Bounded-degree spanning trees in randomly perturbed graphs

Michael Krivelevich * Matthew Kwan† Benny Sudakov‡

Abstract

We show that for any fixed dense graph G and bounded-degree tree T on the same number of vertices, a modest random perturbation of G will typically contain a copy of T. This combines the viewpoints of the well-studied problems of embedding trees into fixed dense graphs and into random graphs, and extends a sizeable body of existing research on randomly perturbed graphs. Specifically, we show that there is $c = c(\alpha, \Delta)$ such that if G is an n-vertex graph with minimum degree at least αn, and T is an n-vertex tree with maximum degree at most Δ, then if we add cn uniformly random edges to G, the resulting graph will contain T asymptotically almost surely (as $n \to \infty$). Our proof uses a lemma concerning the decomposition of a dense graph into super-regular pairs of comparable sizes, which may be of independent interest.

1 Introduction

A classical theorem of Dirac [7] states that any n-vertex graph ($n \geq 3$) with minimum degree at least $n/2$ has a Hamilton cycle: a cycle that passes through all the vertices of the graph. More recently, there have been many results showing that this kind of minimum degree condition is sufficient to guarantee the existence of different kinds of spanning subgraphs. For example, in [11, Theorem 1′] Komlós, Sárközy and Szemerédi proved that for any Δ and $\gamma > 0$, any n-vertex graph (n sufficiently large) with minimum degree $(1/2 + \gamma)n$ contains every spanning tree which has maximum degree at most Δ. This has also been generalized further in two directions. In [13] Komlós, Sárközy and Szemerédi improved their result to allow Δ to grow with n, and in [5] Böttcher, Schacht and Taraz gave a result for much more general spanning subgraphs than trees (see also a result [6] due to Csaba).

The constant $1/2$ in the above Dirac-type theorems is tight: in order to guarantee the existence of these spanning subgraphs we require very strong density conditions. But the situation is very different for a "typical" large graph. If we fix an arbitrarily small $\alpha > 0$ and select a graph uniformly at random among the (labelled) graphs with n vertices and $\alpha\binom{n}{2}$ edges, then the degrees will probably each be about αn. Such a random graph is Hamiltonian with probability $1 - o(1)$ (we say it is Hamiltonian *asymptotically almost surely*, or *a.a.s.*). This follows from a stronger result by Pósa [19] that gives a *threshold* for Hamiltonicity: a random n-vertex, m-edge graph is Hamiltonian a.a.s. if $m \gg n \log n$, and fails to be Hamiltonian a.a.s. if $m \ll n \log n$. Although the exact threshold

*School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 6997801, Israel. Email: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2010115 and by grant 912/12 from the Israel Science Foundation.

†Department of Mathematics, ETH, 8092 Zurich. Email: matthew.kwan@math.ethz.ch

‡Department of Mathematics, ETH, 8092 Zurich. Email: benjamin.sudakov@math.ethz.ch. Research supported in part by SNSF grant 200021-149111.

for bounded-degree spanning trees is not known, in [18] Montgomery proved that for any Δ and any tree T with maximum degree at most Δ, a random n-vertex graph with $\Delta n(\log n)^5$ edges a.a.s. contains T. Here and from now on, all asymptotics are as $n \to \infty$, and we implicitly round large quantities to integers.

In [3], Bohman, Frieze and Martin studied Hamiltonicity in the random graph model that starts with a dense graph and adds m random edges. This model is a natural generalization of the ordinary random graph model where we start with nothing, and offers a "hybrid" perspective combining the extremal and probabilistic settings of the last two paragraphs. The model has since been studied in a number of other contexts; see for example [2, 16, 15]. A property that holds a.a.s. with small m in our random graph model can be said to hold not just for a "globally" typical graph but for the typical graph in a small "neighbourhood" of our space of graphs. This tells us that the graphs which fail to satisfy our property are in some sense "fragile". We highlight that it is generally very easy to transfer results from our model of random perturbation to other natural models, including those that delete as well as add edges (this can be accomplished with standard coupling and conditioning arguments). We also note that a particularly important motivation is the notion of *smoothed analysis* of algorithms introduced by Spielman and Teng in [20]. This is a hybrid of worst-case and average-case analysis, studying the performance of algorithms in the more realistic setting of inputs that are "noisy" but not completely random.

The statement of [3, Theorem 1] is that for every $\alpha > 0$ there is $c = c(\alpha)$ such that if we start with a graph with minimum degree at least αn and add cn random edges, then the resulting graph will a.a.s. be Hamiltonian. This saves a logarithmic factor over the usual model where we start with the empty graph. Note that some dense graphs require a linear number of extra edges to become Hamiltonian (consider for example the complete bipartite graph with partition sizes $n/3$ and $2n/3$), so the order of magnitude of this result is tight.

Let $\mathbb{G}(n,p)$ be the binomial random graph model with vertex set $[n] = \{1, \ldots, n\}$, where each edge is present independently and with probability p. (For our purposes this is equivalent to the model that uniformly at random selects a $p\binom{n}{2}$-edge graph on the vertex set $[n]$). In this paper we prove the following theorem, extending the aforementioned result to bounded-degree spanning trees.

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, and T is an n-vertex tree with maximum degree at most Δ, and $R \in \mathbb{G}(n, c/n)$, then a.a.s. $T \subseteq G \cup R$.*

One of the ingredients of the proof is the following lemma, due to Alon and two of the authors ([1, Theorem 1.1]). It shows that the random edges in R are already enough to embed almost-spanning trees.

Lemma 2. *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 [14]. 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 vertices needing leaves with leftover vertices. We show that this is possible by verifying that a certain Hall-type condition typically holds.

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. In order to complete this forest into our spanning tree T, we need to join up distinguished "special 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 "super-regular pairs" (basically, dense subgraphs with edges very well-distributed in G). After embedding our small forest and making a number of further adjustments, we can find our desired special paths using super-regularity and a tool called the "blow-up lemma".

2 Proof of Theorem 1

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 later (these c_i will in turn determine c). Let $V = [n]$ be the common vertex set of G, R and the R_i.

Remark 3. At several points in the proof we will prove that (with high probability) there exist certain substructures in the random edges R, and we will assume by symmetry that these substructures (or some corresponding vertex sets) are themselves uniformly random. A simple way to see why this kind of reasoning is valid is to notice that if we apply a uniformly random vertex permutation to a random graph $R \in \mathbb{G}(n, p)$, then by symmetry the resulting distribution is still $\mathbb{G}(n, p)$. So we can imagine that we are finding substructures in an auxiliary random graph, then randomly mapping these structures to our vertex set.

2.1 Case 1: T has many leaves

For our first case suppose there are at least λn leaves in T (for some fixed $\lambda = \lambda(\alpha) > 0$ which will be determined in the second case). Then consider a tree T' with some λn leaves removed. By Lemma 2 we can a.a.s. embed T' into R_1 (abusing notation, we also denote this embedded subgraph by T'). Let $A \subseteq V$ be the set of vertices of T' which had a leaf deleted from them, and for $a \in A$ let $\ell(a) \leq \Delta$ be the number of leaves that were deleted from a. Let $B \subseteq V$ be the set of λn vertices not part of T'. In order to complete the embedding of T' into an embedding of T it suffices to find a set of disjoint stars in $G \cup R_2$ each with a center vertex $a \in A$ and with $\ell(a)$ leaves in B. We will prove the existence of such stars with the following Hall-type condition.

Lemma 4. *Consider a bipartite graph G with bipartition $A \cup B$, and consider a function $\ell : A \to \mathbb{N}$ such that $\sum_{a \in A} \ell(a) = |B|$. Suppose that*

$$|N(S)| \geq \sum_{s \in S} \ell(s) \quad \text{for all } S \subseteq A.$$

Then there is a set of disjoint stars in G, each of which has a center vertex $a \in A$ and $\ell(a)$ leaves in B.

Lemma 4 can be easily proved by applying Hall's marriage theorem to an auxiliary bipartite graph which has $\ell(a)$ copies of each vertex $a \in A$.

In this section, and at several points later in the paper, we will need to consider the intersection of random sets with fixed sets. The following concentration inequality (taken from [9, Theorem 2.10]) will be useful.

Lemma 5. *For a set Γ with $|\Gamma| = N$, a fixed subset $\Gamma' \subseteq \Gamma$ with $\Gamma' = m$, and a uniformly random n-element subset $\Gamma_n \subseteq \Gamma$, define the random variable $X = |\Gamma_n \cap \Gamma'|$. This random variable is said to have the* hypergeometric distribution *with parameters N,n and m. Noting that $\mathbb{E}X = mn/N$, we have*

$$\Pr(|X - \mathbb{E}X| \geq \varepsilon \mathbb{E}X) \leq \exp(c_\varepsilon \mathbb{E}X)$$

for some constant c_ε depending on ε.

Now, in accordance with Remark 3 we can assume A and B are uniformly random disjoint subsets of V of their size (note $|A| \geq \lambda n/\Delta$ and $|B| = \lambda n$), not depending on G. For each $a \in A$, the random variable $|N_G(a) \cap B|$ is hypergeometrically distributed with expected value at least $\lambda\alpha n$, and similarly each $|N_G(b) \cap A|$ ($b \in B$) is hypergeometric with expectation at least $\lambda\alpha n/\Delta$. Let $\beta = \lambda\alpha/(2\Delta)$; by Lemma 5 and the union bound, in G each $a \in A$ a.a.s. has at least βn neighbours in B, and each $b \in B$ a.a.s. has at least βn neighbours in A. That is to say, the bipartite graph induced by G on the bipartition $A \cup B$ has minimum degree at least βn. We treat A and B as fixed sets satisfying this property. It suffices to prove the following lemma.

Lemma 6. *For any $\beta > 0$ and $\Delta \in \mathbb{R}$ there is $c = c(\beta)$ such that the following holds. Suppose G is a bipartite graph with bipartition $A \cup B$ ($|A|, |B| \leq n$) and minimum degree at least βn. Suppose $\ell : A \to [\Delta]$ is some function satisfying $\sum_{a \in A} \ell(a) = |B|$. Let R be the random bipartite graph where each edge between A and B is present independently and with probability c/n. Then a.a.s. $G \cup R$ satisfies the condition in Lemma 4.*

Proof. Let $S \subseteq A$ be nonempty. If $|S| \leq \beta n/\Delta$ then $|N_G(S)| \geq |N_G(a)| \geq \beta n \geq \sum_{s \in S} \ell(s)$ for any $a \in S$. If $|S| \geq |A| - \beta n$ then $|A \backslash S| \leq |N_G(b)|$ for each $b \in B$, so $|N_G(S)| = |B| \geq \sum_{s \in S} \ell(s)$.

The remaining case is where $\beta n/\Delta \leq |S| \leq |A| - \beta n$. In this case $\sum_{s \in S} \ell(s) = |B| - \sum_{a \in A \backslash S} \ell(a) \leq |B| - \beta n$. Now, for any subsets $A' \subseteq A$ and $B' \subseteq B$ with $|A'| = \beta n/\Delta$ and $|B'| = \beta n$, the probability there is no edge between A' and B' is $(1 - c/n)^{(\beta n/\Delta)(\beta n)} \leq e^{-c\beta^2 n/\Delta}$. 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. If this is true, it follows that $|N_R(S)| \geq |B| - \beta n$, because if S had more than βn non-neighbours in B then this would give us a contradictory pair of subsets with no edge between them. We have proved that a.a.s. $|N_R(S)| \geq \sum_{s \in S} \ell(s)$ for all $S \subseteq A$, as required. $\square$

2.2 Case 2: T has few leaves

Now we address the second case where there are fewer than λn leaves in T. The argument is broken into a number of subsections, which we outline here. The first step (carried out in Section 2.2.1) is to divide G into a bounded number of pairs of "partner clusters" with edges well-distributed between them. Specifically, the edges between each pair of partner clusters will each have a property called "super-regularity". The significance of this is that one can use a tool called the "blow up lemma" to easily embed bounded-degree spanning structures into super-regular pairs.

Guided by our partition into pairs of clusters, we then embed most of T into $G \cup R$ in a step-by-step fashion. The objective is to do this in such a way that, afterwards, the remaining connected components of T can each be assigned to their own cluster-pair, so that we can finish the embedding in each cluster-pair individually, with the blow-up lemma.

As outlined in the introduction, the fact that T has few leaves means that it is mostly comprised of long *bare paths*: paths where every vertex has degree two. The first step in the embedding is to embed the non-bare-path parts of T using only the random edges in R. That is, we obtain a forest F by removing bare paths from T, then we embed F into R using Lemma 2. This step is performed in Section 2.2.2.

After embedding F, it remains to connect certain "special pairs" of vertices with paths of certain lengths in $G \cup R$, using the vertices that were not yet used to embed F. (That is, we need to embed the bare paths we deleted from T to obtain F). There are a few problems we need to overcome before we can apply the blow-up lemma to accomplish this. First, the vertices in special pairs, in general, lie in totally different clusters in our partition, so we cannot hope to embed each path in a single cluster-pair. In Section 2.2.3, we correct this problem by finding very short paths from the vertices of each special pair to a common cluster-pair.

The next issue is that the relative sizes of the cluster-pairs will not in general match the number of special pairs they contain. For example, if a cluster-pair contains N special pairs which we need to connect with paths of length k, then we need that cluster-pair to have exactly $(k-1)N$ vertices that were not used for anything so far. To fix this problem, in Section 2.2.4 we already start to embed some of the paths between special pairs, mainly using the random edges in R. We choose the vertices for these paths in a very specific way, to control the relative quantities of remaining vertices in the clusters.

After these adjustments, in Section 2.2.5 we are able to complete the embedding of T. We show that finding paths between distinguished vertices is equivalent to finding cycles with certain properties in a certain auxiliary graph. These cycles can be found in a straightforward manner using the blow-up lemma and super-regularity.

2.2.1 Partitioning into super-regular pairs

As outlined, we first need to divide G into a bounded number of pairs of "partner clusters" with well-distributed edges.

Definition 7. For a disjoint pair of vertex sets (X, Y) in a graph G, 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* in G 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 $\overline{h} = 2 - h$; if (V_1, V_2) is (ε, δ)-dense and moreover each $v \in V_h$ has at least $\delta \left|V_{\overline{h}}\right|$ neighbours in $V_{\overline{h}}$, then we say (V_1, V_2) is *(ε, δ)-super-regular.*

Lemma 8. *For $\alpha, \varepsilon > 0$ with ε sufficiently small relative to α, there are $\delta = \delta(\alpha) > 0$, $\rho = \rho(\alpha)$ and $Q = Q(\alpha, \varepsilon)$, such that the following holds. Let G be an n-vertex graph with minimum degree at least αn. Then there is $q \leq Q$ and a partition of $V(G)$ into clusters V_i^h ($1 \leq i \leq q$, $h = 1, 2$) such that each pair $\left(V_i^1, V_i^2\right)$ is (ε, δ)-super-regular. Moreover, each $\left|V_i^h\right| / \left|V_j^g\right| \leq \rho$.*

We emphasize that in Lemma 8 we do *not* guarantee that the sets V_i^h are of the same size, just that the variation in the sizes of the clusters is bounded independently of ε.

To prove Lemma 8, we will apply Szemerédi's regularity lemma to obtain a reduced *cluster graph*, then decompose this cluster graph into small stars. In each star T_i, the center cluster will give us V_i^1 and the leaf clusters will be combined to form V_i^2. We will then have to redistribute some of the vertices between the clusters to ensure super-regularity. Before giving the details of the proof, we give a statement of (a version of) Szemerédi's regularity lemma and some auxiliary lemmas for working with regularity and super-regularity.

Lemma 9 (Szemerédi's regularity lemma, minimum degree form)**.** *For every $\alpha > 0$, and any $\varepsilon > 0$ that is sufficiently small relative to α, there are $\alpha' = \alpha'(\alpha) > 0$ and $K = K(\varepsilon)$ such that the following holds. For any graph G of minimum degree at least $\alpha|G|$, there is a partition of $V(G)$ into clusters $V_0, V_1, \dots V_k$ ($k \le K$), and a spanning subgraph G' of G, satisfying the following properties. The "exceptional cluster" V_0 has size at most εn, and the other clusters have equal size sn. The minimum degree of G' is at least $\alpha' n$. There are no edges of G' within the clusters, and each pair of non-exceptional clusters is ε-regular in G' with density zero or at least α'. Moreover, define the* cluster graph *C as the graph whose vertices are the k non-exceptional clusters V_i, and whose edges are the pairs of clusters between which there is nonzero density in G'. The minimum degree of C is at least $\alpha' k$.*

This version of Szemerédi's regularity lemma is just the "degree form" of [10, Theorem 1.10], plus a straightforward claim about minimum degree. Our statement follows directly from [17, Proposition 9].

We now give some simple lemmas about (ε, δ)-denseness. Note that (ε, δ)-denseness is basically a one-sided version of ε-regularity that is more convenient in proofs about super-regularity. In particular, an ε-regular pair with density δ is $(\varepsilon, \delta - \varepsilon)$-dense. Here and in later sections, most of the theorems about (ε, δ)-dense pairs correspond to analogous theorems for ε-regular pairs with density about δ.

Lemma 10. *Consider disjoint vertex sets $V^1, V_1^2, \dots, V_r^2$, such that each V_i^2 is the same size and each $\left(V^1, V_i^2\right)$ is (ε, δ)-dense. Let $V^2 = \bigcup_{i=1}^r V_i^2$; then $\left(V^1, V^2\right)$ is $(\varepsilon, \delta/r)$-dense.*

Proof. Let $U^h \subseteq V^h$ with $\left|U^h\right| \ge \varepsilon\left|V^h\right|$. Let V_i^2 be the cluster which has the largest intersection with U^2, so we have $\left|U^2 \cap V_i^2\right| \ge \varepsilon\left|V^2\right|/r = \varepsilon\left|V_i^2\right|$, and there are therefore at least $\delta\left|U^1\right|\left|U^2 \cap V_i^2\right| \ge (\delta/r)\left|U^1\right|\left|U^2\right|$ edges between U^1 and U^2. □

Lemma 11. *Let $\left(V^1, V^2\right)$ be an (ε, δ)-super-regular pair. Suppose we have $W^h \supseteq V^h$ with $\left|W^h\right| \le (1 + f\varepsilon)\left|V^h\right|$, and suppose that each vertex in $W^h \backslash V^h$ has at least $\delta'\left|W^{\overline{h}}\right|$ neighbours in $W^{\overline{h}}$. Then $\left(W^1, W^2\right)$ is an (ε', δ')-super-regular pair, where*

$$\varepsilon' = \max\{2f, 1 + f\}\,\varepsilon, \quad \delta' = \min\{1/4, 1/(1 + f\varepsilon)\}\,\delta.$$

Proof. Suppose $U^h \subseteq W^h$ with $\left|U^h\right| \ge \varepsilon'\left|W^h\right|$. Then $\left|U^h \cap V^h\right| \ge \varepsilon'\left|W^h\right| - f\varepsilon\left|V^h\right| \ge \varepsilon\left|V^h\right|$ so there are at least $\delta\left|U^1 \cap V^1\right|\left|U^2 \cap V^2\right|$ edges between U^1 and U^2. But note that $\left|U^h \cap V^h\right| \ge \left|U^h\right| - f\varepsilon\left|W^h\right| \ge (1 - (f\varepsilon)/\varepsilon')\left|U^h\right| \ge (1/2)\left|U^h\right|$, so $d\left(U^1, U^2\right) \ge \delta/4$ and $\left(W^1, W^2\right)$ is (ε', δ')-dense.

Next, note that each $v \in V^h$ has $\delta\left|V^{\overline{h}}\right| \ge \delta'\left|W^{\overline{h}}\right|$ neighbours in $W^{\overline{h}}$, and by assumption each $v \in W^h \backslash V^h$ has $\delta'\left|W^{\overline{h}}\right|$ neighbours in $W^{\overline{h}}$, proving that $\left(W^1, W^2\right)$ is (ε', δ')-super-regular. □

Lemma 12. *Every (ε, δ)-dense pair $\left(V^1, V^2\right)$ contains a $(\varepsilon/(1-\varepsilon), \delta-\varepsilon)$-super-regular sub-pair $\left(W^1, W^2\right)$, where $\left|W^h\right| \geq (1-\varepsilon)\left|V^h\right|$.*

Lemma 12 follows easily from the same proof as [4, Proposition 6].

Now we prove Lemma 8. First we need a lemma about a decomposition into small stars.

Lemma 13. *Let G be an n-vertex graph with minimum degree at least αn. Then there is a spanning subgraph S which is a union of vertex-disjoint stars, each with at least two and at most $1 + 1/\alpha$ vertices.*

Proof. Let S be a union of such stars with the maximum number of vertices. Suppose there is a vertex v uncovered by S. If v has a neighbour which is a center of one of the stars in S, and that star has fewer than $1 + \lfloor 1/\alpha \rfloor$ vertices, then we could add v to that star, contradicting maximality. (Here we allow either vertex of a 2-vertex star to be considered the "center"). Otherwise, if v has a neighbour w which is a leaf of one of the stars in S, 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 center of a star with $1 + \lfloor 1/\alpha \rfloor > 1/\alpha$ vertices. But these stars would comprise more than n vertices, which is again a contradiction. We conclude that S covers G, as desired. □

Proof of Lemma 8. Apply our minimum degree form of Szemerédi's regularity lemma, with some ε' to be determined (which we will repeatedly assume is sufficiently small). Consider a cover of the cluster graph C by stars $T_1, \dots, T_q$ of size at most $1 + 1/\alpha'$, as guaranteed by Lemma 13. Let W_i^1 be the center 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 "center"). Each edge of the cluster graph C corresponds to an $(\varepsilon', \alpha'/2)$-dense pair, and each W_i^h is the union of at most $1/\alpha'$ clusters V_i. By Lemma 10, with $\delta' = (\alpha')^2/2$ each pair $\left(W_i^1, W_i^2\right)$ is (ε', δ')-dense in G' (therefore G).

Apply Lemma 12 to obtain sets $V_i^h \subseteq W_i^h$ such that $\left|V_i^h\right| = (1-\varepsilon')\left|W_i^h\right|$ and each $\left(V_i^1, V_i^2\right)$ is $(2\varepsilon', \delta'')$-super-regular, for $\delta'' = \delta'/2$. Combining the exceptional cluster V_0 and all the $W_i^h \backslash V_i^h$, there are at most $2\varepsilon' n$ "bad" vertices which are not part of a super-regular pair.

Each bad vertex v has at least $\alpha' n - 2\varepsilon' n$ neighbours to the clusters V_i^h. The clusters V_i^h which contain fewer than $\delta''\left|V_i^h\right|$ of these neighbours, altogether contain at most $\delta'' n$ such neighbours. Each V_i^h has size at most sn/α', so for small ε' there are at least

$$\frac{\alpha' n - 2\varepsilon' n - \delta'' n}{sn/\alpha'} \geq \left(\alpha'\right)^2/(2s)$$

clusters V_i^h which have at least $\delta''\left|V_i^h\right|$ neighbours of v. Pick one such V_i^h uniformly at random, and put v in $V_i^{\overline{h}}$ (do this independently for each bad v). By a concentration inequality, after this procedure a.a.s. at most $2(2\varepsilon' n)/\left(\left(\alpha'\right)^2/(2s)\right) \leq \left(8\varepsilon'/(\alpha')^2\right)\left|W_i^h\right| \leq \left(9\varepsilon'/(\alpha')^2\right)\left|V_i^h\right|$ bad vertices have been added to each V_i^h. By Lemma 11, for small ε' each $\left(V_i^1, V_i^2\right)$ is now $(O(\varepsilon'), \delta''/4)$-super-regular. To conclude, let $\delta = \delta''/4$ and $\rho = 2\alpha'$, and choose some ε' small relative to ε. □

We apply Lemma 8 to our graph G, with $\varepsilon_1 = \varepsilon_1(\delta)$ to be determined. For $\mathbf{i} = (i, h)$, let $V_{\mathbf{i}} = V_i^h$, and let $|V_{\mathbf{i}}| = s_{\mathbf{i}} n$.

2.2.2 Embedding a subforest of T

In this section we will start to embed T into $G \cup R$. Just as in Section 2.1, we will use the decomposition of R into "phases" R_1, R_2, R_3, R_4 (but we start with R_2 for consistency with the section numbering).

Recall that a *bare path* in T is a path P such that every vertex of P has degree exactly two in 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 due to one of the authors ([14, Lemma 2.1]).

Lemma 14. *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 $\ell = \lambda n$. If we choose k large enough and choose λ small enough relative to k, then T contains a collection of $n/(2(k-1))$ disjoint bare k-paths. (The definite value of k will be determined later; it will be odd and depend on α but not ε). If we delete the interior vertices of these paths, then we are left with a forest F on $n/2$ vertices.

Now, embed F into the random graph R_2 (also, with some abuse of notation, denote this embedded subgraph by F). There are $n/(2(k-1))$ "special pairs" of "special vertices" of $F \subseteq R_2$ that need to be connected with k-paths of non-special vertices. We call such connecting paths "special paths". Let $X \subseteq V$ be the set of special vertices and let $W = V \backslash F$ be the set of "free" vertices. By symmetry we can assume

$$X \cup F\backslash X \cup W$$

is a uniformly random partition of V into parts of sizes $n/(k-1)$, $n/2 - n/(k-1)$ and $n/2$ respectively. We can also assume that the special pairs correspond to a uniformly random partition of X into pairs. Note that $2|W| = (k-1)|X|$.

Let $X_{\mathbf{i}} = V_{\mathbf{i}} \cap X$ and $W_{\mathbf{i}} = V_{\mathbf{i}} \backslash F$ be the set of free vertices remaining in $V_{\mathbf{i}}$. By Lemma 5 and the union bound, a.a.s. each $|W_{\mathbf{i}}| \sim s_{\mathbf{i}} n/2$ and $|X_{\mathbf{i}}| \sim s_{\mathbf{i}} n/(k-1)$.

Here and in future parts of the proof, it is critical that after we take certain subsets of our super-regular pairs, we maintain super-regularity or at least density (albeit with weaker ε and δ). We will ensure this in each situation by appealing to one of the following two lemmas.

Lemma 15. *Suppose $\left(V^1, V^2\right)$ is (ε, δ)-dense, and let $W^h \subseteq V^h$ with $\left|W^h\right| \geq \gamma \left|V^h\right|$. Then $\left(W^1, W^2\right)$ is $(\varepsilon/\gamma, \delta)$-dense.*

Proof. If $U^h \subseteq W^h$ with $\left|U^h\right| \geq (\varepsilon/\gamma)\left|W^h\right|$, then $\left|U^h\right| \geq \varepsilon \left|V^h\right|$. The result follows from the definition of (ε, δ)-denseness. $\square$

Lemma 16. *Fix $\delta, \varepsilon' > 0$. There is $\varepsilon = \varepsilon(\varepsilon') > 0$ such that the following holds. Suppose $\left(V^1, V^2\right)$ is (ε, δ)-dense, let n_h satisfy $\left|V^h\right| \geq n_h = \Omega(n)$ (not necessarily uniformly over ε and δ) and for each h let $W^h \subseteq V^h$ be a uniformly random subset of size n_h. Then $\left(W^1, W^2\right)$ is a.a.s. $(\varepsilon', \delta/2)$-dense. If moreover $\left(V^1, V^2\right)$ was (ε, δ)-super-regular then $\left(W^1, W^2\right)$ is a.a.s. $(\varepsilon', \delta/2)$-super-regular.*

Lemma 16 is a simple consequence of a highly nontrivial (and much more general) result proved by Gerke, Kohayakawa, Rödl and Steger in [8], which shows that small random subsets inherit certain regularity-like properties with very high probability. The notion of "lower regularity" in that paper essentially corresponds to our notion of (ε, δ)-density.

Proof of Lemma 16. The fact that $\left(W^1, W^2\right)$ is a.a.s. $(\varepsilon', \delta/2)$-dense for suitable ε, follows from two applications of [8, Theorem 3.6].

It remains to consider the degree condition for super-regularity. For each $v \in W^h$, the number of neighbours of v in $W^{\overline{h}}$ is hypergeometrically distributed, with expected value at least $\delta n_{\overline{h}}$. Since $n_{\overline{h}} = \Omega(n)$, Lemma 5 and the union bound immediately tells us that a.a.s. each such v has $\delta n_{\overline{h}}/2$ such neighbours, as required. $\square$

Let $\overline{(i,h)} = \left(i, \overline{h}\right)$, and let $[(i,h)] = i$. It is an immediate consequence of Lemma 16 that each $\left(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ and $\left(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ are $(\varepsilon_2, \delta_2)$-super-regular, where $\delta_2 = \delta_1/2$ and ε_2 can be made arbitrarily small by choice of ε_1.

Now that we have embedded F, we no longer care about the vertices used to embed $F \backslash X$; 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, update n to be the number of vertices not used to embed $F \backslash X$ (previously $(1/2 + 1/(k-1))n$), and update $s_{\mathbf{i}}$ to satisfy $|W_{\mathbf{i}}| = s_{\mathbf{i}} n$. Note that (asymptotically speaking) this just rescales the $s_{\mathbf{i}}$ so that they sum to $((k-1)/(k+1))$. Therefore the variation between the $s_{\mathbf{i}}$ is still bounded independently of ε; we can increase ρ slightly so that each $|s_{\mathbf{i}}|/|s_{\mathbf{j}}| \le \rho$. We then have $|X_{\mathbf{i}}| \sim 2 s_{\mathbf{i}} n/(k-1)$ for each $\mathbf{i}$, and there are $n/(k+1)$ special pairs.

2.2.3 Fixing the endpoints of the special paths

We will eventually need to start finding and removing special paths between our special pairs. For this, we would like each of the special pairs to be "between" partner clusters $V_{\mathbf{i}}, V_{\bar{\mathbf{i}}}$, so that we can take advantage of the super-regularity in G. Therefore, for every special pair $\{x, y\}$, we will find very short (length-2) paths from x to some vertex $x' \in W_{\mathbf{r}}$ and from y to some $y' \in W_{\bar{\mathbf{r}}}$, for some $\mathbf{r}$. All these short paths will be disjoint. Our short paths effectively "move" the special pair $\{x, y\}$ to $\{x', y'\}$: to complete the embedding we now need to find length-$(k-4)$ special paths connecting the new special pairs.

Note that there is a.a.s. an edge between any two large vertex sets in the random graph R_3, as formalized below.

Lemma 17. *For any $\varepsilon > 0$, there is $c = c(\varepsilon)$ such that the following holds. In a random graph $R \in \mathbb{G}(n, c/n)$, there is a.a.s. an edge between any two (not necessarily disjoint) vertex subsets A, B of size εn.*

Lemma 17 is a standard result and can be proved in basically the same way as Lemma 6. We include a proof for completeness.

Proof. For any such subsets A, B, choose disjoint $A' \subseteq A$ and $B' \subseteq B$ of size $\varepsilon n/2$. The probability there is no edge between A and B is less than the probability there is no edge between A' and B', which is $(1 - c/n)^{(\varepsilon n/2)^2} \le e^{-c\varepsilon^2 n/2}$. There are at most 2^{2n} choices of A and B so for large c, by the union bound there is a.a.s. an edge between any such pair of sets. $\square$

It is relatively straightforward to greedily find suitable short paths in $G \cup R_3$ using the minimum degree condition in the super-regular pairs $\left(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ and Lemma 17. However we need to be careful to find and remove our paths in such a way that afterwards we still have super-regularity between certain subsets. We will accomplish this by setting aside a random subset of each $W_{\mathbf{r}}$ from which the x', y' (and no other vertices in our length-2 paths) will be chosen, then using the symmetry argument of Remark 3 and appealing to Lemma 16. The details are a bit involved, as follows.

Uniformly at random partition each $W_{\mathbf{i}}$ into sets $W_{\mathbf{i}}^{(1)}$ and $W_{\mathbf{i}}^{(2)}$ of size $|W_{\mathbf{i}}|/2$. We will take the x',y' from the $W_{\mathbf{i}}^{(2)}$s and we will take the intermediate vertices in our length-2 paths from the $W_{\mathbf{i}}^{(1)}$s. Note that in G, a.a.s. each $x \in X_{\mathbf{i}}$ has at least $(\delta_2/3)s_{\bar{\mathbf{i}}}n$ neighbours in $W_{\mathbf{i}}^{(1)}$. This follows from Lemma 5 applied to the hypergeometric random variables $\left|N_G(x) \cap W_{\bar{\mathbf{i}}}^{(1)}\right|$ (which have mean at least $\delta_2 s_{\bar{\mathbf{i}}}n/2$ by super-regularity), plus the union bound. Arbitrarily choose an ordering of the special pairs, and choose some "destination" index $\mathbf{r}$ for each special pair, in such a way that each $\mathbf{r}$ is chosen for a $(1+o(1))/(2q)$ fraction of the special pairs. (Recall that q is the number of pairs of partner clusters). We will find our length-2 paths greedily, by repeatedly applying the following argument.

Suppose we have already found length-2 paths for the first $i-1$ special pairs, and let S be the set of vertices in these paths. Let $\{x, y\}$ be the ith special pair, with $x \in X_{\mathbf{i}}$ and $y \in X_{\mathbf{j}}$. Let $\mathbf{r}$ be the destination index for this special pair. We claim that if k is large then $\left|N_G(x) \cap \left(W_{\bar{\mathbf{i}}}^{(1)} \backslash S\right)\right| = \Omega(n)$ and $\left|W_{\mathbf{r}}^{(2)} \backslash S\right| = \Omega(n)$. It will follow from Lemma 17 that there is a.a.s. an edge between $N_G(x) \cap \left(W_{\bar{\mathbf{i}}}^{(1)} \backslash S\right)$ and $W_{\mathbf{r}}^{(2)} \backslash S$ in R_3. This gives a length-2 path between x and some $x' \in W_{\mathbf{r}}^{(2)}$ disjoint to the paths so far, in $G \cup R_3$. By identical reasoning there is a disjoint length-2 path between y and some $y' \in W_{\bar{\mathbf{r}}}^{(2)}$ passing through $W_{\bar{\mathbf{j}}}^{(1)}$.

To prove the claims in the preceding argument, note that we always have $\left|W_{\bar{\mathbf{i}}}^{(1)} \cap S\right| \le |X_{\mathbf{i}}|$ and $\left|W_{\mathbf{r}}^{(2)} \cap S\right| \le (1+o(1))|X|/(2q)$. Let $s = ((k-1)/(k+1))/2q$ be the average $s_{\mathbf{i}}$, and recall that ρ controls the relative sizes of the $s_{\mathbf{i}}$. For large k,

$$\left|N_G(x) \cap \left(W_{\bar{\mathbf{i}}}^{(1)} \backslash S\right)\right| \ge \frac{\delta_2 s_{\bar{\mathbf{i}}}}{3}n - |X_{\mathbf{i}}| \sim \frac{\delta_2 s_{\bar{\mathbf{i}}}}{3}n - \frac{2s_{\mathbf{i}}}{k-1}n \ge \left(\frac{\delta_2}{3} - \frac{2\rho}{k-1}\right)s_{\bar{\mathbf{i}}}n = \Omega(n),$$

and

$$\left|W_{\mathbf{r}}^{(2)} \backslash S\right| \ge \frac{|W_{\mathbf{i}}|}{2} - (1+o(1))\frac{|X|}{2q} \sim \frac{s_{\mathbf{i}}}{2}n - \frac{2s}{k-1}n \ge \left(\frac{1}{2} - \frac{2\rho}{k-1}\right)s_{\mathbf{i}}n = \Omega(n),$$

as required. We emphasize that since δ_2 and ρ are independent of ε (depending only on α), we can also choose k independent of ε.

After we have found our length-2 paths, let $X_{\mathbf{i}}' \subseteq W_{\mathbf{i}}$ be the set of "new special vertices" x', y' in $W_{\mathbf{i}}$. Note that we can in fact assume that each $X_{\mathbf{i}}'$ is a uniformly random subset of $W_{\mathbf{i}}^{(2)}$ of its size (which is $|X|/(2q)$). This can be seen by a symmetry argument, as per Remark 3. Because here the situation is a bit complicated, we include some details. Condition on the random sets $W_{\mathbf{i}}^{(2)}$. For each $\mathbf{i}$, the vertices in $X_{\mathbf{i}}'$ are each chosen so that they are adjacent in R_3 to some vertex in one of the $W_{\mathbf{j}}^{(1)}$. Note that the distribution of R_3 is invariant under permutations of $W_{\mathbf{i}}^{(2)}$. We can choose such a vertex permutation π uniformly at random; then $\pi(X_{\mathbf{i}}')$ is a uniformly random subset of $W_{\mathbf{r}}^{(2)}$ of its size, which has the right adjacencies in $\pi(R_3)$ (which has the same distribution as R_3).

Now we claim that even after removing the vertices of our length-2 paths we a.a.s. maintain some important properties of the clusters. Let $W'_{\mathbf{i}} \subseteq W_{\mathbf{i}} \backslash X'_{\mathbf{i}}$ be the subset of $W_{\mathbf{i}}$ that remains after the vertices in the length-2 paths are deleted.

Claim 18. *Each* $\left(X'_{\mathbf{i}}, W'_{\bar{\mathbf{i}}}\right)$ *and* $\left(W'_{\mathbf{i}}, W'_{\bar{\mathbf{i}}}\right)$ *are a.a.s.* $(\rho(k-1)\varepsilon_2/(1-\delta_2/2), \delta_2/2)$*-super-regular, satisfying*

$$(2\rho)^{-1} \leq \frac{k-1}{2} \cdot \frac{|X'_{\mathbf{i}}|}{|W'_{\mathbf{i}}|} \leq 2\rho.$$

Proof. Since each $W_{\mathbf{i}}^{(2)}$ is a uniformly random subset of $W_{\mathbf{i}}$, we can assume each $X'_{\mathbf{i}}$ is a uniformly random subset of $W_{\mathbf{i}}$ of its size. So, each $|N_G(w) \cap X'_{\mathbf{i}}|$ is hypergeometrically distributed with mean at least $\delta_2|X'_{\mathbf{i}}|$. By Lemma 5 and the union bound, in G a.a.s. each $w \in W_{\bar{\mathbf{i}}}$ has at least $(\delta_2/2)|X'_{\mathbf{i}}|$ neighbours in $X'_{\mathbf{i}}$. Next, note that each $|X'_{\mathbf{i}}| \sim 2sn/(k-1)$. Since $s/s_{\bar{\mathbf{i}}} \geq 1/\rho$, we have $\rho(k-1)|X'_{\mathbf{i}}| \geq |W_{\bar{\mathbf{i}}}|$ and by Lemma 15 each $(X'_{\mathbf{i}}, W_{\bar{\mathbf{i}}})$ is $(\rho(k-1)\varepsilon_2, \delta_2)$-dense. It follows that each $(X'_{\mathbf{i}}, W_{\bar{\mathbf{i}}})$ is $(\rho(k-1)\varepsilon_2, \delta_2/2)$-super-regular.

Now, for large k note that $|W_{\mathbf{i}} \backslash W'_{\mathbf{i}}| \leq (\delta_2/2)|W_{\mathbf{i}}|$. Indeed, recalling the choice of vertices for the length-2 paths, note that

$$\frac{|W_{\mathbf{i}} \backslash W'_{\mathbf{i}}|}{|W_{\mathbf{i}}|} = \frac{|X_{\bar{\mathbf{i}}}| + |X|/2q}{|W_{\mathbf{i}}|} \sim \frac{\frac{2}{k-1}(s_{\bar{\mathbf{i}}} + s)}{s_{\mathbf{i}}} \leq \frac{4\rho}{k-1},$$

which is smaller than $\delta_2/2$ for large k. For $v \in W_{\mathbf{i}}$ we have $|N_G(v) \cap W_{\bar{\mathbf{i}}}| \geq \delta_2 |W_{\bar{\mathbf{i}}}|$ by super-regularity, so

$$\left|N_G(v) \cap W'_{\bar{\mathbf{i}}}\right| \geq |N_G(v) \cap W_{\bar{\mathbf{i}}}| - |W_{\mathbf{i}} \backslash W'_{\mathbf{i}}| \geq (\delta_2/2)|W_{\bar{\mathbf{i}}}| \geq (\delta_2/2)\left|W'_{\bar{\mathbf{i}}}\right|.$$

Combining this with Lemma 15, a.a.s. each $\left(W'_{\mathbf{i}}, W'_{\bar{\mathbf{i}}}\right)$ is $(\varepsilon_2/(1-\delta_2/2), \delta_2/2)$-super-regular and each $\left(X'_{\mathbf{i}}, W'_{\bar{\mathbf{i}}}\right)$ is $(\rho(k-1)\varepsilon_2/(1-\delta_2/2), \delta_2/2)$-super-regular.

Finally, the claims about the relative sizes of the $W'_{\mathbf{i}}$, $X'_{\mathbf{i}}$ for large k are simple consequences of the asymptotic expressions

$$|W'_{\mathbf{i}}| \sim s_{\mathbf{i}} n - \frac{2}{k-1}(s_{\bar{\mathbf{i}}} + s)n, \quad |X'_{\mathbf{i}}| \sim \frac{2s}{k-1} n. \qquad \square$$

Now we can redefine each $X_{\mathbf{i}}$ to be the set of new special vertices taken from $W_{\mathbf{i}}$, and we can update the $W_{\mathbf{i}}$ by removing all the vertices used in our length-2 paths (that is, set $X_{\mathbf{i}} = X'_{\mathbf{i}}$ and $W_{\mathbf{i}} = W'_{\mathbf{i}}$). All the special pairs are between some $X_{\mathbf{i}}$ and $X_{\bar{\mathbf{i}}}$; let $Z_{[\mathbf{i}]} = Z_{[\bar{\mathbf{i}}]}$ be the set of such pairs (recall that $[(i,h)] = i$). Update the sets $X = \bigcup_{\mathbf{i}} X_{\mathbf{i}}$ and $W = \bigcup_{\mathbf{i}} W_{\mathbf{i}}$, and let $Z = \bigcup_i Z_i$ be the set of all special pairs, so $|Z| = |X|/2$.

We also update all the other variables as before: update n to be the number of vertices in $X \cup W$ (previously $n - 4n/(k+1)$), and redefine the $s_{\mathbf{i}}$ to still satisfy $|W_{\mathbf{i}}| = s_{\mathbf{i}} n$. Also, update k to be the new required length of special paths (previously $k-4$), so we still have $2|W| = (k-1)|X|$.

Finally, let $\varepsilon_3 = \rho(k-1)\varepsilon_2/(1-\delta_2/2)$ and $\delta_3 = \delta_2/2$ and double the value of ρ. By Claim 18, each $(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}})$ and $(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}})$ are then $(\varepsilon_3, \delta_3)$-super-regular, and each

$$\rho^{-1} \leq \frac{k-1}{2} \cdot \frac{|X_{\mathbf{i}}|}{|W_{\mathbf{i}}|} \leq \rho.$$

2.2.4 Adjusting the relative sizes of the clusters

The next step is to use the random edges in R_4 to start removing special paths. Our objective is to do this in such a way that after we have removed the vertices in those paths from the $X_{\mathbf{i}}$ and $W_{\mathbf{i}}$, we will have $2|W_{\mathbf{i}}| = (k-1)|X_{\mathbf{i}}|$ for each $\mathbf{i}$. We require this to apply the blow-up lemma in the final step.

The way we will remove special paths is with Claim 19, to follow.

A "special sequence" is a sequence of vertices $x, w_1, \dots, w_{k-1}, y$, such that $\{x, y\}$ is a special pair and each w_j is in some $W_{\mathbf{i}}$. For any special path we might find, the sequence of vertices in that path will be special. In fact, the special sequences are precisely the allowed sequences of vertices in special paths.

We also define an equivalence relation on special sequences: two special sequences $x, w_1, \dots, w_{k-1}, y$ and and $x', w'_1, \dots, w'_{k-1}, y'$ are equivalent if x and x' are in the same cluster $X_{\mathbf{i}}$ (therefore $y, y' \in X_{\bar{\mathbf{i}}}$), and if w_j and w'_j are in the same cluster $W_{\mathbf{i}_j}$ for all j. A "template" is an equivalence class of special sequences, or equivalently a sequence of clusters $X_{\mathbf{i}}, W_{\mathbf{i}_1}, \dots, W_{\mathbf{i}_{k-1}}, X_{\bar{\mathbf{i}}}$ (repetitions permitted).

The idea is that we can specify (a multiset of) desired templates, and provided certain conditions are satisfied we can find special paths with these templates, using the remaining vertices in $G \cup R_4$. Say a special sequence $x, w_1, \dots, w_{k-1}, y$ with $x \in X_{\mathbf{i}}$ and $y \in X_{\bar{\mathbf{i}}}$ is "good" if $w_1 \in W_{\bar{\mathbf{i}}}$ and $w_{k-1} \in W_{\mathbf{i}}$. We make similar definitions for good special paths and good templates. It is easier to find good special paths because we can take advantage of super-regularity at the beginning and end of the paths.

Recall that $R_4 \in \mathbb{G}(n, c_4/n)$; we still have the freedom to choose large c_4 to make it possible to find our desired special paths in $G \cup R_4$. Recall from previous sections that s and ρ control the sizes of the clusters, that Q controls the number of clusters, and that each $\left(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ and $\left(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ are $(\varepsilon_3, \delta_3)$-super-regular, where we can make ε_3 as small as we like relative to α.

Claim 19. *For any $\gamma > 0$, if ε_3 is sufficiently small relative to γ and δ_3, there is $c_4 = c_4(s, \rho, \delta_3, \gamma, Q, k)$ such that the following holds.*

Suppose we specify a multiset of templates (each of which we interpret as sequences of clusters), in such a way that each $X_{\mathbf{i}}$ (respectively, each $W_{\mathbf{i}}$) appears in these templates at most $(1-\gamma)|X_{\mathbf{i}}|$ times (respectively, at most $(1-\gamma)|W_{\mathbf{i}}|$ times). Then, we can a.a.s. find special paths in $G \cup R_4$ with our desired templates in such a way that after the removal of their vertices, each $\left(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ and $\left(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\right)$ are still $(\varepsilon_3/\gamma, \delta_3/4)$-super-regular.

Note that our condition on the templates says precisely that at least a γ-fraction of each $X_{\mathbf{i}}$ and $W_{\mathbf{i}}$ should remain after deleting the vertices of the special paths.

We will prove Claim 19 with a simpler lemma.

Lemma 20. *For any $\delta, \xi > 0$ and any integer $k > 2$, there are $\varepsilon(\delta, \xi) > 0$ and $c = c(\delta, \xi, k)$ such that the following holds.*

Let G be a graph on the vertex set $[(k+1)n]$, together with a vertex partition into disjoint clusters $X, W_1 \dots, W_{k-1}, Y$, such that $|X| = |Y| = n$ and each $|W_i| = n$. Suppose that (X, W_1) and (W_{k-1}, Y) are (ε, δ)-dense and that each $x \in X$ is bijectively paired with a vertex $y \in Y$, comprising *a* special pair (x, y). *Let* $R \in \mathbb{G}((k+1)n, c/n)$. *Then, in* $G \cup R$ *we can a.a.s. find* $(1-\xi)n$ *vertex-disjoint paths running through* $X, W_1, \ldots, W_{k-1}, Y$ *in that order, each of which connects the vertices of a special pair (these are* special paths*).*

Proof of Lemma 20. If c is large, in R we can a.a.s. find an edge between any subsets $W_i' \subseteq W_i$ and $W_{i+1}' \subseteq W_{i+1}$ with $|W_i'|, \left|W_{i+1}'\right| \geq \delta\xi n/(k-2)$, for any $1 \leq i < k-1$. We can prove this fact with basically the same argument as in the proof of Lemma 17. In fact, it immediately follows that a.a.s. between any such W_i' and W_{i+1}' there is a matching with more than $\min\left\{|W_i'|, \left|W_{i+1}'\right|\right\} - \delta\xi n/(k-2)$ edges, in R. This is because in a maximal matching, the subsets of unmatched vertices in W_i' and W_{i+1}' have no edge between them.

It follows that (a.a.s.) if we choose any subsets $W_i' \subseteq W_i$ $(1 \leq i \leq k-1)$ with $|W_1'| = \cdots = \left|W_{k-1}'\right| = \delta\xi n$, there is a path in R running through the $W_1', \ldots, W_{k-1}'$ in that order. Indeed, we have just proven there is a matching of size greater than $|W_i'| - \delta\xi n/(k-2)$ between each W_i' and W_{i+1}'; the union of such matchings must include a suitable path.

Now, consider a ξn-element subset Z' of the special pairs, and consider a ξn-element subset W_i' of each W_i. We would like to find a special path using one of the special pairs in Z and a vertex from each W_i'. If we could (a.a.s.) find such a special path for all such subsets Z' and W_i' then we would be done, for we would be able to choose our desired $(1-\xi)n$ special paths greedily. That is, given a collection of fewer than $(1-\xi)n$ disjoint special paths running through the clusters, we could find an additional disjoint special path with the leftover vertices.

Let $X' \subseteq X$ and $Y' \subseteq Y$ be the subsets containing the vertices in the special pairs in Z'. By (ε, δ)-denseness (with $\varepsilon \leq \xi$) there are at most εn vertices in X' (respectively, in Y') with fewer than $\delta\xi n$ neighbours in W_1' (respectively, in W_{k-1}'), in G. So, if $\varepsilon < \xi/2$, then there must be a special pair (x, y) such that $|N_G(x) \cap W_1'|, \left|N_G(y) \cap W_{k-1}'\right| \geq \delta\xi n$. By the discussion at the beginning of the proof, there is a.a.s. a path in R running through $N_G(x) \cap W_1', W_2', \ldots, W_{k-2}', N_G(y) \cap W_{k-1}'$ in that order. Combining this path with x and y gives us a special path between x and y, as desired. □

A corollary of Lemma 20 is that we can find special paths corresponding to a single template:

Corollary 21. *For any* $b, \delta, \xi > 0$ *and any integer* $k > 2$*, there are* $\varepsilon(\delta, \xi) > 0$ *and* $c = c(b, \delta, \xi, k)$ *such that the following holds.*

Let G *be a graph on some vertex set* $[N]$ $(n \geq bN)$*, together with a vertex partition into clusters. Consider a sequence of clusters* $X, W_1, \ldots, W_{k-1}, Y$ *(there may be repetitions in the* W_i*s), such that* $|X|, |Y| \geq n$ *and if* W_i *appears* t_i *times in the sequence then* $|W_i| \geq t_i n$*. Suppose that* (X, W_1) *and* (W_{k-1}, Y) *are* (ε, δ)*-dense, and 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, in* $G \cup R$ *we can a.a.s. find* $(1-\xi)n$ *vertex-disjoint paths running through* $X, W_1, \ldots, W_{k-1}, Y$ *in that order, each of which connects the vertices of a special pair (these are* special paths*).*

Proof. Uniformly at random choose t_i disjoint n-vertex subsets of each W_i. This gives a total of $k-1$ disjoint subsets; arrange these into a sequence $W_1', \ldots, W_{k-1}'$ in such a way that $W_i' \subseteq W_i$. Also uniformly at random choose n-vertex subsets $X' \subseteq X$ and of $Y' \subseteq Y$. By Lemma 16, (X', W_1') and $\left(W_{k-1}', Y'\right)$ are a.a.s. (ε', δ)-dense, where we can make ε' small by choice of ε. We can therefore directly apply Lemma 20. □

We can now prove Claim 19.

Proof of Claim 19. The basic idea is to split the clusters into many random subsets, one for each good template, and to apply Corollary 21 many times. (Note that there are fewer than $Q^k = O(1)$ different templates). We also need to guarantee super-regularity in what remains. For this, we set aside another random subset of each cluster, which we will not touch (we used a similar idea in Section 2.2.3).

If we are able to successfully find our special paths, those with template τ will contain some number $t_{\mathbf{i}}^{\tau}|W_{\mathbf{i}}|$ of vertices from $W_{\mathbf{i}}$. (Note that $t_{\mathbf{i}}^{\tau}$ depends only on the multiset of desired templates, and not on the specific special paths we find). After removing the vertices in our found special paths, there will be $l_{\mathbf{i}}|W_{\mathbf{i}}| = (1 - \sum_{\tau} t_{\mathbf{i}}^{\tau})|W_{\mathbf{i}}| \geq \gamma|W_{\mathbf{i}}|$ vertices remaining in $W_{\mathbf{i}}$. Similarly, the special paths with good template τ will take some number $s_i^{\tau}|Z_i|$ of special pairs from each Z_i, leaving $r_i|Z_i| \geq \gamma|Z_i|$ special pairs.

Now, for some small $\xi > 0$ to be determined, we want to partition each $W_{\mathbf{i}}$ into parts $W_{\mathbf{i}}^{\tau}$ of size at least $t_{\mathbf{i}}^{\tau}|W_{\mathbf{i}}|/(1-\xi)$, and a "leftover" part $W_{\mathbf{i}}'$ of size $|W_{\mathbf{i}}| - \sum_{\tau}|W_{\mathbf{i}}^{\tau}|$. Similarly, we want to partition each Z_i into parts Z_i^{τ} of size at least $s_i^{\tau}|Z_i|/(1-\xi)$ and a leftover part Z_i'. Let $X_{\mathbf{i}}^{\tau} \subseteq X_{\mathbf{i}}$ and $X_{\mathbf{i}}' \subseteq X_{\mathbf{i}}$ contain the special vertices in the special pairs in the $Z_{[\mathbf{i}]}^{\tau}$ and $Z_{[\mathbf{i}]}'$ respectively.

Choose the sizes of the $W_{\mathbf{i}}^{\tau}$ and Z_i^{τ} such that each $|W_{\mathbf{i}}^{\tau}| \geq \xi sn/Q^k$ and $|Z_i^{\tau}| \geq \xi sn/\big((k-1)Q^k\big)$, and also each $|W_{\mathbf{i}}'| \geq (l_{\mathbf{i}}/2)|W_{\mathbf{i}}|$ and $|Z_i'| \geq (r_i/2)|Z_i|$ (this will be possible if ξ is small enough to satisfy $(1-\gamma)/(1-\xi) + \xi \leq (1-\gamma/2)$). Given these part sizes choose our partitions uniformly at random. (Note that this means the marginal distribution of each part is uniform on subsets of its size).

By Lemma 16, for any $\varepsilon' > 0$ we can choose ε_3 such that each $\left(X_{\mathbf{i}}^{\tau}, W_{\bar{\mathbf{i}}}^{\tau}\right)$ is a.a.s. $(\varepsilon', \delta_3/2)$-dense. Moreover a.a.s. each $w \in W_{\bar{\mathbf{i}}}$ has at least $(\delta_3/2)|W_{\mathbf{i}}'|$ neighbours in $W_{\mathbf{i}}'$ and at least $(\delta_3/2)|X_{\mathbf{i}}'|$ neighbours in $X_{\mathbf{i}}'$, and each $x \in X_{\mathbf{i}}$ has at least $(\delta_3/2)\left|W_{\bar{\mathbf{i}}}'\right|$ neighbours in $W_{\bar{\mathbf{i}}}'$. We can find our desired special pairs with good template τ using the clusters $X_{\mathbf{i}}^{\tau}$, $W_{\mathbf{i}}^{\tau}$ and Corollary 21, provided ε' is small (in the notation of Corollary 21, we need $\varepsilon' \leq \varepsilon(\delta_3/2, \xi)$).

After deleting the vertices in these special paths, by Lemma 15 each $\big(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ and $\big(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ are $(\varepsilon_3/\gamma, \delta_3)$-dense. By the condition on the sizes of the $W_{\mathbf{i}}'$s and Z_i's and the bounds on the neighbourhood sizes in the last paragraph, each $w \in W_{\bar{\mathbf{i}}}$ now has at least $(\delta_3/4)|W_{\mathbf{i}}|$ neighbours in $W_{\mathbf{i}}$ and at least $(\delta_3/4)|X_{\mathbf{i}}|$ neighbours in $X_{\mathbf{i}}$, and each $x \in X_{\mathbf{i}}$ has at least $(\delta_3/4)\big|W_{\bar{\mathbf{i}}}\big|$ neighbours in $W_{\bar{\mathbf{i}}}$. So each $\big(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ and $\big(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ are $(\varepsilon_3/\gamma, \delta_3/4)$-super-regular, as required. $\square$

Having proved Claim 19, we now just need to show that there are suitable good templates so that after finding special paths with those templates and removing their vertices, we end up with $2|W_{\mathbf{i}}| = (k-1)|X_{\mathbf{i}}|$ for each $\mathbf{i}$.

Let

$$m_i = \min\big\{|Z_i|,\, 2\big|W_{(i,1)}\big|/(k-1),\, 2\big|W_{(i,2)}\big|/(k-1)\big\}.$$

Impose that k is odd; we want to leave $L_i^Z := \lfloor m_i/2 \rfloor$ special pairs in each Z_i and $L_i^W := ((k-1)/2)\lfloor m_i/2 \rfloor$ vertices in each $W_{(i,h)}$, leaving a total of $L = \sum_i \big(2L_i^Z + 2L_i^W\big)$ vertices. (The reason we divide m_i by 2 will become clear in the proof of Claim 22, to follow). Recall from Section 2.2.3

that each $\rho^{-1} \le ((k-1)/2)\big(|Z_{[\mathbf{i}]}|/|W_{\mathbf{i}}|\big) \le \rho$, so $m_i \ge \rho^{-1}|Z_i|$ and

$$L_i^Z/|Z_i| \ge (1+o(1))\rho^{-1}/2, \quad L_{[\mathbf{i}]}^W/|W_{\mathbf{i}}| \ge (1+o(1))\rho^{-2}/2.$$

Provided we can find suitable templates, we can therefore apply Claim 19 with $\gamma = \rho^{-2}/3$. We can actually choose our templates almost arbitrarily; we wrap up the details in the following claim.

Claim 22. *For large k, we can choose good templates in such a way that if we remove the vertices of special paths corresponding to these templates, there will be L_i^Z special pairs left in each Z_i and $L_{[\mathbf{i}]}^W$ vertices left in each $W_{\mathbf{i}}$.*

Proof. It is convenient to describe our templates by a collection of $M := (n-L)/(k+1)$ disjoint good special sequences. The equivalence classes of those special sequences will give our templates.

Arbitrarily choose subsets $Z_i' \subseteq Z_i$ and $W_{\mathbf{i}}' \subseteq W_{\mathbf{i}}$ with $|Z_i'| = |Z_i| - L_i^Z$ and $|W_{\mathbf{i}}'| = |W_{\mathbf{i}}| - L_{[\mathbf{i}]}^W$. Let $X_{\mathbf{i}}' \subseteq X_{\mathbf{i}}$ be the subset of special vertices in $X_{\mathbf{i}}$ coming from special pairs in $Z_{[\mathbf{i}]}'$. Let $Z' = \bigcup_i Z_i'$ be the set of all special pairs in our subsets, let $X' = \bigcup_{\mathbf{i}} X_{\mathbf{i}}'$ be the set of all special vertices in those special pairs, and let $W' = \bigcup_{\mathbf{i}} W_{\mathbf{i}}'$ be the set of free vertices in our subsets. Note that $|Z'| = M = |X'|/2 = |W'|/(k-1)$.

Recall that $\rho^{-1} \le ((k-1)/2)(|X_{\mathbf{i}}|/|W_{\mathbf{i}}|) \le \rho$ and note that $|W_{\mathbf{i}}'| \ge |W_{\mathbf{i}}| - m_i/2 \ge |W_{\mathbf{i}}|/2$. So, if k is large then for each $\mathbf{i}$ we have

$$\frac{|W_{\mathbf{i}}'|}{|X_{\mathbf{i}}'|} \ge \frac{|W_{\mathbf{i}}|/2}{|X_{\mathbf{i}}|} \ge \frac{k-1}{4\rho} \ge 1.$$

That is, for each $\mathbf{i}$ there are at least as many vertices in $W_{\mathbf{i}}'$ as in $X_{\mathbf{i}}'$. (The only reason we divided m_i by 2 in the definitions of L_i^Z and L_i^W was to guarantee this). We now start filling our special sequences with the vertices in $X' \cup W'$ (note $|X' \cup W'| = (k+1)M$, as required). For each special sequence we choose its first and last vertex to be the vertices of a special pair in Z' (say we choose $\{x, y\} \in Z_i'$, and we choose $x \in X_{(i,1)}$ and $y \in X_{(i,2)}$ to be the first and last vertices in the special sequence respectively). Then we choose arbitrary vertices from $W_{(i,1)}'$ and $W_{(i,2)}'$ to be the second and second-last vertices in the sequence (in our case, a vertex from $W_{(i,2)}'$ should be second and a vertex from $W_{(i,1)}'$ should be second-last). We have just confirmed that there are enough vertices in the $W_{\mathbf{i}}$ for this to be possible. After we have determined the first, second, second last and last vertices of every sequence, we can use the remaining $(k-3)M$ vertices in W' to complete our special sequences arbitrarily. $\square$

Remark 23. If we are careful, instead of choosing our templates arbitrarily we can strategically choose them in such a way that each cluster is only involved in a few different templates. In the proof of Claim 19, we would then not need to divide each cluster into nearly as many as Q^{k+1} subsets. This would mean that we can avoid the use of the powerful results of [8] and make do with a simpler and weaker variant of Lemma 16.

We now apply Claim 19 using the templates from Claim 22, and remove the vertices of the special paths that we find. Update the $W_{\mathbf{i}}$, $X_{\mathbf{i}}$ and Z_i; we finally have $2|W_{\mathbf{i}}| = (k-1)\lfloor m_{[\mathbf{i}]}/2 \rfloor = (k-1)|X_{\mathbf{i}}|$ for each $\mathbf{i}$, and each $\big(X_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ and $\big(W_{\mathbf{i}}, W_{\bar{\mathbf{i}}}\big)$ are $(\varepsilon_4, \delta_4)$-super-regular, where $\delta_4 = \delta_3/4$ and $\varepsilon_4 = 3\rho^2\varepsilon_3$.

2.2.5 Completing the embedding with the blow-up lemma

We have now finished with the random edges in 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. We first give a statement of the blow-up lemma, due to Komlós, Sárközy and Szemerédi.

Lemma 24 (Blow-up Lemma [12])**.** *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 $n_1, \ldots, n_r \in \mathbb{N}^+$. Let $V_1, \ldots, V_r$ be pairwise disjoint sets of sizes $n_1, \ldots, n_r$. We construct two graphs on the vertex set $V = \bigcup_{i=1}^r V_i$ as follows. The first graph $b_{n_1,\ldots,n_r}(C)$ (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 $B \subseteq b_{n_1,\ldots,n_r}(C)$ (a "super-regular blow-up") is obtained by putting edges between each such V_i and V_j so 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.

Now we describe the setup to apply the blow-up lemma. For each $i \in [q]$, define

$$S_i^0 = X_{(i,1)},\ S_i^1 = W_{(i,2)},\ S_i^2 = W_{(i,1)},\ S_i^3 = X_{(i,2)}.$$

We construct an auxiliary graph G_i as follows. Start with the subgraph of G induced by $S_i^0 \cup S_i^1 \cup S_i^2 \cup S_i^3$, and remove every edge not between some S_i^ℓ and $S_i^{\ell+1}$. Identify the two vertices in each special pair to get a super-regular "cluster-cycle" with 3 clusters S_i^Z, S_i^1 and S_i^2. To be precise, the elements of S_i^Z are the $|Z_i|$ (ordered!) special pairs $(x_1, x_2) \in X_{(i,1)} \times X_{(i,2)}$; such a special pair is adjacent to $v \in S_i^h$ if x_h is adjacent to v. The purpose of defining G_i is that a cycle containing exactly one vertex (x, y) from S_i^Z corresponds to a special path connecting x and y. See Figure 1.

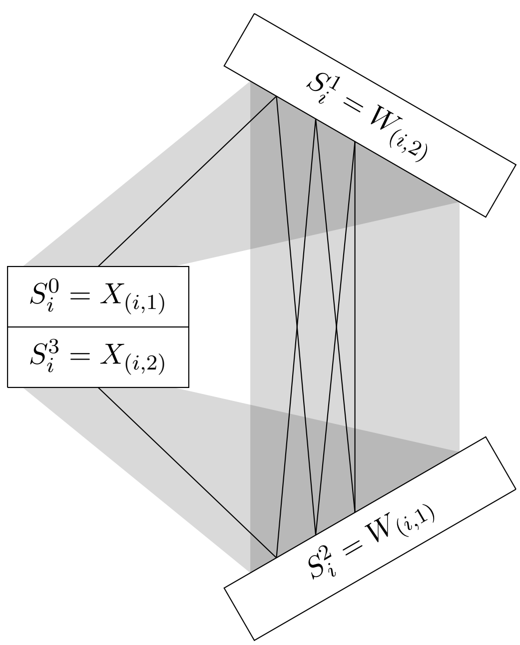

Figure 1. An example special path is shown, for $k = 7$.

Let $n_i^Z = \left|S_i^Z\right| = |Z_i|$. Note that G_i is a $(\varepsilon_4, \delta_4)$-super-regular blow-up of a 3-cycle C_3, with part sizes

$$n_i^Z, \quad n_i^1 := ((k-1)/2)n_i^Z, \quad n_i^2 := ((k-1)/2)n_i^Z.$$

The corresponding complete blow-up $b_{n_i^Z, n_i^1, n_i^2}(C_3)$ contains n_i^Z vertex-disjoint k-cycles (each has a vertex in S_i^Z and its other $k-1$ vertices alternate between S_i^1 and S_i^2). By the blow-up lemma, G_i also contains n_i^Z vertex-disjoint k-cycles. Since k is odd, each of these must use at least one vertex from S_i^Z, but since $\left|S_i^Z\right| = n_i^Z$, each cycle must then use exactly one vertex from S_i^Z. These cycles correspond to special paths which complete our embedding of T.

3 Concluding Remarks

We have proved that any given bounded-degree spanning tree typically appears when a linear number of random edges are added to an arbitrary dense graph. There are a few interesting questions that remain open. Most prominent is the question of embedding more general kinds of spanning subgraphs into randomly perturbed graphs. It would be particularly interesting if the general result of [5] (concerning arbitrary spanning graphs with bounded degree and low bandwidth) could be adapted to this setting. There is also the question of universality: whether it is true that a randomly perturbed dense graph typically contains every bounded-degree spanning tree at once. Finally, it is possible that our use of Szemerédi's regularity lemma could be avoided, thus drastically improving the constants $c(\alpha)$ and perhaps allowing us to say something about random perturbations of graphs which have slightly sublinear minimum degree.

Acknowledgement. Parts of this work were carried out when the first author visited the Institute for Mathematical Research (FIM) of ETH Zurich, and also when the third author visited the School of Mathematical Sciences of Tel Aviv University, Israel. We would like to thank both institutions for their hospitality and for creating a stimulating research environment.

References

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

[2] T. Bohman, A. Frieze, M. Krivelevich, and R. Martin, *Adding random edges to dense graphs*, Random Structures & Algorithms **24** (2004), no. 2, 105–117.

[3] T. Bohman, A. Frieze, and R. Martin, *How many random edges make a dense graph Hamiltonian?*, Random Structures & Algorithms **22** (2003), no. 1, 33–42.

[4] J. Böttcher, M. Schacht, and A. Taraz, *Spanning 3-colourable subgraphs of small bandwidth in dense graphs*, Journal of Combinatorial Theory, Series B **98** (2008), no. 4, 752–777.

[5] ———, *Proof of the bandwidth conjecture of Bollobás and Komlós*, Mathematische Annalen **343** (2009), no. 1, 175–205.

[6] B. Csaba, *On embedding well-separable graphs*, Discrete Mathematics **308** (2008), no. 19, 4322–4331.

[7] G. A. Dirac, *Some theorems on abstract graphs*, Proceedings of the London Mathematical Society **3** (1952), no. 1, 69–81.

[8] S. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger, *Small subsets inherit sparse ε-regularity*, Journal of Combinatorial Theory, Series B **97** (2007), no. 1, 34–56.

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

[10] 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.

[11] J. Komlós, G. N. Sárközy, and E. Szemerédi, *Proof of a packing conjecture of Bollobás*, Combinatorics, Probability and Computing **4** (1995), no. 03, 241–255.

[12] ———, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.

[13] ———, *Spanning trees in dense graphs*, Combinatorics, Probability and Computing **10** (2001), 397–416.

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

[15] M. Krivelevich, M. Kwan, and B. Sudakov, *Cycles and matchings in randomly perturbed digraphs and hypergraphs*, Combinatorics, Probability and Computing, to appear. (2015).

[16] M. Krivelevich, B. Sudakov, and P. Tetali, *On smoothed analysis in dense graphs and formulas*, Random Structures & Algorithms **29** (2006), no. 2, 180–193.

[17] D. Kühn, D. Osthus, and A. Taraz, *Large planar subgraphs in dense graphs*, Journal of Combinatorial Theory, Series B **95** (2005), no. 2, 263–282.

[18] R. Montgomery, *Embedding bounded degree spanning trees in random graphs*, arXiv preprint arXiv:1405.6559 (2014).

[19] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics **14** (1976), no. 4, 359–364.

[20] D. A. Spielman and S.-H. Teng, *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*, Journal of the ACM **51** (2004), no. 3, 385–463.