Edge-statistics in Ramsey graphs

Talk by Matthew Kwan
Notes by Sanjana Das
January 24, 2023

§1 Introduction

§1.1 Ramsey's theorem

Definition 1.1. A homogeneous set in a graph is a set of vertices where either all possible edges are present, or none are (i.e., a clique or an independent set).

The following two theorems state that in any graph, we can always find a homogeneous set of logarithmic size; but we can't necessarily do better than this.

Theorem 1.2 (Ramsey 1930, Erdős–Szekeres 1935)
Every n-vertex graph has a homogeneous set of size at least $\frac{1}{2}n\log_2 n$.

Theorem 1.3 (Erdős 1947)
A *random* n-vertex graph — specifically, the Erdős–Rényi random graph $\mathcal{G}(n,1/2)$ — typically has no homogeneous set of size greater than $2\log_2 n$.

It's a major open question what the right constant — between $1/2$ and 2 — is. But the aspect of Theorem 1.3 that we'll emphasize today is that this is an instance where randomness really is essential — it's actually an open question to get an explicit construction of a graph that doesn't have homogeneous sets of size $C\log_2 n$ for *any* constant C! The best known result is $(\log n)^{C(\log\log\log n)/(\log\log\log\log n)}$, by Li (2018); the result gives an efficient deterministic algorithm that produces a graph with no homogeneous sets of this size. (This is of interest in computer science because it's connected to derandomization of randomized algorithms.)

Definition 1.4. We say a graph is C-Ramsey if it has no homogeneous set of size larger than $C\log_2 n$.

Question 1.5. Are all C-Ramsey graphs 'random-like' — more specifically, what characteristic properties of random graphs do they possess?

The motivation for this is that it'd explain the phenomenon that it's easy to construct Ramsey graphs with randomness, and hard to construct them without it.

§1.2 Quasirandomness

First, what properties of random graphs do we want to think about? One natural notion to consider is that of *quasirandomness*.

Definition 1.6. We say a graph is quasirandom with density ρ if every vertex subset consisting of k vertices induces about $\rho\binom{k}{2}$ edges (for all k).

It's definitely not true that every Ramsey graph is quasirandom.

Example 1.7

The graph consisting of two disjoint copies of $\mathcal{G}(n/2, 1/2)$ is 4-Ramsey, but is certainly not quasirandom.

However, the next-best thing is true — we can't necessarily say that the densities of our graph on all vertex subsets are roughly the same, but we *can* say that they're all between ε and $1-\varepsilon$ (which is still very strong).

Theorem 1.8 (Erdős–Szemerédi 1972)

For every C, there exists $\varepsilon > 0$ such that for any C-Ramsey graph G, for every subset $U \subseteq V(G)$ with $k \geq \sqrt{n}$ vertices, we have

$$\varepsilon\binom{k}{2} \leq e(G[U]) \leq (1-\varepsilon)\binom{k}{2}.$$

§1.3 Induced subgraphs

However, a lot of the work on this question has considered a very different point of view on the properties of random graphs we're interested in — specifically, that they induce lots of different subgraphs. Here's some history of such results.

Theorem 1.9 (Promel–Rödl 1999)

Every C-Ramsey graph induces a copy of every graph on $c \log n$ vertices.

(The order of quantifiers is 'for every C, there exists c,' and we use n to denote the number of vertices.) This resolved a conjecture of Erdős–Hajnal.

Theorem 1.10 (Shelah 1908)

Every C-Ramsey graph contains at least 2^{cn} non-isomorphic induced subgraphs.

Theorem 1.11 (Kwan–Sudakov 2019)

Every C-Ramsey graph has at least $cn^{5/2}$ induced subgraphs which are distinguishable by their number of vertices and edges.

Theorem 1.12

Every C-Ramsey graph has an induced subgraph with at least $cn^{2/3}$ different degrees.

§1.4 The Erdős–McKay conjecture

Erdős and his collaborators made a series of conjectures on this topic (including the above results); most of these have been resolved since then, but one of these conjectures remained open until now.

Conjecture 1.13 (Erdős–McKay) — For all C, there exists c such that for any $0 \le x \le cn^2$, every C-Ramsey graph on n vertices contains an induced subgraph with exactly x edges.

(Erdős offered \$100 for this conjecture.)

The following theorem proves this for x up to n^c (instead of cn^2).

Theorem 1.14 (Alon–Krivelevich–Sudakov 2003)

For all C, there exists c such that for any $0 \le x \le n^c$, every C-Ramsey graph contains an induced subgraph with exactly x edges.

But in fact, we can get x all the way up to nearly the number of edges in the graph.

Theorem 1.15 (Kwan–Sah–Sauermann–Sawhney)

For all C and ε, every C-Ramsey graph G contains an induced subgraph with exactly x edges for all $0 \le x \le (1-\varepsilon)e(G)$.

In particular, Theorem 1.8 says that a C-Ramsey graph is reasonably dense, so this implies the conjecture. And this theorem is actually a consequence of a stronger theorem about random sampling in Ramsey graphs.

§2 Proof ideas

§2.1 The role of randomness

Usually in combinatorics, we use the probabilistic method to compare the behavior of a random object with its *average* behavior. In particular, a key thing we care about is *concentration* — showing that things tend to be close to their expected values. The notion of quasirandomness is close to this, since it states that any subset of vertices has about the right number of edges (where 'right' corresponds to the average number of edges for a subset of that size).

But most of the proofs of the theorems we've discussed use randomness for the opposite purpose — to show that things are *diverse* (in some sense). Here's an example of such a proof.

Theorem 2.1 (Bukh–Sudakov 2007)

Every C-Ramsey graph has an induced subgraph with $c\sqrt{n}$ different degrees.

Proof sketch. We'll choose U randomly and show that for most vertices u and v, we have

$$\mathbb{P}[\deg_U(u) = \deg_U(v)] \lesssim \frac{1}{\sqrt{n}}.$$

Intuitively, the reason for this is that $\deg_U(u) - \deg_U(v)$ can be written as a sum of independent random variables (for each vertex in the graph, placing it in U either contributes to the degree of u or the degree of v — or both or neither, in whicih case it cancels out and has no effect). And it takes integer values and has standard deviation about $\sqrt{n}$; and it's known that a sum of n independent random variables can't be more concentrated than $1/\sqrt{n}$.

For this to work, we need u and v to have fairly different neighborhoods (for example, if they had exactly the same neighborhood, then they'd always have the *same* degree), so that it's really true that the sum

involves about n independent random variables. And this is where we use the C-Ramsey property — if we had lots of vertices with the same neighborhood, then they'd result in large bipartite or anti-bipartite subgraphs, which we could use to get large cliques or independent sets. □

The unifying philosophy behind these types of results is that in order to show there's many different types of things, we study a *random* one; so to prove the Erdős–McKay conjecture, we imagine choosing U randomly and study the point probabilities $\mathbb{P}[e(G[U]) = x]$.

To illustrate a key difficulty, we can compare this random variable with the one in the proof of Theorem 2.1 (which involved a difference of degrees). That difference could be written as a sum of independent random variables — if we let ξ be the characteristic vector of U, then $\deg_U(u) - \deg_U(v)$ is a linear function of ξ. But on the other hand, $e(G[U])$ is a *quadratic* function of ξ — specifically, we have

$$e(G[U]) = \sum_{uv \in E} \xi_u \xi_v.$$

This is still low-degree, so we can try to use similar techniques, but everything becomes way harder. So we need to develop new techniques for studying low-degree polynomials of independent random variables. (That's one of the things that the authors did which is of interest beyond this particular application.)

§2.2 The Fourier transform

The first idea is to use a Fourier analytic point of view. Letting $X = e(G[U])$ be the random variable we're interested in, we can study its Fourier transform φ_X, the function defined by $t \mapsto \mathbb{E}[e^{itX}]$. Then the Fourier inversion formula states that

$$\mathbb{P}[X = x] = \int_{-\pi}^{\pi} e^{-itx} \varphi_X(t)\, dt. \tag{1}$$

We wanted to bound $\mathbb{P}[X = x]$, and we'll do so by trying to estimate the right-hand side. For this, we have two results that handle 'most' values of t.

Proposition 2.2 (Kwan–Sauermann 2020)
For all $1/n \ll |t| \ll 1$, we have $\varphi_X(t) \approx 0$.

This has the direct probabilistic consequence that $\mathbb{P}[X = x] \lesssim 1/n$ — for a generic quadratic form we have the bound $\mathbb{P}[X = x] \lesssim 1/\sqrt{n}$, so this is an improvement.

To understand intuitively what this statement means, the Fourier transform measures the resonant frequencies of X — so for example, if the density function were periodic, we'd have a spike in the Fourier transform at that frequency. So saying that the Fourier transform is negligible in some range of densities means is like saying that the distribution is smooth at the corresponding scale.

The idea of the proof is to use decoupling techniques to convert a quadratic problem into a linear one. Decoupling is a technique to reduce the degree of a polynomial of independent random variables; it's generally very lossy, but for Ramsey graphs (and the quadratic forms associated to them), it's actually not that lossy. Morally, the reason for this is that the quadratic form has high rank (the adjacency matrix of a Ramsey graph very robustly has high rank, and this remains true under random sampling), and for certain things you need to control in decoupling, it helps to have high rank.

Proposition 2.3
For some Gaussian random variable Z, we have $\varphi_X(t) \approx \varphi_Z(t)$ for $t \ll 1/n$.

Morally, this statement implies a sort of central limit theorem for X.

The proof of this is completely different. Here the idea is that X (which is the number of edges in a random vertex subset) is 'almost linear,' in that if we change a vertex u to be either in or out of the subset, this changes X by essentially $\deg(u)/2$. Linearity means that if you change one input by a certain amount, the change in the output is completely predictable (i.e., fixed), and here that's *almost* true. And then we can use Taylor expansions to approximate φ_X with the Fourier transform of a linear random variable.

Together, Propositions 2.2 and 2.3 mean that we understand the behavior of the characteristic function φ_X everywhere except when t is close to $1/n$ or close to a constant.

§2.3 The gap at $t \approx 1/n$

The first gap at $t \approx 1/n$ turns out to be crucial. If everything were nice and Gaussian, then at $t \approx 1/n$, the characteristic function would be close to negligible. But it turns out that this is just not true — our random variable can actually have two-scale behavior, where there are spikes around $1/n$ — we know the large-scale behavior of X is linear (corresponding to the central limit theorem we get from Proposition 2.3), but there'll be wiggles that come from the interaction of linear and quadratic behavior.

For example, imagine that G is a regular graph. Then the number of edges in an induced subgraph $G[U]$ is well-approximated by the number of *vertices*. But even if we condition on the number of vertices, there's still some randomness that gives us some fluctuation, and that's what corresponds to the wiggles.

So because of this spike at $1/n$, we need a combination of Fourier and physical arguments. Instead of using Fourier analysis on $e(G[U])$ directly, we instead condition on certain information, so that the conditional distributions we get no longer have two-scale behavior and we *can* use Fourier analysis. This gives us point probability estimates, and we integrate over the outcomes of the conditioned information in physical space.

§2.3.1 Conditioning

First, what information should we condition on? If we had a *regular* graph, we'd condition on the number of vertices. In general, what's relevant is the arithmetic structure of the sequence of degrees — for example, if we had two different degrees (and half of our vertices had the first degree, while half had the second), we'd have to separately condition on the number of vertices with each.

In general, there's the notion of RLCD (which stands for 'regularized LCD') from random matrix theory. (We won't define what LCD is — very roughly, it's how much we need to scale up a given unit vector to get something that's close to an integer vector.) We imagine scaling down the degree vector of our graph to have unit length, and then ask how much we need to scale it up to get an integer vector — for example, if our graph is regular then all entries of our unit vector would be $1/\sqrt{n}$, so we'd need to scale it up by $\sqrt{n}$ (while if our unit vector were $(1, 0, 0, \ldots)$ then we wouldn't need to scale it up at all).

And we can take this notion and extract something to condition on (sort of like dividing vertices into buckets and conditioning on the number of vertices in each). And after this conditioning, we've removed the linear component of our random variable, and now it's purely quadratic — previously we had a linear thing with lots of little spikes, but now we've isolated to just one of those little spikes.

Now we have a quadratic polynomial. This doesn't need to satisfy a central limit theorem, but we do have the Gaussian invariance principle — this says that its limiting distribution is some quadratic polynomial of Gaussians (e.g., the distribution of $Z_1^2 + Z_2^2 + Z_3^2$ for independent Gaussians Z_1, Z_2, and Z_3). So Fourier analysis lets us compare the random variable we're interested in to a certain quadratic polynomial of Gaussians, which means our goal is now to understand this quadratic polynomial of Gaussians.

§2.3.2 Anticoncentration for quadratic polynomials of Gaussians

To understand our quadratic polynomial, we need to develop new theory — the best-possible bound for *general* quadratic polynomials is as follows.

Theorem 2.4 (Carbery–Wright)
If $f \in \mathbb{R}[x_1, \dots, x_n]$ is quadratic and $Z_1, \dots, Z_n \sim \mathcal{N}(0,1)$ are independent Gaussians, then

$$\sup_{x \in \mathbb{R}} \mathbb{P}[|f(Z_1, \dots, Z_n) - x| \leq \varepsilon\sigma] \leq O(\sqrt{\varepsilon}),$$

where σ is the standard deviation of $f(Z_1, \dots, Z_n)$.

This is sharp — for example, equality holds for $f(x_1, \dots, x_n) = x_1^2$. But for our purposes, we need a bound of ε and not $\sqrt{\varepsilon}$, so we need to extend this. And the authors show that this behavior (where we get $\sqrt{\varepsilon}$) is atypical — it only happens when the quadratic polynomial is close to low-rank.

Theorem 2.5 (Kwan–Sah–Sauermann–Sawhney)
Suppose that the quadratic part of f robustly has high rank. Then

$$\sup_{x \in \mathbb{R}} \mathbb{P}[|f(Z_1, \dots, Z_n) - x| \leq \varepsilon\sigma] \leq O(\varepsilon).$$

(By 'robustly,' we mean that it's far from a low-rank matrix in some matrix norm.)

To use this, we need to separately prove that the quadratic forms we get from Ramsey graphs have robustly high rank, but this is true. And then this lets us deal with the spike in the Fourier transform at $1/n$.

§2.4 The gap at constant t

There were two regions of the Fourier transform φ_X that we couldn't understand — one at frequencies around $1/n$, and the other at constant frequencies. We've now seen how to deal with the spike at $1/n$, but to use (1), we'd also need a handle on constant frequencies.

Since we don't have control on constant frequencies, we don't exactly get to control the point probabilities $\mathbb{P}[X = x]$. Instead, we get to control probabilities on short intervals — i.e., probabilities of the form $\mathbb{P}[X - x \in I]$ for short intervals I. The constant frequencies correspond to e.g., whether X is biased to be even or odd, and we can't control that.

Then the final step of the proof is showing that within short intervals, we don't have too much bias — i.e., the behavior of X in short intervals is reasonably smooth. We do this in physical space (we're done with the Fourier analytic parts of the argument); we basically want to show that the probability of having x edges isn't too far from the probability for $x + 1$.

The way we do this is by something called the *switching method* — where we study the relative probabilities of two events by designing ways to move from one event to the other, and understanding the number of ways to do this in both directions.

There's an operation that slightly perturbs the set U to go from x to $x + 1$. In combinatorics, the switching method is usually used in a fairly bare-handed way, using uniform estimates — you usually consider *any* point in one event and bound the number of ways to switch. But here it isn't quite as simple — the authors could only show these types of statements on *average* (you can understand the moments of the number of ways to switch in various ways, and use this to get conclusions on the relative probabilities).

§2.5 Conclusion

That's more or less a summary of the proof — at a high level it's Fourier analytic, but you need to condition on certain things to get rid of spikes, and you need to use other methods to get rid of very high frequencies. And the conclusion we get is the following result on randomly sampled induced subgraphs.

Theorem 2.6

Fix a Ramsey graph G, and let $X = e(G[U])$ for a random vertex subset U. Then for $x \approx \mathbb{E}[X]$, we can estimate the point probabilities $\mathbb{P}[X = x]$ up to a constant factor, and they're roughly what they would be if X were Gaussian.

§3 Further questions

Finally, we'll discuss some further questions. A Ramsey graph is a graph with no small homogeneous sets for the strongest possible definition of 'small' (i.e., logarithmic in n). Perhaps it's possible to weaken this assumption.

Question 3.1. Can we weaken the assumption that the graph G is Ramsey — what if the largest homogeneous set is of size at most n^c (for some $c > 0$)?

(This is an open conjecture by Erdős–Hajnal.)

Also, an important fact about Ramsey graphs used in our proof is that there's high rank; so it's interesting to consider the quantitative relationship between the size of the largest homogeneous set and the rank. (This is also related to the log rank conjecture, though that's in a very different regime.)

A third question is that this work is about graphs, and we can ask if we can extend things to hypergraphs. It's not completely obvious how to generalize the definition of Ramsey graphs to hypergraphs, but there are reasonable assumptions we can make on the hypergraph instead (e.g., about disorder or high rank, or looking random-like in some sense). The bigger obstruction is that we'll now have higher degree polynomials. And the proof for quadratic polynomials relies on the diagonalization of quadratic forms, which we can't do with higher degrees.