

Sharp threshold for the appearance of certain spanning trees in random graphs

Dan Hefetz *

Michael Krivelevich †

Tibor Szabó ‡

June 30, 2011

Abstract

We prove that a given tree T on n vertices with bounded maximum degree is contained almost surely in the binomial random graph $G(n, \frac{(1+\varepsilon)\log n}{n})$ provided that T belongs to one of the following two classes: (1) T has linearly many leaves; (2) T has a path of linear length all of whose vertices have degree two in T .

1 Introduction

In this paper we consider the problem of embedding a copy of a given tree T on n vertices into the binomial random graph $G(n, p)$. We will restrict our attention to the (already challenging enough) case of trees of bounded maximum degree.

The problem of embedding large or *nearly* spanning bounded degree trees in random graphs on n vertices (where by a nearly spanning tree we mean a tree T whose number of vertices is at most $(1 - c)n$ for some constant $c > 0$) is a rather well studied subject (see, e.g., [9], [1], [11], [10], [12]). In particular, Alon, Sudakov and the second author proved in [2] that for given $\varepsilon > 0$ and integer d there exists $C = C(d, \varepsilon) > 0$ such that a.s.¹ the random graph $G(n, p)$ with $p = C/n$ admits a copy of a tree T on $(1 - \varepsilon)n$ vertices with maximum degree at most d (in fact, it was proved in [2] that such a random graph contains a.s. a copy of *every* such tree). A better bound on the aforementioned constant C and the resilience version of this result have been obtained in [4] and in [5], respectively.

In contrast, apart from some sporadic special cases, not much is known about the case of embedding *spanning* trees. Of course, no spanning tree appears until the random graph becomes connected, which typically happens at $p(n) = \log n/n$ (in fact, $p = \log n/n$ is known to be the

*School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, England. Email: d.hefetz@qmul.ac.uk.

†School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 69978, Israel. Email: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2006322 and by grant 1063/08 from the Israel Science Foundation.

‡Institute of Mathematics, Free University Berlin, 14195 Berlin, Germany. Email: szabo@math.fu-berlin.de.

¹An event \mathcal{E}_n occurs almost surely, or a.s. for brevity, in the probability space $G(n, p)$ if $\lim_{n \rightarrow \infty} Pr[G \sim G(n, p) \in \mathcal{E}_n] = 1$.

connectivity threshold in the following very strong sense: if $p(n) = (\log n - \omega(1))/n$, then a.s. $G(n, p)$ is not connected, whereas if $p(n) = (\log n + \omega(1))/n$ then a.s. $G(n, p)$ is connected; as usual, $\omega(1)$ stands for any function tending to infinity with n arbitrarily slowly). This simple argument provides an immediate lower bound for the edge probability $p(n)$ sufficient for an almost sure appearance of any given spanning tree T in $G(n, p)$. Krivelevich proved in [16] that for any given spanning bounded degree tree T , if $p(n) > n^{-1+\varepsilon}$ for an arbitrarily small constant $\varepsilon > 0$, then $G(n, p)$ contains a.s. a copy of T . (We repeatedly write “a given tree” to stress the order of quantifiers – first a tree is given and only then a random graph $G(n, p)$ is exposed. Our aim is to find a copy of one single tree T . Stronger, universality-type statements, asserting that a.s. a random graph $G(n, p)$ contains simultaneously *every* tree T from a given class, are usually much harder to obtain.) The authors of [2] have observed that if a tree T has at least αn leaves for some constant $\alpha > 0$, then a.s. $G(n, C \log n/n)$ contains a copy of T for some sufficiently large $C = C(\alpha) > 0$. The proof is not that hard and utilizes the embedding result for nearly spanning trees from the same paper. Here is a brief sketch. We represent the random graph $G \sim G(n, p)$ as a union of two independent random graphs G_1 and G_2 , where $G_i \sim G(n, p_i)$ and $1 - p = (1 - p_1)(1 - p_2)$. We set $p_1 = C_1/n$, where C_1 is a sufficiently large constant. Let L denote the set of leaves of T and let $T' = T \setminus L$. By the embedding results for the nearly spanning case, the first random graph G_1 contains a.s. a copy of T' . Now we expose the second random graph G_2 and use its edges to embed the leaves of T and the edges that connect them to their already embedded parents. This can be done using Hall-type arguments provided that $p_2 = C_2 \log n/n$, where $C_2 = C_2(\alpha) > 0$ is a sufficiently large constant. It is instructive to notice that this proof yields the value of $p = C \log n/n$, where the constant $C = C(\alpha)$ has to grow as α becomes smaller (as, in particular, the random graph G_2 has to have an edge connecting every vertex from the set of vertices of diminishing size outside the image of the embedding of T' to one of the vertices slated to serve as a parent of a leaf of T).

While this was not stated explicitly in [2], one can observe that a very similar approach works for another class of bounded degree spanning trees. First, we need to introduce the notion of a bare path.

Definition 1.1 *A path P in a tree T is called bare if all vertices of P have degree exactly two in T .*

Assume now that a given tree T on n vertices admits a bare path P of length at least αn , for some constant $\alpha > 0$. Then one can start by embedding the forest $T' = T \setminus P$ in a random graph $G(n, C_1/n)$, and then use the edges of the random graph $G(n, C_2 \log n/n)$ to find a copy of P between its already embedded endpoints. The latter task amounts to finding a.s. a Hamilton path between a given pair of vertices in a random graph. This can be achieved using known tools (say, those from [13]). Here too the constant $C_2 = C_2(\alpha)$ has to grow as $\alpha \rightarrow 0$.

In this paper, we rectify this problem to some extent by showing that for the two aforementioned classes of bounded degree spanning trees, any given member of the class appears a.s. in $G(n, p)$ already at $p(n) = (1 + \varepsilon) \log n/n$, that is, very shortly after the the binomial random graph first becomes connected. Our main results are manifested in the following two theorems.

Theorem 1.2 *Let α and ε be positive real numbers and let d be a positive integer. Let T be a tree on n vertices, with maximum degree at most d and with at least αn leaves. Then a.s. the*

random graph $G(n, (1 + \varepsilon) \log n/n)$ contains a copy of T .

Remark 1.3 Let T be an arbitrary bounded degree tree on n vertices with $\Theta(n)$ leaves (say, the complete d -ary tree on n vertices for some fixed d). Since $G(n, p)$ is a.s. disconnected for every $p \leq (1 - \varepsilon) \log n/n$, it follows from Theorem 1.2 that there is a sharp threshold at $\log n/n$ for the appearance of T in $G(n, p)$.

Let G be a graph and let \mathcal{F} be a family of graphs. The graph G is said to be *universal for the family \mathcal{F}* , or \mathcal{F} -universal for brevity, if G contains every $F \in \mathcal{F}$ as a subgraph.

Theorem 1.4 Let α and ε be positive real numbers and let d be a positive integer. Let \mathcal{L} be the family of all trees on n vertices with maximum degree at most d which admit a bare path of length αn . Then the random graph $G(n, (1 + \varepsilon) \log n/n)$ is a.s. \mathcal{L} -universal.

The rest of this paper is organized as follows. In the next section we gather several tools needed for the subsequent proofs of our main results. In Section 3 we prove Theorems 1.2 and 1.4. The final section of the paper is devoted to concluding remarks.

For the sake of simplicity and clarity of presentation, we do not make a particular effort to optimize the constants obtained in our proofs. We also omit floor and ceiling signs whenever these are not crucial. Most of our results are asymptotic in nature and whenever necessary we assume that the number of vertices n is sufficiently large. Moreover, when proving that a certain graph property holds a.s. we will sometimes refrain from explicitly denoting the number of vertices of the graph, as it is in fact a sequence of integers which tends to infinity. Throughout the paper, \log stands for the natural logarithm, unless stated otherwise. Our graph-theoretic notation is standard and follows that of [18]. In particular, we use the following.

For a graph G , let $V(G)$ and $E(G)$ denote its sets of vertices and edges respectively, and let $v(G) = |V(G)|$ and $e(G) = |E(G)|$. For disjoint sets $A, B \subseteq V(G)$, let $E_G(A, B)$ denote the set of edges of G with one endpoint in A and one endpoint in B , and let $e_G(A, B) = |E_G(A, B)|$. For a set $U \subseteq V(G)$ and a vertex $w \in V(G)$, let $N_G(w, U) = \{u \in U : wu \in E(G)\}$ denote the set of neighbors of w in U and let $d_G(w, U) = |N_G(w, U)|$ denote the degree of w into U . For sets $U, W \subseteq V(G)$ let $N_G(W, U) = \bigcup_{w \in W} N_G(w, U)$. We abbreviate $N_G(w, V(G))$ to $N_G(w)$, and let $d_G(w) = |N_G(w)|$ denote the degree of w in G . We also abbreviate $N_G(W, V(G))$ to $N_G(W)$. The minimum degree and the maximum degree of a graph G are denoted by $\delta(G)$ and $\Delta(G)$ respectively. Often, when there is no risk of confusion, we omit the subscript G from the notation above. For a set $S \subseteq V(G)$, let $G[S]$ denote the subgraph of G , induced on the vertices of S . A graph G is said to be *Hamilton connected* if, for every two vertices $u, w \in V(G)$, there is a Hamilton path in G whose endpoints are u and w .

2 Preliminaries

In this section we prove several results which will be useful in proving our main theorems. Since there are quite a few such results and they vary in nature, this section is divided into several subsections.

2.1 Properties of $G(n, p)$

In this subsection we prove some simple properties of the binomial random graph $G(n, p)$. In our proofs we will make use of certain known bounds on the tail of the binomial distribution (see e.g. [15]). In particular we will use the following bound.

Lemma 2.1 [6, Lemma 2.1] *If $X \sim \text{Bin}(n, p)$ and $k \geq np$ then $\Pr(X \geq k) \leq (enp/k)^k$.*

The properties of interest are described in the following three lemmas.

Lemma 2.2 *Let $0 < \varepsilon < 1$ and $0 \leq \beta \leq \varepsilon/7$ be real numbers and let $p = p(n) = (1 + \varepsilon) \frac{\log n}{n}$. Let U be a given set of size $|U| \leq \beta n$. Then a.s. the random graph $G = G(n, p) = ([n], E)$ satisfies all of the following properties:*

(P1) $\Delta(G) \leq 10 \log n$.

(P2) $d_G(u, [n] \setminus U) \geq \eta \log n$ for every $u \in [n]$, where $0 < \eta = \eta(\varepsilon) < 1/2$ is a real number satisfying $\eta \log(6\eta^{-1}) = \varepsilon/3$; in particular $\delta(G) \geq \eta \log n$.

(P3) Every subset $A \subseteq [n]$ of cardinality $|A| \leq \frac{n(\log \log n)^2}{\log n}$ spans at most $\frac{|A| \log n}{\log \log n}$ edges in G .

(P4) For every two disjoint subsets $A, B \subseteq [n]$ of cardinality $|A| \leq \frac{n(\log \log n)^2}{\log n}$ and $|B| = |A| \sqrt{\log n}$, we have $e_G(A, B) \leq \frac{|A| \log n}{\log \log n}$.

(P5) For every two disjoint subsets A, B of $[n]$ of cardinality $|A| = |B| = \frac{n(\log \log n)^{3/2}}{\log n}$, we have $e_G(A, B) > 0$.

Proof Properties **(P1)** - **(P5)** follow by standard first moment calculations and standard bounds on the tail of the binomial distribution.

(P1): For a given vertex $v \in [n]$, the degree of v is distributed binomially with parameters $n - 1$ and p . Therefore, it follows by Lemma 2.1 that

$$\Pr[d_G(v) \geq 10 \log n] \leq \left(\frac{enp}{10 \log n} \right)^{10 \log n} \leq (2e/10)^{10 \log n} = o(1/n).$$

Applying the union bound over all vertices of $[n]$ proves that the required bound on the maximum degree holds a.s.

(P2): For a given vertex $u \in [n]$, the degree of u in $[n] \setminus U$ is distributed binomially with parameters $n - \beta n$ and p if $u \in U$ and with parameters $n - \beta n - 1$ and p if $u \in [n] \setminus U$.

Denote $\delta = \eta \log n$ and fix some vertex $u \in [n]$. Then

$$\begin{aligned}
Pr[d_G(u, [n] \setminus U) \leq \eta \log n] &\leq \sum_{i=0}^{\delta} Pr[Bin(n - \beta n - 1, p) = i] \\
&\leq (\delta + 1) Pr[Bin((1 - \beta)n - 1, p) = \delta] \\
&\leq \log n \binom{(1 - \beta)n}{\delta} p^\delta (1 - p)^{(1 - \beta)n - 1 - \delta} \\
&\leq \log n \left(\frac{enp}{\delta(1 - p)} \right)^\delta (1 - p)^{(1 - \beta)n - 1} \\
&\leq \left(\frac{6}{\eta} \right)^{\eta \log n} e^{-p(1 - \beta)n} \\
&= \exp \{ [\eta \log(6/\eta) - (1 + \varepsilon)(1 - \beta)] \log n \} \\
&\leq \frac{1}{n^{1 + \varepsilon/3}},
\end{aligned}$$

where the first inequality above follows by the monotonicity of the binomial distribution along its lower tail.

Applying the union bound over all vertices of $[n]$ proves that the required bound on the minimum degree holds a.s.

(P3): Let $A \subseteq [n]$ be any subset of size $1 \leq a \leq \frac{n(\log \log n)^2}{\log n}$. Let X_A be the random variable that counts the number of edges of G with both endpoints in A . Then $X_A \sim Bin\left(\binom{a}{2}, p\right)$ and thus $\mathbb{E}(X_A) = \binom{a}{2}p$. Let E_3 denote the event “there exists a set $A \subseteq [n]$, of size $1 \leq a \leq \frac{n(\log \log n)^2}{\log n}$, such that $e_G(A) > \frac{a \log n}{\log \log n}$ ”. Using Lemma 2.1 we get

$$\begin{aligned}
Pr[E_3] &\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \binom{n}{a} Pr \left[X_A \geq \frac{a \log n}{\log \log n} \right] \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\frac{en}{a} \left(\frac{e \binom{a}{2} p}{a \log n / (\log \log n)} \right)^{\log n / \log \log n} \right]^a \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\frac{en}{a} \left(\frac{3a \log \log n}{n} \right)^{\log n / \log \log n} \right]^a \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\exp \left\{ 1 + \log(n/a) - \frac{\log n}{\log \log n} (\log(n/a) - 2 \log \log \log n) \right\} \right]^a \\
&= o(1).
\end{aligned}$$

(P4): Let $A \subseteq [n]$ be any subset of cardinality $1 \leq a \leq \frac{n(\log \log n)^2}{\log n}$ and let B be any subset of $[n] \setminus A$ of cardinality $b = a\sqrt{\log n}$. Let X_{AB} be the random variable that counts the

number of edges of G with one endpoint in A and the other in B . Then $X_{AB} \sim \text{Bin}(ab, p)$ and thus $\mathbb{E}(X_{AB}) = abp = a^2 p \sqrt{\log n}$. Let E_4 denote the event “there exist two disjoint subsets $A, B \subseteq [n]$, of sizes $1 \leq a = |A| \leq \frac{n(\log \log n)^2}{\log n}$ and $b = |B| = a\sqrt{\log n}$, such that $e_G(A, B) > \frac{a \log n}{\log \log n}$ ”. Using Lemma 2.1 we get

$$\begin{aligned}
Pr[E_4] &\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \binom{n}{a} \binom{n}{b} Pr \left[X_{AB} \geq \frac{a \log n}{\log \log n} \right] \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\frac{en}{a} \left(\frac{en}{b} \right)^{\sqrt{\log n}} \left(\frac{ea^2 p \sqrt{\log n}}{a \log n / (\log \log n)} \right)^{\log n / \log \log n} \right]^a \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\frac{en}{a} \left(\frac{en}{b} \right)^{\sqrt{\log n}} \left(\frac{6a\sqrt{\log n} \log \log n}{n} \right)^{\log n / \log \log n} \right]^a \\
&\leq \sum_{a=1}^{\frac{n(\log \log n)^2}{\log n}} \left[\exp\{1 + \log(n/a) + \sqrt{\log n} (1 + \log(n/b))\} \right. \\
&\quad \left. - \frac{\log n}{\log \log n} (\log(n/a) - 0.6 \log \log n) \right]^a \\
&= o(1).
\end{aligned}$$

(P5): Let E_5 denote the event: “there exist two disjoint subsets $A, B \subseteq [n]$ of size $|A| = |B| = \frac{n(\log \log n)^{3/2}}{\log n}$ such that $e_G(A, B) = 0$ ”. Then

$$\begin{aligned}
Pr[E_5] &\leq \binom{n}{|A|} \binom{n}{|B|} (1-p)^{|A||B|} \\
&\leq \left(\frac{n}{\frac{n(\log \log n)^{3/2}}{\log n}} \right)^2 e^{-p|A||B|} \\
&\leq \left(\frac{e \log n}{(\log \log n)^{3/2}} \right)^{\frac{2n(\log \log n)^{3/2}}{\log n}} \exp \left\{ -\frac{(1+\varepsilon) \log n}{n} \cdot \frac{n^2 (\log \log n)^3}{(\log n)^2} \right\} \\
&\leq \exp \left\{ \frac{2n(\log \log n)^{5/2}}{\log n} - \frac{n(\log \log n)^3}{\log n} \right\} \\
&= o(1).
\end{aligned}$$

□

Lemma 2.3 Let $0 < \beta_1 < \beta_2 \leq 1$ be real numbers and let $p = p(n) = (1 - \beta_1) \frac{\log n}{n}$. Let $W \subseteq [n]$ be a given subset of size $|W| \leq n^{1-\beta_2}$ and let $r \in [n] \setminus W$ be a given vertex. Then a.s. the random graph $G = G(n, p) = ([n], E)$ satisfies all of the following properties:

(Q1) $\Delta(G) \leq 10 \log n$.

(Q2) Let $0 < \gamma < 1/2$ be a real number satisfying $\gamma \log(3/\gamma) = (\beta_2 - \beta_1)/3$, then $d_G(w, [n] \setminus W) \geq \gamma \log n$ for every $w \in W$.

(Q3) $|\{u \in N_G(w) \setminus W : N_G(u) \cap (W \setminus \{w\}) \neq \emptyset\}| \leq 2/\beta_2$ for every $w \in W$.

Proof Properties (Q1) - (Q3) follow by standard first moment calculations and standard bounds on the tail of the binomial distribution.

(Q1): This follows immediately from Property P1 of Lemma 2.2.

(Q2): It suffices to prove this under the assumption $|W| = n^{1-\beta_2}$. For a given vertex $w \in W$, the degree of w in $[n] \setminus W$ is distributed binomially with parameters $n - n^{1-\beta_2}$ and p .

Denote $\delta = \gamma \log n$ and fix some vertex $w \in W$. Then

$$\begin{aligned}
Pr[d_G(w, [n] \setminus W) \leq \gamma \log n] &= \sum_{i=0}^{\delta} Pr[Bin(n - n^{1-\beta_2}, p) = i] \\
&\leq (\delta + 1) Pr[Bin(n - n^{1-\beta_2}, p) = \delta] \\
&\leq \log n \binom{n - n^{1-\beta_2}}{\delta} p^\delta (1-p)^{n - n^{1-\beta_2} - \delta} \\
&\leq \log n \left(\frac{enp}{\delta(1-p)} \right)^\delta (1-p)^{n - n^{1-\beta_2}} \\
&\leq \left(\frac{3}{\gamma} \right)^{\gamma \log n} e^{-(1-o(1))pn} \\
&= \exp \{ [\gamma \log(3/\gamma) - (1 - \beta_1 - o(1))] \log n \} \\
&\leq \frac{1}{n^{1-(\beta_1+\beta_2)/2}},
\end{aligned}$$

where the first inequality above follows by the monotonicity of the binomial distribution along its lower tail.

Since $\beta_1 < \beta_2$, applying the union bound over all vertices of W proves that the required lower bound on the degree a.s. holds for every $w \in W$.

(Q3): Fix some $w \in W$ and let $u \in [n] \setminus W$ be an arbitrary vertex. It follows that

$$\begin{aligned}
Pr[uw \in E(G), N_G(u) \cap (W \setminus \{w\}) \neq \emptyset] &< p \cdot |W|p \\
&\leq n^{-1-\beta_2} \log^2 n \\
&< n^{-1-\beta_2/2}.
\end{aligned}$$

Hence, the probability that there are at least $2/\beta_2$ such vertices u is at most

$$\binom{n}{2/\beta_2} \left(n^{-1-\beta_2/2} \right)^{2/\beta_2} \leq 1/n.$$

Applying the union bound over all vertices $w \in W$ yields the desired result. \square

2.2 Splitting trees and random graphs

In order to embed a spanning tree T in $G(n, p)$, we will want to split both T and $G(n, p)$ into several parts. Starting with the former, we prove the following.

Lemma 2.4 *Let α and ε be positive real numbers and let d be a positive integer. Let $n_0 = n_0(\alpha, \varepsilon)$ be a sufficiently large integer and let $n \geq n_0$ be an integer. Let T be a tree on n vertices, with maximum degree at most d and with at least αn leaves. Then there exists a real number $\beta = \beta(\varepsilon, \alpha, d) > 0$, a vertex $u \in V(T)$ and subtrees T_1 and T_2 of T such that $V(T_1) \cup V(T_2) = V(T)$, $V(T_1) \cap V(T_2) = \{u\}$, $|V(T_1)| \leq \varepsilon n$ and the number of leaves of T_1 is at least βn .*

Proof Let $k = k(\varepsilon)$ be the smallest positive integer for which $2^{-k} \leq \varepsilon$ holds. It is well known (and easy) that any tree T has a $(1, 1/2)$ -separator, that is, a vertex $u \in V(T)$ such that every connected component of $T \setminus \{u\}$ has at most $|V(T)|/2$ vertices. Let u be such a vertex and let v_1, \dots, v_r be the neighbors of u in T ; note that $r \leq d$. For every $1 \leq i \leq r$ let T^i denote the tree of $T \setminus \{u\}$ which contains v_i (we consider v_i to be the root of T^i). It is clear that there exists a $1 \leq j \leq r$ such that the number of leaves of T^j is at least $\alpha n/d$. Moreover, $T \setminus T^j$ is a tree and $|V(T^j)| \leq |V(T)|/2$ holds by the choice of u . Repeating this process k times (each time with an appropriate tree T^j) we obtain a tree T_1 with at most $2^{-k}n \leq \varepsilon n$ vertices and with at least $\alpha n/d^k$ leaves. Moreover, adding the root of T_1 to $T \setminus T_1$ yields the desired tree T_2 . \square

The following lemma handles splitting $G(n, p)$.

Lemma 2.5 (*Clustered Local Lemma*) *Let $G = (V, E)$ be a graph on n vertices with maximum degree Δ . Let $Y \subseteq V$ be a set of $m = a + b$ vertices where a and b are positive integers. Assume that $d_G(v, Y) \geq \delta$ holds for every $v \in V$. If $\Delta^2 \cdot \left\lceil \frac{m}{\min\{a, b\}} \right\rceil \cdot 2 \cdot e^{1 - \frac{\min\{a, b\}^2}{5m^2} \cdot \delta} < 1$, then there exists a partition $Y = A \cup B$ of Y such that*

- (i) $|A| = a$ and $|B| = b$.
- (ii) $d_G(v, A) \geq \frac{a}{3m} d_G(v, Y)$ for every $v \in V$.
- (iii) $d_G(v, B) \geq \frac{b}{3m} d_G(v, Y)$ for every $v \in V$.

In the proof of Lemma 2.5 we will make use of the following well known results.

Lemma 2.6 (**Lovász Local Lemma** (see e.g. [3])) *Let A_1, A_2, \dots, A_n be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most d , and that $\Pr(A_i) \leq p$ for all $1 \leq i \leq n$. If $ep(d+1) \leq 1$, then $\Pr(\bigwedge_{i=1}^n \bar{A}_i) > 0$.*

Lemma 2.7 (**Hoeffding's inequality** [14]) *Let X_1, \dots, X_n be independent random variables and let $S = \sum_{i=1}^n X_i$. Let $a_1, \dots, a_n, b_1, \dots, b_n$ be real numbers such that $a_i \leq X_i \leq b_i$ holds for every $1 \leq i \leq n$. Then for every $t > 0$*

$$\Pr(|S - \mathbb{E}(S)| \geq t) \leq 2 \exp \left\{ - \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

Proof of Lemma 2.5 Assume without loss of generality that $a \leq b$. Let $Y = Q_1 \cup \dots \cup Q_a$ be an arbitrary partition of Y into a parts of nearly equal size (that is, $|Q_i| = \lfloor \frac{m}{a} \rfloor$ or $|Q_i| = \lceil \frac{m}{a} \rceil$ for every $1 \leq i \leq a$, in particular no Q_i is empty). We construct the set A by selecting one vertex from every Q_i , independently and uniformly at random. The size of A is then clearly precisely a and the size of $B := Y \setminus A$ is precisely b . Hence Property (i) is satisfied. For every $v \in V$, let E_v denote the *bad event* “ v violates Property (ii) or Property (iii)”, that is, $d_G(v, A) < \frac{a}{3m} d_G(v, Y)$ or $d_G(v, B) < \frac{b}{3m} d_G(v, Y)$.

In order to bound from above the probability of a bad event E_v we consider the random variables $d_G(v, A)$ and $d_G(v, B)$. Let $\Gamma_v := \{1 \leq i \leq a : N_G(v) \cap Q_i \neq \emptyset\}$; note that $|\Gamma_v| \leq d_G(v, Y)$. For every $i \in \Gamma_v$ let X_i be the indicator random variable for the event “ $N_G(v, A) \cap Q_i \neq \emptyset$ ”. Then $\Pr[X_i = 1] = \frac{|N_G(v) \cap Q_i|}{|Q_i|}$ for every $i \in \Gamma_v$ and $d_G(v, A) = \sum_{i=1}^{|\Gamma_v|} X_i$. It follows that $\mathbb{E}(d_G(v, A)) = \sum_{i=1}^{|\Gamma_v|} \frac{|N_G(v) \cap Q_i|}{|Q_i|}$. Hence

$$\mathbb{E}(d_G(v, A)) \geq \sum_{i=1}^{|\Gamma_v|} \frac{|N_G(v) \cap Q_i|}{\lceil m/a \rceil} = \frac{d_G(v, Y)}{\lceil m/a \rceil} \geq \frac{a}{m+a} d_G(v, Y),$$

and similarly

$$\mathbb{E}(d_G(v, A)) \leq \sum_{i=1}^{|\Gamma_v|} \frac{|N_G(v) \cap Q_i|}{\lfloor m/a \rfloor} = \frac{d_G(v, Y)}{\lfloor m/a \rfloor} \leq d_G(v, Y)/2,$$

where the last inequality follows since $a \leq m/2$.

Clearly $d_G(v, B) = d_G(v, Y) - d_G(v, A)$ and thus $\mathbb{E}(d_G(v, B)) = d_G(v, Y) - \mathbb{E}(d_G(v, A))$. Hence, applying Lemma 2.7 with $n = |\Gamma_v|$ and with $a_i = 0$ and $b_i = 1$ for every $i \in \Gamma_v$ we obtain

$$\begin{aligned} \Pr(E_v) &\leq \Pr\left(d_G(v, A) < \frac{a}{3m} d_G(v, Y)\right) + \Pr\left(d_G(v, B) < \frac{b}{3m} d_G(v, Y)\right) \\ &\leq \Pr\left(d_G(v, A) < \frac{a}{3m} d_G(v, Y)\right) + \Pr\left(d_G(v, A) > \frac{2m+a}{3m} d_G(v, Y)\right) \\ &\leq \Pr\left(d_G(v, A) - \mathbb{E}(d_G(v, A)) < \left(\frac{a}{3m} - \frac{a}{m+a}\right) d_G(v, Y)\right) \\ &\quad + \Pr\left(d_G(v, A) - \mathbb{E}(d_G(v, A)) > \left(\frac{2}{3} + \frac{a}{3m} - \frac{1}{2}\right) d_G(v, Y)\right) \\ &\leq \Pr\left(|d_G(v, A) - \mathbb{E}(d_G(v, A))| > \frac{a}{3m} d_G(v, Y)\right) \\ &\leq 2 \exp\left\{-\frac{2a^2 d_G(v, Y)^2}{9m^2 |\Gamma_v|}\right\} \\ &\leq 2 \exp\left\{-\frac{a^2 \delta}{5m^2}\right\}, \end{aligned}$$

where the fourth inequality follows since $a \leq m/2$.

Next, we bound from above the maximum degree of the dependency graph of bad events. Let $v, w \in V$ be two distinct vertices. It is clear that if $\Gamma_v \cap \Gamma_w = \emptyset$, then E_v and E_w are independent

events. As previously noted $|\Gamma_v| \leq d_G(v, Y) \leq \Delta$. Since $|Q_i| \leq \lceil \frac{m}{a} \rceil$ for every $i \in \Gamma_v$, it follows that E_v is independent of all but at most $\sum_{i \in \Gamma_v} \sum_{u \in Q_i} (d_G(u) - 1) \leq \Delta^2 \lceil \frac{m}{a} \rceil - 1$ of the events $E_u : u \in V \setminus \{v\}$.

The existence of the required partition $Y = A \cup B$ thus follows by Lemma 2.6. \square

2.3 Embedding almost spanning trees in random and pseudo-random graphs

As noted in the previous section, in order to embed a spanning tree T in $G(n, p)$, we will first want to embed a large subtree of T in a certain subgraph of $G(n, p)$. Moreover, we will want this embedding to cover certain “problematic” vertices of $G(n, p)$. We prove the following embedding statement which might be of independent interest.

Proposition 2.8 *Let β_1 and $0 < \beta_2 \leq 1$ be real numbers such that $\beta_2 > 2\beta_1$. Let $0 < a \leq b < 1$ be real numbers and let d be a positive integer. Let $T = (V, E)$ be a rooted tree with maximum degree d , where $an \leq |V| \leq bn$. Let $r' \in V$ be the root of T , let $W \subseteq [n]$ be a given subset of size $|W| \leq n^{1-\beta_2}$ and let $r \in [n] \setminus W$. Then a.s. there exists an embedding $\phi : V \rightarrow [n]$ of T in the random graph $G(n, p)$ with $p = \frac{(1-\beta_1) \log n}{n}$ such that $\phi(r') = r$ and $W \subseteq \phi(V)$.*

Before proving Proposition 2.8 we introduce some terminology that will be used in the course of our proof. Let G be a graph, let T be a tree, and let $S \subseteq V(T)$ be an arbitrary set. An S -partial embedding of T in G is an injective mapping $f : S \rightarrow V(G)$, such that $f(x)f(y) \in E(G)$ whenever $\{x, y\} \subseteq S$ and $xy \in E(T)$. For every vertex $v \in f(S)$ we denote $v' = f^{-1}(v)$. If $S = V(T)$, we call an S -partial embedding of T in G simply an embedding of T in G . We say that the vertices of S are *embedded*, whereas the vertices of $V(T) \setminus S$ are called *new*. An embedded vertex is called *closed* if all its neighbors in T are embedded as well. An embedded vertex that is not closed, is called *open*. The vertices of $f(S)$ are called *taken*, whereas the vertices of $V(G) \setminus f(S)$ are called *available*. With some abuse of this terminology, for a closed (respectively open) vertex $u \in S$, we will sometimes refer to $f(u)$ as being closed (respectively open) as well.

Proof of Proposition 2.8 Since the existence of the required mapping ϕ is a monotone increasing property, we can assume without loss of generality that $\beta_1 > 0$ (while $\beta_2 > 2\beta_1$ still holds).

We expose G in two rounds, that is, we split $G = G_1 \cup G_2$, where $G_1 \sim G(n, p_1)$ with $p_1 = (1 - 2\beta_1) \log n / n$ and $G_2 \sim G(n, p_2)$ with $(1 - p_2)(1 - p_1) = 1 - p$. It follows that $p_2 \geq \beta_1 \log n / n$.

We first expose the edges of G_1 with one endpoint in W and the other in $[n] \setminus (W \cup \{r\})$. For every $w \in W$ let $Z_w := N_{G_1}(w, [n] \setminus (W \cup \{r\}))$. Let $0 < \gamma < 1/4$ be a real number satisfying $2\gamma \log(3/(2\gamma)) = (\beta_2 - 2\beta_1)/3$. It follows by Properties **(Q1)** and **(Q2)** from Lemma 2.3 that a.s. $2\gamma \log n \leq |Z_w| \leq 10 \log n$ holds for every $w \in W$. For every $w \in W$ let $X_w := \{u \in Z_w : N_{G_1}(u) \cap (W \setminus \{w\}) = \emptyset\}$. It follows by Property **(Q3)** from Lemma 2.3 that a.s. $|X_w| \geq 2\gamma \log n - 2/\beta_2 \geq \gamma \log n$. Denote $X = \bigcup_{w \in W} X_w$, then a.s. $|X| \leq 10 \log n \cdot n^{1-\beta_2}$. Next, we expose all edges of $H_1 := G_1[[n] \setminus (W \cup X \cup \{r\})]$. Let H denote the graph obtained from H_1 by repeatedly deleting all vertices of degree less than $(1 - 2\beta_1)(1 - b) \log n / 2$; clearly

$\delta(H) \geq (1 - 2\beta_1)(1 - b) \log n/2$. We claim that a.s. $|V(H_1) \setminus V(H)| \leq \frac{1-b}{4} \cdot n$. Indeed, otherwise there exists a set $B \subseteq V(H_1) \setminus V(H)$ of size $|B| = \frac{1-b}{4} \cdot n$ such that there are at most $\frac{1-b}{4} \cdot n \cdot (1 - 2\beta_1)(1 - b) \log n/2$ edges of G_1 with one endpoint in B and the other in $V(H_1) \setminus B$. The expected number of such edges in G_1 is $|B||V(H_1) \setminus B|p_1 = \frac{1-b}{4} (1 - 2\beta_1) \left(1 - \frac{1-b}{4} - o(1)\right) n \log n$. It follows by standard bounds on the tail of the binomial distribution that the probability that such a set B exists is at most

$$\binom{|V(H_1)|}{|B|} \exp \left\{ -c \cdot \frac{1-b}{4} (1 - 2\beta_1) \left(1 - \frac{1-b}{4} - o(1)\right) n \log n \right\} = o(1),$$

where $c = c(b, \beta_1) > 0$ is an appropriately chosen constant.

Next, we expose the edges of G_1 with one endpoint in X and the other in $V(H)$. It follows by Property **(Q2)** from Lemma 2.3 that a.s. there exists a real number $\gamma' > 0$ such that $d_{G_1}(u, V(H)) \geq \gamma' \log n$ holds for every $u \in X$ (indeed, for $N := |V(H) \cup X| \geq 3n/4$ there exists a constant $\beta_3 > 2\beta_1$ such that $|X| \leq N^{1-\beta_3}$; hence we can apply Lemma 2.3 with N , X and β_3 instead of n , W and β_2).

We conclude that a.s. there exist positive real numbers γ and γ' , a family $\{X_w : w \in W\}$ of pairwise disjoint subsets of $[n] \setminus (W \cup \{r\})$, and a subgraph $H \subseteq G_1[[n] \setminus (W \cup X \cup \{r\})]$ which satisfy the following properties:

- (i) $wu \in E(G_1)$ for every $w \in W$ and every $u \in X_w$.
- (ii) $\gamma \log n \leq |X_w| \leq 10 \log n$ for every $w \in W$.
- (iii) $|V(H)| \geq \left(1 - \frac{1-b}{4} - o(1)\right) n$.
- (iv) $\delta(H) \geq (1 - 2\beta_1)(1 - b) \log n/2$.
- (v) $d_{G_1}(u, V(H)) \geq \gamma' \log n$ for every $u \in X$.

Let $M \subseteq V(H)$ be a set of size $\frac{1-b}{4} \cdot n$ and let $\gamma'' = \gamma''(\gamma', b, \beta_1) > 0$ be a real number such that $d_{G_1}(v, M) \geq \gamma'' \log n$ holds for every $v \in V(H) \cup X$. Such M and γ'' exist by Lemma 2.5 (with $V = V(H) \cup X$ and $Y = V(H)$).

Finally, we expose the edges of G_1 with one endpoint in $\{r\}$ and the other in M . Standard bounds on the tail of the binomial distribution show that a.s. $d_{G_1}(r, M) \geq \gamma''' \log n$ holds for some positive real number $\gamma''' = \gamma'''(\beta_1, b)$. Let $\zeta = \min\{(1 - 2\beta_1)(1 - b)/2, \gamma, \gamma', \gamma'', \gamma'''\}$.

Next, we try to embed T in $G_2[V(H) \cup W \cup X \cup \{r\}]$. We embed T vertex by vertex starting with r' . At each point during the embedding process the part of T which was already embedded will be a subtree T' of T rooted at r . We will give priority to including the vertices of W in the embedding: whenever there will be a possibility to embed a path of length two from an open vertex of the current embedding to a not yet embedded vertex $w \in W$, we will do so. We will then use the available neighbors of w in X_w to embed the children of the preimage of w in T ; this will close w . The edges of G_2 are exposed during the embedding process, as will be described shortly. Upon encountering certain difficulties, we will make use of the edges of G_1 . At any point during the embedding of T we denote the set of embedded vertices by S and the current S -partial embedding by ϕ . Moreover, we denote $I = W \setminus \phi(S)$. Initially $S = \{r'\}$ and $\phi(r') = r$.

As long as $V \setminus S \neq \emptyset$ we proceed as follows. Let $v' \in S$ be an arbitrary open vertex and let $v = \phi(v')$. Assume first that there exist vertices $x', y' \in V \setminus S$ such that $v'x', x'y' \in E$ (that is, v' has a non-leaf neighbor which has not been embedded yet). Expose all edges of G_2 with one endpoint in $\{v\}$ and the other in X . Assume further that there exists some $w \in I$ and some $u \in X_w$ such that $vu \in E(G_2)$. Choose such w and u arbitrarily, add x' and y' to S , and update ϕ by setting $\phi(x') = u$ and $\phi(y') = w$. If y' is open, embed all of its children in T into arbitrary vertices of $X_w \setminus \{u\}$. This is possible since $|X_w \setminus \{u\}| > d \geq d_T(y')$ holds by Property (ii) above (we will not use the vertices of X_w except when attempting to embed and subsequently close w). Update S and ϕ accordingly; note that w is now closed. Assume then that this is not the case, that is, every non-leaf neighbor of v' has already been embedded or there are no edges in G_2 with one endpoint in $\{v\}$ and the other in $\bigcup_{w \in I} X_w$. Expose all edges of G_2 with one endpoint in $\{v\}$ and the other in $V(H) \setminus (M \cup \phi(S))$. Let $\{v'_1, \dots, v'_t\} = N_T(v') \setminus S$ be the new neighbors of v' in T . If $d_{G_2}(v, V(H) \setminus (M \cup \phi(S))) \geq t$, then, for every $1 \leq i \leq t$, we set $\phi(v'_i) = v_i$ for an arbitrary vertex $v_i \in N_{G_2}(v, V(H) \setminus (M \cup \phi(S)))$ (using distinct vertices to embed distinct v'_i) and update S accordingly. Note that v is now closed. Otherwise we declare an *emergency*. During this emergency we try to embed v'_1, \dots, v'_t into $N_{G_1}(v, M \setminus \phi(S))$. If this attempt is successful, that is, if $d_{G_1}(v, M \setminus \phi(S)) \geq t$, then v is closed and we update S and ϕ accordingly. Otherwise we declare a *failure*.

In order to complete the proof of the proposition it suffices to prove that a.s. there are no failures and that a.s. $W \subseteq \phi(V)$. Starting with the former assume there is a failure at v . It follows that $d_{G_1}(v, M \setminus \phi(S)) < t \leq d$. Denote $k := d_{G_1}(v, M)$; recall that $k \geq \zeta \log n$ holds by the choice of M and since $v \notin W$ as it is open. Denote $A_v := N_{G_1}(v, M)$ and $B_v := N_{G_1}(A_v, V(H) \cup X \cup \{r\})$. It follows by Property **(Q1)** of Lemma 2.3 that a.s. $|B_v| \leq 100 \log^2 n$. Since we use the vertices of M only during emergencies, it follows that we already treated at least $(k-d)/d$ emergencies at vertices of B_v . Let $w \in B_v$ be an arbitrary vertex at which we have treated an emergency. Observe that when the state of emergency was declared at w there were $m \geq |V(H)| - |\phi(S)| - |M| - |W| \geq \frac{1-b}{3} \cdot n$ vertices available to embed the new neighbors of the preimage of w . Since the degree in G_2 of w into this set was distributed as $Bin(m, p_2)$, it follows by standard bounds on the tail of the binomial distribution that the probability of declaring an emergency at w is at most

$$Pr[Bin(m, p_2) < d] \leq e^{-\frac{1}{3} \cdot \frac{1-b}{3} \beta_1 \log n} \leq n^{-c},$$

where $c = c(b, \beta_1) > 0$ is an appropriately chosen real number. Note that this bound is independent of the occurrence of any other emergency.

Hence, the probability that there are at least $(k-d)/d$ emergencies at vertices of B_v is at most

$$\left(\frac{|B_v|}{\frac{k-d}{d}}\right) n^{-c \cdot \frac{k-d}{d}} < (100 \log^2 n \cdot n^{-c})^{c' \log n} = o(1/n),$$

where $c' = c'(c, d, \zeta) > 0$ is an appropriate constant.

Applying the union bound over all vertices of $\phi(V)$ proves that a.s. there are no failures.

Next we prove that a.s. $W \subseteq \phi(V)$. Fix some $w \in W$ and assume that $w \notin \phi(V)$. It follows that $w \in I$ holds throughout the embedding process. Consider an arbitrary point during the embedding process. At this given point let $v' \in S$ be an open vertex and let $x', y' \in V \setminus S$ be vertices for which $v'x', x'y' \in E$. Since $\phi(y') \neq w$, it follows that either $d_{G_2}(\phi(v'), X_w) = 0$ or

$d_{G_2}(\phi(v'), X_w) > 0$ but also $d_{G_2}(\phi(v'), X_z) > 0$ for some $z \in I \setminus \{w\}$. The probability of this happening is at most

$$(1 - p_2)^{|X_w|} + p_2^{|X_w|} \cdot \sum_{z \in I \setminus \{w\}} p_2^{|X_z|} \leq e^{-p_2^{|X_w|}} + \Theta(n^{-1-\beta_2} \log^4 n) \leq e^{-K \log^2 n/n},$$

where $K = K(\gamma, \beta_1) > 0$ is an appropriate constant.

Since $\Delta(T) \leq d$ and $|V| \geq an$, it follows that there exists some constant $c'' = c''(d, a) > 0$ such that there are at least $c''n$ vertices $v' \in V$ which are neither leaves nor parents of leaves. For every such vertex v' there are vertices $x', y' \in V$ such that $v'x', x'y' \in E$. In each embedding step we embed at most $d+1$ vertices. Hence there are at least $\frac{c''n}{d+1}$ attempts to embed vertices of V into w , where all attempts are mutually independent. It follows that

$$\Pr[w \notin \phi(V)] \leq \left(e^{-K \log^2 n/n}\right)^{c''n/(d+1)} = o(1/n).$$

Applying the union bound over all vertices of W proves that a.s. $W \subseteq \phi(V)$. This concludes the proof of the proposition. \square

Our proof of Proposition 2.8 does not apply to pseudo-random graphs. For such graphs we prove the following embedding criterion.

Proposition 2.9 *Let $\gamma, \varepsilon > 0$ be real numbers. Let $n_0 = n_0(\gamma, \varepsilon)$ be a sufficiently large positive integer, let $n \geq n_0$ be an integer and let $d = \sqrt{\log n}$. Let T be a tree on $n' \leq (1 - \varepsilon)n$ vertices with maximum degree at most d . Let $H = (V, E)$ be a graph on n vertices such that $\delta(H) \geq \gamma \log n$. If, moreover, H satisfies Properties **(P3)**, **(P4)** and **(P5)** from Lemma 2.2, then H contains a copy of T .*

The proof of Proposition 2.9 is via a simple application of the following corollary of a theorem of Haxell [12].

Theorem 2.10 [4] *Let d, m and M be positive integers, and let $0 \leq \ell \leq 2dm$. Assume that H is a non-empty graph satisfying the following two conditions.*

- (i) *For every $X \subseteq V(H)$ with $0 < |X| \leq m$, $|N_H(X)| \geq d|X| + 1$.*
- (ii) *For every $X \subseteq V(H)$ with $m < |X| \leq 2m$, $|N_H(X)| \geq d|X| + M$.*

Then H contains every tree T with $M + \ell$ vertices and maximum degree at most d , provided that T has at least ℓ leaves.

Proof of Proposition 2.9

Let ℓ denote the number of leaves of T and let $M = n' - \ell$. Let $m = \max \left\{ \frac{n(\log \log n)^{3/2}}{\log n}, \frac{\ell}{2d} \right\}$; note that $0 \leq \ell \leq 2dm$. Hence, by Theorem 2.10, it suffices to prove that H satisfies Properties (i) and (ii) above. First, assume for the sake of contradiction that there exists a set $X \subseteq V$

of size $0 < |X| \leq \min \left\{ m, \frac{n(\log \log n)^2}{\log n} \right\}$ such that $|N_H(X)| \leq d|X|$. Let $Y \supseteq N_H(X) \setminus X$ be an arbitrary subset of $V \setminus X$ of size $|Y| = d|X|$. Since H satisfies Property **(P4)** from Lemma 2.2, it follows that $e_H(X, Y) \leq \frac{|X| \log n}{\log \log n}$. On the other hand, since H satisfies Property **(P3)** from Lemma 2.2 and $\delta(H) \geq \gamma \log n$, it follows that $e_H(X, Y) \geq \gamma|X| \log n - \frac{2|X| \log n}{\log \log n} > \frac{|X| \log n}{\log \log n}$. This is clearly a contradiction. Next, assume for the sake of contradiction that there exists a set $X \subseteq V$ of size $\frac{n(\log \log n)^2}{\log n} < |X| \leq m$ such that $|N_H(X)| \leq d|X|$; note that $m = \frac{\ell}{2d}$ holds in this case. Let $Y = V \setminus (X \cup N_H(X))$. Note that $|Y| \geq n - (d+1)m \geq n - \ell \geq \varepsilon n \geq \frac{n(\log \log n)^{3/2}}{\log n}$ and $e_H(X, Y) = 0$. Since $|X| \geq \frac{n(\log \log n)^{3/2}}{\log n}$ and H satisfies Property **(P5)** from Lemma 2.2, it follows that $e_H(X, Y) > 0$. This is clearly a contradiction. Finally, assume for the sake of contradiction that there exists a set $X \subseteq V$ of size $m < |X| \leq 2m$ such that $|N_H(X)| < d|X| + M$. Let $Y = V \setminus (X \cup N_H(X))$. Note that $|Y| \geq n - 2m(d+1) - M \geq \varepsilon n + \ell - \frac{d+1}{d}\ell \geq \varepsilon n/2 \geq \frac{n(\log \log n)^{3/2}}{\log n}$ and $e_H(X, Y) = 0$. Since $|X| \geq m \geq \frac{n(\log \log n)^{3/2}}{\log n}$, this contradicts Property **(P5)** from Lemma 2.2. \square

Finally, we cite a known criterion for embedding not too large trees into sparse random graphs.

Theorem 2.11 (Theorem 1.1 from [2]) *Let $d \geq 2$, let $0 < \varepsilon < 1/2$ and let*

$$c \geq \varepsilon^{-1} 10^6 d^3 \log d \log^2(2\varepsilon^{-1}).$$

Then a.s. the random graph $G(n, c/n)$ contains every tree of maximum degree at most d on $(1 - \varepsilon)n$ vertices.

2.4 Embedding star forests in bipartite graphs

Assume we have embedded all vertices of some tree T , except for a set L of some of its leaves, into a graph G . Let $U \subseteq V(G)$ denote the image of $V(T) \setminus L$ under this embedding. In order to embed the vertices of L as well, we will need to connect certain vertices of U with vertices of $V(G) \setminus U$. The following lemma asserts that this is indeed possible, given that T and G satisfy certain conditions.

Lemma 2.12 *Let $F = (A_F \cup B_F, E_F)$ be a bipartite graph and let d be a positive integer such that $1 \leq d_F(u, B_F) \leq d$ holds for every $u \in A_F$ and $d_F(u, A_F) = 1$ holds for every $u \in B_F$. Let $G = (A \cup B, E)$ be a bipartite graph and assume that there exist positive integers δ_A , δ_B and s such that the following properties hold:*

- (i) $|A| = |A_F|$ and $|B| = |B_F|$.
- (ii) $d_G(u, B) \geq \delta_A$ for every $u \in A$ and $d_G(u, A) \geq \delta_B$ for every $u \in B$.
- (iii) $e_G(X, Y) < \min\{\delta_A|X|, \delta_B|Y|\}$ for every $X \subseteq A$ of size $|X| \leq s$ and every $Y \subseteq B$ of size $|Y| \leq d|X|$.
- (iv) $e_G(X, Y) > 0$ for every $X \subseteq A$ and $Y \subseteq B$ of sizes $|X|, |Y| > s$.

Then, for any bijection $f : A_F \rightarrow A$, there exists an embedding of F in G which maps every $u \in A_F$ to $f(u)$.

In the proof of Lemma 2.12 we will make use of the following polygamous version of Hall's Theorem (see e.g. [8]).

Proposition 2.13 *Let $G = (A \cup B, E)$ be a bipartite graph, where $A = \{a_1, \dots, a_k\}$ and $B = \{b_1, \dots, b_r\}$, and let d_1, \dots, d_k be positive integers. Then, there exists a spanning subgraph H of G such that $\deg_H(a_i) = d_i$ for every $1 \leq i \leq k$ and $\deg_H(b_j) = 1$ for every $1 \leq j \leq r$ if and only if $|N_G(S)| \geq \sum_{i: a_i \in S} d_i$ for every $S \subseteq A$.*

Proof of Lemma 2.12 Assume for the sake of contradiction that there exists a set $X \subseteq A$ which does not satisfy Hall's condition, that is, $|N_G(X)| < \sum_{u \in X} d_F(f^{-1}(u))$. Note that $\sum_{u \in X} d_F(f^{-1}(u)) \leq d|X|$ holds by our assumption that $d_F(u, B_F) \leq d$ for every $u \in A_F$. First, assume that $|X| \leq s$. It follows by Property (iii) above that $\delta_A|X| > e_G(X, N_G(X)) = \sum_{u \in X} d_G(u, B) \geq \delta_A|X|$. This is clearly a contradiction. Assume then that $|X| > s$. Let $Y \subseteq B \setminus N_G(X)$ be an arbitrary set of size $|Y| = \sum_{u \in A \setminus X} d_F(f^{-1}(u))$ (such a set exists since $\sum_{u \in A} d_F(f^{-1}(u)) = |B_F| = |B|$ and $|N_G(X)| < \sum_{u \in X} d_F(f^{-1}(u))$). Since, by assumption $1 \leq d_F(u, B_F) \leq d$ holds for every $u \in A_F$, it follows that $|A \setminus X| \leq |Y| \leq d|A \setminus X|$. If $|X| < |A| - s$, then $|Y| \geq |A \setminus X| > s$ which contradicts Property (iv) above since $|X| > s$ but $e_G(X, Y) = 0$. Assume then that $|X| \geq |A| - s$. Since $N_G(Y) \subseteq A \setminus X$, it follows that $e_G(A \setminus X, Y) = e_G(N_G(Y), Y) = \sum_{u \in Y} d_G(u, A) \geq \delta_B|Y|$. This contradicts Property (iii) above since $|A \setminus X| \leq s$ and $|Y| \leq d|A \setminus X|$. \square

2.5 Hamilton connectivity

In this subsection we prove a sufficient condition for a graph to be Hamilton connected.

Lemma 2.14 *Let $\beta, c > 0$ be real numbers, let $n_0 = n_0(\beta, c)$ be a sufficiently large positive integer and let $n \geq n_0$. Let G be a graph on n vertices which satisfies Properties (P3), (P4), and (P5). Let $H = (V, E)$ be an induced subgraph of G on cn vertices. If $\delta(H) \geq \beta \log n$, then H is Hamilton connected.*

In our proof of Lemma 2.14 we will make use of the following sufficient condition for a graph to be Hamilton connected.

Theorem 2.15 [13] *There exists an integer n_0 such that for every integers $n \geq n_0$ and $12 \leq d \leq e^{\sqrt[3]{\log n}}$ the following holds. If $G = (V, E)$ is a graph on n vertices which satisfies the following two properties:*

(H1) *For every $S \subseteq V$, if $|S| \leq \frac{n \log \log n \log d}{d \log n \log \log \log n}$, then $|N(S)| \geq d|S|$;*

(H2) *There is an edge in G between any two disjoint subsets $A, B \subseteq V$ with $|A|, |B| \geq \frac{n \log \log n \log d}{4130 \log n \log \log \log n}$;*

then G is Hamilton connected.

Proof of Lemma 2.14 Note that H satisfies Properties **(P3)**, **(P4)**, and **(P5)** from Lemma 2.2. It follows from Theorem 2.15 that it suffices to prove that H satisfies Properties **(H1)** and **(H2)** for some $12 \leq d \leq e^{\sqrt[3]{\log n}}$. Fix $d = \sqrt{\log n}$. Starting with **(H1)**, we claim that $|N_H(A)| \geq |A|\sqrt{\log n}$ holds for every $A \subseteq V$ of size $a \leq \frac{n}{\log n}$. Indeed, assume for the sake of contradiction that there exists a set $A \subseteq V$ of size $a \leq \frac{n}{\log n}$ such that $|N_H(A)| < a\sqrt{\log n}$. Since H is an induced subgraph of G , $\delta(H) \geq \beta \log n$, and H satisfies Property **(P3)** from Lemma 2.2, it follows that $e_H(A, V \setminus A) = e_G(A, V \setminus A) \geq a\beta \log n - 2\frac{a \log n}{\log \log n} \geq a\beta \log n / 2$. Let $B \subseteq V \setminus A$ be an arbitrary set of size $a\sqrt{\log n}$ which contains $N_H(A) \setminus A$. It follows from the discussion above that $e_H(A, B) \geq a\beta \log n / 2 > \frac{a \log n}{\log \log n}$. This contradicts the fact that H satisfies Property **(P4)** from Lemma 2.2. Since $\frac{n}{\log n} \geq \frac{cn \log \log(cn) \log(\sqrt{\log n})}{\sqrt{\log n} \log(cn) \log \log \log(cn)}$, it follows that H satisfies Property **(H1)**.

Next, we claim that H satisfies Property **(H2)**. Indeed, since H satisfies Property **(P5)** from Lemma 2.2 and since

$$\frac{n(\log \log n)^{3/2}}{\log n} \leq \frac{cn \log \log(cn) \log(\sqrt{\log n})}{4130 \log(cn) \log \log \log(cn)},$$

it follows that H satisfies Property **(H2)**. This concludes the proof of the lemma. \square

3 Proof of the main results

Proof of Theorem 1.2

We expose G in three rounds, that is, we split $G = G_1 \cup G_2 \cup G_3$, where $G_1, G_3 \sim G(n, \frac{\varepsilon \log n}{4n})$ and $G_2 \sim G(n, \frac{(1+\varepsilon/2)\log n}{n})$. Note that we thus indeed have $G \sim G(n, p)$ with $p \leq \frac{(1+\varepsilon)\log n}{n}$.

Let $r' \in V(T)$ be a vertex at which T can be split into two rooted subtrees T_1 and T_2 (both rooted at r') such that the subtree T_1 has γn vertices and $\gamma_1 n$ leaves for some $\beta < \gamma_1 < \gamma \leq \varepsilon/14$, where $\beta = \beta(\varepsilon/14, \alpha, d) > 0$ is the real number whose existence is guaranteed by Lemma 2.4. Let L_1 be the set of leaves of T_1 and let $T'_1 := T_1 \setminus L_1$.

We first expose the edges of G_1 . Let $f_1 : V(T'_1) \rightarrow [n]$ be an embedding of T'_1 in G_1 . Such an embedding exists a.s. by Theorem 2.11 since $|V(T'_1)| \leq n/2$. Let $U = f_1(V(T'_1))$ denote the image of the embedding f_1 , let $L_0 = f_1(N_{T_1}(L_1))$ denote the images of the parents of the leaves of T_1 under this embedding and let $r = f_1(r')$ denote the image of the root of T_1 . Note that $\gamma_1 n / d \leq |L_0| \leq |L_1| = \gamma_1 n$.

Now we expose the edges of G_2 with one endpoint in L_0 and the other in $[n] \setminus U$. It follows from Property **(P2)** of Lemma 2.2 that there exists a real number $\mu = \mu(\varepsilon, \gamma) > 0$ such that a.s. $d_{G_2}(v, [n] \setminus U) \geq \mu \log n$ holds for every $v \in L_0$. Moreover, it follows from Property **(P5)** of Lemma 2.2 that a.s. $e_{G_2}(A, B) > 0$ holds for every $A \subseteq L_0$ and every $B \subseteq [n] \setminus U$ of size $|A| = |B| = \frac{n(\log \log n)^{3/2}}{\log n}$.

Now we expose the edges of G_3 with one endpoint in L_0 and the other in $[n] \setminus U$. Let $W = \{w \in [n] \setminus U : d_{G_3}(w, L_0) < \frac{\varepsilon \log n}{8n} |L_0|\}$. We claim that a.s. W is a ‘‘small’’ set and that

$d_{G_2}(u, [n] \setminus (U \cup W)) \geq \mu \log n/2$ holds for every vertex $u \in L_0$. Indeed, starting with the former, for every $u \in [n] \setminus U$, let A_u denote the event: $u \in W$. For any $u \in [n] \setminus U$ we have $d_{G_3}(u, L_0) \sim \text{Bin}(|L_0|, \varepsilon \log n/(4n))$. Hence, it follows by standard bounds on the tail of the binomial distribution (see e.g. [15]) that

$$\begin{aligned} \Pr[A_u] &= \Pr \left[d_{G_3}(u, L_0) < \frac{\varepsilon \log n}{8n} |L_0| \right] = \Pr [d_{G_3}(u, L_0) < \mathbb{E}(d_{G_3}(u, L_0))/2] \\ &\leq \exp \left\{ -\frac{1}{8} |L_0| \frac{\varepsilon \log n}{4n} \right\} \leq n^{-\frac{\gamma_1 \varepsilon}{32d}}, \end{aligned}$$

where the last inequality follows from the aforementioned lower bound on $|L_0|$.

By Markov's inequality we conclude that a.s. $|W| \leq n^{1-\frac{\gamma_1 \varepsilon}{33d}}$.

Next, we prove the latter. Let $v \in L_0$ be an arbitrary vertex; recall that $d_{G_2}(v, [n] \setminus U) \leq \Delta(G_2) \leq 10 \log n$ holds by Property **(P1)** from Lemma 2.2 and, as previously noted, that $d_{G_2}(v, [n] \setminus U) \geq \mu \log n$. Let $u \in N_{G_2}(v, [n] \setminus U)$ be an arbitrary vertex. As previously noted $\Pr[A_u] \leq n^{-\frac{\gamma_1 \varepsilon}{32d}}$. Hence,

$$\begin{aligned} \Pr[d_{G_2}(v, [n] \setminus (U \cup W)) < \mu \log n/2] &\leq \Pr[d_{G_2}(v, W) \geq \mu \log n/2] \\ &\leq \sum_{\substack{S \subseteq N_{G_2}(v, [n] \setminus U) \\ |S| = \mu \log n/2}} \prod_{w \in S} \Pr[A_w] \\ &\leq \binom{10 \log n}{\mu \log n/2} (n^{-\frac{\gamma_1 \varepsilon}{32d}})^{\mu \log n/2} \\ &\leq \exp \left\{ 10 \log n - \frac{\gamma_1 \varepsilon \mu}{64d} \log^2 n \right\} \\ &= o(1/n), \end{aligned}$$

where the second inequality follows since A_{u_1} and A_{u_2} are independent events for every two distinct vertices $u_1, u_2 \in [n] \setminus U$.

It follows by a union bound argument that a.s. $d_{G_2}(v, [n] \setminus (U \cup W)) \geq \mu \log n/2$ holds for every $v \in L_0$.

Let $X \subseteq [n] \setminus (U \cup W)$ be a set of size $\gamma_1 n/2$ such that $d_{G_2 \cup G_3}(u, X) \geq \zeta \log n$ holds for every vertex $u \in L_0$, where $\zeta = \zeta(\mu, \gamma_1) > 0$ is an appropriately chosen real number. Such a set X exists by Lemma 2.5. Note that $d_{G_2 \cup G_3}(x, L_0) \geq \frac{\varepsilon \log n}{8n} |L_0| \geq \frac{\gamma_1 \varepsilon}{8d} \log n$ holds for every vertex $x \in X$ since $X \cap W = \emptyset$ by construction.

Now we expose the rest of the edges of G_2 . Let $f_2 : V(T_2) \rightarrow ([n] \setminus (U \cup X)) \cup \{r\}$ be an embedding of T_2 in $G_2([n] \setminus (U \cup X)) \cup \{r\}$ such that $f_2(r') = r$ and $W \subseteq f_2(V(T_2))$. Such an embedding exists a.s. by Proposition 2.8 since $|V(T_2)| \leq (1 - \gamma)n + 1 \leq (1 - \gamma_1/3)(1 - \gamma + \gamma_1/2)n \leq (1 - \gamma_1/3)|[n] \setminus (U \cup X)|$ (this gives $b < 1$ in Proposition 2.8) and since $|[n] \setminus (U \cup X)| \geq (1 - \varepsilon/9)n$ (this gives $\beta_1 < 0$ in Proposition 2.8 which ensures that $\beta_2 > 2\beta_1$ holds).

Let $\nu = \min\{\frac{\gamma_1 \varepsilon}{8d}, \zeta\}$ and let $Y = [n] \setminus (f_1(V(T'_1)) \cup f_2(V(T_2)))$. At this point of the embedding process all that is left to embed are the leaves of T_1 , to be embedded bijectively into Y . Note that $Y \supseteq X$ and that $Y \cap W = \emptyset$. It follows from the former and from our choice of the set X that $d_{G_2 \cup G_3}(u, Y) \geq \nu \log n$ holds for every $u \in L_0$. It follows from the latter

and from the definition of W that $d_{G_2 \cup G_3}(u, L_0) \geq \nu \log n$ holds for every $u \in Y$. Let $F = T_1[L_1 \cup N_{T_1}(L_1)]$ and let $H = (L_0 \cup Y, E_{G_2 \cup G_3}(L_0, Y))$. Since, by the aforementioned properties and by Properties **(P4)** and **(P5)** of Lemma 2.2, F and H satisfy the conditions of Lemma 2.12 (with $\delta_A = \delta_B = \nu \log n$, $d = \Delta(T)$ and $s = \frac{n(\log \log n)^2}{\log n}$), it follows that there exists an embedding f_3 of F in H such that $f_3(u) = f_1(u)$ for every $u \in N_{T_1}(L_1)$. It is clear that the mapping $\phi : V(T) \rightarrow [n]$ defined by

$$\phi(u) := \begin{cases} f_1(u) & \text{if } u \in V(T'_1) \\ f_2(u) & \text{if } u \in V(T_2) \\ f_3(u) & \text{if } u \in L_1 \end{cases}$$

is an embedding of T in G . This concludes the proof of the theorem. \square

Proof of Theorem 1.4

It follows by Lemma 2.2 that a.s. $G(n, (1 + \varepsilon) \log n/n)$ satisfies properties **(P1)**–**(P5)**. Hence, in order to prove that $G(n, (1 + \varepsilon) \log n/n)$ is a.s. \mathcal{L} -universal, it suffices to prove the any graph which satisfies properties **(P1)**–**(P5)** is \mathcal{L} -universal.

Let $G = (V, E)$ be a graph on n vertices which satisfies properties **(P1)**–**(P5)**. Let T be a tree on n vertices with maximum degree at most d which admits a bare path of length αn . We will prove that G contains T as a subgraph.

Let $t = \alpha n + 1$ and let $P = (v_1, v_2, \dots, v_t)$ be a bare path of length αn in T . Let T_P be the tree obtained from T by contracting the path P to a single edge between v_1 and v_t . Let $V = V_1 \cup V_2$ be a partition of V which satisfies the following properties:

- (i) $|V_1| = \alpha n/2$ and $|V_2| = (1 - \alpha/2)n$;
- (ii) $d_G(v, V_1) \geq \mu \log n$ for every $v \in V$;
- (iii) $d_G(v, V_2) \geq \mu \log n$ for every $v \in V$;

where $\mu = \mu(\alpha, \varepsilon) > 0$ is an appropriately chosen real number. Such a partition exists by Lemma 2.5.

Let $\phi : V(T_P) \rightarrow V_2$ be an embedding of T_P in $G[V_2]$. Such an embedding exists by Proposition 2.9 since $|V(T_P)| \leq (1 - \alpha/2)|V_2|$ and since $G[V_2]$ satisfies Property (iii) above and Properties **(P3)**–**(P5)** from Lemma 2.2. Let $U_2 := \phi(V(T_P))$ and let $U_1 := (V \setminus U_2) \cup \{\phi(v_1), \phi(v_t)\}$. In order to complete the embedding of T in G , it suffices to prove that there is a Hamilton path in $G[U_1]$ whose endpoints are precisely $\phi(v_1)$ and $\phi(v_t)$. In order to do so it suffices to prove the stronger result asserting that $G[U_1]$ is Hamilton connected. This however readily follows from Lemma 2.14 since $G[U_1]$ satisfies Properties **(P3)**–**(P5)** from Lemma 2.2 and since $\delta(G[U_1]) \geq \mu \log n$ holds by Property (ii) above. \square

4 Concluding remarks

We have proven that a bounded degree tree T on n vertices is contained almost surely in a random graph $G(n, (1 + \varepsilon) \log n/n)$, where $\varepsilon > 0$ is arbitrarily small but fixed, provided that T has linearly many leaves or alternatively in the case where T contains a bare path of linear length (in which case $G(n, (1 + \varepsilon) \log n/n)$ a.s. contains all such trees simultaneously). These results are optimal as for $p(n) = (1 - \varepsilon) \log n/n$ the random graph $G(n, p)$ is almost surely disconnected and thus does not contain spanning trees at all.

Our proofs can be tightened rather easily to show that the embedding results we obtained hold already for $p(n) = (1 + \varepsilon(n)) \log n/n$, where $\varepsilon(n)$ is some concrete function tending to 0 with n . Since we are rather doubtful the bounds on the error term $\varepsilon(n)$ obtained in this fashion would be close to being optimal, we chose not to pursue this goal.

Our proof of Theorem 1.4 uses randomness in a rather limited way and thus applies to pseudo-random graphs as well. Namely, we in fact proved that any graph G on n vertices satisfying Properties **(P1)**–**(P5)** contains any given bounded degree tree which admits a bare path on $\Theta(n)$ vertices. Hence the random graph $G(n, (1 + \varepsilon) \log n/n)$ is a.s. *universal* for this class of trees. Our proof of Theorem 1.2 relies on the multiple rounds of exposure in an essential way and thus cannot be applied in a pseudo-random setting. It would be interesting to prove universality-type results for the trees covered by Theorem 1.2.

It was proved in [7] that, if $p = c \log n/n$ for any constant $c > 2e^2$, then a.s. $G(n, p)$ contains almost every tree on n vertices. Since a typical tree on n vertices has $\Omega(n)$ leaves, it seems plausible that one could adjust Theorem 1.2 to show that a.s. almost every tree is contained in $G(n, (1 + \varepsilon) \log n/n)$. Since the maximum degree of a random tree is a.s. $(1 + o(1)) \log n / \log \log n$ (see e.g. [17]), this is not immediate. However, it is indeed doable. We sketch the main required changes below.

Using similar (yet more careful) arguments to the ones used in the proof of Lemma 2.4, one can show its assertion holds for any tree with $\Omega(n)$ leaves (without any assumption on its maximum degree). Moreover, using known properties of random trees, one can ensure that a.s. T_1 will have $\Omega(n)$ parents of leaves and T_2 will have $\Omega(n)$ vertices which are neither leaves nor parents of leaves. Due to this property of T_2 and since a.s. $\Delta(T_2) \leq \Delta(T) \leq (1 + o(1)) \log n / \log \log n$, one can show that Proposition 2.8 still applies to T_2 (with very few and straightforward changes to its proof). Finally, due to the aforementioned property of T_1 , the proof of Theorem 1.2 is also mostly unaffected (except for small changes). The only substantial difference is that we cannot use Theorem 2.11 to embed T'_1 (since its maximum degree is unbounded). Instead we embed T'_1 at the same stage (and in a similar fashion) as T_2 . Hence, we split G into two graphs only, $G_1 \sim G(n, \frac{(1+\varepsilon/2)\log n}{n})$ and $G_2 \sim G(n, \frac{\varepsilon \log n}{2n})$. We embed $T'_1 \cup T_2$ in G_1 and the remainder of T in G_2 . We expose G_1 vertex by vertex, embed T'_1 first and then T_2 . Let $r := f_1(r')$ and let $U = f_1(V(T'_1))$. As in the current proof we embed T_2 in $G_1[(\{r\} \cup U) \cup \{r\}]$. That is, we need to use the neighbors of r in G_1 twice: once for embedding the neighbors of r' in T'_1 and once for embedding its neighbors in T_2 . Let A and B be disjoint sets of roughly equal size such that $A \cup B = [n] \setminus \{r\}$. when embedding T'_1 we expose only the edges of $E_{G_1}(\{r\}, A)$ and then, when embedding T_2 , we expose the edges of $E_{G_1}(\{r\}, B)$.

References

- [1] M. Ajtai, J. Komlós and E. Szemerédi, The longest path in a random graph, *Combinatorica* 1 (1981), 1–12.
- [2] N. Alon, M. Krivelevich and B. Sudakov, Embedding nearly-spanning bounded degree trees, *Combinatorica* 27 (2007), 629–644.
- [3] N. Alon and J. H. Spencer, **The Probabilistic method**, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 3rd edition, 2008.
- [4] J. Balogh, B. Csaba, M. Pei, and W. Samotij, Large bounded degree trees in expanding graphs, *Electronic J. Combinatorics* 17 (2010), R6.
- [5] J. Balogh, B. Csaba and W. Samotij, Local resilience of almost spanning trees in random graphs, *Random Structures and Algorithms* 38 (2011), 121–139.
- [6] J. Beck, On size Ramsey number of paths, trees and circuits. I, *Journal of Graph Theory*, 7 (1983) 115–129.
- [7] E. A. Bender and N. C. Wormald, Random trees in random graphs, *Proc. Amer. Math. Soc.* 103 (1988), 314–320.
- [8] R. Diestel, **Graph Theory**, 4th edition, Springer, 2010.
- [9] W. Fernandez de la Vega, Long paths in random graphs, *Studia Sci. Math. Hungar.* 14 (1979), 335–340.
- [10] W. Fernandez de la Vega, Trees in sparse random graphs, *Journal of Combinatorial Theory Ser. B* 45 (1988), 77–85.
- [11] J. Friedman and N. Pippenger, Expanding graphs contain all small trees, *Combinatorica* 7 (1987), 71–76.
- [12] P. Haxell, Tree embeddings, *Journal of Graph Theory* 36 (2001), 121–130.
- [13] D. Hefetz, M. Krivelevich and T. Szabó, Hamilton cycles in highly connected and expanding graphs, *Combinatorica* 29 (2009), 547–568.
- [14] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. American Stat. Assoc.* 58 (301) (1963), 13–30.
- [15] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, 2000.
- [16] M. Krivelevich, Embedding spanning trees in random graphs, *SIAM J. Discrete Math.* 24 (2010), 1495–1500.
- [17] J. W. Moon, On the maximum degree in a random tree, *Michigan Math. J.* 15 (1968), 429–432.
- [18] D. B. West, **Introduction to Graph Theory**, Prentice Hall, 2nd edition, 2001.