1 Properties of random graphs

In this talk we will be studying the *binomial* (or *Erdős–Rényi*) random graph model, denoted $G(n,p)$. This is a random n-vertex graph, obtained by making each of the $\binom{n}{2}$ pairs of vertices an edge with probability p, with all these choices made independently.

The study of such random graphs goes back to work of Erdős and Rényi from the 1960s. A fundamental phenomenon that they observed is that random graphs exhibit *thresholds* or *phase transitions*. Loosely speaking, these are results of the following type: for a given property of graphs (e.g. being connected, or containing a triangle, or having chromatic number at least 100, or containing a Hamiltonian path), the probability that $G(n,p)$ has this property is very close to 0 if p is smaller than some function $p_*(n)$, and very close to 1 if p is larger than $p_*(n)$. There is at this point a rich theory of powerful and general meta-theorems, which imply that such threshold phenomena occur for essentially all natural properties of graphs. But rather than stating these general results, I will just give a few examples of such threshold results, to give a sense of their flavor. Let's start with the connectivity theorem of Erdős and Rényi.

Theorem 1 (Erdős–Rényi 1959)**.** *For any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \Pr(G(n,p) \textit{ is connected}) = \begin{cases} 0 & \textit{if } p \leq (1-\varepsilon)\frac{\log n}{n}, \\ 1 & \textit{if } p \geq (1+\varepsilon)\frac{\log n}{n}. \end{cases}$$

Such a result really consists of two statements, which need to be proved separately, and which are usually called the *0-statement* and *1-statement*. In general, completely different arguments need to be used for the two statements, since they are really saying two different things; to prove the 1-statement you need to explain why the random graph is almost surely *connected*, whereas to prove the 0-statement you need to explain why it is almost surely *disconnected*. It is often the case, and the connectivity theorem is a good example of this, that the 0-statement is substantially easier to prove than the 1-statement—often there is a simple "explanation" for why a graph may fail to have a certain property, and then an elementary computation implies the 0-statement. In the case of connectivity, one can show that if $p \leq (1-\varepsilon)\log n/n$, then with high probability $G(n,p)$ has some isolated vertices, and hence is certainly not connected.

Another important threshold result is the following. We denote by v_H, e_H the number of vertices and edges, respectively, of H, and define its *maximum density* to be

$$m(H) = \max_{J \subseteq H} \frac{e_J}{v_J},$$

where the maximum runs over all non-empty subgraphs of H.

Theorem 2 (Erdős–Rényi 1960, Bollobás 1981 in full generality)**.** *Let H be a graph. Then*

$$\lim_{n\to\infty} \Pr(H \textit{ is a subgraph of } G(n,p)) = \begin{cases} 0 & \textit{if } p = o(n^{-1/m(H)}), \\ 1 & \textit{if } p = \omega(n^{-1/m(H)}). \end{cases}$$

Moreover, if $p = cn^{-1/m(H)}$ for some constant c, then $\Pr(H$ is a subgraph of $G(n,p))$ is bounded away from both 0 and 1.

To understand this, we can estimate the expected number of copies of H in $G(n,p)$. There are, up to a constant factor, n^{v_H} many choices for where H can appear in $G(n,p)$. For each such "potential" copy, the probability that it is truly a copy of H is p^{e_H}, as we need e_H random trials to all succeed, and each one succeeds independently with probability p. Therefore, the expected number of copies of H is $\Theta(n^{v_H} p^{e_H})$. Thus, if $p = o(n^{-v_H/e_H})$, then this quantity is $o(1)$, and we expect $G(n,p)$ to have no copies of H. Of course, if H has some subgraph J which is denser than H, we can run the same argument for copies of J, which explains why the *maximal* density $m(H)$ is the relevant parameter.

With an application of Markov's inequality, the heuristic argument above can be turned into a full proof of the 1-statement of Theorem 2, but as before, the 1-statement is noticeably harder to prove.

To conclude this section, let me say a little bit more about the typical structure of a sparse random graph. Denote $d := pn$; up to lower-order terms, this is the expected average degree of the random graph $G(n,p)$. It turns out that the size of d controls many aspects of the typical behavior of $G(n,p)$. For example, we have the following dichotomy.

- If d is a small constant (e.g. 10^{-10}), then $G(n,p)$ almost surely has a very simple structure: all of its connected components have size $O(\log n)$, and the graph is 3-colorable.
- On the other hand, if d is a large constant (e.g. 10^{10}), then the structure of $G(n,p)$ is almost surely very complicated: there is a single, highly connected, "giant component" containing the majority of the vertices, and the chromatic number of the graph is very large.

2 Ramsey properties of graphs

The following foundational result is known as Ramsey's theorem; it is the basis of a large, highly active field of mathematics called Ramsey theory.

Theorem 3 (Ramsey 1930)**.** *For every integer t, there exists an integer n with the property that any 2-coloring of $E(K_n)$ contains a monochromatic K_t.*

This theorem can be equivalently stated in the following form, which looks superficially stronger.

Theorem 4. *For every graph H, there exists a graph G with the property that any 2-coloring of $E(G)$ contains a monochromatic copy of H.*

One says that G is *Ramsey* for H if the property above holds, that is, if every 2-coloring of $E(G)$ contains a monochromatic copy of H. Equivalently, G is Ramsey for H if the edges of G *cannot* be partitioned into two H-free subgraphs. This is a "robust" version of the statement that G has many copies of H: not only does G have many copies, but they are so tightly interlinked that one cannot destroy all of them by splitting $E(G)$ into two parts.

There is an equivalent way of viewing this definition, which is a bit more abstract (and thus somewhat harder to understand at first) but which yields a very useful perspective. Let us denote by $\mathcal{C}_H(G)$ the collection of copies of H within the graph G. We can view $\mathcal{C}_H(G)$ as an r-uniform hypergraph, where $r = e_H$, as follows. Recall that an r-uniform hypergraph consists of a vertex set V and a collection of *hyperedges*, which are subsets of V of size r. Thus, a graph is the same as a 2-uniform hypergraph, and an r-uniform hypergraph is essentially the same as an $(r-1)$-dimensional simplicial complex. In any case, we can turn $\mathcal{C}_H(G)$ into a hypergraph by defining V to be the set of *edges* of G, and an r-subset of V is a hyperedge of $\mathcal{C}_H(G)$ if and only if that set of r edges define a copy of H in G.

Now, an edge-coloring of G is simply a coloring of the *vertices* of $\mathcal{C}_H(G)$, and a monochromatic copy of H in G is simply a hyperedge of $\mathcal{C}_H(G)$ which is monochromatic. Thus, G is *not* Ramsey for H if and only if there is a coloring of the vertices of $\mathcal{C}_H(G)$ with the property that no hyperedge is monochromatic. In other words, G is *not* Ramsey for H if and only if $\mathcal{C}_H(G)$ is bipartite, i.e. has a proper 2-coloring. Equivalently, G *is* Ramsey for H if and only if the chromatic number of $\mathcal{C}_H(G)$ is at least 3.

3 Ramsey properties of random graphs

As one might expect, we can combine the two topics we have been discussing and ask the following question: for a given graph H, what is the threshold that determines whether $G(n,p)$ is almost surely Ramsey for H?

People started asking this question in the late 1980s, and it was resolved in seminal work of Rödl and Ruciński in the mid-1990s. But before stating the answer, let's try to guess it based on some intuitive reasoning.

Recall the hypergraph $\mathcal{C}_H(G)$, consisting of all H-copies in G. If G is itself the random graph $G(n,p)$, then we obtain a *random* hypergraph $\mathcal{C}_H(G(n,p))$. We are interested in understanding the threshold at which the chromatic number of this hypergraph becomes "large", i.e. at least 3; as discussed above, this is precisely the same as saying that $G(n,p)$ is Ramsey for H.

If we use our intuition from random *graphs*, we should expect the chromatic number to be "small" if the average degree of $\mathcal{C}_H(G(n,p))$ is at most a small constant, and should expect the chromatic number to be "large" if the average degree is at least some large constant.

What is the average degree in $\mathcal{C}_H(G(n,p))$? Well, recall that vertices of $\mathcal{C}_H(G(n,p))$ correspond to *edges* of $G(n,p)$, and the expected number of copies of H containing a given edge of $G(n,p)$ is easily seen to be $\Theta(n^{v_H-2}p^{e_H-1})$. Define the *maximal* 2-*density* of H by

$$m_2(H) \coloneqq \max_{J \subseteq H} \frac{e_J - 1}{v_J - 2}.$$

Using similar computations to before, we can conclude the following.

Proposition 5. *If $p \leq cn^{-1/m_2(H)}$, where c is a small constant, then the average degree of $\mathcal{C}_H(G(n,p))$ is at most a small constant. On the other hand, if $p \geq Cn^{-1/m_2(H)}$, where C is a large constant, then the average degree of $\mathcal{C}_H(G(n,p))$ is at least a large constant.*

Because of this, the following theorem should not be surprising.

Theorem 6 (Rödl–Ruciński 1995)**.** *For every graph H with $m_2(H) > 1$, there exist constants $C > c > 0$ such that*

$$\lim_{n\to\infty} \Pr(G(n,p) \text{ is Ramsey for } H) = \begin{cases} 0 & \text{if } p \leq cn^{-1/m_2(H)}, \\ 1 & \text{if } p \geq Cn^{-1/m_2(H)}. \end{cases}$$

The assumption $m_2(H) > 1$ is equivalent to saying that H is not a forest; the threshold for forests is also well-understood, but there are a few annoying edge cases and for simplicity I prefer to avoid them.

While the Rödl–Ruciński theorem agrees exactly with our intuition developed above, that is not to say that it is easy to prove. Moreover, in contrast to the examples we saw before, neither the 0-statement or the 1-statement is easy. Indeed, Rödl and Ruciński developed a wide array of ingeneous techniques for each of the two statements.

In the last decade, there has been a flurry of activity in developing powerful, general tools which allow one to prove the 1-statement of many results, including Theorem 6. These began with the transference principles of Conlon–Gowers, Schacht, and Friedgut–Rödl–Schacht, and more recently the hypergraph container method of Balogh–Morris–Samotij and Saxton–Thomason, which was applied in this setting by Nenadov–Steger. With these techniques, it is now quite straightforward to prove the 1-statement of Theorem 6, and, moreover, to view it as a special case of a general theory, rather than relying on the complex, ad hoc arguments used by Rödl and Ruciński.

However, the 0-statement of Theorem 6 has resisted any such simplification. To understand why, suppose we fix some H, and fix some G which is Ramsey for H. If G is a subgraph of $G(n,p)$, then certainly $G(n,p)$ is Ramsey for H. Therefore,

$$\Pr(G(n,p) \text{ is Ramsey for H}) \geq \Pr(G \text{ is a subgraph of } G(n,p)).$$

By Theorem 2, we know that this latter quantity is bounded away from zero if and only if $p = \Omega(n^{-1/m(G)})$. Hence, if $m(G) \leq m_2(H)$, then the 0-statement of Theorem 6 *cannot be true.*

This explains (part of) the reason why the 0-statement of Theorem 6 is difficult: although it is a probabilistic statement, it is *at least as hard* as the following statement, which Rödl and Ruciński termed the "deterministic lemma".

Lemma 7 (Deterministic lemma, Rödl–Ruciński 1995)**.** *Let H be a graph with $m_2(H) > 1$. If G is Ramsey for H, then $m(G) > m_2(H)$.*

As explained above the deterministic lemma is a natural necessary condition for the 0-statement of Theorem 6 to hold. More surprisingly, Rödl–Ruciński proved that this simple necessary condition is also *sufficient*.

Lemma 8 (Probabilistic lemma, Rödl–Ruciński 1995)**.** *Let H be a graph, and suppose that $m(G) > m_2(H)$ for all graphs G which are Ramsey for H (i.e. the deterministic lemma is true). There exists a constant $c > 0$ such that if $p \le cn^{-1/m_2(H)}$, then*

$$\lim_{n\to\infty} \Pr(G(n,p) \text{ is Ramsey for } H) = 0,$$

i.e. the 0-statement of Theorem 6 is true.

To understand why the probabilistic lemma is true, let us return to our random graph intuition. Recall that if we have a random graph whose average degree is at most a small constant, then all of its connected components are small. One can show that something similar happens in the random hypergraph $\mathcal{C}_H(G(n,p))$: if its average degree is at most a small constant (which is equivalent to saying $p \le cn^{-1/m_2(H)}$), then all of its connected components have *constant* size. Therefore, to prove that $G(n,p)$ is not Ramsey for H—or equivalently that $\mathcal{C}_H(G(n,p))$ is 2-colorable—it suffices to prove that each of these constant-sized connected components is 2-colorable. But this is precisely the statement of the deterministic lemma! Indeed, by Theorem 2, we know that the only constant-sized graphs G that can appear in $G(n,p)$ are those with $m(G) \le m_2(H)$, and hence each of the connected components of $\mathcal{C}_H(G(n,p))$ corresponds to a graph G satisfying $m(G) \le m_2(H)$. By the deterministic lemma, all of these are non-Ramsey for H, meaning that each connected component of $\mathcal{C}_H(G(n,p))$ is 2-colorable.

This perspective also gives a nice way of thinking about the significance of the deterministic lemma, and the meaning of Theorem 6. By the deterministic lemma, even when $p = Cn^{-1/m_2(H)}$, every "small" portion of $G(n,p)$ is not Ramsey for H. Instead, the Ramseyness comes from the global structure of $G(n,p)$, rather than from any local portion. This is closely analogous to the fact that random graphs can have high chromatic number, despite having low chromatic number whenever one considers only a small portion of the graph.

Incidentally, it is the failure of the deterministic lemma which makes Theorem 6 somewhat more subtle for forests. For example, one can check that if H is a path on four vertices and G is obtained from C_5 by adding a leaf to each vertex of the C_5, then G is Ramsey for H and satisfies $m(G) = 1 = m_2(H)$. Thus, the deterministic lemma is simply false for this choice of H, and therefore so is the 0-statement of Theorem 6.

4 Asymmetric Ramsey properties

Up to now, we have been concerned with *symmetric* Ramsey properties, wherein one is interested in finding a monochromatic copy of H in either of the two colors. But one can equally well study *asymmetric* Ramsey properties. Concretely, given a pair of graphs (H, L), let us say that a graph G is *Ramsey for* (H, L) if every red/blue coloring of $E(G)$ contains a red copy of H or a blue copy of L.

As before, for a fixed pair (H, L), we might ask what the threshold is for $G(n,p)$ to be Ramsey for (H, L). Intuitively, it makes sense that being Ramsey for (H, L) is of "intermediate difficulty" between being Ramsey for H and being Ramsey for L. That is, if

$m_2(H) \geq m_2(L)$, then we expect the threshold to fall somewhere between $n^{-1/m_2(L)}$ and $n^{-1/m_2(H)}$. In particular, the following result should not be surprising.

Theorem 9. *If $m_2(H) = m_2(L) > 1$, then there exist constants $C > c > 0$ such that*

$$\lim_{n\to\infty} \Pr(G(n,p) \text{ is Ramsey for } (H,L)) = \begin{cases} 0 & \text{if } p \leq cn^{-1/m_2(H)}, \\ 1 & \text{if } p \geq Cn^{-1/m_2(H)}. \end{cases}$$

The 1-statement of Theorem 9 follows from the Nenadov–Steger proof of Theorem 6 using the container method, whereas the 0-statement was proved recently by Kuperwasser and Samotij.

Because of Theorem 9, we should feel free to assume from now on that $m_2(H) > m_2(L)$ (the names of the graphs represent *heavy* and *light*). We expect there to be some quantity, sandwiched between $m_2(L)$ and $m_2(H)$, which determines the threshold for being Ramsey for $G(n,p)$. Let us try to get some insight as to what this quantity should be; in particular, let us assume that $n^{-1/m_2(L)} \ll p \ll n^{-1/m_2(H)}$, and examine the structure of $G(n,p)$.

Since $p \gg n^{-1/m_2(L)}$, we know that the hypergraph $\mathcal{C}_L(G(n,p))$ has high average degree, and expect it to be very complicated. On the other hand, since $p \ll n^{-1/m_2(H)}$, the opposite is true for $\mathcal{C}_H(G(n,p))$—almost all edges of $G(n,p)$ lie in no copies of H, and we expect the structure of $\mathcal{C}_H(G(n,p))$ to be very simple. Note that if an edge of $G(n,p)$ lies in no copy of H, then it is essentially irrelevant for the Ramsey problem: we may safely color it red without fear of creating a red copy of H, and thus we can effectively simply ignore this edge. In other words, we can focus on the subgraph $G' \subseteq G(n,p)$ consisting of all edges which lie in a copy of H. We now consider the hypergraph $\mathcal{C}_L(G')$, which is an induced subhypergraph of $\mathcal{C}_L(G(n,p))$. Heuristically, we expect the Ramsey property to now hinge on the "complexity" (or average degree) of $\mathcal{C}_L(G')$.

Computing the average degree of $\mathcal{C}_L(G')$ is difficult. But let us pretend that G' is a *uniformly random* subgraph of $G(n,p)$ of a given density, meaning that $\mathcal{C}_L(G')$ is a random induced subhypergraph of $\mathcal{C}_L(G(n,p))$. Then computing its average degree becomes a simple yet tedious computation, and it boils down to the following: our heuristic model for $\mathcal{C}_L(G')$ will have constant average degree if and only if $p = \Theta(n^{-1/m_2(H,L)})$, where the *mixed* 2*-density* is defined by

$$m_2(H,L) := \max_{J \subseteq H} \frac{e_J}{v_J - 2 + 1/m_2(L)}.$$

This reasoning motivates the following conjecture.

Conjecture 10 (Kohayakawa–Kreuter 1997)**.** *For all graphs H, L with $m_2(H) \geq m_2(L) > 1$, there exist constants $C > c > 0$ such that*

$$\lim_{n\to\infty} \Pr(G(n,p) \text{ is Ramsey for } (H,L)) = \begin{cases} 0 & \text{if } p \leq cn^{-1/m_2(H,L)}, \\ 1 & \text{if } p \geq Cn^{-1/m_2(H,L)}. \end{cases}$$

It is not hard to verify that $m_2(L) \leq m_2(H,L) \leq m_2(H)$, and hence Theorem 9 implies Conjecture 10 in the case $m_2(H) = m_2(L)$. Conjecture 10 has been verified in a number of other special cases, including for all pairs of cycles by Kohayakawa and Kreuter, and for all pairs of complete graphs by Marciniszyn, Skokan, Spöhel, and Steger.

The first truly general progress towards Conjecture 10 was due to Mousset, Nenadov, and Samotij, who proved the 1-statement. Their proof also uses the container method, plus a sophisticated additional trick which they term "random typing". Roughly speaking, this trick allows one to pass to the subgraph G' discussed above, and thus get around the bottleneck that the container method "wants to" work at one of the densities $n^{-1/m_2(H)}$ or $n^{-1/m_2(L)}$, rather than in some intermediate range.

As regards the 0-statement of Conjecture 10, progress has been more limited; the most general result prior to our work was due to Hyde, who proved the 0-statement in the majority of cases where H and L are d-regular graphs. Our work resolves Conjecture 10 in almost all cases.

Theorem 11 (Kuperwasser–Samotij–W. 2023)**.** *The 0-statement of Conjecture 10 holds if $m_2(L) > \frac{11}{5}$. Additionally, it holds in any of the following cases (as well as many others):*

- *H is a clique, or a clique minus a small number of edges,*
- *H is complete bipartite,*
- *H is a d-regular bipartite graph,*
- *L is non-bipartite,*
- *L is not the union of two forests.*

Thus, roughly speaking, the only remaining cases are those where L is bipartite and very sparse, whereas H is not too dense nor too structured.

However, in my opinion, our main contribution is not Theorem 11. Rather, it is a probabilistic lemma akin to that of Rödl and Ruciński, which reduces the 0-statement of Conjecture 10 to a purely deterministic statement.

Lemma 12 (Probabilistic lemma, Kuperwasser–Samotij–W. 2023)**.** *Suppose that $m(G) > m_2(H, L)$ for all graphs G which are Ramsey for (H, L). Then the 0-statement of Conjecture 10 holds for the pair (H, L).*

Note that, just as in the symmetric case, the statement "$m(G) > m_2(H, L)$ for all graphs G which are Ramsey for (H, L)" is a clear *necessary* condition for Conjecture 10 to be true; the upshot of the probabilistic lemma is that this simple necessary condition (which we also call the deterministic lemma) is also sufficient. Once we have this, proving Theorem 11 is straightforward, as it turns out that proving the deterministic lemma is fairly simple in a wide array of cases, including most of the ones in Theorem 11.

Moreover, if for some reason you care about Conjecture 10 for some *specific* pair of graphs which is not covered by Theorem 11, all you need to do is to prove the deterministic lemma for this pair, which is usually not too hard. Moreover, our proof of the probabilistic lemma actually gives something stronger: for any pair (H, L), there exists some (explicit) constant

$K = K(H, L)$, such that the deterministic lemma only needs to be verified for all graphs on at most K vertices. While the value of K given by our proof is far too large for any practical purpose, it shows that verifying Conjecture 10 for any given pair is *decidable*—it boils down to checking some property for a finite number of graphs.

I won't say much about the proof of the probabilistic lemma, but I'll try to give a few ideas. The main difficulty in applying the techniques of Rödl–Ruciński is that the relevant hypergraph no longer consists of many small components. Indeed, since we are in the regime $p \gg n^{-1/m_2(L)}$, the hypergraph $\mathcal{C}_L(G(n, p))$ has very high average degree and is very complicated. Of course, we don't actually want to color *this* hypergraph, but rather the "combined" hypergraph of all H-copies and L-copies; however, the complexity of $\mathcal{C}_L(G(n, p))$ makes getting a handle on the "combined" hypergraph very difficult.

Instead, we use an alternative approach. Let us say that G is *minimally Ramsey* for (H, L) if it is Ramsey for (H, L), but any proper subgraph of it is not. Clearly, every Ramsey graph contains a minimally Ramsey subgraph. In particular, if we suppose for contradiction that $G(n, p)$ is Ramsey for (H, L), then there must exist some minimally Ramsey $G \subseteq G(n, p)$. By the deterministic lemma, such a G must be "large", i.e. not of constant size.

Ideally, we would like to union-bound over all potential choices for G, and say that with high probability none of them appear in $G(n, p)$. But this is impossible, as there are far too many choices for G. To get around this, we define a collection $\mathcal{S}$ of *supporting graphs*, satisfying two key properties. First, $\mathcal{S}$ is small, and in fact so small that we can successfully union-bound over it; and second, every minimally Ramsey graph G contains some $S \in \mathcal{S}$ as a subgraph. Hence, once the union bound shows us that $S \nsubseteq G(n, p)$ for all $S \in \S$, we conclude that $G(n, p)$ does not contain any minimally Ramsey G, and thus is not Ramsey for (H, L). This basic idea, of replacing a large, complicated family by a smaller family that "dominates" it, is also the basic idea of the container method. The difficulty in both cases is the construction of the smaller family, and I won't say anything about the construction of $\mathcal{S}$.

The remaining work, of course, is to prove Conjecture 10 in full generality. By the probabilistic lemma, this boils down to proving a simple deterministic lemma: $m(G) > m_2(H, L)$ for all graphs G which are Ramsey for (H, L). In fact, we believe that something much stronger should be true; it is not hard to see that it implies the deterministic lemma in all cases, and it is also an appealing graph partitioning question in its own right.

Conjecture 13 (Kuperwasser–Samotij–W. 2023)**.** *If G is a graph, there exists a forest $F \subseteq G$ such that*

$$m_2(G \setminus F) \leq m(G).$$

In other words, we can partition the edge set of G into a forest F and a graph K such that $m_2(K) \leq m(G)$.

An unexpected fact about Conjecture 13 is that we know how to prove it when $m(G)$ is an integer, using matroid-theoretic techniques that don't seem applicable to the general case. It would be fantastic to prove Conjecture 13 in full generality, and thus settle the Kohayakawa–Kreuter conjecture.