Bounded-degree spanning trees in randomly perturbed graphs

Michael Krivelevich, Matthew Kwan, Benny Sudakov

February 14, 2015

Abstract

abstract

We will prove the following theorem.

Theorem 1. *There is $c = c(\alpha, \Delta)$ such that if G is a graph on the vertex set $[n]$ with minimum degree at least αn, T is a tree on $[n]$ with maximum degree at most Δ, and $R \in \mathbb{G}(n, c/n)$, then a.a.s. $T \subseteq G \cup R$.*

A key ingredient of the proof is the following lemma: with the random edges in R alone, we can embed trees that are not too big.

Lemma 2 ([1, Theorem 1.1])**.** *There is $c = c(\varepsilon, \Delta)$ such that $G \in \mathbb{G}(n, c/n)$ a.a.s. contains every tree of maximum degree at most Δ on $(1 - \varepsilon)n$ vertices.*

We will split the proof of Theorem 1 into two cases in a similar way to [6]. If our spanning tree T has many leaves, then we remove the leaves and embed the resulting non-spanning tree in R (using Lemma 2). To complete this into our spanning tree T, it remains to match the the vertices needing leaves with leftover vertices. This amounts to finding a perfect matching in a certain bipartite graph. The random edges in R provide several almost-perfect matchings on their own; we combine these with the dense graph G to satisfy the conditions of Hall's marriage theorem and guarantee the required perfect matching.

The more difficult case is if T has few leaves. In this case T cannot be very "complicated" and must be a subdivision of a small tree. In particular T must have many long *bare paths*: paths where each vertex has degree exactly two. By removing some of these bare paths we can obtain a non-spanning forest. We would like to embed such a forest into R using Lemma 2, then complete it into the spanning tree T. This would involve joining up distinguished pairs of vertices with disjoint paths of certain lengths. In order to make this task feasible we first use Szemerédi's regularity lemma to divide the vertex set into a bounded number of pieces, each of which induces a "super-regular pair" in G (that is, a dense subgraph with edges very well-distributed). When we embed our forest, some of the paths we need to add will be between pieces, and some will be within pieces. We carefully choose which bare paths to remove, and choose how to embed the resulting forest, in such a way that very few such paths need to be between pieces. We first greedily embed this small number of paths between pieces. Then the super-regularity of the pieces allows us to use a tool called the blow-up lemma to finish joining up the paths, after making some adjustments using the random edges in R.

Now we proceed to the details of the proof. It is convenient to work in the model where R contains each edge with probability $cn/\binom{n}{2}$ independently. We split the random edges into multiple independent "phases": say $R \supseteq R_1 \cup R_2 \cup R_3 \cup R_4$, where $R_i \in \mathbb{G}(n, c_i/n)$ for some large $c_i = c_i(\alpha, \Delta)$ to be determined (these c_i will in turn determine c).

For our first case suppose there are at least λn leaves in T (for some $\lambda = \lambda(\alpha) > 0$ to be determined). Then consider the tree T' with some λn leaves removed. By Lemma 2 we can embed T' into R_1. Let A' be the set of vertices of T' which had a leaf deleted from them, and let B be the set of λn vertices not part of T'. We can assume A' and B are uniformly random disjoint sets of the appropriate size (note $|A'| \geq \lambda n/\Delta$ and $|B| = \lambda n$). For each $b \in B$, the number of neighbours in A' is hypergeometrically distributed with expected value at least $\lambda \alpha n/\Delta$. Let $\beta = \lambda\alpha/(2\Delta)$; by a concentration inequality (for example, see [2, Theorem 2.10]) and the union bound, a.a.s. each $a' \in A'$ has at least βn neighbours in B. Similarly each $b \in B$ a.a.s. has at least βn neighbours in A'. From now on we treat A' and B as fixed sets satisfying these properties.

For any graph H on $[n]$, we define a bipartite graph $\mathbb{B}(H)$ by throwing away all edges not between B and A', then making multiple copies of the vertices in A'. Specifically, let A be a set consisting of ℓ copies of each vertex $a' \in A'$ which needs ℓ extra leaves (so $|A| = \lambda n$). Let $\mathbb{B}(H)$ be the bipartite graph on the vertex set $A \cup B$ which has an edge between $a \in A$ (a copy of $a' \in A'$) and $b \in B$ if there is an edge between a' and b in $\mathbb{B}(H)$. If we can find a perfect matching in $\mathbb{B}(G) \cup \mathbb{B}(R_2) \subseteq \mathbb{B}(G \cup R)$, this gives us a way to extend our embedding of T' to an embedding of T, as desired.

Let $\mathbb{B}_{A,B}(p)$ be the random binomial graph where each of the $|A||B|$ possible edges between A and B are present with probability p. For any $c_{\mathbb{B}}$, we can choose large c_2 such that $1 - c_2/n \leq (1 - c_{\mathbb{B}}/n)^{\Delta}$ for large n. This means $\mathbb{B}(R_2)$ stochastically dominates $\mathbb{B}_{A,B}(c_{\mathbb{B}}/n)$, so the following lemma provides the desired perfect matching.

Lemma 3. *Let A and B be n-vertex sets. There is $c_{\mathbb{B}} = c_{\mathbb{B}}(\alpha)$ such that if G is a bipartite graph with bipartition $A \cup B$ and minimum degree at least αn, and if $R \in \mathbb{B}_{A,B}(c_{\mathbb{B}}/n)$, then a.a.s. $G \cup R$ has a perfect matching.*

In order to prove Lemma 3 we need another lemma.

Lemma 4. *Let $\varepsilon > 0$, and let A and B be n-vertex sets. There is $c = c(\varepsilon)$ such that $R \in \mathbb{B}_{A,B}(c/n)$ has a matching of size $(1-\varepsilon)n$.*

Proof. For any subsets A' and B' of size εn, the probability there is no edge between A' and B' is $(1-p)^{(\varepsilon n)^2} \leq e^{-c\varepsilon^2 n}$. There are at most 2^{2n} choices of such A', B', so for large c, by the union bound there is a.a.s. an edge between any such pair of sets. It follows that a maximum-size matching a.a.s. has at least $(1-\varepsilon)n$ edges. □

Proof of Lemma 3. We apply Hall's marriage theorem. If $W \subseteq A$ with $|W| \leq \alpha n$ then $|N_G(W)| \geq \alpha n \geq |W|$ by the minimum degree condition on G. Similarly, every vertex outside $N_G(W)$ has at least αn neighbours outside W, so if $|W| \geq (1-\alpha)n$ then $|N_G(W)| = n \geq |W|$. So, consider any W with $\alpha n \leq |W| \leq (1-\alpha)n$. Split R into phases $R \supseteq R_1 \cup \dots \cup R_r$, with $R_i \in \mathbb{B}_{A,B}(c'/n)$, for some $c' = c'(\alpha)$ and $r = r(\alpha)$ to be determined. By Lemma 4, if c' is large we can find a $(1-\alpha/2)n$-edge matching M_i in each R_i. Condition on the vertices $W_i := W \cap M_i$ matched by each M_i. The distribution of R_1 is invariant under permutations of B, so we can assume each $N_{M_i}(W)$ is an independently and uniformly random $|W_i|$-element subset of B. We will use this kind of argument implicitly throughout this paper, to assume that objects embedded in random graphs are themselves random.

Is this OK, or do you think I should explain these symmetry arguments more?

For each $b \in B$, let $X_{b,i}$ be the indicator random variable for the event that $b \notin N_{M_i}(W)$. The random variables $X_{b,i}$ are said to be *negatively associated* (see [3], in particular property P_7 and Application 3.1(c)). Hence, the random variables $X_b = \min_i X_{b,i}$ (which are indicators for the events $b \notin N_G(W)$) are also negatively associated. This means in particular that we can apply Chernoff bounds even though the X_b are not independent. We have

$$\mathbb{E}\sum_{b\in B} X_b = n\prod_{i=1}^{r}((n-|W_i|)/n) \le (1-\alpha/2)^r n,$$

so by the Chernoff bound,

$$\Pr(|N(W)|<|W|) = \Pr\left(\sum_{b\in B} X_b > n-|W|\right) \le \Pr\left(\sum_{b\in B} X_b > \alpha n\right) = \exp\bigl(-\Omega\bigl((1-\alpha/2)^{-r}\alpha^2\bigr)\bigr).$$

since there are at most 2^n choices for W, for large r the union bound gives $|N_R(W)| \ge |W|$ for all required W. □

Now we address the second case where there are fewer than λn leaves in T. As outlined, we first need to divide the vertex set into groups such that the edges of G are well distributed within those groups.

Definition 5. For a disjoint pair of vertex sets (X,Y) in a graph, let its *density* $d(X,Y)$ be the number of edges between X and Y, divided by $|X||Y|$. A pair of vertex sets (V_1,V_2) is said to be *ε-regular* if for any U_1,U_2 with $U_h \subseteq V_h$ and $|U_h| \ge \varepsilon|V_h|$, we have $|d(U_1,U_2)-d(V_1,V_2)| \le \varepsilon$. If alternatively $d(U_1,U_2) \ge \delta$ for all such pairs U_1,U_2 then we say (V_1,V_2) is *(ε,δ)-semi-super-regular*. Let $\bar{h} = 2-h$; if (V_1,V_2) is (ε,δ)-semi-super-regular and moreover each $v \in V_h$ has at least $\delta|V_{\bar{h}}|$ neighbours in $V_{\bar{h}}$, then we say (V_1,V_2) is *(ε,δ)-super-regular*.

Is there already some word to describe what I'm calling "semi-super-regularity"? It is basically a one-sided version of regularity.

Lemma 6. *For $\alpha,\varepsilon > 0$ with ε sufficiently small relative to α, there are $\delta = \delta(\alpha) > 0$, $M = M(\alpha,\varepsilon)$, $a = a(\alpha,\varepsilon) > 0$ and $b = b(\alpha,\varepsilon) > 0$ such that the following holds. Let G be a graph on $[n]$ with minimum degree at least αn. We can choose $q \le M$ and partition $[n]$ into sets V_i^h ($1 \le i \le q$, $h = 1,2$) such that each $an \le \left|V_i^h\right| \le bn$ and each pair $\left(V_i^1,V_i^2\right)$ is (ε,δ)-super-regular.*

To prove Lemma 6, we apply Szemerédi's regularity lemma to obtain a reduced *cluster graph*, then decompose this cluster graph into small stars with the following auxiliary lemma, each of which will roughly correspond to a pair $\left(V_i^1,V_i^2\right)$. We will then redistribute some of the vertices to ensure super-regularity.

Lemma 7. *Let G be an n-vertex graph with minimum degree at least αn. Then there is a spanning subgraph Q which is a disjoint union of vertex-disjoint stars, each with at least two vertices and at most $1/\alpha$ leaves.*

Proof. Let Q be a maximal-size union of such stars. Suppose there is a vertex v uncovered by Q. If v has a neighbour which is a centre of one of the stars in Q, and that star has fewer than $1/\alpha$ leaves, then we could add v to that star, contradicting maximality. Otherwise, if v has a neighbour w which is a leaf of one of the stars in Q, then we could remove that leaf from its star and create a new 2-vertex star with edge vw, again contradicting maximality. The remaining case is where each of the (at least αn) neighbours of v is a centre of a star with $1/\alpha$ leaves. But these stars would comprise more than n vertices, which is again a contradiction. We conclude that Q covers G, as desired. □

Proof of Lemma 6. We apply the "degree form" of Szemerédi's regularity lemma ([4, Theorem 1.10]) with density threshold $\alpha' = \alpha/3$ and some regularity parameter ε', which we will repeatedly assume is very small relative to α'. There is large $M = M(\alpha)$, a partition of $[n]$ into clusters $V_0, \dots, V_k$ ($k \le M$) and a subgraph $G' \subseteq G$ satisfying the following conditions. The "exceptional cluster" V_0 has size at most $\varepsilon' n$, and the other clusters have equal size sn (so $1-\varepsilon' \le ks \le 1$). The minimum degree of G' is at least $(\alpha' + \varepsilon')n$. There are no edges of G' within clusters, and each pair of non-exceptional clusters is ε'-regular in G' with density at least α'.

Define the cluster graph C as the bipartite graph whose vertices are the non-exceptional clusters V_i, and whose edges are the pairs of clusters between which there is nonzero density in G'. The fact that G is dense implies that C is dense as well, as follows. In G' each V_i has at least $((\alpha' + \varepsilon')n)(sn)$ edges to other clusters. There are at most $(\varepsilon' n)(sn)$ edges to the exceptional cluster V_0 and at most $(sn)^2$ edges to each other cluster. So, $d_C(V_i) \ge ((\alpha' + \varepsilon)n - \varepsilon' n)sn/(sn)^2 \ge \alpha' k$. That is, C has minimum degree at least $\alpha' k$.

Now, let Q be a star cover of C by stars $T_1, \dots, T_q$ of size at most $1/\alpha'$, as guaranteed by Lemma 7. Let W_i^1 be the centre cluster of T_i, and let W_i^2 be the union of the leaf clusters of T_i (for two-vertex stars, arbitrarily choose one vertex as the "leaf" and one as the "centre"). Consider any $U^h \subseteq W_i^h$ with $\left|U^h\right| \ge \varepsilon' W_i^h$, and let ℓ be the number of leaves in T_i. Let V_p be the leaf of T_i with the most vertices in U^2, so that $\left|U^2 \cap V_p\right| \ge \left|U^2\right|/\ell \ge \varepsilon' sn$. By ε'-regularity there are at least $(\alpha' - \varepsilon')\left|U^1\right|\left(\left|U^2\right|/\ell\right) \ge \left((\alpha')^2/2\right)\left|U^1\right|\left|U^2\right|$ edges between $U^2 \cap V_p$ and U^1. That is, $d\left(U^1, U^2\right) \ge (\alpha')^2/2$, so with $\delta' = (\alpha')^2/2$ each pair $\left(W_i^1, W_i^2\right)$ is (ε', δ')-semi-super-regular.

There are still some vertices left over from exceptional clusters, and we cannot guarantee that the $\left(W_i^1, W_i^2\right)$ are super-regular (there may be vertices without enough neighbours in their "partner" cluster). We now make some adjustments to fix these problems.

If D_i^h is the set of vertices of W_i^h with degree less than $\delta'\left|W_i^{\bar{h}}\right|$, then there are fewer than $\delta'\left|D_i^h\right|\left|W_i^{\bar{h}}\right|$ edges between D_i^h and $W_i^{\bar{h}}$. So, we must have $\left|D_i^h\right| < \varepsilon'\left|W_i^h\right|$, by semi-super-regularity. Let $S_i^h = W_i^h \backslash D_i^h$. Note that each $v \in S_i^h$ has at least $(1-\varepsilon)\delta'\left|S_i^h\right| \ge \delta'\left|S_i^h\right|/2$ neighbours in its partner cluster $S_i^{\bar{h}}$. Now, let V_i^h correspond to a "maximal extension" of the clusters S_i^h, as follows. Each $V_i^h \supseteq S_i^h$, each V_i^h is disjoint, each $v \in V_i^h$ has at least $\delta'\left|S_i^{\bar{h}}\right|/2$ neighbours in its partner cluster V_i^{2-h}, each $\left|V_i^h \backslash S_i^h\right| \le (4\varepsilon'/\alpha')\left|V_i^h\right|$, and none of the V_i^h can be enlarged retaining these properties.

Let $P = V_0 \cup \bigcup_{i=1}^q \left(D_i^1 \cup D_i^2\right)$ be the set of at most $2\varepsilon' n$ vertices not in any S_i^h. If $\left|V_i^h \backslash S_i^h\right| = (4\varepsilon'/\alpha')\left|V_i^h\right|$ then we say V_i^h is "full". There can be at most $2\varepsilon' n/(4\varepsilon'/\alpha') = \alpha' n/2$ vertices in full clusters. Now suppose there is some v not in any V_i^h. That v has $\alpha' n/2$ neighbours in non-full clusters, so some non-full cluster V_i^h has $\alpha'\left|V_i^h\right|/2 \ge \delta'\left|S_i^h\right|/2$ of these neighbours. But this is a contradiction, because we could add v to $V_i^{\bar{h}}$ and it would still satisfy all the required properties, contradicting maximality.

We conclude that the V_i^h partition the vertex set. If $a = s(1-\varepsilon')$ and $b = s/(\alpha'(1-4\varepsilon'/\alpha'))$, then each $an \le |V_i^h| \le bn$. For each i, if $U^h \subseteq V_i^h$ with $|U^h| \ge (4/\alpha'+1)\varepsilon'|V_i^h|$, then $|U^h \cap W_i^h| \ge \varepsilon'|W_i^h|$ so

$$d(U^1, U^2) \ge d(U^1 \cap W_i^1, U^2 \cap W_i^2)/(4/\alpha'+1)^2 \ge \delta'/(4/\alpha'+1)^2.$$

That is, each pair (V_i^1, V_i^2) is $((4/\alpha'+1)\varepsilon', \delta)$-semi-super-regular, for $\delta = \delta'/(4/\alpha'+1)^2$. Finally, each $v \in V_i^h$ has at least $\delta'|S_i^{\bar{h}}|/2 \ge \delta'|V_i^{\bar{h}}|/(2(4\varepsilon'/\alpha'+1)) \ge \delta|V_i^h|$ neighbours in its partner cluster V_i^h, so in fact each pair (V_i^1, V_i^2) is $((4/\alpha'+1)\varepsilon', \delta)$-super-regular. With ε' small enough so that $(4/\alpha'+1)\varepsilon' \le \varepsilon$, we are done. □

We apply Lemma 6 to our graph G, with $\varepsilon = \varepsilon(\delta)$ to be determined. Proceeding with the proof outline, we need the fact that T is almost entirely composed of bare paths, as is guaranteed by the following lemma.

Lemma 8 ([6, Lemma 2.1])**.** *Let T be a tree on n vertices with at most ℓ leaves. Then T contains a collection of at least $(n-(2\ell-2)(k+1))/(k+1)$ vertex-disjoint bare paths of length k each.*

In our case $\ell = \lambda n$. For any $\phi = \phi(\alpha) > 0$ (which we will determine later), if we choose $k = k(\phi)$ large enough and $\lambda = \lambda(k)$ small enough, then T contains a collection of $(1-\phi)n/(k-1)$ disjoint bare k-paths. If we delete the interior vertices of these paths, then we are left with a forest F on at most ϕn vertices. We define a "reduced" tree $r(T)$ whose vertices are the trees in F, with an edge between two of these vertices if the corresponding trees were originally joined by a bare path. Note that $r(T)$ has $(1-\phi)n/(k-1)$ edges and therefore $(1-\phi)n/(k-1)+1$ vertices.

Now, every subforest of $r(T)$ corresponds to a subforest of T, obtained by "expanding" vertices back into trees and edges back into bare paths. We want to choose an appropriate subforest $r(F') \subseteq r(T)$, and embed the corresponding subforest $F' \subseteq T$ into R_2. We want to do both these things in such a way that most of the bare paths in $T \backslash F'$ have both their endpoints in the same cluster V_i^h. The way we will accomplish our goal depends on the structure of $r(T)$, as follows.

Lemma 9. *For $\tau > 0$ and $q \in \mathbb{N}$, there is $\gamma = \gamma(\tau, q)$ such that one of the following holds for any tree T on n vertices (where n is sufficiently large).*

(a) T has at least γn leaves,

(b) There is a spanning subforest $F = \bigcup_{i=1}^q F_i$ of T, such that the F_i are disjoint subforests each with at least γn edges, and such that there are at most τn edges in T that are not in F.

Proof. We use Lemma 8. For sufficiently small γ, either $r(T)$ has γn leaves or it has γn disjoint bare paths of length $2(q-1)$. In the second case, divide some γn such bare paths evenly into $q-1$ forests of paths $F_1, \dots, F_{q-1}$ (we can assume $q > 1$, or else (b) trivially holds). Then, let $F_q = T \backslash \bigcup_{i=1}^{q-1} F_i$. For small γ, there are at most $2\gamma n \le \tau n$ edges of T not in some F_i, each F_i ($i \le q-1$)has at least $(2(q-1)-1)\gamma n/(q-1) \ge \gamma n$ edges, and F_q has $n - ((q-1)/\tau + 2)\gamma n \ge \gamma n$ edges. □

It seems pretty ugly that I apply Lemma 8 twice... I spent some time thinking about whether this can really be reduced to one application but I couldn't seem to do it.

We apply Lemma 9 to $r(T)$ with some $\tau = \tau(\alpha) > 0$ to be determined. In case (a), $r(T)$ has at least $\gamma((1-\phi)n/(k-1)+1) \geq \gamma n/(2k)$ leaves. Let $r(T') \subseteq r(T)$ be a subtree obtained by deleting $\gamma n/(2k)$ leaves. Then T' has at most $(1-\gamma/2)n$ vertices (it is "missing" at least k vertices from each bare path associated with our $\gamma n/(2k)$ leaves of $r(T)$). For large c_1, we can a.a.s. embed T' into R_1; for such an embedding, let X be the set of $\gamma n/(2k)$ vertices of T' that need a bare path attached to them. By the permutation-invariance of the distribution of R_1, we can assume the vertex set of the embedding of T' is a uniformly random set of size $|T'|$, and X is a uniformly random $\gamma n/(2k)$-vertex subset. By a concentration inequality for the hypergeometric distribution, a.a.s. each V_i^h has at least $a\gamma n/(3k)$ and at most $b\gamma n/k$ elements of X, and there are at least $a\gamma n/3$ vertices in each V_i^h that were not used for T'. For each $x \in X$, let $T_x \subseteq F$ be the tree that needs to be connected to x via a k-path. For each V_i^h, embed $\bigcup_{x\in V_i \cap X} T_x$ randomly into $R_2\left[V_i^h \backslash T'\right]$. In order to ensure this is a.a.s. possible and that there are still at least $a\gamma/3$ vertices left over, choose $\phi = a\gamma/6$ (we also need c_2 to be large). Now, all that remains is to find k-paths that connect each $x \in X$ to some specific vertex y (let Y be the set of all such y). Call the pairs (x, y) "special pairs", and call a k-path connecting a special pair a "special path". By construction, all special pairs (x, y) have x and y in the same V_i^h.

In case (b), we have a decomposition of the vertex set of $r(T)$ into $2q$ forests $r\left(F_i^h\right)$. Each of these forests $r\left(F_i^h\right)$ corresponds to a forest $F_i^h = \bigcup V\left(r\left(F_i^h\right)\right) \subseteq F$ (that is, we expand the vertices of $r\left(F_i^h\right)$ to trees but throw away the edges of $r\left(F_i^h\right)$). With large c_1 and $\phi < a/2$, we can a.a.s. embed each F_i^h into $R_1\left[V_i^h\right]$ with $an/2$ vertices left over. We have now embedded F into R_1 but nothing else; it remains to join up $(1-\phi)n/(k-1)$ special pairs of vertices with special paths. For each special pair, arbitrarily designate one vertex as an "x-type" endpoint and the other as "y-type". Let X be the set of x-type endpoints and let Y be the set of y-type endpoints. There are at most $\tau((1-\phi)n/k+1)$ special pairs (x, y) which have x and y in different groups V_i^h (call these "annoying" special pairs), and each V_i^h includes at least $\gamma((1-\phi)n/(k-1)+1) \geq \gamma n/(2k)$ special pairs.

Now, in both cases, for each $j = (i, h) \in [q] \times [2]$, let $W_j^0 = V_i^h \cap X$ and $W_j^4 = V_i^h \cap Y$. Consider the sets U_j^h obtained by deleting from V_i^h every vertex that has been used towards our embedding of T so far. That is, let $U_i^h = V_i^h \backslash \left(V(T') \cup \bigcup_{x\in X} V(T_x)\right)$ in case (a) and $U_i^h = V_i^h \backslash V((F) \cup \bigcup_{i=1}^r V(P_i))$ in case (b). Partition each U_i^h into 3 equal-size sets $W_{(i,\bar{h})}^1, W_{(i,h)}^2, W_{(i,\bar{h})}^3$ uniformly at random (if $3 \nmid \left|U_i^h\right|$ then allow one or two of the parts to have an extra vertex). Each of the pairs $\left(W_j^i, W_j^{i+1}\right)$ is a.a.s. $(\varepsilon, \delta/2)$-super-regular, by the following lemma.

Lemma 10. *For some $a > 0$, let (V_1, V_2) be an (ε, δ) super-regular pair in a graph G, where each $|V_h| \geq n$. For any $n_h \geq an$, let W_h be a uniformly random n_h-vertex subset of V_h. Then (W_1, W_2) is a.a.s. $(\varepsilon, \delta/2)$-super-regular.*

Proof. Since $W_h \subseteq V_h$, by definition (W_1, W_2) is $(\varepsilon, \delta/2)$-semi-super-regular. Each vertex in W_h has at least $\delta|V_{\bar{h}}|$ vertices in $V_{\bar{h}}$; by a concentration inequality for the hypergeometric distribution and the union bound, at least $\delta|W_{\bar{h}}|/2$ of those remain in $W_{\bar{h}}$. □

Now, we deal with the annoying special pairs in case (b). For small τ, there are at most

$$\tau((1-\phi)n/k+1) \leq \delta an/(4k)$$

such. Consider a maximal collection $P_1, \dots, P_m$ of special paths in $G \cup R_2$ connecting annoying special pairs. Let $G' = G \backslash \bigcup_{i=1}^m P_i$ (we are deleting vertices, not just edges). Since we have removed fewer than $\delta an/4$ vertices, G' has minimum degree at least $\delta an/4$. Now, suppose there is some annoying special pair (x, y) left unconnected. Let $x \dots u$ be a path of length $k-3$ in G. Let U be a set of $\delta an/8$ neighbours of u and let Y be a set of $\delta an/8$ neighbours of Y, in such a way that U and Y are disjoint with each other and with our new $(k-3)$-path (we can do this because all the vertices we have used in the P_i and the new path comprise at most $\delta an/4$ vertices). Now, note that (if c_2 is large), there is a.a.s. an edge in R_2 between any pair of disjoint sets of size $\delta an/4$ (the proof of this is basically identical to the proof of Lemma 4). But such an edge between U and Y would allow us to augment our collection of paths $P_1, \dots, P_m$, contradicting its maximality. Therefore a.a.s. we can connect all the annoying special pairs. Let $S_j^i = W_j^i \backslash \bigcup_{i=1}^m P_i$, so the pairs $\left(S_j^i, S_j^{i+1}\right)$ are $(\varepsilon/2, \delta/4)$-super-regular. In case (a), we can just let $S_j^i = W_j^i$. For large k, in both cases $2\left|S_j^0\right| = 2\left|S_j^4\right| \leq \left|S_j^1\right|, \left|S_j^3\right|$, and there is $a' = a'(a, \gamma, k)$ such that each $a'n \leq \left|S_j^h\right| \leq bn$.

After some adjustments using R_3, the "cluster-paths" $\left(S_j^0, S_j^1, S_j^2, S_j^3, S_j^4\right)$ will eventually decompose into our required special paths, using R_4 and the following lemma. (We also impose that k is even).

Lemma 11. *For any $\delta, a > 0$ and any even integer $k > 6$, there is $\varepsilon = \varepsilon(\delta) > 0$ and $c = c(k, \delta)$ such that the following holds. Let G be a graph on the vertex set $[n]$, together with a vertex partition $[n] = V_0 \cup V_1 \cup V_2 \cup V_3 \cup V_4$, such that each (V_i, V_{i+1}) is (ε, δ)-superregular. Suppose each vertex $x \in V_0$ is bijectively paired with a vertex $y \in V_4$, comprising a special pair (x, y). Let $n_i = |V_i|$ and suppose $(k-1)n_0 = n_1 + n_2 + n_3$, $an \leq n_i$, and $2n_0 \leq n_1, n_3$. Let $R \in \mathbb{G}(n, c/n)$. Then $G \cup R$ a.a.s. contains a spanning vertex-disjoint collection of n_0 special paths.*

After some adjustments, Lemma 11 will in turn follow from Lemma 13 (upcoming shortly), which is proved with the blow-up lemma.

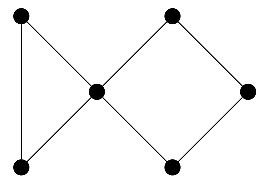
Lemma 12 (Blow-up Lemma [5])**.** *Let $\delta, \Delta > 0$ and $r \in \mathbb{N}$. There is $\varepsilon = \varepsilon(r, \delta, \Delta)$ such that the following holds. Let C be a graph on the vertex set $[r]$, and let $n_1, \dots, n_r$ be arbitrary positive integers. Let $V_1, \dots, V_r$ be pairwise disjoint sets of sizes $n_1, \dots, n_r$. We construct two graphs on the vertex set $V = \bigcup_{i=1}^r V_i$ as follows. The first graph $b(C)$ (the "complete blow-up") is obtained by putting the complete bipartite graph between V_i and V_j whenever $\{i, j\}$ is an edge in C. The second graph $b_{\varepsilon,\delta}(C)$ (the "super-regular blow-up") is obtained by putting edges between each such V_i and V_j such that (V_i, V_j) is an (ε, δ)-super-regular pair. If a graph H with maximum degree bounded by Δ can be embedded into $b(C)$, then it can be embedded into $b_{\varepsilon,\delta}(C)$.*

Our promised lemma is as follows.

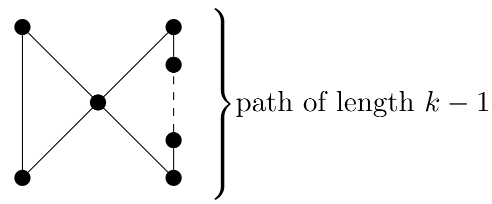
Lemma 13. *For any $\delta > 0$, there is $\varepsilon = \varepsilon(\delta) > 0$ such that the following holds. Let G be a graph with a vertex partition $V(G) = V_0 \cup V_1 \cup V_2 \cup V_3 \cup V_4$, such that each (V_i, V_{i+1}) is (ε, δ)-superregular. Suppose each vertex $x \in V_0$ is bijectively paired with a vertex $y \in V_4$, comprising a special pair (x, y). Let $n_i = |V_i|$ and suppose $(k-1)n_0 = n_1 + n_2 + n_3$, $n_0 = n_4 \leq n_1, n_3$ and $n_0 + n_2 = n_1 + n_3$ for some even k. Then G contains a spanning vertex-disjoint collection of n_0 special paths.*

Proof. Let W_0 be the set of n_0 special pairs, and let W_1 and W_2 be additional n_0-element sets (disjoint to everything so far). We construct a new multipartite graph $b_{\varepsilon,\delta}(C)$ on the vertex set $W_0 \cup W_1 \cup W_2 \cup V_1 \cup V_2 \cup V_3$ as follows. Every vertex of W_i is adjacent to every vertex of W_j $(i \neq j)$.

A vertex $(x,y) \in W_0$ is adjacent to a vertex $v \in V_1$ if x is adjacent to v in G, and (x,y) is adjacent to $v \in W_3$ if y is adjacent to v in G. The adjacencies between V_1 and V_2, and between V_2 and V_3, are as in G. Then $b_{\varepsilon,\delta}(C)$ is a (ε,δ)-super-regular blow-up of the following graph.

$C =$ 

Now, let H be a disjoint union of n_0 copies of the following graph with maximum degree 4.

$H' =$ 

The conditions on the n_i ensure that H is embeddable in the complete blow-up $b(C)$. Specifically, to embed each copy of H' in $b(C)$, embed the triangular part using the clusters W_i, then embed the $(k-1)$-path zigzagging through the clusters V_i. By the blow-up lemma, for small ε the super-regular blow-up $b_{\varepsilon,\delta}(C)$ contains H as well. This embedding of H into $b_{\varepsilon,\delta}(C)$ must be of the same type as our embedding into $b(C)$: the triangular part of H' can only be embedded using the clusters W_i, and embedding all n_0 of these triangular parts uses up all the vertices in the W_i. Therefore the embeddings of the k-cycle part of H' give us the special paths in G we require.

This seems like a silly way to apply the blow-up lemma... I add the 3-cycle just to make sure my k-cycles are embedded in the "right" way. Is there a better way to guarantee this? □

It remains to explain the "adjustments" necessary to apply the above two lemmas. For both, we make use of the following lemma $O(1)$ times, repeatedly removing sets of "adjusting" special paths.

Lemma 14. *For any $\delta < 0$, $f < 1$, and any integer $k > 2$ there is $\delta' = \delta'(\delta) > 0$ and $c = c(\delta, f, k)$ such that the following holds. Let G be a graph on some vertex set $[N]$, together with a vertex partition into $O(1)$ clusters, such that some of the pairs of clusters (V,W) are (ε,δ)-super-regular. Consider any sequence of clusters $V_0,\dots,V_k$, such that (V_0,V_1) and (V_{k-1},V_k) are (ε,δ)-super-regular pairs, and such that if a cluster appears in the sequence t times, then it has at least tn vertices. Suppose further that each $x \in V_0$ is bijectively paired with a vertex $y \in V_k$, comprising a special pair (x,y). Let $R \in \mathbb{G}(n, c/n)$. Then for any $m \le fn$, there are m vertex-disjoint special paths of the form $v_1 \dots v_k$ ($v_i \in V_i$) in $G \cup R$. Let P be the union of our special paths; we can choose these paths in such a way that $(V \backslash P, W \backslash P)$ is (ε,δ')-super-regular for each (V,W) that was (ε,δ)-super-regular (in G).*

We also give a simplified form of the same lemma, which we will use in the proof of Lemma 14.

Lemma 15. *For any $\delta > 0$, $g < 1$ and $k \in \mathbb{N}$, there is $c = c(\delta, p, k)$ such that the following holds. Let G be a graph on the vertex set $[kn]$, together with a vertex partition $[kn] = V_0 \cup \cdots \cup V_k$. Suppose that each $|V_i| = n$, and each $v \in V_0$ (respectively, $v \in V_k$) has at least δn neighbours in V_2 (respectively, in V_{k-1}). Suppose further that each $x \in V_0$ is bijectively paired with a vertex $y \in V_k$, comprising a special pair (x, y). Let $R \in \mathbb{G}(n, c/n)$. Then there are a.a.s. gn vertex-disjoint special paths of the form $v_1 \dots v_k$ ($v_i \in V_i$) in $G \cup R$.*

Proof. It suffices to show that we can a.a.s. find εn such disjoint paths, for any small $\varepsilon = \varepsilon(\delta) > 0$. This is because we can break R into phases $R \supseteq R_1 \cup \cdots \cup R_r$, where r is chosen so that $(1-\varepsilon)^r \leq (1-g)$. Each phase we can cover an ε-fraction of the vertices with suitable k-paths; discard these paths for the next phase and repeat.

We proceed with this plan, for $\varepsilon = \delta^3/\big(2\sqrt{2}\big)$. For any $\gamma = \gamma(\delta) > 0$, by Lemma 4 we can find a matching of $(1 - (1-\gamma)/(k-2))n$ edges between each V_i and V_{i+1} ($1 \leq i < k-1$). This gives us a disjoint union of γn paths $P_1, \dots, P_{\gamma n}$ in $G \cup R$, where $P_i = v_1^i \dots v_{k-1}^i$ (with $v_j^i \in V_j$). We can assume that $v_j^1, \dots, v_j^{\gamma n}$ is a uniformly random ordering of a uniformly random set of γn elements of V_j (independently for $2 \leq j \leq k-1$).

Now, let the special pairs be $\big(x^1, y^1\big), \dots, (x^n, y^n)$. Note that $\Pr\big(v_1^1 \sim x^1, v_{k-1}^1 \sim y^1\big) \geq \delta^2$ and more generally,

$$\Pr\big(v_1^i \sim x^i, v_{k-1}^i \sim y^i \mid v_1^1, v_{k-1}^1, \dots, v_1^{\gamma n}, v_{k-1}^{\gamma n}\big) \geq \Big((\delta n)^2 - i^2\Big)/n^2 \geq \delta^2/2$$

for all $i \leq \gamma n$, with $\gamma = \delta n/\sqrt{2}$. By the Chernoff bound, there are a.a.s. $\delta^3 n/\big(2\sqrt{2}\big)$ special paths $x_i, v_1^i, \dots, v_{k-1}^i, y_i$. □

Proof of Lemma 14. First note that we can assume each V_i is distinct. For, we can take every cluster that appears multiple times (say t times) in the sequence and split that cluster into t equal-sized subclusters uniformly at random. Since $t \leq k = O(1)$, by Lemma 10 super-regularity is a.a.s. preserved by these splits. That is, if (V, W) was (ε, δ)-super-regular in G, and $V' \subseteq V$, $W' \subseteq W$ are clusters after the splits, then (V', W') is a.a.s. $(\varepsilon, \delta/2)$-super-regular.

So we proceed with distinct V_i. For each V_i, set aside a uniformly random subset D_i of size $(|V_i| - m)/2$. Let $U_i = V_i \backslash D_i$ so that $|V_i \backslash D_i| \geq n/2$ and $m \leq 2f|V_i \backslash D_i|/(1+f)$. Apply Lemma 15 with $g = 2f/(1+f)$ to find m special paths in $G \cup R$ through the clusters $V_0 \backslash D_0, \dots V_k \backslash D_k$. Note that each $|V_i \backslash P| = 2|D_i|$. For each (ε, δ)-super-regular pair (V_i, V_j) (respectively (V_i, W)), by Lemma 10 (D_i, D_j) (respectively (D_i, W)) is a.a.s. $(\varepsilon, \delta/2)$-super-regular, so that $(V_i \backslash P, V_j \backslash P)$ (respectively $(V_i \backslash P, W)$) is $(\varepsilon, \delta/4)$-super-regular, as required. □

Proof of Lemma 11. First consider the case where $n_0 + n_2 > n_1 + n_3$. By the constraints on the n_i, we have $n_2 \leq (k-5)n_0$ and $n_0 + n_2 - (n_1 + n_3) \leq (k-6)n_0$. With $f = \frac{k-6}{k-4}$, we can find $\left\lfloor \frac{n_0 + n_2 - (n_1 + n_3)}{k-4} \right\rfloor$ disjoint special paths through the clusters

$$V_0, V_1, \underbrace{V_2, \dots, V_2}_{k-3}, V_3, V_4.$$

Let $q = (k-4)\Big\langle \frac{n_0 + n_2 - (n_1 + n_3)}{k-4} \Big\rangle$ (where $\langle \cdot \rangle$ gives the fractional part of a real number). From the vertices that remain after removing the paths so far, apply Lemma 14 again with a small σ to find a single path through the clusters

$$V_0, \underbrace{V_1, \ldots, V_1}_{(k-2-q)/2}, \underbrace{V_2, \ldots, V_2}_{(k-2+q)/2}, V_3, V_4$$

or through

$$V_0, V_1, \underbrace{V_2, \ldots, V_2}_{(k-2+q)/2}, \underbrace{V_3, \ldots, V_3}_{(k-2-q)/2}, V_4$$

(one of these is possible). After removing all the adjusting paths the conditions for Lemma 13 are satisfied, and Lemma 13 ensures that we can finish connecting the special pairs.

The case $n_0 + n_2 < n_1 + n_3$ is almost identical. Remove many paths through the clusters

$$V_0, V_1, \underbrace{V_3, \ldots, , V_3}_{k-2}, V_4,$$

then through the clusters

$$V_0, \underbrace{V_1, \ldots, , V_1}_{k-2}, V_3, V_4,$$

then clean up by removing a single path before applying Lemma 13. □

Now, we conclude the proof of Theorem 1: recall that we are trying to remove paths from the clusters S_j^i using R_3, in order to apply Lemma 11 with R_4. This proceeds in the same way as the adjusting path arguments in the proof of Lemma 11.

Suppose $\left|S_j^1 \cup S_j^2 \cup S_j^3\right| + m \geq (k-1)\left|S_j^0\right|$ and $\left|S_r^1 \cup S_r^2 \cup S_r^3\right| - m \leq (k-1)\left|S_{j'}^0\right|$ for some j, r, where one of these inequalities is an equality. Then use Lemma Lemma 14 with $f = 1 - a'/(bk)$ to find $\lfloor m/(k-3) \rfloor$ paths through the clusters

$$S_r^0, S_r^1, \underbrace{S_j^2, \ldots, S_j^2}_{k-3}, S_r^3, S_r^4.$$

After removing these paths (plus an additional "cleanup path") to obtain clusters F_j^i, we have $\left|F_g^1 \cup F_g^2 \cup F_g^3\right| = (k-1)\left|F_g^0\right|$ for some $g \in \{j, r\}$, and still each $\left|F_g^1\right|, \left|F_g^3\right| \geq 2\left|F_g^0\right|$. After removing paths in this way $q = O(1)$ times, the conditions for Lemma 11 are satisfied for each cluster-path.

These "adjusting path" arguments are extremely unwiedly. Maybe there's some different way to express the situation that makes it easier to state what I'm doing?

References

[1] Noga Alon, Michael Krivelevich, and Benny Sudakov, *Embedding nearly-spanning bounded degree trees*, Combinatorica **27** (2007), no. 6, 629–644.

[2] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Cambridge University Press, 2000.

[3] Kumar Joag-Dev and Frank Proschan, *Negative association of random variables with applications*, The Annals of Statistics (1983), 286–295.

[4] J Komlós and M Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty (D Miklós, VT Sos, and T Szonyi, eds.), Bolyai Society mathematical studies, vol. 2, János Bolyai Mathematical Society, Budapest, 1996, pp. 295–352.

[5] János Komlós, Gábor N Sárközy, and Endre Szemerédi, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.

[6] Michael Krivelevich, *Embedding spanning trees in random graphs*, SIAM Journal on Discrete Mathematics **24** (2010), no. 4, 1495–1500.