Bounded-degree spanning trees in randomly perturbed graphs

Michael Krivelevich, Matthew Kwan, Benny Sudakov

February 22, 2015

Abstract

abstract

1 Introduction

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 (Section 2.1), 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 where T has few leaves, which we attack in Section 2.2. 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 these bare paths we obtain a small forest which we can embed into R using Lemma 2 (Section 2.2.2). In order to complete this into our spanning tree T, we need to join up distinguished pairs of vertices with disjoint paths of certain lengths. In order to make this task feasible, in Section 2.2.1 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). In Section 2.2.3 we use R to find most of our desired paths, in such a way that it only remains to join up pairs of vertices within the same superregular pair. We make some further adjustments with R (Section 2.2.4), after which the super-regularity of the pairs allows us to find the rest of our paths using a tool called the blow-up lemma (Section 2.2.5).

2 Proof of Theorem 1

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).

2.1 Case 1: T has many leaves

For our first case suppose there are at least λn leaves in T (for some $\lambda = \lambda(\alpha) > 0$ to be determined in the second case). 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.

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

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

so by the Chernoff bound,

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

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

Instead of proving this via negative association and a Chernoff bound, its possible to use a concentration inequality for the hypergeometric distribution several times. But I tried writing both ways and this way seems cleaner. □

2.2 Case 2: T has few leaves

Now we address the second case where there are fewer than λn leaves in T.

2.2.1 Partitioning into super-regular pairs

As outlined, we first need to divide G into a bounded number of pairs of clusters with edges well-distributed between them. This partition will inform the way we use R to embed T.

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 $|U_h| \geq \varepsilon|V_h|$, we have $|d(U_1,U_2) - d(V_1,V_2)| \leq \varepsilon$. If alternatively $d(U_1,U_2) \geq \delta$ for all such pairs U_1,U_2 then we say (V_1,V_2) is *(ε,δ)-dense*. Let $\bar{h} = 2-h$; if (V_1,V_2) is (ε,δ)-dense 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*.

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 $an \le |V_i^h| \le bn$ and each pair (V_i^1, V_i^2) 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 (V_i^1, V_i^2). 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. $\square$

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 $|U^h| \ge \varepsilon' |W_i^h|$, 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 $|U^2 \cap V_p| \ge |U^2|/\ell \ge \varepsilon' sn$. By ε'-regularity there are at least $(\alpha' - \varepsilon')|U^1|(|U^2|/\ell) \ge \left((\alpha')^2/2\right)|U^1||U^2|$ edges between $U^2 \cap V_p$ and U^1. That is, $d(U^1, U^2) \ge (\alpha')^2/2$, so with $\delta' = (\alpha')^2/2$ each pair (W_i^1, W_i^2) is (ε', δ')-semi-super-regular.

There are still some vertices left over from exceptional clusters, and we cannot guarantee that the (W_i^1, W_i^2) 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 $(1-\varepsilon)\delta'\left|S_i^h\right| \geq \delta'\left|S_i^h\right|/2$ 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 $V_i^{\bar{h}}$, each $\left|V_i^h \backslash S_i^h\right| \leq (4\varepsilon'/\alpha')\left|V_i^h\right|$, 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 $\alpha'\left|V_i^h\right|/2 \geq \delta'\left|S_i^h\right|/2$ 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 $an \leq \left|V_i^h\right| \leq bn$. For each i, if $U^h \subseteq V_i^h$ with $\left|U^h\right| \geq (4/\alpha'+1)\varepsilon'\left|V_i^h\right|$, then $\left|U^h \cap W_i^h\right| \geq \varepsilon'\left|W_i^h\right|$ so

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

That is, each pair $\left(V_i^1, V_i^2\right)$ 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 $\delta'\left|S_i^{\bar{h}}\right|/2 \geq \delta'\left|V_i^{\bar{h}}\right|/(2(4\varepsilon'/\alpha'+1)) \geq \delta\left|V_i^{\bar{h}}\right|$ neighbours in its partner cluster $V_i^{\bar{h}}$, so in fact each pair $\left(V_i^1, V_i^2\right)$ is $((4/\alpha'+1)\varepsilon', \delta)$-super-regular. With ε' small enough so that $(4/\alpha'+1)\varepsilon' \leq \varepsilon$, we are done. $\square$

We apply Lemma 6 to our graph G, with $\varepsilon = \varepsilon(\delta)$ to be determined. For $\mathbf{i} = (i, h)$, let $V_{\mathbf{i}} = V_i^h$, and let $|V_{\mathbf{i}}| = s_{\mathbf{i}} n = s_i^h n$.

2.2.2 Embedding a small subforest of T

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 $k = k(\phi, \alpha)$ large enough and $\lambda = \lambda(k)$ small enough, then T contains a collection of $\psi n := (1-\phi)n/(k-1)$ disjoint bare k-paths. (We also impose that k is odd, for reasons that will become clear later). If we delete the interior vertices of these paths, then we are left with a forest F on ϕn vertices.

Now, embed F into R_1. There are ψn "special pairs" of vertices of $F \subseteq R_1$ that need to be connected with k-paths. We call such connecting paths "special paths". For each special pair, arbitrarily choose one vertex to be "x-type" and the other to be "y-type". Let X and Y be the set of x-type and y-type vertices, respectively. By symmetry we can assume

$$X \cup Y \cup F \backslash (X \cup Y) \cup V \backslash F$$

is a uniformly random partition of V into parts of sizes ψn, ψn, $(\phi - 2\psi)n$, $(1-\phi)n$, and the special pairs correspond to a random bijection between X and Y. For $\mathbf{i} = (i, h)$, let $X_{\mathbf{i}} = X_i^h = V_{\mathbf{i}} \cap X$ and $Y_{\mathbf{i}} = Y_i^h = V_{\mathbf{i}} \cap Y$, and let $Z_{\mathbf{i},\mathbf{j}}$ be the set of special pairs in $X_{\mathbf{i}} \times Y_{\mathbf{j}}$. By a concentration inequality for the hypergeometric distribution and the union bound, a.a.s. each $|V_{\mathbf{i}} \backslash (F \backslash (X \cup Y))| \sim (1 - \phi + 2\psi)s_{\mathbf{i}} n$. Each $|X_{\mathbf{i}}| \sim \psi s_{\mathbf{i}} n$, and therefore each $|Z_{\mathbf{i},\mathbf{j}}| \sim \psi s_{\mathbf{i}} s_{\mathbf{j}} n$.

Now we have embedded F, we no longer care about the vertices used to embed $F \backslash (X \cup Y)$; they will never be used again. It is convenient to imagine, for the duration of the proof, that instead of embedding parts of T, we are "removing" vertices from $G \cup R$, gradually making the remaining graph easier to deal with. So, remove from each $V_{\mathbf{i}}$ all vertices used in $F \backslash (X \cup Y)$. Update n to be the number of vertices not used to embed $F \backslash (X \cup Y)$ (previously $(1 - \phi + 2\psi)n$), and update each $s_{\mathbf{i}}$ to still satisfy $s_{\mathbf{i}} n = |V_{\mathbf{i}}|$ (this does not change $s_{\mathbf{i}}$ asymptotically). We now have

$$|Z_{\mathbf{i},\mathbf{j}}| \sim \psi(1-\phi)/(1-\phi+2\psi)s_{\mathbf{i}} s_{\mathbf{j}} n = s_{\mathbf{i}} s_{\mathbf{j}} n/(k+1).$$

For $\mathbf{i} = (i, h)$, also let $W_i^h = W_{\mathbf{i}} = V_{\mathbf{i}} \backslash (X \cup Y)$.

Here and in future phases, it is critical that after removing vertices from the $V_{\mathbf{i}}$ we do not destroy super-regularity. This will mostly be guaranteed by making choices randomly and applying the following lemma.

Lemma 9. *Fix any $a > 0$. 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 $(\varepsilon, \delta/2)$-dense. 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 $W_{\bar{h}}$. $\square$

Let $X_{\mathbf{i},\mathbf{j}}$ and $Y_{\mathbf{i},\mathbf{j}}$ be the set of x-type and y-type vertices in pairs in $Z_{\mathbf{i},\mathbf{j}}$, respectively. An immediate consequence of Lemma 9 is that each $\left(X_{(i,h),\mathbf{j}}, W_i^{\bar{h}}\right)$, $\left(W_i^1, W_i^2\right)$ and $\left(W_i^h, Y_{\mathbf{j},(i,\bar{h})}\right)$ are $(\varepsilon, \delta/2)$-super-regular.

2.2.3 Embedding special paths not compatible with the super-regular partition

Next, we want to eliminate almost all special pairs which are not between clusters of a super-regular pair. This is achieved using the following lemma (which will be used again later).

Lemma 10. *For any $\delta > 0$, $f < 1$, and any integer $k > 2$ there is $c = c(\delta, f, k)$ 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 (U, W) are (ε, δ)-super-regular. Consider any sequence of clusters $X, V_1, \ldots, V_r, Y$, and any partition $t_1 + \cdots + t_r = k - 1$ (with each $t_i \in \mathbb{N}$), such that $|V_i| \geq t_i n$ and $|X|, |Y| \geq n$. Suppose that (X, V_1) and (V_r, Y) are (ε, δ)-super-regular pairs, and suppose further that each $x \in X$ is bijectively paired with a vertex $y \in Y$, comprising a special pair (x, y). Let $R \in \mathbb{G}(N, c/N)$.

Then, for any $m \le fn$, in $G \cup R$ there are m vertex-disjoint special paths, each with t_i vertices in each V_i. These paths can be chosen such that after their removal, each (U, W) that was previously (ε, δ)-super-regular is now $(\varepsilon, \delta/4)$-super-regular.

It is convenient to prove Lemma 10 using a similar, but simpler, version of the lemma.

Lemma 11. *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 $[(k+1)n]$, together with a vertex partition $[(k+1)n] = X \cup V_1 \cdots \cup V_{k-1} \cup Y$. 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 $x \in X$ is bijectively paired with a vertex $y \in Y$, 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 $xv_1 \dots v_{k-1}y$ ($v_i \in V_i$) 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 $(1-\varepsilon)^r \le (1-g)$. 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} ($1 \le i < k-1$). 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 $v_j^i \in V_j$). For each 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 $2 \le j \le k-1$).

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

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

for all $i \le \gamma n$, 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 10. Let $V_0 = X$ and $V_k = Y$. For each V_i, set aside a uniformly random subset D_i of size $(|V_i| - t_i m)/2$. Divide each $V_i \backslash D_i$ ($1 \le i \le k-1$) randomly into t_i equal-sized pieces and order the resulting $k-1$ clusters as $U_1, \dots, U_{k-1}$ such that $U_1 \subseteq V_1$ and $U_{k-1} \subseteq V_r$. By Lemma 9, (X, U_1) and (U_{k-1}, Y) are $(\varepsilon, \delta/2)$-super-regular. Each $|U_i| \ge n/(2(k-1))$ and also note

$$\begin{aligned} |V_i \backslash D_i| &= |V_i|/2 + t_i m/2, \\ m &\le f|V_i|/t_i \\ &\le 2f|V_i \backslash D_i|/t_i - fm, \\ (1+f)m &\le 2f|V_i \backslash D_i|/t_i. \end{aligned}$$

Let $f' = 2f/(1+f)$ so each $f'|U_i| \ge m$. Apply Lemma 11 to find m special paths in $G \cup R$ with the required intersection with each V_i. After removing these paths, there are $2|D_i|$ vertices left in each V_i. For each (ε, δ)-super-regular pair (V_i, V_j) (respectively (V_i, W)) in the original graph, by Lemma 9 (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. □

Now we return to the proof of Theorem 1. For each $\mathbf{i} = (i,h)$ and each $\mathbf{j} = (j,g) \neq (i,\bar{h})$ and some very small $\xi = \xi(a,\delta) > 0$, we want to find $(1-\xi)s_{\mathbf{i}}s_{\mathbf{j}}n/(k+1)$ special paths connecting special pairs in $V_{\mathbf{i},\mathbf{j}}$, where each of these special paths has one interior vertex from $W_i^{\bar{h}}$, one interior vertex from W_j^g and the other $k-3$ interior vertices from W_i^h. Provided k is large we can find our desired paths with at most $(2q)^2$ applications of Lemma 10 (splitting R_2 into $(2q)^2$ phases).

Remove these paths, and update each $V_{\mathbf{i}}, W_{\mathbf{i}}, X_{\mathbf{i}}, Y_{\mathbf{i}} X_{\mathbf{i},\mathbf{j}}, Y_{\mathbf{i},\mathbf{j}}, Z_{\mathbf{i},\mathbf{j}}$ accordingly. (but we do not update n or the $s_{\mathbf{i}}$). Each $\left(X_{(i,h),\mathbf{j}}, W_i^{\bar{h}}\right)$, $\left(W_i^1, W_i^2\right)$ and $\left(W_i^h, Y_{\mathbf{j},(i,\bar{h})}\right)$ are (ε, δ')-regular, where $\delta' = (\delta/2)/4^{(2q)^2}$. Let $a' = a'(a)$ give a lower bound on the sizes of the $W_{\mathbf{i}}$: $|W_{\mathbf{i}}| \geq a'n$. (for large k, $a' = a^2/2$ will do).

Now, for every $\mathbf{i} = (i,h)$ and $\mathbf{j} = (j,g) \neq (i,\bar{h})$, $|Z_{\mathbf{i},\mathbf{j}}| = \xi s_{\mathbf{i}}s_{\mathbf{j}}n/(k+1)$. The special pairs in all such $Z_{\mathbf{i},\mathbf{j}}$ comprise a total of at most $\xi n/(k+1)$ special pairs. For small ξ this is less than $\delta' a' n/(2(k-1))$ special pairs, which we deal with greedily, as follows. For some such special pair (x,y) with $x \in X_i^h$, choose a length-$(k-3)$ path $xw_1, \ldots, w_{k-3}$, where the w_ℓ alternate between W_i^h and $W_i^{\bar{h}}$ with $w_{k-3} \in W_i^h$ (recall, k is odd). Then choose a set $U_x \subseteq W_i^{\bar{h}}$ of $\delta' a' n/4$ neighbours of w_{k-3} and a disjoint set $U_y \subseteq W_j^{\bar{g}}$ of $\delta' a' n/4$ neighbours of y, where $y \in V_j^g$. Now, note that (if c_3 is large), there is a.a.s. an edge in R_3 between any pair of disjoint sets of size $\delta' a' n/4$ (the proof of this is basically identical to the proof of Lemma 4). Such an edge allows us to complete a k-path from x to y, which we remove. We can repeat this for each unconnected special pair in $Z_{\mathbf{i},\mathbf{j}}$. Note that these special paths comprise less than $\delta' a' n/2$ vertices total, which is so few that each $\left(W_i^1, W_i^2\right)$ is $(\varepsilon, \delta'/2)$-super-regular, and each $x \in X_i^h$ (respectively $y \in Y_i^h$) has at least $\delta a' n/2$ neighbours in $W_i^{\bar{h}}$, throughout the whole process.

Now, update all the relevant variables. We are left with a set of clusters $X_{\mathbf{i}}, W_{\mathbf{i}}, Y_{\mathbf{i}}$ partitioning the vertices of $G \cup R$, such that each $\left(X_i^h, W_i^{\bar{h}}\right)$, $\left(W_i^1, W_i^2\right)$ and $\left(W_i^h, Y_i^{\bar{h}}\right)$ is $(\varepsilon, \delta'/2)$-regular, and each special pair is in some $X_i^h \times Y_i^{\bar{h}}$. Moreover, with X (respectively Y) being the set of x-type (respectively y-type) vertices remaining, and W being the set of non-special vertices remaining, we have $(k-1)|X| = (k-1)|Y| = |W|$. For large k each $an \leq |X_{\mathbf{i}}|, |Y_{\mathbf{i}}|, |W_{\mathbf{i}}| \leq bn$ for some a, b, and each $|X_{\mathbf{i}}|, |Y_{\mathbf{i}}| \leq |W_{\mathbf{j}}|$.

2.2.4 Adjusting the relative sizes of the clusters

The next step is to remove paths such that in the graph that remains, we have $|W_{\mathbf{i}}| = (k-1)|X_{\mathbf{i}} \cup Y_{\mathbf{i}}|/2$ for each $\mathbf{i}$. This is easily achievable through repeated application of Lemma 10, but is a bit fiddly to explain.

Suppose $(k-1)|X_{\mathbf{i}} \cup Y_{\mathbf{i}}|/2 + m \geq |W_{\mathbf{i}}|$ and $(k-1)|X_{\mathbf{j}} \cup Y_{\mathbf{j}}|/2 + m \leq |W_{\mathbf{j}}|$, where one of these inequalities is an equality. Let $\mathbf{i} = (i,h)$ and $\mathbf{j} = (j,g)$. Use Lemma 10 and R_4 to find $\lfloor m/((k-1)/2-1) \rfloor$ special paths connecting special pairs in $X_i^h \times Y_i^{\bar{h}}$, each with $(k-1)/2$ interior vertices in $W_i^{\bar{h}}$, one interior vertex in W_i^h, and the remaining $(k-1)/2 - 1$ interior vertices in W_j^g.

Then, let $r = ((k-1)/2-1)\langle m/((k-1)/2-1) \rangle$, where $\langle \cdot \rangle$ denotes the fractional part of a real number. With another application of Lemma 10, find a single special path of the same type, except it has $(k-1)/2 - 1 - r$ interior vertices in W_i^h and r interior vertices in W_j^g. We now have either $(k-1)|X_{\mathbf{i}} \cup Y_{\mathbf{i}}|/2 = |W_{\mathbf{i}}|$ or $(k-1)|X_{\mathbf{j}} \cup Y_{\mathbf{j}}|/2 \leq |W_{\mathbf{j}}|$. After at most $2q$ iterations of this argument, we will have each $(k-1)|X_{\mathbf{i}} \cup Y_{\mathbf{i}}|/2 = |W_{\mathbf{i}}|$ as desired. Each $\left(X_i^h, W_i^{\bar{h}}\right)$, $\left(W_i^1, W_i^2\right)$ and $\left(W_i^h, Y_i^{\bar{h}}\right)$ is now (ε, δ'')-super-regular, where $\delta'' = \delta/4^{2q}$.

2.2.5 Completing the embedding with the blow-up lemma

We have now finished with R; what remains is sufficiently well-structured that we can embed the remaining paths using the blow-up lemma and the super-regularity in G. The idea is illustrated in Figure 1.

Uniformly at random divide each W_i^h into sets A_i^h and B_i^h of size $(k-1)\left|X_i^h\right|/2$ and $(k-1)\left|Y_i^h\right|/2$ respectively. For $\mathbf{i}=(i,h)$, let $S_{\mathbf{i}}^0=X_i^h$, $S_{\mathbf{i}}^1=B_i^{\bar{h}}$, $S_{\mathbf{i}}^2=A_i^h$ and $S_{\mathbf{i}}^3=Y_i^{\bar{h}}$, so that each $\left(S_{\mathbf{i}}^0,S_{\mathbf{i}}^2,S_{\mathbf{i}}^3,S_{\mathbf{i}}^3\right)$ is a sequence of clusters of sizes $n_{\mathbf{i}},(k-1)n_{\mathbf{i}}/2,(k-1)n_{\mathbf{i}}/2,n_{\mathbf{i}}$ for some $n_{\mathbf{i}}$, and each $\left(S_{\mathbf{i}}^\ell,S_{\mathbf{i}}^{\ell+1}\right)$ is $(\varepsilon,\delta''/2)$-super-regular.

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 $b_{\varepsilon,\delta}(C)$ (a "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 $b_{\varepsilon,\delta}(C)$.*

For each $\mathbf{i}$, we construct an auxiliary graph $G_{\mathbf{i}}$ as follows. Start with the 4-partite subgraph of G induced by $S_{\mathbf{i}}^0\cup S_{\mathbf{i}}^2\cup S_{\mathbf{i}}^3\cup S_{\mathbf{i}}^3$, and remove every edge not between some $S_{\mathbf{i}}^\ell$ and $S_{\mathbf{i}}^{\ell+1}$. Contract each special pair to a single vertex, to obtain a super-regular "cluster-cycle" with 3 clusters $S_{\mathbf{i}}^*,S_{\mathbf{i}}^1,S_{\mathbf{i}}^2$ of sizes $n_{\mathbf{i}},(k-1)n_{\mathbf{i}}/2,(k-1)n_{\mathbf{i}}/2$. This is a (ε,δ'')-super-regular blow-up of a 3-cycle C_3.

The complete blow-up $b(C_3)$ contains $n_{\mathbf{i}}$ disjoint k-cycles (each has a vertex in $S_{\mathbf{i}}^*$ and its other $k-1$ vertices alternate between $S_{\mathbf{i}}^1$ and $S_{\mathbf{i}}^2$). By the blow-up lemma, for small ε the (ε,δ'')-super-regular blow-up $G_{\mathbf{i}}$ of C_3 also contains $n_{\mathbf{i}}$ disjoint k-cycles. Since k is odd, each of these must use at least one vertex from $S_{\mathbf{i}}^*$, but since $|S_{\mathbf{i}}^*|=n_0$, each cycle must then use exactly one vertex from $S_{\mathbf{i}}^*$. Each of these cycles corresponds to a special path running through $S_{\mathbf{i}}^0,S_{\mathbf{i}}^2,S_{\mathbf{i}}^3,S_{\mathbf{i}}^3$, and these special paths complete our embedding of T.

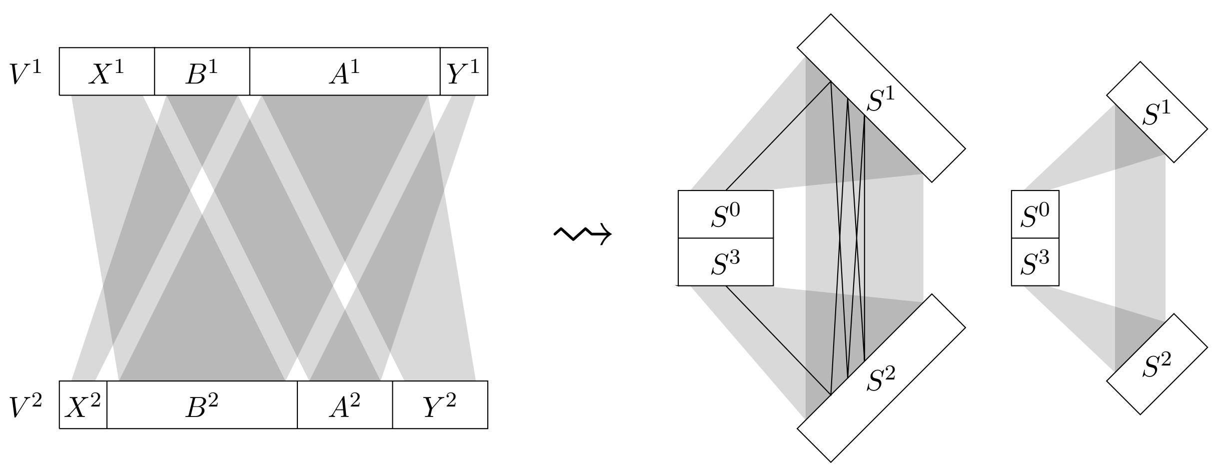

Figure 1. From super-regular pairs to cluster-cycles. An example special path is shown, for $k=7$.

References

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

[2] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Cambridge University Press, 2000.

[3] K. Joag-Dev and F. 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, V. Sos, and T. Szonyi, eds.), Bolyai Society mathematical studies, vol. 2, János Bolyai Mathematical Society, Budapest, 1996, pp. 295–352.

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

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