}$, and therefore $A^i \subseteq A^{+\epsilon}$.

Now consider the sequences

$$A^1 \subseteq \bigcup_{i \leq 2} A^i \subseteq \bigcup_{i \leq 3} A^i \subseteq \cdots \subseteq A^{+1/3} \subseteq A^{+1/2} \subseteq A^+.$$ 

Then by Theorem 3.42, there is an internal $A^+$ with $\bigcup_{i \in \mathbb{N}} A^i \subseteq A^+$, and $A^+ \subseteq A^{+1/i}$ for all $i$. Therefore $\mu(A^+) \leq c + 1/i$ for all $i$, so $\mu(A^+) = c$.

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

---

94

CHAPTER 4. INTEGRALS IN ULTRAPRODUCTS
4.3 Probability Spaces

What about non-internal sets? We cannot hope to extend $\mu$ to make sense for all the subsets of $[V_n]_U$: 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 $[V_n]_U$. There will have to be some unmeasurable sets which are not in the domain of $\mu$.

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 4.9.** We say $B \subseteq [V_n]_U$ is $\mu$-null if, for every $\epsilon > 0$, there is an internal set $A_\epsilon$ such that $B \subseteq A_\epsilon$ and $\mu(A_\epsilon) < \epsilon$.

It is almost immediate that a finite union of $\mu$-null sets is also $\mu$-null. Less trivially, even a countable union of $\mu$-null sets is $\mu$-null.

**Lemma 4.10.** If each $B_i$ is $\mu$-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 4.8 there is an internal set $A^+ \supseteq \bigcup_{i \in \mathbb{N}} A_i \supseteq \bigcup_{i \in \mathbb{N}} B_i$ with $\mu(A^+) = \sum_i \mu(A'_i) \leq \sum_i \mu(A_i) < \epsilon$. \qed

More generally, we could extend $\mu$ to any set which is within a null set of an internal set:

**Definition 4.11.** We say $B \subseteq [V_n]_U$ is $\mu$-approximable if there is an internal set $A$ so that $A \triangle B$ is $\mu$-null. We call $A$ a $\mu$-approximation of $B$.

We write $\mathcal{B}(\mu)$ for the set of $\mu$-approximable sets. We define $\mu : \mathcal{B}(\mu) \to [0, 1]$ by $\mu(B) = \mu(A)$ where $A$ is any $\mu$-approximation of $B$.

Note that this definition gives a unique value for the measure of each $\mu$-approximable set: if $A$ and $A'$ are two $\mu$-approximations of $B$ then $A \triangle A' \subseteq (A \triangle B) \cup (A' \triangle B)$ is an internal $\mu$-null set, and therefore has measure 0, so $\mu(A) = \mu(A')$.

The $\mu$-approximable sets are a suitable family of sets to use as the basis for probability theory.

**Theorem 4.12.** The $\mu$-approximable sets are a $\sigma$-algebra.
Theorem 4.13. \( \mu \) is a probability measure on \( \mathcal{B}(\mu) \).

Proof. The only thing to check is countable additivity. Suppose each \( B_i \) is \( \mu \)-approximable and \( B_i \cap B_j = \emptyset \) whenever \( i \neq j \). For each \( B_i \), fix a \( \mu \)-approximation \( A_i \). Observe that when \( i \neq j \), \( \mu(A_i \cap A_j) = 0 \) since \( B_i \cap B_j = \emptyset \), \( A_i \cap A_j \subseteq (A_i \triangle B_j) \cup (A_j \triangle B_i) \), and is therefore \( \mu \)-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 \( \mu(A^+) = \sum_i \mu(A'_i) \). Since

\[
A^+ \setminus \bigcup_{i \in \mathbb{N}} B_i \subseteq (A^+ \setminus \bigcup_{i \in \mathbb{N}} A_i) \cup \bigcup_{i \in \mathbb{N}} (A_i \setminus B_i)
\]

is \( \mu \)-null, \( A^+ \) is a \( \mu \)-approximation of \( \bigcup_{i \in \mathbb{N}} B_i \), so

\[
\mu\left( \bigcup_{i \in \mathbb{N}} B_i \right) = \mu(A^+) = \sum_i \mu(A_i) = \sum_i \mu(B_i).
\]

The \( \mu \)-approximable sets are naturally complete: any subset of a measure 0 set is measurable.

Lemma 4.14. 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 \( \mu \)-approximation of \( Y \). Since \( Y \in \mathcal{B}(\mu) \) with measure 0, there is an internal set \( A \) with \( \mu(A) = 0 \) so that
4.4. PROBABILITY SPACES ON $[V_n]_U^k$

$Y \triangle A$ is $\mu$-null. Then, for each $\epsilon > 0$, there is a $B_\epsilon$ 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 $\mu$-approximation of $X$, so $\mu(X) = 0$.

\begin{proof}
If $X \subseteq [V_n]_U$ 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 $\mu(X) = 0$.
\end{proof}

**Corollary 4.15.** If $X \subseteq [V_n]_U$ 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 $Y_\epsilon$ were internal, this would be immediate because $X$ would be $\mu$-null. Instead, note that $X \subseteq \bigcap_\epsilon Y_\epsilon = Y$ and, since $\mathcal{B}(\mu)$ is a $\sigma$-algebra, $Y \in \mathcal{B}(\mu)$ and $\mu(Y) = 0$. Therefore $X \in \mathcal{B}(\mu)$ and $\mu(X) = 0$.

\begin{proof}
\end{proof}

### 4.4 Probability Spaces on $[V_n]_U^k$

The sets we're most interested in the probability of are not subsets of $[V_n]_U$ itself, but rather subsets of $[V_n]_U^k$ for various $k$. For instance, we are interested in the measure of $\{(x, y) \mid \{x, y\} \in [E_n]_U\} \subseteq [V_n]_U^2$, or the measure of $\{(x, y, z) \mid \{x, y\} \in [E_n]_U \text{ and } \{x, z\} \in [E_n]_U \text{ and } \{y, z\} \in [E_n]_U\} \subseteq [V_n]_U^3$ (that is, the density of the set of triangles).

The sets $[V_n]_U^k$ are themselves ultraproducts—$[V_n]_U^k$ is the same as $[V_n^k]_U$. 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 $[W_n]_U = [V_n]_U^k = [V_n]_U^k$.

**Definition 4.16.** For any $k \in \mathbb{N}$, we define $\mu_k$ on internal subsets $X = [X_n]_U \subseteq [V_n]_U^k$ by

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

A set $B \subseteq [V_n]_U^k$ is $\mu_k$-null if, for every $\epsilon > 0$, there is an internal set $A_\epsilon$ such that $B \subseteq A_\epsilon$ and $\mu_k(A_\epsilon) < \epsilon$.

A set $B \subseteq [V_n]_U^k$ is $\mu_k$-approximable if there is an internal set $A$ such that $A \triangle B$ is $\mu$-null, and we call $A$ a $\mu_k$-approximation of $B$.

We write $\mathcal{B}(\mu_k)$ for the set of $\mu_k$-approximable sets and define $\mu_k : \mathcal{B}(\mu_k) \to [0, 1]$ by $\mu_k(B) = \mu_k(A)$ where $A$ is any $\mu_k$-approximation of $B$.

As in the previous section, we have

**Theorem 4.17.** $\mathcal{B}(\mu_k)$ is a $\sigma$-algebra and $\mu_k$ is a probability measure on $\mathcal{B}(\mu_k)$.
However the spaces $[V_n]^k_U$ are more interesting than just $[V_n]_\mathcal{U}$ by itself, because these spaces are related to each other—the measure on $[V_n]^2_U$ 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 $\mu_2$: by an integral. For instance, when $X \subseteq [V_n]^2_U$, 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 $\mu_2$ 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. $\mu_2$ 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 $([V_n]^k_U, \mathcal{B}(\mu_k), \mu_k)$ 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.
4.4. PROBABILITY SPACES ON $[V]_N^K$

For the moment, however, we want to understand how the spaces $([V]_N^K, B(\mu_k), \mu_k)$ relate to each other. First, we need to describe several ways we could use the description of a set by coordinates to define other sets.

**Definition 4.18.** If $X \subseteq V^r$ and $\pi : [1, r] \rightarrow [1, r]$ is a bijection then

$$X^\pi = \{(x_1, \ldots, x_r) \mid (x_{\pi(1)}, \ldots, x_{\pi(r)}) \in X\} \in B_r.$$  

If $k \leq r$ and $x_1, \ldots, x_k \in V$, the slice of $X$ corresponding to $x_1,\ldots, x_k$ is

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

We can now define the properties we expect our family of measures to have:

**Definition 4.19.** Let $V$ be a set and suppose that, for each $k$, we have a probability measure space $(V^k, B_k, \mu_k)$. The spaces $\{(V^k, B_k, \mu_k)\}_{k \in \mathbb{N}}$ are a **Keisler graded probability space** if:

- (Symmetry) whenever $\pi : [1, k] \rightarrow [1, k]$ is a bijection and $X \in B_k$,
  - $X^\pi \in B_k$, and
  - $\mu_k(X^\pi) = \mu_k(X),$

- (Products) whenever $B \in B_k$ and $C \in B_r$, $B \times C \in B_{k+r},$

- (Fubini Property) whenever $X \in B_{k+r}$
  - the set of $x_1, \ldots, x_k \in V$ such that $X_{x_1,\ldots,x_k} \in B_r$ belongs to $B_k$ and has $\mu_k$-measure 1, and
  - $\mu_{k+r}(X) = \int \mu_r(X_{x_1,\ldots,x_k}) \, d\mu_k(x_1, \ldots, x_k).$

These properties seem technical, but they are exactly the properties we would expect a family of probability measures on $V^k$ to have. The first two properties guarantee that there are enough measurable sets.

However we are allowed to—as we typically will—have additional sets beyond those required by the product property. The Fubini property puts a restriction on these new sets, requiring that these additional sets be assigned measures in a way that is consistent with the measures assigned to sets of lower arity: their slices have to exist in the lower arity $\sigma$-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 4.20. If \( \{(V^k, \mathcal{B}_k, \mu_k)\}_{k \in \mathbb{N}} \) is a Keisler graded probability space then for any measurable function on \( V^{k+r} \),

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

Proof. For any \( \epsilon > 0 \), we may choose a simple function \( g = \sum_{i \leq d} c_i \chi_{B_i} \) 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

\[
\int f \, d\mu_{k+r} = \int g \, d\mu_{k+r} + e \, |e| < \epsilon
\]

where \( |e| < \epsilon \)

\[
= \sum_{i \leq d} c_i \mu_{k+r}(B_i) + e \text{ where } |e| < \epsilon
\]

\[
= \sum_{i \leq d} c_i \int \mu_r((B_i)_{x_1,\ldots,x_k}) \, d\mu_k + e \text{ where } |e| < \epsilon
\]

\[
= \int \sum_{i \leq d} c_i \mu_r((B_i)_{x_1,\ldots,x_k}) \, d\mu_k + e \text{ where } |e| < \epsilon
\]

\[
= \int \left( \int f(x_1, \ldots, x_{k+r}) \, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right) \, d\mu_k(x_1, \ldots, x_k) + e' \text{ where } |e'| < 2\epsilon.
\]

Since this holds for every \( \epsilon > 0 \),

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

\( \square \)

The key fact we need is that the natural measure spaces on the ultraproduct form a Keisler graded probability space.

Theorem 4.21. The collection of probability measure spaces \( \{(V^k, \mathcal{B}(\mu_k), \mu_k)\}_{k \in \mathbb{N}} \) 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 \( A = [A_n]_U \) be internal. For any bijection \( \pi : [1, k] \to [1, k] \), since \( A^\pi = [A^\pi_n]_U \) and \( (\mu_k)_n(A^\pi_n) = (\mu_k)_n(A_n) \), we have \( \mu(A^\pi) = \mu(A) \).

Next, consider an arbitrary \( X \in \mathcal{B}^k \) and take a \( \mu_k \)-approximation \( A \). Then, for any \( \epsilon > 0 \), we have an internal set \( \mathcal{A}_\epsilon \) with \( \mu_k(A_{\epsilon}) < \epsilon \) and \( (A \triangle X) \subseteq \mathcal{A}_\epsilon \). But also \( \mu_k(A^\pi_{\epsilon}) = \mu_k(A_{\epsilon}) \) \( < \epsilon \) and \( (A^\pi \triangle X^\pi) \subseteq A^\pi_{\epsilon} \), so \( A^\pi \) is a \( \mu_k \)-approximation to \( X^\pi \), so \( X^\pi \in \mathcal{B}^k \) and \( \mu_k(X^\pi) = \mu_k(A^\pi) = \mu_k(A) = \mu_k(X) \).
4.4. **Probability Spaces on \([V_N]_U^K\)

Products of internal sets are certainly internal: if \(A = [A_n]_U \subseteq [V_n]_U^K\) and \(B = [B_n]_U \subseteq [V_n]_U^K\) are internal then \(A \times B = [A_n \times B_n]_U\).

Suppose that, \(X \in B_k\) and \(Y \in B_r\). Consider a \(\mu_k\)-approximation \(A\) of \(X\) and \(\mu_r\)-approximation \(B\) of \(Y\). Then for any \(\epsilon > 0\), we may choose internal sets \(A_r \supseteq (A \triangle X)\) and \(B_\epsilon \supseteq (B \triangle Y)\) with \(\mu_k(A_r) < \epsilon/2\) and \(\mu_r(B_\epsilon) < \epsilon/2\). Then

\[
(A \times B) \triangle (X \times Y) \subseteq [(A \triangle X) \times [V_n]_U^K \cup ([V_n]_U^K \times (B \triangle Y))
\]

\[
= (A_r \times [V_n]_U^K) \cup ([V_n]_U^K \times B_\epsilon),
\]

and \(\mu_{k+r}((A_r \times [V_n]_U^K) \cup ([V_n]_U^K \times B_\epsilon)) < \epsilon\).

Finally, we turn to the Fubini property. Once again, we start with internal sets. When \(A = [A_n]_U \in B^{k+r}\) is internal, \(A_{x_1,\ldots,x_k}\) is also internal for any choice of \(x_1,\ldots,x_k\), and therefore is in \(B^r\).

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

\[
\int (\mu_r)_n((A_n)_{x_1,\ldots,x_k})d(\mu_k)_n(x_1,\ldots,x_k) = \frac{1}{|V_n|^k} \sum_{(x_1,\ldots,x_k) \in V_n^k} \frac{|(A_n)_{x_1,\ldots,x_k}|}{|V_n|^r}
\]

\[
= \frac{1}{|V_n|^{k+r}} \sum_{(x_1,\ldots,x_k) \in V_n^k} |(A_n)_{x_1,\ldots,x_k}|
\]

\[
= \frac{|A_n|}{|V_n|^{k+r}} = \mu_{k+r}^n(A_n).
\]

Moreover, we have \(\lim_{n \to U} \mu_{k+r}^n(A_n) = \mu_{k+r}(A_n)\), so what we need to do is show that

\[
\lim_{n \to U} \int (\mu_r)_n((A_n)_{x_1,\ldots,x_k})d(\mu_k)_n(x_1,\ldots,x_k) = \int \mu_r((A_n)_{x_1,\ldots,x_k})d\mu_k(x_1,\ldots,x_k).
\]

This does not immediately follow from our definitions—to prove this, we have to look carefully at how integrals are calculated. Integrals are approximated by finite sums: we can break the interval \([0,1]\) into small sub-intervals: let \(J = \{1/\epsilon\}\) 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
\[
(\mu^k + \epsilon_n)(A_n) = \int (\mu^k)(A_{x_1, \ldots, x_k})d(\mu^k_n)(x_1, \ldots, x_k)
\]
\[
= \sum_{j \leq J} \int_{C_j} (\mu^k)(A_{x_1, \ldots, x_k})d(\mu^k_n)(x_1, \ldots, x_k)
\]
\[
\approx \sum_{j \leq J} (\mu^k_n)(C_j)(j + 1/2)\epsilon.
\]
More precisely, since each interval $I_j$ has radius $\epsilon/2$, we have
\[
\left| \int (\mu^k)(A_{x_1, \ldots, x_n})d(\mu^k_n)(x_1, \ldots, x_n) - \sum_{j \leq J} (\mu^k_n)(C_j)(j + 1/2)\epsilon \right| < \epsilon/2.
\]

Consider the corresponding sets in the ultraproduct, $C^j = [C^j_n]_U$. Note that $C^j$ is not quite $\{(x_1, \ldots, x_k) \mid \mu^r(A_{x_1, \ldots, x_k}) \in I_j\}$—because of the behavior at limits, when $\mu^r(A_{x_1, \ldots, x_k}) = j\epsilon$, we could have either $(x_1, \ldots, x_k) \in I_j$ or $(x_1, \ldots, x_k) \in I_{j-1}$. Instead, what we have is that $\mu^r(A_{x_1, \ldots, x_k})$ is in the closure of $I_j$: if $(x_1, \ldots, x_k) = ([x^1_n]_U, \ldots, [x^k_n]_U) \in C^j$ then
\[
\mu^r(A_{x_1, \ldots, x_n}) = \lim_{n \to U} (\mu^r_n)(A_n)_{x^1_n, \ldots, x^k_n} \in \overline{I}_j = [j\epsilon, (j + 1)\epsilon].
\]
Moreover, $\mu^k(C^j) = \lim_{n \to U} (\mu^k_n)(C^j_n)$. Therefore
\[
\int \mu^r(A_{x_1, \ldots, x_k})d\mu^k(x_1, \ldots, x_k) = \sum_{j \leq J} \int_{C_j} \mu^r(A_{x_1, \ldots, x_k})d\mu^k(x_1, \ldots, x_k)
\]
and so
\[
\left| \int \mu^r(A_{x_1, \ldots, x_k})d\mu^k(x_1, \ldots, x_k) - \sum_{j \leq J} \mu^k(C^j)(j + 1/2)\epsilon \right| < \epsilon/2.
\]

Putting this together,
\[
\mu^{k+r}(A) = \lim_{n \to U} (\mu^{k+r}_n)(A_n)
\]
\[
\approx \lim_{n \to U} \sum_{j \leq J} (\mu^k_n)(C^j_n)(j + 1/2)\epsilon
\]
\[
= \sum_{j \leq J} \int_{C_j} \mu^r(A_{x_1, \ldots, x_k})d\mu^k(x_1, \ldots, x_k)
\]
\[
= \int \mu^r(A_{x_1, \ldots, x_k})d\mu^k(x_1, \ldots, x_k)
\]
where the $\approx$ indicates an error of size $<\epsilon/2$. Since we can make $\epsilon$ 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 $\mu_{k+r}$-approximation of $X$. We will show that, for almost every $x_1,\ldots,x_k$, $A_{x_1,\ldots,x_k}$ is a $\mu_r$-approximation of $X_{x_1,\ldots,x_k}$.

Let $D \subseteq V^k_d$ consist of the “defective” points—those points for which $A_{x_1,\ldots,x_k}$ is not a $\mu_r$-approximation of $X_{x_1,\ldots,x_k}$. Then $A_{x_1,\ldots,x_k} \triangle X_{x_1,\ldots,x_k}$ is not $\mu_r$-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 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 4.15, $\mu(D_d) = 0$.

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

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

So the spaces $([V_n]_d^K, \mathcal{B}(\mu_k), \mu_k)$ have the relationships we expect. In particular, we can calculate measures by integrating over one of the coordinates. For example, let $T_{C_3} = \{(x, y, z) \mid \{x, y\} \in E$ and $\{x, z\} \in E$ 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 above justifies the definition of a *measurable graph*. Ultraproducts are our main example, and whenever we see a measurable graph, an ultraproduct will not be far behind, but is convenient to sometimes forget about the extra structure of an ultraproduct.

**Definition 4.22.** A *measurable graph* is a Keisler graded probability space \( \{(V^k, \mathcal{B}_k, \mu_k)\}_{k \in \mathbb{N}} \) together with a symmetric set \( E \in \mathcal{B}_2 \). We say a measurable graph is *atomless* if, for all \( v \in V \), \( \mu_1(\{v\}) = 0 \).

We often say “\( G = (V, E, \mu_1) \) is a measurable graph”, leaving the spaces \( \mathcal{B}_k \) and measures \( \mu_k \) 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 \( \mu_k \) 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 4.23.** 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 \( (x_1, \ldots, x_k) \in V^k_1 \) 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 \( G = [G_n]_U \) is an ultraproduct, \( T_H([G_n]_U) \) is certainly internal: it is the set \( \bigcap \{v_i, v_j\} \in E \{\{x_1, \ldots, x_k\} \mid \{x_i, x_j\} \in [E_n]_U\} \). By the Lős Theorem, \( T_H = [T_H(G_n)]_U \), where \( T_H(G_n) \) is the set of copies of \( H \) in \( G_n \).

The definition of the Keisler graded probability space ensures that each of the sets \( \{x_1, \ldots, x_k\} \mid \{x_i, x_j\} \in [E_n]_U \} \) is measurable, and therefore their intersection is as well.

**Definition 4.24.** We define \( t_H(G) \), the subgraph density of \( H \) in \( G \), to be \( \mu(T_H) \).

When \( G = [G_n]_U \), since the \( T_H \) are internal, we have \( t_H([G_n]_U) = \lim_{n \to U} \mu_n(T_H(G_n)) = \lim_{n \to U} t_H(G_n) \).
The quantity $t_{K_2}(E)$ has a particular significance, since it is essentially the density of $E$ itself. Indeed, by definition,

$$t_{K_2}(E) = \mu_2(\{(x, y) \mid \{x, y\} \in E\}).$$

We can generalize $t_H$ to symmetric functions, which we can think of as weighted graphs:

**Definition 4.25.** 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 4.26.** 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

$$T_H(G, x_1, \ldots, x_d) = \{(x_{d+1}, \ldots, x_k) \in V^{k-d} \mid \pi_{x_1,\ldots,x_k} is an actual copy of \ H\}$$

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 $H' = (W_0, F \upharpoonright \binom{W_0}{2})$ 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: $t_{K_2}(G, x) = \int \chi_E(x, y) \, d\mu_1(y) = \deg_G(x)$. More generally, this lets us count things like

$$t_{K_3}(G, x) = \int \chi_E(x, y) \chi_E(x, z) \chi_E(y, z) \, d\mu_2(z),$$

the number of pairs $\{y, z\}$ which form a triangle with $x$, and

$$t_{K_3}(G, x, y) = \int \chi_E(x, z) \chi_E(y, z) \, d\mu_1(z),$$
the number of ways to extend a pair \( x, y \) to a triangle. The Fubini property promises that
\[
\int t_{K_3}(G, x) \, d\mu_1 = t_{K_3}(G)
\]
—that is, if we add up, over all vertices, how many extensions there are to a triangle, we get the number of triangles—and
\[
\int \chi_E(x, y) t_{K_3}(G, x, y) \, d\mu_2 = t_{K_3}(G)
\]
—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 \( \epsilon \)-quasirandom) measurable graphs.

Definition 4.27. \( G = (V, E, \mu_1) \) is quasirandom if
\[
t_{C_4}(G) = \left( t_{K_2}(G) \right)^4.
\]
The proof that \( t_{C_4}(G) \geq (t_{K_2}(G))^4 \) for any graph \( G \) still applies, so again quasirandomness means that \( t_{C_4}(G) \) is as small as possible: the value of \( t_{K_2}(G) \) forces there to be a certain number of cycles of length 4, and a quasirandom graph has only as many as it has to.

Theorem 4.28. \([G_n]_U \) is quasirandom if and only if, for every \( \epsilon > 0 \), \( \{ n \mid G_n \text{ is } \epsilon \text{-quasirandom} \} \in \mathcal{U} \).

Proof. We have \( t_{C_4}([G_n]_U) = \lim_{n \to U} t_{C_4}(G_n) \) and \( (t_{K_2}([G_n]_U))^4 = \lim_{n \to U} (t_{K_2}(G_n))^4 \).

If, for every \( \epsilon > 0 \), \( \{ n \mid G_n \text{ is } \epsilon \text{-quasirandom} \} \in \mathcal{U} \) then, for every \( \epsilon > 0 \),
\[
t_{C_4}([G_n]_U) = \lim_{n \to U} t_{C_4}(G_n) \leq \lim_{n \to U} (t_{K_2}(G_n))^4 + \epsilon = (t_{K_2}([G_n]_U))^4 + \epsilon,
\]
so \( t_{C_4}([G_n]_U) = (t_{K_2}([G_n]_U))^4 \).

If there is an \( \epsilon > 0 \) such that \( \{ n \mid G_n \text{ is } \epsilon \text{-quasirandom} \} \notin \mathcal{U} \) then
\[
t_{C_4}([G_n]_U) = \lim_{n \to U} t_{C_4}(G_n) \geq \lim_{n \to U} (t_{K_2}(G_n))^4 + \epsilon = (t_{K_2}([G_n]_U))^4 + \epsilon.
\]

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.
4.6. QUASIRANDOMNESS

Theorem 4.29. If $G = (V, E, \mu_1)$ is quasirandom then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $B_1$,

$$\mu_2(E \cap (X \times Y)) = \mu_2(E)\mu_1(X)\mu_1(Y).$$

The proof we gave of the analogous theorem in the finite setting, Theorem 2.28, goes through essentially unchanged. Rather than repeat that here, we will wait for Corollary 5.19 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 4.30. If $G = (V, E, \mu_1)$ is quasirandom then for every finite graph $H = (W, F)$, $t_H(G) = p^{|F|}$ where $p = t_{K_2}(G) = \mu_2(E)$.

Proof. First, consider the where $H = C_3$, the triangle, to get the main idea. In this case

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

That is, $t_{C_3}(G)$ is the probability that, when we select three random vertices $x, y, z$, 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 4.29—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 $\bullet \bullet$, 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 4.29 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

$$t_{C_3}(G_U) = p^2 \int \chi_E(z, x) \, d\mu_3 = p^3.$$

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

$$t_H(G) = \mu_{|W|}(\{(x_1, \ldots, x_{|W|}) \mid \text{for each } \{v_i, v_j\} \in F, \{x_i, x_j\} \in E_U\})$$

$$= \int \prod_{\{v_i, v_j\} \in F} \chi_E(x_i, x_j) \, d\mu_{|W|}(x_1, \ldots, x_{|W|}).$$

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.
4.7. CONSEQUENCES FOR FINITE GRAPHS

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

4.7 Consequences for Finite Graphs

We can use Theorem 4.30 to give a proof of Theorem 2.32. The technique is typical of the way we obtain results about finite graphs using ultraproducts.
\textbf{Theorem 2.32}. For every finite graph \(H = (W, F)\), each \(\varepsilon > 0\), there is a \(\delta > 0\) so that if \(G = (V, E)\) is \(\delta\)-quasirandom and \(V\) is sufficiently large, \(|t_H(G) - t_{K_2}(G)|^F| < \varepsilon\).

\textit{Proof.} Suppose not. Then there is a finite graph \(H = (W, F)\) and an \(\varepsilon > 0\) 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

\[|t_H(G_n) - (t_{K_2}(G_n))^F| \geq \varepsilon.\]

Let \([G_n]_U\) be an ultraproduct of the sequence \(\langle G_n \rangle_{n \in \mathbb{N}}\). Let \(p = \lim_{n \to U} t_{K_2}(G_n) = t_{K_2}([G_n]_U)\). Since \(G_n\) is \(1/n\)-quasirandom, \(t_{C_4}([G_n]_U) = \lim_{n \to U} t_{C_4}(G_n) = p^4\), so \([G_n]_U\) is quasirandom.

By Theorem 4.30, \(\lim_{n \to U} t_H(G_n) = t_H([G_n]_U) = p^F\). Therefore, taking \(\delta\) small enough,

\[\{n \mid |t_{K_2}(G_n) - p| < \delta\} \cap \{n \mid |t_H(G_n) - p^F| < \varepsilon/2\} \in U.\]

But consider some \(n\) in both of these sets: we have

\[|t_H(G_n) - (t_{K_2}(G_n))^F| \leq |p^F - (t_{K_2}(G_n))^F| + |t_H(G_n) - p^F| \leq \varepsilon/2 + \varepsilon/2 < \varepsilon,\]

which is a contradiction. \(\square\)

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 4.30 and translate them, step by step, into arguments in the finite setting. The difficulty is that the arguments would become peppered with additional \(\varepsilon\)'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 \(\varepsilon\)-quasirandomness ensures \(|\mu_2([X \times Y] \cap E) - p\mu(X)\mu(Y)| < \delta\) for some \(\delta\). The main service of the ultraproduct in this argument is hiding all those \(\varepsilon\)'s and \(\delta\)'s by letting them reach the limit value of 0.

\section*{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 atomless measurable graph resembles an ultraproduct.
There is a canonical way to sample a random finite graph from $G = (V, E, \mu_1)$: choose finitely many vertices $v_1, \ldots, v_n$ according to the measure $\mu_1$ on $V$, and let $G_n = (\{v_1, \ldots, v_n\}, E | \{v_1, \ldots, v_n\})$. Our requirement that $\mu_1(\{v\}) = 0$ for each $v \in V$ means that when $i \neq j$, $P(v_i = v_j) = 0$, so, with probability 1, $G_n$ is a well-defined finite graph whose properties we can consider.

**Lemma 4.31.** Let $G = (V, E)$ be an atomless measurable graph. For any finite graph $H$ and any $\epsilon > 0$, there is an $m$ so that whenever we sample an $n$ vertex graph $G_n$ from $G$ with $n \geq m$,

$$P[|t_H(G) - t_H(G_n)| < \epsilon] > 1 - \epsilon.$$  

**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 $E(t_H(G_n))$. For each $\pi : W \rightarrow \{1, \ldots, n\}$, let $1_{\pi}$ be the random variable which is 1 if $\pi(i) = v_\pi(i)$ is an actual copy of $H$ and 0 otherwise. When $\pi$ is injective, observe that $1_{\pi}$ is simply the probability that, when we select the points $v_{\pi(1)}, \ldots, v_{\pi(k)}$ from $V$, that we get an actual copy of $H$. Therefore

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

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

By the linearity of expectation,

$$E(t_H(G_n)) = \frac{1}{n^{|W|}} \sum_{\pi} E(1_{\pi}).$$

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

$$E(t_H(G_n)) = t_H(G) + O\left(\frac{1}{n}\right).$$

As in the proof of Theorem 2.7 we use McDiarmid’s inequality. This time the random variables are the $v_i$ and $f(v_1, \ldots, v_n)$ 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 $\pi$, and therefore changing a single vertex can change the subgraph density by at most $|W|/n$. Then McDiarmid’s inequality says

$$P(|E(t_H(G_n)) - t_H(G_n)| \geq \epsilon) \leq 2e^{-\frac{2\epsilon^2}{n^{|W|}}} = 2e^{-\frac{2\epsilon^2n}{|W|^2}}.$$
Once we pick $\epsilon$, the value $\frac{2\epsilon^2}{|W|^\alpha}$ is fixed, so by choosing $n$ large enough, we can make this bound as small as we like. \hfill \square

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

**Corollary 4.32.** Let $G = (V, E)$ be an atomless measurable graph. For any finite list of finite graphs $H_1, \ldots, H_k$ and any $\epsilon > 0$, there is an $m$ so that whenever we sample an $n$-vertex graph $G_n$ from $G$ with $n \geq m$, for each $i \leq k$,

$$P[|t_{H_i}(G) - t_{H_i}(G_n)| < \epsilon] > 1 - \epsilon.$$  

**Proof.** Apply the lemma with $\epsilon/k$. \hfill \square

**Corollary 4.33.** Let $G = (V, E)$ be an atomless measurable graph. For each $n$, let $G_n$ be an $n$-vertex graph sampled from $G$ (with the $G_n$ sampled independently from each other). Then, with probability 1, for every finite graph $H$,

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

**Proof.** There are countably many finite graphs, so it suffices to show that, for each one individually, the probability that $\lim_{n \to \infty} t_H(G_n) \neq t_H(G)$ 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 $n \geq m$, $|t_H(G_n) - t_H(G)| < \epsilon$.

For a given $m$, the probability that there is any $n \geq m$ with $|t_H(G_n) - t_H(G)| \geq \epsilon$ is at most

$$\sum_{n \geq m} P(|t_H(G_n) - t_H(G)| \geq \epsilon) \leq \sum_{n \geq m} 2e^{-\frac{\epsilon^2}{|W|^\alpha n}} \leq \int_m^{\infty} 2e^{-\frac{\epsilon^2}{|W|^\alpha x}} dx = \frac{2|W|^\alpha}{\epsilon^2} e^{-\frac{\epsilon^2}{|W|^\alpha m}}.$$  

In particular, choosing $m$ large enough, this probability is smaller than $\epsilon$. \hfill \square

Note that this corollary gives us true limits, not just ultralimits: every ultraproduct of the $G_n$ has the same subgraph densities as $G$.

### 4.9 The Possible Subgraph Densities are Compact

One reason to consider atomless measurable graphs, even if our main interest is finite graphs, is that they compactify the space of subgraph densities: rather than speaking of sequences of graphs with subgraph densities approaching some limiting values of interest, we can focus on a single measurable graph achieving these limiting values.
4.9. THE POSSIBLE SUBGRAPH DENSITIES ARE COMPACT

Theorem 4.34. Let $\zeta$ be a function assigning, to each finite graph $H$, a density $\zeta(H)$. Suppose that for every finite list of finite graphs $H_1, \ldots, H_k$ and every $\epsilon > 0$, there is a finite or atomless measurable graph $G$ so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G)| < \epsilon$.

Then there is a finite or atomless measurable graph $G$ such that $t_H(G) = \zeta(H)$ for all finite graphs $H$.

A more formal approach to this idea, which we do not investigate in detail because we will not otherwise need it, is to call $\zeta$ a possible subgraph density if there exists some finite or atomless measurable graph $G$ so that, for every $H$, $\zeta(H) = t_H(G)$. There is a natural topology on the space of possible subgraph densities—take a basic open set to be $\{\zeta \mid \zeta(H) \in I\}$ for an open interval $I$. (This topology is even metrizable—if we put the finite graphs in an order $H_1, H_2, \ldots$, we can define $d(\zeta, \zeta') = \sum_i \frac{|\zeta(H_i) - \zeta'(H_i)|}{2^i}$.) Then the theorem says that the space of possible subgraph densities are compact.

Proof. One way to prove this is to drop through the world of finite graphs. Fix an ordering of the finite graphs, $H_1, H_2, \ldots$. For each $k$, choose a finite or atomless measurable graph $G'_{k}$ so that, for all $i \leq k$, $|\zeta(H_i) - t_{H_i}(G'_{k})| < 1/2k$. If $G'_{k}$ is finite, let $G_k = G'_{k}$. If $G'_{k}$ is atomless, by Corollary 4.32 we may sample a graph $G_k$ from $G'_{k}$ with $|t_{H_i}(G_k) - t_{H_i}(G'_{k})| < 1/2k$ for each $i \leq k$, so $|\zeta(H_i) - t_{H_i}(G_k)| < 1/k$.

Then for each $H_i$, we have

$$t_{H_i}([G_k]|_{U}) = \lim_{k \to U} t_{H_i}(G_k) = \zeta(H_i).$$

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

Theorem 4.35. Let $\zeta$ 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 $H \in \mathcal{X}$, $T_H(\mathcal{X}) = \emptyset$.

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 $H \in \mathcal{X}$. 

CHAPTER 4. INTEGRALS IN ULTRAPRODUCTS

Proof. This follows by the same proof, noting that when we sample $G_n$ from $G$, if $T_H(G) = \emptyset$ then $T_H(G_n) = \emptyset$ (because $G_n$ 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 Los’ Theorem).

4.10 Shifting Measures

We want to consider what happens if we take a measurable graph $G = (V, E, \mu_1)$ and modify $\mu_1$ 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 4.36. If $\mu$ and $\nu$ are both measures on $\mathcal{B}$, we say $\nu$ is absolutely continuous with respect to $\mu$, 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 Theorem). If $\nu \ll \mu$ then there is a measurable function $f$, called the Radon-Nikodym derivative of $\nu$, such that, for every measurable $B \in \mathcal{B}$,

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

**Lemma 4.37.** 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 $\nu$. 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 \( \nu_k \) follows by the definition and the symmetry of \( \mu_k \).

The Fubini Property holds since

\[
\nu_{k+r}(X) = \int \chi_X(x_1, \ldots, x_{k+r}) \prod_{i \leq k+r} f(x_i) \, d\mu_{k+r} \\
= \int \left[ \int \chi_X(x_1, \ldots, x_{k+r}) \prod_{k < i \leq k+r} f(x_i) \, d\mu_r(x_{k+1}, \ldots, x_{k+r}) \right] \prod_{i \leq k} f(x_i) \, d\mu_k(x_1, \ldots, x_k) \\
= \int \nu_r(X_1, \ldots, x_k) \prod_{i \leq k} f(x_i) \, d\mu_k \\
= \int \nu_r(X_1, \ldots, x_k) \, d\nu_k.
\]

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

\[
\nu'_{k+1}(B) = \int \nu'_k(B_{x_1}) \, d\nu'_1(x_1) \\
= \int \nu_k(B_{x_1}) \, d\nu_1(x_1) \\
= \nu_{k+1}(B).
\]

In light of this lemma, when we have a measurable graph \( G \) with measure \( \mu \) 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 4.38.** For any \( t \geq 3 \), if \( G \) is a measurable graph with \( t_{K_2}(G) > 1 - \frac{1}{t-1} \) then \( |T_{K_t}(G)| > 0 \).

**Proof.** Suppose not, so there is a measurable graph \( G = (V, E, \mu) \) with \( t_{K_2}(G) > 1 - \frac{1}{t-1} \) and \( T_{K_t}(G) = \emptyset \). By Theorem 4.35 there is a measurable graph \( G \) with \( T_{K_2}(G) = \emptyset \) maximizing the value \( t_{K_2}(G) \) among all such measurable graphs.
Claim 1. For almost every $x$, $\deg_G(x) = t_{K_2}(G)$.

Proof. Suppose not. The idea is that we will pick a set $B$ of vertices whose degree is too high and tweak the measure $\mu_1$ 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 $c = t_{K_2}(G)$. Since $G$ is not evenly distributed, choose some $\epsilon > 0$ so that the set of $x$ with $\deg_G(x) - t_{K_2}(G) > \epsilon$ has positive measure, and let $B$ be the set of such $x$. Let $d = \mu_1(B)$.

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

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

Then

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

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

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

$$t_{K_2}(G_\delta) = \int \chi_E(x, y) \ d(\nu_\delta)_2$$

$$= \frac{1}{(1 + \delta)^2} \left[ \int \chi_E(x, y) \ d\mu_2 + \frac{\delta}{d} \int_{B \times V} \chi_E(x, y) \ d\mu_1(x) \ d\nu_\delta(y) \right]$$

$$= \frac{1}{(1 + \delta)^2} \left[ \int \chi_E(x, y) \ d\mu_2 + 2\frac{\delta}{d} \int_{B \times V} \chi_E(x, y) \ d\mu_2 + \frac{\delta^2}{d^2} \int_{B \times B} \chi_E(x, y) \ d\mu_2 \right]$$

$$= \frac{1}{(1 + \delta)^2} \left[ c + 2\delta(c + \epsilon) + \delta^2 \frac{\int_{B \times B} \chi_E(x, y) \ d\mu_2}{d^2} \right].$$

In particular,

$$\lim_{\delta \to 0} \frac{t_{K_2}(G_\delta) - t_{K_2}(G)}{\delta} = \lim_{\delta \to 0} \frac{c + 2\delta(c + \epsilon) + \delta^2 \frac{\int_{B \times B} \chi_E(x, y) \ d\mu_2}{d^2} - c(1 + \delta)^2}{\delta(1 + \delta)^2}$$

$$= \lim_{\delta \to 0} \frac{2\epsilon + \delta \int_{B \times B} \chi_E(x, y) \ d\mu_2}{(1 + \delta)^2} - c\delta$$

$$= 2\epsilon.$$
4.12. MINIMIZING DENSITY

so when δ is sufficiently small, we must have $t_{K_2}(G_δ) - t_{K_2}(G) > 0$. But $T_{K_t}(G_δ) = T_{K_t}(G) = ∅$, contradicting the maximality of $t_{K_2}(G)$ 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 $i \leq m$, $\{y \in V' \mid \{x_i, y\} \notin E\}$ 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 $> 1 - \frac{m}{t-1} \geq 0$, and in particular is non-empty, so we may choose $x_m+1 \in V'$ so that $x_1, \ldots, x_m, x_{m+1}$ is a copy of $K_{m+1}$. This gives us a copy of $K_t$, showing that $|T_{K_t}(G)| > 0$.

This immediately gives a finite version.

**Corollary 4.39.** 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 $t_{K_2}(G) > 1 - \frac{1}{t-1} + \epsilon$ then $|T_{K_t}(G)| > 0$.

**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 $t_{K_2}(G_n) > 1 - \frac{1}{t-1} + \epsilon$ and $|T_{K_t}(G_n)| = 0$.

Then take any ultrafilter $U$ and consider $G = [G_n]_U$: $t_{K_2}(G) = \lim_{n \to U} t_{K_2}(G_n) \geq 1 - \frac{1}{t-1} + \epsilon$ while $T_{K_t}(G) = ∅$ 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 ultraproduct 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 $t_{K_2}(G) = 1 - \frac{1}{t-1}$ while having no copies of $K_t$.

### 4.12 Minimizing Density

More generally, if $t_{K_2}(G) > 1 - \frac{1}{t-1}$, we could ask what the smallest possible value of $t_{K_t}(G)$ is.

For triangles, we have the following bound.

**Lemma 4.40 (Goodman’s Bound).** When $G$ is a measurable graph,

$$t_{K_3}(G) \geq t_{K_2}(G) \cdot (2t_{K_2}(G) - 1).$$
The proof we will give here—like most of the results in this section—is due to Razborov \cite{82}. 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

\[
\begin{align*}
t_{K_3} + t_{K_2} &= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \, d\mu_3 + \int \chi_E(x,y) \, d\mu_2 \\
&= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \, d\mu_3 + \int \chi_E(x,y)(1 - \chi_E(x,z)) \, d\mu_3 \\
&\quad + \int \chi_E(x,y)\chi_E(x,z) \, d\mu_3 \\
&= \int \chi_E(x,y)\chi_E(x,z)\chi_E(y,z) \, d\mu_3 + \int \chi_E(x,y)(1 - \chi_E(x,z))(1 - \chi_E(y,z)) \, d\mu_3 + \int \chi_E(x,y)\chi_E(x,z) \, d\mu_3 \\
&\quad + \int \chi_E(x,y)\chi_E(x,z) \, d\mu_3 \\
&= \int \chi_E(x,y)(1 - \chi_E(x,z))(1 - \chi_E(y,z)) \, d\mu_3 + 2 \int \chi_E(x,y)\chi_E(x,z) \, d\mu_3 \\
&\geq 2 \left( \int \chi_E(x,y) \, d\mu_1 \right)^2 \, d\mu_1(x) \\
&\geq 2\left( \int \chi_E(x,y) \, d\mu_2 \right)^2 \\
&= 2(t_{K_2}(G))^2.
\end{align*}
\]

Rearranging, we have

\[
t_{K_3} = 2(t_{K_2}(G))^2 - t_{K_2} = t_{K_2}(G)(2t_{K_2}(G) - 1).
\]

This gives a lower bound: a graph with \(t_{K_2}(G)\) edges needs at least \(t_{K_3}(G)(2t_{K_2}(G) - 1)\) triangles.

Consider the evenly \(t\)-partite graphs from the previous section: the graph \(P_t\) has \(t_{K_2}(P_t) = 1 - \frac{1}{t}\) and \(t_{K_3}(P_t) = (1 - \frac{1}{t})(1 - \frac{2}{t})\). Noticing that

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

we have

\[t_{K_3}(P_t) = 2(1 - \frac{1}{t})^2 - (1 - \frac{1}{t}) = \frac{1}{t}.\]
4.12. MINIMIZING DENSITY

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 \( c = t_{K_2}(G) \) 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 4.41.** \( g_r(x) = t_{K_r}(P_{t,a}) \) where \( t, a \) are chosen so that \( t_{K_2}(P_{t,a}) = x \).

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 4.42.** For every \( r \) and every \( x \), the graph \( P_{t,a} \) with \( t, a \) chosen so that \( t_{K_2}(P_{t,a}) = x \) minimizes \( t_{K_r}(G) \) among all graphs with \( t_{K_2}(G) = x \).

The case for triangles, where \( r = 3 \) is shown in [83], and the case where \( r = 4 \) in [78].

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 \( g_3 \). Given \( t, a \), we have

\[
t_{K_2}(P_{t,a}) = a(1-a) + (1-a)(a + \frac{(1-a)}{t}(t-1)) = -a^2(1 + \frac{1}{t}) + 2a\frac{1}{t} + 1 - \frac{1}{t}.
\]

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
\[
a = \frac{2t \pm \sqrt{4t^2 + 4(1 - \frac{1}{t} - x)(1 + \frac{1}{t})}}{2(1 + 1/t)} = \frac{2 \pm \sqrt{4 + 4(t - 1 - tx)(1 + t)}}{2t + 2} = \frac{1 \pm \sqrt{t(t - (t + 1)x)}}{t + 1}.
\]

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 \(t_{k_2}(P_{t,a}) = x\), we have
\[
t_{k_3}(P_{t,a}) = 3a(1 - a)2t - 1 \frac{t - 1}{t} + (1 - a)3t - 1 t - 2 \frac{t}{t}
\]
\[
= \left( t + \sqrt{t(t - (t + 1)x)} \right)^2 \left[ \frac{3 - \sqrt{t(t - (t + 1)x) t - 1}}{t + 1} + \frac{t + \sqrt{t(t - (t + 1)x)} (t - 1)(t - 1)}{t + 1} \frac{t^2}{t^2} \right]
\]
\[
= \frac{(t - 1)(t + \sqrt{t(t - (t + 1)x)})^2}{t^2(t + 1)^2} \left[ \frac{3t - 3t \sqrt{t(t - (t + 1)x)} + (t - 2)(t + \sqrt{t(t - (t + 1)x)})}{t + 1} \frac{t^2}{t} \right]
\]
\[
= \frac{(t - 1)(t + \sqrt{t(t - (t + 1)x)})^2}{t^2(t + 1)^2} \left[ \frac{t + t^2 - 2t \sqrt{t(t - (t + 1)x)} + \sqrt{t(t - (t + 1)x)}}{t + 1} \frac{t + 1}{t + 1} \right]
\]
\[
\quad = \frac{(t - 1)(t + \sqrt{t(t - (t + 1)x)})^2}{t^2(t + 1)^2} (t - 2 \sqrt{t(t - (t + 1)x)}).
\]

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
\[
g_3(x) = \frac{(t - 1)(t + \sqrt{t(t - (t + 1)x)})^2}{t^2(t + 1)^2} (t - 2 \sqrt{t(t - (t + 1)x)}).
\]

Then the example of the graph \(P_{t,a}\) shows that it is possible to have \(t_{k_2}(G) = x\) and \(t_{k_3}(G) = g_3(x)\).

**Lemma 4.43.** Let \(G = (V, E, \mu_1)\) be a measurable graph with \(t_{k_1}(G) = c\), let \(B \subseteq V\), and let \(u = \int_B t_{k_1}(G, x) d\mu_1\). Let \(\nu_\delta\) be the measure
\[
\nu_\delta(S) = \frac{1}{1 + \delta} \left[ \mu(S \setminus B) + (1 + \frac{\delta}{\mu(B)}) \mu(S \cap B) \right]
\]
and let \( G_\delta = (V, E, \nu_\delta) \). Let \( h \) be the function with \( h(\delta) = t_{K_1}(G_\delta) \). Then \( h'(0) = t(u - c) \).

**Proof.** Observe that

\[
t_{K_1}(G_\delta) = \int \prod_{1 \leq i < j \leq t} \chi_E(x_i, x_j) d(\nu_\delta)
\]

\[
= \frac{1}{(1 + \delta)^t} \left[ \int \prod_{1 \leq i < j \leq t} \chi_E(x_i, x_j) d\mu + t \frac{\delta}{\mu(B)} \int \chi_B(x_1) \prod_{1 \leq i < j \leq t} \chi_E(x_i, x_j) d\mu + \cdots \right.
\]

\[
+ \frac{\delta^t}{(\mu(B))^t} \int \prod_{i \leq t} \chi_B(x_i) \prod_{1 \leq i < j \leq t} \chi_E(x_i, x_j) d\mu \right].
\]

In particular, since \( h(0) = c = t_{K_1}(G) \),

\[
h'(0) = \lim_{\delta \to 0} \frac{t_{K_1}(G_\delta) - t_{K_1}(G)}{\delta}
\]

\[
= \lim_{\delta \to 0} \frac{c + tu\delta + \delta^2(\cdots) - c(1 + \delta)^t}{\delta(1 + \delta)^t}
\]

\[
= t(u - c).
\]

\( \square \)

**Theorem 4.44.** For any \( x \in (1/2, 2/3) \) and any atomless measurable graph \( G \), if \( t_{K_2}(G) = x \) then \( t_{K_3}(G) \geq g_3(x) \).

**Proof.** Suppose not, so there is an atomless measurable graph \( G \) with \( t_{K_2}(G) \in (1/2, 2/3) \) and \( g_3(t_{K_2}(G)) - t_{K_3}(G) > 0 \). By Theorem 4.34, we may choose a measurable graph \( G \) maximizing \( g_3(t_{K_2}(G)) - t_{K_3}(G) \). Let \( c = t_{K_2}(G) \) and \( d = t_{K_3}(G) \).

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,

\[
2g_3'(c)t_{K_2}(G, x) - 3t_{K_3}(G, x) = 2g_3'(c)c - 3d.
\]
we have no way to make more copies and thereby do better.

\[ B^+ = \{ x \mid 2g_3'(c)t_{K_2}(G, x) - 3t_{K_3}(G, x) - (2g_3'(c)c - 3d) \geq \epsilon \} \]

or

\[ B^- = \{ x \mid 2g_3'(c)t_{K_2}(G, x) - 3t_{K_3}(G, x) - (2g_3'(c)c - 3d) \leq -\epsilon \} \]

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

We will apply Lemma 4.43. Let \( u_2 = \int_{B^+} t_{K_2}(G, x) \, d\mu_1 \) and \( u_3 = \int_{B^+} t_{K_3}(G, x) \, d\mu_1 \). By assumption,

\[ 2g_3'(c)u_2 - 3u_3 - (2g_3'(c)c - 3d) = \int_{B^+} 2g_3'(c)t_{K_2}(G, x) - 3t_{K_3}(G, x) - (2g_3'(c)c - 3d) \, d\mu_1 \geq \epsilon \mu(B^+). \]

Letting \( \nu_\delta \) and \( G_\delta \) be as in the statement of Lemma 4.43, we take \( h_2(\delta) = t_{K_2}(G_\delta) \) and \( h_3(\delta) = t_{K_3}(G_\delta) \). Consider the function \( p(\delta) = g_3(t_{K_2}(G_\delta)) - t_{K_3}(G_\delta) \). Then

\[ p'(0) = g_3'(c)h_2'(0) - h_3'(0) = 2g_3'(c)(u_2 - c) - 3(u_3 - d) = 2g_3'(c)u_2 - 3u_3 - (2g_3'(c)c - 3d) \geq \epsilon \mu(B^+) > 0. \]

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

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 \( \mu_2(E), 3t_{K_3}(G, x, y) \leq g_3'(c). \)

**Proof.** Suppose not. Then there is an \( \epsilon > 0 \) so that, letting \( B = \{(x, y) \in E \mid 3t_{K_3}(G, x, y) > g_3'(c) + \epsilon \}, \mu_2(B) > 0 \). 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 \( t_{K_2}(G_\delta) = c - \delta \) and

\[
t_{K_3}(G_\delta) = \int \chi_{E \setminus B_\delta}(x, y)\chi_{E \setminus B_\delta}(x, z)\chi_{E \setminus B_\delta}(y, z) \, d\mu_3
= \int \chi_E(x, y)\chi_E(x, z)\chi_E(y, z) \, d\mu_3 - 3 \int \chi_{B_\delta}(x, y)\chi_E(x, z)\chi_E(y, z) \, d\mu_3
+ 3 \int \chi_{B_\delta}(x, y)\chi_{B_\delta}(x, z)\chi_E(y, z) \, d\mu_3 - \int \chi_{B_\delta}(x, y)\chi_{B_\delta}(x, z)\chi_{B_\delta}(y, z) \, d\mu_3.
\]
4.12. MINIMIZING DENSITY

In particular,

\[
\limsup_{\delta \to 0^+} \frac{t_{K_3}(G_\delta) - d}{\delta} = \frac{1}{\delta} \left( -3 \int \chi_{B_\delta}(x, y) \chi_E(x, z) \chi_E(y, z) \, d\mu_3 \\
+ 3 \int \chi_{B_\delta}(x, y) \chi_{B_\delta}(x, z) \chi_E(y, z) 
- \int \chi_{B_\delta}(x, y) \chi_{B_\delta}(x, z) \chi_{B_\delta}(y, z) \, d\mu_3 \right) < -g'_3(c) - \epsilon.
\]

Letting \( p(\delta) = g_3(t_{K_2}(G_\delta)) - t_{K_3}(G_\delta) \), we have

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

Then there is a small value of \( \delta \) so that \( g_3(t_{K_3}(G_\delta)) - t_{K_3}(G_\delta) > g_3(c) - d \), contradicting the maximality of \( G \).

Next, we need an inequality:

**Claim 4.**

\[
3 \int \chi_E(x, y) t_{K_3}(G, x) \, d\mu_2 + 3 \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) t_{K_3}(G, x, y) \, d\mu_3 \geq 2d.
\]

**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, \( P_1 = T_{K_4}(G) \).) Then \( 24d = 24\mu_4(P_1) + 12\mu_4(P_2) + 6\mu_4(P_3) + 6\mu_4(P_4) \).

Similarly,

\[
24 \int \chi_E(x, y) t_{K_3}(G, x) \, d\mu_2 = 24\mu_4(P_1) + 16\mu_4(P_2) + 2\mu_4(P_3)
\]

and

\[
24 \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) \chi_E(x, y) t_{K_3}(G, x, y) \, d\mu_3 = 2\mu_4(P_3) + 6\mu_3(P_4),
\]
Finally, we notice that

\[
3 \int \chi_E(x, y) t_{K_3}(G, x) \, d\mu_2 + 3 \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) \chi_E(x, y) t_{K_3}(G, x, y) \, d\mu_3
\]

\[
= 72 \mu_4(P_1) + 48 \mu_4(P_2) + 12 \mu_4(P_3) + 18 \mu_4(P_4)
\]

\[
\geq 48 \mu_4(P_1) + 24 \mu_4(P_2) + 12 \mu_4(P_3) + 12 \mu_4(P_4)
\]

\[
= d.
\]

Observe that the first claim means that

\[
\int \chi_E(x, y)(2g_3'(c)t_{K_3}(G, x) - 3t_{K_3}(G, x)) \, d\mu_2 = c(2g_3(c) - 3d)
\]

and the second claim means that

\[
\int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) \chi_E(x, y) t_{K_3}(G, x, y) \, d\mu_3 \leq \frac{g_3'(c)}{3} \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) \chi_E(x, y)
\]

Therefore

\[
2d \leq 3 \int \chi_E(x, y) t_{K_3}(G, x) \, d\mu_2 + 3 \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) t_{K_3}(G, x, y) \, d\mu_3
\]

\[
\leq -c(2g_3(c) - 3d) + 2g_3'(c) \int \chi_E(x, y) \chi_E(x, z) \, d\mu_3 + g_3'(c) \int (1 - \chi_E(x, z))(1 - \chi_E(y, z)) \chi_E(x, y)
\]

\[
- c(2g_3(c) - 3d) + g_3'(c) \int \chi_E(x, y) \chi_E(x, z) \, d\mu_3 + g_3'(c) \int (1 - \chi_E(x, z)) \chi_E(x, y) \, d\mu_3
\]

\[
= -c(2g_3(c) - 3d) + g_3'(c) \int \chi_E(x, y) \chi_E(x, z) \, d\mu_3 + g_3'(c) \int \chi_E(x, y) \, d\mu_2
\]

\[
= -2g_3'(c)c^2 + g_3'(c)c + 3cd + g_3'(c)d.
\]

Reorganizing gives us

\[
g_3'(c)c(2c - 1) \leq d(g_3'(c) + 3c - 2).
\]

When \( x \in [1/2, 2/3] \), \( g_3'(x) \geq 1 \), so \( g_3'(c) + 3c - 2 > 0 \) and therefore

\[
d \geq \frac{g_3'(c)(2c - 1)}{g_3'(c) + 3c - 2}.
\]

Finally, we notice that \( g_3 \) satisfies the differential equation

\[
x(2x - 1)g_3'(x) = g_3(x)(g_3'(x) + 3x - 2),
\]
Then let $\phi$ be a counterexample and, for each $\delta > 0$ so that whenever $G$ has at least $n$ vertices and $|t_{K_3}(G) - x| < \delta$, $t_{K_3}(G) \geq g_3(x) - \epsilon$. But this contradicts the theorem.

**Corollary 4.45.** For each $x \in (1/2, 2/3)$ and each $\epsilon > 0$, there is an $n$ and $\delta > 0$ so that whenever $G$ has at least $n$ vertices and $|t_{K_3}(G) - x| < \delta$, $t_{K_3}(G) \geq g_3(x) - \epsilon$.

**Proof.** Suppose not. Let $x$ and $\epsilon$ be a counterexample and, for each $n$, choose $G_n$ with at least $n$ vertices so that $|t_{K_3}(G_n) - x| < 1/n$ but $t_{K_3}(G_n) \geq g_3(x) - \epsilon$, and take any ultrafilter $\mathcal{U}$.

Then $t_{K_3}([G_n]_\mathcal{U}) = \lim_{n \to \mathcal{U}} t_{K_3}(G_n) = x$ and $t_{K_3}([G_n]_\mathcal{U}) = \lim_{n \to \mathcal{U}} t_{K_3}(G_n) \geq g_3(x) - \epsilon$. 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 4.46.** Suppose that, for every pair of natural numbers $y$ and $z$, $\phi_{y,z}(x_1, \ldots, x_m)$ is a first-order formula. Choose parameters $[b^1_n]_\mathcal{U}, \ldots, [b^{m-k}_n]_\mathcal{U}$, and let

$$X = \{([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}) \mid \forall y \exists z \phi_{y,z}([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}, [b^1_n]_\mathcal{U}, \ldots, [b^{m-k}_n]_\mathcal{U})\}.$$

Then $([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}) \in X$ if and only if, for every $y$, there is a $z$ so that $\{n \mid \phi_{y,z}(a^1_n, \ldots, a^k_n, b^1_n, \ldots, b^{m-k}_n)\} \notin \mathcal{U}$.

In particular, note that $X$ is not (typically) internal: in order to belong to $X$, we need the choice of $z$ to be independent of $n$.

**Proof.** Consider some $([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}) \in [V^k_n]_\mathcal{U}$. Suppose $([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}) \notin X$, so there is some $y$ so that, for each $z$, $\phi_{y,z}([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}, [b^1_n]_\mathcal{U}, \ldots, [b^{m-k}_n]_\mathcal{U})$ is false. Then, by Łoś’s Theorem, for each $z$, $\{n \mid \phi_{y,z}(a^1_n, \ldots, a^k_n, b^1_n, \ldots, b^{m-k}_n)\} \notin \mathcal{U}$.

Conversely, suppose there is some $y$ so that, for every $z$, $\{n \mid \phi_{y,z}(a^1_n, \ldots, a^k_n, b^1_n, \ldots, b^{m-k}_n)\} \notin \mathcal{U}$. Then, by Łoś’s Theorem again, $\phi_{y,z}([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}, [b^1_n]_\mathcal{U}, \ldots, [b^{m-k}_n]_\mathcal{U})$ is false for each $z$, so $([a^1_n]_\mathcal{U}, \ldots, [a^k_n]_\mathcal{U}) \notin X$. 

\[ \square \]
We are especially interested in the case where $m = 0$.

**Corollary 4.47** (Transfer). Suppose that for every pair of natural numbers $y$ and $z$, $\sigma_{y,z}$ is a first-order sentence. Then the following are equivalent:

- in $[G_n]_U$, for every $y$ there is a $z$ so that $\sigma_{y,z}$ is true,

- for every $y$ there is a $z$ so that $\{n \mid \sigma_{y,z} \text{ is true in } G_n\} \in U$.

**Corollary 4.48.** Suppose that for every pair of natural numbers $y$ and $z$, $\sigma_{y,z}$ is a first-order sentence. Then the following are equivalent:

- in every atomless ultraproduct $[G_n]_U$, for every $y$ there is a $z$ so that $\sigma_{y,z}$ is true,

- for every $y$ there is a $Z$ and an $n$ so that whenever $G_n$ is a graph with $\geq n$ vertices, there is a $z \leq Z$ so that $\sigma_{y,z}$ is true in $G_n$.

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

Conversely, suppose that there is an $y$ so that, for each $Z$, there are arbitrarily large graphs $G_n$ where each $\sigma_{y,z}$ with $z \leq Z$ is false. Then, for each $n$, take $G_n$ to be a graph with $\geq n$ vertices and such that, for each $z \leq n$, $\sigma_{y,z}$ is false. Let $[G_n]_U$ be any ultraproduct. Since, for each $z$, $\{n \mid \sigma_{y,z} \text{ is true in } G_n\}$ is finite, the lemma says that each $\sigma_{y,z}$ is false in $[G_n]_U$. \(\square\)

Note that we are avoiding writing “$\forall y \exists z \sigma_{y,z}$” because it is important to remember that the quantifiers over $y$ and $z$ 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 4.49. 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 $\epsilon$, there is an internal set $M_{f,r,\epsilon}$ such that

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

and

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

Proof. We construct $M_{f,r,\epsilon}$ almost by definition: for each $n$, take $M_{f,e,n} \subseteq V^*_{r,n}$ 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 $\delta$-quasirandom with $t_{K_2}(G) = p$ and $V$ sufficiently large, $|t_H(G) - t_{K_2}(G)|^{|F|} < \epsilon$.

For a fixed value of $H$, if we replace $\epsilon$ with $E = \lceil 1/\epsilon \rceil$ and $\delta$ with $D = \lceil 1/\delta \rceil$, we can take $\sigma_{E,D}$ to be “either $|t_{C_4}(G) - (t_{K_2}(G))^4| > 1/D$ then $|t_H(G) - t_{K_2}(G)|^{|F|} < 1/E$”. Understood in the ultraproduct, this corresponds to the weaker statement “either $|t_{C_4}(G) - (t_{K_2}(G))^4| \geq 1/D$ or $|t_H(G) - t_{K_2}(G)|^{|F|} \leq 1/E$”. Since we proved that $|t_{C_4}(G) - (t_{K_2}(G))^4| = 0$ implies $|t_H(G) - t_{K_2}(G)|^{|F|} = 0$, 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 [69], and measures of this kind are called Loeb measures. The notion of a graded probability space was introduced by Keisler [60] and used, in a context very close to the one we are concerned with, by Keisler’s student Hoover [53].
The Furstenberg correspondence—a correspondence between finite sets and dynamical systems introduced by Furstenberg to give his proof of Szemerédi’s Theorem \[38\]—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 \[57\]), but a general framework extending first-order logic by predicates for measures exists as well \[42\].
Chapter 5

Structure and Randomness

Throughout this chapter, we will continue to work in a Keisler graded probability space \( \{(V^k, B_k, \mu_k)\}_{k \in \mathbb{N}} \).

5.1 The \( L^2 \) Norm

We are going to develop tools for working with quasirandom graphs: we will develop a “norm” which will let us measure how random a graph, and more generally, a function is. This will give us an explanation for why \( t_{C_4}(G) \) is significant in ways that, say, \( t_{C_3}(G) \) is not: \( t_{C_3} \) does not have an associated norm.

As a preview, we recall the properties of the \( L^2 \) norm, which will be a model for the seminorm we are interested in.

**Definition 5.1.** When \( f : V^k \to \mathbb{R} \) is measurable, we define \( ||f||_{L^2(\mu_k)} = \sqrt{\int (f(x_1, \ldots, x_k))^2 d\mu_k} \).

We define \( L^2(\mu_k) \) to be the set of measurable functions such that \( ||f||_{L^2(\mu_k)} \) is finite (i.e. the integral exists).

When considering the space \( L^2(\mu_k) \), it is standard to identify functions “up to the \( L^2 \) norm”—that is, when \( ||f - g||_{L^2(\mu_k)} \), we treat \( f \) and \( g \) as equivalent, and even write \( f = g \). Note that this is weaker than pointwise equality—two functions could agree in this sense, but actually have \( f(x) \neq g(x) \) for some values of \( x \). However this is the same as requiring that \( \{x \mid f(x) \neq g(x)\} \) have measure 0. So when we write \( f = g \), we mean equality except possibly on a set of measure 0. This is consistent with our view that sets of measure 0 are negligible.

It is a standard fact that the \( L^2 \) norm is a norm; that is:
• for any $f$, $\|f\|_{L^2(\mu_k)} \geq 0$,
• if $\|f\|_{L^2(\mu_k)} = 0$ then $f = 0$ (the function constantly equal to 0, or at least equal to 0 except on a set of measure 0),
• for any real number $c$, $\|c \cdot f\|_{L^2(\mu_k)} = |c| \cdot \|f\|_{L^2(\mu_k)}$, and
• $\|f + g\|_{L^2(\mu_k)} \leq \|f\|_{L^2(\mu_k)} + \|g\|_{L^2(\mu_k)}$.

The last property is the triangle inequality.

One might ask why we use the power 2—why not work with $(\int (f(x_1, \ldots, x_k))^p \, d\mu_k)^{1/p}$ for some other value of $p$? One good reason is that the $L^2$ norm associated to an inner product.

**Definition 5.2.** When $f, g \in L^2(\mu_k)$, $\langle f, g \rangle_{L^2(\mu_k)} = \int f(x_1, \ldots, x_k)g(x_1, \ldots, x_k) \, d\mu_k$.

With this definition, $\|f\|_{L^2(\mu_k)} = \sqrt{\langle f, f \rangle_{L^2(\mu_k)}}$.

This inner product is bilinear:

$$\langle f, ag + bh \rangle = a\langle f, g \rangle + b\langle f, h \rangle$$

and

$$\langle af + bh, g \rangle = a\langle f, g \rangle + b\langle h, g \rangle.$$

The Cauchy-Schwarz inequality can be stated in the form

$$\langle f, g \rangle_{L^2(\mu_k)} \leq \|f\|_{L^2(\mu_k)} \cdot \|g\|_{L^2(\mu_k)}.$$ 

5.2 The $U^2$ Seminorm

We will define a norm-like operation based on $t_{C_4}$:

**Definition 5.3.**

$$\|f\|_{U^2} = \sqrt{\int \int f(x, y)f(x', y)f(x, y')f(x', y') \, d\mu_4}.$$ 

Of course, when $f = \chi_E$ is the characteristic function of a graph, $\|f\|_{U^2} = \sqrt{t_{C_4}(E)}$. We will establish the basic properties of this operation and then use those to see how it relates to quasirandomness.

For an arbitrary measurable function $f$, this integral might not exist. It would be enough to restrict to functions $f$ whose fourth power is integrable—the $L^4(\mu_2)$ functions for which $\int |f|^4 \, d\mu_2$ exists—but it will be sufficient for us, and simpler, to impose a stricter restriction and consider only the almost everywhere bounded functions.
Definition 5.4. $L^\infty(\mu_k) \subseteq L^2(\mu_k)$ is the space of functions $f$ such that there is a real number $c > 0$ such that $\mu_k(\{(x_1, \ldots, x_k) \mid |f(x_1, \ldots, x_k)| > c\}) = 0$.

That is, $L^\infty(\mu_k)$ is the functions which are restricted to $[-c, c]$ for some $c$, except on a set of measure 0.

$\|\cdot\|_{U^2}$ is known as the Gowers $U^2$ seminorm. It is the first of the Gowers uniformity seminorms. (The others will appear later, when we generalize to hypergraphs.)

As the name suggests, $\|\cdot\|_{U^2}$ is a seminorm, not a norm: it is possible to have $\|f\|_{U^2} = 0$ even when $f$ is non-zero function, but the other three properties of a norm (non-negativity, multiplication by scalars, and the triangle inequality) are satisfied, as we will prove shortly.

Definition 5.5. An operation $\|\cdot\|$ is a seminorm if it satisfies:

- for every $f$, $\|f\| \geq 0$,
- for any real number $c$ and any $f$, $\|c \cdot f\| = |c| \cdot \|f\|$,
- for any $f$ and $g$, $\|f + g\| \leq \|f\| + \|g\|$.

We could define operations like $\|\cdot\|_{U^2}$ corresponding to other graphs, like $\|f\|_{C_3} = \sqrt[3]{\int f(x, y)f(y, z)f(z, x) d\mu_3}$. However for most graphs, this fails to be a seminorm. Even the form suggests immediately that the result will not always be non-negative: there is no squaring going on, and indeed, it is not hard to find $f$ so that $\|f\|_{C_3}$ is negative.

Example 5.6. Let $A \subseteq V$ be a set with $\mu(A) = 1/2$ and define $f(x, y)$ by:

$$f(x, y) = \begin{cases} 
-1 & \text{if } x \in A \text{ and } y \in A \\
-1 & \text{if } x \notin A \text{ and } y \notin A \\
0 & \text{if } x \in A \text{ and } y \notin A \\
0 & \text{if } x \notin A \text{ and } y \in A 
\end{cases}$$

Then, for any $x, y, z$, $f(x, y)f(y, z)f(z, y)$ is 0 unless all three belong to $A$ or all three do not belong to $A$, in which case $f(x, y)f(y, z)f(z, y)$ is $-1$. In particular, $\|f\|_{C_3} = -1/4 < 0$.

However $C_4$ is not the only graph which gives us a seminorm, and it will be useful to consider a second example in parallel.

Definition 5.7. When $f \in L^\infty(\mu_2)$,

$$\|f\|_V = \sqrt{\int f(x, y)f(x, z) d\mu_3}.$$
We call this the $V$ seminorm because it corresponds to the graph $\mathcal{V}$. The $V$ seminorm is not as common (or important) as the $U^2$ seminorm, but belongs to a larger family of seminorms which extends the Gowers uniformity seminorms.

Before proving that $\| \cdot \|_{U^2}$ and $\| \cdot \|_V$ are seminorms, it is helpful to introduce the inner product-like notions corresponding to them. Unlike an ordinary inner product, the $U^2$ version of an inner product has four arguments.

**Definition 5.8.**

$$\langle f_1, f_2 \rangle_V = \int f_1(x,y)f_2(x,z)\,d\mu_3$$

and

$$\langle f_1, f_2, f_3, f_4 \rangle_{U^2} = \int f_1(x,y)f_2(x,y')f_3(x',y)f_4(x',y')\,d\mu_4.$$ 

Analogous to the equality $\| f \|_{L^2(\mu_k)}^2 = \langle f, f \rangle_{L^2(\mu_k)}$, we have $\| f \|_{V}^2 = \langle f, f \rangle_V$ and $\| f \|_{U^2}^4 = \langle f, f, f, f \rangle_{U^2}$.

Linearity in each coordinate follows by the linearity of integrals; for instance,

$$\langle f, ag + bh \rangle_V = a\langle f, g \rangle_V + b\langle f, h \rangle_V$$

and

$$\langle f_1, f_2, f_3, ag + bh \rangle_{U^2} = a\langle f_1, f_2, f_3, g \rangle_{U^2} + b\langle f_1, f_2, f_3, h \rangle_{U^2}.$$ 

Most importantly, the $V$ and $U^2$ norms satisfies a version of Cauchy-Schwarz we will use over and over again.

**Theorem 5.9 (Gowers-Cauchy-Schwarz).**

- $|\langle f_1, f_2 \rangle_V| \leq \|f_1\|_V \cdot \|f_2\|_V$, and
- $|\langle f_1, f_2, f_3, f_4 \rangle_{U^2}| \leq \|f_1\|_{U^2} \cdot \|f_2\|_{U^2} \cdot \|f_3\|_{U^2} \cdot \|f_4\|_{U^2}.$

**Proof.** Both parts follow by the use of Cauchy-Schwarz.

For the first part,

$$|\langle f_1, f_2 \rangle_V|^2 = \left| \int f_1(x,y) f_2(x,z) \, d\mu_3 \right|^2 \leq \int \left( \int f_1(x,y) \, d\mu(y) \right)^2 \left( \int f_2(x,z) \, d\mu(z) \right) \, d\mu(x) \leq \int \left( \int f_1(x,y) \, d\mu(y) \right)^2 \, d\mu \int \left( \int f_2(x,y) \, d\mu(y) \right)^2 \, d\mu = \|f_1\|_V^2 \cdot \|f_2\|_V^2,$$
so taking square roots of both sides,
\[ \|(f_1, f_2)\|^2 \leq \|f_1\| \cdot \|f_2\| \cdot \|v\|. \]

The second part is similar, but we use Cauchy-Schwarz twice.
\[
\|(f_1, f_2, f_3, f_4)\|_{U^2}^4
= \left| \int f_1(x, y) f_2(x, y') f_3(x', y) f_4(x', y') \, d\mu_4 \right|^4
\]
\[
= \int \left( \int f_1(x, y) f_3(x', y) \, d\mu(y) \right) \left( \int f_2(x, y') f_4(x', y') \, d\mu(y') \right) \, d\mu_2(x, x')^4
\]
\[
\leq \left| \int \left( \int f_1(x, y) f_3(x', y) \, d\mu(y) \right)^2 \, d\mu_2(x, x')^2 \cdot \left( \int \left( \int f_2(x, y') f_4(x', y') \, d\mu(y') \right)^2 \, d\mu_2(x, x')^2 \right]
\]
\[
\leq \left| \int f_1(x, y) f_1(x, y') f_3(x', y) f_3(x', y') \, d\mu_4 \right|^2 \cdot \left| \int f_2(x, y) f_2(x, y') f_4(x', y) f_4(x', y') \, d\mu_4 \right|^2.
\]

Consider one of these pieces:
\[
\left| \int f_1(x, y) f_1(x, y') f_3(x', y) f_3(x', y') \, d\mu_4 \right|^2
\]
\[
= \left| \int \left( \int f_1(x, y) f_1(x, y') \, d\mu(x) \right) \cdot \left( \int f_3(x', y) f_3(x', y) \, d\mu(x') \right) \, d\mu_2(y, y')^2
\]
\[
\leq \int \left( \int f_1(x, y) f_1(x, y') \, d\mu(x) \right)^2 \, d\mu_2(y, y') \cdot \int \left( \int f_3(x', y) f_3(x', y) \, d\mu(x') \right)^2 \, d\mu_2(y, y')
\]
\[
= \int f_1(x, y) f_1(x, y') f_3(x', y) f_3(x', y') \, d\mu_4 \cdot \int f_3(x, y) f_3(x, y') f_3(x', y) f_3(x', y') \, d\mu_4
\]
\[
= \|f_1\|_{U^2}^4 \cdot \|f_3\|_{U^2}^4.
\]

Similarly,
\[
\left| \int f_2(x, y) f_2(x, y') f_4(x', y) f_4(x', y') \, d\mu_4 \right|^2 \leq \|f_2\|_{U^2}^4 \cdot \|f_4\|_{U^2}^4.
\]

Putting these together,
\[
\langle f_1, f_2, f_3, f_4 \rangle^4 \leq \|f_1\|_{U^2}^4 \cdot \|f_3\|_{U^2}^4 \cdot \|f_2\|_{U^2}^4 \cdot \|f_4\|_{U^2}^4,
\]
and therefore
\[
\|(f_1, f_2, f_3, f_4)\| \leq \|f_1\|_{U^2} \cdot \|f_2\|_{U^2} \cdot \|f_3\|_{U^2} \cdot \|f_4\|_{U^2}.
\]
\[\square\]
CHAPTER 5. STRUCTURE AND RANDOMNESS

This argument gives some idea what makes $C_4$ and the $V$ graph special: they correspond to “doubling vertices”. If we begin with a single edge $\{x, y\}$ and “double” the $y$ vertex, we get the $V$ graph—the single vertex $x$ adjacent to two different vertices. If we take the $V$ graph and now double the $x$ vertex, we get the graph $C_4$: two copies of $y$ and two copies of $x$, with both $y$ vertices adjacent to both $x$ vertices.

This doubling is exactly what lets us apply the Cauchy-Schwarz inequality, and helps explain why $C_4$ and $V$ give us seminorms while most other graphs don’t.

Theorem 5.10. $\| \cdot \|_V$ and $\| \cdot \|_{U^2}$ are seminorms.

Proof. For any $L^\infty$ function $f$, $\|f\|_V = \sqrt{\int (\int f(x,y) \, d\mu(y))^2 \, d\mu(x)}$; since the quantity $(\int f(x,y) \, d\mu(y))^2$ is non-negative, $\|f\|_V$ is always defined and non-negative. Similarly, $\|f\|_{U^2} = \sqrt[4]{\int (\int f(x,y) \, f(x',y) \, d\mu(y))^2 \, d\mu_2(x,x')}$; since the quantity $(\int f(x,y) \, f(x',y) \, d\mu(y))^2$ is non-negative, we ensure that $\|f\|_{U^2}$ is always defined and non-negative.

For any real number $c$,

$$\|c \cdot f\|_V = \sqrt{\int c^2 f(x,y) f(x,z) \, d\mu_3} = |c| \cdot \|f\|_V$$

and

$$\|c \cdot f\|_{U^2} = \sqrt[4]{\int c^4 f(x,y) f(x',y) f(x,y') f(x',y') \, d\mu_4} = |c| \cdot \|f\|_{U^2}.$$ 

To see the triangle inequality for $\| \cdot \|_V$, observe that

$$\|f + g\|_V^2 = \langle f + g, f + g \rangle_V$$

$$= \langle f, f \rangle_V + 2 \langle f, g \rangle_V + \langle g, g \rangle_V$$

$$\leq \|f\|_V^2 + 2 \|f\|_V \cdot \|g\|_V + \|g\|_V^2$$

$$= (\|f\|_V + \|g\|_V)^2,$$

so $\|f + g\|_V \leq \|f\|_V + \|g\|_V$.

Similarly, to see the triangle inequality for $\| \cdot \|_{U^2}$, we work with the fourth power:

$$\|f + g\|_{U^2}^4 = \langle f + g, f + g, f + g, f + g \rangle_{U^2}$$

$$= \langle f, f, f, f \rangle_{U^2} + \langle f, f, g, g \rangle_{U^2} + \langle f, g, g, g \rangle_{U^2} + \langle g, g, g, g \rangle_{U^2}$$

$$\leq \|f\|_{U^2}^4 + \|f\|_{U^2}^4 \|g\|_{U^2}^4 + \cdots + \|g\|_{U^2}^4$$

$$= (\|f\|_{U^2} + \|g\|_{U^2})^4.$$
Therefore
\[ \|f + g\|_{U^2} \leq \|f\|_{U^2} + \|g\|_{U^2}. \]

Lemma 5.11. For any \( f \), \( |\int f \, d\mu_2| \leq \|f\|_{V} \leq \|f\|_{U^2} \leq \|f\|_{L^2(\mu_2)}. \)

Proof. Write \( 1 \) for the function constantly equal to \( 1 \). Then
\[
\left| \int f(x, y) \, d\mu_2 \right| = \left| \int f(x, y) 1(x, z) \, d\mu_3 \right|
= |\langle f, 1 \rangle_V|
\leq \|f\|_V \cdot \|1\|_V
= \|f\|_V
\]
since \( \|1\|_V = \int 1 \, d\mu_3 = 1. \)

Similarly,
\[
\|f\|_V^2 = \int f(x, y) f(x, z) \, d\mu_3
= \langle f, f, 1, 1 \rangle_{U^2}
\leq \|f\|^2_{U^2} \cdot \|1\|^2_{U^2}
= \|f\|^2_{U^2}
\]
since \( \|1\|_{U^2} = \int 1 \, d\mu_4 = 1. \) Taking the square root of both sides gives
\[ \|f\|_V \leq \|f\|_{U^2}. \]

Finally,
\[
\|f\|_{U^2}^4 = \int f(x, y) f(x, y') f(x', y) f(x', y') \, d\mu_4
\leq \|f\|_{L^2}^4
\]
by Cauchy-Schwarz. \( \square \)

The outer inequality \( |\int f \, d\mu_2| \leq \|f\|_{U^2} \) is an extension of our old observation that \( t_{K_2}(G)^4 \leq t_{C_4}(G). \) This suggests that we might think of an arbitrary function \( f \) as quasirandom exactly when \( \|f\|_{U^2} = |\int f \, d\mu_2|. \)

The \( V \) and \( U^2 \) norms satisfy an additional inequality which characterizes what makes them useful:

Theorem 5.12. For any measurable set \( B \) and any \( f \in L^\infty(\mu_2), \)
\[ \|f(x, y) \chi_B(x)\|_V \leq \|f\|_V \]
CHAPTER 5. STRUCTURE AND RANDOMNESS

and

\[ \|f(x, y)\chi_B(x)\|_{U^2} \leq \|f\|_{U^2} \text{ and } \|f(x, y)\chi_B(y)\|_{U^2} \leq \|f\|_{U^2}. \]

Applying the \( U^2 \) version of this lemma twice, once in the \( x \) coordinate and once in the \( y \) coordinate, shows that

\[ \|f(x, y)\chi_B(x)\chi_C(y)\|_{U^2} \leq \|f\|_{U^2}. \]

That is, when we restrict a function to a rectangle, the \( U^2 \) norm does not increase.

**Proof.** Let \( B \) be measurable and let \( f \in L^\infty(\mu_2) \) be given. Letting \( \overline{B} = V \setminus B \), we have

\[ f(x, y) = f(x, y)\chi_B(x) + f(x, y)\chi_{\overline{B}}(x). \]

For the \( V \) norm, observe that

\[
\|f\|_V^2 = \langle f, f \rangle_V \\
= \langle f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}} \rangle_V \\
= \langle f\chi_B, f\chi_B \rangle_V + \langle f\chi_B, f\chi_{\overline{B}} \rangle_V + \langle f\chi_{\overline{B}}, f\chi_B \rangle_V + \langle f\chi_{\overline{B}}, f\chi_{\overline{B}} \rangle_V \\
= \|f\chi_B\|_V^2 + 2 \int f(x, y)\chi_B(x)(x, z)\chi_{\overline{B}}(x) \, d\mu_3 + \|f\chi_{\overline{B}}\|_V^2 \\
= \|f\chi_B\|_V^2 + \|f\chi_{\overline{B}}\|_V^2. \\
\]

For the \( U^2 \) norm, the argument is similar except that there are more terms to worry about. We prove the first inequality for the \( U^2 \) norm since the last inequality follows symmetrically by the same argument.

\[
\|f\|_{U^2}^4 = \langle f, f, f, f \rangle_{U^2} \\
= \langle f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}}, f\chi_B + f\chi_{\overline{B}} \rangle_{U^2} \\
= \langle f\chi_B, f\chi_B \rangle_{U^2} + \langle f\chi_B, f\chi_{\overline{B}} \rangle_{U^2} + \langle f\chi_{\overline{B}}, f\chi_B \rangle_{U^2} + \langle f\chi_{\overline{B}}, f\chi_{\overline{B}} \rangle_{U^2} + \cdots + \langle f\chi_{\overline{B}}, f\chi_{\overline{B}}, f\chi_{\overline{B}}, f\chi_{\overline{B}} \rangle_{U^2}. \\
\]

There are a total of 16 terms we need to consider in this sum. We will show that each term is non-negative.

Each term has the form

\[ \int f(x, y)\chi_{S_1}(x)f(x', y)\chi_{S_2}(x')f(x, y')\chi_{S_3}(x)f(x', y')\chi_{S_4}(x') \, d\mu_4 \]

where \( S_1, S_2, S_3, S_4 \) are each either \( B \) or \( \overline{B} \).
5.2. THE $U^2$ SEMINORM

If $S_1 \neq S_3$ or $S_2 \neq S_4$ then this integral is 0, like the middle term in the $V$ norm argument above. For instance,

$$\int f(x,y)\chi_B(x)f(x',y)\chi_B f(x',y')\chi_B(x')d\mu_4 = 0$$

because for every $x$, one of $\chi_B(x)$ and $\chi_B(x)$ must be 0, so the product is always 0.

However there are four terms,

$$\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')f(x',y')\chi_{S_1}(x)f(x',y')\chi_{S_2}(x')d\mu_4$$

which do not cancel out so simply. But

$$\int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')f(x',y')\chi_{S_1}(x)f(x',y')\chi_{S_2}(x')d\mu_4$$

$$= \int \left( \int f(x,y)\chi_{S_1}(x)f(x',y)\chi_{S_2}(x')d\mu(y) \right)^2 d\mu_2(x,x'),$$

which is always non-negative because the integrand is squared.

Since all the terms are non-negative,

$$||f||^4_{U^2} \geq \langle f\chi_B, f\chi_B, f\chi_B, f\chi_B \rangle_{U^2} = ||f\chi_B||^4_{U^2}.$$

This implies that when $||f||_{U^2} = 0$, $f$ is orthogonal to any “rectangle” of the form $X \times Y$.

Lemma 5.13. If $||f||_{U^2} = 0$ then whenever $X \subseteq V$ and $Y \subseteq V$ are sets in $B(\mu)$,

$$\int f(x,y)\chi_X(x)\chi_Y(y)d\mu_2 = 0.$$

Proof. We have

$$\left| \int f(x,y)\chi_X(x)\chi_Y(y)d\mu_2 \right| \leq ||f(x,y)\chi_X(x)\chi_Y(y)||_{U^2} \leq ||f||_{U^2} = 0.$$
CHAPTER 5. STRUCTURE AND RANDOMNESS

5.3 Seminorms, Even Distribution, and Quasirandomness

We would like to identify the relationship between the $U^2$ norm and quasirandomness. When $E \subseteq V$ is a symmetric measurable set, the definitions say that $E$ is quasirandom exactly when $||\chi_E||_{U^2} = \mu_2(E)$.

However this turns out not to be the most important perspective on the relationship. We will ultimately want to be able to take any $L_\infty(\mu_2)$ function $f$ and decompose it into a structured part and a random part—to write $f = f^+ + f^-$ where $||f^-||_{U^2} = 0$ and $f^+$ has some sort of nice description. This is similar to the spectral perspective, where we write $f$ as a sum of eigenfunctions; indeed, one way to give a description of $f^+$ will be as the sum of the eigenfunctions.

When $f = \chi_E$ is quasirandom—and, more generally, when $||f||_{U^2} = \int f \, d\mu_2$—this decomposition will be particularly useful because $f^+$ will turn out to be a constant function.

Before proving this, it will be useful to develop the analogous theory for the $V$ norm, both to preview the techniques and as a step in the proof.

The $V$ norm corresponds to a weaker notion of randomness than quasirandomness.

Definition 5.14. $f$ is evenly distributed if, for almost every $x$, $\int f(x, y) \, d\mu(y) = \int f(x, y) \, d\mu_2$.

Like the $V$ norm itself, the definition of even distribution is asymmetric in the two variables, so we will most often apply it when $f$ itself is symmetric.

In particular, when $f = \chi_E$ is the characteristic function of a graph, $f$ is evenly distributed when (outside of a set of measure 0), every point $x$ has the “right number” of neighbors.

For instance, the bipartite graph with two equal parts—the graph where there is an $A \subseteq V$ with $\mu(A) = 1/2$ and $E = A \times (V \setminus A) \cup (V \setminus A) \times A$—is evenly distributed, since $\mu_2(E) = 1/2$ and, for every $x$, $\mu(E_x) = 1/2$. (But this graph is very far from being quasirandom—for instance, it has no triangles.)

Theorem 5.15. Let $f \in L_\infty(\mu_2)$ and take $p = |\int f \, d\mu_2|$. The following are equivalent:

- $f$ is evenly distributed,
- $||f||_V = p$. 
5.3. SEMINORMS, EVEN DISTRIBUTION, AND QUASIRANDOMNESS

- $||f - p||_V = 0$.

Proof. If $f$ is evenly distributed then

$$||f||_V = \sqrt{\int f(x,y)f(x,z)\,d\mu_3}$$

$$= \sqrt{\int (\int f(x,y)\,d\mu(y))^2 \,d\mu}$$

$$= \sqrt{\int p^2 \,d\mu}$$

$$= p.$$

If $||f||_V = p$ then

$$||f - p||_V^2 = \langle f - p, f - p \rangle_V$$

$$= \langle f, f \rangle_V - 2 \langle f, p \rangle_V + \langle p, p \rangle_V$$

$$= p^2 - 2 \int f(x,y)p \,d\mu_3 + p^2$$

$$= p^2 - 2p \int f(x,y) \,d\mu_2 + p^2$$

$$= 0.$$

If $||f - p||_V = 0$ then

$$0 = ||f - p||_V^2$$

$$= \int \left( \int f(x,y)\,d\mu(y) - p \right)^2 \,d\mu(x).$$

Therefore $\{ x \mid \int f(x,y)\,d\mu(y) = p \}$ must have measure 1, so $f$ is evenly distributed.

Knowing that $\chi_E$ is evenly distributed—that is, that $t_V(E) = t_{K_2}(E)^2$—is not enough to guarantee that $E$ contains the right number of copies of all graphs, as the example of the bipartite graph above shows. However, equidistributed graphs satisfy an analog of Theorem 4.30 for a limited family of graphs:

**Theorem 5.16.** If $G$ is evenly distributed then whenever $H = (W, F)$ is a tree—that is, $H$ contains no cycles—$t_H(G) = p^{|F|}$.
CHAPTER 5. STRUCTURE AND RANDOMNESS

Problem 5.17. Adapt the proof of Theorem \ref{thm:4.30} to prove this. Adapt the proof of Theorem \ref{thm:2.32} to prove the finite analog: for every tree $H$ and every $\epsilon > 0$, there is a $\delta > 0$ so that if $\left| t_{V}(G) - t_{K_2}(G)^2 \right| < \delta$ and $G$ is sufficiently large, then $\left| t_{H}(G) - t_{K_2}(G)^{|F|} \right| < \epsilon$.

Theorem 5.18. Let $f \in L^\infty(\mu_2)$ be symmetric and take $p = \left| \int f \, d\mu_2 \right|$. The following are equivalent:

- $\|f\|_{U^2} = p$,
- $\|f - p\|_{U^2} = 0$.

Proof. First, suppose $\|f\|_{U^2} = p$. Since $p \leq \|f\|_V \leq \|f\|_{U^2}$, we also have $\|f\|_V = p$, so $f$ is evenly distributed. We can calculate

$$
\|f - p\|_{U^2}^4 = \langle f - p, f - p, f - p, f - p \rangle_{U^2}
= \langle f, f, f, f \rangle_{U^2} - 3 \langle f, f, f, p \rangle_{U^2} + \cdots + \langle p, p, p, p \rangle_{U^2}.
$$

There are sixteen total terms in this sum which we need to consider; we will show that all of them are equal to $p^4$. The first and last terms are certainly equal to $p^4$.

Consider the four terms with $f$ three times and $p$ once. For instance

$$
\langle f, f, f, p \rangle_{U^2} = \int f(x, y) f(x, y') f(x', y) p \, d\mu_4
= p \int f(x, y) f(x', y) \int f(x, y') \, d\mu_4(x, x', y)
= p^2 \int f(x, y) f(x', y) \, d\mu_3
= p^2 \|f\|_V^2
= p^4.
$$

The other three of these terms are equal to $p^4$ by symmetric arguments.

Consider the six terms with $f$ twice and $p$ twice. These are not all symmetric: four of them are like

$$
\langle f, p, f, p \rangle_{V} = \int f(x, y) p f(x', y) \, d\mu_4
$$
where the two copies of $f$ share a variable, while two are like

$$
\langle f, p, p, f \rangle_{V} = \int f(x, y) p^2 f(x', y') \, d\mu_4.
$$
5.3. **SEMINORMS, EVEN DISTRIBUTION, AND QUASIRANDOMNESS**

For the first group,  
\[ \int f(x, y)pf(x', y)p\,d\mu_4 = p^2 t_V(f) = p^4. \]

For the second group,  
\[ \int f(x, y)p^2 f(x', y')\,d\mu_4 = p^2 (\int f(x, y)\,d\mu_2)^2 = p^4. \]

There are four terms where \( f \) only appears once, all of which are symmetric with \( \langle f, p, p, p \rangle_V = \int f(x, y)p^3\,d\mu_4 = p^4. \)

Putting these sixteen terms together,  
\[ ||f - p||_{U^2}^4 = p^4 - 4p^4 + 6p^4 - 4p^4 + p^4 = 0. \]

Conversely, suppose \( ||f - p||_{U^2} = 0 \). Since \( ||f||_{U^2} = ||f - p + p||_{U^2} \leq ||f - p||_{U^2} + ||p||_{U^2} = 0 + p \), we have \( ||f||_{U^2} \leq p \). Since we always have \( p \leq ||f||_{U^2} \), it follows that \( ||f||_{U^2} = p \). \( \square \)

This gives us an almost trivial proof of a version of Theorem 2.28. (Carefully comparing the finite proof we gave with the work above will show that this is really the same argument, with all the work packaged into various lemmata.)

**Corollary 5.19.** If \( E \) is quasirandom then whenever \( X \subseteq V \) and \( Y \subseteq V \) are sets in \( \mathcal{B}(\mu) \),  
\[ \mu_2(E \cap (X \times Y)) = \mu_2(E)\mu(X)\mu(Y). \]

**Proof.** Let \( p = \mu_2(E) \). Since \( E \) is quasirandom, \( ||\chi_E - p||_{U^2} = 0 \), so by Lemma 5.13 we have  
\[ 0 = \int (\chi_E(x, y) - p)\chi_X(x)\chi_Y(y)\,d\mu_2 = \int \chi_E(x, y)\chi_X(x)\chi_Y(y)\,d\mu_2 - p \int \chi_X(x)\chi_Y(y)\,d\mu_2, \]

which is the same as saying that  
\[ \mu(E \cap (X \times Y)) = p\mu(X)\mu(Y). \]

\( \square \)
5.4 Rectangles

$L^2(\mu_2)$ is a vector space, and functions with $U^2$ norm 0 form a subspace, so we expect to be able to identify a dual to the space of functions with $U^2$ norm 0. That will let us decompose functions, writing

$$f = f^+ + f^-$$

where $||f^-||_{U^2} = 0$ and $f^+$ is orthogonal to all functions with $U^2$ norm 0.

**Definition 5.20.** We say $f$ is dual to $U^2$ if whenever $||r||_{U^2} = 0$, $\int r \, d\mu_2 = 0$.

By Lemma 5.13, whenever $X,Y \subseteq V$, $\chi_X(x)\chi_Y(y)$ is dual to $U^2$. On the other hand, if $f$ is dual to $U^2$ and non-zero then, in particular, $\int f f \, d\mu_2 = ||f||^2_{L^2(\mu_2)} > 0$, so we must have $||f||_{U^2} \neq 0$.

The following result then tells us that when $f$ is dual to $U^2$, at least $f$ correlates with a rectangle.

**Theorem 5.21.** When $f \in L^\infty(\mu_2)$, $||f||_{U^2} > 0$ if and only if there is a rectangle $X \times Y$ such that $|\int_{X \times Y} f \, d\mu_2| > 0$.

**Proof.** Suppose $||f||_{U^2} > 0$, so

$$0 < ||f||_{U^2}^4$$

$$= \int f(x,y)f(x',y)f(x,y')f(x',y') \, d\mu_4$$

$$= \int \left[ \int f(x,y)f(x',y)f(x,y')f(x',y') \, d\mu_2(x,y) \right] \, d\mu_2(x',y').$$

Let us assume that $\mu_2(\{(x,y) \mid |f(x,y)| > 1\}) = 0$ (the case where this holds with 1 replaced by some value $d$ follows the same argument after scaling).

There must be some $x',y'$ with $|f(x',y')| \leq 1$ so that

$$\int f(x,y)f(x',y)f(x',y') \, d\mu_2 \geq \int f(x,y)f(x',y)f(x',y') \, d\mu_2 = \epsilon > 0.$$

(And also $\mu(\{x \mid |f(x,y')| > 1\}) = \mu(\{y \mid |f(x',y)| > 1\}) = 0$.)

For any interval $[a,b]$, let $X_{a,b} = \{x \mid f(x,y') \in [a,b]\}$ and $Y_{[a,b]} = \{y \mid f(x',y) \in [a,b]\}$. If we choose $[a,b]$ and $[c,d]$ so that $|b-a|$ and $|d-c|$ are small, the value of this integral on $X_{[a,b]} \times Y_{[c,d]}$ does not depend much on
5.4. RECTANGLES

\[
\int_{X[a,b] \times Y[c,d]} f(x,y)f(x',y') \, d\mu_2 \leq |ac| \int_{X[a,b] \times Y[c,d]} f(x,y) \, d\mu_2 \\
+ |b-a| \cdot |c-d| \mu(X[a,b]) \mu(Y[c,d]).
\]

In particular, if we take \( K \geq \sqrt{\frac{2}{\epsilon}} \),

\[
\epsilon \leq \left| \int f(x,y)f(x',y)f(x,y') \, d\mu_2 \right|
\leq \sum_{i \leq K} \sum_{j \leq K} \left| \int_{X[i\sqrt{\frac{1}{2}},(i+1)\sqrt{\frac{1}{2}}] \times Y[j\sqrt{\frac{1}{2}},(j+1)\sqrt{\frac{1}{2}}]} f(x,y)f(x',y') \, d\mu_2 \right|
\leq \sum_{i \leq K} \sum_{j \leq K} \frac{ij\epsilon}{2} \left| \int_{X[i\sqrt{\frac{1}{2}},(i+1)\sqrt{\frac{1}{2}}] \times Y[j\sqrt{\frac{1}{2}},(j+1)\sqrt{\frac{1}{2}}]} f(x,y) \, d\mu_2 \right| + \frac{\epsilon}{2}.
\]

In particular, this means there must be some \( i, j \) so that

\[
\left| \int_{X[i\sqrt{\frac{1}{2}},(i+1)\sqrt{\frac{1}{2}}] \times Y[j\sqrt{\frac{1}{2}},(j+1)\sqrt{\frac{1}{2}}]} f(x,y) \, d\mu_2 \right| > 0.
\]

\begin{proof}

This suggests that the rectangles will be the building blocks for the functions which are dual to \( U^2 \).
\end{proof}

**Definition 5.22.** \( B^0_{2,1} \subseteq B_2 \) consists of sets of the form \( \bigcup_i X_i \times Y_i \) where each \( X_i, Y_i \in B_1 \).

What we really need is for \( B^0_{2,1} \) to be an algebra—that is, to be closed under complements, finite unions, and finite intersections. Normally this requires more than just closure under finite unions, but since the complement of a rectangle is a finite union of rectangles, our definition of \( B^0_{2,1} \) is sufficient.

**Lemma 5.23.** \( B^0_{2,1} \) is an algebra—that is:
• whenever $B \in B_{2,1}$, also $V \setminus B \in B_{2,1}$,

• whenever $B_1, \ldots, B_k \in B_{2,1}$, also $\bigcup_{i \leq k} B_i \in B_{2,1}$,

• whenever $B_1, \ldots, B_k \in B_{2,1}$, also $\bigcap_{i \leq k} B_i \in B_{2,1}$.

Proof. It is helpful to note that when $B = \bigcup_{i \leq k} X_i \times Y_i$, we can rewrite $B$ as a union $\bigcup_{i \leq k'} X'_i \times Y'_i$ so that for any $i, j \leq k'$, either $X'_i = X'_j$ or $X'_i \cap X'_j = \emptyset$, and either $Y'_i = Y'_j$ or $Y'_i \cap Y'_j = \emptyset$. To see this, we identify all the intersections of the $X_i$ and the $Y_i$—all sets of the form $X_s = \bigcap_{i \in s} X_i \cap \bigcap_{i \not\in s} (V \setminus X_i)$ for some $s \subseteq [1, k]$, and all $Y_s = \bigcap_{i \in s} Y_i \cap \bigcap_{i \not\in s} (V \setminus Y_i)$ for some $s \subseteq [1, k]$. (The sets $X_s$ are precisely the atoms of the finite algebra of sets generated by the sets $\{X_i\}$, and similarly for the $Y_s$.)

Suppose $s \neq s'$. Then there is an $i \in s \triangle s'$; without loss of generality, assume $i \in s$. Then $X_s \subseteq X_i$ but $X_{s'} \cap X_i = \emptyset$. Therefore $X_s \cap X_{s'} = \emptyset$.

So we can take $B$ to be the union of those $X_s \times Y_s$ such that there is some $i$ with $X_s \subseteq X_i$ and $Y_s \subseteq Y_i$.

Without loss of generality, let us write $B = \bigcup_{i \leq k} X_i \times Y_i$ with the additional condition that for any $i, j \leq k$, either $X_i = X_j$ or $X_i \cap X_j = \emptyset$, and either $Y_i = Y_j$ or $Y_i \cap Y_j = \emptyset$. Then we can see that $V^2 \setminus B$ is also a union of rectangles: for each $X_i$, let $Y'_i = \bigcup_{j \leq k, X_j = X_i} Y_j$, and set

$$V^2 \setminus B = \bigcup_{i \leq k} X_i \times (V \setminus Y'_i).$$

The second part holds since a union of unions is a union, and the third part follows from the first two since

$$\bigcap_{i \leq k} B_i = V^2 \setminus \left[ \bigcup_{i \leq k} (V^2 \setminus B_i) \right].$$

We really want to work with $\sigma$-algebras, so we have to extend $B_{2,1}$ slightly, to those sets which can be approximated using $B_{2,1}$:

**Definition 5.24.** $B_{2,1} \subseteq B_2$ consists of those sets $B \in B_2$ such that, for every $\epsilon > 0$, there is a $B_\epsilon \in B_{2,1}$ such that $\mu_2(B \triangle B_\epsilon) < \epsilon$.

**Theorem 5.25.** $B_{2,1}$ is a $\sigma$-algebra.
5.5. GRIDS

Proof. Clearly $\emptyset$ and $V^2 \in \mathcal{B}_{2,1}$. If $B \in \mathcal{B}_{2,1}$ then, for every $\epsilon > 0$, $\mu_2((V^2 \setminus B) \Delta (V^2 \setminus B')) = \mu_2(B \Delta B') < \epsilon$.

We can show that $\mathcal{B}_{2,1}$ is closed under unions of two elements (and therefore under finite unions and intersections): if $B, B' \in \mathcal{B}_{2,1}$ then, for any $\epsilon > 0$, we may find $B_\epsilon, B'_\epsilon$ with $\mu_2(B \setminus B_\epsilon) < \epsilon/2$ and $\mu_2(B' \setminus B'_\epsilon) < \epsilon/2$. Then

$$\mu_2((B \cup B') \setminus (B \cup B'_\epsilon)) \leq \mu_2(B \setminus B_\epsilon) + \mu_2(B' \setminus B'_\epsilon) < \epsilon.$$

Suppose we have a sequence with $B_i \in \mathcal{B}_{2,1}$ for all $i \in \mathbb{N}$. By replacing each $B_i$ with $B_i \setminus \bigcup_{j<i} B_j$, we may assume the $B_i$ are pairwise disjoint. For each $i$, choose $B_{2^{-i}} \in \mathcal{B}_{2,1}$ with $\mu_2(B_i \setminus B_{2^{-i}}) < \epsilon 2^{-i-1}$. Since the $B_i$ are pairwise disjoint, $\mu_2(\bigcup_{i \geq N} B_i) = \sum_{i \geq N} \mu_2(B_i) \leq 1$, so we may choose $N$ large enough that $\mu_2(\bigcup_{i \geq N} B_i) < \epsilon/2$. Then

$$\mu_2(\bigcup_i B_i \setminus \bigcup_{i < N} B_{2^{-i}}) \leq \sum_{i < N} \mu_2(B_i \setminus B_{2^{-i}}) + \mu_2(\bigcup_{i \geq N} B_i) \leq \epsilon \sum_{i < N} 2^{-i-1} + \epsilon/2 < \epsilon.$$

\[\square\]

5.5 Grids

We can now define a subspace of $L^\infty(\mu_2)$ which, we will eventually show, exactly captures the notion of being dual to $U^2$.

Definition 5.26. $L^\infty(\mathcal{B}_{2,1})$ is the subspace of $L^\infty(\mu_2)$ consisting of functions which are measurable with respect to $\mathcal{B}_{2,1}$.

Suppose $V = \bigcup_{i \leq k} B_i$ is a finite partition (so when $i \neq j$, $B_i \cap B_j = \emptyset$). Then the corresponding rectangles form a partition of $V^2$. We will—temporarily—define a “grid” on $\{B_i\}_{i \leq k}$ to be a step function on these rectangles—that is, a function of the form

$$\sum_{i,j \leq k} \gamma_{ij} \chi_{B_i}(x) \chi_{B_j}(y)$$

for some choices of constants $\gamma_{i,j}$. (This definition is temporary because we will give it its proper name in the next section.)

Lemma 5.27. For any function $f$ and any finite partition $V = \bigcup_{i \leq k} B_i$, the grid

$$\sum_{i,j \leq k} \alpha_{ij} \chi_{B_i}(x) \chi_{B_j}(y)$$

We can now define a subspace of $L^\infty(\mu_2)$ which, we will eventually show, exactly captures the notion of being dual to $U^2$. 

Definition 5.26. $L^\infty(\mathcal{B}_{2,1})$ is the subspace of $L^\infty(\mu_2)$ consisting of functions which are measurable with respect to $\mathcal{B}_{2,1}$.

Suppose $V = \bigcup_{i \leq k} B_i$ is a finite partition (so when $i \neq j$, $B_i \cap B_j = \emptyset$). Then the corresponding rectangles form a partition of $V^2$. We will—temporarily—define a “grid” on $\{B_i\}_{i \leq k}$ to be a step function on these rectangles—that is, a function of the form

$$\sum_{i,j \leq k} \gamma_{ij} \chi_{B_i}(x) \chi_{B_j}(y)$$

for some choices of constants $\gamma_{i,j}$. (This definition is temporary because we will give it its proper name in the next section.)

Lemma 5.27. For any function $f$ and any finite partition $V = \bigcup_{i \leq k} B_i$, the grid

$$\sum_{i,j \leq k} \alpha_{ij} \chi_{B_i}(x) \chi_{B_j}(y)$$
CHAPTER 5. STRUCTURE AND RANDOMNESS

with \( \alpha_{i,j} = \frac{\int_{B_i \times B_j} f(x,y) \, d\mu_2}{\mu_2(B_i \times B_j)} \) is the grid minimizing the \( L^2(\mu_2) \) distance from \( f \).

Proof. Consider an arbitrary grid \( \sum_{i,j \leq k} \gamma_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) \) and let \( e_{i,j} = \gamma_{i,j} - \alpha_{i,j} \). We will show that choosing \( e_{i,j} = 0 \) minimizes the \( L^2(\mu_2) \) distance. The main point is that \( \alpha_{i,j} \) is chosen so that

\[
\int_{B_i \times B_j} (f - \alpha_{i,j})^2 \, d\mu_2 = 0.
\]

Then

\[
||f - \sum_{i,j \leq k} \gamma_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||^2_{L^2(\mu_2)} = \int (f - \sum_{i,j \leq k} \gamma_{i,j} \chi_{B_i}(x) \chi_{B_j}(y))^2 \, d\mu_2
\]

\[
= \sum_{i,j \leq k} \int (f - \gamma_{i,j})^2 \, d\mu_2
\]

\[
= \sum_{i,j \leq k} \int (f - \alpha_{i,j} - e_{i,j})^2 \, d\mu_2
\]

\[
= \sum_{i,j \leq k} \int (f - \alpha_{i,j})^2 - 2(f - \alpha_{i,j})e_{i,j} + e_{i,j}^2 \, d\mu_2
\]

\[
= \sum_{i,j \leq k} \int (f - \alpha_{i,j})^2 + e_{i,j}^2 \, d\mu_2.
\]

In particular, choosing all values of \( e_{i,j} \) to be equal to 0—so \( \gamma_{i,j} = \alpha_{i,j} \)—minimizes the \( L^2(\mu_2) \) distance.

\[\Box\]

Theorem 5.28. If \( f \in L^\infty(B_{2,1}) \) then, for every \( \epsilon > 0 \), there is a partition

\( V = \bigcup_{i \leq k} B_i \) such that, taking \( \alpha_{i,j} = \frac{\int_{B_i \times B_j} f(x,y) \, d\mu_2}{\mu_2(B_i \times B_j)} \), we have

\[
||f - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon.
\]

Proof. Let \( f \) and \( \epsilon > 0 \) be given. It is a standard result about measurable functions that \( f \) is approximated by simple functions:

\[
||f - \sum_{i \leq d} \gamma_{i} \chi_{D_i}(x,y)||_{L^2(\mu_2)} < \epsilon/2
\]

with \( D_i \in B_{2,1} \). Each \( D_i \) can be approximated by finite unions of rectangles: for each \( D_i \), we have

\[
\mu(D_i \triangle \bigcup_{j \leq d} X_{i,j} \times Y_{i,j}) < \frac{\epsilon}{2\gamma_i},
\]
5.5. GRIDS

so

$$\|f - \sum_{i \leq d, j \leq d} \gamma_i \chi_{X_{i,j}}(x) \chi_{Y_{i,j}}(y)\|_{L^2(\mu_2)} < \epsilon.$$ 

We rearrange the $X_{i,j}, Y_{i,j}$ into a partition $\{B_i\}_{i \leq k}$ as in the proof of Lemma 5.23. Since $\sum_{i \leq d, j \leq d} \gamma_i \chi_{X_{i,j}}(x) \chi_{Y_{i,j}}(y)$ is a grid, we have

$$\|f - \sum_{i,j} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)\|_{L^2(\mu_2)} \leq \|f - \mathbb{E}(f | \{B_i \times B_j\}_{i,j \leq k})\|_{L^2(\mu_2)} < \epsilon.$$ 

We actually want a small refinement of this result which is more complicated to state, but more useful. In particular, we want to insist that the sets $B_i$ in the partition be internal. The property of the internal sets we need is that they are a dense algebra in $B_1$—that is, the internal sets are closed under complement, finite union, and finite intersection, and every set in $B_1$ can be approximated by internal sets. In practice, the case where are always interested in is when $B$ is the internal sets.

**Corollary 5.29.** If $f \in L^\infty(B_2, \lambda) \cap B \subseteq B_1$ is a dense algebra then, for every $\epsilon > 0$, there is a partition $V = \bigcup_{i \leq k} B_i$ with each $B_i \in B$ and $\mu_1(B_i) > 0$ such that, taking $\alpha_{i,j} = \frac{\int_{B_i \times B_j} f(x,y) \, d\mu_2}{\mu_2(B_i \times B_j)}$, we have

$$\|f - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)\|_{L^2(\mu_2)} < \epsilon.$$ 

**Proof.** Follow the same proof, using the fact that each $D_i$ can be can be approximated by rectangles from $\mathcal{B}$, and taking sets of measure 0 and combining them with a set of positive measure.

The simplest examples of such functions are things like the characteristic function of a bipartite graph: we have $V = B_1 \cup B_2$ and $f(x,y) = \chi_{B_1}(x) \chi_{B_2}(y) + \chi_{B_2}(x) \chi_{B_1}(y)$.

When $0 \leq f \leq 1$, we can represent these functions by drawing grids. For instance, the bipartite function can be drawn as a grid where we think of this grid as representing $V^2$, with one copy of $V$ on each axis. The filled in boxes are where the function is equal to 1 and the empty boxes are where
the function is equal to 0. On the other hand, 

\[ B_1 \quad B_2 \]

\[ B_1 \quad B_2 \]

is the characteristic function of the complement—two complete graphs, one on \( B_1 \) and one on \( B_2 \), with no edges between them.

In Section 2.3 we constructed the graphs \( G_p \) which interpolated between these two examples. The graph \( G_p \) partitioned \( V \) into four regions and (when \( p < 1/2 \)) corresponded to a grid like \[
\begin{array}{cc}
\begin{array}{cc}
\begin{array}{cc}
\begin{array}{cc}
\text{black} & \text{white} \\
\text{white} & \text{black}
\end{array}
\end{array}
\end{array}
\end{array}
\]
where edges in the black regions are always present, edges in the white regions are never present, edges in the light grey regions are present with probability \( p \), and edges in the dark gray regions are present with probability \( 1 - p \).

More generally, a grid can have different regions of varying sizes. In the grid \[
\begin{array}{cc}
\begin{array}{cc}
\begin{array}{cc}
\begin{array}{cc}
\text{black} & \text{light grey}
\end{array}
\end{array}
\end{array}
\end{array}
\]
we have \( V = B_1 \cup B_2 \cup B_3 \cup B_4 \) where \( \mu_1(B_2) \) is smaller than the measure of the other regions and the \( \alpha_{i,j} \) are various values in \([0, 1]\) (for instance, we can see that \( \alpha_{3,3} \) is a small but non-zero value, while \( \alpha_{1,1} \) is a dark grey corresponding to a value close to but not equal to 1).

Whenever \( f \in L^2(B_{2,1}) \) and \( 0 \leq f \leq 1 \), Theorem 5.28 says that \( f \) can always be approximated by grids like these. In particular, sets in \( B_{2,1} \) can be approximated by grids which are mostly “black or white”—that is, we can arrange for most of the measure to be on rectangles where the constant is either close to 0 or close to 1.

**Lemma 5.30.** Suppose \( E \in B_{2,1} \), \( V = \bigcup_{i \leq k} B_i \) is a partition, and there is a grid so that

\[ ||\chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) ||_{L^2(\mu_2)} < \epsilon^{3/2}. \]

Then

\[ \mu_2(\bigcup_{i,j \leq k, \alpha_{i,j} \in (\epsilon, 1-\epsilon)} B_i \times B_j) < \epsilon. \]

**Proof.** Since \( \chi_E(x, y) \in \{0, 1\} \) for all \( x, y \), whenever \( x \in B_i \) and \( y \in B_j \) so that \( \alpha_{i,j} \in (\epsilon, 1-\epsilon) \) we have

\[ |\chi_E(x, y) - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)| = |\chi_E(x, y) - \alpha_{i,j}| \geq \epsilon. \]
Therefore, letting \( U = \bigcup_{i,j \leq k, \alpha_{i,j} \in (\epsilon, 1-\epsilon)} B_i \times B_j \),
\[
\epsilon^3 > \left\| \chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) \right\|_{L^2(\mu)}^2
= \int \left( \chi_E - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) \right)^2 \, d\mu
\geq \int_U \epsilon^2 \, d\mu
= \mu_2(U) \epsilon^2.
\]
So \( \mu_2(U) < \epsilon \).

The existence of some rectangles where the constant \( \alpha_{i,j} \) is in the interval \((\epsilon, 1-\epsilon)\) is sometimes unavoidable, as the following example shows. The idea is to use the diagonal:

The idea is that no matter how many rectangles we use to approximate this, the part along the diagonal will still be “fuzzy”:

\[\text{Theorem 5.31. There is a Keisler graded probability space } \{(V, B_k, \mu_k)\}_{k \in \mathbb{N}} \text{ and an } E \in B_{2,1} \text{ so that whenever } V = \bigcup_{i \leq k} B_i \text{ is a partition there are } i,j \leq k \text{ so that } \mu_2((B_i \times B_j) \cap E) = 1/2.\]

**Proof.** Naturally, we will construct \( G \) as an ultraproduct. Let \( V_n = \{1, 2, \ldots, n\} \). We let \( E_n \) consist of those pairs \((x, y)\) with \( x < y \).

Let \( V = [V_n]_\mathcal{U} \) and \( E = [E_n]_\mathcal{U} \). First we must show that \( E \in B_{2,1} \)—that is, for every \( k \), \( E \) can be approximated up to \( 1/k \) by unions of rectangles. Given \( k > 0 \), we can partition each \( V_n \) with \( n \) big enough into intervals \( I_1^n, \ldots, I_k^n \) of size close to \( 1/k \): take \( I_1^n = \{1, \ldots, \lfloor n/k \rfloor\}, I_2^n = \{\lfloor n/k \rfloor + 1, \ldots, \lfloor 2n/k \rfloor\}, \) and so on. Take \( I^i = [I_1^n]_\mathcal{U} \), so \( \mu_1(I^i) = 1/k \). Observe that
\[
\bigcup_{i<j \leq k} I^i \times I^j \subseteq E \subseteq \bigcup_{i \leq j \leq k} I^i \times I^j.
\]

*The example is not symmetric, but it can be turned into a bipartite graph with a bit more effort—instead of working with a single copy of \( \{1, 2, \ldots, n\} \), we work with the disjoint union of two copies, and take an edge to be a pair \((x, y)\) with \( x \) from the first copy, \( y \) from the second, and \( x < y \).
Therefore
\[
\mu_2(\mathcal{E} \Delta \bigcup_{i<j \leq k} I^i \times I^j) = \mu_2(\mathcal{E} \setminus \bigcup_{i<j \leq k} I^i \times I^j) \\
\leq \mu_2((\bigcup_{i<j \leq k} I^i \times I^j) \setminus (\bigcup_{i \leq j \leq k} I^i \times I^j)) \\
= \mu_2(\bigcup_{i \leq k} I^i \times I^i) \\
= \frac{1}{k}.
\]

On the other hand, consider some \( \{B_i\}_{i \leq k} \). On the sets \( B_i \times B_i \), we always have \( \mu_2((B_i \times B_i) \cap \mathcal{E}) = \frac{1}{2} \mu_2(B_i \times B_i) \).

This example will turn out to be a central one in the next chapter, where we will be concerned with conditions that avoid this situation, and discover that in some sense this is the only example.

## 5.6 Conditional Expectation

The \( \sigma \)-algebra \( \mathcal{B}_{2,1} \) is a sub-\( \sigma \)-algebra of \( \mathcal{B}_2 \). Whenever we have two \( \sigma \)-algebras like this, we can talk about the conditional expectation of functions from the larger \( \sigma \)-algebra with respect to the smaller one.

For the purposes of this section, we forget the setting of a Keisler graded probability space and work in a more general setting: we suppose we have a measure space \( (V, \mathcal{B}, \mu) \) and a sub-\( \sigma \)-algebra \( \mathcal{D} \subseteq \mathcal{B} \).

We write \( L^2 \) for the set of functions \( f \) which are measurable with respect to \( \mathcal{B} \) and such that \( \int f^2 \, d\mu \) is finite. We write \( L^2(\mathcal{D}) \) for the subset of \( L^2 \) consisting only of functions which are measurable with respect to \( \mathcal{D} \).

In the presence of a measure, we can define the conditional expectation.

**Theorem 5.32.** For any \( f \in L^2(\mathcal{B}) \):

- There is a function \( g \in L^2(\mathcal{D}) \) such that, for all \( h \in L^2(\mathcal{D}) \),
  \[
  \|f - g\|_{L^2} \leq \|f - h\|_{L^2},
  \]
  and
- If \( g_0, g_1 \in L^2(\mathcal{D}) \) both have the property that for all \( h \in L^2(\mathcal{B}) \), \( \|f - g_i\|_{L^2} \leq \|f - h\|_{L^2} \), then \( \|g_0 - g_1\|_{L^2} \).
Proof. The proofs of both parts follow from a “quantitative” version of the second part. Let $f$ be given and let $\alpha = \inf_{h \in L^2(D)} \|f - h\|_{L^2}$.

What we will show is that if $g_0, g_1 \in L^2(D)$ are “almost as good as $\alpha$” at approximating $f$ then $g_0$ and $g_1$ must be near each other. More precisely, if $\|f - g_i\|_{L^2} \leq \alpha + \epsilon$ for each $i$ then $\|g_0 - g_1\|_{L^2} \leq \sqrt{8\alpha \epsilon} + 4\epsilon^2$.

For observe that

$$
\|f - \frac{g_0 + g_1}{2}\|_{L^2}^2 = \langle f - \frac{g_0 + g_1}{2}, f - \frac{g_0 + g_1}{2} \rangle_{L^2} \\
= \langle f, f \rangle_{L^2(\mu_2)} - \langle f, g_0 \rangle_{L^2} - \langle f, g_1 \rangle_{L^2} - \frac{1}{4} \langle g_0, g_0 \rangle_{L^2} + \frac{1}{4} \langle g_1, g_1 \rangle_{L^2} + \frac{1}{2} \langle g_0, g_1 \rangle_{L^2} \\
= \frac{1}{2} \|f - g_0\|_{L^2}^2 + \frac{1}{2} \|f - g_1\|_{L^2}^2 - \frac{1}{4} \|g_0 - g_1\|_{L^2}^2 \\
\leq (\alpha + \epsilon)^2 - \frac{1}{4} \|g_0 - g_1\|_{L^2}^2.
$$

Since we must have $\|f - \frac{g_0 + g_1}{2}\|_{L^2} \geq \alpha$, it follows that $\|g_0 - g_1\|_{L^2} \leq \sqrt{8\alpha \epsilon} + 4\epsilon^2$.

This immediately gives the second part: if we had such a $g_0, g_1$, we would have $\|g_0 - g_1\|_{L^2} < \delta$ for all $\delta$.

To prove the first part, for each $n$ we may choose $g_n$ with $\|f - g_n\|_{L^2} < \alpha + 1/n$. This is a Cauchy sequence, since for each $\epsilon > 0$, when $n$ is small enough, $\|g_n - g_m\|_{L^2} < \epsilon$ for all $m \geq n$. Then by the completeness of $L^2$, we may choose $g$ to be the limit of this sequence. \qed

Definition 5.33. When $f \in L^2$, we write

$$
\mathbb{E}(f \mid D)
$$

for the function in $L^2(D)$ given by the preceding lemma.

When $B \in \mathcal{B}$, we write $\mathbb{E}(B \mid D)$ as an abbreviation of $\mathbb{E}(\chi_D \mid D)$.

We refer to the conditional expectation interchangeably as the projection onto $D$ (or, more properly, onto $L^2(D)$), since the conditional expectation operation is indeed a projection of a vector space onto a subspace.

$\mathbb{E}(f \mid D)$ represents the “best approximation” to $f$ using only the sets from $D$.

Lemma 5.34. For any $f \in L^2(\mathcal{B})$ and $g \in L^2(D)$,

- $\int fg \, d\mu = \int \mathbb{E}(f \mid D)g \, d\mu$,
- $\int (f - \mathbb{E}(f \mid D))g \, d\mu = 0$. 
CHAPTER 5. STRUCTURE AND RANDOMNESS

Proof. The two parts are equivalent (using the linearity of the integral and moving an integral to the other side of the equality), so it suffices to prove the second part.

Suppose not—suppose \(|f - \mathbb{E}(f | D) - c| = c > 0\). Let \(h = \mathbb{E}(f | D) + \frac{c}{||g||_{L^2}} g\). \(h\) is also in \(L^2(D)\), and we will show it is a better approximation of \(f\):

\[
||f - h||^2_{L^2} = \int (f - \mathbb{E}(f | D) - \frac{c}{||g||_{L^2}} g)^2 d\mu
\]

\[
= ||f - \mathbb{E}(f | D)||^2_{L^2} - 2 \frac{c}{||g||_{L^2}} \int (f - \mathbb{E}(f | D)) g d\mu + \frac{c^2}{||g||_{L^2}^2} ||g||_{L^2}
\]

\[
= ||f - \mathbb{E}(f | D)||^2_{L^2} - \frac{c^2}{||g||_{L^2}^2}
\]

\[
< ||f - \mathbb{E}(f | D)||^2_{L^2},
\]

which contradicts the definition of \(\mathbb{E}(f | D)\).

We can think of \(||\mathbb{E}(f | D)||_{L^2}\) as a measurement of what portion of \(f\) has been explained by \(D\).

Lemma 5.35.

\[
||f||^2_{L^2} = ||\mathbb{E}(f | D)||^2_{L^2} + ||f - \mathbb{E}(f | D)||^2_{L^2}.
\]

Proof.

\[
||f||^2_{L^2} = \int (f - \mathbb{E}(f | D) + \mathbb{E}(f | D))^2 d\mu
\]

\[
= ||f - \mathbb{E}(f | D)||^2_{L^2} + 2 \int (f - \mathbb{E}(f | D)) \mathbb{E}(f | D) d\mu + ||\mathbb{E}(f | D)||^2_{L^2}
\]

\[
= ||\mathbb{E}(f | D)||^2_{L^2} + ||f - \mathbb{E}(f | D)||^2_{L^2}.
\]

The “grids” of the previous section were examples of conditional expectation: the finite collection \(D = \{B_i \times B_j\}_{i,j \leq k}\) is a sub-\(\sigma\)-algebra of \(B_{2,1}\) (because the collection if finite, there are no countable unions or intersections to consider). Because this collection is finite, \(L^2(D)\) is exactly the functions we called grids, and Lemma 5.27 shows exactly that the choice of coefficients \(\alpha_i = \frac{\int_{B_i \times B_j} f(x,y) d\mu_2}{\mu_2(B_i \times B_j)}\) gives the function in \(L^2(D)\) minimizes the \(L^2(\mu_2)\) distance to \(f\), and therefore in this case

\[
\mathbb{E}(f | D) = \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y).
\]
5.6. CONDITIONAL EXPECTATION

It is sometimes convenient to note that when we have bounds on \( f \), these bounds pass over to \( \mathbb{E}(f \mid D) \).

**Lemma 5.36.** If \( f \leq a \) almost everywhere then \( \mathbb{E}(f \mid D) \leq a \) almost everywhere.

Dually, if \( a \leq f \) almost everywhere then \( a \leq \mathbb{E}(f \mid D) \) almost everywhere.

**Proof.** We prove the first part since the second is symmetric. Let \( g = \mathbb{E}(f \mid D) \). Suppose that for some \( \epsilon > 0 \), \( \mu(\{ x \mid g(x) > a \}) = \delta > 0 \). Then let \( g'(x) = \max\{g(x), a\} \). Since \( g \) is measurable with respect to \( D \), the level set \( \{ x \mid g(x) > a \} \) is in \( D \), so \( g' \) is measurable with respect to \( D \) as well. But

\[
||f - g'||_{L^2} \leq ||f - g||_{L^2} - \epsilon^2 \delta < ||f - g||_{L^2},
\]

contradicting the minimality of \( ||f - g||_{L^2} \). \( \square \)

In particular, if \( f = \chi_B \) then \( \mathbb{E}(f \mid D) \) is bounded between 0 and 1. Slightly more generally, if \( f \in L^\infty(B) \) then \( \mathbb{E}(f \mid B) \in L^\infty(D) \).

**Lemma 5.37.** \( \mathbb{E} \) is linear:

\[
\mathbb{E}(af + bg \mid D) = a\mathbb{E}(f \mid D) + b\mathbb{E}(g \mid D).
\]

**Proof.** Consider any \( h \in L^2(D) \). Let \( z = a\mathbb{E}(f \mid D) + b\mathbb{E}(g \mid D) \) and \( h' = h - z \). Note that \( h - z \in L^2(D) \), so \( \langle af + bg - z, h - z \rangle_{L^2} = 0 \). Then

\[
||af + bg - h||_{L^2}^2 = ||(af + bg - z) - (h - z)||_{L^2}^2
= ||af + bg - z||_{L^2}^2 - 2\langle af + bg - z, h - z \rangle_{L^2} + ||h - z||_{L^2}^2
= ||af + bg - z||_{L^2}^2 + ||h - z||_{L^2}^2
\geq ||af + bg - z||_{L^2}^2,
\]

so \( z = \mathbb{E}(af + bg \mid D) \). \( \square \)

**Lemma 5.38.** \( ||\mathbb{E}(f \mid D)||_{L^2} \leq ||f||_{L^2} \)

**Proof.**

\[
||f||^2_{L^2} = \langle \mathbb{E}(f \mid D) + (f - \mathbb{E}(f \mid D)), \mathbb{E}(f \mid D) + (f - \mathbb{E}(f \mid D)) \rangle_{L^2}^2
= ||\mathbb{E}(f \mid D)||^2_{L^2} + ||f - \mathbb{E}(f \mid D)||^2_{L^2},
\]

since \( \langle \mathbb{E}(f \mid D), f - \mathbb{E}(f \mid D) \rangle_{L^2} = 0 \). \( \square \)
5.7 Conditional Expectation on Rectangles

Naturally, we want to apply the work of the previous section with the \(\sigma\)-algebras \(\mathcal{B}_{2,1} \subseteq \mathcal{B}_2\).

We should notice that projections preserve the symmetry of functions, by a similar argument to the ones we were using.

Lemma 5.39. If \(f \in L^\infty(\mu_2)\) is symmetric, so is \(\mathbb{E}(f \mid \mathcal{B}_{2,1})\).

Proof. Let \(f^+ = \mathbb{E}(f \mid \mathcal{B}_{2,1})\). Suppose not, so \(\{(x, y) \mid f^+(x, y) \neq f^+(y, x)\}\) has positive measure. Then there is some \(\delta > 0\) so that \(B = \{(x, y) \mid f^+(x, y) - f^+(y, x) > \delta\}\) has positive measure. This set belongs to \(\mathcal{B}_{2,1}\), as does \(B^{\text{op}} = \{(y, x) \mid (x, y) \in B\}\), so

\[
\int_B f^+(x, y) - f^+(y, x) \, d\mu_2 = \int_B f^+(x, y) \, d\mu_2 - \int_B f^+(y, x) \, d\mu_2 \\
= \int_B f^+(x, y) \, d\mu_2 - \int_{B^{\text{op}}} f^+(x, y) \, d\mu_2 \\
= \int_B f(x, y) \, d\mu_2 - \int_{B^{\text{op}}} f(x, y) \, d\mu_2 \\
= \int_B f(x, y) \, d\mu_2 - \int_B f(x, y) \, d\mu_2 \\
= 0.
\]

But this is a contradiction, since \(\int_B f^+(x, y) - f^+(y, x) \, d\mu_2 > \delta \mu(B)\). \(\square\)

In this case we have:

Lemma 5.40. \(\|f\|_{U^2} = 0\) if and only if \(\|\mathbb{E}(f \mid \mathcal{B}_{2,1})\|_{L^2(\mu_2)} = 0\).

Note that \(\|\mathbb{E}(f \mid \mathcal{B}_{2,1})\|_{L^2(\mu_2)} = 0\) is the same as \(\mathbb{E}(f \mid \mathcal{B}_{2,1}) = 0\)—that is, except on a set of measure 0, \(\mathbb{E}(f \mid \mathcal{B}_{2,1})(x, y) = 0\). Therefore the \(U^2\) norm is 0 exactly if the projection is the trivial function which is 0 almost everywhere.

Proof. By Theorem ??, \(\|\mathbb{E}(f \mid \mathcal{B}_{2,1})\|_{L^2} = 0\) exactly when, for every \(g \in L^2(\mathcal{B}_{2,1})\), \(\langle f, g \rangle_{L^2(\mu_2)} = 0\).

Every rectangle \(\chi_X(x)\chi_Y(y)\) is in \(L^2(\mathcal{B}_{2,1})\), so if \(\|\mathbb{E}(f \mid \mathcal{B}_{2,1})\|_{L^2(\mu_2)} = 0\) then for every rectangle \(X \times Y\), \(\langle f, \chi_X(x)\chi_Y(y) \rangle_{L^2(\mu_2)} = 0\), so by Lemma 5.21 \(\|f\|_{U^2} = 0\).

Conversely, suppose \(\|\mathbb{E}(f \mid \mathcal{B}_{2,1})\|_{L^2(\mu_2)} > 0\), so there is a \(g \in L^2(\mathcal{B}_{2,1})\) such that \(\langle f, g \rangle_{L^2(\mu_2)} > 0\). Consider sets of the form \(B_{a,b} = \{(x, y) \mid a < x < b, \ a < y < b\}\)
5.7. CONDITIONAL EXPECTATION ON RECTANGLES

\( g(x, y) \leq b \). These all belong to \( B_{2,1} \) and for any partition \((-\infty, \infty) = \bigcup_{i<k} (a_i, a_{i+1}] \)

\[
\int fg \, d\mu_2 = \sum_i \int fg \chi_{B_{a_i, a_{i+1}}} \, d\mu_2 \approx \sum_i \frac{a_i + a_{i+1}}{2} \int f \chi_{B_{a_i, a_{i+1}}} \, d\mu_2,
\]

there must be some \( a, b \) so that \(|f \chi_{B_{a,b}} \, d\mu_2| = \epsilon > 0\).

Since \( B_{a,b} \in B_{2,1} \), there must be some \( B \in B_{0,2} \) so that \( \mu_2(B_{a,b} \triangle B) < \epsilon/||f||_{L^2(\mu_2)} \), and therefore

\[
\left| \int f \chi_B \, d\mu_2 \right| > 0
\]
as well.

Since \( B \) is a finite union of rectangles, there must be some rectangle with \(|f \chi_{X \times Y} \, d\mu_2| > 0\). Therefore, by Lemma 5.21, \( ||f||_{L^2} > 0 \).

It is worth noting that one half of this proof is much harder than the other: when the \( U^2 \) norm is non-zero, this quickly implies that the \( L^2(\mu_2) \) norm of the projection is non-zero. In the other direction, it requires a lot more work to get from knowing that the \( L^2(\mu_2) \) norm of the projection is large to conclude that the \( U^2 \) norm is large. Later we will see that this reflects a quantitative fact: knowing that the \( U^2 \) norm is larger than some \( \epsilon > 0 \) will tell us that the \( L^2(\mu_2) \) norm of the projection must also be larger than some \( \delta > 0 \) which we can calculate from \( \epsilon \). But in the reverse direction, we will see examples where even though the \( L^2(\mu_2) \) norm of the projection is as large as we want—say, \( 1 \)—the \( U^2 \) norm can be arbitrarily small.

**Lemma 5.41.** \( f \) belongs to \( L^2(B_{2,1}) \) if and only if \( f \) is dual to \( U^2 \).

**Proof.** Suppose \( ||r||_{U^2} = 0 \) and \( |\int fr \, d\mu_2| = \epsilon > 0 \). We may choose \( g = \sum_{i \leq k} \alpha_i \chi_{X_i}(x) \chi_{Y_i}(y) \) so that \( ||f - g||_{L^2(\mu_2)} < \frac{\epsilon}{2||r||_{L^2(\mu_2)}} \).

\[
|\int gr \, d\mu_2| = \int (f + (g - f))r \, d\mu_2
= \int fr \, d\mu_2 + \int (g - f)r \, d\mu_2
> \epsilon - \frac{\epsilon}{2||r||_{L^2(\mu_2)}} ||r||_{L^2(\mu_2)}
> \epsilon/2.
\]
CHAPTER 5. STRUCTURE AND RANDOMNESS

But

\[ \int gr \, d\mu_2 = \sum_{i \leq k} \alpha_i \int \chi_{X_i}(x) \chi_{Y_i}(y) \, d\mu_2 = 0, \]

a contradiction.

Suppose \( f \) is dual to \( U^2 \). Let \( f^+ = E(f \mid B_{2,1}) \) and \( f^- = f - f^+ \). Then

\[ 0 = \int f f^- \, d\mu_2 = \int f^+ f^- \, d\mu_2 + \int f^- f^- \, d\mu_2 = 0 + ||f^-||_{L^2(\mu_2)}^2. \]

So \( ||f^-||_{L^2(\mu_2)} = 0 \), so \( f = f^+ \), so \( f \) is measurable with respect to \( B_{2,1} \).

So, for our purposes, a function is “structured”—totally non-random—if it belongs to \( L^2(B_{2,1}) \). These are exactly the functions approximated by grids we discussed above.

The \( U^2 \) norm \( ||f||_{U^2} \) is positive exactly when \( f \) has some correlation with \( B_{2,1} \). However it is not accurate to say that the \( U^2 \) norm measures how much \( f \) correlates with \( B_{2,1} \); in fact, even if \( f \in L^2(B_{2,1}) \), so \( f \) is entirely described by \( B_{2,1} \), the \( U^2 \) norm can take on any value between \( | \int f \, d\mu_2 | \) and \( ||f||_{L^2(\mu_2)} \).

Instead, the \( U^2 \) norm reflects something about the “complexity” of \( f \). For example, consider graphs \( E \) with \( \mu_2(E) = 1/2 \). We always have \( ||\chi_E||_{U^2} \geq 1/2 \) in this case, so it makes sense to focus on the \( U^2 \) norm of \( f(x, y) = \chi_E(x, y) - 1/2 \), which counts how many “extra” rectangles the graph \( E \) has, beyond the ones it must have. When \( E \in B_{2,1} \), \( f \in L^2(B) \), so we will have \( ||f||_{U^2} > 0 \).

When \( E \) can be described using a small number of rectangles, \( ||f||_{U^2} \) will be larger. For example, consider the case where \( E \) is the bipartite graph \( B_1 \times B_2 \cup B_2 \times B_1 \) with \( \mu(B_1) = \mu(B_2) = 1/2 \). This can be represented by the grid

which only has four “boxes”. The \( U^2 \) norm of \( f \) is 1/2.

On the other hand, suppose \( E \) comes from a grid like

where the boxes are filled in randomly. Again, let \( f(x, y) = \chi_E(x, y) - 1/2 \). Then \( ||f||_{U^2} \) is smaller—roughly 0.279 (depending on the exact distribution of squares). And, as the number of squares increases, the \( U^2 \) norm continues
to decrease. (If we fill in a \( k \times k \) grid randomly with black and white boxes to get the graph \( E \), the \( U^2 \) norm of \( \chi_E - 1/2 \) will be roughly \( \frac{1}{29/4^{3/4}} \).)

Note that \( ||f||_{L^2(\mu_2)} = ||\chi_E - 1/2||_{L^2(\mu_2)} = 1/2 \), so \( \chi_E \) is always “far” from the constantly 1/2 function in the \( L^2(\mu_2) \) norm. However, when \( E \) is built out of many rectangles—when \( E \) is “complicated”—\( \chi_E \) is close to 1/2 in the \( U^2 \) norm.

To see this, consider the \( 100 \times 100 \) grid:

```
Visually, the grid is beginning to blur to grey. A finer grid, say 1000 \times 1000, would be indistinguishable to the eye from a constant grey. There is some sense that a graph built from a very large number of small, randomly chosen squares is close to the function which is constantly 1/2. The norm which reflects that these are close is the \( U^2 \) norm.

More generally, suppose we begin with a graph \( E \) which may or may not belong to \( \mathcal{B}_{2,1} \), but where \( ||\chi_E||_{U^2} \) is just a bit larger than \( \mu_2(E) \). Since \( ||\chi_E||_{U^2} \geq \mu_2(E) \), this means that \( E \) is not quasirandom. We can decompose \( \chi_E = \mathbb{E}(\chi_E | \mathcal{B}_{2,1}) + (\chi_E - \mathbb{E}(\chi_E | \mathcal{B}_{2,1})) \). It could be that \( ||\chi_E||_{U^2} \) is “not too big” because the structured part \( ||\mathbb{E}(\chi_E | \mathcal{B}_{2,1})||_{L^2(\mu_2)} \) is itself not too big. It could also be that \( E \) is entirely, or almost entirely, structured—that \( ||\mathbb{E}(\chi_E | \mathcal{B}_{2,1})||_{L^2(\mu_2)} \) is close to \( \mu_2(E) \)—but that \( \mathbb{E}(\chi_E | \mathcal{B}_{2,1}) \) is very complicated.

5.8 Which Rectangles are Needed

Given \( f \in L^2(\mu_2) \), the question of whether \( ||\mathbb{E}(f | \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0 \) seems like it ought to depend on the particular choice of \( \sigma \)-algebra \( \mathcal{B}_1 \). For instance, if we replace \( \mathcal{B}_1 \) with some \( \mathcal{B}'_1 \subseteq \mathcal{B}_1 \), fewer functions are measurable with respect to \( \mathcal{B}'_{2,1} = \mathcal{B}'_1 \times \mathcal{B}'_1 \). We might think that a function \( f \) could be measurable with respect to \( \mathcal{B}_{2,1} \) but not \( \mathcal{B}'_{2,1} \).

Yet our work above shows that this is not the case—\( ||\mathbb{E}(f | \mathcal{B}_{2,1})||_{L^2(\mu_2)} \neq 0 \) if and only if \( ||f||_{U^2} \neq 0 \), and the integral \( \int f(x, y) f(x', y) f(x, y') f(x', y') \, d\mu_4 \) does not depend on the particular \( \sigma \)-algebra \( \mathcal{B}_1 \). The explanation is that the definition of a Keisler graded probability space forces certain sets to belong to \( \mathcal{B}_1 \), and these must be sufficient to express the projection of \( f \) onto \( \mathcal{B}_{2,1} \).

**Lemma 5.42.** Let \( f \in L^\infty(\mu_2) \) with \( ||f||_{U^2} > 0 \) and let \( \mathcal{B} \subseteq \mathcal{B}_1 \) be any \( \sigma \)-algebra large enough that, for almost every \( a \in V \), we have the property
for every interval $I \subseteq \mathbb{R}$, \{ $x \mid f(a, x) \in I$ $\}$ and \{ $x \mid f(x, a) \in I$ $\}$ are both in $B$.

Then $||E(f \mid B \times B)||_{L^2(\mu_2)} > 0$.

Proof. Without loss of generality, we assume that \{( $x, y$) $\mid |f(x, y)| \leq 1$ $\}$ has measure 1. (If not, there is some $c > 0$ so that $f/c$ has this property, and we can work with $f/c$ instead.)

Since $0 < ||f||_{U^2}^4 = \int f(x, y) \left[ \int f(x, y') f(x', y) f(x', y') d\mu_2(x', y') \right] d\mu(x, y)$, there must be a set of $x', y'$ of positive measure so that $\left| \int f(x, y) f(x, y') f(x', y) d\mu_2(x, y) \right| = \epsilon > 0$.

We can choose some $x', y'$ whose level sets are in $B$.

Choosing $n > 1/\epsilon$, we may divide $[-1, 1]$ into intervals, $[-1, 1] \subseteq \bigcup_{i \leq n} [-1 + i \epsilon/2, -1 + (i + 1)\epsilon/2]$. For each $i \leq n$, let $A_i = \{ x \mid f(x, y') \in [-1 + i \epsilon/2, -1 + (i + 1)\epsilon/2] \}$ and $B_i = \{ y \mid f(x', y) \in [-1 + i \epsilon/2, -1 + (i + 1)\epsilon/2] \}$.

Then

$$
\epsilon = \left| \int f(x, y) f(x, y') f(x', y) d\mu_2(x, y) \right|
= \left| \sum_{i,j \leq n} \int_{A_i \times B_j} f(x, y) f(x, y') f(x', y) d\mu_2(x, y) \right|
\leq \sum_{i,j \leq n} \int_{A_i \times B_j} f(x, y)(-1 + i \epsilon/2)(-1 + j \epsilon/2) d\mu_2(x, y) + \epsilon/2.
$$

In particular, there must be some rectangle $A_i \times B_j \in B \times B$ such that $|\int_{A_i \times B_j} f(x, y) d\mu_2| > 0$.

Theorem 5.43. Let $f \in L^\infty(\mu_2)$ be given and let $B \subseteq B_1$ be any $\sigma$-algebra containing such that, for almost every $a \in V$,

for every interval $I \subseteq \mathbb{R}$, the sets \{ $x \mid f(a, x) \in I$ $\}$ and \{ $x \mid f(x, a) \in I$ $\}$.

Then $E(f \mid B \times B) = E(f \mid B_{2,1})$. 

\[ \square \]
5.9. **ROTH’S THEOREM**

Proof. Let \( f^- = f - \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B}) \).

Since \( \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B}) \) is \( \mathcal{B} \times \mathcal{B} \)-measurable, it must be the case (by Fubini’s Theorem) that for every \( a \in \mathcal{V} \) and every interval \( I \), the sets \( \{ x \mid f(a, x) \in I \} \) and \( \{ x \mid f(x, a) \in I \} \) are in \( \mathcal{B} \). So for almost every \( a \), it is also the case that \( \{ x \mid f^-(a, x) \in I \} = \{ x \mid f(a, x) - \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B})(a, x) \in I \} \) is in \( \mathcal{B} \), and similarly for \( \{ x \mid f^-(x, a) \in I \} \).

If \( \mathbb{E}(f \mid \mathcal{B} \times \mathcal{B}) \neq \mathbb{E}(f \mid \mathcal{B}_2) \) then \( \|\mathbb{E}(f^- \mid \mathcal{B}_2)\|_{L^2(\mu_2)} \neq 0 \), so by the previous lemma, \( \|\mathbb{E}(f^- \mid \mathcal{B} \times \mathcal{B})\|_{L^2(\mu_2)} \neq 0 \), which is a contradiction. \( \Box \)

**5.9 Roth’s Theorem**

We can now prove Roth’s Theorem—that is, Szemerédi’s Theorem for arithmetic progressions of length 3, which we state as Theorem 5.48 below.

First, we note that subgraph density depends only on \( \mathcal{B}_2 \)—that is, if we want to examine how dense a subgraph is, we can focus only on the projection to \( \mathcal{B}_2 \).

**Lemma 5.44.** For any symmetric \( f \in L^\infty(\mu_2) \), \( t_H(f) = t_H(\mathbb{E}(f \mid \mathcal{B}_2)) \).

**Proof.** Let \( H = (W, F) \). Let \( f^+ = \mathbb{E}(f \mid \mathcal{B}_2) \) and \( f^- = f - f^+ \). Then

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

We successively replace each \( f \) in this product with \( f^+ \): for \( S \subseteq F \), we show by induction on \( |S| \) that

\[
  t_H(f) = \int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S} f(x_i, x_j) \, d\mu_{|W|}.
\]

When \( S = \emptyset \), this is the definition of \( t_H(f) \). Suppose \( S = S' \cup \{(v, v')\} \). Then

\[
  t_H(f) = \int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f(v, v') \, d\mu_{|W|}
  = \int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^+(v, v') \, d\mu_{|W|}
  + \int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^-(v, v') \, d\mu_{|W|},
\]

where we apply the induction hypothesis to \( f^+(v, v') \) and \( f^-(v, v') \) respectively.
so it suffices to show that

\[
\int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^-(v, v') \, d\mu_{|W|} = 0.
\]

Here we use Fubini’s Theorem:

\[
\int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^-(v, v') \, d\mu_{|W|} = \int \int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^-(v, v') \, d\mu(x_i, x_j) d\mu_{|W|} = 0.
\]

For any fixed \(|W| = 2\) tuple \((x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{|W|})\), \(\prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j)\) belongs to \(B_{2,1}\) — in fact, we can decompose this function into the form

\[
c g_0(x) g_1(x').
\]

Therefore

\[
\int \prod_{\{v_i,v_j\} \in S} f^+(x_i, x_j) \prod_{\{v_i,v_j\} \in F \setminus S'} f(x_i, x_j) f^-(v, v') \, d\mu(x_i, x_j) = 0.
\]

\[\square\]

The same argument would apply to induced subgraph density.

**Definition 5.45.** When \(f \in L^\infty(\mu_2)\) is symmetric and \(H = (W, F)\) with \(W = \{w_1, \ldots, w_k\}\),

\[
t_{H}^{\text{ind}}(f) = \int \prod_{\{w_i,w_j\} \in F} f(x_i, x_j) \prod_{\{w_i,w_j\} \in (\mathcal{W}) \setminus F} (1 - f(x_i, x_j)) \, d\mu_k.
\]

**Lemma 5.46.** For any symmetric \(f \in L^\infty(\mu_2)\), \(t_H^{\text{ind}}(f) = t_H^{\text{ind}}(\mathbb{E}(f \mid B_{2,1}))\).

We can now set out to prove Roth’s Theorem, the \(k = 3\) case of Szemerédi’s Theorem.

We will follow the plan introduced in Chapter 1: the main part will be showing that certain graphs have many triangles.

**Theorem 5.47.** Suppose \(t_{C_3}(E) = 0\). Let \(B \subseteq B_2\) be a dense algebra. Then for every \(\epsilon > 0\) there is a \(B \in \mathcal{B}\) so that \(\mu_2(B) < \epsilon\) and

\[
T_{C_3}(E \setminus B) = \emptyset.
\]
5.9. ROTH’S THEOREM

That is, if there are very few triangles, we can remove a set that is as small as we like and thereby remove all triangles. We will be interested in the case where $B$ is the internal sets.

**Proof.** Let $f^+ = E(E \mid B_{2,1})$. We have $t_{C_3}(f^+) = t_{C_3}(E) = 0$ by the previous lemma. Let $E^+ = \{(x, y) \mid f^+(x, y) > 0\}$; then we must also have $t_{C_3}(E^+) = 0$. Also

$$
\mu_2(E \setminus E^+) = \int \chi_E(x, y)(1 - \chi_{E^+}(x, y)) \, d\mu_2
$$

$$
= \mu_2(E) - \int E(\chi_E \mid B_{2,1}) \chi_{E^+} \, d\mu_2
$$

$$
= \mu_2(E) - \int E(\chi_E \mid B_{2,1}) \, d\mu_2
$$

$$
= \mu_2(E) - \int \chi_E \, d\mu_2
$$

$$
= 0.
$$

By Theorem 5.29, we may choose positive measure sets $\{B_i\}_{i \leq d}$ in $B$ so that $V = \bigcup_{i \leq d} B_i$ and, for suitable constants,

$$
||\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x)\chi_{B_j}(y)||_{L^2(\mu_2)} < \frac{\sqrt{\epsilon}}{4}.
$$

Indeed, we may assume that $\alpha_{i,j} = \frac{\mu_2((B_i \times B_j) \cap E^+)}{\mu_2(B_i \times B_j)}$. Let $E' = \bigcup_{i,j \leq d, \alpha_{i,j} > 3/4} B_i \times B_j$.

If $(x, y) \in E^+ \setminus E'$ then $\chi_{E^+}(x, y) = 1$ while $\alpha_{i,j} \leq 3/4$, and therefore $\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x)\chi_{B_j}(y) \geq 1/4$. Therefore

$$
\mu_2(E^+ \setminus E') \leq 16 \int (\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x)\chi_{B_j}(y))^2 \, d\mu_2
$$

$$
= 16||\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x)\chi_{B_j}(y)||^2_{L^2(\mu_2)}
$$

$$
< \epsilon,
$$

so $\mu_2(E^+ \setminus E') < \epsilon$, so $\mu_2(E \setminus E') < \epsilon$ as well.

We will let $B = E \setminus E'$, so $E \setminus B = E \cap E' \subseteq E'$. Towards a contradiction, suppose $T_{C_3}(E') \neq 0$. $E'$ is a union of rectangles from the grid $\{B_i\}_{i \leq d}$, so there must be some triangle $(i, j), (j, k), (i, k)$ so that $B_i \times B_j \times B_k \subseteq T_{C_3}(E')$, and therefore $t_{C_3}(E') \geq \mu(B_i)\mu(B_j)\mu(B_k) > 0$. 
Since \( \mu_2(E^+ \cap (B_i \times B_j)) > 3/4 \), at most \( \frac{1}{4} \mu(B_i) \mu(B_j) \mu(B_k) \) of the triangles in \( B_i \times B_j \times B_k \) do not have \((x, y) \in E^+ \). Similarly, at most \( \frac{1}{4} \mu(B_i) \mu(B_j) \mu(B_k) \) fail to have \((x, z) \in E^+ \) and at most \( \frac{1}{4} \mu(B_i) \mu(B_j) \mu(B_k) \) fail to have \((y, z) \in E^+ \).

So \( \ell_{C_3}(E^+) \geq \frac{1}{4} \mu(B_i) \mu(B_j) \mu(B_k) > 0 \), which contradicts the assumption. So \( T_{C_3}(E') = \emptyset \), and therefore \( B = E \setminus E' \) is the promised set. \( \square \)

**Theorem 5.48** (Roth’s Theorem). For every \( \epsilon > 0 \), there is an \( N \) so that whenever \( n \geq N \) and \( A \subseteq \{1, 2, \ldots, n\} \) is a set with \( |A| / n \geq \epsilon \), there is an \( a \in A \) and a \( d > 0 \) such that \( a, a + d, a + 2d \in A \).

**Proof.** Suppose not. Then there is some \( \epsilon > 0 \) which is a counterexample: for every \( N \) there is an \( n > N \) and an \( A_n \subseteq \{1, 2, \ldots, n\} \) with \( |A_n| / n \geq \epsilon \) but \( A_n \) contains no arithmetic progression \( a, a + d, a + 2d \).

We define graphs \( G_n = (V_n, E_n) \) as follows:

- \( V_n \) is the disjoint union of \( X_n, Y_n, \) and \( Z_n \) where \( X_n = Y_n = Z_n = \{1, 2, \ldots, 3n\} \),
- \((x, y) \in X_n \times Y_n \) belongs to \( E_n \) if \( x + 2y \mod 3n \in A_n \),
- \((x, z) \in X_n \times Z_n \) belongs to \( E_n \) if \( 2z - x \mod 3n \in A_n \),
- \((y, z) \in Y_n \times Z_n \) belongs to \( E_n \) if \( z + y \mod 3n \in A_n \).

We go up to \( 3n \) rather than \( n \) to avoid some minor technical issues where the modulus would cause fake progressions from overlowing past \( n \).

Suppose \((x, y, z) \in T_{C_3}(E_n) \). Let \( a = x + 2y \mod 3n \) and \( d = z - (x + y) \mod 3n \). Then \( a \in A_n \); in particular, \( a \leq n \). Also \( a + d = x + 2y + z - x - y = y + z \mod 3n \in A_n \), so \( a + d \mod 3n \leq n \) as well. Similarly, \( a + 2d \mod 3n \in A_n \). If \( d > 0 \) then \( d \mod n \) gives an arithmetic progression.

However if \( d = 0 \) then \( z = x + y \mod 3n \). Let \( T_n = \{(x, y, x + y \mod 3n) \mid x + 2y \in A_n \} \) (the “trivial” triangles).

For any fixed \( y \in Y_n \), \( |\{x \mid x - 2y \mod 3n \in A_n \}| = |A_n| \geq \epsilon n^2 \), so

\[
\frac{|E_n \cap (X_n \times Y_n)|}{9n^2} \geq \frac{\epsilon}{9}.
\]

Similarly, \( \frac{|E_n \cap (Y_n \times Z_n)|}{9n^2} \geq \frac{\epsilon}{9} \) and \( \frac{|E_n \cap (X_n \times Z_n)|}{9n^2} \geq \frac{\epsilon}{9} \), so

\[
\frac{|E_n|}{9n^2} \geq \frac{\epsilon}{3}.
\]
5.10. MORE APPLICATIONS

Relatedly, we need one additional structure on the finite sets: consider the functions \( \rho_n^x : X_n \times Y_n \to X_n \times Z_n \) and \( \rho_n^y : X_n \times Y_n \to Y_n \times Z_n \) given by \( \rho_n^x(x, y) = (x, x + y \mod 3n) \) and \( \rho_n^y(x, y) = (y, x + y \mod 3n) \). Note that both these maps are one-to-one, and so in particular measure-preserving.

Pick any nonprincipal ultrafilter \( \mathcal{U} \) and let \( G = [G_n]_\mathcal{U} \) and \( E = [E_n]_\mathcal{U} \). We also set \( X = [X_n]_\mathcal{U}, Y = [Y_n]_\mathcal{U}, Z = [Z_n]_\mathcal{U}, \rho^x = [\rho_n^x]_\mathcal{U}, \) and \( \rho^y = [\rho_n^y]_\mathcal{U}. \)

Suppose \( t_{C_3}(E) = 0 \). Then there is an internal set \( B = [B_n]_\mathcal{U} \) with \( \mu_2(E \setminus B) < \epsilon/12 \) such that \( T_{C_3}(E \setminus B) = 0. \)

It will be more convenient to divide \( B \) into the three pieces corresponding to the three pieces of \( E \): set \( B_{XY} = B \cap (X \times Y), B_{XZ} = B \cap (X \times Z), \) and \( B_{YZ} = B \cap (Y \times Z) \). Consider those \( (x, y) \in X \times Y \) such that

- \( (x, y) \in E, \)
- \( (x, y) \notin E \cap B_{XY}, \)
- \( (x, x + y) \notin E \cap B_{XZ}, \)
- \( (y, x + y) \notin E \cap B_{YZ}. \)

More formally, this is the set

\[
E^- = (E \cap (X \times Y)) \setminus (B_{XY} \cup (\rho^x)^{-1}(B_{XZ}) \cup (\rho^y)^{-1}(B_{YZ})).
\]

Note that when \( (x, y) \in E \cap (X \times Y), \) also \( \rho^x(x, y) \in E \) and \( \rho^y(x, y) \in E \) (by the definition of \( E \) above), so if \( (x, y) \) is in \( E \setminus E^- \) then either \( (x, y) \in E \setminus B_{XY}, \rho^x(x, y) \in E \setminus B_{XZ}, \) or \( \rho^y(x, y) \in E \setminus B_{YZ}. \) Therefore

\[
\mu_2(E^-) \geq \mu^2(E) - \mu^2(B) > \epsilon/3 - 3\epsilon/4 > \epsilon/12.
\]

Therefore \( E^- \) is non-empty, so there is an \( (x, y) \in E^- \), so \( (x, y, x + y) \in T_{C_3}(E \setminus B) \), which is a contradiction.

So \( t_{C_3}(E) > 0. \) Since \( \mu_3(T) = \lim_{n \to \infty} \frac{|T_n|}{27n^3} < \lim_{n \to \infty} \frac{n^2}{27n^3} = 0, \) there is an \( ([x_n]_\mathcal{U}, [y_n]_\mathcal{U}, [z_n]_\mathcal{U}) \in T_{C_3}(E) \setminus T. \) Therefore for some (indeed, almost every) \( n, \) there is an \( (x_n, y_n, z_n) \in E_n \setminus T_n, \) giving an arithmetic progression in \( A_n, \) which is a contradiction. \( \square \)

5.10 More Applications

Theorem 5.47 generalizes to arbitrary graphs with only notational changes.
Theorem 5.49. Let $H = (W, F)$ be a finite graph and suppose $t_H(E) = 0$. Let $\mathcal{B} \subseteq \mathcal{B}_2$ be a dense algebra. Then for every $\epsilon > 0$ there is a $B \in \mathcal{B}$ so that $\mu_2(B) < \epsilon$ and

$$T_H(E \setminus B) = \emptyset.$$ 

Proof. Let $f^+ = \mathbb{E}(E \mid \mathcal{B}_2)$. By Lemma 5.44, $t_H(f^+) = t_H(E) = 0$, so letting $E^+ = \{(x, y) \mid f^+(x, y) > 0\}$, also $t_H(E^+) = 0$ and $\mu_2(E \setminus E^+) = 0$.

By Theorem 5.29 we may choose a partition $V = \bigcup_{i \leq d} B_i$ so that each $B_i \in \mathcal{B}$, $\mu_1(B_i) > 0$, and, taking $\alpha_{i,j} = \frac{\mu((B_i \times B_j) \cap E^+)}{\mu_2(B_i \times B_j)}$,

$$\|\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)\|_{L^2(\mu_2)} \leq \frac{\sqrt{\epsilon}}{\sqrt{|F|+1}}.$$ 

Set $E' = \bigcup_{i,j \leq d, \alpha_{i,j} > -\frac{1}{|F|+1}} B_i \times B_j$. If $(x, y) \in E^+ \setminus E'$ then $\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) > -\frac{1}{|F|+1}$, so

$$\mu_2(E^+ \setminus E') \leq (|F|+1) \int \chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y) d\mu_2 = (|F|+1) \|\chi_{E^+} - \sum_{i,j \leq d} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)\|_{L^2(\mu_2)}^2 < \epsilon.$$ 

So we take $B = E \setminus E'$, and therefore $E \setminus B = E \cap E' \subseteq E'$. Suppose $T_H(E') \neq \emptyset$, so there is a copy $\pi : W \rightarrow V$ of $H$ in $(V, E')$. For each $w \in W$, there is an atom $B_w \in \mathcal{B}'$ so that $w \in B_w$. Then for every $\pi' : W \rightarrow V$ such that $\pi'(w) \in B_w$, $\pi'$ must be a copy of $H$, so $t_H(E') \geq \prod_{w \in W} \mu_1(B_w) > 0$.

We now show that $t_H(E^+) \geq \frac{1}{|F|+1} t_H(E')$, which gives the desired contradiction. Fix an edge $e = (w, w') \in F$ and consider the copies $\pi$ of $H$ in $(V, E')$ such that $(\pi(w), \pi(w')) \notin E^+$. This is

$$\int \chi_{E'}(1 - \chi_{E^+}) \prod_{f \in F \setminus \{e\}} \chi_{E'} d\mu|W|.$$ 

Since all the copies of $\chi_{E'}$ are measurable with respect to $\mathcal{B}'$, this is equal to

$$\int \chi_{E'}(1 - \mathbb{E}(\chi_{E^+} \mid (\mathcal{B}')^2)) \prod_{f \in F \setminus \{e\}} \chi_{E'} d\mu|W|.$$ 

By the definition of $E'$, this is

$$\leq \frac{1}{|F|+1} t_H(E').$$
If \( \pi \) is a copy of \( H \) in \((V, E')\) but not \((V, E^+)\) then there must be some such edge, so the set of such \( \pi \) has measure
\[
\leq \sum_{e \in F} \frac{1}{|F| + 1} t_H(E') = \frac{|F|}{|F| + 1} t_H(E').
\]
That leaves at least \( \frac{1}{|F| + 1} t_H(E') \) copies which must also be copies in \((V, E^+)\), so \( t_H(E^+) \geq \frac{1}{|F| + 1} t_H(E') > 0 \). This gives the desired contradiction. \( \square \)

There is a finite analog of this theorem, known as graph removal.

**Corollary 5.50** (Graph Removal). For every \( \epsilon > 0 \) and every finite graph \( H = (W, F) \), there is a \( \delta > 0 \) so that whenever \( G = (V, E) \) is a graph with \( t_H(G) < \delta \), there is a \( B \subseteq \binom{V}{2} \) so that \( \frac{|B|}{\binom{|V|}{2}} < \epsilon \) and \( T_H(E \setminus B) = \emptyset \).

**Proof.** Suppose not, so for some \( \epsilon > 0 \) and some finite graph \( H = (W, F) \), for each \( n \) there is a \( G_n = (V_n, E_n) \) with \( t_H(G_n) < 1/n \) but whenever \( B \subseteq \binom{V}{2} \) with \( \frac{|B|}{\binom{|V|}{2}} < \epsilon \), \( T_H((V_n, E_n \setminus B)) \neq \emptyset \).

Note that \( t_H(G_n) < 1/n \) and \( T_H(G_n) \neq \emptyset \) implies that \( |V_n| > n \), so \( \lim_{n \to \infty} |V_n| = \infty \).

Consider any ultraproduct \( [G_n]_U = (V, E) \) and the corresponding Keisler graded probability space. Let \( B \subseteq B_2 \) be the internal sets. Then there is a \( B \in B \) with \( \mu_2(B) < \epsilon \) and \( T_H(E \setminus B) = \emptyset \). Since \( B \) is internal, \( B = [B_n]_U \). In particular, we may choose an \( n \) with \( \frac{|B_n|}{|V_n|} < \epsilon \) and such that \( T_H(E_n \setminus B_n) = \emptyset \), contradicting the assumption. \( \square \)

Frequently Theorem 5.29 is combined with the observation that most pairs in the resulting approximation are “regular”: recall the notion of \( \epsilon \)-regularity, as specialized to a bipartite graph:

If \( E \subseteq X \times Y \), the triple \((X, Y, E)\) is \( \epsilon \)-regular if whenever \( X' \subseteq X \) and \( Y' \subseteq Y \) with \( \mu_1(X') \geq \epsilon \mu_1(X) \) and \( \mu_1(Y') \geq \epsilon \mu_1(Y) \),
\[
|d_E(X', Y') - d_E(X, Y)| < \epsilon.
\]

**Lemma 5.51.** If \( f \in L^\infty(B_2, \lambda) \) and \( B \subseteq B_1 \) is a dense algebra then, for every \( \epsilon > 0 \), there is a partition \( V = \bigcup_{i \leq k} B_i \) with each \( B_i \in B \), \( \mu_1(B_i) > 0 \), and such that \( \mu_2(\bigcup_{i \leq k} B_i \times B_j, E) \) is not \( \epsilon \)-regular \( B_i \times B_j \) < \( \epsilon \).

**Proof.** By Theorem 5.29 choose a a \( V = \bigcup_{i \leq k} B_i \) so that each \( B_i \in B \), \( \mu_1(B_i) > 0 \), and, taking \( \alpha_{i,j} = \frac{\mu_2(B_i \times B_j \cap E)}{\mu_2(B_i \times B_j)} \),
\[
||E(\chi_{B_i} | B_2, \lambda) - \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)||_{L^2(\mu_2)} < \epsilon.
\]
Let $S_0$ be the algebra of sets generated by $\{B_i\}_{i,j \leq k}$ and let $B_0$ be the algebra of sets generated by rectangles from $S_0$, so $E(\chi_E \mid B_0) = \sum_{i,j \leq k} \alpha_{i,j} \chi_{B_i}(x) \chi_{B_j}(y)$.

Consider some pair $i, j$ so that $(B_i, B_j, E)$ is not $\epsilon$-regular, and choose $X_{i,j} \subseteq B_i$ and $Y_{i,j} \subseteq B_j$ so that $|E(X_{i,j}, Y_{i,j}) - \alpha_{i,j}| \geq \epsilon$. Let $S_{i,j}$ be the algebra of sets generated by $S_0$ together with $\{X_{i,j}, Y_{i,j}\}$ and let $B_{i,j}$ be the algebra of sets generated by rectangles from $S_{i,j}$. In particular, the atoms of $B_{i,j}$ are the same as those of $B_0$ except that $B_i \times B_j$ has been split into four pieces---$X_{i,j} \times Y_{i,j}$, $(B_i \setminus X_{i,j}) \times Y_{i,j}$, $X_{i,j} \times (B_j \setminus Y_{i,j})$, and $(B_i \setminus X_{i,j}) \times (B_j \setminus Y_{i,j})$. Therefore,

$$||E(\chi_E \mid B_{i,j}) - E(\chi_E \mid B_0)||^2_{L^2(\mu_2)} \geq (d_E(X_{i,j}, Y_{i,j}) - \alpha_{i,j})^2 \mu_1(X_{i,j}) \mu_1(Y_{i,j}) \geq \epsilon^4 \mu(B_i) \mu(B_j).$$

(In the first line, we are discarding the other three sub-rectangles since we only care about the inequality.)

Let $Z = \{(i,j) \mid (B_i, B_j, E) \text{ is not } \epsilon\text{-regular}\}$. Let $S_1$ be the algebra generated by $\bigcup_{(i,j) \in Z} S_{i,j}$ and let $B_1$ be the algebra of sets generated by rectangles from $S_1$. Then, by our choice of $\{B_i\}_{i \leq k}$,

$$\epsilon^5 > ||E(\chi_E \mid S_1) - E(\chi_E \mid S_0)||^2_{L^2(\mu_2)}$$

$$= \sum_{i,j \leq k} \int_{B_i \times B_j} (E(\chi_E \mid S_1) - E(\chi_E \mid S_0))^2 \, d\mu_2$$

$$\geq \sum_{(i,j) \in Z} \int_{B_i \times B_j} (E(\chi_E \mid S_{i,j}) - E(\chi_E \mid S_0))^2 \, d\mu_2$$

$$\geq \epsilon^4 \sum_{(i,j) \in Z} \mu(B_i) \mu(B_j).$$

A variant of the Turán-type theorems discussed in the previous chapter is to ask what subgraphs appear in a graph with certain edge densities as well as additional properties.

**Theorem 5.52.** Suppose that $G = (V, E)$ is an atomless measurable graph with $t_{K_2}(G) > 1/8$ such that whenever $X \subseteq V$ with $\mu(X) > 0$, $E \cap (X \times X) \neq \emptyset$. Then $T_{K_4}(G) \neq \emptyset$.

**Proof.** The idea is that we will look at a partition into blocks, taking advantage of the fact every part must contain edges within it. If we find two parts $B_i, B_j$ where $d_E(B_i, B_j) > 1/2$ then we will be able to find a copy of $K_4$ with
two vertices in each part. Otherwise all parts have densities \( \leq 1/2 \), and we will use Mantel’s Theorem to find three parts \( B_i, B_j, B_k \) which form a triangle and have all three densities \( d_E(B_i, B_j), d_E(B_i, B_j), d_E(B_j, B_k) > \epsilon/2 \). Then we will find a copy of \( K_4 \) with one vertex each in \( B_i, B_j \) and two in \( B_k \).

Let \( \epsilon = t_{K_4}(G) - 1/8 > 0 \). By the preceding lemma, we may choose a partition \( \{B_i\}_{i \leq d} \) such that \( \mu_2(\bigcup_{i,j \leq k} (B_i, B_j, E)) \) is not \( \epsilon/6 \)-regular \( B_i \times B_j) < \epsilon/6 \).

First, suppose there is an \( i, j \) with \( \alpha_{i,j} > 1/2 \). There must be some \( \delta > 0 \) so that \( \alpha_{i,j} \geq 1/2 + \delta \), so consider \( B' = \{x \in B_i \mid \mu_1(\{y \in B_j \mid \{x, y\} \in E\}) \geq (1/2 + \delta)\mu_1(B_j)\} \). We must have \( \mu_1(B') > 0 \), so we may choose \( x_1, x_2 \in B' \) with \( \{x_1, x_2\} \in E \). Since more than half of \( B_j \) is a neighbor of \( x_1 \) and more than half is a neighbor of \( x_2 \), \( \mu_1(\{y \in B_j \mid \{x_1, y\} \in E \} \} > 0 \), so we may find \( y_1, y_2 \) with \( \{y_1, y_2\} \in E \) so that both are neighbors of both \( x_1 \) and \( x_2 \). Then \( x_1, x_2, y_1, y_2 \) is a copy of \( K_4 \).

So suppose not—for every \( i, j \), \( \alpha_{i,j} \leq 1/2 \). Let \( R \) be the set of pairs \( (i,j) \) so that \( (B_i, B_j, E) \) is \( \epsilon/6 \)-regular and \( \alpha_{i,j} > \epsilon/2 \). Let \( E^* \) be the graph given by \( E^* = \bigcup_{(i,j) \in R} B_i \times B_j \). That is, there is an edge in \( E^* \) between \( x \) and \( y \) if they belong to a regular pair with density which is not “too small”. This graph must have density \( \geq 1/4 + \epsilon/2 \): since the original graph had density \( > 1/8 \) and a \( \mu_2((B_i \times B_j) \cap E) \leq \frac{1}{2} \mu_2((B_i \times B_j) \cap E^*) \) for each pair \( i, j \), and we lose measure at most \( \epsilon/2 \) by discarding the pairs outside of \( R \). So, by Mantel’s Theorem, this graph contains a triangle—there are parts \( B_{i_1}, B_{i_2}, B_{i_3} \) (not necessarily distinct) so that \( \alpha_{i_1,i_2}, \alpha_{i_1,i_3}, \alpha_{i_2,i_3} > \epsilon/2 \).

Choose \( x \in B_{i_1} \) so that \( \mu_1(\{y \in B_{i_2} \mid \{x, y\} \in E\}) \geq \epsilon/3 \) and \( \mu_1(\{z \in B_{i_3} \mid \{x, z\} \in E\}) \geq \epsilon/3 \). Let \( C = \{y \in B_{i_2} \mid \{x, y\} \in E\} \) and \( D = \{z \in B_{i_3} \mid \{x, z\} \in E\} \). Then \( \|d_E(C, D) - \alpha_{i_2,i_3}\| < \epsilon/6 \), so in particular we may choose a \( y \in C \) so that \( \mu_1(\{z \in D \mid \{y, z\} \in E\}) > 0 \). Then we may choose \( z_1, z_2 \) belonging to \( \{z \in D \mid \{y, z\} \in E\} \) so that \( \{z_1, z_2\} \in E \), and therefore \( x, y, z_1, z_2 \) are a copy of \( K_4 \).

As usual, we obtain a corresponding finite version.

**Corollary 5.53.** For every \( \epsilon > 0 \) there is a \( \delta > 0 \) so that if \( G = (V, E) \) with \( |V| > 1/\delta, t_{K_4}(G) \geq 1/8 + \epsilon \), and every set \( X \subseteq V \) with \( \frac{|X|}{|V|} > \delta \) contains an edge, then \( G \) contains a copy of \( K_4 \).

**Proof.** Suppose not. Fix some \( \epsilon \) for which this fails. For each \( N \), there is a \( G_N = (V_N, E_N) \) with \( |V_N| > N, t_{K_4}(G) \geq 1/8 + \epsilon \), and every \( X \subseteq V_N \) with \( \frac{|X|}{|V_N|} > \delta \) contains an edge, but \( G_N \) contains no copy of \( K_4 \).

Let \( G = [G_N]_H \). Then \( G \) contradicts the preceding theorem. \( \square \)
5.11 Szemerédi Regularity

Lemma 5.51 also has a finitary version—indeed, this is the seed from which all the work discussed in this book grew. Before proving it, we should note that the property of being an \( \epsilon \)-regular bipartite graph is preserved by the ultraproduct.

**Lemma 5.54.** If \((B_n]_\mathcal{U}, [C_n]_\mathcal{U}, [E_n]_\mathcal{U})\) is \( \epsilon \)-regular then \( \{n \mid (B_n, C_n, E_n) \text{ is } \epsilon \text{-regular}\} \in \mathcal{U}. \)

**Proof.** The contrapositive is easier: suppose \( \{n \mid (B_n, C_n, E_n) \text{ is not } \epsilon \text{-regular}\} \in \mathcal{U}. \) Then for each such \( n \), there is \( X_n \supseteq B_n \) and \( Y_n \subseteq C_n \) with \( \frac{|X_n|}{|B_n|} \geq \epsilon, \frac{|Y_n|}{|C_n|} \geq \epsilon, \) and \( |d_{E_n}(X_n, Y_n) - d_{E_n}(B_n, C_n)| \geq \epsilon. \)

Therefore, taking \( X = [X_n]_\mathcal{U} \) and \( Y = [Y_n]_\mathcal{U} \), \( |d_E(X, Y) - d_E(B, C)| \geq \epsilon, \) so \( (B, C, E) \) is not \( \epsilon \)-regular. \( \Box \)

**Theorem 5.55** (Szemerédi’s Regularity Lemma). For every \( \epsilon > 0 \) and every \( k_0 \), there is an \( N \) so that whenever \( G = (V, E) \) is a finite graph with \( |V| \geq N \), there is a partition \( V = \bigcup_{i \leq k} B_i \) such that:

- \( k \leq N, \)
- \( \text{for each } i, |B_i| \leq |V|/k_0, \)
- \( \text{there is a set } R \subseteq [1, k] \times [1, k] \) of regular pairs so that:
  - if \( (i, j) \in R \) then \((B_i, B_j, E)\) is \( \epsilon \)-regular,
  - \( |\bigcup_{(i, j) \notin R} B_i \times B_j| < \epsilon |V|^2. \)

The first condition says that the size of the partition is not too big—in particular, \( V \) can be chosen to be much, much larger than \( N \). It is common to replace the second requirement with the requirement that there be at least \( k_0 \) parts and that the partition be an “equipartition”—that the pieces \( B_i \) differ in size by at most 1. Such a version can be obtained from this one by a certain amount of additional fiddling with the pieces.

**Proof.** Suppose the theorem were false, so there is an \( \epsilon > 0 \) and an \( n \) and, for every \( N \), a graph \( G_N = (V_N, E_N) \) with \( |V_N| \geq N \) so that there is no such partition of size \( \leq N. \)

Let \( G = [G_N]_\mathcal{U}, V = [V_N]_\mathcal{U}, \) and \( E = [E_N]_\mathcal{U}. \) By Theorem 5.29 choose a partition \( V = \bigcup_{i \leq k} B_i \) so that each \( B_i \) is internal, each \( \mu_1(B_i) > 0, \) and

\[ \|\mathbb{E}(\chi_E \mid \mathcal{B}_{2,1}) - \mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i, j \leq k})\|_{L^2(\mu_2)} < \epsilon^5. \]
If any of the $B_i$ have measure $> 1/k_0$, we split them into smaller pieces. Then, as in the proof of Lemma 5.51, $\mu_2(\bigcup_{0 \leq i, j \leq k}(B_i \times B_j))$ is not $\epsilon$--regular $B_i \times B_j < \epsilon$.

Each $B_i = [B_{i,n}]_U$. There are only finitely many $i$, so we may find an $n$ so that whenever $(B_i, B_j, E)$ is $\epsilon$--regular, $(B_{i,n}, B_{j,n}, E_n)$ is $\epsilon$--regular, and $\mu_1(B_{i,n}) \leq 1/k_0$.

This does not fully capture the strength of Theorem 5.29. In the ultraproduct, we get a partition which is within $\epsilon$ of being optimal among all partitions. But if we examine the proof of Lemma 5.51, we only compared the partition we got to one other partition. (Indeed, a partition whose size can be calculated—it has size at most $k^{2k}$.)

In the finite world, we can’t hope for a uniform (independent of $|V|$) bound on an almost optimal partition. For suppose we had some bound $N$; then consider some number $M$ much larger than $N$, and a graph $|V|$ with many more than $M$ vertices where we partition $V = \bigcup_{i \leq M} C_i$ where the $C_i$ all have the same size, and then for each rectangle $C_i \times C_j \cup C_j \cup C_i$, we flip a coin and either place the whole rectangle in $E$ (if the coin is heads) or none of it (if the coin is tails). (This is basically the example from the end of Section 5.7.) Then the partition into only $N$ parts does provide any meaningful information—on each rectangle we will have $\mu_2((B_i \times B_j) \cap E) \approx \frac{1}{2} \mu_2(B_i \times B_j)$ (because the rectangle is subdivided into many rectangles $C_i \times \bar{C}_j$ which are randomly assigned to be in or out of the graph). Yet there is some excellent partition, namely $V = \bigcup_{i \leq M} C_i$.

So we can’t hope to find a bound on optimal partitions. But we can hope to find a bound on partitions which are optimal among partitions which are not too much bigger.

**Theorem 5.56.** For every $\epsilon > 0$, every $n$, and every function $F : \mathbb{N} \to \mathbb{N}$ there is an $N$ so that whenever $G = (V, E)$ is a finite graph with $|V| \geq N$, there is a partition $V = \{B_i\}_{i \leq k}$ such that:

- $k \leq N$,
- if $k' \leq F(k)$ and $V = \bigcup_{j \leq k'} D_j$ is a partition so that each $D_j \subseteq B_i$ for some $i \leq k$,

$$||\mathbb{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j, j' \leq k'}) - \mathbb{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i, i' \leq k})||_{L^2(\mu_2)} < \epsilon.$$

**Proof.** Suppose not. Let $\epsilon > 0$ and $F$ be a counterexample and, for each $N$, let $G_N = (V_n, E_n)$ with $|V_N| \geq N$ be a graph with no such partition of size $\leq N$. 

Let $G = [G_N]_U$, $V = [V_N]_U$, and $E = [E_N]_U$. By Theorem 5.29, choose a partition $V = \bigcup_{i \leq k} B_i$ so that each $B_i$ is internal, each $\mu_1(B_i) > 0$, and

$$||\mathbb{E}(\chi_E \mid B_{2,1}) - \mathbb{E}(\chi_E \mid \{B_i \times B_j\}_{i,j \leq k})||_{L^2(\mu_2)} < \epsilon.$$ 

Let each $B_i = [B_{i,N}]_U$. In almost every $G_N$ with $N \geq k$, the $B_{i,N}$ form a partition, so by assumption there must be some $k'_N \leq F(k)$ and some partition $V_N = \bigcup_{j \leq k'_N} \{D_{j,N}\}_{j \leq k'_N}$ so that each $D_{j,N} \subseteq B_{i,N}$ for some $i \leq k$ and

$$||\mathcal{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j,j' \leq k'_N}) - \mathcal{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} \geq \epsilon.$$

$F(k)$ is finite, so there must be some $k'$ so that $\{N \mid k'_N = k'\} \in U$. For each $j \leq k'$, let $D_j = [D_{j,N}]_U$. For each $j$ and each $N$, there is an $i_{j,N} \leq k$ so that $D_{j,N} \subseteq B_{i_{j,N},N}$, so there is an $i_j$ so that $\{N \mid i_{j,N} = i_j\} \in U$, so $D_j \subseteq B_{i,j}$. So the partition $V = \{D_j\}_{j \leq k'}$ refines $\{B_i\}_{i \leq k}$, and therefore

$$||\mathcal{E}(\chi_E \mid \{B_i \times B_{i'}\}_{i,i' \leq k})||_{L^2(\mu_2)} + \epsilon \leq ||\mathcal{E}(\chi_E \mid \{D_j \times D_{j'}\}_{j,j' \leq k'_N})||_{L^2(\mu_2)} \leq ||\mathbb{E}(\chi_E \mid B_{2,1})||_{L^2(\mu_2)},$$

contradicting the choice of $\{B_i\}_{i \leq k}$.

This version was introduced by Alon, Fischer, Krivelevich, and Szegedy in [5] and has gone by a variety of names, like “strong”, “robust”, and “metastable”.

### 5.12 Higher Order Transfer

The relationship between Theorem 5.29 and Theorem 5.56 is an example of a more general equivalence.

Note that Theorem 5.29 has the wrong form to apply the transfer theorem (Corollary 4.47): transfer applies to statements of the form “for every natural number $y$ there is a natural number $z$ so that $\sigma_{y,z}$ is true”, while Theorem 5.29 can be interpreted as saying

for every $\epsilon > 0$ there is a $k$ so that for every $d$

$$\text{there is a partition } V = \bigcup_{i \leq k} B_i \text{ so that whenever } V = \bigcup_{i \leq d} C_i, ||f - \mathbb{E}(f \mid \{B_i \times B_j\})|| < ||f - \mathbb{E}(f \mid \{C_i \times C_j\})|| + \epsilon.$$ 

The inner part (“there is a partition...”) can be expressed in a first-order way, but the outer quantifier structure involves three quantifiers: “for every $\epsilon$ there is a $k$ so that for every $d$”. A generalization of the transfer theorem holds for such statements, and resembles the strong regularity lemma.
5.12. HIGHER ORDER TRANSFER

Theorem 5.57. Suppose that, for every triple of natural numbers \( y, z, w \), \( \phi_{y,z,w}(x_1, \ldots, x_m) \) is a first-order formula. Choose parameters \([b_1^0]_U, \ldots, [b^{m-k}]_U\), and let

\[
X = \{([a_1^1]_U, \ldots, [a_n^k]_U) \mid \forall y \exists z \forall w \phi_{y,z,w}([a_1^1]_U, \ldots, [a_n^k]_U, [b_1^0]_U, [b_1^1]_U, \ldots, [b^{m-k}]_U)\}.
\]

Then \([a_1^1]_U, \ldots, [a_n^k]_U \in X\) if and only if for every \( y \) and every \( W : \mathbb{N} \to \mathbb{N} \), there is a \( z \) so that

\[
\{n \mid \phi_{y,z,W(\zeta)}(a_1^1, \ldots, a_n^k, b_1^0, \ldots, b^{m-k}_n)\} \in U.
\]

Proof. Consider some \([a_1^1]_U, \ldots, [a_n^k]_U \in [V^k]_U\). Suppose \([a_1^1]_U, \ldots, [a_n^k]_U \notin X\), so there is some \( y \) so that, for every \( z \) there is a \( w \) with \( \phi_{y,z,w}([a_1^1]_U, \ldots, [a_n^k]_U, [b_1^0]_U, [b_1^1]_U, \ldots, [b^{m-k}]_U)\) false. Let \( W \) be the function with \( W(z) = w \). Suppose that, for this \( y \) and \( W \), we had some \( z \) so that

\[
\{n \mid \phi_{y,z,W(\zeta)}(a_1^1, \ldots, a_n^k, b_1^0, \ldots, b^{m-k}_n)\} \in U.
\]

But this would mean that \( \phi_{y,z,w}([a_1^1]_U, \ldots, [a_n^k]_U, [b_1^0]_U, [b_1^1]_U, \ldots, [b^{m-k}]_U)\), which contradicts our choice of \( w \).

Conversely, suppose that \([a_1^1]_U, \ldots, [a_n^k]_U \in X\). Then for any \( y \) and \( W \), we can ignore \( W \)—there is some \( z \) so that, for every \( w \), \( \phi_{y,z,w}([a_1^1]_U, \ldots, [a_n^k]_U, [b_1^0]_U, [b_1^1]_U, \ldots, [b^{m-k}]_U)\) holds. Then, for every \( w \),

\[
\{n \mid \phi_{y,z,w}(a_1^1, \ldots, a_n^k, b_1^0, \ldots, b^{m-k}_n)\} \in U.
\]

So, in particular,

\[
\{n \mid \phi_{y,z,W(\zeta)}(a_1^1, \ldots, a_n^k, b_1^0, \ldots, b^{m-k}_n)\} \in U.
\]

\(\square\)

Corollary 5.58 (Metastable Transfer). Suppose that, for every triple of natural numbers \( y, z, w \), \( \sigma_{y,z,w}(x_1, \ldots, x_m) \) is a first-order sentence. Then the following are equivalent:

- in \([G_n]_U\), for every \( y \) there is a \( z \) so that for every \( w \) \( \sigma_{y,z,w} \) is true,
- for every \( y \) and every \( W \), there is a \( z \) so that \( \{n \mid \sigma_{y,z,W(\zeta)} \text{ is true in } G_n\} \in U \).

Corollary 5.59. Suppose that, for every triple of natural numbers \( y, z, w \), \( \sigma_{y,z,w}(x_1, \ldots, x_m) \) is a first-order sentence. Then the following are equivalent:
• in every infinite ultraproduct \([G_n]_U\), for every \(y\) there is a \(z\) so that for every \(w\), \(\sigma_{y,z,w}\) is true,

• for every \(y\) and \(W\) there is a \(Z\) and an \(n\) so that whenever \(G_n\) is a graph with \(\geq n\) vertices, there is a \(z \leq Z\) so that \(\sigma_{y,z,W(z)}\) is true in \(G_n\).

5.13 Remarks

\(V\) and \(C_4\) are not the only graphs which give rise to seminorms, and the property that \(t_H\) gives a (power of) a seminorm seems to be very close to saying that the presence of \(H\) characterizes some notion of randomness \([50]\).

Szemerédi’s regularity lemma first appeared as a step in his proof of Szemerédi’s Theorem, and was isolated in \([100]\). It immediately became a central tool in extremal graph theory; Theorem 5.52 is a representative early application, also due to Szemerédi \([99]\), but there are many others. While the proof of the regularity lemma and its generalizations is a central topic for us, delving into its applications would take us far afield, and a number of excellent surveys have been written on the topic \([64, 65, 86]\).

The regularity lemma has notoriously poor bounds: the bound on the number of pieces is known to be bounded by a function of the order of a tower of exponents of size \(1/\epsilon\), and this cannot be improved \([32, 47, 75]\). As a result, alternative proofs have been sought for many of its consequences.

Graph removal in particular has been extensively studied. Graph removal and its variants turn out to have applications in computer science, where they are related to questions of “property testing” \([44]\)—roughly speaking, the question of whether a property of a graph can be tested by a computer program in a reasonable way. Many variants of graph removal have been proven—for instance, where one modifies a graph with few induced copies of a graph to contain no copies \([5]\), where one deals with infinite families of graphs \([2]\), directed graphs \([3]\). Proofs of some of these results have been given which do not use the regularity lemma \([31]\) and give somewhat better bounds. Some of these and other results are surveyed in \([22]\).

The generalization of Szemerédi regularity, Theorem 5.56, was introduced in \([5]\) in order to prove a graph removal with induced subgraphs. (Another variant, which corresponds to choosing the function \(F(k) = k\), is sometimes called the “weak” regularity lemma \([35]\).) The approach here was identified by Tao \([107]\), who introduced the term “metastable” for such results in a slightly different context \([106]\). The more general notion of higher order transfer between “metastable” formulations—that is, formulations of the
form “for every $\epsilon$ and every function $F$”—and $\Pi_3$ statements was first noticed by Kohlenbach [62] and has subsequently been extensively studied [7, 8, 27, 61, 63].
Chapter 6

Combinatorial Structure

6.1 VC Dimension

If we have a graph \( G = (V, E) \) (finite or infinite) and a finite set of vertices \( \{x_1, \ldots, x_n\} \subseteq V \), and vertex \( y \in V \) picks out a subset, namely \( \{x_i \mid \{x_i, y\} \in E\} = \{x_1, \ldots, x_n\} \cap N_G(y) \). As we consider different vertices in \( V \), we identify different subsets of \( \{x_1, \ldots, x_n\} \) in this way. In a random graph (for example, \( \mathcal{R}_{1/2}(V) \) where \( |V| \) is much larger than \( n \)), we expect to obtain all \( 2^n \) subsets of \( \{x_1, \ldots, x_n\} \) as we vary \( y \) across different values.

We are interested in graphs where this does not happen: graphs where we do not find all the subsets in this way.

**Definition 6.1.** Suppose \( G = (V, E) \) is a graph (finite or infinite). For any \( x \in V \), write \( E_x = \{y \mid \{x, y\} \in E\} \)—that is, \( E_x \) is the neighborhood of \( x \).

If \( X \subseteq V \) (typically \( X \) is finite), we say \( E \) shatters \( X \) if, for every \( S \subseteq X \), there is a \( y \in V \) so that \( X \setminus E_y = S \).

The VC dimension of \( G \) is the largest \( n \) such that there exists some \( \{x_1, \ldots, x_n\} \subseteq V \) which is shattered by \( E \), or \( \infty \) if there is no such \( n \).

VC stands for Vapnik-Chervonenkis, the names of the two computational learning theorists who introduced the notion.

Having high VC dimension, particularly infinite VC dimension, is a notion of complexity. Note that VC dimension is defined by a worst case scenario: we have high VC dimension if we find a single choice of set \( \{x_1, \ldots, x_n\} \) which is shattered. This reflects the idea that if we combine a simple graph and a complex one (say, by taking the disjoint union of their vertices), we ought to get a complex graph.

Graphs with low VC dimension tend to be very structured. For instance, the complete bipartite graph \( K_{n,n} \) (with \( n \geq 3 \)) has VC dimension 1: if we
take a set \( \{x_1, x_2\} \), if these vertices are from different parts then no \( y \) has \( \{x_1, x_2\} \cap E_y = \{x_1, x_2\} \), and if they are from the same part then no \( y \) has \( \{x_1, x_2\} \cap E_y = \{x_1\} \).

Similarly, consider the bipartite graph where \( V_0 = V_1 = \{x_1, x_2, \ldots, x_n\} \) and \( V \) is the disjoint union of \( V_0 \) and \( V_1 \), and \( E \) is the set of pairs \( \{x, y\} \) with \( x \in V_0, \ y \in V_1, \) and \( x < y \). This has VC dimension 1: suppose we take two vertices \( \{x_1, x_2\} \). If they are in different parts, we cannot shatter the set because there would be no \( y \) with \( \{x_1, x_2\} \cap E_y = \{x_1, x_2\} \), and if they are in the same part—without loss of generality, assume \( V_0 \)—they have an order—say, \( x_1 < x_2 \)—and then we cannot have \( \{x_1, x_2\} \cap E_y = \{x_2\} \).

At the other extreme, \( R_{1/2}(V) \) tends to have large VC dimension; in particular, with probability 1, \( |R_{1/2}(V_n)| \) will have VC dimension \( \infty \) (assuming \( |V_n| \to \infty \)).

The notion of VC dimension has been invented several separate times in the literature. In particular, Shelah introduced it in model theory \cite{Shelah}, where it goes by the name NIP (“not the independence property”). More precisely, a theory is NIP if every definable graph has finite VC dimension. Although beyond our scope here, the point is that a wide variety of theories have the NIP property, and therefore any graph defined in a model of one of those theories will have finite VC dimension.

VC dimension is usually considered in a slightly more abstract setting.

**Definition 6.2.** A set system is a set \( V \) and a collection \( F \) of subsets of \( V \). When \( X \subseteq V \), we say \( F \) shatters \( X \) if, for every \( S \subseteq X \), there is an \( F \in F \) with \( X \cap F = S \).

The **VC dimension** of \((V, F)\) is the largest \( n \) such that there exists some \( X \subseteq V \) with \( |X| = n \) so that \( X \) is shattered by \( n \).

A graph \((V, E)\) corresponds to the set system on \( V \) with \( F = \{E_y \mid y \in V\} \). (Conversely, given a set system \((V, F)\), we can obtain a bipartite graph whose vertices are \( V \cup F \) with an edge between \( x \) and \( F \) exactly when \( x \in F \).)

Unlike the definition in a graph, this definition looks asymmetric. If we are only concerned with whether a set system has finite or infinite VC dimension, however, there is a symmetry.

**Lemma 6.3.** Suppose \((V, F)\) has infinite VC dimension. Then for every \( d \), there exist \( F_0, \ldots, F_{d-1} \in F \) so that, for every \( S \subseteq \{F_0, \ldots, F_{d-1}\} \), there is an \( x \in V \) so that \( \{F \in \{F_0, \ldots, F_{d-1}\} \mid x \in F\} = S \).

**Proof.** Since \((V, F)\) has infinite VC dimension, we may choose a set \( \{x_1, \ldots, x_{2d}\} \) which is shattered by \( F \). For each \( j < d \), we choose \( F_j \) containing exactly
those $x_i$ such that, when we write $i$ in base 2, the $j$-th digit is a 1. Then for any $\{F_j, \ldots, F_{j_k}\} \subseteq \{F_0, \ldots, F_{d-1}\}$, we may let $i = \sum_{k \leq s} 2^j$, and we have $\{F_j, \ldots, F_{j_k}\} = \{F \in \{F_0, \ldots, F_{d-1}\} \mid x_i \in F\}$.

VC dimension gives a striking dividing line. For any finite $X \subseteq V$, we can define $\Pi_G(X) = \{S \mid \exists F \in \mathcal{F} X \cap F = S\}$ and $\pi_G(n) = \max(|\Pi_G(X)| \mid X \subseteq V, |X| = n)$. That is, $\Pi_G(X)$ is the collection of subsets of $X$ we can obtain by looking at neighborhoods and $\pi_G(n)$ is the largest possible size of $\Pi_G(X)$ among sets of size $n$. Since $\Pi_G(X)$ is a set of subsets of $X$, $|\Pi_G(X)| \leq 2^{|X|}$, so $\pi_G(n) \leq 2^n$. The VC dimension of $(V, \mathcal{F})$ is the smallest $n$ with $\pi_G(n + 1) < 2^{n+1}$.

It turns out that there are only two possible behaviors for the function $\pi_G$: either the function is the exponential $2^n$ or it is bounded by a polynomial.

**Theorem 6.4 (Sauer-Shelah).** If the VC dimension of $(V, \mathcal{F})$ is $d$ then, for all $n$, $\pi_G(n) \leq \sum_{i=0}^{d} \binom{n}{i}$.

There are a very large number of proofs of this theorem, of which we only give one.

**Proof.** We show something slightly stronger: for any finite $X \subseteq V$, there are at least $|\mathcal{F}|$ subsets of $X$ which are shattered by $\mathcal{F}$. The theorem will follow since, if there is a set $X \subseteq V$ with $|X| = n$ and $|\Pi_G(X)| > \sum_{i=0}^{d} \binom{n}{i}$, there are only $\sum_{i=0}^{d} \binom{n}{i}$ subsets of $X$ of size $\leq d$, so there must be a subset of $X$ of size $> d$ shattered in $G$.

We show this by induction on $|X|$. If $X = \emptyset$ and $\mathcal{F} > 0$ then $X$ is shattered by definition.

Suppose that $|X| = m + 1$ and choose some $x \in X$. Let $\mathcal{F}_x = \{v \in \mathcal{F} \mid x \in v\}$ and let $\mathcal{F}_- = \mathcal{F} \setminus \mathcal{F}_x$. Then, by the inductive hypothesis, $\mathcal{F}_x$ shatters at least $|\mathcal{F}_x|$ subsets of $X \setminus \{x\}$ while $\mathcal{F}_-$ shatters at least $|\mathcal{F}_-|$ subsets of $X \setminus \{x\}$.

We now count subsets of $X$ shattered by $\mathcal{F}$. First, if $S \subseteq X$ is shattered by $\mathcal{F}_-$, $X$ is shattered by $\mathcal{F}$. Suppose $S \subseteq X$ is shattered by $\mathcal{F}_x$. If $S$ is also shattered by $\mathcal{F}_-$ then $S \cup \{x\}$ is shattered by $\mathcal{F}$. Otherwise $S$ is not shattered by $\mathcal{F}_-$ but is shattered by $\mathcal{F}$. In particular, for each set shattered by $\mathcal{F}_x$, we obtain an additional set shattered by $\mathcal{F}$, so the number of sets shattered by $\mathcal{F}$ is at least $|\mathcal{F}_-| + |\mathcal{F}_x| = |\mathcal{F}|$. 

\[\square\]
6.2 Closure of VC Dimension

It is convenient to note that having finite VC dimension is preserved under various ways of combining set systems.

**Lemma 6.5.** If $(V, F)$ is a set system with finite VC dimension, the set system consisting of complements, $(V, \{V \setminus F \mid F \in F\})$, also has finite VC dimension.

*Proof.* Let $X \subseteq V$ be a finite set not shattered by $F$. Then there is an $S \subseteq X$ so that $X \cap F \neq S$ for any $F \in F$. Let $T = X \setminus S$; then $X \cap (V \setminus F) \neq T$ for any $F \in F$. \hfill \Box

**Lemma 6.6.** Let $(V, F)$ and $(V, G)$ be two set systems on the same sets of points, both having finite VC dimension. Then the intersection set system, $(V, \{F \cap G \mid F \in F \text{ and } G \in G\})$ also has finite VC dimension.

*Proof.* A slick proof is to make use of the Sauer-Shelah theorem. Let $d$ bound the VC dimension of both set systems, and choose $n$ much larger than $d$. Consider a subset $X \subseteq V$ of size $n$ and consider how many subsets of $X$ have the form $X \setminus (F \cap G)$ with $F \in F$ and $G \in G$.

First, consider the subsets of $X$ of the form $X \cap F$ with $F \in F$: by Sauer-Shelah, there are at most $\sum_{i \leq d} \binom{n}{i}$ of these. For each set $X \cap F$, consider the number of subsets of the form $X \cap F \cap G$ with $G \in G$: there are, again, at most $\sum_{i \leq d} \binom{|X \cap F|}{i}$ of these.

So, in total, there can be at most $(\sum_{i \leq d} \binom{|X \cap F|}{i}) \leq \sum_{i \leq d} \binom{n}{i}$ subsets of $X$ of the form $X \cap (F \cap G)$. When $n$ is large enough, this is less than $2^n$, so in particular $X$ cannot be shattered. \hfill \Box

**Lemma 6.7.** Let $(V, F)$ and $(V, G)$ be two set systems on the same sets of points, both having finite VC dimension. Then the union set system, $(V, \{F \cup G \mid F \in F \text{ and } G \in G\})$ also has finite VC dimension.

*Proof.* We could give an argument similar to the previous lemma, but we can also note that the union is the complement of the intersection of the complements, so this follows from the previous two lemmata. \hfill \Box

The main application we will need is actually the symmetric difference.

**Corollary 6.8.** Let $(V, F)$ and $(V, G)$ be two set systems on the same sets of points, both having finite VC dimension. Then the symmetric difference set system, $(V, \{F \triangle G \mid F \in F \text{ and } G \in G\})$ also has finite VC dimension.

This again follows from the lemmata above because the symmetric difference can be formed using union, intersection, and complement.
6.3 \( \epsilon \)-Nets

Suppose \( G \) is a measurable graph and \( y \) is a vertex with \( \mu(E_y) \geq \epsilon > 0 \). If we select a large number of points \( x_1, \ldots, x_n \) at random (with \( n \) large relative to \( 1/\epsilon \)), we expect that at least one of these points will probably belong to \( E_y \). If we have several vertices, \( y_1, \ldots, y_m \) with each \( \mu(E_{y_i}) \geq \epsilon \), we would expect to need to make \( n \) larger if we want to find a point in every set \( E_{y_i} \)—even though we’re likely to get a point in \( E_{y_1} \), and a point in \( E_{y_2} \), and so on, the likelihood that we miss at least one of the sets \( E_{y_i} \) increases with \( m \). In particular, when there are uncountably many values of \( y \), it could become very likely that we miss at least one of them. For instance, in \( \mathbb{R}_{1/2} \cup \), almost every \( y \) has \( \mu(E_y) = 1/2 \), but we cannot pick an \( x \) belonging to all these \( E_y \) simultaneously.

If the graph has finite VC dimension, however, we can hope to select a reasonable number of \( x \)’s so that we have found at least one member of every \( E_{y_i} \). We call such a set of \( x \)’s an \( \epsilon \)-net—like a net, it “catches” every set \( E_y \) (except those which are so small that they slip through).

**Definition 6.9.** A set \( X \subseteq V \) is an \( \epsilon \)-net for \( G = (V, E, \mu) \) if, for every \( y \in V \) with \( \mu(E_y) \geq \epsilon \), \( X \cap E_y \neq \emptyset \).

Of course, one can also speak of \( \epsilon \)-nets for any set system.

**Theorem 6.10.** If \( G = (V, E, \mu) \) has finite VC dimension then there is an \( \epsilon \)-net.

The proof here is similar to the one given in [73].

**Proof.** Let \( d \) be the VC dimension of \( G \). We will show that there is a large enough size \( r \) (depending on \( d \) and \( \epsilon \)) so that if \( X \) is a set of size \( r \) chosen randomly according to \( \mu \), there is positive probability that \( X \) is an \( \epsilon \)-net.

When we choose \( X \) of size \( r \), let \( p \) be the probability that \( X \) is not an \( \epsilon \)-net. We will consider a related probability: the probability that we choose two sets \( X_0 \) and \( X_1 \), each of size \( r \), such that there is some \( y \) such that \( X_1 \) catches \( E_y \) but \( X_0 \) does not—that is, some \( y \) so that \( X_0 \cap E_y = \emptyset \) while \( |X_1 \cap E_y| \geq \epsilon r/2 \). We will count the probability of choosing this set in two different ways, one of them involving \( p \), to get a bound on \( p \).

The first way of counting is by first picking a set \( Z \) of size \( 2r \) and then picking \( X_0 \subseteq Z \) of size \( r \) as a random subset of \( Z \). Suppose we have chosen a set \( Z \). For any given \( y \) with \( |E_y \cap Z| \geq \epsilon r/2 \), we can ask for the possibility that, when we choose \( X_0 \subseteq Z \) so that \( E_y \cap X_0 = \emptyset \).
For any fixed \( y \) with \( |E_y \cap Z| \geq \epsilon r/2 \), the probability that we choose \( r \) elements from \( z \) while avoiding all \( \epsilon r/2 \) elements of \( E_y \cap Z \) is at most \((1 - \epsilon/4)^r\).

Of course, there are many values of \( y \). However—after we’ve chosen \( Z \)—this outcome depends only on the set \( E_y \cap Z \). The Sauer-Shelah Theorem tells us that there are only \( \Pi_G(2^r) \leq C(2^r) \) possibilities (for some \( C \)) for the set \( E_y \cap Z \). Therefore the probability that we have \( E_y \cap Z = \emptyset \) for at least one value of \( y \) is bounded by \( C(2^r)(1 - \epsilon/4)^r \). In particular, this decreases exponentially in \( r \), so when \( r \) is large enough we can ensure that this probability is at most \( 1/2 \).

We now count this a second way: we first choose the set \( X_0 \) so that \( X_0 \) is not an \( \epsilon \)-net, with the unknown probability \( p \), and then choose \( X_1 \). If \( X_0 \) is not an \( \epsilon \)-net, there is some \( y \) with \( \mu(E_y) \geq \epsilon \) but \( X_0 \cap E_y = \emptyset \). Fix some such \( y \). Then (by the Hoeffding inequality), when \( r \) is large enough, the probability that we choose \( X_1 \) so that \(|X_1 \cap E_y| \geq \epsilon r/2 \) is at least \( 3/4 \).

Therefore the probability that we choose \( X_0 \) and \( X_1 \) so there is some \( y \) with \( X_0 \cap E_y = \emptyset \) but \(|X_1 \cap E_y| \geq \epsilon r/2 \) is at least \( 3p/4 \) but at most \( 1/2 \). Therefore \( 3p/4 \leq 1/2 \), so \( p \leq 2/3 < 1 \), so in particular there is a positive probability that \( X_0 \) is an \( \epsilon \)-net.

\[ \Box \]

### 6.4 VC Dimension and Rectangles

The existence of \( \epsilon \)-nets gives us a quick tool for constructing approximations using rectangles.

**Theorem 6.11.** If \( G = (V, E, \mu) \) is a measurable graph with finite VC dimension then \( E \in B_{2,1} \).

**Proof.** It suffices to show that, for every \( \epsilon > 0 \), there is an approximation of \( E \) to within \( \epsilon \) using rectangles.

Consider the set system for symmetric differences, \( (V, \{E_x \triangle E_y \mid x, y \in V\}) \). This set system also has finite VC dimension, and therefore we may choose an \( \epsilon \)-net \( Z = \{z_1, \ldots, z_m\} \). Since this is an \( \epsilon \)-net, whenever \( \mu_1(E_x \triangle E_y) \geq \epsilon \), there is an \( i \) with \( z_i \in E_x \triangle E_y \).

For each \( s \subseteq Z \), let \( V_s = \{x \in V \mid E_x \cap Z = s\} \). For each \( s \subseteq Z \) with \( V_s \neq \emptyset \), choose some \( y_s \in V_s \). Observe that if \( x \in V_s \) then \( \mu_1(E_x \triangle E_{y_s}) < \epsilon \).
Then let \( E' = \bigcup_{s \subseteq Z, V_s \neq \emptyset} V_s \times E_{ys} \). We claim that \( \mu_2(E \triangle E') < \epsilon \):

\[
\begin{align*}
\mu_2(E \triangle E') &= \sum_{s \subseteq Z} \int_{V_s} \mu_1(E_{xs} \triangle E_{ys}) \, d\mu_1 \\
&< \sum_{s \subseteq Z} \mu_1(V_s) \epsilon \\
&= \epsilon.
\end{align*}
\]

This gives us a corresponding improvement of Szemerédi’s Regularity Lemma (Theorem 5.55) under the assumption of VC dimension. Recall that the regularity lemma promises a uniform bound \( N \) so that every finite graph \((V, E)\) has a partition into pieces \( V = \bigcup_{i \leq k} B_i \) with \( k \leq N \) so that most pairs \((B_i, B_j, E)\) are \( \epsilon \)-regular.

In applications of the regularity lemma, there are a number of ways in which one could hope it might be improved. First, one would like to improve the bounds—the bound on \( N \) is a tower of exponents in \( 1/\epsilon \). Second, one could hope that most of the pairs are not only \( \epsilon \)-regular, but have density \( d_{E}(B_i, B_j) \) close to either 0 or 1. Third, one would like to eliminate the “irregular pairs”—one would like to have every pair \((B_i, B_j, E)\) be \( \epsilon \)-regular, instead of most. And finally, one could hope to strengthen regularity to homogeneity: we could hope to have either \( B_i \times B_j \subseteq E \) or \((B_i \times B_j) \cap E = \emptyset\).

VC dimension gives us the first two of these improvements. In keeping with our focus on limiting cases, we do not worry about the first, and only prove that graphs with bounded VC dimension satisfy a version of the strong regularity lemma (Theorem 5.56).

**Theorem 6.12.** For every \( d \) and \( \epsilon > 0 \) there is an \( N \) so that whenever \((V, E)\) is a graph with \(|V| \geq N\), there is a partition \( V = \bigcup_{i \leq k} B_i \) such that:

- \( k \leq N \),
- \[ \|\chi_{E} - \mathbb{E}(\chi_{E} | \{B_i \times B_j\}_{i,j \leq k})\|_{L^2(\mu_2)} < \epsilon. \]

**Proof.** Suppose not. Let \( d \) and \( \epsilon > 0 \) be given so that this fails, so for each \( N \) we have a counterexample \( G_N = (V_N, E_N) \) with \(|V_N| \geq N\). By the previous theorem, \([E_n]_U\) belongs to \( B_{2,1} \), so there is a partition \([V_N]_U = \bigcup_{i \leq k} B_i \) with each \( B_i \) internal so that

\[
\|\chi_{E} - \mathbb{E}(\chi_{E} | \{B_i \times B_{i'}\}_{i,i' \leq k})\|_{L^2(\mu_2)} < \epsilon/2.
\]
CHAPTER 6. COMBINATORIAL STRUCTURE

Then we may pick representatives $B_i = [B_{i,N}]_{\mu}$ and find an $N$ so that

$$||\chi_{E_N} - \mathbb{E}(\chi_{E_N} | \{B_{i,N} \times B_{i',N}\}_{i,i' \leq k})||_{L^2(\mu_2)} < \epsilon,$$

giving a contradiction. \[\square\]

Note that (as in Lemma 5.51), this implies that most of the $(B_i, B_j, E)$ are $\epsilon$-regular.

6.5 A Converse of Sorts

Neither the existence of $\epsilon$-nets nor belonging to $B_{2,1}$ can plausibly be equivalent to having finite VC dimension. These consequences are measure-theoretic, while VC dimension is determined by finite sets, so we can always take a graph with finite VC dimension and add an additional graph with infinite VC dimension, but put no measure on the additional part.

To obtain an equivalence, we have to ask about what happens under arbitrary measures.

**Theorem 6.13.** $G = (V,E)$ has finite VC dimension if and only if, for every $\epsilon > 0$, there is an $n$ so that, for any $\mu$ making $(V,E,\mu)$ a measurable graph, the probability that a subset of $V^n$ is an $\epsilon$-net is at least $1 - \epsilon$.

**Proof.** We have already shown that graphs with finite VC dimension have this property.

For the converse, suppose $G = (V,E)$ has infinite VC dimension; we will show that, for $\epsilon = 1/2$, for each $n$ there is a measure $\mu$ where $(V,E,\mu)$ is a measurable graph and the probability that a subset of $V^n$ is an $\epsilon$-net is 0. Let $d \geq 2n$ and take a subset $S$ of $V$ of size $d$ which is shattered, and let $\mu$ be the uniform distribution on $S$ (with measure 0 on $V \setminus S$). Then whenever $X \subseteq S$ with $|X| = n$, we may choose, for instance, a $y \in V$ with $E_y \cap S = S \setminus X$, so $\mu(E_y) = \frac{|E_y \cap S|}{|S|} \geq 1/2$, but $E_y \cap X = \emptyset$, so $X$ is not a 1/2-net. \[\square\]

That is, VC dimension is equivalent to the existence of $\epsilon$-nets in a uniform way under all possible measures.

6.6 Stability

One of the first questions asked about the regularity lemma is whether it can be improved so that there are no irregular pairs—that is, all the graphs
(B_i, B_j, E) are \( \epsilon \)-regular. A standard example, the \textit{half graph}, shows that this cannot, in general, be avoided: let \( G_n \) be the graph where \( V_n \) is the disjoint union of two copies of \( \{1, \ldots, n\} \) and a pair \( \{x, y\} \in E_n \) when \( x \) is in the first copy, \( y \) is in the second, and \( x < y \). Then \([G_n]_U\) is a bipartite graph with parts \( L \) and \( R \), an internal bijection \( \pi : L \rightarrow R \), and an ordering \( < \) on each part (with \([v_n]_U < [w_n]_U\) exactly when \( \{n \mid v_n < w_n\} \in U \)) so that \( \pi(v) < \pi(w) \), and \([E_n]_U\) consists of pairs \( v, w \) where \( \pi(v) < w \). The idea is that, in any partition of \( L \) and \( R \), the rectangles along the diagonal would be irregular. (The details are rather tedious, since in an arbitrary partition the two sides could have different partitions, which makes it difficult to identify the diagonal.)

However this is, in a precise sense, the only obstacle.

**Definition 6.14.** If \( G = (V, E) \) is a graph, a \textit{strict ladder of length} \( d \) in \( G \) is a pair of sequences \( x_1, \ldots, x_d, y_1, \ldots, y_d \in V \) so that \( \{x_i, y_j\} \in E \) if and only if \( i < j \).

We say \( G \) is \textit{stable} if there is a \( d \) so that there does not exist a ladder of length \( d \).

The word “strict” refers to the fact that we have a requirement when \( i = j \) as well; a simple “ladder” would allow either \( \{x_i, y_i\} \in E \) or \( \{x_i, y_i\} \notin E \).

Stability is a strengthening of having finite VC dimension: finite VC dimension says that, given \( \{x_1, \ldots, x_d\} \) with \( d \) sufficiently large, there is \textit{some} subset we cannot obtain by looking at the intersection with \( E_y \). Stability prohibits specific combinations of subsets from appearing.

As the example of the half graph shows, stability is strictly stronger—the half graph has finite VC dimension but is unstable.

Like VC dimension, stability gives a bound of sorts on the number of subsets of a set we can hope to obtain, though it has a different character: stability implies that there are only countably many subsets of any countable set.

**Theorem 6.15.** Suppose \( (V, E) \) is stable. Then whenever \( B \subseteq V \) is countable, there are only countably many subsets of \( B \) of the form \( E_y \cap B \).

Stated in this form, this follows directly from the Erdős-Makkai Theorem:

**Theorem 6.16.** Let \( (B, \mathcal{F}) \) be a set system with \( |\mathcal{F}| > |B| \) and \( B \) infinite. Then there exist \( x_1, \ldots, x_d \in B \) and \( F_0, \ldots, F_d \in \mathcal{F} \) so that \( x_i \in F_j \) iff \( i \leq j \).

**Proof.** By induction on \( d \). For \( d = 0 \) this is immediate: take any \( F_0 \in \mathcal{F} \).

Suppose the claim holds for \( d \). Choose any \( F_{d+1} \in \mathcal{F} \) with \( F_{d+1} \neq B \). Consider the set system \( \mathcal{F}' = \{ F \cap F_{d+1} \mid F \in \mathcal{F} \} \).
First, suppose $|F'| > |B|$. For each $b \in F_{d+1}$, consider $F_b = \{ F \in F' \mid b \not\in F \}$. Observe that, other than $F_{d+1}$ itself, every $F \in F'$ belongs to at least one $F_b$, so $F' = \{ F_{d+1} \} \cup \bigcup_{b \in B} F_b$. Since $|F'| > |B|$, this means there is some $b$ with $|F_b| > |B|$. We apply the inductive hypothesis to $(F_{d+1}, F_b)$, obtaining $x_1, \ldots, x_d, F_0, \ldots, F_d$ with each $x_1, \ldots, x_d \in F_{d+1}$ and $x_i \in F_j$ if and only $i < j$. Taking $x_{d+1} = b$ completes the construction.

Otherwise, $|F'| = |B|$. Then consider instead the family $F^c = \{ B \setminus F \mid F \in F \}$. Then $|F^c| > |B|$ as well, and we may apply the same argument as the first case with $B \setminus F_{d+1}$ to obtain $x_1, \ldots, x_{d+1}$ and $B \setminus F_1, \ldots, B \setminus F_{d+1}$ so that $x_i \in B \setminus F_{j+1}$ if and only if $i \leq j$. Then the sequence $x_{d+1}, \ldots, x_1, F_{d+1}, \ldots, F_0$ witnesses the claim.

In an ultraproduct (and, more generally, structures which satisfy the right saturation properties) this is equivalent to stability.

**Theorem 6.17.** Suppose $([V_n]_\mathcal{U}, [E_n]_\mathcal{U})$ is unstable. Then there is a countable $B \subseteq [V_n]_\mathcal{U}$ so that there are uncountably many subsets of $B$ of the form $([E_n]_\mathcal{U})_g$.

**Proof.** The idea is that instability together with the compactness of ultraproducts allows us to construct a copy of the half graph. For each $d$ we have sequences $x^{d,1}, \ldots, x^{d,d}, y^{d,1}, \ldots, y^{d,d}$ so that $\{ x^{d,i}, y^{d,j} \} \in E$ if and only if $i < j$.

Choosing representatives, take $x^{d,i} = [x^{d,i}]_\mathcal{U}$ and $y^{d,i} = [y^{d,i}]_\mathcal{U}$. Choose a sequence of sets $I_0 \supseteq I_1 \supseteq \cdots$ in $\mathcal{U}$ such that if $n \in I_d$ then $\{ x^{d,i}, y^{d,j} \} \in E_n$ if and only if $i < j$. For each real number $r \in [0, 1]$, define $x^r_n, y^r_n$ for $n \in I_{d+1} \setminus I_d$ by taking $x^r_n = x^{d,\lceil dr \rceil}_n$ and $y^r_n = y^{d,\lceil dr \rceil}_n$. For any $r < s$, observe that $\{ n \mid \{ x^r_n, y^r_n \} \in E_n \} \in I_{\lfloor (r - s) \rfloor} = \mathcal{U}$, so $\{ x^r_n \} \in [E_n]_\mathcal{U}$.

Let $B = \{ [x^q_n]_\mathcal{U} \mid q \in \mathbb{Q} \cap [0, 1] \}$. Then for each real number $r$, $E_{[y^q_n]_\mathcal{U}} \cap B = \{ [x^q_n]_\mathcal{U} \mid q < r \}$. This gives uncountably many distinct subsets of $B$.

It would be tempting, by analogy to VC dimension, to define $d$-stability to mean that $d$ is the largest size for which a strict ladder exists. In fact, this term is reserved for a slightly different configuration.

**Definition 6.18.** We write $\{0, 1\}^d$ for the set of binary sequences of length $d$—that is, functions $\sigma : [0, d - 1] \to \{0, 1\}$ and $\{0, 1\}^<d = \bigcup_{k<d} \{0,1\}^k$. When $\sigma \in \{0, 1\}^d$, we set $|\sigma| = d$ and call $|\sigma|$ the length of $d$. We write $\sigma \subseteq \tau$ if $\sigma$ is an initial segment of $\tau$—that is, if $\tau \mid [0, |\sigma| - 1]$.

We say $G = (V, E)$ is $d$-stable if there do not exist collections of points $\{x_\sigma\}_{\sigma \in \{0, 1\}^d}$ and $\{y_\eta\}_{\eta \in \{0, 1\}^<d}$ such that whenever $\eta \sqsubset \sigma$, $x_\sigma \in E_{y_\eta}$ if and only if $\eta^{-1}(1) \subseteq \sigma$. 

6.6. Stability

We can think of the $y_\eta$ as labeling the nodes of a binary tree, indicating that elements in $E_{y_\eta}$ should go to the left and elements not in $E_{y_\eta}$ should go to the right. Starting at the root, every element $x$ follows a path: left or right depending on whether it belongs to $E_{y_{\langle 0 \rangle}}$, then left or right again depending on $E_{y_{\langle 0 \rangle}}$ or $E_{y_{\langle 1 \rangle}}$, and so on. $d$-stability means that, however we label the internal nodes of the tree of height $d$, there is a path which no element $b$ will follow.

The $d$ in stability and the $d$ in $d$-stability don’t necessarily match, but up to that numeric difference, the notions are equivalent.

**Lemma 6.19.** If $G$ is $d$-stable then $G$ is stable.

**Proof.** Suppose $G$ is not stable. For any $d$, we will show that $G$ is not $d$-stable. The idea is that if we have a ladder of length $2^d$, we can use the $x$’s to label the leaves of a tree of height $d$, and the $y$’s to label the internal nodes; for instance, when $d = 2$:

```
 y2
  / \
 y1  y3
 /   \
x0  x1  x2  x3
```

Making this precise for a general $d$ is just a matter of working through the indexing. Start with a ladder $x_0, \ldots, x_{2^d-1}, y_0, \ldots, y_{2^d-1} \in V$ so that $\{x_i, y_j\} \in E$ if and only if $j \leq i$. (We have shifted the indexing by one and swapped the names of the $x$’s and $y$’s in order to match the definition of the tree better.) We can associate each $i < 2^d$ with a sequence of length $d$ by writing it in binary with $\sigma(0)$ being the most significant bit—that is, let $\sigma_i$ be the sequence such that $i = \sum_{j<d} 2^{d-1-j} \sigma_i(j)$, and take $x_{\sigma_i} = x_i$.

Then for each $\eta$, we take $y_\eta = y \sum_{j<|\eta|} 2^{d-1-j} \eta(j) + 2^d - \eta |\eta|$, and therefore $\{x_{\sigma}, y_\eta\} \in E$, while if $\eta \setminus \langle 1 \rangle \subseteq \sigma$ then $x_{\sigma} = x_i$ for some $i \geq \sum_{j<|\eta|} 2^{d-1-j} \eta(j) + 2^d - \eta |\eta|$, and therefore $\{x_{\sigma}, y_\eta\} \notin E$. \hfill \Box

The converse also holds, but is more difficult to prove. We follow the proof in [52], which needs a Ramsey-like theorem for trees.

**Definition 6.20.** The binary tree of height $n$, written $\{0,1\}^{<n}$, is the collection of all sequences of 0’s and 1’s of length less than $n$—that is, $\{0,1\}^{<n} = \bigcup_{0 \leq m < n} \{0,1\}^{[1,..,m]}$.

We write $\sigma \subseteq \tau$ if $\tau \upharpoonright \text{dom}(\sigma) = \sigma$. 

When $m = 0$, the notation $\{0,1\}^{|1|\ldots|0|}$ means the function from the empty set to $\{0,1\}$—that is, the empty function, which we write $\langle \rangle$; we call this the root of the tree. $\sigma \subseteq \tau$ means that $\tau$ extends $\sigma$.

**Definition 6.21.** A subtree of $\{0,1\}^{<n}$ of height $m$ is a function $f : \{0,1\}^{<m} \rightarrow \{0,1\}^{<n}$ such that $f(\sigma) \subseteq f(\tau)$ if and only if $\sigma \subseteq \tau$.

For instance, $f(\sigma)$ must extend $f(\langle \rangle)$ for all $\sigma \in \{0,1\}^{<m}$, and the two nodes $f(\sigma \setminus \langle 0 \rangle)$ and $f(\sigma \setminus \langle 1 \rangle)$ must be two distinct, incomparable extensions of $f(\sigma)$.

**Lemma 6.22.** Suppose that $c : \{0,1\}^{<n_0+n_1} \rightarrow \{0,1\}$. Then for some $i \in \{0,1\}$, there is a subtree $f : \{0,1\}^{<n_i} \rightarrow \{0,1\}^{<n_0+n_1}$ such that $c(f(\sigma)) = i$ for all $\sigma \in \{0,1\}^{<n_i}$.

That is, we can find a “monochromatic subtree”.

**Proof.** We proceed by induction on $n_0 + n_1$. When $n_0 + n_1 = 0$ all these trees are empty, so this is trivial.

Suppose $c(\langle \rangle) = 0$. (The case where $c(\langle \rangle) = 1$ is symmetric.) There are two natural subtrees of height $n_0 + n_1 - 1$—the nodes extending $\langle 0 \rangle$ and the nodes extending $\langle 1 \rangle$.

Formally, consider two colorings $c_j : \{0,1\}^{<n_0+n_1-1} \rightarrow \{0,1\}$ given by $c_j(\sigma) = c(\langle j \rangle \setminus \sigma)$. We may apply the inductive hypothesis to each of $c_0$ and $c_1$ with $n_0 - 1, n_1$; if, for either $j$, we obtain $f_j : \{0,1\}^{<n_j} \rightarrow \{0,1\}^{<n_0+n_1-1}$ so that $c_j(f_j(\sigma)) = 1$ for all $\sigma \in \{0,1\}^{<n_j}$ then we may take $f(\sigma) = (j \setminus f_j(\sigma))$.

Suppose not. Then we obtain two functions $f_j : \{0,1\}^{<n_0-1} \rightarrow \{0,1\}^{<n_0+n_1-1}$ so that $c_j(f_j(\sigma)) = 0$ for all $\sigma \in \{0,1\}^{<n_0-1}$. Then we define $f : \{0,1\}^{<n_0} \rightarrow \{0,1\}^{<n_0+n_1}$ by $f(\langle \rangle) = \langle \rangle$ and $f((j) \setminus \sigma) = (j \setminus f_j(\sigma))$. \qed

**Theorem 6.23.** If $G$ is stable then there is a $d$ so that $G$ is $d$-stable.

**Proof.** We will show that if $G$ fails to be $3 \cdot 2^{d-1} - 2$-stable then we can find a strict ladder of length $d$. Specifically, we will show the following by induction on $d$:

Let $h = 3 \cdot 2^{d-1} - 2$. Suppose we have a tree of vertices $\{x_\sigma\}_{\sigma \in \{0,1\}^h}$ and $\{y_\eta\}_{\eta \in \{0,1\}^{<h}}$ such that when $\eta \subseteq \sigma$, $\{x_\sigma, y_\eta\} \in E$ if and only if $\eta \setminus (1) \subseteq \sigma$. Then there exist $x_1, \ldots, x_d, y_1, \ldots, y_d$ so that $\{x_i, y_j\} \in E$ if and only if $i < j$, for each $i$ there is a $\sigma$ so that $x_i = x_\sigma$, and for each $j$ there is an $\eta$ so that $y_j = y_\eta$. 
Let such a tree be given. When $d = 1$, $h = 1$, so we may take $y_1 = y_{\langle \rangle}$ and $x_1 = x_{\langle \rangle}$.

Suppose $d > 1$. Let $h' = 3 \cdot 2^{d-2} - 2$, so that $1 + h' + (h' + 1) = h$. Choose some $\sigma \ni \langle \rangle$ and define $c : \{0, 1\}^{<h-1} \to \{0, 1\}$ by setting $c(\eta) = 1$ if $\{x_\sigma, y_{\langle \rangle} \in E \text{ and } 0\}$ otherwise.

Suppose that there is a function $f : \{0, 1\}^{<h'} \to \{0, 1\}^{<h-1}$ such that, for each $\eta \in \{0, 1\}^{<h'}$, $c(f(\eta)) = 1$. Then for each $\eta \in \{0, 1\}^{<h'}$, set $y_\eta = y_{\langle \rangle} - f(\eta)$. For each $\tau \in \{0, 1\}^{h'}$, we have $\tau = \eta \prec (j)$ for some $j \in \{0, 1\}$; choose any leaf $\tau' \ni \langle \rangle \prec f(\eta) \prec (j)$ and set $x_\tau = x_\tau'$. Applying the inductive hypothesis to $\{x_\tau\} \tau \in \{0, 1\}^{h'}$, we obtain $x_1, \ldots, x_{d-1}, y_1, \ldots, y_{d-1}$ with $\{x_i, y_j\} \in E$ if and only if $i < j$. We may set $x_0 = x_\sigma$ and $y_0 = y_{\langle \rangle}$; then we also have $\{x_i, y_j\} \notin E$ for all $i$, and $\{0, y_i\} \in E$ for all $i > 0$.

Suppose there is no such $f$. Then the previous lemma gives us a function $f : \{0, 1\}^{<h'+1} \to \{0, 1\}^{<h-1}$ such that, for each $\eta \in \{0, 1\}^{<h'+1}$, $c(f(\eta)) = 0$. For each $\eta \in \{0, 1\}^{<h'}$, set $y_\eta = y_{\langle \rangle} - f(\eta)$. For each $\tau \in \{0, 1\}^{h'}$, we have $\tau = \eta \prec (j)$ for some $j \in \{0, 1\}$; choose any leaf $\tau' \ni \langle \rangle \prec f(\langle \rangle \prec (j))$. Then, by the inductive hypothesis again, we find $x_1, \ldots, x_{d-1}, y_1, \ldots, y_{d-1}$ with $\{x_i, y_j\} \in E$ if and only if $i < j$. We may set $x_d = x_\sigma$ and $y_d = y_{f(\langle \rangle)}$. Then we have $\{x_i, y_d\} \notin E$ for all $i$ and $\{x_i, y_d\} \in E$ for all $i < d$.

## 6.7 Stable Regularity

Since stable graphs have finite VC dimension, we know they belong to $B_{2,1}$. However stable graphs belong to $B_{2,1}$ in an unusually clean way: we can find approximations with no irregular pairs.

**Definition 6.24.** We say $A$ is $\epsilon$-good for $E$ if, for every $b \in V$, either $\frac{\mu(A \cap E_b)}{\mu(A)} < \epsilon$ or $\frac{\mu(A \cap E_b)}{\mu(A)} > 1 - \epsilon$.

**Lemma 6.25.** If $E$ is stable and $\mu(A) > 0$ then there is an $\epsilon$-good $A' \subseteq A$ with $\mu(A') \geq \epsilon^d \mu(A)$.

**Proof.** We construct a tree of subsets of $A$ inductively so that when $\eta$ is a sequence of length $n$, $\mu(A_\eta) \geq \epsilon^n \mu(A)$. Let $A_{\langle \rangle} = A$. Given $A_\eta$, if $A_\eta$ is not good then there is some $y_\eta$ so that $\epsilon < \frac{\mu(A_\eta \cap E_{y_\eta})}{\mu(A_\eta)} < 1 - \epsilon$. Then we may take $A_{\eta \prec (0)} = A_\eta \cap E_{y_\eta}$ and $A_{\eta \prec (1)} = A_\eta \setminus E_{y_\eta}$.

For each $\sigma \in 2^d$, pick an $x_\sigma \in A_\sigma$, which we can do since each $A_\eta$ has positive measure. Then when $\eta \sqsubseteq \sigma$, we have $x_\sigma \in E_{y_\eta}$ if and only if $\eta \prec (1) \sqsubseteq \sigma$, contradicting $d$-stability. \[\square\]
Therefore we have $B$ whenever $B$ is any other set, $A$ induces a partition of $B$.

**Definition 6.26.** When $B$ is $\delta$-good and $A$ is a set, write $A_{B,\delta}^0 = \{a \in A \mid \frac{\mu(A \cap E_a)}{\mu(A)} < \delta\}$ and $A_{B,\delta}^1 = \{a \in A \mid \frac{\mu(A \cap E_a)}{\mu(A)} > 1 - \delta\}$.

When $1/4 > \delta \geq \epsilon > 0$, we say $A$ is $(\epsilon, \delta)$-excellent if $A$ is $\epsilon$-good and whenever $B$ is $\delta$-good, either $\frac{\mu(A_{B,\delta}^0)}{\mu(A)} < \epsilon$ or $\frac{\mu(A_{B,\delta}^1)}{\mu(A)} < \epsilon$.

**Lemma 6.27.** If $A$ is $\delta^2$-good and $B$ is $(\delta^2, \delta^2)$-excellent then $(A, B, E)$ is $3\delta$-regular.

**Proof.** Let us assume $\frac{\mu(A_{B,\delta}^0)}{\mu(A)} < \delta^2$, since the other case is similar. Then $d_E(A, B) \geq (1 - \delta^2)^2$, since

$$d_E(A, B) = \frac{\mu(E \cap (A \times B))}{\mu(A)\mu(B)} \geq \frac{1}{\mu(A)\mu(B)} \int_{A_{B,\delta}^1} \mu(B \cap E_a) \, d\mu(a) \geq \frac{1}{\mu(A)\mu(B)} (1 - \delta^2)\mu(A_{B,\delta}^1) \geq \frac{(1 - \delta^2)}{\mu(A)\mu(B)}.$$

Whenever $A' \subseteq A$ and $B' \subseteq B$ with $\mu(A') \geq \delta\mu(A)$ and $\mu(B') \geq 3\delta\mu(B)$, we have $\mu(A' \cap A_{B,\delta}^0) < \delta^2\mu(A) \leq \delta\mu(A')/3$, and for $a \in A' \cap A_{B,\delta}^1$ we have $\mu(B' \setminus E_a) \leq \mu(B \setminus E_a) < 3\delta\mu(B) \leq \delta\mu(B')/3$, so $d_E(A', B') \geq (1 - \delta/3)^2$. Therefore

$$|d_E(A', B') - d_E(A, B)| \leq ((1 - \delta/3)^2 - (1 - \delta^2)^2) < \delta < 3\delta.$$

\qed

**Lemma 6.28.** If $E$ is stable, $\mu(A) > 0$, and $\delta < 1/2^d$ then there is an $(\epsilon, \delta)$-excellent $A' \subseteq A$ with $\mu(A') \geq \epsilon^d\mu(A)$.

We could weave together this proof with the proof that good sets exist to get a bound of $\mu(A') \geq \epsilon^d\mu(A)$.
6.8. INTERNAL CARDINALITY AND PSEUDOFINITE DIMENSION

Proof. Again, we will construct a tree of subsets of \( A \) inductively so that when \( \eta \) is a sequence of length \( n \), \( \mu(A_\eta) \geq \epsilon^n d \mu(A) \).

Let \( A_\emptyset = A \). Given \( A_\eta \) good, if \( A_\eta \) is not \((\epsilon, \delta)\)-excellent then there is a \( \delta \)-good set \( B \) so that both \( A_{B,\delta}^0 \) and \( A_{B,\delta}^1 \) have measure \( \geq \epsilon \). We take \( A_{\eta^-}(0) \) to be an \( \epsilon \)-good subset of \( A_{B,\delta}^0 \) and \( A_{\eta^-}(1) \) to be an \( \epsilon \)-good subset of \( A_{B,\delta}^1 \).

For each \( \sigma \in 2^d \), choose \( x_\sigma \in A_\sigma \). We claim that, for each \( \eta \in 2^{<d} \), we can choose \( y_\eta \) so that when \( \eta \wpreceq \sigma \), \( x_\sigma \in E_{y_\eta} \) exactly when \( \eta^-\langle 1 \rangle \wpreceq \sigma \).

For each \( \eta \in 2^{<d} \), there is a set \( B_\eta \) so that we have \( \frac{\mu(B_\eta \cap E_{x_\eta})}{\mu(B_\eta)} < \delta \) when \( \eta^-\langle 0 \rangle \wpreceq \sigma \) and \( \frac{\mu(B_\eta \cap E_{x_\eta})}{\mu(B_\eta)} > 1 - \delta \) when \( \eta^-\langle 1 \rangle \wpreceq \sigma \). Since \( \delta < 1/2^d \), it follows that

\[
0 < \mu(B_\eta \cap \bigcap_{\sigma \wpreceq \eta^-\langle 0 \rangle} (V \setminus E_{x_\sigma}) \cap \bigcap_{\sigma \wpreceq \eta^-\langle 1 \rangle} E_{x_\sigma}),
\]

so we may choose \( y_\eta \) in this set.

The existence of such a tree contradicts \( d \)-stability, so we find some \( \eta \) with \( A_\eta \) \((\epsilon, \delta)\)-excellent. \( \square \)

Theorem 6.29. If \( E \) is stable then there is a partition \( V = \bigcup_{i \leq k} A_i \) so that every pair \((A_i, A_j, E)\) is \( \epsilon \)-regular.

Proof. Successively choose \( A_1 \subseteq V \), \( A_2 \subseteq (V \setminus A_1) \), \( A_3 \subseteq (V \setminus (A_1 \cup A_2)) \) to be \((\epsilon^2/18, \epsilon^2/9)\)-excellent.

When \( m \) is large enough, \( \mu(V \setminus \bigcup_{i \leq m} A_i) < (\epsilon^2/18)\mu(A_1) \), so \( A_1 \cup \bigcup_{i \leq m} A_i \) is \((\epsilon^2/9, \epsilon^2/9)\)-excellent. For convenience, replace \( A_1 \) with \( A_1 \cup \bigcup_{i \leq m} A_i \).

Then each pair \( A_i, A_j \) is \( \epsilon \)-regular. \( \square \)

Examination of these arguments shows that in fact \( k \) can be a polynomial in \( 1/\epsilon \).

6.8 Internal Cardinality and Pseudofinite Dimension

So far we have focused on the density structure of finite graphs and their ultraproducts—that is, on sets with positive measure. However when we have a large finite set, we can also consider behavior at different orders of magnitudes.

Definition 6.30. When \( \mathcal{U} \) is an ultrafilter, the nonstandard real numbers are the ultraproduct \( [\mathbb{R}]_\mathcal{U} \).
That is, a nonstandard natural number is a sequence \( \langle r_n \rangle_{n \in \mathbb{N}} \) where we say \( \langle r_n \rangle_{n \in \mathbb{N}} \) represents the same sequence as \( \langle s_n \rangle_{n \in \mathbb{N}} \) if \( \{ n \mid r_n = s_n \} \in \mathcal{U} \). As always, we write \( [r_n]_\mathcal{U} \) for the equivalence class of \( \langle r_n \rangle_{n \in \mathbb{N}} \).

It is common to abbreviate the nonstandard reals as \( {}^*\mathbb{R} \); this notation can be misleading, because it implies that there is a single clearly defined object called \( {}^*\mathbb{R} \).

Just like finite sets have sizes which are a natural number, internal sets have “internal cardinality”, which is a nonstandard natural number:

**Definition 6.31.** The nonstandard natural numbers are the ultraproduct \([\mathbb{N}]_\mathcal{U}\)—that is, those \( [r_n]_\mathcal{U} \) such that \( \{ n \mid r_n \in \mathbb{N} \} \in \mathcal{U} \).

If \( S \subseteq [V_n]^k \) is internal, so \( S = [S_n]_\mathcal{U} \), the internal cardinality is \( |S| = |[S_n]_\mathcal{U}| \).

Internal cardinality interacts with internal functions in the way we expect:

**Lemma 6.32.** If \( A \) and \( B \) are internal sets,

- there is an internal injection \( f : A \to B \) if and only if \( |A| \leq |B| \),
- there is an internal surjection \( f : A \to B \) if and only if \( |B| \leq |A| \),
- there is an internal bijection \( f : A \to B \) if and only if \( |A| = |B| \).

The nonstandard real numbers have many of the properties we expect; we can add, multiply, and divide them, for instance. We can also compare them: we say \( [r_n]_\mathcal{U} < [s_n]_\mathcal{U} \) if \( \{ n \mid r_n < s_n \} \in \mathcal{U} \).

**Definition 6.33.** \( [r_n]_\mathcal{U} \) is bounded if there is a standard real number so that \( \{ n \mid |r_n| < r \} \in \mathcal{U} \).

If \( [r_n]_\mathcal{U} \) is bounded then \( \text{st}([r_n]_\mathcal{U}) \), the standard part of \( [r_n]_\mathcal{U} \), is \( \lim_{n \to \mathcal{U}} r_n \).

For instance, we have \( \mu_k(S) = \text{st}(|S|_\mathcal{U}) \); from this perspective, the measure \( \mu_k \) is the counting measure, suitably interpreted.

The nonstandard natural numbers satisfy a sort of induction principle for internal sets.

**Theorem 6.34.** Let \( S \) be a set of nonstandard natural numbers such that \( [0]_\mathcal{U} \in S \) and, whenever \( [k_n]_\mathcal{U} \in S \), \( [k_n + 1]_\mathcal{U} \in S \). If \( S \) is internal then \( S = [\mathbb{N}]_\mathcal{U} \).

We certainly need to demand that \( S \) be internal; for instance, if \( S \) consisted of the constant sequences \( [k]_\mathcal{U} \) for \( k \in \mathbb{N} \), \( S \) would contain \( 0 \) and be closed under successor, but would not contain \( [n]_\mathcal{U} \).
Proof. Since $S$ is internal, $S = [S_n]_U$. Suppose $\{n \mid S_n \neq N\} \in U$. Then, for each such $n$, there is a least $k_n \in N \setminus S_n$. Since $[0]_{n \in U} \in S$, $\{n \mid k_n > 0\} \in U$, so we may consider $[k_n - 1]_U$ and we have $[k_n - 1]_U \in S$. But $[k_n]_U \notin S$. 

The following variant is often more useful:

**Corollary 6.35.** If $S \subseteq [N]_U$ is internal and non-empty then $S$ has a least element.

Rather than focusing on sets of positive density, we can use the internal cardinality to characterize the “dimension” of a set, relative to $[V_n]_U$. For instance, if each $V_n$ is a set with $|V_n| = n$, we could consider subsets $X_n \subseteq V_n$ with $|X_n| \approx \sqrt{n}$. We can say that $X_n$ has “dimension 1/2”. This makes sense when we consider that a subset of $V_n$ with size $\approx n$ is a 1 dimensional set, a subset of $V_n^2$ with size $\approx n^2$ is a 2 dimensional set, and analogously, a subset of $V_n^k$ with size roughly $n^r$ should be an $r$ dimensional set—even when $r$ is a real number.

More precisely, to calculate the dimension we look at the logarithm: if $|X_n| \approx n^r$ then $\log |X_n| \approx r \log n$, so $r \approx \frac{\log |X_n|}{\log |V_n|}$.

**Definition 6.36.** Let $X = [X_n]_U$ be an interval subset of $V^k = [V_n^k]_U$. The (coarse) pseudofinite dimension of $X$ is defined to be

$$st \left( \frac{\log |X|}{\log |V|} \right) = \lim_{n \to U} \frac{\log |X_n|}{\log |V_n|}.$$ 

We write $\delta(X)$ for the pseudofinite dimension of $X$.

This is the only pseudofinite dimension we will consider, but we note that there is a broader family of pseudofinite dimensions; within that family, this is known as the “coarse” pseudofinite dimension.

**Lemma 6.37.** When $V = [V_n]_U$,

- $\delta(V^k) = k$,
- if $A, B \subseteq V$ are internal then $\delta(A \cup B) = \max\{\delta(A), \delta(B)\}$,
- if $Y \subseteq V^{k+m}$ and $Z \subseteq V^k$ are internal, $\delta(Z) = r$, and for all $\bar{v} \in Z$, $\delta(Y_{\bar{v}}) \leq s$ then $\delta(\bigcup_{\bar{v} \in Z} Y_{\bar{v}}) \leq r + s$,
- if $Y \subseteq V^{k+m}$ and $Z \subseteq V^k$ are internal, $\delta(Z) = r$, and for all $\bar{v} \in Z$, $\delta(Y_{\bar{v}}) \geq s$ then $\delta(\bigcup_{\bar{v} \in Z} Y_{\bar{v}}) \geq r + s$,
The last property is sometimes called the “fiber property”: it says that if we have a collection of sets \{Y_\vec{v}\}_{\vec{v} \in Z}, the size of the union is bounded by adding the number of sets in the collection and the size of each set. (Consider that the cardinality of the union should be bounded by the product of the cardinalities, and the logarithmic nature of dimension converts that to a sum.)

Proof. The first part follows directly from the definition. For the second part, if \(A = [A_n]_\mathcal{U}\) and \(B = [B_n]_\mathcal{U}\), note that \(\delta(A \cup B) = \lim_{n \to \mathcal{U}} \frac{\log |A_n \cup B_n|}{\log |V_n|}\),

and similarly for \(B_n\). We also have \(\log |A_n \cup B_n| \leq \log 2 \max\{|A_n|, |B_n|\}\) + \(\log 2\), and since \(\frac{\log 2}{\log |V_n|} \to 0\), we have \(\delta(A \cup B) \leq \max\{\delta(A), \delta(B)\}\) as well.

For the third part, we have

\[
\delta\left(\bigcup_{\vec{v} \in Z} Y_\vec{v}\right) = \lim_{n \to \mathcal{U}} \frac{\log |\bigcup_{\vec{v} \in Z_n} (Y_n)_\vec{v}|}{\log |V_n|}
\]

\[
\leq \lim_{n \to \mathcal{U}} \frac{\log |Z_n| \cdot \max_{\vec{v} \in Z_n} |(Y_n)_\vec{v}|}{\log |V_n|}
\]

\[
\leq r + s.
\]

The fourth part is analogous.

\[\square\]

**Lemma 6.38.** Let \(B \subseteq [V_n]_\mathcal{U}\) and \(E \subseteq [V_n]^2_\mathcal{U}\) be internal sets with \(E\) symmetric, and suppose that for each \(x \in B\), \(\delta(\{y \in B \mid (x,y) \in E\}) = 0\). Then there is an internal set \(I\) and internal function \(c : B \to I\) such that:

- \(\delta(I) = 0\),
- if \(b,b' \in B\) and \(c(b) = c(b')\) then \((b,b') \notin E\).

The finite version of this is not surprising: suppose that \(|B| = n\) and for each \(x \in B\) we have \(|\{y \in B \mid (x,y) \in E\}| \leq m\). Then we expect to be able to partition \(B\) into sets \(B_1, \ldots, B_{m+1}\) so that if \(x, x' \in B_i\) then \((x, x') \notin E\). Indeed, we could construct such a partition greedily: place \(B\) in an order and when we decide which partition component to put the \(j\)-th element of \(B\) in, there are at most \(m\) of them which have been ruled out, so we place this element in whichever one remains. The infinite version follows by doing this in the ultraproduct.
6.9. STABLE ERDŐS-HAJNAL

Proof. There are really two parts to this. The first is noting that, since $B$ is internal and $\delta(E_x) = 0$ for every $x \in B$, there is actually a supremum on the sizes of the $E_x$ which itself has dimension 0.

Indeed, let $S \subseteq [\mathbb{N}]_U$ be the set of nonstandard natural numbers $k$ such that, for all $x \in B$, $|E_x| \leq k$. This is an internal set (for instance, it can be expressed by a first-order formula using the predicate which holds of $x,k$ when $|E_x| \leq n$) and contains every $k$ with $\delta(k) > 0$, so has a least element $k$. If $\delta(k) > 0$ then $\delta(k - 1) > 0$, so $k \in S$ as well, so we must have $\delta(k) = 0$ for this least element.

Choose an internal set $I$ with $|I| = k + 1$. Now consider the set $S \subseteq [\mathbb{N}]_U$ of natural numbers $m$ such that there is an internal $U \subseteq B$ and $c : U \to I$ with $|U| = m$ so that if $b,b' \in U$ and $c(b) = c(b')$ then $(b,b') \notin E$. Clearly $[0] \in S$ since $|\emptyset| = [0]$ the existence of such a function for the empty set is trivial.

Consider the least $m \notin S$. If $m > |B|$ then we are done. Suppose $m \leq |B|$. Then $m - 1 \in S$, so we may choose $U \subseteq B$ with $|U| = m - 1$ and $c : U \to I$. Choose any $b \in B \setminus U$. Since $|E_b| < |I|$, there must be some $i \in I$ so that there is no $b' \in U$ with $(b,b') \in E$ and $c(b') = i$, so we may extend $U$ to $U \cup \{b\}$ by adding $c(b) = i$. This contradicts the assumption that $m \notin S$. \qed

6.9 Stable Erdős-Hajnal

Coarse pseudofinite dimension is the right setting for considering the Erdős-Hajnal conjecture, which says that, for every finite graph $H$, any graph with no induced copies of $H$ either has a large clique or an anti-clique.

Definition 6.39. If $G = (V,E)$ is a graph, we say $S \subseteq V$ is $E$-homogeneous if either $(S^2) \subseteq E$ or $(S^2) \cap E = \emptyset$.

That is, $S$ is either a clique in $E$ (every pair is an edge) or an anti-clique (no pair is an edge).

Conjecture 6.40 (Erdő-Hajnal Conjecture). For every finite graph $H$ there is a constant $d(H) > 0$ so that whenever $G = (V,E)$ is a finite graph with $T_H^{ind}(G) = \emptyset$, there is an $E$-homogeneous set $S \subseteq V$ with $\delta(S) \geq d(H)$.

While this is open, it is known that stable graphs—that is, graphs omitting the ladder of size $d$ for all $d$—do have large homogeneous sets. The ultraprodut version of this is the following.

Theorem 6.41. If $[G_n]_U = (V,E)$ is stable then there is an internal set $S$ with $\delta(S) > 0$ so that $S$ is $E$-homogeneous.
Proof. Let $G = [G_n]_U = (V, E)$. We will first show that there is a set $B$ with $\delta(B) > 0$ such that either for every $y \in B$, $\delta(\{x \in B \mid (x, y) \notin E\}) = 0$ or for every $y \in B$, $\delta(\{x \in B \mid (x, y) \notin E\}) = 0$.

We construct a tree of sets: let $A_0 = V$. Suppose we have constructed $A_\eta$ with $\delta(A_\eta) > 0$. If there is a $y \in A_\eta$ such that both $\delta(\{x \in A_\eta \mid (x, y) \in E\}) > 0$ and $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) > 0$ then let $y_\eta = y$, $A_{\eta^{-1}}(0) = \{x \in A_\eta \mid (x, y) \in E\}$ and $A_{\eta^{-1}}(1) = \{x \in A_\eta \mid (x, y) \notin E\}$.

If we can construct such a tree of size $d$, for each leaf $\sigma$ we have $\delta(A_\sigma) > 0$, so we may choose any $x_\sigma \in A_\sigma$ and we have a tree contradicting $d$-stability.

So we eventually obtain an $A_\eta$ so that, for every $y \in A_\eta$, either $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) = 0$ or $\delta(\{x \in A_\eta \mid (x, y) \in E\}) = 0$. Let $A_\eta^+$ be those $y$ such that $\delta(\{x \in A_\eta \mid (x, y) \in E\}) = 0$ and $A^-_\eta$ be those $y$ such that $\delta(\{x \in A_\eta \mid (x, y) \notin E\}) = 0$, so $A_\eta = A_\eta^+ \cup A^-_\eta$.

If $\delta(A^-_\eta) = 0$, we may take $B = A_\eta \setminus A^-_\eta$. Otherwise, for each $y \in A^+_\eta$, since $\delta(\{x \in A^-_\eta \mid (x, y) \in E\}) = 0$, $\delta(\{x \in A^+_\eta \mid (x, y) \notin E\}) = 0$. Therefore $\delta(\{(x, y) \in E \mid x \in A^-_\eta \text{ and } y \in A^+_\eta\}) = \delta(A^-_\eta) + \delta(A^+_\eta) > 0$.

But for each $x \in A^-_\eta$, $\delta(\{y \in A^+_\eta \mid (x, y) \in E\}) = 0$, so $\delta(\{(x, y) \in E \mid x \in A^-_\eta \text{ and } y \in A^+_\eta\}) \leq \delta(A^-_\eta) + 0 = \delta(A^-_\eta)$. Therefore $\delta(A^+_\eta) = 0$, so we may take $B = A_\eta \setminus A^+_\eta$.

Without loss of generality, let us assume we have a $B$ with $\delta(B) > 0$ such that for every $y \in B$, $\delta(\{x \in B \mid (x, y) \in E\}) = 0$. Then by Lemma 6.38, we have an $I$ with $\delta(I) = 0$ and a $c : B \to I$ so that if $c(b) = c(b')$ then $(b, b') \notin E$. Consider $Y = \{(i, b) \mid c(b) = i\}$. Since $\delta(B) = \delta(\bigcup_{i \in I} Y_i) \leq 0 + \sup_{i \in I} \delta(Y_i)$, there must be some $i \in I$ so that $\delta(Y_i) = \delta(B)$. Then $Y_i$ is the desired anti-clique.

As always, we can extract a finitary conclusion:

**Theorem 6.42.** For every $d$ there is a $\delta > 0$ so that whenever $G = (V, E)$ is $d$-stable and sufficiently large, there is an $S \subseteq V$ with $|S| \geq |V|^\delta$ so that $S$ is $E$-homogeneous.

Proof. Suppose not. Then there is a $d$ so that for every $n$ there is a $G_n = (V_n, E_n)$ with $n$ vertices which is $d$-stable but every clique or anti-clique has size $< |V||e|/n$.

Let $G = [G_n]_U$. Since being $d$-stable is a first-order property and all the $G_n$ are $d$-stable, $G$ is $d$-stable. Then there is an internal homogeneous set $S = [S_n]_U$ with $\delta(S) > 0$. Therefore we may choose an $n$ so $S_n$ is homogeneous and $|S_n| \geq |V||e|/n$, contradicting our assumption. □
6.10 Compression Schemes

We can think of stability as one way for graphs with finite VC dimension to be particularly nice. We now consider a different way that a graph of finite VC dimension might be particularly nice.

**Definition 6.43.** Let $E \subseteq V \times W$ be a graph. We say $E$ is compressible if there is a $k$ and a $C \subseteq V \times W^k$ such that whenever $B \subseteq W$ is finite with $|B| \geq 2$ and $a \in V$, there are $(b_1, \ldots, b_k) \in B^k$ so that $(a, b_1, \ldots, b_k) \in C$ and for every $b \in B$, either

- $C_{b_1, \ldots, b_k} \subseteq E_b$, or
- $C_{b_1, \ldots, b_k} \cap E_b = \emptyset$.

We call $C$ a compression scheme for $E$.

The idea is that $b_1, \ldots, b_k$ is a bounded amount of information which encodes the set $E_a \cap B$.

For instance, suppose $V = W = [0, 1]$ and $E = \{(x, y) \mid x < y\}$. Then we can take $k = 2$ and $C = \{(a, b_1, b_2) \mid b_1 \leq a < b_2\} \cup \{(a, b_1, b_1) \mid a < b_1\}$. Whenever $B \subseteq W$ is a finite set and $a \in V$, if $a \leq \min B$ then we take $b_1 = \min B$ and have $(a, b_1, b_1) \in C$, so $C_{b_1, b_1} = [0, b_1) \subseteq [0, b)$ for all $b \in B$. If $a \geq \max B$ then we take $b_1 = \max B$ and $b_2 \in B$ to be any other element; then $(a, b_2, b_1) \in C$ and $C_{b_1, b_2} = [b_1, 1]$, which is disjoint from $E_b$ for all $b \in B$. Otherwise, we choose $b_1 = \max B \cap [0, a]$ and $b_2 = \min B \cap (a, 1]$, so $(a, b_1, b_2) \in C$ and $C_{b_1, b_2} = [b_1, b_2)$, which is contained in $E_b$ for every $b \in B$ with $b_2 \leq b$ and disjoint from $E_b$ for every $b \in B$ with $b \leq b_1$.

On the other hand, any graph with infinite VC dimension fails to be compressible.

**Lemma 6.44.** If $E \subseteq V \times W$ is compressible then $E$ has finite VC dimension.

**Proof.** Suppose $C \subseteq V \times W^k$ be a compression scheme for $E$. Let $B \subseteq W$ with $2|B| > |B|^k$.

Observe that each choice of $b_1, \ldots, b_k \in B$ defines a subset of $B$, namely $S_{b_1, \ldots, b_k} = \{b \mid E_b \supseteq C_{b_1, \ldots, b_k}\}$. We will show that, for every $a$, $E_a \cap B$ is one of these sets; since there are not enough such sets to shatter $B$, it follows that $E$ has finite VC dimension.

Take any $a \in V$. Since $C$ is a compression scheme, there must be $b_1, \ldots, b_k$ with $(a, b_1, \ldots, b_k) \in C$. If $b \in E_a \cap B$ then we have $a \in E_b \cap C_{b_1, \ldots, b_k}$, so $E_b \cap C_{b_1, \ldots, b_k}$, so we must have $E_b \supseteq C_{b_1, \ldots, b_k}$, so $b \in S_{b_1, \ldots, b_k}$. Conversely, if $b \in S_{b_1, \ldots, b_k}$ then $a \in C_{b_1, \ldots, b_k} \subseteq E_b$. Therefore $E_a \cap B = S_{b_1, \ldots, b_k}$. 

\qed
In fact, modulo some technicalities about coding, the converse holds: compressibility and finite VC dimension are equivalent \[74\]. However we are interested in a slightly stronger notion: when a certain set system associated with the compression scheme also has finite VC dimension.

**Definition 6.45.** When \( C \subseteq V \times W^k \) is a compression scheme for \( E \subseteq V \times W \), we say \( d \in W \) crosses \( \vec{b} \) if both \( C_{\vec{b}} \cap E_d \neq \emptyset \) and \( C_{\vec{b}} \setminus E_d \neq \emptyset \).

For each \( \vec{b} \in W^k \), we define \( Cr(C, \vec{b}) \) to be the set of \( d \in W \) which cross \( \vec{b} \).

We say \( E \) is *distally compressible* if there is a compression scheme for \( E \) such that the set system \( \{Cr(C, \vec{b})\}_{\vec{b} \in W^k} \) has finite VC dimension.

It is rather difficult to produce an example of a set with finite VC dimension which is not distally compressible. In fact, I know of only one essential example, and it involves a graph of measure 0.

**Theorem 6.46 ([18]).** Let \( p \) be a prime and let \( \mathbb{F} \) be the algebraic closure of \( \mathbb{F}_p \). Let \( P \subseteq \mathbb{F}^2 \) be the set of points and let \( L \) be the set of lines in \( \mathbb{F}^2 \). Then the incidence relation \( I \subseteq P \times L \) consisting of pairs \( (p, \ell) \) where the point \( p \) is on line \( \ell \) is not distally compressible.

This example is known to be stable.

In fact, distal compressibility is derivated from the model theoretic notion of *distality* \[94\]. A combinatorial characterization of distality given in \[15\] says that a theory is distal if every graph defined in the theory has a compression scheme which is itself definable in the theory. The original purpose of distality was to capture the notion of NIP theories (that is, theories where every definable set has finite VC dimension) which are “purely unstable”; indeed, non-trivial theories cannot be both distal and stable. However distality is fundamentally different in character from stability; in particular, it is not monotone in the theory—we could have a theory which is not distal because the compression schemes are not definable, but by adding additional formulas, the theory becomes distal. (And in the process, we might expand a stable theory to an unstable one because the new formulas are unstable.)

Distally compressible graphs satisfy a strong form of the Erdős-Hajnal conjecture, at least if we allow bipartite sets rather than single ones.

**Theorem 6.47.** If \( G \) is a distally compressible measurable graph then there is an \( \delta > 0 \) and sets \( X, Y \) with \( \mu_1(X) \geq \delta \) and \( \mu_2(Y) \geq 1/4 \) so that which are \( E \)-homogeneous—that is, either \( X \times Y \subseteq E \) or \( (X \times Y) \cap E = \emptyset \).
Indeed, the proof will show that we can even ensure that \( Y \) has measure close to 1/2.

**Proof.** Fix a compression scheme \( C \subseteq V \times W^k \) so that \((W, \{Cr(C, \vec{b})\}_{\vec{b} \in W^k})\) has finite VC dimension. Let \( \epsilon = 1/2 \). Let \( \delta = 1/c^k \) where \( c \) is large enough that there is an \( \epsilon \)-net of size \( \leq c \).

Let \( B \subseteq W \) be an \( \epsilon \)-net with \( |B| \leq c \), so for every \( \vec{b} \in W^k \) (note that \( \vec{b} \) need not be from \( B^k \)), if \( \mu_1(Cr(C, \vec{b})) \geq \epsilon \) then \( B \cap Cr(C, \vec{b}) \neq \emptyset \).

For every \( a \in V \), there must be some \( \vec{b} \in B^k \) so that \( a \in C_\vec{b} \) and \( Cr(C, \vec{b}) \cap B = \emptyset \). Therefore there must be some \( \vec{b} \in B^k \) so that \( Cr(C, \vec{b}) \cap B = \emptyset \) and \( \mu_1(C_\vec{b}) \geq 1/|B|^k \geq \delta \).

Let \( D_0 = \{ d \mid C_\vec{b} \subseteq E_d \} \) and \( D_1 = \{ d \mid C_\vec{b} \cap E_d = \emptyset \} \). We have \( W = D_0 \cup D_1 \cup Cr(C, \vec{b}) \). Since we chose \( B \) to be an \( \epsilon \)-net, we must have \( \mu_1(Cr(C, \vec{b})) < \epsilon \), so there is an \( i \in \{0, 1\} \) with \( \mu_1(D_i) > \frac{1}{2}(1 - \epsilon) \). Since both \( C_\vec{b} \times D_i \) are homogeneous, we are done. \( \square \)

**Theorem 6.48** (Regularity for Distal Graphs). For each \( \epsilon > 0 \) there is an \( n \) so that whenever \( G = (V, E) \) is a graph with \( E \) is distally compressible, there is a partition \( V = \bigcup_{i \leq k} A_i \) so that either \( A_i \times A_j \subseteq E \) or \( A_i \times A_j \cap E = \emptyset \) except for a set \( R \) of \( (i, j) \) so that \( \mu_2(\bigcup_{(i, j) \in R} A_i \times A_j) < \epsilon \).

**Proof.** We begin with the trivial partition \( V = V \). Given a partition \( V = \bigcup_{i \leq k} A_i \), let \( R(\{A_i\}) \) be the set of \( (i, j) \) so that \( A_i, A_j \) is not \( E \)-homogeneous and let \( m(\{A_i\}) = \mu_2(\bigcup_{(i, j) \in R(\{A_i\})} A_i \times A_j) \).

Consider some partition \( V = \bigcup_{i \leq k} A_i \). Take any pair \( A_i \times A_j \) which is not \( E \)-homogeneous. The previous theorem gives us subsets \( B_i \subseteq A_i \) and \( B_j \subseteq A_j \) with \( \mu_1(B_i) \geq \delta \mu_1(A_i) \) and \( \mu_1(B_j) \geq \delta \mu_1(A_j) \). Consider the partition \( V = \bigcup_{i \leq k+2} A'_i \) where we replace \( A_i \) with \( B_i, A_i \setminus B_i \) and \( A_j, A_j \setminus B_j \). Then

\[
m(\{A'_i\}) \leq m(\{A_i\}) - \frac{\delta}{4} \mu_1(A_i) \mu_1(A_j).
\]

Applying this repeatedly, we eventually obtain a partition with \( m(\{A_i\}) < \epsilon \). \( \square \)

### 6.11 Remarks

VC dimension is a notion which has been reinvented a very large number of times in a variety of contexts, including computational learning theory [110], model theory [92], and combinatorics [89]. In model theory, the notion of “finite VC dimension” is often called “NIP” (standing for “not the
independence property”). (More precisely, one typically calls a theory NIP if all its formulas define sets of finite VC dimension; the development of NIP theories and VC dimension were completely separate for almost twenty years before Laskowski identified the connection [68].)

One might ask whether the VC dimension characterizes the rate of growth of the function of $\pi_G(n)$. In fact, it does not: there are examples where $G$ has VC dimension $d$ but the function $\pi_G(n)$ grows at a much slower than $n^d$. The infimum of those $r$ such that $\pi_G(n) \in O(n^r)$ is called the VC density of $G$, and can be much lower than the VC dimension (and need not be an integer).

The proof we give of the existence of an $\epsilon$-net is essentially taken from [73], which is an excellent reference for the geometric consequences of finite VC dimension. The existence of $\epsilon$-approximations is known as the Glivenko-Cantelli property; the relationship between various forms and strengthenings of the Glivenko-Cantelli property and combinatorial characterizations has been extensively studied [6, 26, 101, 102].

Stability was the first model theoretic property of its type to be discovered, and has been a central topic in model theory for several decades [80, 81]. The connection between properties like stability, bounded VC dimension, and distality on the one hand, and regularity on the other, has been extensively investigated in the setting of particular (usually algebraic) structures having these properties [33, 34, 105] as well as from the analytic and model theoretic perspective [1, 16, 17, 18, 21, 70, 71, 72]. Many of these results include quantitative bounds of the sort we have omitted here.

The notion of pseudofinite dimension is introduced in [57] and developed in [39, 40, 43, 56].
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


