A tutorial on the spread method

Talk by Tom Kelly
Notes by Sanjana Das

March 15, 2024

§1 Three theorems

Today we're going to discuss some applications of the Park–Pham theorem. We'll focus on three theorems (which we'll work with throughout). First, we'll set up some notation.

Definition 1.1. For a graph or hypergraph G and some $p \in [0,1]$, we use G_p to dentoe the random subgraph of G where we keep each edge independently with probability p.

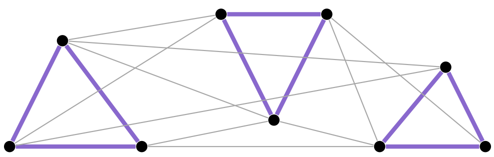
Example 1.2

The Erdős–Rényi random graph $\mathcal{G}_{n,p}$ is simply $(K_n)_p$ — we start with all edges, and include each with probability p.

Notation 1.3. We use $\delta(G)$ to denote the minimum degree of G.

All three theorems consider the question of when G_p contains a specific object. The first theorem considers when G_p (for a graph G) contains a triangle factor.

Definition 1.4. A triangle factor in a graph is a collection of vertex-disjoint triangles that covers all the vertices (meaning that each vertex is in exactly one triangle).



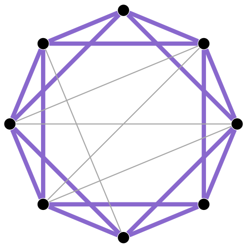
Theorem 1.5

For every $\varepsilon > 0$, there exists C such that if G is an n-vertex graph with $\delta(G) \geq (\frac{2}{3} + \varepsilon)n$ and we have $p \geq C(\log n)^{1/3}n^{-2/3}$, then G_p asymptotically almost surely has a triangle factor.

(Lots of the theorems we'll discuss need some divisibility condition on n — for example, if an n-vertex graph has a triangle factor, then we need $3 \mid n$. But we'll typically omit these conditions for brevity — for example, the statement of Theorem 1.5 really should have the extra hypothesis that $3 \mid n$.)

This was first proven by Allen, Böttcher, Davies, Jenssen, Morris, Roberts, and Skokan (2022), and then generalized by Pham, Sah, Sawhney, and Simkin. (In fact, the authors proved it even without the ε.)

The next theorem considers when G_p contains the square of a Hamiltonian cycle — for a graph G to contain a Hamiltonian cycle means that we can number the vertices $1, \ldots, n$ so that G contains the edges 12, 23, 34, ..., $n1$, and for G to contain the *square* of a Hamiltonian cycle means that we can number the vertices so that it contains all these edges as well as the edges 13, 24, 35, ..., $n2$. (In general, the kth power of a graph H is defined as the graph where we connect all vertices whose distance in H was at most k.)



Theorem 1.6

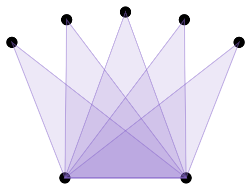
For every $\varepsilon > 0$, there exists C such that if G is an n-vertex graph with $\delta(G) \geq (\frac{2}{3} + \varepsilon)n$ and we have $p \geq Cn^{-1/2}$, then G_p asymptotically almost surely contains the square of a Hamiltonian cycle.

For the third theorem, we'll have the same setup, but with hypergraphs instead of graphs.

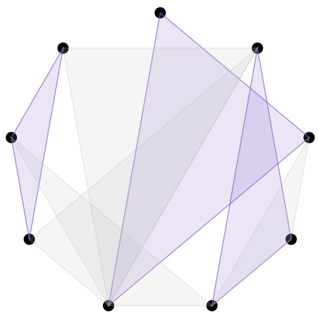
Notation 1.7. For a 3-graph (i.e., 3-uniform hypergraph) G, we define

$$\delta_2(G) = \min_{u,v} \#(\text{edges containing } u \text{ and } v).$$

In other words, saying that $\delta_2(G) \geq d$ means that for every pair of vertices, there's at least d edges containing both of them (where an edge in a 3-graph is a triple of vertices).



And the object we'll consider is a *perfect matching* — i.e., a bunch of disjoint edges that cover all the vertices (we again need $3 \mid n$ for a 3-graph to have a perfect matching).



Theorem 1.8

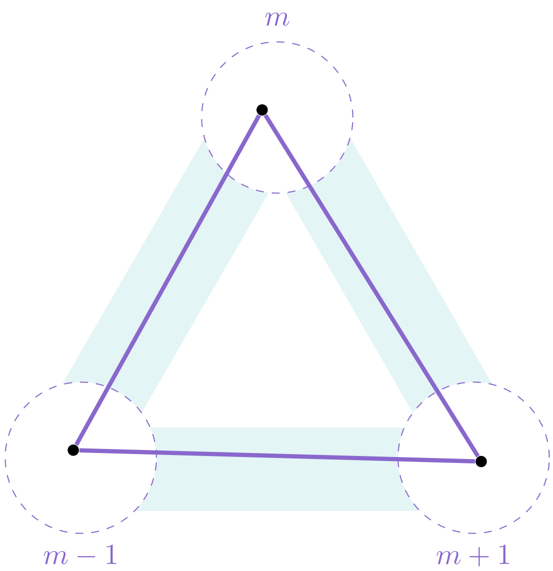
For every $\varepsilon > 0$, there exists C such that if G is an n-vertex 3-graph with $\delta_2(G) \geq (\frac{1}{2} + \varepsilon)n$ and we have $p \geq C(\log n)n^{-2}$, then G_p asymptotically almost surely has a perfect matching.

This was proven in the same paper by Pham, Sah, Sawhney, and Simkin, as well as in a paper by Kang, Kelly, Kühn, Osthus, and Pfenninger.

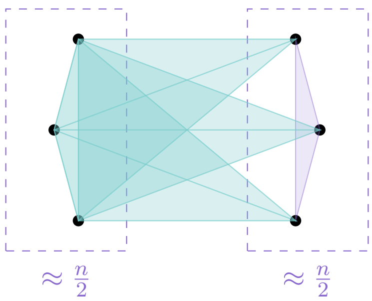
§1.1 The case $p = 1$

First, these theorems are already nontrivial in the case $p = 1$, where they're about finding some structure in a *fixed* graph with a certain minimum degree. Theorem 1.5 with $p = 1$ follows from the Corrádi–Hajnal theorem from the 1960s, or more generally the Hajnal–Szemerédi theorem from the 1970s (which considers the problem of finding a K_r-factor). Theorem 1.6 with $p = 1$ follows from results of Komlós, Sárközy, and Szemerédi from the 1990s, in a series of influential papers using the blowup lemma. And Theorem 1.8 with $p = 1$ follows fro work of Rödl, Ruciński, and Szemerédi from the late 2000s using absorption methods. All of these are important results on their own.

In particular, even from the case $p = 1$ we can see that the constant $\frac{2}{3}$ in Theorem 1.5 (in the bound on the minimum degree) is the best possible — as a counterexample, imagine a slightly unbalanced (complete) tripartite graph, with part sizes m, $m - 1$, and $m + 1$ (where $n = 3m$). This graph has minimum degree $2m - 1 \approx \frac{2}{3}n$, but it doesn't have a triangle factor — any triangle uses one vertex from each part, but the parts don't have equal sizes.



The constant $\frac{1}{2}$ in Theorem 1.8 is also tight. For a construction, we can split the vertices into two parts (a left part and a right part) of roughly equal sizes, such that the left part has odd size. And we include only the edges that have an even number of vertices (i.e., 0 or 2) from the left part. The minimum 2-degree of this hypergraph is roughly $\frac{1}{2}n$, but it doesn't have a perfect matching — if it did, then the number of vertices in the left half would be even.



§2 The threshold versions and the Park–Pham theorem

First we're going to consider the case where G is the complete graph. Then G_p becomes the Erdős–Rényi random graph $\mathcal{G}_{n,p}$, and the three theorems are all about thresholds — we're asking for the threshold for $\mathcal{G}_{n,p}$ to have a triangle factor or square of a Hamiltonian cycle, or for its 3-uniform version to have a perfect matching. (The theorems are really stating *upper* bounds on these thresholds — for example, Theorem 1.5 states that the threshold for $\mathcal{G}_{n,p}$ to have a triangle factor is at most $C(\log n)^{1/3}n^{-2/3}$.)

§2.1 Threshold lower bounds

First, we can see where the values of p in the theorems come from by showing matching *lower* bounds on the corresponding thresholds (which means the conditions on p are tight, up to constants).

For Theorem 1.5, if $\mathcal{G}_{n,p}$ has a triangle factor, then certainly every vertex should be contained in a triangle. We can first do an expected value calculation — if we fix a vertex v, then the expected number of triangles containing v is roughly p^3n^2 (since there's roughly n^2 choices for the other two vertices, and p^3 is the probability that all three edges of the triangle are present). Then by a coupon collector-type argument, we can show that for *every* vertex to be contained in a triangle (with high probability) we need $p^3n^2 \gtrsim \log n$, which rearranges to $p \gtrsim (\log n)^{1/3}n^{-2/3}$.

Remark 2.1. This argument can be made rigorous to show that $(\log n)^{1/3}n^{-2/3}$ is a lower bound for the threshold. It turns out that it's also an upper bound for the threshold (as stated in Theorem 1.5) — this was first proved in a famous paper by Johansson, Kahn, and Vu from 2008, which is pretty technical and difficult. But today we'll see that it can be derived from the Park–Pham theorem in a way that's much easier.

Theorem 1.6 is about the threshold for $\mathcal{G}_{n,p}$ to have the square of a Hamiltonian cycle. Here it's again easy to get a *lower* bound on this threshold — we can just compute the expected number of squares of Hamiltonian cycles in $\mathcal{G}_{n,p}$. There's $n!$ ways to arrange the vertices, and for each such arrangement, to get the corresponding square of a Hamiltonian cycle we need $2n$ edges to be present, which occurs with probability p^{2n}. Using a crude bound on $n!$ gives

$$\mathbb{E}[\#\text{squares of Hamiltonian cycles}] \approx n!\, p^{2n} \approx \left(\frac{np^2}{e}\right)^n$$

(we're overcounting a bit — several arrangements may correspond to the same Hamiltonian cycle — but it doesn't matter). So for this expectation to be at least 1 we need $np^2/e \geq 1$, meaning that $p \geq \sqrt{e/n}$.

Remark 2.2. Theorem 1.6 gives an upper bound of $Cn^{-1/2}$ for the threshold, which matches this lower bound up to constants. But it's an open question whether this is a *sharp* threshold — if $p \geq \sqrt{(e+\varepsilon)/n}$, then is it true that $\mathcal{G}_{n,p}$ asymptotically almost surely has the square of a Hamiltonian cycle?

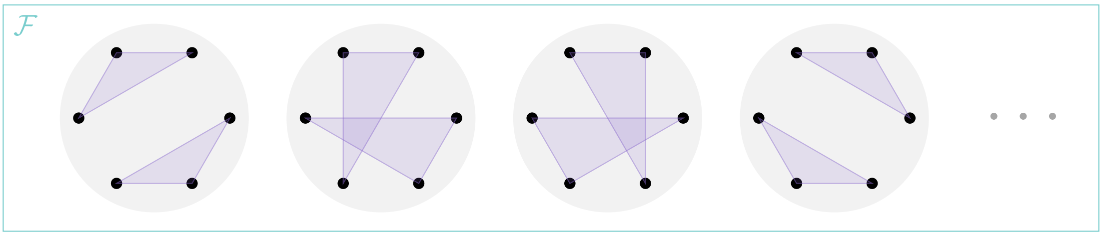
Finally, Theorem 1.8 is about the threshold for $\mathcal{G}^{(3)}_{n,p}$ (the random 3-graph where we include each edge with probability p) to have a perfect matching. Here the lower bound on this threshold is very similar to the one for Theorem 1.5 (every vertex has to be included in some edge, and we can deal with this via a coupon collector argument), so we won't go through the calculation again. Proving the matching upper bound was known as Shamir's problem, and it was also solved by Johansson, Kahn, and Vu in the same paper.

Remark 2.3. It's known that all these problems have *sharp* thresholds, due to Friedgut's theorem. For Theorems 1.5 and 1.8 we know what these sharp thresholds are; but for Theorem 1.6 we don't.

§2.2 The Park–Pham theorem

Now we're going to discuss how to prove these threshold upper bounds (i.e., the cases of the three theorems where G is complete), using the Park–Pham theorem. This started off as a conjecture due to Kahn and Kalai from the mid-late 2000s (called the Kahn–Kalai conjecture); then the fractional version of this conjecture was proven by Frankston, Kahn, Narayanan, and Park, and then the full version was proven by Park and Pham (both in the last five years). And this gives a much shorter way to derive these thresholds (compared to the original proofs, e.g. by Johansson–Kahn–Vu).

We're only going to discuss a special case of the theorem. Let $\mathcal{F}$ be a set of n-vertex k-graphs (for example, you can think of $\mathcal{F}$ as the set of perfect matchings). We can think of any n-vertex k-graph as a subset of $E(K_n^{(k)})$, where $K_n^{(k)}$ is the complete k-graph on n vertices and $E(K_n^{(k)})$ is the set of its edges; then we can think of $\mathcal{F}$ as a subset of $\mathcal{P}(E(K_n^{(k)}))$, which is the set of all n-vertex k-graphs.



Definition 2.4. We say a probability measure $\mathbb{P}$ on $\mathcal{F}$ is q-spread if for $M \in \mathcal{F}$ sampled according to $\mathbb{P}$, for all $S \subseteq E(K_n^{(k)})$ we have $\mathbb{P}[M \supseteq S] \leq q^{|S|}$.

So we're sampling M randomly from $\mathcal{F}$ according to our distribution $\mathbb{P}$ (e.g., if $\mathcal{F}$ is the set of all perfect matchings, then M is a random perfect matching). And we're fixing some set of edges S and looking at the probability that our random M contains this set of edges (e.g., that a given set of edges is in our perfect matching); and the spreadness condition means that this probability shouldn't be too large relative to $|S|$ (e.g., no specific set of edges should be too likely to be in our random perfect matching).

And a special case of the Park–Pham theorem and the fractional version by Frankston–Kahn–Narayanan–Park is the following.

Theorem 2.5

There is an absolute constant K such that if $\mathcal{F} \subseteq \mathcal{P}(E(K_n^{(k)}))$ supports a q-spread measure and $|F| \leq \ell$ for all $F \in \mathcal{F}$, then for $p \geq Kq \log \ell$, the random k-graph $\mathcal{G}_{n,p}^{(k)}$ will asymptotically almost surely contain some $F \in \mathcal{F}$ as a sub-hypergraph.

So we assume that $\mathcal{F}$ supports a q-spread measure and that p is a bit larger than q (specifically, a log factor larger, depending on the size of the elements of $\mathcal{F}$). And then we get that $\mathcal{G}_{n,p}^{(k)}$ contains some element of $\mathcal{F}$ (e.g., if $\mathcal{F}$ is the set of perfect matchings, then this means $\mathcal{G}_{n,p}^{(k)}$ has a perfect matching).

§2.3 The threshold for Theorem 1.8

We'll first use this to prove the threshold upper bound for Theorem 1.8 (i.e., for when a random 3-uniform hypergraph contains a perfect matching). We'll actually determine this threshold for *any* (constant) uniformity k — the situation in Theorem 1.8 is the special case $k = 3$.

Here we'll take $\mathcal{F}$ to be the set of all perfect matchings in $K_n^{(k)}$, and $\ell = n/k$ (since any perfect matching is a set of n/k edges).

Claim 2.6 — The uniform distribution on $\mathcal{F}$ is $O(n^{-(k-1)})$-spread.

Proof. Fix a set $S = \{e_1, \dots, e_s\} \subseteq E(K_n^{(k)})$ consisting of s edges, and imagine we choose a perfect matching M uniformly at random; we want to upper-bound the probability that M contains all these s edges, i.e.,

$$\mathbb{P}[M \supseteq S] = \frac{\#(\text{perfect matchings containing } \{e_1, \dots, e_s\})}{\#(\text{perfect matchings})}.$$

And both the numerator and denominator are things we can count easily — the denominator is

$$\#(\text{perfect matchings}) = \binom{n-1}{k-1}\binom{n-k-1}{k-1}\cdots\binom{k-1}{k-1}$$

(since there's $\binom{n-1}{k-1}$ ways to choose the remaining $k-1$ vertices of the edge containing the first vertex; then once we remove these k vertices, there's $\binom{n-k-1}{k-1}$ ways to choose the remaining vertices of the edge containing the next remaining vertex; and so on). Similarly, if $e_1, \dots, e_s$ are vertex-disjoint, then

$$\#(\text{perfect matchings containing } \{e_1, \dots, e_s\}) = \binom{n-sk-1}{k-1}\binom{n-(s+1)k-1}{k-1}\cdots\binom{k-1}{k-1}.$$

(If they're not vertex-disjoint, then this number is 0.) This means

$$\mathbb{P}[M \supseteq S] = \frac{1}{\binom{n-1}{k-1}\binom{n-k-1}{k-1}\cdots\binom{n-(s-1)k-1}{k-1}},$$

which means $\mathbb{P}[M \supseteq S]^{1/s} = O(n^{-(k-1)})$ for all s (where the implicit constant depends on k, but not s). (Very roughly, the point is that there's s terms in the denominator, and each is something like n^{k-1}.) □

Then applying the Park–Pham theorem immediately gives Theorem 1.8 in the case where G is complete (and a similar statement for higher uniformities), which we'll state explicitly as a corollary.

Corollary 2.7

For each $k \in \mathbb{N}$, there exists C such that if $p \geq C(\log n)n^{-(k-1)}$, then $\mathcal{G}_{n,p}^{(k)}$ asymptotically almost surely contains a perfect matching.

§2.4 Some canonical applications of the Park–Pham theorem

The Park–Pham theorem (Theorem 2.5) involves a log factor, so lots of thresholds which have a log in them can be proven using it. But we chose Theorems 1.5 and 1.6 as examples of thresholds that *don't* have a log — Theorem 1.5 has a fractional power of a log, and Theorem 1.6 doesn't have a log at all. So to prove them, we'll have to go beyond basic applications of the Park–Pham theorem — we'll have to do something more complicated to get such values of p.

But before we do so, we'll briefly mention a few canonical applications of the Park–Pham theorem.

Definition 2.8. For a graph (or hypergraph) F, its 0-density (denoted $m(F)$) and 1-density (denoted $m_1(F)$) are defined as

$$m(F) = \max_{F' \subseteq F} \frac{e(F')}{v(F')} \quad \text{and} \quad m_1(F) = \max_{F' \subseteq F} \frac{e(F')}{v(F') - 1}.$$

The first application we'll mention is a generalization of Corollary 2.7.

Fact 2.9 — If F is a k-graph with $v(F) = n$ and $\Delta_1(F) = O(1)$, then the uniform distribution on the set of copies of F in $K_n^{(k)}$ is $O(n^{-1/m_1(F)})$-spread.

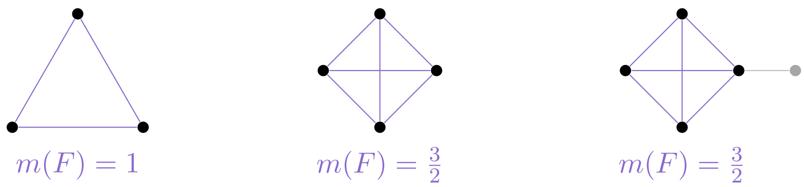
We use $\Delta_1(F)$ to denote the maximum number of edges that a single vertex is contained in, i.e., $\Delta_1(F) = \max_u \#(\text{edges containing } u)$. For example, if we take F to be a perfect matching, then this minimum degree condition is satisfied as $\Delta_1(F) = 1$, and we have $m_1(F) = 1/(k-1)$ (the maximum comes from taking F' to be a single edge), so this recovers Claim 2.6.

Then if F satisfies this maximum degree condition, we can use the Park–Pham theorem to get an upper bound on the threshold for containing a copy of F. This actually resolves two of the main problems that motivated the Kahn–Kalai conjecture in the first place — one was Shamir's problem (on the threshold for containing a perfect matching). The other, solved by Montgomery, was a problem about the threshold for $\mathcal{G}_{n,p}$ to have a bounded-degree spanning tree. If we take F to be any *fixed* bounded-degree spanning tree, then we have $m_1(F) = 1$ (any subgraph F' of F is a forest and therefore has $e(F') \le v(F') - 1$), so this fact combined with the Park–Pham theorem gives a threshold upper bound of $(\log n)/n$. (And we can get a matching lower bound by considering isolated vertices.)

The next application we'll consider is when F is a fixed graph (or hypergraph).

Fact 2.10 — If $v(F) = O(1)$, the uniform distribution on copies of F in $K_n^{(k)}$ is $O(n^{-1/m(F)})$-spread.

This can be used to get the threshold for $\mathcal{G}_{n,p}$ to contain any fixed graph — for example, a single triangle (where $m(F) = 1$), a K_4 (where $m(F) = \frac{3}{2}$), or a K_4 with a pendant edge (where $m(F)$ is still $\frac{3}{2}$, taking F' to be the K_4). Note that here there isn't an extra log factor because ℓ is a constant (F has a constant number of vertices and therefore edges), so the threshold we get is just $n^{-1/m(F)}$. (We get a matching lower bound by considering the expected number of copies of F'.)



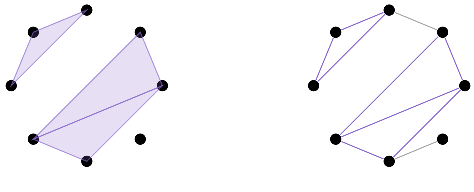
(The threshold when F is a fixed graph can be computed by simpler methods — e.g., the second moment method — but the point is that the Park–Pham theorem covers this case as well.)

§2.5 Coupling and the threshold for Theorem 1.5

To prove the threshold version of Theorem 1.5, we'll need a coupling due to Riordan (we're only going to state a special case of this coupling), which relates triangles in a random graph to edges in a random 3-graph.

Theorem 2.11 (Riordan)

If $p \le (\log n)^2 n^{-2/3}$, then we can couple $\mathcal{G}_{n,p}$ with $\mathcal{G}_{n,\pi}^{(3)}$ for some $\pi \asymp p^3$ such that asymptotically almost surely, every edge in $\mathcal{G}_{n,\pi}^{(3)}$ is an edge in $\mathcal{G}_{n,p}$.



Then combining this coupling together with the threshold for 3-uniform perfect matchings (as in Theorem 1.8) gives the threshold upper bound for Theorem 1.5 — the value of $p \asymp (\log n)^{1/3} n^{-2/3}$ in Theorem 1.5 satisfies this hypothesis, and $p^3 \asymp (\log n) n^{-2}$ is exactly the value in Theorem 1.8.

Remark 2.12. Kahn proved a *sharp* threshold for perfect matchings in k-graphs (as in Theorem 1.8), and this together with coupling can be used to get a sharp threshold for k-clique factors. But we don't know how to get a sharp threshold for clique factors without coupling.

§2.6 A smoother version of spread

We've now seen how it's possible to get a fractional power of log in our threshold (as in Theorem 1.5); but what about getting rid of the log entirely (as in Theorem 1.6)? For this we'll need a more complicated notion of spread, due to Sam Spiro.

Definition 2.13. Let $\mathcal{F} \subseteq \mathcal{P}(E(K_n^{(k)}))$, and let $r_1, \ldots, r_{\ell+1} \in \mathbb{N}$ be a decreasing sequence. We say a probability distribution $\mathbb{P}$ on $\mathcal{F}$ is $(q\colon r_1, \ldots, r_{\ell+1})$-spread if for M sampled according to $\mathbb{P}$, we have

$$\mathbb{P}[|M \cap S| = t] \le q^t$$

for all $S \subseteq E(K_n^{(k)})$ with $|S| \in [r_{i+1}, r_i]$ and all $t \in [r_{i+1}, r_i]$.

In the original notion of spread, we looked just at the probability that M *contained* S; now we're looking at the probability that M intersects S in a certain number of edges. So we've got a bunch of windows $[r_{i+1}, r_i]$, and instead of only trying to control the probability M *fully* contains S, we're trying to more generally control the probability that M intersects a *large part* of S (where the meaning of 'large part' is determined by the size of the window — we want $|S|$ and the intersection size to lie in the same window).

First, this version of spread is related to the original one in the following way.

Proposition 2.14

Suppose that $|F| \le r_1$ for all $F \in \mathcal{F}$.

(a) If $\mathbb{P}$ is $(q\colon r_1, \ldots, r_{\ell+1})$-spread with $r_{\ell+1} = 1$, then $\mathbb{P}$ is also q-spread.

(b) If $\mathbb{P}$ is q-spread and $\frac{1}{2} r_i < r_{i+1} < r_i$ for each i, then $\mathbb{P}$ is also $(4q\colon r_1, \ldots, r_{\ell+1})$-spread.

The first direction is immediate (taking t to be $|S|$ in this 'smoother' notion of spread; the conditions on r_1 and $r_{\ell+1}$ ensure that our windows together cover all possible sizes of S). The more interesting direction is the second, which states that the original notion of spread implies this smoother one if we choose our windows dyadically (so the sequence decreases by $\frac{1}{2}$ each step) — in other words, it doesn't really matter whether we think about containing a set S of size t or intersecting a set S of size at most $2t$ in t elements. (The proof is that if $|S| \le 2t$, then there are at most $\binom{2t}{t} \le 4^t$ subsets of S with size t, and the original notion of spread gives that M contains each of these subsets with probability at most q^t; so by a union bound, the probability it contains any of them (which must occur for $|M \cap S|$ to be t) is at most $4^t q^t$.)

And Spiro proved a strengthening of the theorem we saw earlier which replaces the original notion of spread with this smoother notion.

Theorem 2.15 (Spiro)

Suppose that $\mathcal{F} \subseteq \mathcal{P}(E(K_n^{(k)}))$ supports a $(q\colon r_1, \ldots, r_{\ell+1})$-spread measure where $|F| \le r_1$ for all $F \in \mathcal{F}$ and $r_{\ell+1} = 1$. Then for $p \gg q\ell$, $\mathcal{G}_{n,p}^{(k)}$ asymptotically almost surely contains some $F \in \mathcal{F}$.

This, combined with Proposition 2.14(b), gives the special case of the Park–Pham theorem that we stated (in Theorem 2.5) — if $\mathcal{F}$ supports a q-spread measure and $|F| \le r_1$ for all $F \in \mathcal{F}$, then the same measure satisfies this smoother version of spread with $\ell \approx \log r_1$. But the point is that if we can get better control — if we can get this smoother version of spread with super-dyadic windows — then we can get a better bound (in particular, we can improve or get rid of the log factor).

Remark 2.16. Spiro's proof uses the same idea as the Frankston–Kahn–Narayanan–Park proof of the fractional Kahn–Kalai conjecture — you reveal a bit at a time and track the sizes of the fragments, going for ℓ steps. Spiro's theorem combined with Proposition 2.14(b) implies the Frankston–Kahn–Narayanan–Park theorem; however, it *doesn't* imply the general Park–Pham theorem (it implies the version we stated, but we haven't stated the full theorem). So there's the remaining question of whether there's a common generalization of both of them.

§2.7 The threshold for Theorem 1.6

The threshold for $\mathcal{G}_{n,p}$ to contain the square of a Hamiltonian cycle was proven by Kahn, Narayanan, and Park around five years ago; this was before Spiro's result, and their result influenced Spiro's. But now we can think about their proof in terms of Spiro's framework — their proof ends up showing that the uniform distribution on squares of Hamiltonian cycles in K_n is $(C_1 n^{-1/2} \colon 2n, C_2 n^{1/2}, 1)$-spread (for some constants C_1 and C_2). (This proof is a couple of pages.)

And once you know that this is the right thing to prove, with Spiro's theorem this gives that the threshold for containing the square of a Hamiltonian cycle is $n^{-1/2}$ (here there's no log because there's only 3 terms in the sequence, which is a constant).

§3 Proving the general theorems

We've now discussed how to prove the threshold versions of all three theorems (i.e., the case where G is complete). But what about the general case?

We're still going to prove these theorems by constructing spread distributions; but now constructing these distributions is harder (we can't just take the uniform distribution anymore). To prove Theorem 1.8, we first prove the following statement (on the existence of a spread distribution).

Theorem 3.1

If G is an n-vertex 3-graph with $\delta_2(G) \ge (\frac{1}{2} + \varepsilon)n$, then there exists a (C/n^2)-spread distribution on the perfect matchings of G.

(Here the quantifiers are 'for every $\varepsilon > 0$, there exists C.')

There's at least three different proofs of Theorem 3.1 (we'll discuss one later). In fact, Theorem 3.1 is true even without the ε; then you can't use the non-random case as a black box, but you can use similar ideas (with an iterative absorption strategy — you sort of use the same ideas as the original proof on a very small random subset). If we allow the ε, then we *can* use the non-random version as a black box (and we can even use the non-random version with an ε as well, which is slightly easier to prove). Both of these proofs use iterative absorption, though in different ways. And there's a recent paper that gives a different way to prove this (which we'll discuss).

Remark 3.2. We can sort of think of *counting* the number of perfect matchings as checking the spread condition when S is maximal. But handling smaller S can be challenging.

To prove Theorem 1.5, you can again use Riordan's coupling (we won't really discuss this).

And to prove Theorem 1.6, we first prove the following statement on the existence of a spread distribution.

Theorem 3.3

If $\delta(G) \geq (\frac{2}{3}+\varepsilon)n$, then there exists a $(C_1 n^{-1/2}\colon 2n, C_2 n^{1/2}, 1)$-spread distribution on squares of Hamiltonian cycles in G (for constants C_1 and C_2 potentially depending on ε).

§3.1 Vertex spreadness

One way to produce the spread distributions in Theorems 3.1 and 3.3 is using the notion of *vertex spreadness*, which was introduced in the Pham–Sah–Sawhney–Simkin paper. The notion of vertex spreadness considers random *embeddings* of one graph H into another graph G (chosen according to some distribution).

Definition 3.4. An embedding $\varphi\colon H \hookrightarrow G$ is an injective map $\varphi\colon V(H) \to V(G)$ that preserves adjacency.

All three problems we're considering are really embedding problems — in Theorem 1.6 we're trying to embed the square of a Hamiltonian cycle into our graph, in Theorem 1.5 we're trying to embed a collection of disjoint triangles, and in Theorem 1.8 we're trying to embed a collection of disjoint 3-edges into a 3-graph (the notion of embeddings makes sense for hypergraphs as well).

Definition 3.5. A probability distribution $\mathbb{P}$ on embeddings $\varphi\colon H \hookrightarrow G$ is q-vertex-spread if for all distinct $x_1, \ldots, x_s \in V(H)$ and $y_1, \ldots, y_s \in V(G)$, we have

$$\mathbb{P}[\varphi(x_i) = y_i \text{ for all } i] \leq q^s.$$

So we imagine that we've got some graph H, and we're trying to embed it randomly into G (according to some distribution). And we fix some s vertices $x_1, \ldots, x_s$ in H, and s target vertices $y_1, \ldots, y_s$ in G; and we consider the probability that all s vertices in H are sent to their corresponding targets — and spreadness means this probability should not be too large.

Remark 3.6. Similarly to edge spreadness, vertex spreadness also generalizes counting in some sense — if we've got a q-spread distribution, then taking $\{x_1, \ldots, x_s\}$ and $\{y_1, \ldots, y_s\}$ to be the entire vertex sets gives a lower bound on the total number of embeddings, i.e., the number of copies of H in G.

The notion of vertex spreadness was introduced for reasons related to bounded-degree spanning trees. But it turns out that when we're dealing with embeddings, it's the right notion to work with, moreso than edge spreadness. One illustration of this is the following proposition.

Proposition 3.7

For any k-graph H, the uniform distribution on embeddings $\varphi\colon H \hookrightarrow K_n^{(k)}$ is (e/n)-vertex-spread.

So if we're trying to prove statements about general graphs (instead of just complete ones), the right thing to do in some sense is to show that there *exists* some (C/n)-vertex-spread distribution — maybe the uniform distribution won't work anymore, but we still want to find *some* distribution with this property.

And a reason that vertex spreadness is useful is that the existence of a vertex-spread distribution implies the existence of an edge-spread one — combined with Proposition 3.7, this is sort of a generalization of Fact 2.9 (replacing the complete graph with an arbitrary one).

Proposition 3.8 (Kelly–Müyesser–Pokrovskiy)

For all constants C and Δ, there exists a constant C' such that if $\Delta(H) \leq \Delta$ and there exists a (C/n)-vertex-spread distribution on embeddings $\varphi\colon H \hookrightarrow G$, then there also exists a $C'n^{-1/m_1(H)}$-spread distribution on copies of H in G.

For bounded-degree spanning trees (which is the problem Park, Sah, Sawhney, and Simkin introduced vertex spreadness to study), the authors showed that there exists a (C/n)-vertex-spread distribution on embeddings, and then you can just do things vertex-by-vertex (since trees are nice in this way). But this proposition tells you that you can go from vertex spreadness to edge spreadness for *any* graph (with bounded degree — e.g., trees or matchings).

This means that for Theorem 1.8 (or rather, Theorem 3.1, the statement about edge spreadness that would imply it), it suffices to construct a *vertex*-spread distribution. For Theorem 1.6 (or rather, Theorem 3.3) we need a bit more, because we actually want the smoother notion of spread.

Proposition 3.9 (Kelly–Sah–Sawhney)

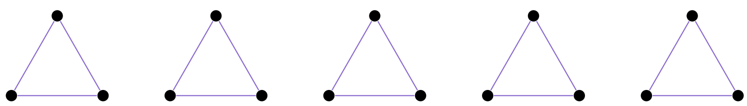
For any constant C, there exist constants C_1 and C_2 such that if there is a (C/n)-vertex-spread distribution on embeddings $\varphi\colon C_n^2 \hookrightarrow G$, then there is a $(C_1 n^{-1/2}\colon 2n, C_2 n^{1/2}, 1)$-spread distribution on copies of C_n^2 in G.

Here we use C_n to denote the n-vertex cycle and C_n^2 to denote its square, so we can think of a square of a Hamiltonian cycle in G as a copy of C_n^2 in G. (In particular, C_n^2 is not a constant, unlike C_1 and C_2.)

Then to prove Theorem 3.3, it really does suffice to construct a C/n-vertex-spread distribution on embeddings of C_n^2.

§3.2 A vertex-spread distribution for triangle factors

Finally, we'll discuss how to construct such C/n-vertex-spread embeddings. We'll focus on the case where the graph we're embedding is a triangle factor, which we'll write as $(n/3)K_3$ (this notation denotes the union of $n/3$ disjoint triangles); the ideas of the proof generalize to C_n^2 as well.



Theorem 3.10 (Kelly–Müyesser–Pokrovskiy)

If $\delta(G) \geq (\frac{2}{3} + \varepsilon)n$, there exists a (C/n)-vertex-spread distribution on embeddings $(n/3)K_3 \hookrightarrow G$.

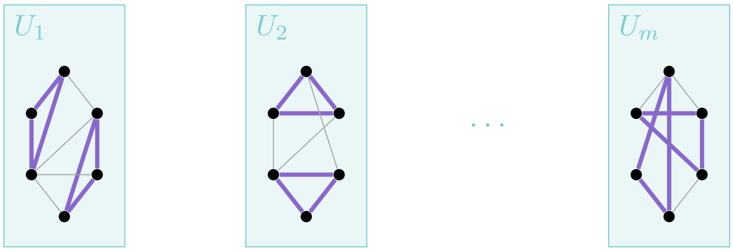
§3.2.1 A weaker version

First, we'll prove a weaker version (which is off by a log factor) as a warmup.

Proposition 3.11

If $\delta(G) \geq (\frac{2}{3} + \varepsilon)n$, there exists a $(C(\log n)/n)$-vertex-spread distribution on embeddings $(n/3)K_3 \hookrightarrow G$.

This isn't hard to prove, but it's a nice illustration of how we can prove things about spreadness. The idea is that we're first going to partition $V(G)$ into $m = n/(C\log n)$ parts $U_1, \ldots, U_m$, each of size $|U_i| \approx C\log n$, uniformly at random (for some constant C). And then we're going to arbitrarily choose a triangle factor from *each* part and put them together.



To show that this works, we first need to check that such a partition preserves the minimum degree property, so that we really can find a triangle factor in each part.

Claim 3.12 — We have $\mathbb{P}[\delta(G[U_i]) \geq (\frac{2}{3} + \frac{\varepsilon}{2})\,|U_i|\ \text{ for all } i] \geq \frac{1}{2}$.

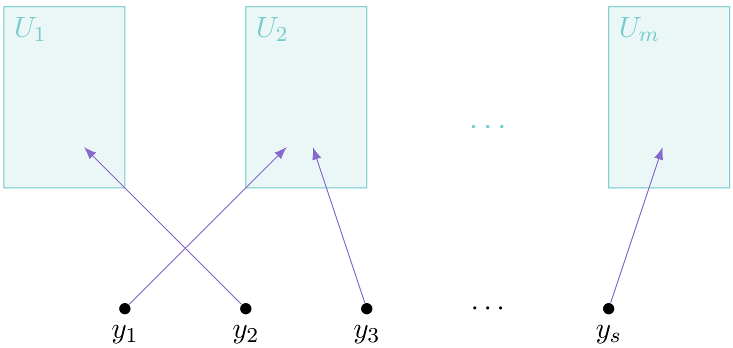
Proof. The proof is essentially just a Chernoff bound plus a union bound, so we won't go through the details. The point is that the expected degree of each vertex will be roughly at least $(\frac{2}{3} + \varepsilon)C\log n$ (since its degree in G is at least $(\frac{2}{3}+\varepsilon)n$, and each of its neighbors in G gets placed into its part with probability $(C\log n)/n$), and by a Chernoff bound this degree will be close to its expectation with high probability. And because this expectation is on the order of $\log n$ (and the Chernoff bounds give tail bounds with the expectation in the exponent), this probability is high enough that we can union bound over all n vertices. $\square$

Meanwhile, to show vertex spreadness, we'll need the following claim.

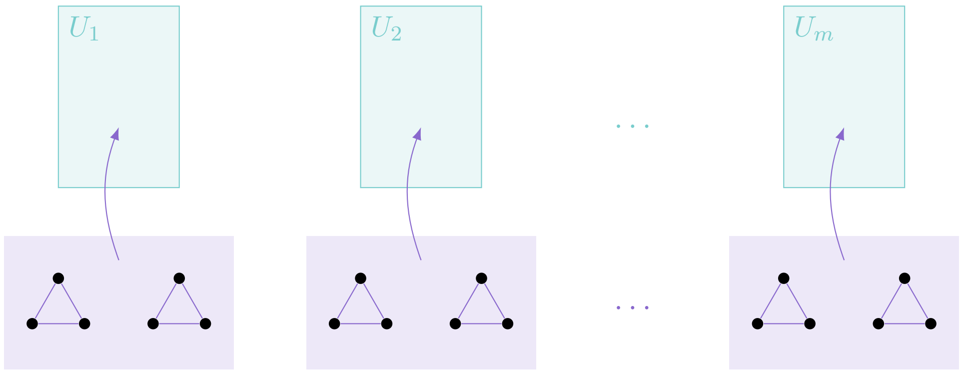
Claim 3.13 — For all $y_1, \ldots, y_s \in V(G)$ and $f\colon [s] \to [m]$, we have

$$\mathbb{P}[y_i \in U_{f(i)} \text{ for all } i] \leq \left(\frac{C\log n}{n}\right)^s.$$

In words, we're looking at our original graph G and fixing s vertices $y_1, \ldots, y_s$, and f represents a target allocation of the vertices $y_1, \ldots, y_s$ to the clusters $U_1, \ldots, U_m$. And we're claiming that the probability that each vertex y_i ends up in the cluster f says it should be in is small. (Proving this is a computation which we won't go through, but the intuition is that each y_i is equally likely to be placed in any of the $(C\log n)/n$ parts, and there's s of them. We might need to adjust C by a constant factor to handle the fact that they aren't exactly independent — e.g., we might want Ce in place of C here — but this doesn't really matter.)



Now we construct our vertex-spread distribution as follows — we first consider choosing a random partition of the vertices into parts $U_1, \ldots, U_m$ (as described earlier), *conditioning* on the outcome in Claim 3.12 (i.e., that $\delta(G[U_i]) \geq (\frac{2}{3} + \frac{\varepsilon}{2})\,|U_i|$ for all i). Then there exists an embedding of a triangle factor into $G[U_i]$ for each i (by the $p = 1$ case of Theorem 1.5), and we can combine these triangle factors to get an embedding of our entire triangle factor $(n/3)K_3$ into G. More precisely, we break the triangle factor $(n/3)K_3$ into m parts of size $C \log n$ (each consisting of $(C \log n)/3$ triangles), and we arbitrarily embed each of these chunks into the corresponding cluster U_i; and combining these gives an embedding of the entire triangle factor.

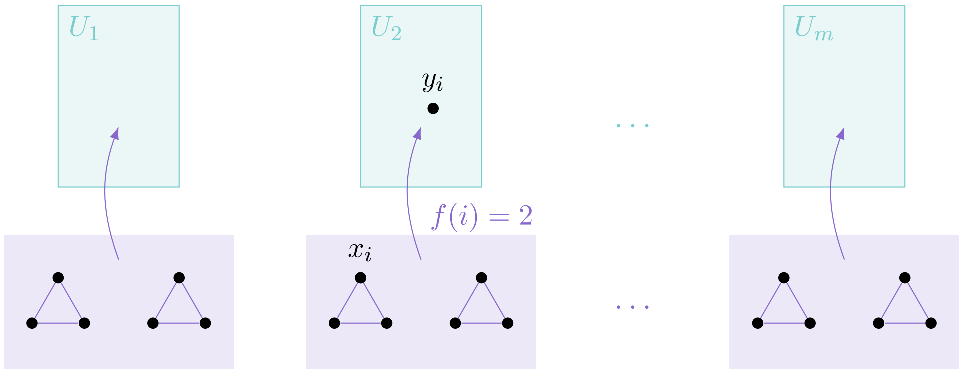


To show that this distribution is vertex-spread, suppose we fix vertices $x_1, \ldots, x_s$ of our triangle factor and $y_1, \ldots, y_s$ of G. We want to bound the probability that $\varphi(x_i) = y_i$ for each i — specifically,

$$\mathbb{P}[\varphi(x_i) = y_i \text{ for all } i \mid \text{Claim 3.12}]$$

(where we write the condititioning on Claim 3.12 as shorthand for conditioning on the outcome described in Claim 3.12 — i.e., that each part has reasonable minimum degree).

First, we can define the function $f\colon [s] \to [m]$ tracking which part U_j each x_i is embedded into. Then in order to have $y_i = \varphi(x_i)$, we must certainly have $y_i \in U_{f(i)}$.



So the probability we're considering is

$$\mathbb{P}[\varphi(x_i) = y_i \text{ for all } i \mid \text{Claim 3.12}] \leq \mathbb{P}[y_i \in U_{f(i)} \text{ for all } i \mid \text{Claim 3.12}].$$

And removing the conditioning costs us a factor of at most $\mathbb{P}[\text{Claim 3.12}]^{-1} \leq 2$, since in general we have

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \leq \frac{\mathbb{P}[A]}{\mathbb{P}[B]}.$$

So then the probability we're considering is at most

$$2\mathbb{P}[y_i \in U_{f(i)} \text{ for all } i] \leq 2\left(\frac{C \log n}{n}\right)^s \leq \left(\frac{2C \log n}{n}\right)^s$$

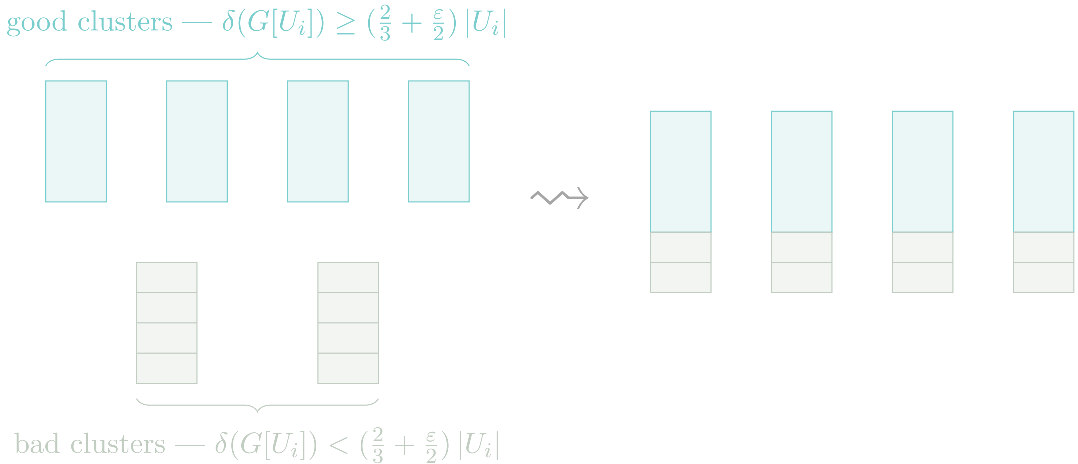
(using Claim 3.13). This proves Proposition 3.11, and all we really had to look at to prove spreadness was the probability each vertex went into the correct cluster.

§3.3 Proof idea for Theorem 3.10

Finally, we'll discuss how to remove the $\log n$ factor (in the spreadness) to prove Theorem 3.10.

The idea is that we'd *like* to run this same process, but partitioning into pieces U_i of *constant* (rather than logarithmic) size — this will correctly give $(C/n)^s$ rather than $((C\log n)/n)^s$ in Claim 3.13, which is what we want. But the issue is that Claim 3.12 won't hold — if we partition randomly into constant-sized pieces, then for each vertex v, the probability its degree is too small will be a *constant*, so even though this constant is tiny, we won't be able to union-bound over vertices.

However, we *can* say that with high probability, *most* of our clusters U_i will have high minimum degree. And the idea is that for the 'bad' clusters that *don't* have high enough minimum degree, we're going to take their vertices and redistribute them to the 'good' clusters — this only makes the good clusters slightly larger (since there's so few bad clusters).



So we're going to condition on a more complicated event that allows us to do this — that all but a small number of clusters satisfy the minimum degree condition *and* have a small number of vertices you can add to them while preserving the minimum degree condition (we'll call such clusters 'good' and the others 'bad'). Then we can take the bad clusters and redistribute their vertices to the good clusters, and then do the same thing as before (finding a triangle factor in each cluster). And we'll still get the key property in Claim 3.13, this time with C/n; this allows us to prove that the distribution is (C/n)-vertex-spread.