Tree tilings in random regular graphs

Sahar Diskin * Ilay Hoshen * Maksim Zhukovskii †

Abstract

We show that for every $\epsilon > 0$ there exists a sufficiently large $d_0 \in \mathbb{N}$ such that for every $d \geq d_0$, **whp** the random d-regular graph $G(n,d)$ contains a T-factor for every tree T on at most $(1-\epsilon)d/\ln d$ vertices. This is best possible since, for large enough integer d, **whp** $G(n,d)$ does not contain a $\frac{(1+\epsilon)d}{\ln d}$-star-factor. Our method gives a randomised algorithm which **whp** finds said T-factor and whose expected running time is $O(n^{1+o(1)})$, as well as an efficient deterministic counterpart.

1. Introduction

Let G be an n-vertex graph and H be an s-vertex graph. An H-*factor* in G is a union of $\lfloor \frac{n}{s} \rfloor$ vertex-disjoint isomorphic copies of H in G.

There has been an extensive study into the threshold of appearance of H-factors in the binomial random graph $G(n,p)$. The case where $H = K_2$ corresponds to finding a perfect matching in $G(n,p)$. The sharp threshold for appearence of a perfect matching was established by Erdős and Rényi [11]. Early results for general H were obtained by Alon and Yuster [3] and Ruciński [28]; they determined the threshold up to a constant factor for a specific family of graphs and gave bounds for the general case. For the case where H is a tree, Łuczak and Ruciński [21] characterised 'pendant' structures, and proved that in the random graph process (that is, when edges are added one after the other uniformly at random), the hitting time of the appearance of an H-factor is the same as the hitting time of the disappearance of these forbidden 'pendant' structures. In particular, one is able to infer the precise threshold for the appearance of an H-factor in this case. In 2008, Johansson, Kahn and Vu [17] determined the threshold (up to a multiplicative constant) for the existence of an H-factor for every strictly-1-balanced[1] graph H and determined the threshold up to a logarithmic factor for an arbitrary graph H. In the case of cliques K_s, Heckel (for $s = 3$) [13] and Riordan (for $s \geq 4$) [26] determined the *sharp* threshold for the appearance of an H-factor. Recently, a hitting time result for the appearance of a K_s-factor was proved [14], and the sharp threshold for the appearance of an H-factor for every strictly-1-balanced graph H was determined [8].

Much less is known in the case of *random d-regular graphs*. The random d-regular graph $G(n,d)$ is a graph chosen uniformly at random among all simple d-regular graphs on the vertex set $[\![n]\!] := \{1,\ldots,n\}$ (throughout the paper, we treat d as fixed and consider the asymptotics in n). Since, for every pair of integers $d \geq 2$ and $k \geq 3$, the number of cycles of length k in $G(n,d)$ is asymptotically distributed as a Poisson random variable with mean $(d-1)^k/(2k)$ (see [29]), **whp**[2] there are $o(n)$ cycles of length k in $G(n,d)$. Thus, we may (and will) restrict our attention to tree factors.

For the case of $H = K_2$, Bollobás [5] proved that whp there exists an H-factor (that is, a perfect matching) in $G(n,d)$ for every $d \geq 3$. There has been some research on the more general case of stars. Naturally, one cannot hope for a $K_{1,t}$-factor for $t > d$, since the graph is d-regular.

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel.

†School of Computer Science, The University of Sheffield, Sheffield S1 4DP, United Kingdom.

[1] A graph H is strictly-1-balanced if $\frac{|E(H)|}{|V(H)|-1} > \frac{|E(J)|}{|V(J)|-1}$ for every proper subgraph $J \subsetneq H$ with $|V(J)| \geq 2$.

[2] With high probability, that is, with probability tending to one as n tends to infinity.

For $d \geq 3$, using a first moment argument, one can show that **whp** $G(n,d)$ does not contain a $K_{1,d}$-factor (see [4, Corollary 2]). Robinson and Wormald [27] showed that for $d \geq 3$, **whp** $G(n,d)$ contains a Hamilton cycle, and thus a $K_{1,2}$-factor. In a subsequent work, Assiyatun and Wormald [4] showed that for $d \geq 4$, **whp** $G(n,d)$ contains a $K_{1,3}$-factor. One may then suspect that for any $d \geq 3$, typically $G(n,d)$ contains a $K_{1,d-1}$-factor. However, using first moment calculations, one can show that this is not the case for $d \geq 5$ (see Appendix A).

There are then two natural avenues to venture into: first, for sufficiently large d, to determine all k such that **whp** $G(n,d)$ contains a factor of stars on k vertices; second, more ambitiously, one could try to find all trees T for which **whp** $G(n,d)$ contains a T-factor.

Considering a related but slightly different problem, Alon and Wormald [2] showed that for any d-regular graph G, there exists an absolute constant c', such that G contains a star-factor, in which every star has at least $c'd/\log d$ vertices (not necessarily all stars are of the same size). We stress that here, and throughout the paper, all logarithms are with respect to the natural basis. They further noted that this is optimal up the choice of the constant $c' > 0$. Indeed, the existence of a factor of stars on at least k vertices implies the existence of a dominating set of size at most $\frac{n}{k-1}$, and for any $\epsilon > 0$ and sufficiently large d, **whp** the smallest dominating set in $G(n,d)$ is of size at least $\frac{(1-\epsilon)n\log d}{d}$ (see [2, page 3]). Let us note here that if one only assumes that G is d-regular, then one cannot hope to obtain a factor of stars of size exactly k for any $3 \leq k = O(d/\log d)$. Indeed, consider for example a d-regular graph G formed by a collection of vertex disjoint copies of K_{d+1} and vertex disjoint copies of complete bipartite graphs $K_{d,d}$. Then, since $\gcd(d+1, 2d) \in \{1,2\}$, for any choice of $k > 2$, one cannot find a factor of stars of size k.

Our first main result shows that typically a random d-regular graph G contains a star-factor with the asymptotically *optimal* possible size. In fact, we extend this result to factors of *any* tree (not necessarily a star).

Theorem 1. *For every constant $0 < \epsilon < 1$, there exists a sufficiently large integer d_0 such that the following holds for any $d \geq d_0$.* ***Whp****, for every tree T on at most $\frac{(1-\epsilon)d}{\log d}$ vertices, there exists a T-factor in $G(n,d)$.*

We note that throughout the paper, we will assume that $|V(G)|$ is divisible by $|V(T)|$, to avoid unnecessary technical details, however all proofs can be directly extended to the general case. Further, we note that we may fix the tree T and show that **whp** there is a T-factor in $G(n,d)$; since there are at most $d^2 \cdot 4^d$ such trees (see [24]) and d is fixed, by the union bound the statement then holds for every tree T.

A detailed sketch of the proof of Theorem 1 is presented in Section 2. Let us briefly recap the main strategy here. We show that **whp**, there exists a balanced partition of $|V(G)|$ into $|V(T)|$ parts so that, for every pair of parts V_i, V_j where $\{i,j\} \in E(T)$, every vertex in $V_i \cup V_j$ has the number of neighbours in the other part concentrated around the mean $d/|V(T)|$. We say that such a partition is *nice*. We obtain this nice partition through four different applications of the algorithmic version of the Lovász Local Lemma, due to Moser and Tardos [23]. In particular, this allows us to find such a nice partition which is *close* to a random partition; further, this gives us a randomised algorithm to find this partition whose average running time is $\tilde{O}(n)$ **whp** (see Theorem 3 and Corollary 3.3). In fact, we show that, in *any* d-regular graph, which does not have short cycles close to each other, a fraction of the possible partitions are close to nice partitions. Utilising a description of the distribution of edges in random graphs with specified degree sequences ([12], see also [22, Theorem 2.2]), we conclude that **whp** almost all partitions of a random regular graph induce multipartite graphs with good expansion properties. This allows us to find a partition with such expansion properties which is also close to a nice partition, ensuring the existence of a perfect matching between pairs of sets. Using an algorithm as in [9], we can find these perfect matchings in time $n^{1+o(1)}$. This gives us our second main result.

Theorem 2. *For every constant $0 < \epsilon < 1$, there exists a sufficiently large integer d_0 such that the following holds for any $d \geq d_0$. There is a randomised algorithm that* ***whp*** *finds a T-factor in $G(n, d)$ in expected time $n^{1+o(1)}$, for every tree T on at most $\frac{(1-\epsilon)d}{\log d}$ vertices.*

Since the events that we consider in our applications of the algorithmic version of the Lovász Local Lemma are determined by $\mathrm{poly}(d)$ random variables over domains of size at most d, [23, Theorem 1.4] shows that there exists a *deterministic* algorithm that **whp** finds these T-factors in polynomial in n time in $G(n, d)$.

Let us make some additional remarks.

- It is not hard to verify that our proof follows through for a uniformly chosen graph on n vertices with a given degree sequence, whose degrees lie in the interval $[d, (1+\delta)d]$ for some small $\delta > 0$. We believe slight modifications of our technique, specifically in Section 5, should allow us to obtain the same result for such graphs whose degrees are between d and $O(d)$.

- We stress that in order to show the existence of a perfect matching between relevant sets in the partition, we need our partition to be close to a random partition, and thus the application of the algorithmic version of the Lovász Local Lemma is crucial, even if we do not aim to get Theorem 2.

A possible simplification, which allows using a non-constructive version of the Lovász Local Lemma, is applying 'sprinkling' due to the contiguity result from [16] instead of applying the direct estimation of probabilities in $G(n, d)$. As soon as a nice partition is obtained, we add independently edges of $G(n, \epsilon' d)$, where $\epsilon' \ll \epsilon$. Although it simplifies the proof, it does not allow deriving Theorem 2 and the generalisation to non-regular random graphs with specified degree sequences. Moreover, this does not allow obtaining any probability bounds, in contrast to our approach. Indeed, our proof gives that the probability a random d-regular graph has a T-factor (for any tree T with $|V(T)| \leq (1-\epsilon)d/\log d$) is at least $1 - n^{-\Theta_d(1)}$. In fact, the latter probability bound is tight. Indeed, consider the vertices $\{1, ..., 10d\} \in [\![n]\!]$, say. The probability they form a connected component without a T-factor in $G(n, d)$ is at least n^{-100d^2}. We thus obtain the following corollary.

Corollary 1.1. *For every constant $0 < \epsilon < 1$, there exists a sufficiently large integer d_0 such that the following holds for any $d \geq d_0$. For any tree T on at most $\frac{(1-\epsilon)d}{\log d}$, the probability that $G(n, d)$ contains a T-factor is $1 - n^{-\Theta_d(1)}$.*

One key complication that arises when using any variant of the Lovász Local Lemma to prove Theorem 1 is that it is impossible to directly apply it, as every 'bad' event has too many dependencies. A similar issue was addressed independently in the paper by Draganić and Krivelevich [10] on connected dominating sets, where they proposed a (significantly different and shorter) proof strategy to show that a d-regular graph without short close cycles has a nice partition. Notably, their method requires $\Theta(n)$ applications of the Lovász Local Lemma (and consequently $O(n^2)$ resamples in the algorithmic version), which precludes a linear time reduction to finding perfect matchings. In contrast, our approach applies the Lovász Local Lemma only a constant number of times, enabling such a reduction.

Let us finish this section with several avenues for future research. While in this general setting Theorem 1 is asymptotically best possible, for the case where T is a path on k vertices one can achieve a better result. Indeed, since $G(n, d)$ is typically Hamiltonian [27], one can **whp** obtain a factor of paths of any size. In fact, this observation can be generalised to all trees of bounded degree (see below). It would be interesting to try characterising, for every value of $k = k(n)$, families of trees T on k vertices for which one can **whp** obtain a T-factor in $G(n, d)$.

As mentioned above, our proof uses results [22, 12] on the distribution of edges in graphs chosen uniformly at random given a degree sequence. It would be interesting to see whether this step can be amended to allow our result to hold for pseudo-random (n,d,λ)-graphs, with $\lambda \ll d$ (see [20] for background and many results on pseudo-random graphs). While a slight adjustment of our methods (with, in fact, a much simpler proof) yields that $\left(n,d,O(\sqrt{d})\right)$-graphs contain a star-factor for any star of size at most $\frac{d}{10\log d}$, this could be far from a complete answer. In fact, we are inclined to believe that Theorem 1 should hold for (n,d,λ)-graphs with $\lambda = o(d)$. Let us mention here that, answering a question of Krivelevich [19], Pavez-Signé [25] showed that for $\lambda = o(d)$, an (n,d,λ)-graph contains a copy of every n-vertex tree with bounded degree and $\Theta(n)$ leaves. Subsequent work by Hyde, Morrison, Müyesser, and Pavez-Signé [15] showed that for $\lambda = o(d/\log^3 n)$, an (n,d,λ)-graph contains a copy of every n-vertex tree with bounded degree — and thus, in particular, contains a T-factor for any tree T of bounded degree.

Another possible direction would be to consider the typical existence of any spanning forests in $G(n,d)$ whose degree is bounded by $(1-\epsilon)d/\log d$. Here, it might be interesting to attempt this first in the model of the binomial random graph, $G(n,p)$, for p above the connectivity threshold, that is, $p \geq \frac{(1+\epsilon)\log n}{n}$. Is it true that it contains any spanning forest F whose degree is bounded by $O(np/\log(np))$ **whp**? This is naturally tightly related to the universality question, with perhaps one key example being the result of Komlós, Sárközy, and Szemerédi [18], showing that that for every positive α, Δ and sufficiently large n, every graph with minimum degree at least $(1/2+\alpha)n$ contains every tree on n vertices with maximum degree at most Δ.

1.1. Organisation

In Section 1.2 we set out some notation which will be of use throughout the paper. We then discuss the proof's structure and strategy in Section 2. In Section 3 we collect some lemmas which we will utilise in subsequent sections. Section 4 is devoted to the proof of the key proposition (Proposition 4.1), and is perhaps the most involved and novel part of the paper. Finally, in Section 5 we prove two typical properties of $G(n,d)$ and show how to deduce Theorem 1 from these properties and Proposition 4.1.

1.2. Notation

Given a graph H, a vertex $v \in V(H)$, and a set $A \subseteq V(H)$, we denote by $d_H(v)$ the degree of v and by $d_H(v,A)$ the number of neighbours of v in A (in H). When the graph in question is clear we may omit the subscript. We write $d(A) = \sum_{v\in A} d(v)$. Given $A, B \subseteq V(H)$, we denote by $e(A,B)$ the number of edges with one endpoint in A and the other endpoint in B. When $A = B$, $e(A) = e(A,A)$ is the number of edges induced by A. We denote by $N(A,B)$ the neighbourhood of A in B, that is, the set of vertices in B which are adjacent to some vertex in A. All logarithms are with the natural base. Moreover, for every positive integer n, define $[\![n]\!] := \{1,2,\ldots,n\}$. We use the fairly standard notation that given sequences $a = (a_n)$ and $b = (b_n) \geq 0$, $a = o(b)$ if, for every $\varepsilon > 0$ there exists n_0 such that $|a_n| \leq \varepsilon b_n$ for all $n \geq n_0$. Given sequences $a' = (a'_d = a'_d(n))$ and $b' = (b'_d = b'_d(n)) \geq 0$, we sometimes also use $a = o_d(b)$ to say that, for every $\varepsilon > 0$ there exists d_0, n_0 such that $|a'_d| \leq \varepsilon b'_d$ for all $d \geq d_0$ and $n \geq n_0$. We systematically ignore rounding signs when it does not affect computations.

2. Proof outline

Unsurprisingly, finding a tree factor is much harder when the size of the tree is close to the optimal size (that is, $d/\log d$). In this section, we will present the proof outline for Theorem 1 in the case when $k \geq \frac{\log d}{10d}$. We will further point out the steps where the proof becomes simpler for trees of smaller size.

Let T be a tree on k vertices and let us label these vertices by $V(T) := [\![k]\!]$. The overall strategy for finding a T-factor in $G \sim G(n,d)$ is quite intuitive. We will find k disjoint sets

$V_1, \dots, V_k \subseteq V(G)$ of equal size and show that **whp** there exists a perfect matching between every V_i and V_j such that $\{i,j\} \in E(T)$. To do so, our proof proceeds in two main steps. In the first step, we find 'good' sets $V_1, \dots, V_k$ (in fact, we show such a partition typically exists in any d-regular graph G without two short cycles close to each other). In the second step, we show the typical existence of a perfect matching between every relevant pair of these sets. The properties achieved in the first step facilitate the execution of the second step.

The first step of the proof, presented in Section 4, is perhaps the most involved and novel part. In this step, we establish key properties of the sets $V_1, \dots, V_k$ which will be crucial in verifying the typical existence of perfect matchings in the second step. First, we show that for every $\{i,j\} \in E(T)$, the degree of every $v \in V_i$ into V_j is around d/k. Notice that, this property alone does not suffice to establish the existence of a perfect matching between V_i and V_j. To that end, we will also make sure that the sets $V_1, \dots, V_k$ are 'close' to uniformly chosen sets.

In the second step, presented in Section 5, we show that **whp**, for every $\{i,j\} \in E(T)$, there exists a perfect matching between V_i and V_j by showing that Hall's condition is satisfied. That is, we will show that **whp**, for every $W \subseteq V_i$, we have $|N(W, V_j)| \geq |W|$. To that end, we utilise a useful bound on the distribution of edges in graphs chosen uniformly at random given a degree sequence (see Theorem 4 in Section 3, see also [22]). First, we will show that **whp** for every two 'small' sets $U, W \subseteq V(G)$ with $|U| = |W|$, there are not too many edges going from U to W. Moreover, since the sets $V_1, \dots, V_k$ were constructed in the first step in a way such that the degree of every vertex $v \in V_i$ to appropriate V_j's is not too small, we obtain a lower bound on $e(W, V_j)$ for every $W \subseteq V_i$. In particular, if $|N(W, V_j)| < |W|$, we will get a contradiction for small sets $W \subseteq V_i$. In the same spirit, Theorem 4 together with the properties of the sets $V_1, \dots, V_k$ allows us to bound $|N(W, V_j)|$ for every 'large' $W \subseteq V_i$ *if* the sets $V_1, \dots, V_k$ were chosen uniformly at random. Indeed, given randomly chosen disjoint sets A and B (that is, sets formed without first exposing $G(n,d)$), the graph $G[A \cup B]$, given its degree sequence, has a uniform distribution. Luckily, the first step ensures that the sets $V_1 \dots, V_k$ behave similarly to uniformly chosen sets.

Let us return to the first step of the proof and describe the broad strategy of showing the typical existence of the 'good' sets $V_1, \dots, V_k$. We begin with a random partition of the vertices into k parts, $S_1, \dots, S_k$. A key tool in establishing the existence of such sets $V_1, \dots, V_k$ is the Lovász Local Lemma. Since we want our sets to be close to the initial random sets $S_1, \dots, S_k$ (so that we may later be able to apply Theorem 4), we will in fact utilise the algorithmic version of the Lovász Local Lemma, due to Moser and Tardos (see Theorem 3 in Section 3). Utilising the algorithmic version of the Lovász Local Lemma, we can show that in each step of the algorithm we resample a small number of random variables assigned to vertices, and thus the initial random sets $S_1, \dots, S_k$ will not be 'far' from $V_1, \dots, V_k$. Let us note here that when applying the algorithmic version of the Lovász Local Lemma, in the initial step of the algorithm one evaluates all 'bad' events (requires $\tilde{O}(n)$ time), and then, at each 'resampling' step, one re-evaluates only those $O(1) = O_d(1)$ events that depend on the resampled random variables. Since the expected number of steps of the algorithm of Moser and Tardos is $O(n)$, this gives the overall expected running time $\tilde{O}(n)$.

Now, for every vertex $v \in V(G)$, sample $X_v \in [\![k]\!]$ uniformly at random, and independently for all vertices. For every $i \in [\![k]\!]$, set $S_i \coloneqq \{v \in V(G) \colon X_v = i\}$. Moreover, for every vertex $v \in V(G)$, denote by B_v the event that there exists $\{i,j\} \in E(T)$ such that $v \in S_i$ and $d(v, S_j) \notin [\delta d/k, Cd/k]$ where $\delta > 0$ is a sufficiently small constant and $C > 0$ is a sufficiently large constant. Notice that if $\neg B_v$ occurs for every $v \in V(G)$, then we get the desired bounds on the degrees which is the first key point in the first step.

Note that, for every vertex $v \in V(G)$ and $j \in [\![k]\!]$, we have $d(v, S_j) \sim \text{Bin}(d, 1/k)$. This distribution is the heart of the obstacle concerning 'large' trees. The reason for it is that whenever $k \leq \frac{d}{10 \log d}$, then the probability that the degree of v into S_j, for some $j \in [\![k]\!]$, is not in the interval $[\delta d/k, Cd/k]$ is at most d^{-8}. Furthermore, the event B_v is determined by $d+1$

random variables X_u, and B_v is independent of all but at most d^2 other events B_u. Therefore, we can apply the Lovász Local Lemma. It is worth noting here that if $k \leq \frac{d}{10 \log d}$, we may omit the requirement that $\{i, j\} \in E(T)$ in the definition of B_v (see Section 4.5). Then, if for every vertex $v \in V(G)$ we have that $\neg B_v$ holds, then $d(v, S_j) \in [\delta d/k, Cd/k]$ for every vertex $v \in V(G)$ and index $j \in [\![k]\!]$.

However, as k gets closer to $\frac{(1-\epsilon)d}{\log d}$, the probability that $\mathrm{Bin}\,(d, 1/k) < \delta d/k$ is not smaller than $d^{-1-\epsilon'}$, for some $\epsilon' > 0$ tending to zero as ϵ tends to zero. Thus, the treatment of this case is much more delicate and involves several rounds of applications of the algorithmic version of the Lovász Local Lemma, in order to refine the initial random partition. In the rest of this section, we describe these rounds.

Notice that, for every $i \in [\![k]\!]$, the event B_v conditioned on $v \in S_i$ is more likely to occur as the degree of the i-th vertex in T gets larger. For this reason, we will treat vertices of small degree and vertices of large degree in T differently (this treatment is in Section 4.1). Assume that $[\![h]\!]$ is the set of vertices of T with 'large' degrees. We slightly decrease the probability that $X_v = i$, for every $i \in [\![h]\!]$. Then, after one application of the Lovász Local Lemma we will be able to get rid of vertices which have more than Cd/k neighbours into S_j, for some $j \in [\![h]\!]$. Next, we consider the neighbourhood of the vertices which have degree less than $\delta d/k$ into S_j, for some $j \in [\![h]\!]$. We 'resample' the vertices in this neighbourhood outside of $S_1, \dots, S_h$ into $S_1, \dots, S_h$, that is, we move them into one of the sets $S_1, \dots, S_h$ uniformly at random. In this way, once again using the Lovász Local Lemma, we will have that $d(v, S_j) \in [\delta d/k, Cd/k]$ for every $v \in V(G)$ and $j \in [\![h]\!]$.

Next, in Section 4.2, we partition the remaining vertices among $S_{h+1}, \dots, S_k$. In this third application of the Lovász Local Lemma, we will ensure that after the resampling we have the property that $d(v, S_j) \in [\delta d/k, Cd/k]$ for every vertex $v \in V(G)$ and for all but at most ϵ^{-2} indices $j \in [\![k]\!]$. At this point, there will still be vertices $v \in S_i$ which have less than $\delta d/k$ neighbours into some S_j where $\{i, j\} \in E(T)$. After the fourth application of the Lovász Local Lemma, we will be able to obtain a partition with no such vertices (that is, $d(v, S_j) \in [\delta d/k, Cd/k]$ for every $v \in S_i$ and $\{i, j\} \in E(T)$).

Finally, in Section 4.4, we adjust the sets $S_1, \dots, S_k$ to be of size n/k each. This is the purpose of the fifth and final round of the Lovász Local Lemma. In this round, we will move vertices from sets of size bigger than $\frac{n}{k}$ to sets of size smaller than $\frac{n}{k}$ in a random manner. We will do so while ensuring the vertices v we move satisfy that $d(v, S_j) \in [\delta d/k, Cd/k]$ for every $j \in [\![k]\!]$. After this round, while keeping the bounds over the degrees of the vertices, we will be able to make each set S_i to be close to n/k up to and additive n/d^{100} error term. Finally, to make the sets exactly of size $\frac{n}{k}$, we introduce a deterministic argument adjusting the sets $S_1, \dots, S_k$ while changing the degree of every vertex to every set S_i by at most one. We thus obtain the required sets $V_1, \dots, V_k$.

3. Preliminaries

We will make use of the following fairly standard Chernoff-type probabilistic bounds (see, for example, Appendix A in [1]).

Lemma 3.1. *Let $p_1, \dots, p_n \in [0, 1]$. For every $i \in [\![n]\!]$, let $X_i \sim$ Bernouli(p_i), and set $X = \sum_{i=1}^n X_i$. Then,*

1. *For every $b > 0$,*
$$\mathbb{P}(X > b\mathbb{E}[X]) \leq \left(\frac{e}{b}\right)^{b\mathbb{E}[X]}.$$

2. *For any $\delta \geq 0$,*
$$\mathbb{P}\left(X \geq (1+\delta)\mathbb{E}[X]\right) \leq e^{-\frac{\delta^2 \mathbb{E}[X]}{2+\delta}}.$$

3. For any $0 \leq t \leq \mathbb{E}[X]$,

$$\mathbb{P}\left(X \leq \mathbb{E}[X] - t\right) \leq e^{-\frac{t^2}{3\mathbb{E}[X]}}.$$

We also require the following bound on the lower tail of the Binomial distribution.

Lemma 3.2. *For every $\xi > 0$ there exists $\delta_0 > 0$ such that for every $\delta \leq \delta_0$ the following holds. Let $t \coloneqq t(\xi, \delta) > 0$ be sufficiently large. Suppose that $np \geq (1+\xi)t$ and $p \leq \frac{1}{2}$. Then,*

$$\mathbb{P}\left(Bin\left(n,p\right) \leq \delta t\right) \leq e^{-(1+2\xi/3)t}.$$

Proof. Note that we may assume that $np = (1+\xi)t$. We have that

$$\mathbb{P}\left(\mathrm{Bin}\left(n,p\right) \leq \delta t\right) = \sum_{i=0}^{\delta t} \binom{n}{i} p^i \left(1-p\right)^{n-i} = \left(1-p\right)^n \cdot \sum_{i=0}^{\delta t} \binom{n}{i} \left(\frac{p}{1-p}\right)^i.$$

For the first term,

$$\left(1-p\right)^n \leq e^{-pn} = e^{-(1+\xi)t}.$$

For the second term,

$$\sum_{i=0}^{\delta t} \binom{n}{i} \left(\frac{p}{1-p}\right)^i \leq 1 + \sum_{i=1}^{\delta t} \left(\frac{enp}{i(1-p)}\right)^i \leq 1 + \sum_{i=1}^{\delta t} \left(\frac{2e(1+\xi)t}{i}\right)^i.$$

Since δ is sufficiently small, the function $g(x) = \left(\frac{2e(1+\xi)t}{x}\right)^x$ is increasing in $x \in (0, \delta t)$ and thus

$$1 + \sum_{i=1}^{\delta t} \left(\frac{2e(1+\xi)t}{i}\right)^i \leq \delta t \left(\frac{11(1+\xi)}{\delta}\right)^{\delta t} \leq e^{\log(12(1+\xi)/\delta) \cdot \delta t},$$

recalling that $t = t(\xi, \delta)$ is sufficiently large. Noting that $\delta \cdot \log\left(12(1+\xi)/\delta\right) \leq \xi/3$ for δ sufficiently small, we have that $\mathbb{P}\left(\mathrm{Bin}\left(n,p\right) \leq \delta t\right) \leq e^{-(1+2\xi/3)t}$, as required. □

We will also make extensive use of the algorithmic version of the Lovász Local Lemma, due to Moser and Tardos [23].

Theorem 3 (Theorem 1.2 of [23], rephrased)**.** *Let U be a finite set. Let $X = (\xi_u)_{u \in U}$ be a tuple of mutually independent random variables. Let $\mathcal{F}$ be a finite set of events determined by X. Suppose that there exists q such that for every event $F \in \mathcal{F}$, $\mathbb{P}_X(F) \leq q$. Moreover, suppose that every $F \in \mathcal{F}$ depends on at most Δ other events $F' \in \mathcal{F}$. Suppose that $\beta \in (0,1)$ satisfies $q \leq \beta(1-\beta)^{\Delta}$. Then, there exists an evaluation of X which does not satisfy any event in $\mathcal{F}$.*

Furthermore, let $(X_n)_{n \in \mathbb{N}}$ be a sequence of mutually independent copies of X, that is, for every $n \in \mathbb{N}$, $X_n \sim X$. Initially, we sample X_0 and let $Z_0 \coloneqq X_0$. At step $t \geq 1$, we pick one $F \in \mathcal{F}$ satisfied by Z_{t-1} (if one exists) in an arbitrary manner. Then, we consider all $u \in U$ such that this witness F depends on the u-th coordinate of Z_{t-1}, and set $(Z_t)_u = (X_t)_u$ for all such u and $(Z_t)_u = (Z_{t-1})_u$ for all other u. If no such $F \in \mathcal{F}$ exists, the process halts and we set $\tau \coloneqq t$. Then, $\mathbb{E}\left[\tau\right] \leq |\mathcal{F}| \frac{\beta}{1-\beta}$.

In fact, we will utilise the following corollary.

Corollary 3.3. *Let U be a finite set. Let $m \in \mathbb{N}$. Let $X = (\xi_u)_{u \in U}$ be a set of mutually independent random variables, supported on $[\![m]\!]$. Let $S_1, \ldots, S_m$ be a partition of U satisfying $S_i = \{u \in U \colon \xi_u = i\}$ for every $i \in [\![m]\!]$. Let $\mathcal{F}$ be a finite set of events determined by $S_1, \ldots, S_m$. Suppose that there exists q such that $\mathbb{P}_X(F) \leq q$ for every event $F \in \mathcal{F}$. Moreover, suppose that every $F \in \mathcal{F}$ is determined by at most Δ_1 random variables ξ_u, and depends on at most Δ_2 other events $F' \in \mathcal{F}$. Furthermore, suppose that $\beta \in (0,1)$ satisfies that $q \leq \beta(1-\beta)^{\Delta_2}$. Then, the probability (under the measure of X) that there exists a partition of U into $U_1, \ldots, U_m$ which does not satisfy any event in $\mathcal{F}$ and $\left|S_i \triangle U_i\right| \leq 2\Delta_1 |\mathcal{F}| \frac{\beta}{1-\beta}$ for every $i \in [\![m]\!]$, is at least $\frac{1}{2}$.*

Proof. Let $\Sigma := (X_n)_{n\in\mathbb{N}}$, let τ and let $Z = (Z_i)_{i=0}^{\tau}$ be the random variables as in the statement of Theorem 3. Let A be the event that $\tau \le 2|\mathcal{F}|\frac{\beta}{1-\beta}$. By Theorem 3, $\mathbb{E}_\Sigma[\tau] \le |\mathcal{F}|\frac{\beta}{1-\beta}$, and thus by Markov's inequality $\mathbb{P}_\Sigma(A) \ge \frac{1}{2}$.

Now, let B be the desired event that there exists a partition of U into $U_1, \dots, U_m$ which does not satisfy any event in $\mathcal{F}$, and $\left|S_i \triangle U_i\right| \le 2\Delta_1|\mathcal{F}|\frac{\beta}{1-\beta}$ for every $i \in [\![m]\!]$. Note that B is determined by X_0. Further, at every step in the algorithm, the number of $u \in U$ for which $(Z_t)_u \ne (Z_{t-1})_u$ is at most Δ_1. Thus, given that $\neg B$ occurs, the number of steps necessary to find an evaluation that does not satisfy any event in $\mathcal{F}$ is more than $\frac{2\Delta_1|\mathcal{F}|\frac{\beta}{1-\beta}}{\Delta_1} = 2|\mathcal{F}|\frac{\beta}{1-\beta}$. Hence, given that $\neg B$ occurs, the probability that the algorithm ran at most $2|\mathcal{F}|\frac{\beta}{1-\beta}$ steps is zero. Therefore,

$$\begin{aligned}\frac{1}{2} \le \mathbb{P}_\Sigma(A) &= \mathbb{P}_\Sigma(A\cap B) + \mathbb{P}_\Sigma(A \cap \neg B)\\ &= \mathbb{P}_\Sigma(A\cap B) \le \mathbb{P}_\Sigma(B) = \mathbb{P}_{X_0}(B),\end{aligned}$$

as claimed. □

Finally, let us conclude this section with a few statements on the distribution of edges in random graphs which are uniformly chosen among all graphs with a given degree sequence.

Theorem 4 (Corollary 8 in [12]). *Given a degree sequence* $\mathbf{d} = (d_1, \dots, d_n)$ *such that* $d_1 \ge d_2 \ge \cdots \ge d_n$, *let* G *be a uniform random graph on vertex set* $[\![n]\!]$ *where vertex* i *has degree* d_i. *Set* $M := \sum_{i=1}^n d_i$. *If an integer* $1 \le \ell \le M/2$ *satisfies* $\sum_{i=1}^{d_1} d_i = o(M - 2\ell)$, *then, for every* $S_1, S_2 \subseteq V(G)$,

$$\mathbb{P}\left(e(S_1,S_2) \ge \ell\right) \le \binom{d(S_1)}{\ell} \cdot \frac{(d(S_2))_\ell}{(M/2)_\ell (2+o(1))^\ell} = \binom{d(S_1)}{\ell} \cdot \frac{\binom{d(S_2)}{\ell}}{\binom{M/2}{\ell}(2+o(1))^\ell}.$$

We will also use the following corollary.

Corollary 3.4. *Under the same setting as in the statement of Theorem 4, if* $d(S_2) \le M/2$, *we further have*

$$\mathbb{P}\left(e(S_1,S_2) \ge \ell\right) \le \binom{d(S_1)}{\ell} \cdot \left(\frac{d(S_2)}{M(1+o(1))}\right)^\ell.$$

.

Proof. By Theorem 4, $\mathbb{P}\left(e(S_1,S_2) \ge \ell\right) \le \binom{d(S_1)}{\ell}\frac{(d(S_2))_\ell}{(M/2)_\ell(2+o(1))^\ell}$. Since $d(S_2) \le \frac{M}{2}$, we have that $\frac{d(S_2)-i}{M/2-i} \le \frac{d(S_2)}{M/2}$ for every $0 \le i \le d(S_2)$. Thus,

$$\begin{aligned}\binom{d(S_1)}{\ell}\frac{(d(S_2))_\ell}{(M/2)_\ell(2+o(1))^\ell} &\le \binom{d(S_1)}{\ell}\left(\frac{d(S_2)}{M/2}\right)^\ell \cdot \frac{1}{(2(1+o(1)))^\ell}\\ &= \binom{d(S_1)}{\ell} \cdot \left(\frac{d(S_2)}{M(1+o(1))}\right)^\ell,\end{aligned}$$

as required. □

We will further utilise the following lemma.

Lemma 3.5. *There exists a sufficiently small constant* $\xi > 0$ *such that the following holds. Given a degree sequence* $\mathbf{d} = (d_1, \dots, d_n)$ *such that* $d_1 \ge d_2 \ge \cdots \ge d_n$, *let* G *be a uniform* *random graph on vertex set* $[\![n]\!]$ *where vertex* i *has degree* d_i*. Set* $M := \sum_{i=1}^{n} d_i$*. For every* $t \in [\![n]\!]$ *and every two disjoint sets* A *and* B *satisfying*

$$M \geq 8(1-3\xi)t, \quad d(A), d(B) \leq 2(1+2\xi)t, \quad \textit{and} \quad d_1 = o\left(\sqrt{t}\right),$$

we have

$$\mathbb{P}\left(e(A,B) \geq (1-2\xi)t\right) \leq 0.95^t.$$

Proof. Set $\ell := (1-2\xi)t$. We have that $M - 2\ell = \Omega(t)$, $\sum_{i=1}^{d_1} d_i \leq d_1^2 = o(t)$, and $d(A), d(B) \leq M/2$. We may thus apply Theorem 4. We have

$$\mathbb{P}\left(e(A,B) \geq \ell\right) \leq \binom{d(A)}{\ell} \cdot \frac{\binom{d(B)}{\ell}}{\binom{M/2}{\ell} \cdot (2+o(1))^{\ell}}.$$

Further,

$$\binom{d(A)}{\ell} \leq \binom{2(1+2\xi)t}{\ell} \leq 4^{(1+2\xi)t} \leq 4^{(1+5\xi)\ell},$$

where the last inequality is true for sufficiently small ξ. We note that the same upper bound holds for $\binom{d(B)}{\ell}$. Moreover,

$$\binom{M/2}{\ell} \geq \binom{4(1-3\xi)\ell}{\ell} = \frac{((4-12\xi)\ell)!}{(\ell)!\,((3-12\xi)\ell)!}.$$

Thus, by Stirling's approximation, for sufficiently small ξ,

$$\begin{aligned}\binom{M/2}{\ell} &\geq \frac{(4-12\xi)^{(4-12\xi)\ell}}{(3-12\xi)^{(3-12\xi)\ell}} \cdot \frac{1}{\Theta\left(\sqrt{\ell}\right)} \\ &\geq \frac{4^{4\ell}}{3^{3\ell}} \cdot \frac{1}{\Theta\left(100^{\xi\ell}\sqrt{\ell}\right)} \geq 9.4^{\ell} \cdot \frac{1}{\Theta\left(100^{\xi\ell}\sqrt{\ell}\right)}.\end{aligned}$$

Hence, assuming ξ is sufficiently small,

$$\mathbb{P}\left(e(A,B) \geq \ell\right) \leq \frac{4^{2(1+5\xi)\ell}}{9.4^{\ell}(2+o(1))^{\ell}} \cdot \Theta\left(100^{\xi\ell}\sqrt{\ell}\right) \leq 0.9^{\ell} \leq 0.95^t.$$

□

4. Planting the seeds

As mentioned in Section 2, the proof of Theorem 1 consists of two main steps. In order to find a T-factor in $G(n,d)$, we will partition the vertices of $G(n,d)$ into $|V(T)|$ sets of the same size, each set represents a different vertex in the tree T, and find a perfect matching between the i-th set and the j-th set for every $\{i,j\} \in E(T)$. In this section, we find a partition of the vertices into $|V(T)|$ sets which satisfies two crucial properties, which in turn will allow us to find the desired perfect matchings in the second step in Section 5.

For every pair of integers d and n, let $\mathcal{G}_d$ be the family of all d-regular graphs on n vertices, such that there are no two cycles of length at most 10 at distance less than 10 from each other. Recall that we treat d as fixed, and consider the asymptotic as $n \to \infty$. Note that **whp** $G(n,d) \in \mathcal{G}_d$ (see, for example, [30]).

The main result of this section, which is the first step in the proof of Theorem 1, is the following.

Proposition 4.1. *For every $\epsilon > 0$, there exist a sufficiently small constant $\delta := \delta(\epsilon) > 0$, a sufficiently large constant $C := C(\epsilon) > 0$, and a sufficiently large integer d_0 such that the following holds for any $d \geq d_0$.*

Let T be a tree on $k \leq (1-\epsilon)\frac{d}{\log d}$ vertices. Let $G \in \mathcal{G}_d$ and suppose that n is divisible by k. Let $S_1, \ldots, S_k$ be a uniformly random partition of $V(G)$: $\mathbb{P}(v \in S_i) = \frac{1}{k}$ for every $i \in [\![k]\!]$, $v \in V(G)$ and the random choice of $i \in [\![k]\!]$ for $v \in V(G)$ is performed independently of all the other vertices.

Then, with probability bounded away from zero, there are disjoint sets $V_1, \ldots, V_k \subseteq V(G)$, each of size $\frac{n}{k}$, with the following properties.

(P1) $|S_i \triangle V_i| = o_d(n/k)$ *for every* $i \in [\![k]\!]$.

(P2) $d(v, V_j) \in \left[\frac{\delta d}{k}, \frac{Cd}{k}\right]$ *for every* $\{i,j\} \in E(T)$ *and* $v \in V_i$.

The proof of Proposition 4.1 is composed of five steps. Let us present here the overview of the proof and the organisation of this section.

In the first step of the proof of Proposition 4.1, appearing in Section 4.1, we take care of the sets V_i under construction which correspond to vertices of high degree in T.

In the second step of the proof of Proposition 4.1, appearing in Section 4.2, we construct the remaining sets in the partition of $V(G)$ (that is, the sets corresponding to vertices of low degree in T). After this step, for every $\{i,j\} \in E(T)$, we will no longer have vertices in the i-th set in the partition whose degree into the j-th set is greater than Cd/k. However, we may still have a small amount of vertices with degree less than $\delta d/k$. In the third step of the proof of Proposition 4.1, appearing in Section 4.3, we get rid of all such vertices. Lastly, in the final step of the proof of Proposition 4.1, appearing in Section 4.4, we will balance the sets of the partition to be of size exactly n/k while we ensure that the requested properties are kept.

As discussed in Section 2, the proof is much simpler whenever one assumes $k \leq \frac{d}{10\log d}$. Indeed, then some of the above steps may be skipped. In Sections 4.1 through 4.4, we focus on the case where $k \geq \frac{d}{10\log d}$. In Section 4.5, we provide a proof for the smaller values of k.

In Sections 4.1 through 4.4, we let T be a tree on $[\![k]\!]$, and (unless explicitly stated otherwise) we assume that $k \geq \frac{d}{10\log d}$.

4.1. Vertices of high degree in the tree

In this section, we build the sets in the partition of $V(G)$ that correspond to vertices of high degree in the tree. Let $\beta := \beta(\epsilon) > 0$ be a sufficiently small constant. Denote by $H_{deg}(T)$ the set of vertices of T whose degree is at least $d^{1-\beta}$, and let $h := |H_{deg}(T)|$, noting that $h < d^{\beta}$, since $k = |V(T)| < d$. Assume WLOG that $[\![h]\!] \subseteq V(T) = [\![k]\!]$ is exactly the set $H_{deg}(T)$. We construct the first h sets of the partition, ensuring that every $v \in V(G)$ will have degree between $\delta d/k$ and Cd/k into each one of these sets.

A key tool here is the algorithmic version of the Lovász Local Lemma. We build the first h sets in two rounds. First, we construct random h sets by assigning each vertex into each one of them with probability $(1-\alpha)/k$ for a suitable choice of α. After this sample, **whp** we will not have vertices with degree larger than Cd/k into any of the sets. However, we will have a small amount of vertices of degree smaller than $\delta d/k$ into some of the sets. We denote this set of vertices by B. In the next round, we will resample the neighbourhood of B outside of the first h sets in the partition, and put each vertex into each one of the first h sets with an appropriate probability, ensuring the expected size of the sets is n/k. As we will see, this probability will be at least $900/k$. This, in turn, will allow us to get rid of vertices with less than $\delta d/k$ neighbours into any of the first h sets in the partition.

The next lemma determines the value of α which should be considered.

Lemma 4.2. *There exists $c \in [\epsilon/4, 5]$ such that $\alpha = d^{-c}$ satisfies*

$$\mathbb{P}\left(Bin\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right) = \frac{\alpha^2}{d \cdot h}.$$

In relation to Theorem 2, we note that for the proof to follow, it is sufficient to n^{-3}-approximate α. As we know that $c \in [\epsilon/4, 5]$, the bisection method (see, for example, [7]) gives us an algorithm of finding an n^{-3}-approximation of α within $O_\epsilon(\log n)$ steps.

Proof. Consider the function $f(\alpha) = \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right) - \frac{\alpha^2}{d \cdot h}$. Let us show that there exists $\epsilon/4 \le c \le 5$ such that $f(d^{-c}) = 0$. Noting that $f(\alpha)$ is continuous, it suffices to show that $f\left(d^{-\epsilon/4}\right) < 0$, and that $f\left(d^{-5}\right) > 0$.

Let us first show that $f(d^{-\epsilon/4}) < 0$. Note that

$$d \cdot \frac{1-d^{-\epsilon/4}}{k} \ge \frac{(1-d^{-\epsilon/4})\log d}{1-\epsilon} \ge (1+\epsilon)\log d,$$

where the first inequality is true by the assumption that $k \le (1-\epsilon)\frac{d}{\log d}$ and the last inequality holds for large enough d. Thus, by Lemma 3.2, for δ sufficiently small with respect to ϵ,

$$\mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-d^{-\epsilon/4}}{k}\right) \le \delta \log d\right) \le e^{-(1+2\epsilon/3)\log d} < \left(d^{-\epsilon/4}\right)^2/(dh),$$

where the last inequality holds for $\beta > 0$ sufficiently small. Thus, we obtain $f\left(d^{-\epsilon/4}\right) < 0$.

Furthermore, for every $\alpha \in [0,1]$,

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right) > \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1}{k}\right) = 0\right) = \left(1-\frac{1}{k}\right)^d \\ \ge e^{-10.5\log d} = d^{-10.5},\end{aligned}$$

where the last inequality uses that $k \ge \frac{d}{10\log d}$. Thus, we have

$$\mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-d^{-5}}{k}\right) \le \delta \log d\right) > d^{-10.5} > \frac{d^{-2\cdot 5}}{d \cdot h},$$

and thus $f(d^{-5}) > 0$.

Recalling that $f(\alpha)$ is a continuous function in α, and $f(d^{-\frac{\epsilon}{4}}) < 0 < f(d^{-5})$, there exists $\epsilon/4 \le c \le 5$ such that $f(d^{-c}) = 0$. $\square$

Throughout the rest of the section, we let α be as in the statement of Lemma 4.2. For every vertex $v \in V(G)$, define the random variable X_v with the following distribution.

$$\mathbb{P}(X_v = i) = \begin{cases} \frac{1-\alpha}{k}, & i \in [\![h]\!] \\ 1 - \frac{(1-\alpha)h}{k}, & i = 0. \end{cases}$$

For every $i \in [\![h]\!]$, let $S_i \coloneqq \{v \in V(G) \colon X_v = i\}$. For every vertex disjoint sets $U_1, \ldots, U_h$, let

$$\begin{aligned} B(U_1, \ldots, U_h) &\coloneqq \{v \in V(G) \colon \exists i \in [\![h]\!], d(v, U_i) \le \delta \log d\}, \\ W(U_1, \ldots, U_h) &\coloneqq \{v \in V(G) \colon \exists i \in [\![h]\!], d(v, U_i) \ge C \log d\}. \end{aligned}$$

Set $B_1 \coloneqq B(S_1, \ldots, S_h)$ and $W_1 \coloneqq W(S_1, \ldots, S_h)$.

Let us first estimate several probabilities that will be useful for us in the proof.

Lemma 4.3. *For every $v \in V(G)$, we have the following.*

1. $\mathbb{P}(v \in W_1) \le d^{-100}$.

2. $\mathbb{P}(v \in B_1) \le \frac{\alpha^2}{d}$.

3. $\mathbb{P}\left(v \in N(B_1) \setminus \bigcup_{i\in[\![h]\!]} S_i\right) \le 2\left(1 - \frac{h(1-\alpha)}{k}\right)\alpha^2$.

Proof. For every $i \in [\![h]\!]$, we have $d(v, S_i) \sim \mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right)$. Using the assumption that $k \ge \frac{d}{10\log d}$, we have $\mathbb{E}[d(v, S_i)] \le 10 \log d$. For the first item of the lemma, by the union bound,

$$\mathbb{P}(v \in W_1) \le h \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \ge C \log d\right) \le d^{-100},$$

where the last inequality holds by Lemma 3.1 for C sufficiently large, recalling that $h \le d^{\beta}$.

We now turn to the second item. By the union bound and by Lemma 4.2,

$$\mathbb{P}(v \in B_1) \le h \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right) = \frac{\alpha^2}{d}.$$

We are thus left with the third item. Note that

$$\begin{aligned}\mathbb{P}\left(v \in N(B_1) \setminus \bigcup_{i\in[\![h]\!]} S_i\right) &= \mathbb{P}\left(v \notin \bigcup_{i\in[\![h]\!]} S_i\right)\mathbb{P}\left(v \in N(B_1) \mid v \notin \bigcup_{i\in[\![h]\!]} S_i\right)\\ &= \left(1 - \frac{h(1-\alpha)}{k}\right)\mathbb{P}\left(v \in N(B_1) \mid v \notin \bigcup_{i\in[\![h]\!]} S_i\right).\end{aligned}$$

By the union bound

$$\begin{aligned}\mathbb{P}\left(v \in N(B_1) \mid v \notin \bigcup_{i\in[\![h]\!]} S_i\right) &= \mathbb{P}\left(\exists u \in N(v), u \in B_1 \mid v \notin \bigcup_{i\in[\![h]\!]} S_i\right)\\ &\le d \cdot \mathbb{P}\left(u \in B_1 \mid u \in N(v), v \notin \bigcup_{i\in[\![h]\!]} S_i\right).\end{aligned}$$

Once again by the union bound and by Lemma 4.2,

$$\begin{aligned}\mathbb{P}\left(u \in B_1 \mid u \in N(v), v \notin \bigcup_{i\in[\![h]\!]} S_i\right) &\le h \cdot \mathbb{P}\left(\mathrm{Bin}\left(d-1, \frac{1-\alpha}{k}\right) \le \delta \log d\right)\\ &\le 2h \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right)\\ &= \frac{2\alpha^2}{d},\end{aligned}$$

where the second inequality is true by the following. For every integer $i \le \delta \log d$, set $x_i = \binom{d-1}{i}\left(\frac{1-\alpha}{k}\right)^i\left(1 - \frac{1-\alpha}{k}\right)^{d-1-i}$ and $y_i = \binom{d}{i}\left(\frac{1-\alpha}{k}\right)^i\left(1 - \frac{1-\alpha}{k}\right)^{d-i}$. Notice that

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}\left(d-1, \frac{1-\alpha}{k}\right) \le \delta \log d\right) &= \sum_{i=0}^{\delta \log d} x_i = \sum_{i=0}^{\delta \log d} y_i \cdot \frac{d-i}{d\left(1 - \frac{1-\alpha}{k}\right)}\\ &\le 2\sum_{i=0}^{\delta \log d} y_i = 2\mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1-\alpha}{k}\right) \le \delta \log d\right).\end{aligned}$$

Thus,

$$\mathbb{P}\left(v \in N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i\right) \leq \left(1 - \frac{h(1-\alpha)}{k}\right) \cdot 2\alpha^2,$$

completing the proof. $\square$

We are now ready for the first (out of several) key step in the proof.

Lemma 4.4. *With probability at least $\frac{1}{2} - o(1)$, there exist disjoint sets $A_1^{(1)}, \dots, A_h^{(1)} \subseteq V(G)$ which satisfy the following. Let $B_2 = B\left(A_1^{(1)}, \dots, A_h^{(1)}\right)$ and let $W_2 = W\left(A_1^{(1)}, \dots, A_h^{(1)}\right)$. Then,*

1. $\left|S_i \triangle A_i^{(1)}\right| \leq \frac{n}{d^{50}}$ *and* $\left||A_i^{(1)}| - \frac{(1-\alpha)n}{k}\right| \leq \frac{n}{d^{50}}$ *for every* $i \in [\![h]\!]$.
2. $W_2 = \varnothing$.
3. $\left|N(B_2) \setminus \bigcup_{i \in [\![h]\!]} A_i^{(1)}\right| \leq 3\alpha^2 n$.

Proof. For every $v \in V(G)$, let F_v be the event that $v \in W_1$. Let $\mathcal{F} := \{F_v\}_{v \in V(G)}$. By Lemma 4.3, for every $v \in V(G)$, we have $\mathbb{P}(F_v) \leq d^{-100} =: q$. Observe that every F_v is determined by $\Delta_1 := d$ random variables (its neighbours). Furthermore, every F_v depends on at most $\Delta_2 := d^2$ other events (revealing whether a vertex v satisfies F_v may only affect the probability that u satisfies F_u for u which is in the second neighbourhood of v). Furthermore, note that $\beta := 4d^{-100}$ satisfies

$$\beta(1-\beta)^{\Delta_2} = 4d^{-100}(1-4d^{-100})^{d^2} \geq 4d^{-100}e^{-d^{-90}} \geq d^{-100} = q. \tag{1}$$

By Corollary 3.3, we obtain that with probability at least $\frac{1}{2}$, there exist sets $A_1^{(1)}, \dots, A_h^{(1)}$ such that $W_2 = \varnothing$ and $\left|S_i \triangle A_i^{(1)}\right| \leq 2 \cdot d \cdot n \cdot \frac{4d^{-100}}{1-4d^{-100}} \leq \frac{n}{d^{51}}$ for every $i \in [\![h]\!]$. Let $A_1^{(1)}, \dots, A_h^{(1)}$ be these sets (if they do not exist, set $A_i^{(1)} = S_i$ for every $i \in [\![h]\!]$). By Lemma 3.1, **whp** $\left||S_i| - \frac{(1-\alpha)n}{k}\right| \leq n^{2/3}$ for every $i \in [\![h]\!]$, and thus we obtain the first and second items of the lemma.

As for the third item, by Lemma 4.3,

$$\mathbb{E}\left[\left|N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i\right|\right] \leq 2\left(1 - \frac{h(1-\alpha)}{k}\right)\alpha^2 n.$$

Moreover, the event that $v \in N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i$ depends on at most d^4 other events $u \in N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i$. Thus,

$$\mathrm{Var}\left(\left|N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i\right|\right) = \sum_{u,v} \mathrm{Cov}\left(v \in N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i, u \in N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i\right) \leq nd^4.$$

Thus, by Chebyshev's inequality, **whp**

$$\left|N(B_1) \setminus \bigcup_{i \in [\![h]\!]} S_i\right| \leq 2\left(1 - \frac{h(1-\alpha)}{k}\right)\alpha^2 n + n^{2/3}. \tag{2}$$

Furthermore, whenever we change the location of a vertex (that is, move it from S_i to S_j), we change the size of B_1 by at most d. Hence,

$$\big||B_2|-|B_1|\big| \le d\cdot \sum_{i\in[\![h]\!]} \big|S_i \triangle A_i^{(1)}\big| \le d\cdot h\cdot \frac{n}{d^{51}},$$

and thus $||N(B_2)|-|N(B_1)|| < \frac{n}{d^{40}}$. Together with (2), we obtain that

$$\left|N(B_2)\setminus \bigcup_{i\in[\![h]\!]} A_i^{(1)}\right| \le 3\alpha^2 n,$$

where we used that $\alpha \ge d^{-5}$. □

Recall that, in order to prove Proposition 4.1, we need to show that the sets $V_1,\dots,V_k$ exist with positive probability (bounded away from zero). We will actually prove that such sets exist with probability $1/4-o(1)$. In particular, by Lemma 4.4, the sets $A_1^{(1)},\dots,A_h^{(1)}$ (satisfying the statement of the lemma) exist with probability $1/2-o(1)$. For every possible tuple of disjoint sets $(S_1,\dots,S_h)$, if there exists a tuple of sets $(A_1^{(1)},\dots,A_h^{(1)})$, satisfying the conclusion of Lemma 4.4, we fix such a tuple. Otherwise, we let $A_i^{(1)} = S_i$ for all $i\in[\![h]\!]$. Further in this section, we assume that the event from Lemma 4.4, that has probability at least $1/2-o(1)$, actually occurs; we call the tuple $(S_1,\dots,S_h)$ *nice* in this case. Note that, under this assumption, the sets $A_i^{(1)}$ satisfy that no vertex has more than $C\log d$ neighbours in each one of them.

We turn to show that with probability bounded away from zero, there exist sets $A_1^{(2)},\dots,A_h^{(2)}$ that are 'not far' from $S_1,\dots,S_h$, and every vertex has between $\delta\log d$ and $2C\log d$ neighbours in each one of $A_1^{(2)},\dots,A_h^{(2)}$. Let U be a set of size $\frac{\alpha n}{1000}$ which contains $N(B_2)\setminus\bigcup_{i\in[\![h]\!]}A_i^{(1)}$ (note that by Lemma 4.4, $\left|N(B_2)\setminus\bigcup_{i\in[\![h]\!]}A_i^{(1)}\right| \le 3\alpha^2 n < \frac{\alpha n}{1000}$). We will make use of the following lemma.

Lemma 4.5. *There exist $\frac{900}{k}\le p_1,\dots,p_h\le\frac{1100}{k}$ such that, for every $i\in[\![h]\!]$,*

$$|A_i^{(1)}| + p_i|U| = \frac{n}{k}.$$

Proof. Recall that for every $i\in[\![h]\!]$, $|A_i^{(1)}| \in \left[\frac{(1-\alpha)n}{k}-\frac{n}{d^{50}}, \frac{(1-\alpha)n}{k}+\frac{n}{d^{50}}\right]$. Thus, p_i should satisfy

$$p_i|U| = \frac{n}{k} - |A_i^{(1)}| \in \left[\frac{\alpha n}{k}-\frac{n}{d^{50}}, \frac{\alpha n}{k}+\frac{n}{d^{50}}\right].$$

Plugging $|U|=\frac{\alpha n}{1000}$ yields

$$p_i \in \left[\frac{1000}{k}-\frac{1000}{\alpha d^{50}}, \frac{1000}{k}+\frac{1000}{\alpha d^{50}}\right] \subseteq \left[\frac{900}{k},\frac{1100}{k}\right].$$

□

Let $p_1,\dots,p_h$ be the probabilities from the statement of Lemma 4.5. Let us note that $\sum_{i\in[\![h]\!]} p_i \le \frac{1100h}{k} < 1$, since $h<d^\beta$ and $k\ge\frac{d}{10\log d}$. Further, for every $v\in U$ we define a random variable X_v such that $\mathbb{P}(X_v=i)=p_i$ for every $i\in[\![h]\!]$. Moreover, for every $i\in[\![h]\!]$, we set $U_i := \{v\in U\colon X_v=i\}$, and let $\tilde{A}_i^{(1)} := A_i^{(1)}\cup U_i$. Let W_3 be the set of vertices $v\in V(G)$ such that there exists $i\in[\![h]\!]$ for which $d(v,\tilde{A}_i^{(1)})\notin(\delta\log d, 2C\log d)$.

Let us first bound the probability that $v\in W_3$. Recall that we have already conditioned on the existence of the sets $A_1^{(1)},\dots,A_h^{(1)}$ satisfying the properties as in the statement of Lemma 4.4. Further, note that the probability measure in the following lemma is induced by the random variables $\{X_v\}_{v\in U}$ given in the previous paragraph.

Lemma 4.6. $\mathbb{P}(v \in W_3) \leq \frac{1}{d^{100}}$ *for every* $v \in V(G)$.

Proof. We begin with the probability that there exists $i \in [\![h]\!]$ for which $d(v, \tilde{A}_i^{(1)}) \geq 2C \log d$. Recall that $d(v, A_i^{(1)}) \leq C \log d$ for every $i \in [\![h]\!]$. Thus, if there is $i \in [\![h]\!]$ for which $d(v, \tilde{A}_i^{(1)}) \geq 2C \log d$, we must have $d(v, U_i) \geq C \log d$. For every $i \in [\![h]\!]$, we have $d(v, U_i) \sim \mathrm{Bin}\,(d(v, U), p_i)$. By Lemma 4.5 and by the bound $d(v, U) \leq d$,

$$\mathbb{P}\left(\exists i \in [\![h]\!], d(v, U_i) \geq C \log d\right) \leq h \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1100}{k}\right) \geq C \log d\right) < d^{-101},$$

where the last inequality is true by Lemma 3.1 for $C > 0$ large enough.

As for the probability that there exists $i \in [\![h]\!]$ for which $d(v, \tilde{A}_i^{(1)}) \leq \delta \log d$, note that for every $v \in V(G)$ and $i \in [\![h]\!]$, we have $A_i^{(1)} \subseteq \tilde{A}_i^{(1)}$ and thus $d(v, \tilde{A}_i^{(1)}) \geq d(v, A_i^{(1)})$. Hence, if $v \notin B_2$, then $d(v, \tilde{A}_i^{(1)}) > \delta \log d$ for every $i \in [\![h]\!]$. Further, note that if $v \in B_2$, then since $N(B_2) \setminus \bigcup_{i \in [\![h]\!]} A_i^{(1)} \subseteq U$, we have that $N(v) \setminus \bigcup_{i \in [\![h]\!]} A_i^{(1)} \subseteq U$. Hence,

$$|N(v) \cap U| = \left| N(v) \setminus \bigcup_{i \in [\![h]\!]} A_i^{(1)} \right| = d - \sum_{i \in [\![h]\!]} d(v, A_i^{(1)}) \geq d - h \cdot C \log d \geq 0.99d. \tag{3}$$

Thus, by the union bound and by Lemma 3.1, we have that

$$\begin{aligned}\mathbb{P}\left(\exists i \in [\![h]\!], d(v, \tilde{A}_i^{(1)}) \leq \delta \log d\right) &\leq h \cdot \mathbb{P}(d(v, U_1) \leq \delta \log d) \\ &\leq h \cdot \mathbb{P}\left(\mathrm{Bin}\left(0.99d, \frac{900}{k}\right) \leq \delta \log d\right) \leq d^{-101},\end{aligned}$$

completing the proof. □

We can now apply Corollary 3.3 and obtain the required sets.

Lemma 4.7. *With probability at least* $1/2 - o(1)$*, there exist disjoint subsets* $A_1^{(2)}, \ldots, A_h^{(2)} \subseteq V(G)$ *which satisfy the following.*

1. $\left| S_i \triangle A_i^{(2)} \right| = o_d(n/k)$ *for every* $i \in [\![h]\!]$.
2. $\left| A_i^{(2)} - \frac{n}{k} \right| = O(n/d^{50})$ *for every* $i \in [\![h]\!]$.
3. $d(v, A_i^{(2)}) \in [\delta \log d, 2C \log d]$ *for every* $i \in [\![h]\!]$ *and* $v \in V(G)$.

Proof. Similarly to the proof of Lemma 4.4, for every $v \in V(G)$, let F_v be the event that $v \in W_3$. Let $\mathcal{F} := \{F_v\}_{v \in V(G)}$. By Lemma 4.6, we have that $\mathbb{P}(F_v) \leq d^{-100} =: q$ for every $v \in V(G)$. Observe that every F_v is determined by $\Delta_1 := d$ random variables. Furthermore, every F_v depends on at most $\Delta_2 := d^2$ other events. Setting $\beta = 4d^{-100}$, we have that $\beta(1-\beta)^{\Delta_2} \geq q$ by (1). Thus, by Corollary 3.3, we obtain that with probability at least $\frac{1}{2}$, there exist sets $A_1^{(2)}, \ldots, A_h^{(2)}$ such that $d(v, A_i^{(2)}) \in [\delta \log d, 2C \log d]$ for every $i \in [\![h]\!]$ and $v \in V(G)$, obtaining the third item. Let $A_1^{(2)}, \ldots, A_h^{(2)}$ be these sets (if they do not exist, set $A_i^{(2)} = \tilde{A}_i^{(1)}$ for every $i \in [\![h]\!]$).

For the first item, again by Corollary 3.3 for every $i \in [\![h]\!]$,

$$\left| \tilde{A}_i^{(1)} \triangle A_i^{(2)} \right| \leq 2 \cdot d \cdot n \cdot \frac{4d^{-100}}{1 - 4d^{-100}} \leq \frac{n}{d^{51}}. \tag{4}$$

Recall that $\tilde{A}_i^{(1)} = A_i^{(1)} \cup U_i$. By Lemma 4.4, $\left|S_i \triangle A_i^{(1)}\right| = o_d(n/k)$, and by Lemma 3.1, **whp** $|U_i| = o_d(n/k)$ and thus $|\tilde{A}_i^{(1)} \triangle S_i| = o_d(n/k)$.

For the second item, we recall that

$$\left| |A_i^{(1)}| - \frac{(1-\alpha)n}{k} \right| \leq \frac{n}{d^{50}} \tag{5}$$

for every $i \in [\![h]\!]$. By construction, $|\tilde{A}_i^{(1)}| = |A_i^{(1)}| + |U_i|$. Now, for every $i \in [\![h]\!]$,

$$|U_i| \sim \mathrm{Bin}\left(\frac{\alpha n}{1000}, p_i\right),$$

where p_i is defined according to Lemma 4.5. In particular, p_i satisfies that $\left|\mathbb{E}[|U_i|] - \frac{\alpha n}{k}\right| = O(n/d^{50})$. By Lemma 3.1, we have that **whp**

$$|U_i| \in \left[\mathbb{E}[|U_i|] - n^{2/3}, \mathbb{E}[|U_i|] + n^{2/3}\right].$$

This, together with (4) and (5) yields that with probability at least $\frac{1}{2} - o(1)$, we have that $\left||A_i^{(2)}| - \frac{n}{k}\right| = O\left(n/d^{50}\right)$. □

4.2. Vertices of low degree in the tree

We have proved that, if $(S_1, \ldots, S_h)$ is nice (this happens with probability at least $1/2 - o(1)$), then there exist sets $A_1^{(2)}, \ldots, A_h^{(2)}$ satisfying the properties as in the statement of Lemma 4.7. Throughout this section, we fix a nice tuple $(S_1, \ldots, S_h)$ and a tuple $(A_1^{(2)}, \ldots, A_h^{(2)})$, satisfying the conclusion of Lemma 4.7.

Recall that the sets $A_1^{(2)}, \ldots, A_h^{(2)}$ correspond to the high-degree vertices in T, and every vertex has between $\delta \log d$ and $2C \log d$ neighbours in each of these sets. As described in the beginning of Section 4, in this section we aim to establish similar sets for low-degree vertices.

Let

$$U \coloneqq V(G) \setminus \bigcup_{i \in [\![h]\!]} A_i^{(2)}. \tag{6}$$

Note that, by Lemma 4.7,

$$\begin{aligned} |U| &\in \left[n - h \cdot \left(\frac{n}{k} + O\left(\frac{n}{d^{50}}\right)\right), n - h \cdot \left(\frac{n}{k} - O\left(\frac{n}{d^{50}}\right)\right)\right] \\ &= \left[\left(1 - \frac{h}{k} - O(d^{-49})\right) n, \left(1 - \frac{h}{k} + O(d^{-49})\right) n\right]. \end{aligned} \tag{7}$$

For every $v \in U$, let X_v be the random variable such that $\mathbb{P}(X_v = i) = \frac{1}{k-h}$, for every index $i \in \{h+1, \ldots, k\}$. All X_v are independent. For every $i \in \{h+1, \ldots, k\}$, set

$$S_i \coloneqq \{v \in U \colon X_v = i\}.$$

For convenience, let us also set $A_i^{(2)} \coloneqq S_i$ for every $i \in \{h+1, \ldots, k\}$ (recall that $A_i^{(2)}$ is already defined for every $i \in [\![h]\!]$).

Given a partition $U_1, \ldots, U_k$ of $V(G)$, let $B(U_1, \ldots, U_k)$ be the set of vertices $v \in V(G)$ satisfying the following. There exist $i \in [\![k]\!]$ and $j \in \{h+1, \ldots, k\}$ such that $\{i, j\} \in E(T), v \in U_i$ and $d(v, U_j) \leq \delta \log d$. Further, let $W(U_1, \ldots, U_k)$ be the set of vertices $v \in V(G)$ that satisfy at least one of the following.

1. There exists $i \in [\![k]\!]$ such that $d(v, U_i) \geq 2C \log d$.

2. There exist more than $1/\epsilon^2$ indices $i \in \{h+1, \dots, k\}$ such that $d(v, U_i) < \delta \log d$.

3. There exists $i \in [\![k]\!]$ such that $d\left(v, U_i \cap B(U_1, \dots, U_k)\right) > \log\log d$.

Set $B_4 \coloneqq B(A_1^{(2)}, \dots, A_k^{(2)})$ and $W_4 \coloneqq W(A_1^{(2)}, \dots, A_k^{(2)})$.

Let us first show that it is quite unlikely for a vertex to be in W_4.

Lemma 4.8. *$\mathbb{P}(v \in W_4) \le d^{-100}$ for every vertex $v \in V(G)$.*

Proof. Fix $v \in V(G)$. Since $h \le d^\beta$ and $d(v, A_i^{(2)}) \le 2C \log d$, for every $i \in [\![h]\!]$, we have

$$d(v, U) \ge d - h \cdot 2C \log d \ge (1 - 0.5\epsilon)d. \tag{8}$$

For the first item, by the union bound, the probability that there exists $i \in \{h+1, \dots, k\}$ such that $d(v, A_i^{(2)}) > 2C \log d$ is at most

$$k \cdot \mathbb{P}\left(\mathrm{Bin}\left(d(v, U), \frac{1}{k-h}\right) \ge 2C \log d\right) \le k \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1}{k-h}\right) \ge 2C \log d\right) < d^{-200},$$

where the last inequality is true whenever C is large enough by Lemma 3.1.

For the second item, note that the event that there exist more than $1/\epsilon^2$ indices $i \in \{h+1, \dots, k\}$ such that $d(v, A_i^{(2)}) < \delta \log d$ implies that there exist $1/\epsilon^2$ indices $i_1, \dots, i_{1/\epsilon^2} \in \{h+1, \dots, k\}$ such that $d\left(v, \bigcup_{j=1}^{1/\epsilon^2} A_{i_j}^{(2)}\right) < \frac{1}{\epsilon^2} \cdot \delta \log d$. Thus, by the union bound over all choices of these indices, this probability is at most $d^{1/\epsilon^2}\mathbb{P}\left(\mathrm{Bin}\left(d(v, U), \frac{1}{\epsilon^2(k-h)}\right) < \delta \cdot \frac{\log d}{\epsilon^2}\right)$. The expectation of this binomial random variable is

$$d(v, U) \cdot \frac{1}{\epsilon^2(k-h)} \ge (1 - 0.5\epsilon)d \cdot \frac{\log d}{\epsilon^2(1-\epsilon)d} \ge (1 + 0.5\epsilon) \cdot \frac{\log d}{\epsilon^2},$$

where we used (8) and $k \le (1-\epsilon)\frac{d}{\log d}$. Thus, for δ small enough, by Lemma 3.2

$$d^{1/\epsilon^2} \cdot \mathbb{P}\left(\mathrm{Bin}\left(d(v, U), \frac{1}{\epsilon^2(k-h)}\right) < \delta \cdot \frac{\log d}{\epsilon^2}\right) \le d^{1/\epsilon^2} \cdot e^{-(1+\epsilon/3)\log d/\epsilon^2} \le d^{-200},$$

where the last inequality is true whenever ϵ is small enough.

For the third item, we fix $i \in [\![k]\!]$ and bound the probability that $d(v, A_i^{(2)} \cap B_4) > \log\log d$ from above. If $i \in [\![h]\!]$, then $d(v, A_i^{(2)}) \le 2C \log d$ and every neighbour u of v in $A_i^{(2)}$ belongs to B_4 with probability at most

$$(k-h) \cdot \mathbb{P}\left(\mathrm{Bin}\left(d(u, U), \frac{1}{k-h}\right) \le \delta \log d\right) \le d^{-\epsilon/2}, \tag{9}$$

where the last inequality is true for δ sufficiently small by Lemma 3.2 and (8). The events that $u_1 \in B_4$ and $u_2 \in B_4$ for two different vertices in $A_i^{(2)} \cap N(v)$ might be dependent if their neighbourhoods intersect. Since $G \in \mathcal{G}_d$, removing at most two vertices from the 2-ball around v ensures that there are no more cycles in the second neighbourhood of v. However, this is still not enough to get rid of dependencies since u_1, u_2 have the common neighbour v. Nevertheless, after revealing the sets $A_i^{(2)}$ that the vertices of $N(v)$ belong to and deleting at most two vertices of $N(v)$ that belong to a cycle, we can upper bound the events $\{u \in B_4\}$, for all $u \in N(v) \cap A_i^{(2)}$, by independent events that do not consider the vertex v. In other words, we bound the probability that $u \in B_4$ from above by $\mathbb{P}\left(\mathrm{Bin}\left(d(u, U) - 1, \frac{1}{k-h}\right) \le \delta \log d\right)$, which satisfies the inequality (9) as well. Thus, we obtain that the probability that there are more than $\log\log d$ vertices in $d(v, A_i^{(2)} \cap B_4)$ is at most

$$\mathbb{P}\left(\mathrm{Bin}\left(2C \log d, d^{-\epsilon/2}\right) \ge \log\log d - 2\right) \le d^{-\Theta(\log\log d)} \le d^{-200},$$

where the first inequality is true by Lemma 3.1.

If $i \in \{h+1, \dots, k\}$, then the probability that a neighbour of v belongs to $A_i^{(2)} \cap B_4$ equals

$$\mathbb{P}\left(u \in A_i^{(2)} \cap B_4 \mid u \in N(v)\right) = \frac{1}{k-h} \cdot \mathbb{P}\left(u \in B_4 \mid u \in N(v) \cap A_i^{(2)}\right) \\ \leq \frac{1}{k-h} \cdot (k-h) \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1}{k-h}\right) < \delta \log d\right) \leq d^{-1-\epsilon/2}, \tag{10}$$

where the last inequality is true by Lemma 3.2 for δ sufficiently small and since $k-h < (1-\epsilon)d/\log d$. Similar to the previous argument (we again delete a possible cycle crossing $N(v)$, and then bound the events that $u \in B_4$ from above by events that do not consider the vertex v in the neighbourhood of u, that is the binomial random variable $\mathrm{Bin}\left(d, \frac{1}{k-h}\right)$ in (10) is replaced by $\mathrm{Bin}\left(d-1, \frac{1}{k-h}\right)$), since $G \in \mathcal{G}_d$, the probability that there are more than $\log\log d$ vertices in $N(v, A_i^{(2)} \cap B_4)$ is at most

$$\mathbb{P}\left(\mathrm{Bin}\left(d, d^{-1-\epsilon/2}\right) \geq \log\log d - 2\right) \leq d^{-\Theta(\log\log d)} \leq d^{-200},$$

where the first inequality is true by Lemma 3.1. □

We also require the following lemma.

Lemma 4.9. ***Whp*** *the following holds.*

1. $|B_4| \leq nd^{-(1+\epsilon/4)}$.
2. *For every* $i \in [\![k]\!]$, *we have* $\left||A_i^{(2)}| - \frac{n}{k}\right| = O(n/d^{49})$.
3. *For every* $i \in [\![k]\!]$, *there are at least* $\frac{n}{3k}$ *vertices* $v \in A_i^{(2)}$ *such that for every* $j \in [\![k]\!]$, $d(v, A_j^{(2)}) \geq \delta \log d$.

Proof. We start with the first item. Fix $v \in V(G)$. We consider two cases separately. Let us first assume $v \notin U$. In this case, we bound the event that $v \in B_4$ by the event that there exists $i \in \{h+1, \dots, k\}$ such that $d(v, A_i^{(2)}) \leq \delta \log d$. By Lemma 3.2, for δ sufficiently small and by (8),

$$\mathbb{P}\left(v \in B_4 \mid v \notin U\right) \leq (k-h) \cdot \mathbb{P}\left(\mathrm{Bin}\left(d(v,U), \frac{1}{k-h}\right) < \delta \log d\right) \leq d^{-\epsilon/2}. \tag{11}$$

Thus, by (7) and for β sufficiently small with respect to ϵ (recalling that $h \leq d^{\beta}$),

$$\mathbb{E}[B_4 \setminus U] \leq \left(\frac{hn}{k} + O\left(\frac{n}{d^{49}}\right)\right) \cdot d^{-\epsilon/2} \leq nd^{-1-\epsilon/3}.$$

Now, if $v \in U$, once again by Lemma 3.2,

$$\begin{aligned}\mathbb{P}\left(v \in B_4 \mid v \in U\right) &\leq \sum_{i=h+1}^{k} \mathbb{P}(v \in A_i^{(2)} \mid v \in U) \sum_{j \in N_T(i) \setminus [\![h]\!]} \mathbb{P}\left(d(v, A_j^{(2)}) < \delta \log d \mid v \in A_i^{(2)}\right) \\ &= \sum_{i=h+1}^{k} \mathbb{P}(v \in A_i^{(2)} \mid v \in U) \sum_{j \in N_T(i) \setminus [\![h]\!]} \mathbb{P}\left(\mathrm{Bin}\left(d(v,U), \frac{1}{k-h}\right) < \delta \log d\right) \\ &\overset{(8)}{\leq} \sum_{i=h+1}^{k} \frac{1}{k-h} \sum_{j \in N_T(i) \setminus [\![h]\!]} d^{-1-\epsilon/2} \leq 2e(T) \cdot \frac{1}{k-h} \cdot d^{-1-\epsilon/2} \leq d^{-1-\epsilon/3}.\end{aligned}$$

Hence, we got $\mathbb{E}[|B_4|] \le 2nd^{-1-\epsilon/3}$. Since $G \in \mathcal{G}_d$, we have by the same arguments verbatim as those in the proof of Lemma 4.4 that $\mathrm{Var}(|B_4|) = O(n)$. Thus, by Chebyshev's inequality, **whp** $|B_4| \le nd^{-(1+\epsilon/4)}$.

The second item holds by Lemma 4.7 for $i \in [\![h]\!]$. For $i \in \{h+1, \ldots, k\}$, we have that $|A_i^{(2)}|$ stochastically dominates $\mathrm{Bin}\left(\left(1 - \frac{h}{k} - O(d^{-49})\right) n, \frac{1}{k-h}\right)$, and $|A_i^{(2)}|$ is stochastically dominated by $\mathrm{Bin}\left(\left(1 - \frac{h}{k} + O(d^{-49})\right) n, \frac{1}{k-h}\right)$, and thus the second item follows from a simple application of Lemma 3.1.

For the third item, given $i \in [\![k]\!]$, denote by R_i the set of vertices $v \in A_i^{(2)}$ such that for every $j \in [\![k]\!]$, $d(v, A_j^{(2)}) \ge \delta \log d$. Note that for every $j \in [\![h]\!]$ and $v \in V(G)$, we have that $d(v, A_j^{(2)}) \ge \delta \log d$ deterministically. For every $j \in \{h+1, \ldots, k\}$, the probability that $v \in V(G)$ satisfies $d(v, A_j^{(2)}) < \delta \log d$ is at most $\mathbb{P}\left(\mathrm{Bin}\left(d(v, U), \frac{1}{k-h}\right) < \delta \log d\right) \le d^{-1-\epsilon/2}$ by Lemma 3.2 together with (8). Thus, by (11), $\mathbb{E}[|R_i|] \ge \frac{n}{2k}$. For every v, the event that $v \in R_i$ depends on at most d^4 other events. Thus, $\mathrm{Var}(|R_i|) = O(n)$, and by Chevyshev's inequality, **whp** for every $i \in [\![k]\!]$ we have that $|R_i| \ge \frac{n}{3k}$. $\square$

We now turn to use Corollary 3.3 to show that there exists a 'good' partition $A_1^{(3)}, \ldots, A_k^{(3)}$, which is not far from $S_1, \ldots, S_k$.

Lemma 4.10. *With probability at least $1/2 - o(1)$, there exists a partition of $V(G)$ into $A_1^{(3)}, \ldots, A_k^{(3)}$ which satisfies the following. Let*

$$B_5 = B\left(A_1^{(3)}, \ldots, A_k^{(3)}\right) \quad \text{and} \quad W_5 = W\left(A_1^{(3)}, \ldots, A_k^{(3)}\right).$$

Then,

1. $\left|A_i^{(3)} \triangle S_i\right| = o_d(n/k)$ *for every* $i \in [\![k]\!]$.
2. $\left||A_i^{(3)}| - \frac{n}{k}\right| = O(n/d^{49})$ *for every* $i \in [\![k]\!]$.
3. $W_5 = \varnothing$.
4. $|B_5| \le nd^{-1-\epsilon/5}$.
5. *For every* $i \in [\![k]\!]$, *there are at least* $\frac{n}{4k}$ *vertices* $v \in A_i^{(3)}$ *which satisfy that for every* $j \in [\![k]\!]$, $d(v, A_j^{(3)}) \ge \delta \log d$.
6. $d(v, A_i^{(3)}) \in [\delta \log d, 2C \log d]$ *for every* $i \in [\![h]\!]$ *and* $v \in V(G)$.

Proof. For every $v \in V(G)$, let F_v be the event that $v \in W_5$. Let $\mathcal{F} := \{F_v\}_{v \in V}$. By Lemma 4.8, we have that for every $v \in V(G)$, $\mathbb{P}(F_v) \le d^{-100} =: q$. Observe that every F_v is determined by at most $\Delta_1 := 1 + d^2$ random variables. Furthermore, every F_v depends on at most $\Delta_2 := d^4$ other events. Moreover, in the same way as in (1), $\beta := 4d^{-100}$ satisfies $\beta(1-\beta)^{\Delta_2} \ge q$. Thus, by Corollary 3.3, we obtain that with probability at least $\frac{1}{2}$, there exist sets $A_1^{(3)}, \ldots, A_k^{(3)}$ such that $W_5 = \varnothing$ and

$$\left|A_i^{(3)} \triangle A_i^{(2)}\right| \le 2 \cdot (1 + d^2) \cdot n \cdot \frac{4d^{-100}}{1 - 4d^{-100}} \le \frac{n}{d^{51}}$$

for every $i \in [\![k]\!]$. Let $A_1^{(3)}, \ldots, A_k^{(3)}$ be these sets (if they do not exist, set $A_i^{(3)} = A_i^{(2)}$ for every $i \in [\![k]\!]$). This, together with Lemmas 4.7 and 4.9, implies the first three properties.

We now move to the fourth property. By Lemma 4.9, we have that **whp** $|B_4| \le nd^{-1-\epsilon/4}$. By construction, **whp** $|B_5| \le |B_4| + d \sum_{i \in [\![k]\!]} \left|A_i^{(3)} \triangle A_i^{(2)}\right|$. Thus, by the above, we have $|B_5| \le nd^{-1-\epsilon/5}$ **whp**.

Similarly, for the fifth property, by Lemma 4.9, we have that **whp** for every $i \in [\![k]\!]$, there are at least $\frac{n}{3k}$ vertices $v \in A_i^{(2)}$ such that $d(v, A_j^{(2)}) \geq \delta \log d$ for every $j \in [\![k]\!]$. Since, by the above, we have moved only $O\left(\frac{n}{d^{50}}\right)$ vertices, we have that **whp**, for every $i \in [\![k]\!]$, there are at least $\frac{n}{4k}$ vertices $v \in A_i^{(3)}$ such that $d(v, A_j^{(3)}) \geq \delta \log d$ for every $j \in [\![k]\!]$.

For the last property, we simply note that for every $i \in [\![h]\!]$, $A_i^{(3)} = A_i^{(2)}$ and thus this follows from Lemma 4.7. $\square$

4.3. Eliminating bad vertices

We now say that a tuple $(S_1, \dots, S_k)$ is *nice* if $(S_1, \dots, S_h)$ is nice and there exist sets $A_i^{(3)}$, $i \in [\![k]\!]$, satisfying the conclusion of Lemma 4.10. Due to Lemmas 4.4 and 4.10, with probability at least $1/4 - o(1)$, the considered random tuple of sets $(S_1, \dots, S_k)$ is nice. As in the previous section, for every nice tuple $(S_1, \dots, S_k)$, we fix the corresponding sets $A_i^{(3)}$, $i \in [\![k]\!]$, satisfying the conclusion of Lemma 4.10. If $(S_1, \dots, S_k)$ is not nice, then we simply set $A_i^{(3)} = S_i$ for all $i \in [\![k]\!]$.

Further in this section, we assume that the event from Lemma 4.10 (that has probability at least $1/4 - o(1)$ due to Lemma 4.4), actually occurs; i.e. the tuple $(S_1, \dots, S_k)$ is *nice*. We recall that $A_1^{(3)}, \dots, A_k^{(3)}$ satisfy that for every vertex $v \in V(G)$ and an index $i \in [\![h]\!]$, we have $d(v, A_i^{(3)}) \in [\delta \log d, 2C \log d]$. Further, for every vertex $v \in V(G)$ and an index $i \in [\![k]\!]$, we have $d(v, A_i^{(3)}) \leq 2C \log d$. We now show that after several resamples, we may obtain sets $A_1^{(4)}, \dots, A_k^{(4)}$ such that for every $\{i, j\} \in E(T)$ and for every $v \in A_i^{(4)}$, the number of neighbours of v in $A_j^{(4)}$ is concentrated around its expectation. More precisely,

Lemma 4.11. *There exist sets $A_1^{(4)}, \dots, A_k^{(4)}$ such that the following holds:*

1. $\left|A_i^{(4)} \triangle A_i^{(3)}\right| \leq nd^{-1-\epsilon/5}$ *for every* $i \in [\![k]\!]$.
2. $d(v, A_j^{(4)}) > \frac{\delta \log d}{2}$ *for every* $\{i, j\} \in E(T)$ *and for every* $v \in A_i^{(4)}$.
3. $d(v, A_j^{(4)}) < 3C \log d$ *for every vertex* $v \in V(G)$ *and every* $j \in [\![k]\!]$.
4. *For every* $i \in [\![k]\!]$, *there are at least* $\frac{n}{5k}$ *vertices* $v \in A_i^{(4)}$ *which satisfy that for every* $j \in [\![k]\!]$, $d(v, A_j^{(4)}) \geq \frac{\delta \log d}{2}$.

Proof. For every vertex $v \in B_5$, denote by $I_v \subseteq \{h+1, \dots, k\}$ the set of indices $j \in \{h+1, \dots, k\}$ such that $d(v, A_j^{(3)}) \leq \delta \log d$. Set $\Gamma_v := \{h+1, \dots, k\} \setminus (I_v \cup N_T(I_v))$. Since $h \leq d^\beta$ and since for every $i \in \{h+1, \dots, k\}$ we have $d_T(i) \leq d^{1-\beta}$, and recalling that $W_5 = \varnothing$,

$$|\Gamma_v| \geq k - h - \epsilon^{-2} \cdot d^{1-\beta} = (1 - o_d(1))k. \tag{12}$$

For every vertex $v \in B_5$, denote by X_v the uniform random variable over the set of integers Γ_v. For every $i \in [\![k]\!]$, set

$$\tilde{A}_i^{(3)} := (A_i^{(3)} \setminus B_5) \cup \{v \in B_5 \colon X_v = i\}.$$

Fix $i \in [\![k]\!]$ and fix $v \in \tilde{A}_i^{(3)}$. Note that for any $j \in [\![k]\!]$,

$$d(v, \tilde{A}_j^{(3)}) \geq d(v, A_j^{(3)}) - d(v, B_5 \cap A_j^{(3)}) \geq d(v, A_j^{(3)}) - \log \log d,$$

where the second inequality is true by the third item in the definition of W_5. We have that at least $\frac{n}{4k}$ vertices in $A_i^{(3)}$ have at least $\delta \log d$ neighbours in $A_j^{(3)}$ for every $j \in [\![k]\!]$. This, together with the above and the bound on $|B_5|$ in Lemma 4.10, immediately implies the fourth item.

For the second item, fix $\{i,j\} \in E(T)$ and fix a vertex $v \in \tilde{A}_i^{(3)}$. If $v \notin B_5$ or $j \in [\![h]\!]$, then we have $d(v, A_j^{(3)}) > \delta \log d$ and thus

$$d(v, \tilde{A}_j^{(3)}) \geq d(v, A_j^{(3)}) - d(v, B_5 \cap A_j^{(3)}) \geq \delta \log d - \log\log d > \frac{\delta \log d}{2}.$$

Assume now that $v \in B_5$ and $j \in \{h+1, \dots, k\}$. In this case, we must have that $i \in \Gamma_v$. Hence, for every $\ell \in [\![k]\!]$ such that $\{i, \ell\} \in E(T)$, we have $d(v, A_\ell^{(3)}) > \delta \log d$. In particular, $\{i,j\} \in E(T)$. Therefore, we also have

$$d(v, \tilde{A}_j^{(3)}) \geq d(v, A_j^{(3)}) - d(v, B_5 \cap A_j^{(3)}) \geq \delta \log d - \log\log d > \frac{\delta \log d}{2}.$$

Note that the above holds deterministically (that is, independently of the values of all X_v).

Next, we show that for a fixed vertex $v \in V(G)$, the probability that there exists an index $i \in [\![k]\!]$ such that $d(v, \tilde{A}_i^{(3)}) \geq 3C \log d$ is at most d^{-199}. Fix a vertex $v \in V(G)$ and fix an index $i \in [\![k]\!]$. We have $d(v, A_i^{(3)}) < 2C \log d$. Thus, if $d(v, \tilde{A}_i^{(3)}) \geq 3C \log d$, then $d(v, B_5 \cap \tilde{A}_i^{(3)}) \geq C \log d$. Notice that for every vertex $u \in B_5$, by (12) we have that $|\Gamma_u| \geq 0.5k$ and thus the probability that $X_u = i$ (which implies that $u \in \tilde{A}_i^{(3)}$) is at most $2/k$. Therefore,

$$\mathbb{P}\left(d(v, B_5 \cap \tilde{A}_i^{(3)}) \geq C \log d\right) \leq \mathbb{P}\left(\mathrm{Bin}\left(d, 2/k\right) \geq C \log d\right) \leq d^{-200},$$

where the last inequality is true whenever C is sufficiently large. Hence, by the union bound over all indices $i \in [\![k]\!]$, we have that the probability that the third item fails for a vertex $v \in V(G)$ is at most $k \cdot d^{-200} \leq d^{-199}$. For every vertex $v \in V(G)$, denote by D_v the event that the third item in the statement fails for v. By the above, $\mathbb{P}(D_v) \leq d^{-199}$. Moreover, the event D_v is independent on all but at most d^4 other events D_u. Therefore, by Corollary 3.3, with probability at least $\frac{1}{2}$ there exist sets $A_1^{(4)}, \dots, A_k^{(4)}$ such that none of the events D_v holds. Fixing these sets, since $|B_5| \leq nd^{-1-\epsilon/5}$, we have that $|A_i^{(4)} \triangle A_i^{(3)}| \leq |B_5| \leq nd^{-1-\epsilon/5}$ for every $i \in [\![k]\!]$, completing the proof. □

4.4. Balancing the sets

We have proved that, if $(S_1, \dots, S_k)$ is nice (this happens with probability at least $1/4 - o(1)$), then there exist sets $A_1^{(4)}, \dots, A_k^{(4)}$ satisfying the properties as in the statement of Lemma 4.11. Throughout this section, we fix a nice tuple $(S_1, \dots, S_k)$ and a tuple $(A_1^{(4)}, \dots, A_k^{(4)})$, satisfying the conclusion of Lemma 4.11.

Recall that the sets $A_1^{(4)}, \dots, A_k^{(4)}$ have good degree distribution in between them, yet their size could be up to $nd^{-1-\epsilon/5}$-far from n/k. We now turn to show that there exist sets $A_1^{(5)}, \dots, A_k^{(5)}$, all 'close' to $A_1^{(4)}, \dots, A_k^{(4)}$, and all of size $\frac{n}{k} \pm O(nd^{-50})$, which still satisfy the 'good degrees' assumption. This will, in turn, allow us to complete the balancing of the sets deterministically, and obtain sets of size exactly $\frac{n}{k}$ which satisfy the 'good degrees' assumption.

To that end, let us reorder the sets such that $A_1^{(4)}, \dots, A_m^{(4)}$ are of size at least $\frac{n}{k}$, and $A_{m+1}^{(4)}, \dots, A_k^{(4)}$ are of size less than $\frac{n}{k}$, for some $m \in [\![k]\!]$. Further, for every $i \in [\![k]\!]$, let $\Delta_i = \left||A_i^{(4)}| - \frac{n}{k}\right|$, noting that by the first item in Lemma 4.11 and by the second item in Lemma 4.10,

$$\Delta_i \leq nd^{-1-\epsilon/6}. \tag{13}$$

We have that for every $i \in [\![k]\!]$, there are at least $\frac{n}{5k}$ vertices $v \in A_i^{(4)}$ such that $d(v, A_j^{(4)}) \in \left[\frac{\delta \log d}{2}, 3C \log d\right]$ for every $j \in [\![k]\!]$. For every $i \in [\![m]\!]$, let $Q_i \subseteq A_i(4)$ be a set of exactly $\frac{n}{5k}$ such vertices, and set $Q := \bigcup_{i \in [\![m]\!]} Q_i$.

For every $i \in [\![m]\!]$ and $v \in Q_i$, set $M_v \sim \text{Bernoulli}(p_i)$ where $p_i = \frac{\Delta_i}{n/5k}$. This Bernoulli random variable represents whether the vertex v is moved to the j-th set, for some $j \in \{m+1, \ldots, k\}$, or not. In addition, let Z_v be the random variable over the set $\{m+1, \ldots, k\}$ defined by $\mathbb{P}(Z_v = j) = \frac{\Delta_j}{\Delta_1 + \ldots + \Delta_m}$ for every $j \in \{m+1, \ldots, k\}$. Note that $\sum_{j \in \{m+1,\ldots,k\}} \mathbb{P}(Z_v = j) = 1$ since $\Delta_1 + \ldots + \Delta_m = \Delta_{m+1} + \ldots + \Delta_k$. The random variable Z_v represents the index $j \in \{m+1, \ldots, k\}$ for which the vertex v may move (it will indeed move to $A_j^{(4)}$ if and only if $M_v = 1$). We stress that for every $v \in Q$, M_v and Z_v are independent, and are also independent over different v. Let

$$\tilde{A}_i^{(4)} := \begin{cases} A_i^{(4)} \setminus \{v \in Q_i \colon M_v = 1\}, & i \in [\![m]\!] \\ A_i^{(4)} \cup \{v \in Q \colon M_v = 1 \text{ and } Z_v = i\}, & i \in \{m+1, \ldots, k\}. \end{cases}$$

Note that by the above construction, if $A_i^{(4)}$ is of size smaller than $\frac{n}{k}$, then we may only move vertices into it, whereas when $A_i^{(4)}$ is of size larger than $\frac{n}{k}$ we may only move vertices outside of it. Further, if the set $A_i^{(4)}$ is of size exactly $\frac{n}{k}$, then $\Delta_i = 0$, and thus the set will remain unchanged.

We first show some typical properties of the sets $\tilde{A}_1^{(4)}, \ldots, \tilde{A}_k^{(4)}$, noting that the probability measure here is induced by the random variables M_v and Z_v.

Lemma 4.12. ***Whp****, the sets $\tilde{A}_1^{(4)}, \ldots, \tilde{A}_k^{(4)}$ satisfy the following for every $i \in [\![k]\!]$.*

1. $\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right| = o_d(n/k)$.
2. $\left||\tilde{A}_i^{(4)}| - \frac{n}{k}\right| \le n^{2/3}$.

Proof. Let us first show that, for every $i \in [\![k]\!]$, the expectation of $|\tilde{A}_i^{(4)}|$ is n/k. This is equivalent to showing that, for every $i \in [\![k]\!]$, the expected number of vertices which were added/removed is Δ_i. Indeed, if $i \in [\![m]\!]$, then all vertices $v \in Q_i \subseteq A_i^{(4)}$ with $M_v = 1$ will no longer be in the i-th set. We have

$$\mathbb{E}[|\{v \in A_i^{(4)} \colon M_v = 1\}|] = \sum_{v \in Q_i} \mathbb{P}(M_v = 1) = |Q_i| \cdot \frac{\Delta_i}{n/5k} = \Delta_i.$$

For $i \in \{m+1, \ldots, k\}$, all vertices $v \in Q$ with $M_v = 1$ and $Z_v = i$ will move to the i-th set. We have

$$\begin{aligned} \mathbb{E}[|\{v \in V(G) \colon M_v = 1 \text{ and } Z_v = i\}|] &= \sum_{j=1}^{m} \sum_{v \in Q_j} \mathbb{P}(M_v = 1 \wedge Z_v = i) \\ &= \sum_{j=1}^{m} \sum_{v \in Q_j} \frac{\Delta_j}{n/5k} \cdot \frac{\Delta_i}{\Delta_1 + \cdots + \Delta_m} \\ &= \sum_{j=1}^{m} |Q_j| \cdot \frac{\Delta_j}{n/5k} \cdot \frac{\Delta_i}{\Delta_1 + \cdots + \Delta_m} \\ &= \sum_{j=1}^{m} \Delta_j \cdot \frac{\Delta_i}{\Delta_1 + \cdots + \Delta_m} \\ &= \Delta_i. \end{aligned}$$

Altogether, we obtain that, for every $i \in [\![k]\!]$,

$$\mathbb{E}\left[\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right|\right] = \Delta_i. \tag{14}$$

Let us now turn to estimate the probability that $\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right|$ deviates from Δ_i by more than $n^{2/3}$. Note that $\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right| \sim \mathrm{Bin}(|Q_i|, p_i)$ for every $i \in [\![m]\!]$. For every $i \in \{m+1, \dots, k\}$, $\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right|$ is distributed according to a sum of $|Q|$ Bernoulli random variables. By (14), the expectation of this sum is Δ_i. Thus, by Lemma 3.1, for any $i \in [\![k]\!]$

$$\mathbb{P}\left(\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right| \geq \Delta_i + n^{2/3}\right) \leq e^{-\frac{n^{1/3}}{4}}.$$

Now, if $\Delta_i < n^{2/3}$, then $\mathbb{P}\left(\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right| \leq \Delta_i - n^{2/3}\right) = 0$ for any $i \in [\![k]\!]$. Otherwise, by Lemma 3.1, for any $i \in [\![k]\!]$

$$\mathbb{P}\left(\left|\tilde{A}_i^{(4)} \triangle A_i^{(4)}\right| \leq \Delta_i - n^{2/3}\right) \leq e^{-\frac{n^{1/3}}{4}}.$$

Thus the probability that there exists $i \in [\![k]\!]$ such that $\left||\tilde{A}_1^{(4)}| - \frac{n}{k}\right| > n^{2/3}$ is at most $d \cdot e^{-\frac{n^{1/3}}{4}} = o(1)$. Using that $\Delta_i = o_d(n/k)$ by (13), we also obtain the first item of this lemma. □

Let $\hat{B}$ be the set of vertices $v \in V(G)$ satisfying at least one of the following.

- That there exists $\{i, j\} \in E(T)$ such that $v \in \tilde{A}_i^{(4)}$ and $d(v, \tilde{A}_j^{(4)}) \notin \left[\frac{\delta \log d}{3}, 4C \log d\right]$.
- v satisfies that $d(v, A_i^{(4)}) > \delta \log d/2$ for every $i \in [\![k]\!]$. Further, there is some $i \in [\![k]\!]$ such that $d(v, \tilde{A}_i^{(4)}) \leq \delta \log d/3$.

Lemma 4.13. *For every $v \in V(G)$, $\mathbb{P}\left(v \in \hat{B}\right) \leq d^{-100}$.*

Proof. For every $j \in \{m+1, \dots, k\}$, let $B_+(v, j) := \{u \in N(v) \cap Q : M_u = 1, Z_u = j\}$. That is, $B_+(v, j)$ is the set of neighbours of v which are moved to the j-th set. Note that $|B_+(v, j)|$ is the sum of independent Bernoulli random variables. Recalling that $d(v, A_i^{(4)}) \leq 3C \log d$ for every $i \in [\![k]\!]$, we have that

$$\begin{aligned}\mathbb{E}\left[|B_+(v, j)|\right] &= \sum_{i=1}^{m} \sum_{u \in N(v) \cap Q_i} \mathbb{P}(M_u = 1, Z_u = j) \leq \sum_{i=1}^{m} 3C \log d \cdot \frac{\Delta_i}{n/5k} \cdot \frac{\Delta_j}{\Delta_1 + \cdots + \Delta_m} \\ &= \frac{3C \log d \cdot \Delta_j}{n/5k} \leq \frac{1}{d^{\epsilon/7}},\end{aligned}$$

where we used $\Delta_j \leq n d^{-1-\epsilon/6}$. Thus, by Lemma 3.1,

$$\mathbb{P}\left(|B_+(v, j)| \geq \log \log d\right) \leq d^{-101}.$$

For every $j \in [\![m]\!]$, let $B_-(v, j) := \{u \in N(v) \cap A_j^{(4)} \cap Q \colon M_u = 1\}$. That is, $B_-(v, j)$ is the set of neighbours of v which are moved out of $A_j^{(4)}$. Recalling that $d(v, A_j^{(4)}) \leq 3C \log d$, we have that $|B_-(v, j)|$ is stochastically dominated by $\mathrm{Bin}\left(3C \log d, \frac{\Delta_i}{n/5k}\right)$. Thus, by Lemma 3.1,

$$\mathbb{P}\left(|B_-(v, j)| \geq \log \log d\right) \leq \mathbb{P}\left(\mathrm{Bin}\left(3C \log d, \frac{1}{d^{\epsilon/7}}\right) \geq \log \log d\right) \leq d^{-101},$$

where we used that $\Delta_i \leq n d^{-1-\epsilon/6}$.

Therefore, for any $v \in V(G)$ and $i \in [\![k]\!]$,

$$\mathbb{P}\left(\left|d(v, \tilde{A}_i^{(4)}) - d(v, A_i^{(4)})\right| \geq \log \log d\right) \leq 2d^{-101}. \tag{15}$$

Therefore, for the second item in the definition of $\hat{B}$, if v satisfies that $d(v, A_i^{(4)}) > \delta \log d/2$, then the probability that $d(v, \tilde{A}_i^{(4)}) \le \delta \log d/3$ is at most $2d^{-101}$. Similarly, for the first item, fix $i \in [\![k]\!]$ and j such that $\{i,j\} \in E(T)$. Fix $v \in \tilde{A}_i^{(4)}$. We have that $d(v, A_j^{(4)}) \in \left[\frac{\delta \log d}{2}, 3C \log d\right]$ for any j such that $\{i,j\} \in E(T)$. Indeed, if $v \in A_i^{(4)}$, this holds by Lemma 4.11. Otherwise, $v \in Q$ and then for every $j \in [\![k]\!]$ we have $d(v, A_j^{(4)}) \in \left[\frac{\delta \log d}{2}, 3C \log d\right]$. Similarly, the probability that $v \in \hat{B}$ is then at most $2d^{-101}$ due to (15). $\square$

We are now ready to apply Corollary 3.3.

Lemma 4.14. *With probability at least $\frac{1}{2} - o(1)$ (in the product measure induced by M_v and Z_v, $v \in Q$), there exist disjoint sets $A_1^{(5)}, \ldots, A_k^{(5)}$ such that the following holds.*

1. *For every $\{i,j\} \in E(T)$ and for every $v \in A_i^{(5)}$, we have that $d(v, A_j^{(5)}) \in \left[\frac{\delta \log d}{3}, 4C \log d\right]$.*
2. *For every $i \in [\![k]\!]$, we have $\left|A_i^{(5)} \triangle \tilde{A}_i^{(4)}\right| = O\left(\frac{n}{d^{50}}\right)$.*
3. *For every $i \in [\![k]\!]$, there are at least $\frac{n}{6k}$ vertices $v \in A_i^{(5)}$ which satisfy that $d(v, A_j^{(5)}) \in \left[\frac{\delta \log d}{3}, 4C \log d\right]$ for every $j \in [\![k]\!]$.*

Proof. For every $v \in V(G)$, let F_v be the event that $v \in \hat{B}$. Let $\mathcal{F} := \{F_v\}_{v \in V(G)}$. By Lemma 4.13, we have that for every $v \in V(G)$, $\mathbb{P}(F_v) \le d^{-100} =: q$. Observe that every F_v is determined by at most $\Delta_1 := 2d + 2$ random variables (M_u and Z_u for every $u \in \{v \cup N(v)\}$). Furthermore, every $W(v)$ depends on at most $\Delta_2 := 2d^4$ other events. Furthermore, similarly to (1), $\beta := 4d^{-100}$ satisfies $\beta(1-\beta)^{\Delta_2} \ge q$. Thus, by Corollary 3.3, we obtain that with probability at least $\frac{1}{2}$, there exist sets $A_1^{(5)}, \ldots, A_k^{(5)}$ such that $\hat{B} = \varnothing$ and $\left|A_i^{(5)} \triangle \tilde{A}_i^{(4)}\right| \le 2 \cdot (2d+2) \cdot n \cdot \frac{4d^{-100}}{1-4d^{-100}} \le \frac{n}{d^{51}}$ for every $i \in [\![k]\!]$. These sets then satisfy, by definition, the first two items. Fix $A_1^{(5)}, \ldots, A_k^{(5)}$ to be these sets (and if they do not exist, set $A_i^{(5)} = \tilde{A}_i^{(4)}$). Then, recall that by Lemma 4.11, for every $i \in [\![k]\!]$ there are at least $\frac{n}{5k}$ vertices $v \in A_i^{(4)}$ which satisfy that $d(v, A_j^{(4)}) \ge \frac{\delta \log d}{2}$ for every $j \in [\![k]\!]$. Thus, by the second item of $\hat{B}$, the sets $A_1^{(5)}, \ldots, A_k^{(5)}$ satisfy that for every $i \in [\![k]\!]$, there are at least $\frac{n}{6k}$ vertices $v \in A_i^{(5)}$ which satisfy that $d(v, A_j^{(5)}) \in \left[\frac{\delta \log d}{3}, 4C \log d\right]$ for every $j \in [\![k]\!]$. $\square$

We have just proved that there exist sets $A_1^{(5)}, \ldots, A_k^{(5)}$ satisfying the properties as in the statement of Lemma 4.14. We are now ready to complete the proof of Proposition 4.1. To that end, let us first show we can move the vertices between the sets $A_1^{(5)}, \ldots, A_k^{(5)}$ to obtain sets $V_1, \ldots, V_k$, each with exactly $\frac{n}{k}$ vertices, while maintaining the degree distribution between the sets.

Lemma 4.15. *There exists sets $V_1, \ldots, V_k$ such that the following holds.*

1. *For every $\{i,j\} \in E(T)$ and for every $v \in V_i$, we have that $d(v, V_j) \in \left[\frac{\delta \log d}{4}, 5C \log d\right]$.*
2. *$\left|V_i \triangle A_i^{(5)}\right| = o_d(n/k)$ for every $i \in [\![k]\!]$.*
3. *$|V_i| = \frac{n}{k}$ for every $i \in [\![k]\!]$.*

Proof. It suffices to show that we can move vertices from sets of size larger than $\frac{n}{k}$ to sets of smaller size, without changing the degree of any vertex into any of the sets by more than one. Consider the following procedure. We start with the sets $A_1^{(5)}, \ldots, A_k^{(5)}$.

Recall that for every $i \in [\![k]\!]$, at least $\frac{n}{6k}$ vertices $v \in A_i^{(5)}$ satisfy $d(v, A_j^{(5)}) \in \left[\frac{\delta \log d}{3}, 4C \log d\right]$ for every $j \in [\![k]\!]$. Furthermore, by the second item in Lemma 4.12 together with the second item in Lemma 4.14, $\left||A_i^{(5)}| - \frac{n}{k}\right| = O(n/d^{50})$ for every $i \in [\![k]\!]$.

We proceed inductively. Suppose we have already moved vertices $v_1, \ldots, v_t$ and that there still exists a set $A_i^{(5)}$ of size larger than n/k. In particular, $t \leq k \cdot O(n/d^{50}) < n/d^{45}$. Note that

$$\frac{n}{6k} - d^2 t \geq \frac{n}{7k} > 0,$$

and thus there exists a vertex $v \in A_i^{(5)}$ which satisfies the following two properties:

- v is not in the second neighbourhood of any $v_1, \ldots, v_t$, and,
- for every $j \in [\![k]\!]$, the number of neighbours of v in $A_j^{(5)}$ lies in the interval $\left[\frac{\delta \log d}{3}, 5C \log d\right]$.

We move the vertex v to an arbitrary set of size smaller than $\frac{n}{k}$. Observe that the above two properties guarantee that throughout the entire process, the degree of any vertex into any set will change by at most one. $\square$

Proof of Proposition 4.1. We have showed that, with probability $1/4 - o(1)$ (in the measure induced by $(S_1, \ldots, S_k)$), there exist sets $V_1, \ldots, V_k$ that satisfy item 1 and item 3 from Lemma 4.15 and that $|V_i \triangle S_i| = o_d(n/k)$ for every $i \in [\![k]\!]$, due to Lemmas 4.10(1), 4.11(1), 4.14(2), and 4.15(2).

Now, let $\tilde{S}_1, \ldots, \tilde{S}_k$ be a uniformly random partition of $V(G)$: every $v \in V(G)$ is assigned to $\tilde{S}_i$ for an index $i \in [\![k]\!]$ chosen uniformly at random, independently of all the other vertices. All that is left then is to observe that there is a coupling $(S'_i, \tilde{S}_i)$ such that $(S'_1, \ldots, S'_k) \stackrel{d}{=} (S_1, \ldots, S_k)$ and $|S'_i \triangle \tilde{S}_i| = o_d(n/k)$ for every $i \in [\![k]\!]$ **whp**.

Indeed, consider the following coupling. Initially, for every $i \in [\![k]\!]$, we set $S'_i = \tilde{S}_i$. We then keep every vertex $v \in S'_i$, for every $i \in [\![h]\!]$, with probability $1 - \alpha$, and with probability α we remove v from S_i. Let N_1 be the set of removed vertices. Recall that the choice of α is according to Lemma 4.2, thus we removed at most $n/d^{1+\epsilon/5}$ vertices from every set S'_i **whp**. Observe that $(S'_1, \ldots, S'_h) \stackrel{d}{=} (S_1, \ldots, S_h)$ and **whp** $|S'_i \triangle \tilde{S}_i| = o_d(n/k)$. However, sets $S'_{h+1}, \ldots, S'_k$ should be still perturbed since the set U from (6) is obtained after two resamples due to Corollary 3.3. Thus, we now consider the first two applications of the algorithmic Lovász Local Lemma. By Lemmas 4.4 and 4.7 and since we defined $A_i^{(1)} := S_i$ for partitions $(S_1, \ldots, S_h)$ that do not satisfy the event from the statement of Lemma 4.4, for every $i \in [\![h]\!]$, the number of vertices which are moved inside/outside of S_i is $o_d(n/k)$. Denote by N_2^+ and N_2^- the sets of vertices which were moved in these first two applications of the algorithmic Lovász Local Lemma from outside of $S'_1 \cup \ldots \cup S'_h$ to $A_1^{(2)} \cup \ldots \cup A_h^{(2)}$ and from $S'_1 \cup \ldots \cup S'_h$ outside of $A_1^{(2)} \cup \ldots \cup A_h^{(2)}$, respectively. We have that $|N_2^-| = o_d(n)$ and $|N_2^+| = o_d(n)$. We then partition the set $N_1 \cup N_2^- \setminus N_2^+$ that has size $o_d(n)$ **whp** uniformly at random into $S''_{h+1}, \ldots, S''_k$. Letting $S'_i := S'_i \cup S''_i$, for every $i \in [\![k]\!] \setminus [\![h]\!]$, and recalling that $k - h = (1 - o_d(1))k$, we get that $(S'_1, \ldots, S'_k) \stackrel{d}{=} (S_1, \ldots, S_k)$ and that $|S'_i \triangle \tilde{S}_i| = o_d(n/k)$ for every $i \in [\![k]\!]$ **whp**, completing the proof. $\square$

4.5. Small trees

For small trees we prove a stronger version of Proposition 4.1:

Proposition 4.16. *Let $k \leq \frac{d}{10 \log d}$. Let $G \in \mathcal{G}_d$, and suppose that n is divisible by k. Then, there exists a sufficiently large constant $C := C(\epsilon) > 0$ and a sufficiently small constant $\delta := \delta(\epsilon) > 0$ such that the following holds.*

Let $S_1, \ldots, S_k$ be a uniformly random partition of $V(G)$: for every $i \in [\![k]\!]$ and for every $v \in V(G)$, the vertex v belongs to S_i with probability $1/k$, independently of all the other vertices.

Then, with probability bounded away from zero, there are disjoint sets $V_1, \ldots, V_k \subseteq V(G)$, each of size $\frac{n}{k}$, with the following properties.

(P1) $|S_i \triangle V_i| = o_d(n/k)$ *for every* $i \in [\![k]\!]$.

(P2) $d(v, V_i) \in \left[\frac{\delta d}{k}, \frac{Cd}{k}\right]$ *for every* $i \in [\![k]\!]$ *and* $v \in V(G)$.

Proof. We begin by assigning to every vertex $v \in V(G)$ a random variable X_v supported on $[\![k]\!]$, where $\mathbb{P}(X_v = i) = \frac{1}{k}$ for every $i \in [\![k]\!]$. For every $i \in [\![k]\!]$, set $S_i := \{v \in V(G) \colon X_v = i\}$. Let W be the set of vertices $v \in V(G)$ such that there exists $i \in [\![k]\!]$ for which $d(v, S_i) \notin [\delta d/k, Cd/k]$. Now, by Lemma 3.1,

$$\begin{aligned}\mathbb{P}\left(\exists i \in [\![k]\!], d(v, S_i) \geq Cd/k\right) &\leq d \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1}{k}\right) \geq Cd/k\right) \\ &\leq d \cdot e^{-\frac{Cd}{4k}} \leq d^{-100},\end{aligned}$$

for sufficiently large C. Further,

$$\begin{aligned}\mathbb{P}\left(\exists i \in [\![k]\!], d(v, S_i) \leq \delta d/k\right) &\leq d \cdot \mathbb{P}\left(\mathrm{Bin}\left(d, \frac{1}{k}\right) \leq \delta d/k\right) \\ &\leq d^2 \binom{d}{\delta d/k} k^{-\delta d/k} (1 - 1/k)^{(1-\delta/k)d} \\ &\leq d^2 \exp\left(\frac{d}{k}(\delta \ln(e/\delta) - 1 + \delta)\right) \\ &\leq d^2 \cdot d^{-10(1-\delta(1+\ln(e/\delta)))} \leq d^{-7},\end{aligned}$$

where we used that $k \leq \frac{d}{10 \log d}$.

For every $v \in V(G)$, let F_v be the event that $v \in W$, and let $\mathcal{F} := \{F_v\}_{v \in V(G)}$. Observe that every F_v is determined by $\Delta_1 := d$ random variables (its neighbours). Furthermore, every F_v depends on at most $\Delta_2 := d^2$ other events (revealing whether a vertex v satisfies F_v may only affect the probability that u satisfies F_u for u which is in the second neighbourhood of v). Furthermore, note that $\beta := 4d^{-4}$ satisfies

$$\beta(1-\beta)^{\Delta_2} = 4d^{-4}(1 - 4d^{-4})^{d^2} \geq 4d^{-4} e^{-d^{-2}} \geq d^{-4} =: q > \mathbb{P}(F_v), \qquad \text{for every } v \in V(G).$$

By Corollary 3.3, we obtain that with probability at least $\frac{1}{2}$, there are sets $A_1, \ldots, A_k$ which satisfy the following for every $i \in [\![k]\!]$.

- $|S_i \triangle A_i| = O(n/d^3) = o_d(n/k)$.
- $d(v, A_i) \in \left[\frac{\delta d}{k} + 1, \frac{Cd}{k} - 1\right]$ for every $v \in V(G)$.

We now fix these sets. Due to the first item and the Chernoff bound, we also get $\left||A_i| - \frac{n}{k}\right| = O(n/d^3)$ **whp**.

We now want to 'balance the sets', that is, obtain sets $V_1, \ldots, V_k$, each of size exactly $\frac{n}{k}$, which satisfy that, for every $i \in [\![k]\!]$, $|S_i \triangle V_i| = o_d(n/k)$ and $d(v, V_i) \in \left[\frac{\delta d}{k} + 1, \frac{Cd}{k} - 1\right]$ for every $v \in V(G)$. To that end, we follow the same proof as in Lemma 4.15 — we iteratively move vertices from sets A_i of size larger than $\frac{n}{k}$ into sets A_j of size smaller than $\frac{n}{k}$, each time choosing a vertex which is not in the second neighbourhood of the previous vertices (and thus the degree of any vertex does not change by more than one into any of the sets), and utilise that $d^2 \cdot O(n/d^3) = o_d(n/k)$. After this procedure, we obtain the required sets. □

5. Growing the trees

In this section, we show how to construct the tree factor, given the vertex partition guaranteed by Propositions 4.1 and 4.16.

In Section 5.1, we collect typical properties of $G(n,d)$ which are important for proving Theorem 1. In Section 5.2, we prove Theorem 1. We do so by applying Proposition 4.1 (or its stronger version, Proposition 4.16, when trees are small) and, using the typical properties we have shown in Section 5.1, we will find a perfect matching between V_i and V_j for every $\{i,j\} \in E(T)$.

5.1. Typical Properties of random regular graphs

The following claim shows that typically there are not 'too many' edges between any two small sets of equal size.

Claim 5.1. *For every $\epsilon, \delta > 0$, there exists $\eta > 0$ such that, for sufficiently large d, **whp** the following holds in $G \sim G(n,d)$. Let $k \le \frac{(1-\epsilon)d}{\log d}$ be an integer. Then, for every two disjoint sets $A, B \subseteq V(G)$ satisfying $|A| = |B| < \eta \cdot \frac{n}{k}$,*

$$e(A,B) < |A| \cdot \delta \cdot \frac{d}{k}.$$

Proof. We have at most n choices for the sizes of the sets A and B. For a fixed size a, we have at most $\binom{n}{a}^2$ choices for the sets of such size. By the union bound over the (at most) n values of a and the number of choices of the sets, it suffices to prove that, for every fixed A and B satisfying $|A| = |B| =: a < \eta \cdot \frac{n}{k}$, we have

$$\mathbb{P}\left(e(A,B) \ge a \cdot \delta \cdot \frac{d}{k}\right) = o\left(\frac{1}{n\binom{n}{a}^2}\right) \qquad \text{uniformly in } a.$$

Fix $a < \eta \cdot \frac{n}{k}$ and set $\ell = a \cdot \delta \cdot \frac{d}{k}$. Note that $2\ell = \frac{2a\delta d}{k} < \frac{2\eta\delta n d}{k^2} < \delta n d$, and thus $M - 2\ell \ge \frac{nd}{3} = \omega(d^2)$. Fix two disjoint sets A and B of size a. Note that $d(A) = ad \le \eta \cdot \frac{nd}{k} < nd/2 = M/2$ where $M := \sum_v d(v)$ (and the second inequality holds for $\eta > 0$ sufficiently small). Thus, we may apply Corollary 3.4 and obtain,

$$\mathbb{P}(e(A,B) \ge \ell) \le \binom{d(A)}{\ell}\left(\frac{d(B)}{M(1+o(1))}\right)^{\ell} \le \left(\frac{ead}{\ell}\right)^{\ell}\left(\frac{ad}{nd(1+o(1))}\right)^{\ell} = \left(\frac{ea^2d}{\ell n(1+o(1))}\right)^{\ell}.$$

Thus,

$$\begin{aligned}
\mathbb{P}(e(A,B) \ge \ell) \cdot n\binom{n}{a}^2 &\le \left(\frac{ea^2d}{\ell n(1+o(1))}\right)^{\ell} \cdot n\binom{n}{a}^2 \le \left(\frac{ea^2d}{\ell n(1+o(1))}\right)^{\ell} \cdot n\left(\frac{en}{a}\right)^{2a} \\
&\le \left(\frac{2eak}{\delta n}\right)^{a\cdot\delta\cdot\frac{d}{k}} \cdot n\left(\frac{en}{a}\right)^{2a} \\
&= \exp\left\{\log n + a\left[2\log\left(\frac{en}{a}\right) - \delta \cdot \frac{d}{k} \cdot \log\left(\frac{\delta n}{2eak}\right)\right]\right\}.
\end{aligned}$$

Observe that the function $f(a) = 2\log\left(\frac{en}{a}\right) - \delta \cdot \frac{d}{k} \cdot \log\left(\frac{\delta n}{2eak}\right)$ is increasing for all $a > 0$ whenever d is large enough. Indeed,

$$f'(a) = -\frac{2}{a} + \frac{\delta d}{ak} \ge \frac{1}{a}\left(\frac{\delta \cdot \log d}{(1-\epsilon)} - 2\right) > 0,$$

where the first inequality is true since $k \leq \frac{(1-\epsilon)d}{\log d}$. Moreover, we have that

$$f\left(\eta \cdot \frac{n}{k}\right) = 2\log\left(\frac{ek}{\eta}\right) - \delta \cdot \frac{d}{k} \cdot \log\left(\frac{\delta}{2e\eta}\right)$$
$$\leq 2\log(d) - \frac{\delta}{1-\epsilon} \cdot \log d \cdot \log\left(\frac{\delta}{2e\eta}\right) < -1,$$

where the last inequality is true whenever η is sufficiently smaller than δ. Thus, if $a \leq n^{1-\xi}$ for some small constant $\xi > 0$, then

$$\begin{aligned}\mathbb{P}(e(A,B) \geq \ell) \cdot n\binom{n}{a}^2 &\leq \exp\left\{\log n + a \cdot f\left(n^{1-\xi}\right)\right\} \leq \exp\left\{\log n + f\left(n^{1-\xi}\right)\right\} \\ &= \exp\left\{\log n + \left[2\log(en^{\xi}) - \delta \cdot \frac{d}{k} \cdot \log\left(n^{\xi} \cdot \frac{\delta}{2ek}\right)\right]\right\} \\ &\leq \exp\left\{\log n + 3\xi \log n - 0.5\xi\delta \cdot \frac{d}{k} \cdot \log n\right\} = o(1),\end{aligned}$$

where the last equality is true for large enough d since $\frac{d}{k} \geq \log d$.

Further, if $n^{1-\xi} \leq a \leq \eta \cdot \frac{n}{k}$, then

$$\mathbb{P}(e(A,B) \geq \ell) \cdot n\binom{n}{a}^2 \leq e^{\log n + a \cdot f(a)} \leq e^{\log n - a} \leq e^{\log n - n^{1-\xi}} = o(1).$$

□

Claim 5.1 is useful in showing Hall's condition between small sets in V_i and V_j (that is, every small set $U \subseteq V_i$ has many neighbours in V_j). We would like to bound the neighbourhoods of large sets as well. This is the essence of the next claim.

Claim 5.2. *For every $\epsilon, \eta > 0$, there exist $\epsilon_1, \epsilon_2 > 0$ such that, for sufficiently large d, the following holds in $G \sim G(n,d)$. Let $2 \leq k \leq \frac{(1-\epsilon)d}{\log d}$ be an integer and let $S_1, \ldots, S_k$ be a uniformly random partition of $V(G)$. Then,* ***whp****, for every $i \neq j \in [\![k]\!]$ and $A \subset S_i$ satisfying $\eta \frac{n}{k} \leq |A| \leq 0.5(1+\epsilon_1)\frac{n}{k}$, we have $|N(A, S_j)| \geq (1+\epsilon_2)|A|$.*

Proof. Let $\xi > 0$ be a sufficiently small constant. We first show that **whp**, for every $i, j \in [\![k]\!]$ (not necessarily different), we have that $d(v, S_j) \in [(1-\xi)d/k, (1+\xi)d/k]$ for all but at most $o_d(n/k)$ of $v \in S_i$. Fix $i, j \in [\![k]\!]$. For every vertex $v \in V(G)$, denote by Z_v the indicator random variable of the event that the number of neighbours of v in S_j is not in the interval $[(1-\xi)d/k, (1+\xi)d/k]$. Set $Z \coloneqq \sum_{v \in V(G)} Z_v \cdot \mathbf{1}_{v \in S_i}$. By Lemma 3.1, for every vertex v,

$$\mathbb{P}(Z_v = 1) = \mathbb{P}\left(\left|\mathrm{Bin}\left(d, \frac{1}{k}\right) - \frac{d}{k}\right| > \xi\frac{d}{k}\right) \leq 2 \cdot e^{-\frac{\xi^2 d}{3k}} \leq d^{-\xi^2/4},$$

where the last inequality follows since $\frac{d}{k} \geq \frac{\log d}{1-\epsilon}$.

Further, the events that $v \in S_i$ and $Z_v = 1$ are independent. Hence, we have

$$\mathbb{E}[Z] \leq \frac{n}{k \cdot d^{\xi^2/4}}. \tag{16}$$

For every vertex $v \in V(G)$, the random variable $Z_v \cdot \mathbf{1}_{v \in S_i}$ is independent of all but at most d^2 other random variables $Z_u \cdot \mathbf{1}_{u \in S_i}$. Therefore,

$$\begin{aligned}\mathrm{Var}(Z) &= \sum_{v,u \in V(G)} \mathrm{Cov}(Z_v \cdot \mathbf{1}_{v \in S_i}, Z_u \cdot \mathbf{1}_{u \in S_i}) \\ &\leq \sum_{v \in V(G)} d^2 \cdot \max_{u \in V(G)} \mathbb{P}(Z_v \cdot \mathbf{1}_{v \in S_i} = Z_u \cdot \mathbf{1}_{u \in S_i} = 1) \leq d^2 n.\end{aligned}$$

By Chebyshev's inequality,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| > \frac{n}{d^{\xi^2/4}k}\right) = o_n(1).$$

Therefore, by the union bound and due to (16), **whp** for every $i, j \in [\![k]\!]$, we have that all but at most $o_d(n/k)$ vertices $v \in S_i$ satisfy $d(v, S_j) \in [(1-\xi)d/k, (1+\xi)d/k]$.

Let $C > 0$ be a large constant. Similarly to the previous argument, we can show that **whp** for every $i, j \in [\![k]\!]$, in every S_i there are $o(n/d^{100})$ vertices of degree larger than $C \cdot d/k$ into S_j. In addition, by a straightforward application of Lemma 3.1 we have that **whp** $|S_i| = \frac{n}{k} + O(n^{0.51})$ for every $i \in [\![k]\!]$.

Assume that the above holds deterministically. That is, for every $i, j \in [\![k]\!]$, we have $d(v, S_j) \in [(1-\xi)d/k, (1+\xi)d/k]$ for all but at most $o_d(n/k)$ vertices $v \in S_i$ and that $d(v, S_j) \le C \cdot d/k$ for all but at most $o(n/d^{100})$ vertices $v \in S_i$.

Note that, for every $i, j \in [\![k]\!]$ and any set $A \subseteq S_i$ of size at least $\eta n/k$, we have

$$e(A, S_j) \ge (|A| - o_d(n/k)) \cdot (1-\xi)\frac{d}{k} > |A| \cdot (1-2\xi)\frac{d}{k}, \tag{17}$$

where the last inequality is true for sufficiently large d. Notice that $e(A, S_j) = e(A, N(A, S_j))$. Thus, if we show that for every choice of B of size $(1+\epsilon_2)|A|$, we have that $e(A, B) \le |A| \cdot (1-2\xi)\frac{d}{k}$, then it implies that $|N(A, S_j)| \ge (1+\epsilon_2)|A|$. Indeed, if $|N(A, S_j)| < (1+\epsilon_2)|A|$, then we may take $N(A, S_j) \subseteq B \subseteq S_j$ to be of size $(1+\epsilon_2)|A|$ and get $e(A, S_j) = e(A, B) \le |A| \cdot (1-2\xi)\frac{d}{k}$, a contradiction to (17).

We now show that **whp** $e(A, B) \le |A| \cdot (1-2\xi)\frac{d}{k}$ for every $i \ne j \in [\![k]\!]$, $A \subseteq S_i$ satisfying $\eta\frac{n}{k} \le |A| \le 0.5(1+\epsilon_1)\frac{n}{k}$ and $B \subseteq S_j$ of size $(1+\epsilon_2)|A|$.

Fix $i \ne j \in [\![k]\!]$, fix $\eta\frac{n}{k} \le a \le 0.5(1+\epsilon_1)\frac{n}{k}$, fix $A \subset S_i$ of size a and fix $B \subseteq S_2$ of size $(1+\epsilon_2)a$. We now prepare the ground for the usage of Lemma 3.5.

Consider the graph $\tilde{G}$ induced by $S_i \cup S_j$. Note that $\tilde{G}$, given its degree sequence, has a uniform distribution. Moreover, all but at most $o_d(n/k)$ vertices from $S_i \cup S_j$ have degree at least $(1-\xi)\frac{d}{k}$ to $S_i \cup S_j$. We also have $|S_i \cup S_j| = \frac{2n}{k} + O(n^{0.51})$. Hence,

$$\sum_{v \in S_i \cup S_j} d_{\tilde{G}}(v) \ge (|S_i \cup S_j| - o_d(n/k)) \cdot 2(1-\xi)\frac{d}{k} \ge 4(1-2\xi) \cdot \frac{n}{k} \cdot \frac{d}{k} \ge 8(1-3\xi) \cdot a \cdot \frac{d}{k},$$

where the last inequality is true since $a \le 0.5(1+\epsilon_1)\frac{n}{k}$ and whenever ϵ_1 is sufficiently smaller than ξ.

Furthermore, since all but at most $o_d(n/k)$ vertices in $S_i \cup S_j$ have degree at most $2(1+\xi)\frac{d}{k}$ into $S_i \cup S_j$ and all but at most n/d^{100} vertices in $S_i \cup S_j$ have degree at most $C \cdot \frac{d}{k}$ into $S_i \cup S_j$, we have

$$\begin{aligned} d_{\tilde{G}}(A) &\le 2(1+\xi) \cdot \frac{d}{k} \cdot |A| + 2C \cdot \frac{d}{k} \cdot o_d\left(\frac{n}{k}\right) + 2d \cdot \frac{n}{d^{100}} \\ &= 2(1+\xi) \cdot a \cdot \frac{d}{k} + 2C \cdot \frac{d}{k} \cdot o_d\left(\frac{n}{k}\right) + 2d \cdot \frac{n}{d^{100}} \le 2(1+2\xi) \cdot a \cdot \frac{d}{k}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} d_{\tilde{G}}(B) &\le 2(1+\xi) \cdot \frac{d}{k} \cdot |B| + 2C \cdot \frac{d}{k} \cdot o_d\left(\frac{n}{k}\right) + 2d \cdot \frac{n}{d^{100}} \\ &\le 2(1+\xi) \cdot (1+\epsilon_2)a \cdot \frac{d}{k} + 2C \cdot \frac{d}{k} \cdot o_d\left(\frac{n}{k}\right) + 2d \cdot \frac{n}{d^{100}} \le 2(1+2\xi) \cdot a \cdot \frac{d}{k}, \end{aligned}$$

where the last inequality is true whenever ϵ_2 is sufficiently smaller than ξ.

Hence, by Lemma 3.5 (applied with $t = a \cdot \frac{d}{k}$),

$$\mathbb{P}\left(e(A,B) \geq (1-2\xi)a \cdot \frac{d}{k}\right) \leq 0.95^{a\cdot\frac{d}{k}} \leq 0.95^{a\log d},$$

since $k \leq \frac{(1-\epsilon)d}{\log d}$. Hence, by the union bound, the probability that there exist such sets A and B such that $e(A,B) > |A| \cdot (1-2\xi)\frac{d}{k}$ is at most

$$\begin{aligned} d^2 \cdot \sum_{a=\eta\frac{n}{k}}^{(0.5+\epsilon_1)\frac{n}{k}} \binom{n/k + O(n^{0.51})}{a}\binom{n/k+O(n^{0.51})}{(1+\epsilon_2)a} 0.95^{a\log d} &\leq \sum_{a=\eta\frac{n}{k}}^{(0.5+\epsilon_1)\frac{n}{k}} e^{3a\log(10n/(ak))} 0.95^{a\log d} \\ &\leq \sum_{a=\eta\frac{n}{k}}^{(0.5+\epsilon_1)\frac{n}{k}} e^{3a\log(10/\eta)} 0.95^{a\log d} \\ &\leq ne^{-a} = o_n(1), \end{aligned}$$

where the last inequality is true if d is sufficiently large. □

5.2. Proof of Theorem 1

Let $\epsilon > 0$ be a constant and let d be a sufficiently large integer. Let T be a tree on $k \leq \frac{(1-\epsilon)d}{\log d}$ vertices. We will show that **whp** $G \sim G(n,d)$ contains a T-factor. Let $\delta = \delta(\epsilon) > 0$ and $C = C(\epsilon) > 0$ be the constants guaranteed by Proposition 4.1. In addition, let $\eta = \eta(\epsilon,\delta)$ be the constant guaranteed by Claim 5.1 and let $\epsilon_1 = \epsilon_1(\eta^2,\epsilon)$ and $\epsilon_2 = \epsilon_2(\eta^2,\epsilon)$, be the constants guaranteed by Claim 5.2.

Now, let $S_1,\dots,S_k$ be such that every $v \in V(G)$ is assigned to S_i for an index $i \in [\![k]\!]$ chosen uniformly at random, independently from all the other vertices. Note that **whp** $G(n,d) \in \mathcal{G}_d$ (see, for example, [30]) and the statements of Claims 5.1 and 5.2, are satisfied. We then fix a deterministic $G \in \mathcal{G}_d$ that satisfies conclusions of both claims.

Let Σ be the set of all partitions of $V(G)$ into k ordered sets. Let $\Sigma' \subset \Sigma$ be the set of all *nice* $(S_1,\dots,S_k)$, i.e. those that satisfy the conclusion of Proposition 4.1. Due to Proposition 4.1, there exists a constant $\gamma > 0$ such that $|\Sigma'|/|\Sigma| \geq \gamma$. On the other hand, let $\Sigma'' \subset \Sigma$ be the set of all good $(S_1,\dots,S_k)$, i.e. those that satisfy the conclusion of Claim 5.2. We know that $|\Sigma''|/|\Sigma| = 1 - o(1)$. We immediately get that there exists a tuple $(S_1,\dots,S_k)$ which is simultaneously nice and good. Since this tuple is nice, there exist sets $V_1,\dots,V_k$ which satisfy all the desired requirements. Under these assumptions, we will be able to show deterministically that there exists a perfect matching between V_i and V_j for every $\{i,j\} \in E(T)$ which implies the existence of a T-factor. One way to show the latter implication is, for example, by induction on k. Assume without loss of generality that $k \in V(T)$ is a leaf and that, by induction assumption, we have a T'-factor in $\cup_{i=1}^{k-1} V_i$ where $T' = T \setminus \{k\}$. We may then complete T' to a T-factor via the perfect matching between V_k and V_i where i is the only neighbour of k in T.

Fix $\{i,j\} \in E(T)$. We will show that Hall's condition is satisfied between V_i and V_j in G. Let $W \subseteq V_i$. We will prove that $|N(W,V_j)| \geq |W|$. By Proposition 4.1, for every $v \in V_i$, we have $d(v,V_j) \in \left[\delta \cdot \frac{d}{k}, C \cdot \frac{d}{k}\right]$. Hence,

$$e(W,V_j) \geq |W| \cdot \delta \cdot \frac{d}{k}. \tag{18}$$

We split the proof into three parts depending on the size of $|W|$.

First of all, we show that if $|W| < \eta \cdot \frac{n}{k}$, then $|N(W,V_j)| > |W|$. Assume towards contradiction that this is false. Then, there exists a set $B \subseteq V_j$ satisfying $N(W,V_j) \subseteq B$ and $|B| = |W|$. By Claim 5.1, we have $e(W,B) = e(W,V_j) < |W| \cdot \delta \cdot \frac{d}{k}$, a contradiction to (18).

Next, assume that $\eta \cdot \frac{n}{k} \le |W| \le 0.5(1+\epsilon_1)\frac{n}{k}$. We have

$$|W \cap S_i| \ge |W| - |V_i \setminus S_i| \ge \eta \cdot \frac{n}{k} - |V_i \setminus S_i| \ge \eta^2 \cdot \frac{n}{k},$$

where the last inequality is true since $|V_i \setminus S_i| = o_d(n/k)$ by Proposition 4.1. Thus,

$$\begin{aligned} |N_{V_j}(W)| &\ge |N_{S_j}(W \cap S_i) \cap V_j| \ge (1+\epsilon_2)|W \cap S_i| - |V_j \setminus S_j| \\ &\ge (1+\epsilon_2)(|W| - |V_i \setminus S_i|) - |V_j \setminus S_j| > |W|, \end{aligned}$$

where the second inequality is true by Claim 5.2 and the last inequality is true since $|V_i \setminus S_i|, |V_j \setminus S_j| = o_d(n/k)$ and $|W| = \Omega(n/k)$.

Finally, assume that $0.5(1+\epsilon_1)\frac{n}{k} < |W| \le \frac{n}{k}$. Assume towards contradiction that $|N(W, V_j)| < |W|$. Let $B \subseteq V_j \setminus N(W, V_j)$ be an arbitrary set of size $|V_i \setminus W|$. Notice that $N(B, V_i) \subseteq V_i \setminus W$, otherwise there exists $v \in B$ which is adjacent to $u \in W$. This in turn implies that $v \in N(W, V_j)$ and, in particular, $v \notin B$ — a contradiction. Moreover,

$$|B| = |V_i| - |W| < |V_i| - 0.5(1+\epsilon_1)\frac{n}{k} = 0.5(1-\epsilon_1)\frac{n}{k}.$$

Therefore, by the previous argument (with i and j reversed), $|N(B, V_i)| > |B|$. However, since $N(B, V_i) \subseteq V_i \setminus W$, we have $|N(B, V_i)| \le |V_i \setminus W| = |B|$ — contradiction.

Acknowledgements The authors thank Itai Benjamini for bringing the question to our attention. The authors wish to thank Michael Krivelevich and Noga Alon for fruitful discussions. In particular, we thank Michael Krivelevich for showing us the alternative argument from [10], after the first version of this paper was uploaded. It allowed us to improve the presentation significantly.

References

[1] N. Alon and J. H. Spencer. *The probabilistic method.* Hoboken, NJ: John Wiley & Sons, fourth edition, 2016.

[2] N. Alon and N. Wormald. High degree graphs contain large-star factors. In *Fete of Combinatorics and Computer Science*, pages 9–21. Springer, 2010.

[3] N. Alon and R. Yuster. Threshold functions for H-factors. *Combinatorics, Probability and Computing*, 2(2):137–144, 1993.

[4] H. Assiyatun and N. Wormald. 3-star factors in random d-regular graphs. *European Journal of Combinatorics*, 27(8):1249–1262, 2006.

[5] B. Bollobás. Random graphs. In *Combinatorics (Swansea, 1981)*, volume 52 of *London Math. Soc. Lecture Note Ser.*, pages 80–102. Cambridge Univ. Press, Cambridge-New York, 1981.

[6] B. Bollobás. *Random graphs.* Cambridge University Press, 2nd edition, 2001.

[7] R. L. Burden and J. D. Faires. *Numerical analysis.* Boston, MA: PWS Publishing Company; London: ITP International Thomson Publishing, 5th ed. edition, 1993.

[8] F. Burghart, A. Heckel, M. Kaufmann, N. Müller, and M. Pasch. Sharp thresholds for factors in random graphs. *arXiv preprint, arXiv:2411.14138*, 2024.

[9] L. Chen, R. Kyng, Y. P. Liu, R. Peng, M. P. Gutenberg, and S. Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 612–623. IEEE, 2022.

[10] N. Draganić and M. Krivelevich. Disjoint connected dominating sets in pseudorandom graphs. *Proceedings of the 57th Symposium on Theory of Computing (STOC'25)*, accepted.

[11] P. Erdős and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.*, 17:359–368, 1966.

[12] P. Gao and Y. Ohapkin. Subgraph probability of random graphs with specified degrees and applications to chromatic number and connectivity. *Random Structures & Algorithms*, 62(4):911–934, 2023.

[13] A. Heckel. Random triangles in random graphs. *Random Struct. Algorithms*, 59(4):616–621, 2021.

[14] A. Heckel, M. Kaufmann, N. Müller, and M. Pasch. The hitting time of clique factors. *Random Structures & Algorithms*, 65:275–312, 2024.

[15] J. Hyde, N. Morrison, A. Müyesser, and M. Pavez-Signé. Spanning trees in pseudorandom graphs via sorting networks. *arXiv preprint arXiv:2311.03185*, 2023.

[16] S. Janson. Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.*, 4:369–405, 1995.

[17] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[18] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of a packing conjecture of Bollobás. *Comb. Probab. Comput.*, 4(3):241–255, 1995.

[19] M. Krivelevich. Crowns in pseudo-random graphs and hamilton cycles in their squares. *arXiv preprint arXiv:2305.08442*, 2023.

[20] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006.

[21] T. Łuczak and A. Ruciński. Tree-matchings in graph processes. *SIAM J. Discrete Math.*, 4(1):107–120, 1991.

[22] B. D. McKay. Subgraphs of random graphs with specified degrees. *Congr. Numer.*, 33:213–223, 1981.

[23] R. A. Moser and G. Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):15, 2010. Id/No 11.

[24] R. Otter. The number of trees. *Annals of Mathematics*, 2(49):583–599, 1948.

[25] M. Pavez-Signé. Spanning trees in the square of pseudorandom graphs. *arXiv preprint arXiv:2307.00322*, 2023.

[26] O. Riordan. Random cliques in random graphs and sharp thresholds for F-factors. *Random Struct. Algorithms*, 61(4):619–637, 2022.

[27] R. W. Robinson and N. C. Wormald. Almost all regular graphs are hamiltonian. *Random Structures & Algorithms*, 5(2):363–374, 1994.

[28] A. Ruciński. Matching and covering the vertices of a random graph by copies of a given graph. *Discrete Math.*, 105(1-3):185–197, 1992.

[29] N. C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B*, 31(2):168–182, 1981.

[30] N. C. Wormald. Models of random regular graphs. In *Surveys in Combinatorics, 1999*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999.

A. $K_{1,d-1}$-factor

Let us show that, for $d \geq 5$, $G(n,d)$ **whp** does not contain a $K_{1,d-1}$-factor. Setting $N = nd$, $M = \frac{n}{d} + \frac{n(d-1)^2}{d} = N - 2n + 2n/d$, we have that the expected number of graphs that correspond to $K_{1,d-1}$-factors in the configuration model (see, for example, [6]) is at most

$$\begin{aligned}
\frac{\binom{n}{n/d} d^{n/d} (n\frac{d-1}{d})! d^{n(d-1)/d} M!/(2^{M/2}(M/2)!}{N!/(2^{N/2}(N/2)!)} &= \left(\sqrt{d} + o(1)\right) \frac{n^n d^n (M/N)^{M/2} e^{n-n/d}}{(n/d)^{n/d} e^{n(d-1)/d} (nd)^{n-n/d}} \\
&= \left(\sqrt{d} + o(1)\right) \left(d^{2/d} \left(1 - \frac{2}{d} + \frac{2}{d^2}\right)^{d/2-1+1/d} \right)^n \\
&= o(1).
\end{aligned}$$

Indeed, $g(d) = d^{2/d} \left(1 - \frac{2}{d} + \frac{2}{d^2}\right)^{\frac{d}{2}-1+\frac{1}{d}}$ decreases in d on $[5, \infty)$, and $g(5) < 1$.