Lecturer: Emmanuel Abbe Scribe: Kiran Vodrahalli

Contents

1 Introduction to Random Graphs

Definition 1.1. $g = (V, E)$ is a graph means V is a finite set, typically $V = [n]$, known as the vertex set. E is the edge set, which is a subset of $\binom{V}{2}$ (pairs of elements, unordered $\implies$ undirected). We don't allow loops or multiedges.

Remark 1.2. If you have a graph g, then we use the notation $V(g)$ and $E(g)$ to denote the vertex and edge sets.

Definition 1.3. We say $g \subseteq g'$ if $V(g) = V(g')$ and $E(g) \subseteq E(g')$.

1.1 Erdős-Rényi random graphs

There are two models which are called Erdős-Rényi:

Definition 1.4. Uniform model.
$\mathcal{G}(n,m)$ where $0 \leq m \leq \binom{n}{2}$. This is a distribution on graphs with the vertex set $V = [n]$: If $G \sim \mathcal{G}(n,m)$, $\mathbb{P}\{G = g\} = 0$ if $|E(g)| \neq m$, and $\frac{1}{\binom{\binom{n}{2}}{m}}$ otherwise. We uniformly draw a graph with exactly m edges.

Definition 1.5. Binomial model.
$G(n,p)$ where $0 \leq p \leq 1$, $V = [n]$. If $G \sim G(n,p)$, $\mathbb{P}\{G = g\} = p^{|E(g)|}(1-p)^{\binom{n}{2}-|E(g)|}$. In other words, we draw each edge independently with probability p.

Remark 1.6. In many cases, these 2 models lead to the same conclusions. If $G_1 \sim \mathcal{G}(n,m)$ then $|E(G_1)| = m$ with probability 1. If $G_2 \sim G(n,p)$, then $|E(G_2)| \sim \text{Bin}\left(\binom{n}{2}, p\right)$ then $\mathbb{E}[|E(G_2)|] = p\binom{n}{2}$ and $\text{var}(|E(G_2)|) = \binom{n}{2}p(1-p)$. If the first and second moments of the binomial are model are the same (which they will often be in this class) since we consider p vanishing or constant, then we get concentration properties.

$$\mathbb{P}\left\{\frac{||E(G_2)| - \mathbb{E}[|E(G_2)|]|}{\sqrt{\text{var}(|E(G_2)|)}} > 0\right\} \to 0$$

Typically we will work with the binomial model.

We know that $|E(G_2)| = \sum_i X_i$ where $X_i \sim \text{Ber}(p)$. By the Law of Large Numbers, $\frac{1}{n}\sum_i X_i \to \mathbb{E}[X_i] = p$ which then gives $\mathbb{P}\{|\sum_i X_i - np| > \epsilon n\} \to 0$. We can think of this as saying the two models get close to each other.

We can represent the graph as a Boolean function f. A graph g can be viewed as a Boolean vector $x \in \{0,1\}^{\binom{n}{2}}$.

Definition 1.7. Graph properties.
A graph property is a subset of graphs. Notation: A_n is a graph property, i.e. a collection of graphs on $V = [n]$.

Example 1.8. $A_n = \{\text{triangle containment}\}$ means formally that A_n is all the graphs such that each graph has $V = [n]$ and contains a triangle.

Definition 1.9. A graph property is

1. symmetric (invariant on the labeling of the vertices)
2. increasing (if you add edges to the graph, the graph will keep its property): if $g \in A$, then $g \cup \{e\} \in A$.
3. decreasing (if you add edges and the graph doesn't have the property, the graph still doesn't have its property): if $g \notin A$, then $g \cup \{e\} \notin A$.

Definition 1.10. Boolean representation of a graph property.
Take $f : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$ for A, graph property. Then $f = \mathbf{1}_A$.

Remark 1.11. If A is increasing, then $\mathbf{1}_A$ is increasing if $x \leq y$, $x, y \in \{0,1\}^{\binom{n}{2}}$ implies $\mathbf{1}_A(x) \leq \mathbf{1}_A(y)$.

Definition 1.12. $\mu_p(A_n) = \mathbb{P}\{G \in A_n\}$ when $G \sim G(n,p)$ (binomial model).

Lemma 1.13. *If A_n is increasing, then $p \in [0,1]$ gets mapped to $\mu_p(A_n)$ is increasing. When p moves, then probability that you have the property can only increase. It is not obvious to show.*

Proof. We begin with continuity. $\mu_p(A_n) = \sum_{g \in A_n} \mathbb{P}\{G = g\} = \sum_{m=1}^{\binom{n}{2}} \sum_{g \in A_n : E(g) = m} p^m (1-p)^{\binom{n}{2} - m}$. This equation is a polynomial in p so this is a continuous function. It is not obvious to say why this is increasing.

To show it is increasing, we use coupling. If you take $0 \leq p, q \leq 1$ and take $G_1 \sim G(n,p)$, $G_2 \sim G(n,q)$. Now construct $G_\Delta \sim G(n,\delta)$. Let us look at $G_1 \cup G_\Delta$ and say they are independent. Then $G_1 \cup G_\Delta \sim G(n, 1-(1-p)(1-\delta))$. Then we can choose δ such that $1-(1-p)(1-\delta) = q$, giving $\delta = \frac{q-p}{1-p}$ (note $q > p$). Then $\mu_q(A_n) = \mathbb{P}\{G_2 \in A_n\} = \mathbb{P}\{G_1 \cup G_\Delta \in A_n\} \geq \mathbb{P}\{G_1 \in A_n\} = \mu_p(A_n)$ by using the increasing property (if you add edges it can only be more likely that you will get the property). □

Pretty much everything in this class will be monotone. Thus, think of $\mu_p(A_n)$ as a probability measure which is increasing on the interval $[0,1]$. As you change the n and n becomes large, a phase transition appears: A threshold such that if p is less than the threshold, it is very low-probability, and if p is greater than the threshold, it is high probability. It turns out that this threshold is in fact very sharp (step-function). For instance, connectivity has such a jump but triangle containment does not.

Definition 1.14. Threshold.
From now on, $\{A_n\}_{n \geq 1}$ is a sequence of symmetric increasing graph properties. $\{\hat{p}_n\}_{n \geq 1}$ is a threshold for $\{A_n\}_{n \geq 1}$ if when you take $p_n << \hat{p}_n$, then $\mu_{p_n}(A) \to 0$ as $n \to \infty$ and if $p_n >> \hat{p}_n$ then $\mu_{p_n}(A) \to 1$ as $n \to \infty$. Our notation means

1. $a_n = o(b_n)$, or $a_n << b_n$ if $\frac{a_n}{b_n} \to 0$
2. $a_n = O(b_n)$ means that for n large enough, there exists c such that $a_n \leq c b_n$. For instance, choosing $a_n = \frac{1}{n}, b_n = \frac{2}{n} - \frac{1}{n^2}$ gives this result.
3. $a_n = \Omega(b_n)$ means that for n large enough, there exists c such that $a_n \geq c b_n$.
4. $a_n = \Theta(b_n)$ holds if $a_n = O(b_n), a_n = \Omega(b_n)$. Alternatively write $a_n \asymp b_n$.

Theorem 1.15. *If A_n is symmetric increasing, then there exists a threshold. (Bollobas-Thomasson)*

Remark 1.16. By the definition of a threshold, this threshold is only unique up to constants. If $\hat{p}$ is a threshold, then it must be that $c\hat{p}$ is also a threshold. Note that this theorem does not tell you that the threshold is sharp. This is called the coarse threshold. As long as it is increasing, you have this for free. After that, we will discuss conditions for the sharp threshold.

A_n is an increasing property, $\hat{p}_n$ is a threshold if $\mu_{\hat{p}_n}(A_n) = 0$ if $p_n << \hat{p}_n$ and 1 otherwise, where $G \sim G(n,p)$.

The first lemma we will use says that the converse of this is also true.

Lemma 1.17. *If A_n is increasing and let $\mathbb{P}_n\{\epsilon\}$ such that $\mu_{\mathbb{P}_n\{\epsilon\}}(A_n) = \epsilon$ for $\epsilon \in (0,1)$. Then $\hat{p}_n \asymp \mathbb{P}_n\{\epsilon\}$ iff $\hat{p}_n$ is a threshold for A_n.*

Remark 1.18. Since $\mu_{\mathbb{P}_n\{\epsilon\}}$ is an increasing function, you get a single preimage.

Proof. First assume that $\hat{p}_n$ is a threshold for A_n. Then if $p << \hat{p}_n$ then $\mu_p(A_n) \to 0$, and if $p >> \hat{p}_n$ then $\mu_p(A_n) \to 1$. This implies that $\hat{p}_n \asymp \mathbb{P}_n\{\epsilon\}$ for all ϵ. If $\hat{p}_n \not\asymp \mathbb{P}_n\{\epsilon\}$ for some ϵ, then either $\frac{\hat{p}_n}{\mathbb{P}_n\{\epsilon\}} \to 0$ or $\to 1$, given a contradiction. Each time, we should say that there exists a subsequence (if limit not defined), and it's on that subsequence we get a contradiction.

For the other direction, assume that $\hat{p}_n \asymp \mathbb{P}_n\{\epsilon\}$ for all ϵ. But $\hat{p}_n$ is not a threshold: Take p s.t. $p << \hat{p}_n$ and $\mu_p(A_n) \to a > 0$, and $p >> \hat{p}_n$ and $\mu_p(A_n) \to 1 - b < 1$ for some $b > 0$. Then $p \le p(a)$ or $p \ge p(1-b)$on the subsequences, thus $a = \mu_{\mathbb{P}\{(\}a)}(A_n)$. □

Now we prove the Bolllobas-Thomasson theorem: Every monotone graph property has a threshold.

Proof. Let A_n be increasing, let $\epsilon \in (0,1)$, $G_1, G_2, \cdots, G_m \sim G(n, \mathbb{P}\{\epsilon\})$ where $\mu_{\mathbb{P}\{\epsilon\}}(A_n) = \epsilon$ and $\sim$ is i.i.d. We have $\mathbb{P}\{\cup_{i=1}^m G_i \in A_n\}$, where the union G' is drawn from $G(n, (1-(1-\mathbb{P}\{\epsilon\})^m))$. Then, $1-(1-\mathbb{P}\{\epsilon\})^m \le m\mathbb{P}\{\epsilon\}$ Thus, if we replace with this other probability, it is more likely and we have $\mathbb{P}\{\cup_{i=1}^m G_i \in A_n\} \le \mu_{m\mathbb{P}\{\epsilon\}}(A_n)$. Then note that $1-\mathbb{P}\{G_1 \notin A_n\}^m = 1-(1-\mathbb{P}\{G_1 \in A_n\})^m = 1-(1-\epsilon)^m \ge 1-\epsilon = \mu_{p(1-\epsilon)}(A_n)$, we need $(1-\epsilon)^m \le \epsilon$, which is implied by $m \ge \frac{\log(\epsilon)}{\log(1-\epsilon)}$. Thus, $\mathbb{P}\{\epsilon\} \le \mathbb{P}\{1-\epsilon\} \le m\mathbb{P}\{\epsilon\}$. If $\epsilon < 1/2$ then $\mathbb{P}\{1-\epsilon\} \asymp \mathbb{P}\{\epsilon\}$ for all $\epsilon \in (0, 1/2]$. □

1.2 Sharp versus Coarse Thresholds

Definition 1.19. Width of a threshold.
For $\epsilon \in (0,1/2]$ define $\delta(\epsilon) = \mathbb{P}\{(\}1-\epsilon) - \mathbb{P}\{\epsilon\}$ where $\mathbb{P}\{\cdot\}$ is defined for a monotone property. From before we had $\delta(\epsilon) = \mathcal{O}(\mathbb{P}\{\epsilon\})$. The question is, $\delta(\epsilon) = \Theta(\mathbb{P}\{\epsilon\})$? Or $\delta(\epsilon) = o(\mathbb{P}\{\epsilon\})$?

Definition 1.20. $\hat{p}$ is coarse if $\exists \epsilon \in (0,1/2]$ s.t. $\delta(\epsilon) = \Theta(\hat{p})$, and sharp if it equals $o(\hat{p})$. From previous lemma, any $\hat{p}$ suffices.

Remark 1.21. Following statements are equivalent: $\hat{p}$ is sharp and $\frac{\mathbb{P}\{1-\epsilon\}}{\mathbb{P}\{\epsilon\}} \to 1$ for all $\epsilon \in (0,1/2]$: In other words, the probabilities must agree on constants for sharp thresholds. We have $\hat{p}_1/\hat{p}_2 \to 1$. $\hat{p}$ is unique up to $\sim$. It is also equivalent to saying $\mu_p(A_n) \to 0$ if $p \le \hat{p}(1-\delta)\ \forall \delta > 0$ or 1 if $p \ge \hat{p}(1+\delta)$. Let us go through some general results about monotone properties.

1.3 General results about monotone properties

1. $\delta\epsilon = \mathcal{O}(\mathbb{P}\{\epsilon\})$
2. $\delta\epsilon = \mathcal{O}(\frac{1}{\log n})$ (Kalai '96)
3. $\delta(\epsilon) = \mathcal{O}(\frac{1}{(\log n)^{2-\delta}})$ (Bourgain and Kalai). It is believed the best you can do is $\mathcal{O}(\frac{1}{(\log n)^2})$. These are only good for very slow thresholds, otherwise the first bound is better.
4. $\delta(\epsilon)/\mathbb{P}\{\epsilon\} \leq c_\epsilon \frac{\log(2/\mathbb{P}\{e\})}{\log(n)}$. For instance, if $\hat{p} = e^{-o(\log n)}$, then this is slower than $1/n$.

The goal now is the following. We would like to understand when we have a coarse or a sharp threshold when we have a monotone property. We will often have much faster decaying thresholds.

Example 1.22. Emptiness.
We will argue this is sort of typical of coarse thresholds. This is the property that $A_n = \{\text{emptyness}\}$, no edges on $V = [n]$. What is the probability this takes place? $\mu_p(A_n) = (1-p)^{\binom{n}{2}}$. If $p_c = c/\binom{n}{2}$, then $\mu_{p_c}(A_n) \sim e^{-c}$ and thus $p_{1/2}$ is a coarse threshold (we have exhibited a threshold as per Bollobas Theorem). This is basically saying we are K_2 free graph. What happens if we do K_3-free graphs (i.e. triangle free)? We handle this next week. We have $|E(G)| \sim Ber(\binom{n}{2}, p)$. We will eventually deal with any finite subgraph. For coarse threshold, you have to allow a bag of some finite subgraphs, then if you take this bag of models to be large enough, you can approximate the monotone property with this local thing.

Example 1.23. Connectivity.
An example which is a sharp threshold is connectivity: You cannot check this locally, you need to check the whole graph. It is a sharp threshold at $\frac{\log n}{n}$.

It is enough to look at in finite graphs that you count expected number of copies. We have $\hat{p} = n^{-\alpha}$, where $\alpha \in \mathbb{Q}$ for K_2.

Example 1.24. How do you connect coding up to what we have discussed now? Recall a graph can be expressed as a binary boolean vector. Consider the property $A_n = \{z \in \{0,1\}^{n-1} : \exists x \in C, x_1 = 1, x_2^n \leq 3\}$. C is a linear code; $x \in C$ is a codeword. Then we want to ask what is the probability we decode the first codeword wrongly. Can we cover a vector which has 1 in the first component ($x_1 = 1$), with a noise pattern where we replace erasures with 1s (this is the $x_2^n \leq 3$)? The probability of erasures in Shannon setting is not erasing. There is a missing piece that you can achieve the result you can achieve capacity. Just by having a code which is linear, this has enoguh symmetry that you can apply the theorem. The other thing you need is to be able to tell where the threshold is. You can locate the threshold with a nice trick.

2 Triangle Containment

Recall last time that every monotone property has a threshold, which is either coarse ($\delta(\epsilon) = \Theta(p(\epsilon))$) or sharp $\delta(\epsilon) = o(p(\epsilon))$.

Example 2.1. "Containing an edge": The # of edges $\sim Bin(\binom{n}{2}, p)$ in $G(n,p)$. We have that $p = c/\binom{n}{2}$ is a coarse threshold. And in fact, we will have that as $n \to \infty$, the # of edges converges in distribution to Pois(c).

The goal today is to show that this type of result will go through.

Definition 2.2. The triangle continment property is $A_n = \{\exists i, j, k \in [n]$ all different s.t. $E_{i,j} = E_{j,k} = Ei, k = 1\}$.

Theorem 2.3. *$\frac{1}{n}$ is a coarse threshold for A_n.*

We showed for edges it was $c/\binom{n}{2} \sim 1/n^2$, here we show it is $1/n$.

2.1 Method of Moments

We say $X \sim \mathbb{R}_+$, $X \sim \mu$.

Theorem 2.4. *Markov's Inequality.*
$\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[X]}{a}$

Proof. Assume X continuous.

$$\mathbb{E}[X] = \int_{R_+} xf(x)dx \geq \int_a^\infty xf(x)dx$$

where $f(x)$ is the probability density function. Then

$$\int_a^\infty xf(x)dx \geq a\int_a^\infty f(x)dx = a\mathbb{P}\{X \geq a\}$$

Thus we get the result by taking the lower bound of x out of the integral. □

Theorem 2.5. *Chebychev Inequality.*
$\mathbb{P}\{|X - \mathbb{E}[X]| \geq a\} = \mathbb{P}\{|X - \mathbb{E}[X]|^2 \geq a^2\} \leq \frac{Var(X)}{a^2}$

which is proven directly by applying Markov's Inequality.

Lemma 2.6. *Let $Z \sim \mathbb{Z}_+$ (non-negative integers). Then*

1. $\mathbb{P}\{Z \geq 1\} \leq \mathbb{E}[Z]$
2. $\mathbb{P}\{Z = 0\} \leq \frac{Var(Z)}{\mathbb{E}[Z]^2}$

Proof. 1. Direct application of Markov's Inequality.

2. $\mathbb{P}\{Z = 0\} = \mathbb{P}\{Z - \mathbb{E}[Z] = -\mathbb{E}[Z]\} \leq \mathbb{P}\{|Z - \mathbb{E}[Z]| \geq \mathbb{E}[Z]\}$.

□

When you rely on these inequalities it is sometimes referred to as method of first and second of moments.

Now we prove the first theorem we stated about the triangles.

Proof. Let $Z_k = \#$ of triangles in $G(n,p) = \sum_{T \in \binom{[n]}{3}} \mathbf{1}(T \subset G)$. $\mathbb{E}[Z_n] = \binom{n}{3}p^3$. Then, if $p << \frac{1}{n}$, $\mathbb{E}[Z_n] \to 0$. By the lemma, $\mathbb{P}\{Z_n \geq 1\} \to 0$ if $p << \frac{1}{n}$. Since we know triangle containment is a monotone property, then we know there is a threshold. Thus $\frac{1}{n}$ is a lower bound on what the threshold can be. Then $Var(Z_n) = \sum_{S,T \in \binom{[n]}{3}} Cov(\mathbf{1}(S \subset G), \mathbf{1}(T \subset G))$. There are three cases:

1. $S \cap T = \emptyset$. There is no covariance, so covariances are all 0.
2. $|S \cap T| = 1$. You have to overlap on one edge, so you must choose how to choose the five edges. The number of cases is $c\binom{n}{4}$, and the probability is p^5 for any such case.
3. $|S \cap T| = 3$; the triangles are on exactly the same location. There will be $\binom{n}{3}$ such cases, and the probability here will be p^3.

Then, $Var(Z_n) \leq c\binom{n}{4}p^5 + \binom{n}{3}p^3$ is an upper bound on the variance. If $p >> \frac{1}{n}$, then you can check that $\frac{Var(Z_n)}{(n^3p^3)^2} \to 0$. By the lemma, we have proved that there are triangles since $\mathbb{P}\{Z = 0\}$ goes to 0, implying that $\frac{1}{n}$ is a threshold. □

Note that if we look at the case $p = \frac{c}{n}$, then $\mathbb{E}[Z_n] = \frac{c^3}{6}$ which implies coarse threshold. Now we ask about the probability distribution.

Theorem 2.7. *When $p = \frac{c}{n}$, $Z_n \to Pois(\frac{c^3}{6})$ in distribution.*

Definition 2.8. Converge in distribution.
If $X_n \sim \mathbb{R}$, then $X_n \to^{(d)} X$ if $\mathbb{P}\{X_n \leq x\} \to \mathbb{P}\{X \leq x\}$ for all x that are continuity points of $n \to \mathbb{P}\{X \leq n\}$.

If X is an integer, then convergence in distribution just means that if $Z_n \to^{(d)} Z$, $\mathbb{P}\{Z_n = k\} \to \mathbb{P}\{Z = k\}$ as $n \to \infty$ for all $k \in \mathbb{Z}_+$.

Now I will explain the true Method of Moments. Before, we only considered expectation and variance. When we try to prove things tend to each other in distribution, we instead try to prove that the moments tend to each other in the limit. And if a distribution is parametrized by its moments, this implies convergence in distribution.

Theorem 2.9. *If Z has a distribution defined by its moments, if Z_n has finite moments for all order of the moments (i.e. $\mathbb{E}[Z_n^k]$ is finite for every n), then if for all $k \geq 1$ $\mathbb{E}[Z_n^k] \to \mathbb{E}[Z^k]$ as $n \to \infty$, this implies that $Z_n \to^{(d)} Z$.*

You can prove the central limit theorem with this kind of thing. You have a combinatoric proof of the Central Limit Theorem by taking the moments and showing they go to Gaussian. You can also prove that the Wigner semicircle theorem (distribution of eigenvalues of random symmetric Bernoulli matrix) with the method of moments.

A common trick is to not really rely on the usual notion of moments, but to rely on factorial moments.

Definition 2.10. A factorial moment $(X)_k = X(X-1)\cdots(X-k+1) = \frac{X!}{(X-k)!}$.

Here is the outline of how we prove the theorem stated before.

Lemma 2.11. *If $Z \sim Pois(c)$, $\mathbb{E}[Z]_k = \lambda^k$.*

We still have the general theorem $\mathbb{E}[Z_n]_k \to \mathbb{E}[Z]_k$ for all k implies $Z_n \to^{(d)} Z$.

Proof. This is an outline for the second theorem.

$$\mathbb{E}[Z_n]_k = \sum_{T_1,T_2,\cdots,T_k \text{ distinct and unordered triangles}} \mathbb{E}[\mathbf{1}(T_1 \in G)\cdots\mathbf{1}(T_k \in G)]$$

which is the probability that $\mathbb{P}\{T_1,\cdots,T_k \in G\}$. This equals

$$S_k + S_k'$$

where S_k is a contribution from the sum over disjoint triangles and S_k' is the rest. So $S_k = \binom{n}{3}\binom{n-3}{3}\binom{n-2*3}{3}\cdots\binom{n-3(k-1)}{3}$. You're also multiplying by p^3 for each of these contributions. Then for k fixed, this product is $\asymp (\frac{c^3}{6})^k$ since we take $p = \frac{c}{n}$.

We skip the part where we show that S_k' is negligible since it must be less than the order of this since there is overlap. □

Can we generalize this? Let us suppose we take some other graph H that looks like a "P". There are 6 edges and we write $\mathbb{E}[Z_n] \asymp \binom{n}{5}p^6 \asymp n^5p^6$ which is a constant if $p = (c/n)^{5/6}$. We claim that $n^{-5/6}$ is a threshold and perhaps conjecture in general that n^{-V_H/E_H} is a threshold.

What ends up being true is that if $p << n^{-5/6}$, $\mathbb{E}[Z_n] \to 0$ and $\mathbb{P}\{Z_n \geq 1\} \to 0$. Then consider $n^{-5/6} << p << n^{-4/5}$. By the same logic as before we see that $n^{-(4/5)}$ should be a threshold. Let H' be the square graph with one diagonal (subgraph of H). This situation implies that H' is not in G with high probability in the setting $p << n^{-5/6}$. If you take the densest subgraph of the graph and look at the ratio between vertices and edges, this will give you the ratio. The catch gives us that $H' \subset H$ but H' is denser.

Definition 2.12. $m(H) = \max\{\frac{e_{H'}}{v_{H'}} : H' \subseteq H, H' \neq \emptyset\}$ where e and v denote edges and vertices in H'.

Theorem 2.13. *Theorem of Finite Graph Containments. (Bollobas)*
For any $H \neq \{\emptyset\}$, $n^{-1/m(H)}$ is a coarse threshold for containing H. This was proved for H balanced (no denser subgraph).

Theorem 2.14. *If H is balanced, $p = cn^{-m(H)}$, then $Z_n \to^{(d)} Pois(\frac{c^{e_H}}{|Aut(H)|})$, where $Aut(H)$ is the automorphism group - counting the members is basically the number of ways to label the graph H.*

Example 2.15. $H = K_t$, then you get $c^k/k!$. If $H = C_k$, then you get $c^k/2k$.

3 Giant Components

Definition 3.1. A giant component is a connected component of size linear in the number of vertices.

Definition 3.2. $LCC(n,p)$ is the size of the largest connected comopnent in $G(n,p)$. $2-LCC(n,p)$ is the size of the second largest connected component.

Remark 3.3. Recall that an event E_n holds with high probability if $\mathbb{P}\{E_n\} \to 1$ as $n \to \infty$. This is also called asymptotically highly probable.

Theorem 3.4. *(Giant Component, Erdös-Rényi 1960)*

1. *If $c < 1$, then with high probability, $LCC(n, \frac{c}{n}) \leq \frac{2.1}{(1-c)^2} \log(n)$.*
2. *If $c > 1$, then with high probability, $LCC(n, p) \leq (1 + o(1))\beta n$ where β is a number between 0 and 1, the solution to $e^{-\beta c} = 1 - \beta$. You can also say that $2 - LCC(n, \frac{c}{n}) \leq \frac{\beta c}{(1-c)^2} \log(n)$.*
3. *If $c = 1$, you have a component with hybrid size.*

Proof. We will prove that there is no giant component with high probability if $c < 1$. Let $v \in [n]$. Define the following branching process: At each generation, the number of descendants is independently drawn from $\text{Bin}(n, \frac{c}{n})$ and also there is no collusion (parent nodes always have different descendants). This defines a Galton-Watson tree. Let C_v be the connected component of node v in $G(n, \frac{c}{n})$. We want to bound $\mathbb{P}\{|C_v| \geq k\} \leq \mathbb{P}\left\{\sum_{i=1}^{k} X_i \geq k - 1\right\}$ where X_i are i.i.d $\text{Bin}(n, \frac{c}{n})$. We can understand this inequality as saying the number of children of the leaves is more than the number of leaves, at a certain point in the branching process (we're also including the original leaves actually in the sum). We define $D_v = C_v - \{v\}$, where D_v is the descendent of node v. First we have

$$\begin{aligned} \mathbb{P}\{|C_v| \geq k\} &= \mathbb{P}\{|D_v| \geq k - 1\} \\ &\leq \mathbb{P}\left\{\sum_{i=1}^{k} X_i \geq k - 1\right\} \\ &= \mathbb{P}\left\{\text{Bin}(kn, \frac{c}{n}) \geq k - 1\right\} \end{aligned} \tag{1}$$

Then we want to apply concentration of the binomial distribution to get the first part of the theorem. Our goal is to show that for k logarithmic (maybe constant):

$$\mathbb{P}\left\{\text{Bin}(kn, \frac{c}{n}) \geq k - 1\right\} \to 0$$

The Chernoff bound gives, for $X \sim \text{Bin}(n, p)$ that

$$\mathbb{P}\{X \geq \mathbb{E}[X] + \delta\} \leq e^{-\frac{\delta^2}{2(np + \delta/2)}}$$

Then note that in the bound, $\mathbb{E}[X] = np = kc$. Adding and subtracting kc gives us

$$\mathbb{P}\{X \geq kc + (k - 1 - kc)\} \leq e^{-(1-c)^2 k/2} \asymp n^{-1.05}$$

by taking $k = \frac{2.1}{(1-c)^2} \log(n)$. Thus $\mathbb{P}\{\exists v \in [n] : |C_v| \geq k\} \leq n^{-1.05} \to 0$. □

Theorem 3.5. *(Connectivity, Erdös-Rényi 1960)*

1. *If $c \leq 1$, then with high probabability $G(n, \frac{c\log(n)}{n})$ is not connected. At $c = 1$, you have constant probability.*

2. *If $c > 1$, then with high probability, $G(n, \frac{c\log(n)}{n})$ is connected.*

In the connected case, the expected number of edges is $\binom{n}{2}\frac{c\log(n)}{n} = \frac{c(n-1)\log(n)}{2}$.

Proof. We first investigate node isolation. This is a property that gives you nodes that are isolated.

Lemma 3.6. 1. *If $c < 1$, $G(n, \frac{c\log(n)}{n})$ has isolated nodes with high probability.*

2. *If $c > 1$, $G(n, \frac{c\log(n)}{n})$ has no isolated nodes.*

Proof. Let $Z = \sum_{v\in[n]} \mathbf{1}(v \text{ is isolated in } G)$.

$$\begin{aligned}
\mathbb{E}[Z] &= n\mathbb{P}\left\{1 \text{ is isolated}\right\} \\
&= n(1-p)^{n-1} \\
&= n\left(1 - \frac{c\log(n)}{n}\right)^{n-1} \\
&\asymp ne^{-c\log(n)} = n^{1-c}
\end{aligned} \tag{2}$$

Thus $\mathbb{E}[Z] \to 0$ if $c > 1$. By a lemma from last week, $\mathbb{P}\left\{Z \geq 1\right\} \to 0$ if $c > 1$. This is how we see c is a threshold for the behavior of connectivity. The $p = \frac{c\log n}{n}$ regime is the boundary line, the behavior is obvious if very far away from this threshold. We can see that if $c > 1$, the number of isolated nodes goes to 0, and if $c < 1$, the number of isolated nodes blows up.

Now we have $Var(Z) = \sum_{u,v\in[n]} Cov(\mathbf{1}(u \text{ is isolated}), Cov(\mathbf{1}(v \text{ is isolated}))$. If $u = v$, then $Var(\mathbf{1}(u \text{ isolated})) = \mathbb{P}\left\{u \text{ isolated}\right\}(1 - \mathbb{P}\left\{u \text{ isolated}\right\}) \asymp n^{-c}(1-n^{-c})$. Then there are n terms like this. If $u \neq v$, then $Var(\mathbf{1}(u \text{ isolated}, v \text{ isolated})) = \mathbb{P}\left\{u \text{ isolated}, v \text{ isolated}\right\} - \mathbb{P}\left\{u \text{ isolated}\right\}^2 = (1-p)(1-p)^{2(n-2)} - (1-p)^{2(n-1)}$. Then note that $Var(Z) \sim \mathbb{E}[Z]$ by throwing out higher moment terms since they are much smaller and asymptotically do not matter. Applying a lemma from before, $\mathbb{P}\left\{Z = 0\right\} \leq \frac{Var(Z)}{\mathbb{E}[Z]^2} \sim \frac{1}{\mathbb{E}[Z]} \to 0$ if $c < 1$. If $c = 1$, $\mathbb{E}[Z] = \Theta(1)$.

□

To finish the proof, we need to show that for $c > 1$ the graph is indeed connected. What we know from the lemma is that there are no isolated nodes, however this does not imply full connectivity. We have to show that full connectivity happens. We want to show $\mathbb{P}\left\{\exists S \subseteq [n] \text{ s.t. } |S| \leq \frac{n}{2}, S \text{ disconnected from } S^c\right\} \to 0$. Note that in a given S_1, with k vertices, the probability that none of the edges appear between S_1 and S_1^c is $(1-p)^{k(n-k)}$.

$$\begin{aligned}
\mathbb{P}\left\{\exists S \subseteq [n] : |S| = k, S \text{ disconnected from } S^c\right\} &\leq \binom{n}{k}\mathbb{P}\left\{S_1 \text{ is disconnected from } S_1^c\right\} \text{ (by union bound)} \\
&\leq \binom{n}{k}(1-p)^{k(n-k)}
\end{aligned} \tag{3}$$

Therefore,

$$\mathbb{P}\left\{\exists S \subseteq [n] \text{ s.t. } |S| \leq \frac{n}{2}, S \text{ disconnected from } S^c\right\} \leq \sum_{k=1}^{n/2} \binom{n}{k}(1-p)^{k(n-k)}$$

Then $p = \frac{c\log(n)}{n}$. If $k = 1$, we have $n(1-p)^{n-1}$: This is exactly the isolated node case, which $\to 0$ if $c > 1$. For any constant, we have this behavior. If $k = n/2$, then this is large: this becomes $2^{n-o(1)}\left(1 - \frac{c\log(n)}{n}\right)^{\frac{n^2}{4}} \asymp 2^n e^{-n\log(n)c} \to 0$. Then you can formally split up the sum and use the constant bound to send one half to 0, and the other half can be shown by the linear size of k. Thus the whole sum goes to 0, and thus the desired probability goes to 0. $\square$

4 Fourier Boolean Analysis

The question we will ask in this section is as follows: Let $f : \mathbb{F}_2^n \to \mathbb{R}$, where we restrict to $[0, 1]$. We would like to know $\mathbb{P}_{X\sim_n \mathbb{F}_2^n}\{f(x) \neq f(x + e_i)\}$. This shows up when we study graph properties.

Definition 4.1. $\mathcal{L}^2(\mathbb{F}_2^n, \mu_p) = \{f : \mathbb{F}_2^n \to \mathbb{R} : \|f\|_2^2 < \infty\}$ where $\|f\|_2^2 = \sum f^2(x)\mu_p(x)$, where μ_p is the product measure: $\mu_p(x) = p^{\sum_{i=1}^n x_i}(1-p)^{1-\sum_{i=1}^n x_i}$. Then $\langle f, g\rangle = \sum_{x\in\mathbb{F}_2^n} f(x)g(x)\mu_p(x)$. Note that $\langle , \rangle$ is an inner product, and $\|\cdot\|$ is the induced norm.

We would like an orthonormal basis for this space. The first one is just the canonical basis, simply for every $s \in \mathbb{F}_2^n$ define Dirac $\delta_S(x) = 1$ if $x = s$ and 0 otherwise. This clearly forms a basis since you can express every function. A more interesting orthonormal basis is the Fourier-Walsh basis. For now, we will let $p = 1/2$. The Fourier-Walsh basis with this setting is given by $\chi_S : \mathbb{F}_2^n \to \{-1, +1\}$, $\chi_S(x) = (-1)^{x\cdot S}$, and $x \cdot S = \sum_{i=1}^n x_i S_i$. Then $\{\chi_S\}_{S\in\mathbb{F}_2^n}$ form an orthonormal basis.

Now we check that it is a basis. First we check $\langle \chi_S, \chi_T\rangle$. We want this to be 0 if $S \neq T, 1$ if $S = T$. We only have 2^n elements, which is the dimension of the space. If all of these functions are orthonormal, then none of them can be linearly dependent and we have a basis. Then we write

$$\begin{aligned}\langle \chi_S, \chi_T\rangle &= \sum_{x\in\mathbb{F}_2^n} \chi_S(x)\chi_T(x)2^{-n} \\ &= \sum_{x\in\mathbb{F}_2^n} (-1)^{X\cdot(S+T)}2^{-n} \\ &= \mathbf{1}(S = T)\end{aligned} \tag{4}$$

since if $S = T, S \oplus T = 0$ and the expression is 1. If $S \neq T$, $S \oplus T = U$. Then by symmetry, there are the same number of U which have an even number of ones and an odd number of ones when we consider all x. Thus, half of the sum is positive and half the sum is negative, and the whole sum is 0 since we are adding a constant 2^{-n}.

Recall that a Fourier basis for $\mathcal{L}^2(\mathbb{R}, \lambda)$ means $\tilde{X}_w \to e^{2\pi i w x}$, and $\langle \tilde{X}_w, f\rangle = \hat{f}_w$. Well, what similarities do these two bases have? We note that the values lie in between $[0, 1]$. The product of two elements $\tilde{X}_{w_1}\tilde{X}_{w_2} = \tilde{X}_{w_1+w_2}$, i.e. a homomorphism between the groups.

These $\tilde{X}_w$ are known as characters, and there is a whole theory about decompositions into characters.

Note since $\{\chi_S\}_{S\in\mathbb{F}_2^n}$ is a basis, we define $\hat{f}(S) = \langle f, \chi_S\rangle$ then $f = \sum_{S\in\mathbb{F}_2^n} \hat{f}(s)\chi_S$.

Lemma 4.2. *Familiar Fourier Transform Properties.*

1. $\langle f, g\rangle = 2^n\langle \hat{f}, \hat{g}\rangle = \sum \hat{f}(s)\hat{g}(s)$. *Note this is analgous to Parseval's Lemma.*
2. $\hat{f * g} = \hat{f}\hat{g}$ *where* $f * g(y) = \sum_{y\in\mathbb{F}_2^n} f(x+y)g(y)2^{-n}$, *where we can write* $x+y$ *since we are in* $\mathbb{F}_2^n$ *instead of* $x-y$.

Proof. When you have the structure of the characters, these properties should hold.

First we explain the Parseval-esque equality.

$$\begin{aligned} 2^n\langle \hat{f}, \hat{g}\rangle &= \sum_s \hat{f}(s)\hat{g}(s) \\ &= \sum_s \sum_x f(x)(-1)^{xs} \sum_y g(y)(-1)^{xy}2^{-n}2^{-n} \\ &= \sum_s \sum_{x,y} f(x)g(y)(-1)^{(x+y)s}2^{-2n} \\ &= \sum_x f(x)g(x)2^{-n} = \langle f, g\rangle \end{aligned} \tag{5}$$

where if $x \neq y$, the term is 0 by orthonormality, so we only get the terms where $y = x$ and we sum over s to eliminate the other 2^{-n}.

2. $\hat{f}(0) = \mathbb{P}\{f(x) = 1\}, \hat{f}(e_i) = \frac{1}{2} - \mathbb{P}\{f(x) = x_i\}$. Now we want to explain the meaning of the Fourier coefficient. We can think of $s \in \mathbb{F}_2^n$ or as a set $s \subseteq [n]$ (the set of indices for which $s_i = 1$). Note $\hat{f}(0) = \sum_{x\in\mathbb{F}_2^n}(-1)^0 f(x)2^{-n} = \sum_x f(x)2^{-n}$. From now on we will consider $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Then $\sum_x f(x)2^{-n} = \mathbb{P}\{x \sim \mathbb{F}_2^n\} f(x) = 1$. We also have $\hat{f}(e_i) = \sum_x (-1)^{e_i\cdot x} f(x)2^{-n} = \sum_x f(x)(-1)^{x_i}2^{-n}$. $x_i = 0$ or $x_i = 1$:

$$\begin{aligned} \sum_x f(x)(-1)^{x_i}2^{-n} &= \sum_{x\in\mathbb{F}_2^n, x_i=0, f(x)=1} f(x)2^{-n} - \sum_{x\in\mathbb{F}_2^n, x_i=1, f(x)=1} f(x)2^{-n} \\ &= \mathbb{P}\{f(x) = 1, x_i = 0\} - \mathbb{P}\{f(x) = 1, x_i = 1\} \\ &= \mathbb{P}\{x_i = 0\} - \mathbb{P}\{f(x) = 0, x_i = 0\} - \mathbb{P}\{f(x) = 1, x_i = 1\} \\ &= \frac{1}{2} - \mathbb{P}\{f(x) = x_i\} \end{aligned} \tag{6}$$

If we have $f(x) = x_1$ (a dictatorship), then only $\hat{f}(e_1) = -\frac{1}{2}$, all other values are $\frac{1}{2}$. The output is fully correlated with x_1. Then we can write $\hat{f}(x) = \frac{1}{2} - \mathbb{P}\{f(x) = \oplus_{i\in s} x_i\}$. It is telling you how much the output of the function correlates with the variables making up the vector x, the x_i.

□

The Fourier coefficients reflect the influence of the variables x_i on the output of $f(x)$.

Definition 4.3. $f : \mathbb{F}_2^n \to \mathbb{R}$, let $I_i(f) = \mathbb{P}_{x \sim_n \mathbb{F}_2^n} \{f(x) \neq f(x \oplus e_i)\}$, and $I(f) = \sum_{i \in [n]} I_i(f)$.

Lemma 4.4. *Influence lemma.*

1. $I_i(f) = 4 \sum_{i \in s} \hat{f}(s)^2$*: You sum over all* i *in* s*, then* $s_i = 1$*.*
2. $I(f) = 4 \sum_{x \in \mathbb{F}_2^n} |s| \hat{f}(s)$*. This follows from the first statement directly.*

Proof. We prove the first statement in the lemma. Let $f_i(x) = f(x + e_i)$. When you do Fourier, you either have a convolution and do products, or you see a norm and use Parseval.

$$\begin{aligned} I_i(f) &= \mathbb{P}\{f(x) \neq f(x + e_i)\} \\ &= \sum_{x \in \mathbb{F}_2^n} (f(x) - f(x + e_i))^2 \, 2^{-n} \\ &= \|f - f_i\|_2^2 \\ &= 2^n \|\hat{f} - \hat{f}_i\|_2^2 \\ &= \sum \hat{f}(s) \hat{f}_i(s) \\ &= \sum_x f(x + e_i)(-1)^{x \cdot s} 2^{-n} \\ &= \sum_y f(y)(-1)^{y \cdot s} 2^{-n} (-1)^{s_i} = \hat{f}(s)(-1)^{s_i} \end{aligned} \tag{7}$$

and thus $\hat{f} - \hat{f}_i = 0$ if $s_i = 0$ and $2\hat{f}(s)$ if $s_i = 1$. Therefore

$$\begin{aligned} I_i(f) &= 2^n \|\hat{f} - \hat{f}_i\|_2^2 \\ &= 2^n * 4\hat{f}^2(s) \\ &= 4 \sum_{s_i = 1} \hat{f}^2(s) \\ &= 4 \sum_{i \in s} \hat{f}^2(s) \end{aligned} \tag{8}$$

□

Now we generalize to $p \neq \frac{1}{2}$.

Definition 4.5. $\chi_i(x) = \sqrt{\frac{p}{1-p}}$ if $x_i = 0$ and $-\sqrt{\frac{1-p}{p}}$ if $x_i = 1$. Then $\chi_s(x) = \prod_{i \in s} \chi_i(x)$.

Theorem 4.6. $\{\chi_s\}_s$ *is an orthonormal basis for* μ_p*.*

Now, what about the influence of the coefficients?

Lemma 4.7. $I_i(f) = \frac{1}{p(1-p)} \sum_{i \in s} \hat{f}^2(s)$ *and* $I(f) = \frac{1}{p(1-p)} \sum_s |s| \hat{f}^2(s)$.

In some areas of computer science and theory this is well known, but it is not used in other areas, and it seems like an important tool. Now how does this tool connect to random graphs, and in fact to coding theory?

Theorem 4.8. *(Russo, Margulis).*
If A is an increasing graph property and $\frac{\partial}{\partial p}\mu_p(A) = I(\mathbf{1}_A)$.

Proof. Fix $i \in [n]$. Then $x_i = \{(x_1^{i-1}, x_{i+1}^n) \in \mathbb{F}_2^{n-1} : f(x_1^{i-1}, 0, x_{i+1}^n) = 1, f(x_1^{i-1}, 1, x_{i+1}^n) = 1\}$, and y_i is defined analagously. Define $p = (p_1, \cdots, p_n)$. Then $\mu_p(A) = \mu_q(x_i) + p_i\mu_i(y_i)$ where $q = (p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_n)$ which works since A is a monotone property. The probability you have the property A is the sum of the cases where you did not activate the i^{th} component and the probability that you did activate the i^{th} component, and this allows us to separate things into independent components. The first term does not depend on p_i so the derivative with respect to p_i is 0. Thus $\frac{\partial}{\partial p_i}\mu_p(A) = 0 + I_i(\mathbf{1}_A)$, and then the chain rule means that you have to sum over these guys and you get the result. □

Now, why must coarse thresholds must be for local properties? If you are coarse, it means that your jump is moderate. If your jump is moderate, it means that $\frac{\partial}{\partial p}\mu_p|_{p(1/2)} < \frac{1}{c}$. All we are saying is that the window $\frac{\delta(1/2)}{p(1/2)} > c$. As long as we know that $1 - p(1/2)$ is going to 1, we can put $1 - p(1/2)$ for free outside, since that is just going to 1. Then we see that $(1 - p(1/2))p(1/2)\frac{\delta(1/2)}{p(1/2)} = I(\mathbf{1}_A) < 1/c$, but this means $\sum_s \hat{f}^2(s)|s| < \frac{1}{c}$ from the theorem from before. Therefore, $\sum_{s:|s|\geq L} \hat{f}(s) \leq \frac{1}{cL}$. We have $\frac{\delta}{p} > c$ which implies $I(\mathbf{1}_A) = I(f) < c'$ and thus $\sum_{|s|<L} \hat{f}^2(c) < c''$, and f concentrates on small graphs. This shows that the Fourier energy must concentrate on small sets, and will show you how to get the kinds of results we have gotten.

5 Hypergraphs, CSPs, planted CSPs, noisy CSPs

Definition 5.1. A hypergraph is a generalization of a graph where "edges" can cover ≥ 2 vertices called hyperedges (often, "hyper" is removed). The set of all possible hypergraphs of order k on n vertices is denoted $\binom{[n]}{k}$: $G = (V, E)$ is a hypergraph, and $E \subseteq \binom{[n]}{k}$. Often, hypergraphs are represented in a factor or bipartite form (put all block edges on one side, and all other edges on the other side).

5.1 Random Hypergraphs

An analog to the Erdös-Renyi model $G_k(n, p)$, and you draw an edge with probability p. The topology of $G_k(n, p)$ is roughly similar to $G_2(n, p)$.

Example 5.2. Connectivity.
Connectivity happens at $\alpha_c = 1/k$ if $p = \frac{\alpha \log(n) n}{\binom{n}{k}}$. Roughly speaking, $2\alpha = c$ where c is the variable used in the previous description of $G_2(n, k)$.

Remark 5.3. Usually, $p = \frac{c}{n}$ or $p = \frac{c\log(n)}{n}$ for Erdos-Renyi with $k = 2$. But, $p = \frac{\alpha n}{\binom{n}{k}}$ or $p = \frac{\alpha n \log(n)}{\binom{n}{k}}$ for $G_k(n, p)$. Note that the expected number of edges in $G(n, \frac{\alpha n}{\binom{n}{k}})$ is $\binom{n}{k}\frac{\alpha n}{\binom{n}{k}} = \alpha n$, so α is edge-density. However, $\binom{n}{2}\frac{c}{n} = \frac{cn}{2}$.

Example 5.4. Giant component.
The back of the envelope calculation requires you to take a node, and count the expected number of neighbors it has (neighbors are vertices). You have $\binom{n-1}{k-1}$ possible edges, pick each edge with probability p. Each time you pick an edge, you get $k-1$ neighbors. Then there are $(k-1)\binom{n-1}{k-1} * p$, if $p = \frac{\alpha n}{\binom{n}{k}} \sim \frac{\alpha n}{n^k/k!}$. So $(k-1)\binom{n-1}{k-1} * p \sim (k-1)k\alpha_c n$, so $\alpha_c = \frac{1}{k(k-1)}$.

5.2 Constraint-Satisfaction Problems (CSP)s

Definition 5.5. A Boolean CSP is defined by

1. A k-hypergraph $G = (V, E)$
2. $\forall e \in E, \phi_e : \mathbb{F}_2^k \to \mathbb{F}_2$

The decision problem is: Does there exist an assignment to $x_1, \cdots, x_n$ where $\phi_e(x[e]) = 1$ for all $e \in E$ where $x[e] = (x_{i_1}, \cdots, x_{i_k})$ when $e = (i_1, \cdots, i_k)$.

Example 5.6. k-SAT.
A clause must have $\phi_e(x[e]) = \mathbf{1}(x[e] \neq s_e)$ for all e, let $s_e \in \mathbb{F}_2^k$. So there is some forbidden pattern which cannot exist. An equivalent form of k-SAT is defined by a Boolean formula (x_1 OR $\bar{x}_2$ OR x_7) AND ($\bar{x}_1$ OR $\bar{x}_6$ OR $\bar{x}_1 1$) AND .... Here, we want to avoid a false output for any of the clauses: For instance, the first clause $(1, 2, 7)$. For this clause, $s_e = (0, 1, 0)$.

Remark 5.7. 2-SAT is in P, but $k \geq 3$, k-SAT is NP-complete (one of Cook's 21 NP-complete problems).

Example 5.8. k-XORSAT. $\phi_e(x[e]) = \mathbf{1}(\oplus_{i \in e} x_i = s_e)$, $s_e \in \{0, 1\}$ is always in P for all k.

Now complexity theory is a worst-case scenario: There are some cases which are very hard, but this is not the case for all versions of the problem! What about average case analysis? We may want to know how easy it is to solve the average case.

5.3 Random CSPs

Definition 5.9. A random CSP in the "uniform" or "binomial" model has $G \sim G_k(n, p)$ and $\phi_e \sim \mu$ on $\mathbb{F}_2^{\mathbb{F}_2^k}$ (power set), all functions which can go from $\mathbb{F}_2^k \to \mathbb{F}_2$.

Claim 5.10. *The SAT conjecture (in the context of a random model).*
For all $k \geq 2, \exists \alpha_k \in \mathbb{R}_+$ such that if $p = \frac{\alpha n}{\binom{n}{k}}$, and $\mathbb{F}_k(n, p)$ is a random k-SAT formula, then for all $\epsilon > 0$, $\mathbb{P}\{\mathbb{F}_k(n, \alpha_k - \epsilon) \text{ is SAT}\} \to 1$ and $\mathbb{P}\{\mathbb{F}_k(n, \alpha_k + \epsilon) \text{ is SAT}\} \to 0$ as $n \to \infty$.

The status of this problem is that for $k = 2$, it is proved. For all $k \geq 2$, it was proved by Friedgut that there is a sequence of α_k $\alpha_k(n)$ which moves with n such that it limits to a step function (so sharp threshold) up to scaling. We do not know if it converges to a single α_k. We do have that $\alpha_k(n) = \Theta(1)$. There is strong evidence that this should converge to a specific number. We have lots of bounds, upper bounds and lower bounds for this α_k. From Berkeley last year, it was proved there is some K such that if $k \geq K$, then $\alpha_k(n)$ does converge. However, $k = 3$ is still open.

The idea of Friedgut's result is to show that being satisfiable is not a local property. First, begin UNSAT is an increasing property (the more you add edges and constraints, the more likely you are to be UNSAT). Now it has a threshold that's coarse or sharp. Second, $F \sim F_k(n, \alpha)$ and choose α such that $\mathbb{P}\{F\text{UNSAT}\} = \frac{1}{2}$ and show $\exists \epsilon > 0\, \mathbb{P}\{F\text{UNSAT}|H \subset F\} \leq \frac{1}{2} + \epsilon$. If you are nonlocal you must have a sharp threshold. Then $\mathbb{P}\{F \text{ UNSAT}|H \subset F\} \leq \mathbb{P}\{F \text{ UNSAT}|x_1 \cdots x_r = 1\}$, where H is a finite formula with r variables. Now F^* is a random formula with r variables frozen to 1. Then the expected number of size k clauses (edges) is $\asymp n$, and the expected number of size $k-1$ clauses (edges) is $\asymp 1$. If you go to $k-2$ and more, it will be vanishing to 0. If you look at a random hypergraph on n variables and r vertices, each time you select a hyperedge with one of your guys, instead of being a 3 edge, it becomes a two edge. The computation you have here is basically $n^{k-1}\frac{\alpha n}{n^k}$. Now you you basically only have edges of order k and $k-1$. The last thing LAST $\leq \mathbb{P}\{F^{**} \text{ unSAT}\}$ means $\binom{n-r}{k}2^k$ possible k-clauses, and you pick them with probability $\frac{\alpha n}{\binom{n}{k}2^k}$ ($1/2^k$ is probability of s_e). We would like to make this look as though we had worked with $n-r$ variables. So we change it to $\frac{\alpha(n-r)}{\binom{n}{k}2^k} + \epsilon$ (it costs us an epsilon). Then we add a constant number of $(k-1)$-clauses to make the problem harder, and we show that if we replace the constant number of $(k-1)$ clauses with a logarithmic number in n of clauses, it doesn't change the difficulty.

The main takeaway from this proof is as follows: If one day you expect a problem where you sharp threshold, you should follow this kind of path. First is it monotone, then you know there must be a threshold. You would like to say it's sharp and not coarse. If the property seems to be local, then you can apply this blackbox theorem and say it must in fact have a coarse threshold. If it is sharp, then you have to follow a path like this. What is the probability it has the property I'm interested in? Now we have to show this won't blow up to one. By conditioning on a fixed subgraph, you restrict the problem in a way such that you hope to see it does not change to much. Then having controlled the extent to which it blows up in this restricted case, you go back to the unrestricted case and show that if you let it be unrestricted, it still won't blow up to 1.

Basically, this is a branch of mathematics where you have an alternative to standard central limit type arguments. Another option is martingale type arguments.

6 Spectral Graph Theory

Hi, I'm Adam Marcus in Fine 1110.

6.1 Graph Laplacian

Definition 6.1. Graph Laplacian matrix.
$A = \sum_{(i,j)\in E}(\delta_i - \delta_j)(\delta_j - \delta_i)^T$, where δ_i is the vector with 1 in the i^{th} coordinate and 0 everywhere else. We can also write it as $L_G = D_G - A_G$ where A_G is the adjacency matrix of the graph and D_G is a diagonal matrix where $D_{i,i} = \deg(v_i)$, vertix i.

L_G is always symmetric and positive semidefinite, since it is a sum of positive semidefinite things. It also has eigenvalue 0 with corresponding eigenvector $[1, 1, 1, \cdots, 1]^T$. Some people like to use the adjacency matrix, but one of the benefits of the Graph Laplacian is that it is PSD. If H is a subgraph of G, then $L_H \leq L_G$, in other words, $L_G - L_H$ is PSD.

$L_G = \sum_{e \in E(G)} L_e$ as well. We can also write $L_G = \sum_i \lambda_i v_i v_i^T$ where v_i are orthonormal and $\lambda_i \geq 0$ (the eigenvalues) and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. $\lambda_2 = 0$ iff G is not connected. Thus λ_2 can be thought of as a quantitative version of connectivity: the conductance of the graph and the worst way of splitting things up so that there is a bottleneck between the two places (we will see these results shortly). We can also write $\lambda_n \leq 2\Delta(G)$, where $\Delta(G)$ is the maximum degree of G.

Let us say we start with a graph G. We can do $G \to L_G$, and then adding an edge $\hat{e}$ gives $L_G \to L_G + L_{\hat{e}}$.

Lemma 6.2. *Matrix Determinant Lemma.*

$$\begin{aligned} \det\left(xI - L_G - L_e\right) &= \det\left(xI - L_G - ee^T\right) \\ &= \det\left(xI - L_G\right)\left(1 - e^T(xI - L_G)^{-1}e\right) \\ &= p(x)\left(1 - \sum_i \frac{\langle e, v_i\rangle^2}{x - \lambda_i}\right) \end{aligned} \tag{9}$$

where $p(x)$ is the original characteristic polynomial.

Now we might wonder what happens when we add an average edge, as we do in probabilistic combinatorics. We want to know how big it becomes. If we add something to add to the average, then we could assert there is something better than the average, but it turns out that the contribution doesn't do anything. $\frac{1}{\binom{n}{2}} \sum_e L_e$, which has $2/n$ on the diagonal and $-1/\binom{n}{2}$ elsewhere. Then $L_G + \frac{1}{\binom{n}{2}} \sum_e L_e = L_{G'}$. But this is no good as a bound. In higher dimensions, you can add a vectors that should both increase the trace by 1, but you change the eigenvectors differently.

6.2 Polynomial interlacing

Let's try something different. Let us define $q(x) = p(x)\left(1 - \sum_i \frac{\langle e, v_i\rangle^2}{x - \lambda_i}\right)$ and assume $p(x)$ has n distinct roots λ_i. Now either a root of $q(x)$ is a root of $p(x)$, or the other thing is zero. Since $x = \lambda_i$ cancels out due to the $x - \lambda_i$ term in the denominator. So roots come from the other term. Then if we add an edge, the average root of the second term must go up by some positive value, by the formula we derived before: We will have that the the averages μ_i must have $\lambda_i < \mu_i < \lambda_{i+1}$, and $\lambda_n < \mu_n$. This property is called **interlacing**. Our question focuses on asking what edges we can add. Now, instead of averaging over random matrices, we are going to try averaging over random polynomials and seeing what happens.

$$\begin{aligned} \frac{1}{\binom{n}{2}} \sum_e \det\left(xI - L_G - ee^T\right) &= \det(xI - L_G)\left(1 - \frac{1}{\binom{n}{2}} \sum_e \sum_i \frac{\langle e, v_i\rangle^2}{x - \lambda_i}\right) \\ &= \tilde{q}(x) \end{aligned} \tag{10}$$

But this $\tilde{q}(x)$ may not make sense, since things could be real or complex, and we don't know how to compare complex numbers properly. However, due to the interlacing property, it turns out this makes sense.

Theorem 6.3. *Let $p_1, \cdots, p_n$ be real rooted polynomials, and let q be a polynomial that interlaces all of the p_i.*

1. *$\hat{p} = \sum_i \alpha_i p_i$, $\alpha_i \geq 0, \sum_i \alpha_i = 1$ has all real roots.*
2. *For all k, $\exists i, j$ such that $\lambda_k(p_i) \leq \lambda_k(\hat{p}) \leq \lambda_k(p_j)$.*

Proof. Let us zoom in on the interval between λ_t and λ_{t+1}. WLOG all $p_i(x) \leq 0$ at t. This implies that all $p_i(x) \geq 0$. Thus at t, $\hat{p}(x) \leq 0$, and at $t+1$, $\hat{p}(x) \geq 0$. Thus somewhere inbetween, $\hat{p}(x) = 0$. $\hat{p}$ has n real roots since this happens n times. Now if we have a bunch of numbers that up to 0, one must be less than or equal to zero and one must be greater than or equal to zero. Then $\det(\lambda I - L_G) = p(x) = x\tilde{p}(x)$. Then, $\frac{1}{\binom{n}{2}} \sum_e \det(\lambda I - L_G - ee^T) = q(x) = x\tilde{q}(x)$. Then $\tilde{q}(x) = \tilde{p}(x) - \frac{2}{n}\tilde{p}'(x)$. There exists some edge e such that $\lambda_2(G + ee^T) \geq \lambda_2(\tilde{p}(x) - \frac{2}{n}\tilde{p}'(x))$. □

A big open question is how to find the correct edge to add.

Theorem 6.4. *Let G have characteristic polynomial $xp(x)$ and H have characteristic polynomial $xq(x)$. Then,*

$$\frac{1}{n!} \sum_{\text{perms } \pi} \det\left(\lambda I - L_G - \pi^T L_H \pi\right) = x\left(\sum_i p^{(i)}(x/2) q^{(n-i-1)}(x/2)\right) \tag{11}$$

We can assert that the polynomial is real rooted.

Now rather than looking for a common interlacer polynomial of the whole thing, let us look for interlacers of subsets, and find a convex combination, and then check if the results have common interlacers, and so on. This is called an **interlacing family**. Then you can go all the way down. This is a way of nicely weakening the idea.

How do we build an interlacing family? Before, we were simply adding some rank 1 matrix (ee^T) to our matrix. Let us add a bunch of rank 1 matrices. We can define a choice vector $\sigma \in [n]^m$, where σ_i is the index of a vector. If you choose these vectors independently, you get an interlacing family.

6.2.1 An Application to Expander Graphs

If G is d-regular, then $D_G = dI$ and $L_G = D_G - A_G$, $0 \leq \lambda_i(L_G) \leq 2d$, $-d \leq \lambda(A_G) \leq d$. So we know that d is an eigenvalue of A_G, and if G is bipartite, then $-d$ is an eigenvalue. If you are looking for a well-expanding bipartite graph, for doing expander codes, then you are essentially forced to have eigenvalues at $d, -d$. So what happens to λ_3 and up? G is a good expander spectrally if you can get the eigenvalues clustered around 0. There is an interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, called the Ramanujan interval. If all eigenvalues are in this interval, then this is a **Ramanujan graph** and an infinite collection is called a Ramanujan family.

Theorem 6.5. *(Alon, Boppana 1996).*
No smaller interval can contain all nontrivial eigenvalues of an infinite collection of d-regular graphs.

The "ultimate" d-regular expander is the d-regular infinite tree, and all eigenvalues of the d-regular infinite tree lies in the Ramanujan interval, and it turns out that this interval is related to the d-regular infinite tree.

Theorem 6.6. *Margulis, Lubotsky-Phillips-Sarnak* (1988).
Ramanujan families exist for $d = p + 1$, where p is a prime number.

This proof used the work of at least 4 Fields Medals. It was extended by Morganstern to $d = p^k + 1$. All these constructions are algebraic. You look at the Cayley graph of the group and use all these high-powered results to show that because of all the symmetries involved, you can't have big eigenvalues. On the other hand, almost everything is almost Ramanujan.

Theorem 6.7. *(Friedman* (2008)*).*
A randomly chosen d-regular graph has its non-trivial eigenvalues in the interval

$$[-2\sqrt{d-1} - \epsilon, 2\sqrt{d-1} + \epsilon] \tag{12}$$

You can use the characteristic polynomial result to show that for all bipartite graphs that the spectrum of eigenvalues is within the Ramanujan interval. We use bipartite since in bipartite graphs, the eigenvalues come in pairs: If λ is an eigenvalue, so is $-\lambda$ of the adjacency matrix.

Therefore,

Theorem 6.8. *There exist bipartite Ramanujan families of degree d for any d.*

Proof. (sketch). Set $G_0 = K_{d,d}$, which is bipartite, d-regular, and Ramanujan for any d. Given G_i, form G_{i+1} using the lifting technique. By the theorem previously stated, this is an interlacing familiy. Using some other results, you can show that G_{i+1} is also Ramanujan (lifts and others), and then proceed by induction to show you have a Ramanujan family. □

Remark 6.9. These results are used in sparsification, traveling salesman, etc. If you could find a way to efficiently find an edge e which would work, you would be able to improve a lot of current computer science algorithms. Currently the only way to find the edge is to brute force search 2^{dn} of them. We only guarantee it works. In fact, even finding one coefficient of one of these polynomials is NP-complete. You would probably have to find some other insight into what's happening. Adding the edge e is building the graph in the direction that the graph is most connected.

Remark 6.10. If you look at all the ways you can add a perfect matching to the bipartite graph, you can find a way to make it an interlacing family so that you can assert there is always some way to do it. Friedman proved his result by adding random matchings to his random graph. This exact method has a way of always working to get you a Ramanujan graph. We don't know how to actually do it without randomness unfortunately.

As far as coding theory goes, there is a situation in graphs where you know what the best thing you could possibly do is: The infinite d-regular tree. But you can't use infinitely many things. You have to use finite resources to get the best thing you could do. This is in some sense the same problem as a code: You are in a finite space for a code, since you cannot space things out exactly the way you want, so you take some penalty. It seems that interlacing polynomials could be interesting to show the existence of new codes.

6.3 Connectivity as a Minimization Problem

Let us start with a toy problem: You are given a graph G and you want to cluster it into two parts. One idea might be to project G onto a line, and then maybe you will split up the components. Of course, this projection might not respect anything that the graph is doing. But maybe there is some way we can ensure that vertices which are in the same cluster end up close together. So the goal is if $x \sim y$ (adjacent) in G, then $|f(x) - f(y)|$ is small. Well in particular, maybe you could try min $\sum_{(x,y)\in E(G)}(f(x) - f(y))^2$. There are a couple problems with this right from the start. First, this minimum is clearly 0, and this can happen whenever $f(x)$ and $f(y)$ are constant functions and equal. Well, how about we modify it so that we minimize over functions which are orthogonal to constant vectors: $f : \langle f, 1\rangle$. Well the infimium will still be zero, since we aren't normalizing with respect to the size of f. We can just add $\|f\|^2 = 1$. Thus our problem is

$$\min_{f:\langle f,1\rangle=0,\|f\|^2=1} \sum_{(x,y)\in E(G)} (f(x) - f(y))^2 \tag{13}$$

So we recognize this as a quadratic form, which we can now write as $g(x) = x^T Ax$ where A is a real symmetric matrix. Let's figure out what the matrix A should be. Consider the case where our graph is a single edge between two vertices. Then $[f(0), f(1)] \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right] [f(0), f(1)]^T = (f(0) - f(1))^2 = af(0)^2 + 2bf(0)f(1) + cf(1)^2$, so we want our matrix to be $A = \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$.

We can write our previous problem as

$$C(G) = \min_{x:\langle x,1\rangle=0} \frac{x^T L_G x}{x^T x} \tag{14}$$

This expression in the minimization is called the Rayleigh quotient.

Theorem 6.11. *$C(G) \neq 0$ iff G is connected.*

Proof. If G is disconnected, then there are two subgraphs G_1, G_2 not connected. Let us say $f(x) = t$ on G_1, $f(x) = s$ on G_2. Then $C(G)$ is 0 after we figure out how to be orthogonal to the all 1s vector, and it will be fine.

If G is connected, then x and y have a path between them, and we want $f(x) \neq f(y)$. Then there is some edge for which $f(z) \neq f(w)$, and if we subtract the two and square it, it will not be zero. □

Thus this minimization problem is able to encode "being connected". It assigns values to how connected the graph is: It is no longer just a qualitative statement. Depending on how you define mixing time, in fact you can define mixing time with respect to $C(G)$ directly.

6.4 Courant-Fisher and Algebraic Connectivity

For A a matrix, vector v non-zero which satisfies $Av = \lambda v$ for some $\lambda \in \mathbb{C}$ is an eigenvector and λ is an eigenvalue. Then $(\lambda I - A)\, v = 0$, thus $\det(\lambda I - A) = 0$. Then setting the characteristic polynomial equal to 0 recovers the eigenvalues. Then to find the eigenvectors, you look at the $\mathrm{Ker}(\lambda I - A)$ and check algebraic multiplicity and geometric multiplicity, and so on. However in the case of real symmetric matrix, we have the spectral theorem.

Theorem 6.12. *Spectral Theorem.*
For any real symmetric matrix A,

1. *All of the eigenvalues are real.*
2. *There exists $v_1, \cdots, v_n$ eigenvectors which form an orthonormal basis.*

Letting $V = [v_1, \cdots, v_n]$, then $VV^T = I$ and $V^TAV = \begin{bmatrix} \lambda_1 & 0 & \cdots \\ 0 & \ddots & \vdots \\ \vdots & \cdots & \lambda_n \end{bmatrix}$, so V diagonalizes A, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Thus we can write

$$A = \sum_i \lambda_i v_i v_i^T \tag{15}$$

Now we can write $x^T L_G x = \sum_i \lambda_i \langle x, v_i \rangle^2$. Since $\lambda_1 = 0$, we can see that

$$\min_{x:\langle x,1\rangle=0} \frac{x^T L_G x}{x^T x} = \lambda_2 \tag{16}$$

We can generalize this to the following theorem:

Theorem 6.13. *Courant-Fisher.*
A has eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and associated eigenvectors $v_1, \cdots, v_n$. Then,

$$\min_{x \perp \{v_1, \cdots, v_k\}} \frac{x^T A x}{x^T x} = \lambda_{k+1} \tag{17}$$

and its argmin is an eigenvector.

Definition 6.14. $\lambda_2(G)$ is called the algebraic connectivity.

Restating our theorem from before, we get that G is connected iff $\lambda_2(L_G) > 0$.

Example 6.15. K_n the complete graph on n vertices. The eigenvalues of K_n are 0 with multiplicity 1 and n with multiplicity $n-1$: Thus all non-trivial eigenvalues are very far away from 0.

Example 6.16. S_n the star on n vertices has eigenvalues 0 with multiplicity 1, 1 with multiplicity $n-2$, n with multiplicity 1.

Example 6.17. The hypercube in dimension H_n has eigenvalue 0 with multiplicity 1, and $2k$ with multiplicity $\binom{n}{k}$.

Example 6.18. Consider the path graph. Its eigenvectors are kind of like a Fourier basis, if you plot the eigenvectors (each of which have n components, $n \to \infty$).

Example 6.19. The cycle graph looks similar to the path graph eigenvalues and eigenvectors, except with more discrete jumps.

Let us see what happens when we add an edge. Suppose G has characteristic polynomial $p_G(x) = \det(xI - L_G)$. G', which has the added edge, has $p_{G'}(x) = \det(xI - L_G - w^T)$. Then we write

$$\begin{aligned} \det(xI - L_G)\det(I - w^T(xI - L_G)^+) &= p(x)(1 - v^T(xI - L_G)^+ v) \\ &= p(x)\left(1 - \sum_j \frac{\langle v, u_j \rangle}{x - \lambda_j}\right) \end{aligned} \tag{18}$$

where (u_j, λ_j) are eigenvector and eigenvalue pairs of G.

6.5 Isoperimetric Constraint

Definition 6.20. Isoperimetric constant.
Given a graph G, set vertices S, then

$$\delta S = \{(u, v) \in E : u \in S, v \notin S\} \tag{19}$$

This essentially bounds S. The isoperimetric constant of S is

$$\Theta(S) = \frac{|\delta S|}{\min |S|, |\bar{S}|} \tag{20}$$

and

$$\Theta_G = \min_S \Theta(S) \tag{21}$$

We claim that

Theorem 6.21. $\Theta(S) \geq \left(1 - \frac{|S|}{|V|}\right) \lambda_2$

Proof. We have $\lambda_2 = \min_{x:\langle x,1\rangle=0} \frac{x^T L_G x}{x^T x}$. We want to find a particular x such that $x^T L_G x \approx |\delta S|$. Let us define x such that $x(v) = 1 - t$ if $v \in S$, and $-t$ otherwise. Then, choosing $t = \frac{|S|}{n}$,

$$\begin{aligned} x^T x &= \sum_{v \in S} (1-t)^2 + \sum_{v \notin S} t^2 \\ &= |S| \left(1 - 2t + t^2\right) + (n - |S|)t^2 \\ &= |S| \left(1 - 2t + t^2\right) + |S|t - |S|t^2 = |S|(1-t) \end{aligned} \tag{22}$$

Then

$$x^T L_G x = \sum_{(u,v) \in \delta S} \left((1-t) + t\right)^2 = |\delta S| \tag{23}$$

Thus $\frac{x^T L_g x}{x^T x} = \frac{|\delta S|}{|S|(1-t)} \geq \lambda_2$, and thus

$$\Theta(S) \geq \lambda_2 \left(1 - \frac{|S|}{n}\right) \tag{24}$$

which implies $\Theta_G \geq \frac{\lambda_2}{2}$. □

Now recall that in H_n, $\lambda_2 = 2$. Thus, this proof implies that for any $S \subseteq H_n$, $|\delta S| \geq |S|$. You can do this combinatorially, but it gets messy and not as nice. Also, this inequality is sharp: The number of edges coming out of one face of the hypercube is exactly the number of vertices that are in it.

6.6 Cheeger's Inequality

Let ρ_G be a normalized version of Θ_G ("conductance"), and ν_2 be a normalized version of λ_2 (weight your Laplacian by dividing by the average degree).

Theorem 6.22. *Cheeger's Inequality.*

$$\frac{\eta_2}{2} \leq \rho_G \leq \sqrt{2\eta_2} \tag{25}$$

Thus, when η_2 is constant size, ρ_G is "trapped" as a function of η_2. This says the conductance and the minimum eigenvalue are basically the same. This η_2 is constant when there is a constant degree over the graph.

This is why having expanders are so important: You want your graphs to be very connective, but on the other hand, you need to have a constant degree. An expander is a constant degree graph which is as close to having properties of the complete graph as you possibly can.

7 Stochastic Block Model

7.1 Recovery

This is analagous to giving better bounds in the connectivity setting. The first theorem we proved is that for the setting where $p = a\frac{\log n}{n}, q = b\frac{\log n}{n}$, that

Theorem 7.1. *We can recover with high probability when* $|\sqrt{a} - \sqrt{b}| \geq \sqrt{2}$.

The intuition for this is to calculate the probability of a bad node (more out of community connections than in community connections) in one of the communities, which by symmetry will appear in the other community with probability $n^{-\frac{\sqrt{a}-\sqrt{b}}{\sqrt{2}}}$. Then the idea is that you swap these paired bad nodes to improve the result. A trick is that if you have 99% of the graph correct, you can clean up the graph by making all the switches, and you will clean up iff $n^{1-\frac{\sqrt{a}-\sqrt{b}}{2}}$, which means you hit the condition. In order to get 99% correct, you can express as an SDP. Let A be the connectivity graph, and $x \in \{-1, 1\}^n$ have 1 to denote one community and -1 denote the other for each of the n nodes. Then

$$\max_{x_i \in \{\pm 1\}, x1 = 0} x^T A x \tag{26}$$

is the problem. The combinatorial part is the restriction ± 1. We can lift this by taking $\mathrm{Tr}(x^T A x) = \mathrm{Tr}(A x x^T) = \mathrm{Tr}(AX)$. This trick is called lifting. Our constraints are now $X > 0$ in the positive definite sense, $X1 = 0$, and $\mathrm{rank}(X) = 1$. The NP-hard part is the $\mathrm{rank}(X) = 1$ aspect. It turns out that throwing out this constraint still gives an SDP that achieves the threshold.

7.2 Detection

This is analgous to giving better bounds in the giant component setting. In the setting where $p = \frac{a}{n}, q = \frac{b}{n}$, there is detection iff $(a-b)^2 > 2(a+b)$, which was given by Massoulié. We (Abbe et al) gave a positive result for weak recovery (this is the same as detection, $> 50\%$ of the communities). First there is a giant if $(a+b)/2 > 1$. From proof from a previous class, we have no giant if $(a+b)/2 < 1$. Now you set up a branching process. Imagine $x_1 \sim Ber(1/2)$. If $x_1 = 0$, then what is the probability $\mathbb{P}\{x_2 = 0 : E_{12} = 1, x_1 = 0\} = \frac{p}{p+q} = \frac{a}{a+b}$.

To achieve the threshold, majority voting works in the case of 2 communities after you set up this broadcasting problem from noisy propagation through the tree. It is believed to work up to 5 communities, where it breaks down; however, nothing has been proven for > 2 communities yet.

7.3 Recovery in the General SBM

Let us take SBM(n, p, W). Define the degree profile of a node $v \in [n]$ by counting its neighbors in each group. Let us say $W = \frac{\log n}{n} Q$, where $Q \in \mathbb{R}^{k \times k}$. The expected number of neighbors a node in community i has in community j is $np_jW_{ij} = \log(n)p_jQ_{ij}$. If v is from community 1, then the degree profile $d_v \sim Pois\left(\log(n)(PQ)_1\right)$. Because the graph is sparse, it still behaves like a Poisson distribution. The first column gives the number of neighbors. If $v \in$ community j, then $d_v \sim Pois\left(\log(n)(PQ)_j\right)$. So now if a node is atypical, we would like to close our eyes, get given a node from one of the four groups, and put it in the middle. We only know how many friends it has in each group. Then we want to tell which group it came from. This is a hypothesis testing problem. Each community has different multinomial Poisson distribution. Then you select the thing with the most likely hypothesis, and it turns out you get the original community. You can do a tournament to select the winner (pairwise comparisons). In each step, you select the correct hypothesis as long as you look at your two big Poisson processes. The mass in the overlap between the distributions tells you how much you screwed up. You have to estimate the area of intersection is $n^{-D_+((PQ)_1 \| (PQ)_2)}$.

This looks a lot like something you have seen before: In channel coding, we ask the question when is x_1 paired with y. When is it that $x_1 \to y$ looks more typical than $x_2 \to y$? It looks like $e^{-nD(P \circ W \| PxW)}$.

Theorem 7.2. *Recovery in the General SBM occurs if*

$$\min_{i<j, i,j \in [k]} D_+((PQ)_i \| (PQ)_j) \geq 1 \tag{27}$$

This divergence is not the KL-divergence.

Definition 7.3. CH-Divergence.

$$D_+(x \| y) = \max_{t \in [0,1]} \sum_{i \in [k]} \left(tx_i + (1-t)y_i - x_i^t y_i^t\right) \tag{28}$$

So this is the fundamental limit of clustering. If $t = 1/2$, this is the Hellinger divergence. If these add to 1, you get the Chernoff divergence. So we call it Chernoff-Hellinger, since it is a generalization of both.

This was originally defined by Csiszor who defined f-divergences.

Definition 7.4. f-divergence.

$$D_f(x\|y) = \sum_{i=1}^{n} y_i f(x_i/y_i) \tag{29}$$

You can recover the KL-divergence by choosing $f(u) = u\log(u)$. The realization in the 1960s was that choosing f to be convex, there are nice properties. These other families have similar properties, but KL-divergence is the correct one in the original setting.

Sham Kakade tried to replicate the clustering result for topic learning recently. In problems where there are a phase transition, you can do similar things.

In general, you would like to infer when a bunch of things have specific featuers, and you get to observe their interactions. This interaction is typically noisy. Now out of these many local interactions, you hope that you have enough to denoise the system to reconstruct the figures you care about. The more edges and interactions you see the better, but also the more noisy the system is, the more difficult it is. These results go beyond the spectral results, and also tells you