1 Ramsey properties of graphs

Let us say that G is *Ramsey* for H 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.

The basic question of graph Ramsey theory is to understand, for a given graph H, which graphs G are Ramsey for it. For example, one of the simplest results is that K_6 is Ramsey for K_3 (but K_5 is not). This immediately implies that every graph G with $K_6 \subseteq G$ is Ramsey for K_3. This motivated Erdős and Hajnal to ask whether this is the *only* reason why a graph can be Ramsey for K_3; that is, do there exist G which are Ramsey for K_3, but with $K_6 \nsubseteq G$?

This question was rapidly answered in the positive independently by Cherlin, Graham, and van Lint. Later, Pósa proved an even stronger theorem, namely the existence of G which is Ramsey for K_3, but with $K_5 \nsubseteq K_3$. Finally, the strongest possible version of this was proved by Folkman.

Theorem 1.1 (Folkman 1970)**.** *There is a graph G which is Ramsey for K_3, but with $K_4 \nsubseteq G$.*

This is quite remarkable: it shows that the global structure of triangles in G is extremely complicated (because G is Ramsey for K_3), whereas the local structure is very simple, since G is K_4-free. Several more general versions of such a statement, yielding "locally sparse" Ramsey graphs, were proven by Nešetřil and Rödl in the 1970s.

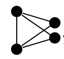
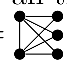
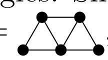
If you squint your eyes and forget the details, Folkman's theorem appears to be describing some sort of expander. Indeed, sparse expanders are graphs whose local structure is very simple (e.g. they may have large girth), but whose global structure is very complex (e.g. they may have sublinear independence number, hence large chromatic number). The constructions of Folkman and Nešetřil–Rödl were completely explicit, but we know that a good way of finding expanders is via randomized constructions. This is one of several motivations for asking about the Ramsey properties of random graphs.

2 Ramsey 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. As suggested above, we 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, focusing on the case $H = K_3$. A key fact we need is a simple result of Erdős and Rényi (and of Bollobás in the most general case), which determines the threshold at which $G(n,p)$ almost surely contains a copy of a fixed graph F: if $p \gg n^{-1/m(F)}$ then $F \subseteq G(n,p)$ a.a.s., whereas if $p \ll n^{-1/m(F)}$ then $F \nsubseteq G(n,p)$ a.a.s., where

$$m(F) = \max_{J \subseteq F} \frac{e_J}{v_J}.$$

For $G(n,p)$ to be Ramsey for K_3, we'd better have that $K_3 \subseteq G$, hence the above implies that we should take $p \gg n^{-1/m(K_3)} = n^{-1}$. However, this is clearly not sufficient: for example, if $p \ll n^{-4/5}$, then $G(n,p)$ contains no copy of $B_2 :=$ , the graph obtained by gluing two triangles along an edge. If $G(n,p)$ contains no copy of B_2, then all the triangles in $G(n,p)$ are edge-disjoint, hence we can easily color the edges of G to destroy all the triangles. Similarly, if $p \ll n^{-5/7}$, then $G(n,p)$ does not contain a copy of either $B_3 :=$  or $P_3 :=$ , which implies that every "connected component of triangles" contains at most two triangles. In this case, it is again easy to color the edges of $G(n,p)$ and destroy all triangles.

Continuing in this fashion, one comes to the conclusion that for the triangles in $G(n,p)$ to "interact a lot", we should have that a typical edge lies in many (i.e. a large constant number of) triangles; this is for essentially the same reason that $G(n,p)$ itself becomes "very complex" once its average degree is a large constant. A simple computation shows that the average edge is on a constant number of triangles exactly when $p \asymp n^{-1/2}$.

Another way of getting at the same answer is to count constraints and degrees of freedom. If our goal is to color $E(G(n,p))$ in such a way that no copy of K_3 is monochromatic, then we have $|E(G(n,p))| \asymp pn^2$ degrees of freedom—one for every edge—and we have as many degrees of freedom as we have copies of K_3 in $G(n,p)$, which is roughly p^3n^3. Heuristically, finding an assignment satisfying all the constraints should be possible if the number of constraints is much smaller than the number of degrees of freedom, and it should be impossible if there are many more constraints than degrees of freedom. Solving $pn^2 \asymp p^3n^3$ again gives $p \asymp n^{-1/2}$, for the same reason: saying that the number of edges is comparable to the number of triangles is the same as saying that a typical edge lies on a constant number of triangles.

More generally, given any graph H, we should expect $G(n,p)$ to become Ramsey for H once a typical edge lies on a constant number of H-copies. Working through some simple computations then suggests that the relvant parameter is the *maximal* 2*-density* of H, defined as

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

where the maximum runs over all subgraphs of H with at least three vertices.

And indeed, this intuition is proven correct by the following foundational theorem.

Theorem 2.1 (Rödl–Ruciński 1995)**.** *For every graph H with $m_2(H) > 1$, there exist constants $C > c > 0$ such that*

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

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.

I tried to motivate the study of these questions via Folkman's theorem, Theorem 1.1. However, as written, Theorem 2.1 is not sufficient to prove Theorem 1.1. Indeed, plugging in $H = K_3$, we see that $m_2(K_3) = 2$, hence Theorem 2.1 implies that $G(n,p)$ is Ramsey for K_3 only once $p = \Omega(n^{-1/2})$. However, $m(K_4) = \frac{3}{2}$, so at this density, there actually are plenty of copies of K_4 in $G(n,p)$, and hence $G(n,p)$ does not automatically yield a K_4-free graph which is Ramsey for K_3. This problem can be remedied by finding sufficiently strong estimates for the rate of convergence in Theorem 2.1, combined with a neat application of Harris's inequality, but we omit the details; we will shortly see a different proof of Theorem 1.1.

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 many examples in random graph theory, 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 2.1. 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 2.1, 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 2.1 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,

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

For a fixed G, this latter quantity is bounded away from zero if and only if $p = \Omega(n^{-1/m(G)})$. Hence, if $m(G) \leqslant m_2(H)$, then the 0-statement of Theorem 2.1 *cannot be true*.

This explains (part of) the reason why the 0-statement of Theorem 2.1 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 2.2 (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 2.1 to hold. More surprisingly, Rödl–Ruciński proved that this simple necessary condition is also *sufficient*.

Lemma 2.3 (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 $p \leqslant cn^{-1/m_2(H)}$, then*

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

i.e. the 0-statement of Theorem 2.1 is true.

To understand why the probabilistic lemma is true, let us return to the intuition above. Recall that we argued that if $p \ll n^{-5/7}$, then all "connected components of triangles" in $G(n,p)$ contain at most two triangles. More generally, one can prove that if $p \leqslant cn^{-1/m_2(H)}$ for a sufficiently small constant $c > 0$, then all "cores of connected components of H-copies" have *constant* size. Therefore, to prove that $G(n,p)$ is not Ramsey for H, it suffices to prove that each of these constant-sized connected components can be edge-colored without creating any monochromatic copy of H. But this is precisely the statement of the deterministic lemma! Indeed, the only constant-sized graphs G that can appear in $G(n,p)$ are those with $m(G) \leqslant m_2(H)$, and hence each of the connected components corresponds to a graph G satisfying $m(G) \leqslant m_2(H)$. By the deterministic lemma, all of these are non-Ramsey for H, which is exactly what we wanted to show.

This perspective also gives a nice way of thinking about the significance of the deterministic lemma, and the meaning of Theorem 2.1. 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 2.1 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 2.1.

3 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 any set of graphs $H_1, \ldots, H_q$, let us say that a graph G is *Ramsey for* $(H_1, \ldots, H_q)$ if every q-coloring of $E(G)$ contains a monochromatic copy of H_i in the ith color, for some i.

For the moment, let us focus on the case $q = 2$. We may assume without loss of generality that $m_2(H_1) \geqslant m_2(H_2)$. It is natural to expect that being Ramsey for (H_1, H_2) is of "intermediate difficulty" between being Ramsey for H_1 and being Ramsey for H_2. Therefore, we should expect the threshold for $G(n,p)$ to be Ramsey for (H_1, H_2) to be somewhere between $n^{-1/m_2(H_2)}$ and $n^{-1/m_2(H_1)}$.

At such an "intermediate" density, a typical edge is in very many copies of H_2, but in very few copies of H_1. Note that if an edge of $G(n,p)$ lies in *no* copy of H_1, then it is essentially irrelevant for the Ramsey problem: we may safely color it with color 1 without fear of creating a monochromatic copy of H_1, and thus we can simply ignore this edge. This suggests that our number of "real" degrees of freedom equals the number of edges in a copy of H_1. Every copy of H_1 or of H_2 gives us a constraint, but there are far more copies of H_2 than of H_1 (because H_2 is sparser than H_1), so the number of constraints essentially equals the total number of copies of H_2.

Equating these two quantities, one can eventually determine that the relevant quantity seems to be the *mixed* 2*-density*, which is defined by

$$m_2(H_1, H_2) := \max_{J \subseteq H_1} \frac{e_J}{v_J - 2 + 1/m_2(H_2)}.$$

This reasoning motivates the following conjecture.

Conjecture 3.1 (Kohayakawa–Kreuter 1997)**.** *Let $H_1, \ldots, H_q$ be graphs with $m_2(H_1) \geqslant \cdots \geqslant m_2(H_q)$ and $m_2(H_2) > 1$. There exist constants $C > c > 0$ such that*

$$\lim_{n\to\infty} \Pr(G(n,p) \text{ is Ramsey for } (H_1, \ldots, H_q)) = \begin{cases} 0 & \text{if } p \leqslant cn^{-1/m_2(H_1,H_2)}, \\ 1 & \text{if } p \geqslant Cn^{-1/m_2(H_1,H_2)}, \end{cases}$$

The motivation above hopefully supports this conjecture in the case $q = 2$. An interesting feature is that adding further colors and graphs $H_3, \ldots, H_q$ changes nothing (except for the constants C and c); the intuitive reason is that, once the edges in H_1-copies start interacting in sufficiently complex ways with the H_2-copies, they will also do the same with copies $H_3, \ldots, H_q$, and hence nothing fundamental about the problem should change by the addition of these extra graphs.

For a while, Conjecture 3.1 had only been verified in a number of special cases, including when all H_i are cycles by Kohayakawa and Kreuter, when all H_i are complete graphs by Marciniszyn, Skokan, Spöhel, and Steger, and when each H_i is a clique or a cycle by Liebenau, Mattos, Mendonça, and Skokan. Recently, Conjecture 3.1 was finally proven in full generality.

Theorem 3.2 (Mousset–Nenadov–Samotij, Kuperwasser–Samotij–W., Bowtell–Hancock–Hyde, Christoph–Martinsson–Steiner–W.)**.** *Conjecture 3.1 is true.*

Before discussing the proof of Conjecture 3.1, let us see two quick applications, including a proof of Folkman's theorem.

Corollary 3.3. *There exists a graph G_0 which is Ramsey for (K_3, K_3, K_3), but not Ramsey for (K_4, K_3).*

Proof. By Theorem 3.2, we know that the threshold for $G(n,p)$ to be Ramsey for (K_3, K_3, K_3) is $p \asymp n^{-1/m_2(K_3,K_3)} = n^{-1/2}$. On the other hand, the threshold for $G(n,p)$ to be Ramsey for (K_4, K_3) is $p \asymp n^{-1/m_2(K_4,K_3)} = n^{-5/12}$. So if we pick $p = n^{-11/24}$, or indeed any function of the form $o(n^{-5/12})$ and $\omega(n^{-1/2})$, then we see asymptotically almost surely, $G(n,p)$ satisfies the desired property. Hence, for sufficiently large n, we may set G_0 to be a sample of $G(n,p)$ and obtain the claimed result. □

Corollary 3.3 is itself a "Folkman-type" statement, in the sense that it proves the existence of a graph G_0 satisfying a Ramsey property, but which is not so complex to satisfy a different Ramsey property. Moreover, as observed by Kuperwasser, a result like Corollary 3.3 immediately yields a proof of Folkman's theorem, Theorem 1.1.

Proof of Theorem 1.1 (Kuperwasser). Let G_0 be the graph from Corollary 3.3. Since G_0 is not Ramsey for (K_4, K_3), there is a coloring of its edges in, say, black and white such that the black graph is K_4-free and the white graph is K_3-free. Let G be the black graph, which is K_4-free by construction. Moreover, every red/blue-coloring of G, when combined with the white graph, gives a three-coloring of $E(G_0)$. By the fact that G_0 is Ramsey for (K_3, K_3, K_3), this coloring must contain a monochromatic triangle, and it cannot be white since the white graph is K_3-free. Hence there is a red or blue K_3, showing that G is Ramsey for K_3. □

4 Proof overview

As with the Rödl–Ruciński theorem, there are three main parts to the proof of Conjecture 3.1: the 1-statement, the probabilistic lemma, and the deterministic lemma. Each of them is substantially harder than the corresponding statement in the proof of Theorem 2.1, so let us discuss them in turn.

First, the 1-statement was proved by Mousset, Nenadov, and Samotij, using the container method. However, a naive application of the container method does not work, roughly because the container method "wants to" work at one of the densities $n^{-1/m_2(H_1)}$ or $n^{-1/m_2(H_2)}$, rather than in some intermediate range. To get around this issue, Mousset–Nenadov–Samotij introduced a sophisticated additional trick which they term "random typing". Roughly speaking, this trick allows one to pass to a subgraph $G' \subseteq G(n,p)$, which lives at the lower density $n^{-1/m_2(H_2)}$, and then apply the container method to this graph. Of course, this cannot be done directly, since discovering the subgraph G' requires revealing all the randomness in $G(n,p)$, and resolving this issue via random typing is the main innovation of their work.

Let us now turn to the 0-statement. Note that for the 0-statement, it suffices to deal with the case of $q = 2$ colors, simply because being Ramsey for $(H_1, \ldots, H_q)$ implies being Ramsey for (H_1, H_2), and hence a lower bound on the threshold probability for the latter event implies the same lower bound for the former event. As mentioned above, the proof of the 0-statement boils down to a deterministic and a probabilistic lemma.

Theorem 4.1 (Deterministic lemma, Christoph–Martinsson–Steiner–W.)**.** *Let $m_2(H_1) \geqslant m_2(H_2) > 1$. If G is Ramsey for (H_1, H_2), then $m(G) > m_2(H_1, H_2)$.*

As before, the deterministic lemma is a clear *necessary* condition for the 0-statement of Conjecture 3.1 to be true. The following probabilistic lemma, proved independently by two groups, demonstrates that this necessary condition is also *sufficient*.

Lemma 4.2 (Probabilistic lemma; Bowtell–Hancock–Hyde, Kuperwasser–Samotij–W.)**.** *Let H_1, H_2 be graphs with $m_2(H_1) \geqslant m_2(H_2) > 1$. The 0-statement of Conjecture 3.1 for (H_1, H_2)*

is equivalent to the deterministic lemma, i.e. to the statement $m(G) > m_2(H_1, H_2)$ if G is Ramsey for (H_1, H_2).

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 we no longer have constant-sized "connected components of H-copies". Indeed, since we are in the regime $p \gg n^{-1/m_2(H_2)}$, the set of H_2-copies is very complicated, and its connected components are very large. Of course, our saving grace is that the H_1-copies are extremely simple, since $p \ll n^{-1/m_2(H_1)}$. However, figuring out how to utilize this saving grace is quite tricky.

Instead, we use an alternative approach. Let us say that G is *minimally Ramsey* for (H_1, H_2) if it is Ramsey for (H_1, H_2), 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_1, H_2), 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 $S \in \mathcal{S}$, we conclude that $G(n, p)$ does not contain any minimally Ramsey G, and thus is not Ramsey for (H_1, H_2). 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 $\mathcal{S}$.

Finally, let us turn to the proof of the deterministic lemma, Theorem 4.1. It is a bit easier to think about in the contrapositive: if $m(G) \leqslant m_2(H_1, H_2)$, then G can be partitioned into an H_1-free graph and an H_2-free graph.

The difficulty with proving such a statement is that we get to assume essentially nothing about H_1 and H_2, and hence it is not clear how to find these H_i-free subgraphs of G. If one assumes certain extra structure on H_1 and H_2 (e.g. that H_2 is non-bipartite, or that H_1 is a clique), then proving such a statement becomes fairly straightforward. However, the tricks that work in such special cases don't seem to work in full generality. As such, it is natural to try to prove a more general statement, which implies Theorem 4.1 without using almost anything about the structure of H_1 and H_2. The following conjecture is such a statement.

Conjecture 4.3 (Kuperwasser–Samotij–W. 2023)**.** *If G is a graph, there exists a forest $F \subseteq G$ such that*

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

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

Indeed, if $m_2(H_1) > m_2(H_2) > 1$, then one can verify that H_2 is not a forest (hence not a subgraph of F), and that $m_2(H_1) > m_2(H_1, H_2)$ (hence, if $m(G) \leqslant m_2(H_1, H_2)$, then H_1 is not a subgraph of K).

Conjecture 4.3 looks very similar to a number of well-known results in the theory of graph partitioning. For example, a famous theorem of Nash-Williams states that a graph G can be partitioned into k forests if and only if $m_1(G) \leqslant k$, where

$$m_1(G) \coloneqq \max_{J \subseteq G} \frac{e_J}{v_J - 1}.$$

One consequence of this is, for example, that G can be partitioned into a forest F and a graph K with $m_1(K) \leqslant \lceil m_1(G) \rceil - 1$.

Nash-Williams's theorem is a special case of the very general matroid partitioning theorem of Edmonds. Using such matroid-theoretic techniques, Kuperwasser, Samotij, and I proved Conjecture 4.3 under the added assumption that $m(G)$ is an integer. However, it seems that such integrality assumptions are a necessary component of any matroid-theoretic proof technique, and as such Conjecture 4.3 seems out of reach at the moment.

However, in order to prove Theorem 4.1, Christoph, Martinsson, Steiner, and I proved two natural weakenings of Conjecture 4.3, which together suffice to prove the deterministic lemma.

Theorem 4.4 (Christoph–Martinsson–Steiner–W.)**.** *Let G be a graph.*

(a) There exists a pseudoforest $P \subseteq G$ such that

$$m_2(G \setminus P) \leqslant m(G).$$

Here, a pseudoforest is a graph with at most one cycle per connected component.

(b) If $m(G) > \frac{3}{2}$, there exists a forest $F \subseteq G$ such that

$$m_{\frac{4}{3}}(G \setminus F) \leqslant m(G),$$

where

$$m_{\frac{4}{3}}(H) \coloneqq \max_{J \subseteq H} \frac{e_J}{v_J - \frac{4}{3}}.$$

These are not the most general results provable by our techniques (in particular the choice of $\frac{4}{3}$ is not of great significance), but they are two simple statements which together imply Theorem 4.1. I believe that there should be a much more general theory of sparsity measures on graphs, with accompanying graph partitioning results, which should hopefully yield general-purpose proofs of Theorems 4.1 and 4.4, and other results such as Nash-Williams's theorem.

As matroid-theoretic techniques do not seem capable of proving "fractional" results like Theorem 4.4, we had to introduce alternative techniques. The basic tool is one we term *allocations*, which a fractional version of a classical technique of Hakimi and Frank–Gyárfás.

Basically, their results allow one to turn upper bounds on $m(G)$ into an orientation of $E(G)$ with bounded out-degrees. Our technique is an extension of this, where every edge is not assigned an orientation, but a "fractional" orientation.

Concretely, let us define an *allocation* to be a function $\theta : V(G)^2 \to [0,1]$ satisfying the following two properties: $\theta(u,v) = 0$ if $uv \notin E(G)$, and $\theta(u,v) + \theta(v,u) = 1$ for every $uv \in E(G)$. In case θ is integer-valued, this is exactly an orientation of $E(G)$, namely we orient the edge uv as $u \to v$ if $\theta(u,v) = 1$, and as $u \leftarrow v$ if $\theta(v,u) = 1$. Extending this intuition, we define the θ-outdegree of a vertex u to be $\sum_{v \in V(G)} \theta(u,v)$. The key lemma relating allocations to densities is the following.

Lemma 4.5. *Let G be a graph and let m be a real number. Then $m(G) \leqslant m$ if and only if there is an allocation $\theta : V(G)^2 \to [0,1]$ such that the θ-outdegree of every vertex is at most m.*

There are a number of ways of proving this lemma, but one of the simplest is as a quick application of the max-flow min-cut theorem.

Now, the basic idea for proving Theorem 4.4 is as follows. We are given a graph G, and let $m := m(G)$. By Lemma 4.5, we obtain an allocation $\theta : V(G)^2 \to [0,1]$ with all θ-outdegrees bounded by m. If θ were an honest edge-orientation, then an obvious way of picking a pseudoforest $P \subseteq G$ is to have every vertex select one out-neighbor; the resulting oriented subgraph has maximum outdegree 1, hence is a pseudoforest. Again, if θ were an orientation, we would have decreased the outdegree of every vertex by 1, hence the subgraph $G \setminus P$ would have an edge-orientation with maximum outdegree at most $m-1$; by the converse of Lemma 4.5, we would conclude that $m(G \setminus P) \leqslant m - 1$, which suffices to prove Theorem 4.4(a). In fact, this argument can be made to work if m is an integer, and essentially recovers a well-known theorem of Hakimi (which is proved in the same way).

Unfortunately, θ need not be a true orienation, so we need to work a lot harder. But the basic strategy is the same, after some "pre-processing". Namely, we begin by modifying the allocation θ, while preserving the bound on the θ-outdegree. For example, it is not hard to see that we can shift the value of θ along any cycle of G without changing the θ-outdegree of any vertex. By performing a number of modifications of this type, we can assume that θ is of a certain special form. We now again have every vertex pick one "outgoing edge" (although even defining this is a bit harder, since a typical edge has only a fractional orientation), and then argue that (a) this yields a pseudoforest, and (b) that its removal yields $m_2(G \setminus P) \leqslant m(G)$. To accomplish this latter goal, we assume for contradiction that some subgraph of $G \setminus P$ is too dense, and show that this contradicts the choice of θ.

The proof of Theorem 4.4(b) is similar, but much more involved. The biggest problem is that we can no longer obtain a forest by simply assigning every vertex one outgoing edge, as this may create cycles. To overcome this, we begin by analyzing the global structure of θ, in order to identify "safe" places to put the leaves of the forest, i.e. vertices with no outgoing edge. Additionally, the verification that $m_{\frac{4}{3}}(G \setminus F) < m(G)$ is also much more involved, and breaks into a large number of cases depending on the exact value of m, as well as on the structure of θ.