

# 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 appearance 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<sup>1</sup> 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  $\llbracket n \rrbracket := \{1, \dots, n\}$  (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**<sup>2</sup> 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.

<sup>1</sup>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  $|V(J)| \geq 2$ .

<sup>2</sup>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 to 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  $\text{poly}(d)$  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  $\{1, \dots, 10d\} \in \llbracket n \rrbracket$ , 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  $(n, d, \lambda)$ -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  $(n, d, O(\sqrt{d}))$ -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  $(n, d, \lambda)$ -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  $(n, d, \lambda)$ -graph contains a copy of every  $n$ -vertex tree with bounded degree and  $\Theta(n)$  leaves. Subsequent work by Hyde, Morrison, Műyesser, and Pavez-Signé [15] showed that for  $\lambda = o(d/\log^3 n)$ , an  $(n, d, \lambda)$ -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 for every positive  $\alpha, \Delta$  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  $\Delta$ .

## 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  $\llbracket n \rrbracket := \{1, 2, \dots, n\}$ . We use the fairly standard notation that given sequences  $a = (a_n)$  and  $b = (b_n) \geq 0$ ,  $a = o(b)$  if, for every  $\epsilon > 0$  there exists  $n_0$  such that  $|a_n| \leq \epsilon b_n$  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  $\epsilon > 0$  there exists  $d_0, n_0$  such that  $|a'_d| \leq \epsilon b'_d$  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  $V(T) := \llbracket k \rrbracket$ . 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  $S_i := \{v \in V(G) : X_v = i\}$ . 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  $\text{Bin}(d, 1/k) < \delta d/k$  is not smaller than  $d^{-1-\epsilon'}$ , for some  $\epsilon' > 0$  tending to zero as  $\epsilon$  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  $\epsilon^{-2}$  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 an 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  $X_i \sim \text{Bernouli}(p_i)$ , 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  $\delta \geq 0$ ,*

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq e^{-\frac{\delta^2 \mathbb{E}[X]}{2 + \delta}}.$$

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

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

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  $t := t(\xi, \delta) > 0$  be sufficiently large. Suppose that  $np \geq (1 + \xi)t$  and  $p \leq \frac{1}{2}$ . Then,*

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

*Proof.* **TOPROVE 0** □

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  $\Delta$  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  $Z_0 := X_0$ . 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  $\tau := t$ . Then,  $\mathbb{E}[\tau] \leq |\mathcal{F}|^{\frac{\beta}{1-\beta}}$ .*

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  $S_1, \dots, S_m$  be a partition of  $U$  satisfying  $S_i = \{u \in U : \xi_u = i\}$  for every  $i \in [m]$ . Let  $\mathcal{F}$  be a finite set of events determined by  $S_1, \dots, S_m$ . 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  $\Delta_1$  random variables  $\xi_u$ , and depends on at most  $\Delta_2$  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  $U_1, \dots, U_m$  which does not satisfy any event in  $\mathcal{F}$  and  $|S_i \triangle U_i| \leq 2\Delta_1 |\mathcal{F}|^{\frac{\beta}{1-\beta}}$  for every  $i \in [m]$ , is at least  $\frac{1}{2}$ .*

*Proof.* **TOPROVE 1** □

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  $d_1 \geq d_2 \geq \dots \geq d_n$ , 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  $1 \leq \ell \leq M/2$  satisfies  $\sum_{i=1}^{d_1} d_i = o(M - 2\ell)$ , then, for every  $S_1, S_2 \subseteq V(G)$ ,*

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

We will also use the following corollary.

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

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

*Proof.* **TOPROVE 2** □

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  $d_1 \geq d_2 \geq \dots \geq d_n$ , let  $G$  be a uniform random graph on vertex set  $\llbracket n \rrbracket$  where vertex  $i$  has degree  $d_i$ . Set  $M := \sum_{i=1}^n d_i$ . For every  $t \in \llbracket n \rrbracket$  and every two disjoint sets  $A$  and  $B$  satisfying*

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

*we have*

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

*Proof.* **TOPROVE 3** □

## 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  $n \rightarrow \infty$ . 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  $S_1, \dots, S_k$  be a uniformly random partition of  $V(G)$ :  $\mathbb{P}(v \in S_i) = \frac{1}{k}$  for every  $i \in \llbracket k \rrbracket$ ,  $v \in V(G)$  and the random choice of  $i \in \llbracket k \rrbracket$  for  $v \in V(G)$  is performed independently of all the other vertices.*

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

(P1)  $|S_i \triangle V_i| = o_d(n/k)$  for every  $i \in \llbracket k \rrbracket$ .

(P2)  $d(v, V_j) \in [\frac{\delta d}{k}, \frac{Cd}{k}]$  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  $\llbracket k \rrbracket$ , 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  $\llbracket h \rrbracket \subseteq V(T) = \llbracket k \rrbracket$  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  $\alpha$ . 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  $\alpha$  which should be considered.

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

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

In relation to Theorem 2, we note that for the proof to follow, it is sufficient to  $n^{-3}$ -approximate  $\alpha$ . 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  $\alpha$  within  $O_\epsilon(\log n)$  steps.

*Proof.* **TOPROVE 4** □

Throughout the rest of the section, we let  $\alpha$  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.

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

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

$$\begin{aligned} B(U_1, \dots, U_h) &:= \{v \in V(G) : \exists i \in [h], d(v, U_i) \leq \delta \log d\}, \\ W(U_1, \dots, U_h) &:= \{v \in V(G) : \exists i \in [h], d(v, U_i) \geq C \log d\}. \end{aligned}$$

Set  $B_1 := B(S_1, \dots, S_h)$  and  $W_1 := W(S_1, \dots, S_h)$ .

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.  $\mathbb{P}(v \in W_1) \leq d^{-100}$ .
2.  $\mathbb{P}(v \in B_1) \leq \frac{\alpha^2}{d}$ .
3.  $\mathbb{P}\left(v \in N(B_1) \setminus \bigcup_{i \in [h]} S_i\right) \leq 2 \left(1 - \frac{h(1-\alpha)}{k}\right) \alpha^2$ .

*Proof.* **TOPROVE 5** □

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  $B_2 = B(A_1^{(1)}, \dots, A_h^{(1)})$  and let  $W_2 = W(A_1^{(1)}, \dots, A_h^{(1)})$ . Then,*

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

*Proof.* **TOPROVE 6** □

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,  $|N(B_2) \setminus \bigcup_{i \in [h]} A_i^{(1)}| \leq 3\alpha^2 n < \frac{\alpha n}{1000}$ ). We will make use of the following lemma.

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

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

*Proof.* **TOPROVE 7** □

Let  $p_1, \dots, p_h$  be the probabilities from the statement of Lemma 4.5. Let us note that  $\sum_{i \in [h]} p_i \leq \frac{1100h}{k} < 1$ , since  $h < d^\beta$  and  $k \geq \frac{d}{10 \log d}$ . 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  $U_i := \{v \in U : X_v = i\}$ , 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.* **TOPROVE 8** □

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  $A_1^{(2)}, \dots, A_h^{(2)} \subseteq V(G)$  which satisfy the following.*

1.  $|S_i \triangle A_i^{(2)}| = o_d(n/k)$  for every  $i \in [h]$ .
2.  $|A_i^{(2)} - \frac{n}{k}| = O(n/d^{50})$  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.* **TOPROVE 9** □

## 4.2. Vertices of low degree in the tree

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

Recall that the sets  $A_1^{(2)}, \dots, A_h^{(2)}$  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

$$U := V(G) \setminus \bigcup_{i \in [h]} A_i^{(2)}. \quad (1)$$

Note that, by Lemma 4.7,

$$\begin{aligned} |U| &\in \left[ n - h \cdot \left( \frac{n}{k} + O\left(\frac{n}{d^{50}}\right) \right), n - h \cdot \left( \frac{n}{k} - O\left(\frac{n}{d^{50}}\right) \right) \right] \\ &= \left[ \left( 1 - \frac{h}{k} - O(d^{-49}) \right) n, \left( 1 - \frac{h}{k} + O(d^{-49}) \right) n \right]. \end{aligned} \quad (2)$$

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  $i \in \{h+1, \dots, k\}$ . All  $X_v$  are independent. For every  $i \in \{h+1, \dots, k\}$ , set

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

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

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

1. There exists  $i \in \llbracket k \rrbracket$  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  $i \in \llbracket k \rrbracket$  such that  $d(v, U_i \cap B(U_1, \dots, U_k)) > \log \log d$ .

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

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

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

*Proof.* TOPROVE 10 □

We also require the following lemma.

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

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

*Proof.* TOPROVE 11 □

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

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

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

*Then,*

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

*Proof.* TOPROVE 12 □

### 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  $i \in [k]$ , 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.  $|A_i^{(4)} \triangle A_i^{(3)}| \leq nd^{-1-\epsilon/5}$  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.* **TOPROVE 13** □

### 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,

$$\Delta_i \leq nd^{-1-\epsilon/6}. \quad (3)$$

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  $i \in \llbracket m \rrbracket$  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  $j \in \{m+1, \dots, k\}$ , or not. In addition, let  $Z_v$  be the random variable over the set  $\{m+1, \dots, k\}$  defined by  $\mathbb{P}(Z_v = j) = \frac{\Delta_j}{\Delta_1 + \dots + \Delta_m}$  for every  $j \in \{m+1, \dots, k\}$ . Note that  $\sum_{j \in \{m+1, \dots, k\}} \mathbb{P}(Z_v = j) = 1$  since  $\Delta_1 + \dots + \Delta_m = \Delta_{m+1} + \dots + \Delta_k$ . The random variable  $Z_v$  represents the index  $j \in \{m+1, \dots, k\}$  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

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

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  $\tilde{A}_1^{(4)}, \dots, \tilde{A}_k^{(4)}$ , noting that the probability measure here is induced by the random variables  $M_v$  and  $Z_v$ .

**Lemma 4.12.** *Whp, the sets  $\tilde{A}_1^{(4)}, \dots, \tilde{A}_k^{(4)}$  satisfy the following for every  $i \in \llbracket k \rrbracket$ .*

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

*Proof.* **TOPROVE 14** □

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  $i \in \llbracket k \rrbracket$ . Further, there is some  $i \in \llbracket k \rrbracket$  such that  $d(v, \tilde{A}_i^{(4)}) \leq \delta \log d/3$ .

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

*Proof.* **TOPROVE 15** □

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  $A_1^{(5)}, \dots, A_k^{(5)}$  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  $i \in \llbracket k \rrbracket$ , we have  $|A_i^{(5)} \triangle \tilde{A}_i^{(4)}| = O\left(\frac{n}{d^{50}}\right)$ .
3. For every  $i \in \llbracket k \rrbracket$ , 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  $j \in \llbracket k \rrbracket$ .

*Proof.* **TOPROVE 16** □

We have just proved that there exist sets  $A_1^{(5)}, \dots, A_k^{(5)}$  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  $A_1^{(5)}, \dots, A_k^{(5)}$  to obtain sets  $V_1, \dots, V_k$ , each with exactly  $\frac{n}{k}$  vertices, while maintaining the degree distribution between the sets.

**Lemma 4.15.** *There exists sets  $V_1, \dots, V_k$  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.  $|V_i \triangle A_i^{(5)}| = o_d(n/k)$  for every  $i \in [k]$ .
3.  $|V_i| = \frac{n}{k}$  for every  $i \in [k]$ .

*Proof.* **TOPROVE 17** □

*Proof.* **TOPROVE 18** □

## 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  $S_1, \dots, S_k$  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  $V_1, \dots, V_k \subseteq V(G)$ , 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.* **TOPROVE 19** □

## 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  $k \leq \frac{(1-\epsilon)d}{\log d}$  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.* **TOPROVE 20** □

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  $S_1, \dots, S_k$  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.* **TOPROVE 21** □

## 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  $\Sigma$  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  $d(v, V_j) \in [\delta \cdot \frac{d}{k}, C \cdot \frac{d}{k}]$ . Hence,

$$e(W, V_j) \geq |W| \cdot \delta \cdot \frac{d}{k}. \quad (4)$$

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 (4).

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

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

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

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

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  $0.5(1+\epsilon_1)\frac{n}{k} < |W| \leq \frac{n}{k}$ . 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  $|N(B, V_i)| \leq |V_i \setminus W| = |B|$  — 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. Oshapkin. 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

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

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  $g(5) < 1$ .