Embedding nearly-spanning bounded degree trees

Noga Alon * Michael Krivelevich † Benny Sudakov ‡

Abstract

We derive a sufficient condition for a sparse graph G on n vertices to contain a copy of a tree T of maximum degree at most d on $(1-\epsilon)n$ vertices, in terms of the expansion properties of G. As a result we show that for fixed $d \geq 2$ and $0 < \epsilon < 1$, there exists a constant $c = c(d, \epsilon)$ such that a random graph $G(n, c/n)$ contains almost surely a copy of every tree T on $(1-\epsilon)n$ vertices with maximum degree at most d. We also prove that if an (n, D, λ)-graph G (i.e., a D-regular graph on n vertices all of whose eigenvalues, except the first one, are at most λ in their absolute values) has large enough spectral gap D/λ as a function of d and ϵ, then G has a copy of every tree T as above.

1 Introduction

In this paper we obtain a sufficient condition for a sparse graph G to contain a copy of every nearly-spanning tree T of bounded maximum degree, in terms of the expansion properties of G. The restriction on the degree of T comes naturally from the fact that we consider graphs of constant degree. Two important examples where our condition applies are random graphs and graphs with a large spectral gap.

The random graph $G(n, p)$ denotes the probability space whose points are graphs on a fixed set of n vertices, where each pair of vertices forms an edge, randomly and independently, with probability p. We say that the random graph $G(n, p)$ possesses a graph property $\mathcal{P}$ *almost surely*, or a.s. for short, if the probability that $G(n, p)$ satisfies $\mathcal{P}$ tends to 1 as the number of vertices n tends to infinity.

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel and IAS, Princeton, NJ 08540, USA. Email: nogaa@post.tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by NSF grant CCR-0324906, by a Wolfensohn fund and by the State of New Jersey.

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2002-133, and by grants 64/01 and 526/05 from the Israel Science Foundation.

‡Department of Mathematics, Princeton University, Princeton, NJ 08544. E-mail: bsudakov@math.princeton.edu. Research supported in part by NSF CAREER award DMS-0546523, NSF grant DMS-0355497, USA-Israeli BSF grant, and by an Alfred P. Sloan fellowship.

The problem of existence of large trees with specified shape in random graphs has a long history with most of the results being devoted to finding a long path. Erdős conjectured that a random graph $G(n, c/n)$ a.s. contains a path of length at least $(1 - \alpha(c))n$, where $\alpha(c)$ is a constant smaller than one for all $c > 1$ and $\lim_{c\to\infty} \alpha(c) = 0$. This conjecture was proved by Ajtai, Komlós and Szemerédi [1] and, in a slightly weaker form, by Fernandez de la Vega [11]. These results were significantly improved by Bollobás [5], who showed that $\alpha(c)$ decreases exponentially in c. Finally Frieze [16] determined the correct speed of convergence of $\alpha(c)$ to zero and proved that $\alpha(c) = (1 + o(1))ce^{-c}$.

The question of existence of large trees of bounded degree other than paths in sparse random graphs was studied by Fernandez de la Vega in [12]. He proved that there exist two constants $a_1 > 0$ and $a_2 > 0$ such that for fixed tree T of order n/a_1 with maximum degree at most d the random graph $G(n, c/n)$ with $c = a_2 d$ almost surely contains T. The constant a_1 in this result is rather large and allows to embed only trees that occupy a small proportion of the random graph. Also, observe that Fernandez de la Vega's result gives the almost sure existence of a *fixed* tree T, and not of all such trees simultaneously.

Our first theorem improves the result of Fernandez de la Vega and generalizes the above mentioned results on the existence of long paths. It shows that the sparse random graph contains almost surely every nearly-spanning tree of bounded degree.

Theorem 1.1 *Let $d \geq 2$, $0 < \varepsilon < 1/2$ and let*

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

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

Results guaranteeing the existence of a long path in a sparse graph can be obtained in a more general situation when the host graph has certain expansion properties. Given a graph $G = (V, E)$ and a subset $X \subset V$ let $N_G(X)$ denote the set of all neighbors of vertices of X in G. Using his celebrated rotation-extension technique, Pósa [18] proved that if for every X in G with $|X| \leq k$ we have that $|N_G(X) \setminus X| \geq 2|X| - 1$, then G contains a path of length $3k - 2$. A remarkable generalization of this result from paths to trees of bounded degree was obtained by Friedman and Pippenger [15]. They proved that if $|N_G(X)| \geq (d + 1)|X|$ for every set X in G with $|X| \leq 2k - 2$, then G contains every tree with k vertices and maximum degree at most d. Note that this result allows to embed only trees whose size is relatively small compared to the size of G. What if we want to embed trees which are nearly-spanning? It turns out that a slightly stronger expansion property, based on the spectral gap condition, is already enough to attain this goal.

For a graph G let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of its adjacency matrix. The quantity $\lambda(G) = \max_{i\geq 2} |\lambda_i|$ is called the *second eigenvalue* of G. A graph $G = (V, E)$ is called an (n, D, λ)-*graph* if it is D-regular, has n vertices and the second eigenvalue of G is at most λ. It is well known (see, e.g., [3] for more details) that if λ is much smaller than the degree D, then G has strong expansion properties, so the ratio D/λ could serve as some kind of measure of expansion of G. Our next result shows that an (n, D, λ)-graph G with large enough spectral gap D/λ contains a copy of every nearly-spanning tree with bounded degree.

Theorem 1.2 *Let $d \geq 2$, $0 < \varepsilon < 1/2$ and let G be an (n, D, λ)-graph such that*

$$\frac{D}{\lambda} \geq \frac{160d^{5/2} \log(2/\varepsilon)}{\varepsilon}.$$

Then G contains a copy of every tree T with $(1 - \varepsilon)n$ vertices and with maximum degree at most d.

Our main results are tightly connected to the notion of universal graphs. For a family $\mathcal{H}$ of graphs, a graph G is $\mathcal{H}$-universal if G contains every member of $\mathcal{H}$ as a (not necessarily induced) subgraph. The construction of sparse universal graphs for various families arises in several fields such as VLSI circuit design, data representation and parallel computing (see, e.g., the introduction of [2] for a short survey and relevant references). Our two main results show that sparse random graphs and pseudo-random graphs on n vertices are universal graphs for the family of bounded-degree trees on $(1-\varepsilon)n$ vertices. Quite an extensive research exists on universal graphs for trees [4], [7], [8], [9], [10], [15]. The most interesting result is that of Bhatt et al. who showed in [4] that there exists a universal graph G on n vertices for the family of trees on n vertices with maximum degree d, whose maximum degree is bounded by a function of d. It is instructive to compare our results with those of [4]: they succeed in embedding spanning trees as opposed to nearly-spanning in our case; on the other hand, their universal graph is a concrete carefully constructed graph that has very dense pieces locally, while we provide a very large family of universal graphs possessing many additional properties that can be useful for obtaining further results on universal graphs.

The results of Theorem 1.1 and 1.2 can be deduced from a more general statement which we present next. We need the following definition.

Definition 1.3 *Given two positive numbers c and $\alpha < 1$, a graph $G = (V, E)$ is called an (α, c)-expander if every subset of vertices $X \subset V(G)$ of size $|X| \leq \alpha|V(G)|$ satisfies:*

$$|N_G(X)| \geq c|X| \ .$$

Theorem 1.4 *Let $d \geq 2$, $0 < \varepsilon < 1/2$. Let $G = (V, E)$ a graph on n vertices of minimum degree $\delta = \delta(G)$ and maximum degree $\Delta = \Delta(G)$. Let n, δ, Δ satisfy:*

1. *(order of graph is large enough)*

$$n \geq \frac{480d^3 \log(2/\varepsilon)}{\varepsilon};$$

2. (maximum degree is not too large compared to the minimum degree)

$$\Delta^2 \leq \frac{1}{K} e^{\delta/(8K)-1} \quad \textit{where} \quad K = \frac{20d^2 \log(2/\varepsilon)}{\varepsilon}.$$

3. (local expansion) Every induced subgraph G_0 of G with minimum degree at least $\frac{\varepsilon\delta}{40d^2 \log(2/\varepsilon)}$ is a $(\frac{1}{2d+2}, d+1)$-expander.

Then G contains a copy of every tree T on at most $(1-\varepsilon)n$ vertices of maximum degree at most d.

The rest of this paper is organized as follows. In the next two sections we show how Theorem 1.4 can be used to embed nearly-spanning trees of bounded degree in random and pseudo-random graphs. We present the proof of Theorem 1.2 first, since it is short and less technical, and then prove Theorem 1.1. In Section 4 we describe the plan of the proof of Theorem 1.4 and discuss some technical tools needed to fulfill this plan. The proof of this theorem appears in Section 5. The last section of the paper contains several concluding remarks and open problems.

Throughout the paper we make no attempts to optimize the absolute constants. To simplify the presentation, we often omit floor and ceiling signs whenever these are not crucial. Throughout the paper, log denotes logarithm in the natural base e.

2 Embedding in pseudo-random graphs

In this section we prove Theorem 1.2. First we need the following lemma that shows that an (n, D, λ)-graph has the local expansion property required by Condition 3 of Theorem 1.4.

Lemma 2.1 *Let $d \geq 2$. Let $G = (V, E)$ be an (n, D, λ)-graph. Denote*

$$D_0 = \frac{2\lambda(d+1)}{\sqrt{d}} \ .$$

Then every induced subgraph G_0 of G of minimum degree at least D_0 is a $\left(\frac{1}{2d+2}, d+1\right)$-expander.

Proof. We will use the following well known estimate on the edge distribution of an (n, D, λ)-graph G (see, e.g., [3], Corollary 9.2.5). For every two (not necessarily disjoint) subsets $B, C \subseteq V$, let $e(B, C)$ denote the number of ordered pairs (u, v) with $u \in B, v \in C$ such that uv is an edge. Note that if $u, v \in B \cap C$, then the edge uv contributes 2 to $e(B, C)$. In this notation,

$$\left| e(B, C) - \frac{|B||C|D}{n} \right| \leq \lambda \sqrt{|B||C|} \ .$$

Let U be a subset of vertices of G such that the induced subgraph $G_0 = G[U]$ has minimum degree at least D_0. Suppose that the claim is false. Then there exists a subset $X \subset U$ of size $|X| = t \leq |U|/(2d+2)$ satisfying $|N_{G_0}(X)| < (d+1)|X|$.

By the above estimate with $B = X$ and $C = N_{G_0}(X)$ we have:

$$D_0 t \le e(B, C) \le \frac{t(d+1)tD}{n} + \lambda t\sqrt{d+1}$$

and therefore

$$\frac{t}{n} \ge \frac{D_0}{(d+1)D} - \frac{\lambda}{\sqrt{d+1}D} \; . \tag{1}$$

Also, note that there are no edges of G from X to $Y = U - (X \cup N_{G_0}(X))$ as G_0 is an induced subgraph of G. From $t = |X| \le |U|/(2d+2)$ and $|N_{G_0}(X)| \le (d+1)t$ it follows that $|Y| \ge dt$. Thus

$$0 = e(X, Y) \ge \frac{t(dt)D}{n} - \lambda\sqrt{t(dt)}\,,$$

implying

$$\frac{t}{n} \le \frac{\lambda}{\sqrt{d}D} \; . \tag{2}$$

Comparing (1) and (2) we obtain

$$\frac{D_0}{(d+1)D} - \frac{\lambda}{\sqrt{d+1}D} \le \frac{\lambda}{\sqrt{d}D} \; .$$

Plugging in the definition of D_0 we derive a contradiction. $\square$

Proof of Theorem 1.2. Since every graph with minimal degree k contains all the trees on k vertices we can assume that $D \le (1-\varepsilon)n$. Let A be the adjacency matrix of G. The trace of A^2 equals the number of ones in A, which is exactly $2|E(G)| = nD$. We thus obtain that

$$nD = Tr(A^2) = \sum_{i=1}^{n} \lambda_i^2 \le D^2 + (n-1)\lambda^2$$

and therefore $\lambda^2 \ge \frac{D(n-D)}{n-1} \ge \varepsilon D$. This together with our assumption on D/λ implies

$$n \ge D \ge \varepsilon \left(\frac{D}{\lambda}\right)^2 \ge \frac{160^2 d^5 \log^2(2/\varepsilon)}{\varepsilon}.$$

Since $\Delta(G) = \delta(G) = D$, from this inequality it follows that G satisfies Conditions 1, 2 of Theorem 1.4. Finally since

$$\frac{\varepsilon D}{40 d^2 \log(2/\varepsilon)} \ge \frac{2(d+1)\lambda}{\sqrt{d}}$$

we can conclude using Lemma 2.1 that G also satisfies the last condition of Theorem 1.4. Thus G contains every tree of size $(1-\varepsilon)n$ with maximum degree at most d. $\square$

3 Embedding in random graphs

To prove Theorem 1.1 we first need to show that a sparse random graph contains a.s. a nearly spanning subgraph with good local expansion properties.

Lemma 3.1 *For every integer $d \geq 2$, real $0 < \theta < 1/2$ and $D \geq 50\theta^{-1}$ the random graph $G\big(n, \frac{4D}{n}\big)$ almost surely contains a subgraph G^* having the following properties:*

1. $|V(G^*)| \geq (1-\theta)n$;
2. $D \leq d_{G^*}(v) \leq 10D$ *for every* $v \in V(G^*)$;
3. *every induced subgraph G_0 of G^* of minimum degree at least $D_0 = 100d \log D$ is a $\big(\frac{1}{2d+2}, d+1\big)$-expander.*

The following statement contains a few easy facts about random graphs.

Proposition 3.2 *Let $G(n,p)$ be a random graph with $np > 20$, then almost surely*

(i) The number of edges between any two disjoint subsets of vertices $A, |A| = a$ and $B, |B| = b$ with $abp \geq 32n$ is at least $abp/2$ and at most $3abp/2$.

(ii) Every subset of vertices of size $a \leq n/4$ spans less than $anp/2$ edges.

Proof. (i) Since the number of edges between A and B is a binomially distributed random variable with parameters ab and p, it follows by the standard Chernoff-type estimates (see, e.g., [3]) that (denoting $t = abp/2$)

$$\mathbb{P}\Big[e(A,B) - abp < -t\Big] \leq e^{-\frac{t^2}{2abp}} = e^{-abp/8}$$

and

$$\mathbb{P}\Big[e(A,B) - abp > t\Big] \leq e^{-\frac{t^2}{2abp}+\frac{t^3}{2(abp)^2}} = e^{-abp/16}.$$

Using that $abp \geq 32n$ we can bound the probability that there are sets A, B with $|e(A,B) - abp| > abp/2$ by $2^n \cdot 2^n \cdot \big(2e^{-2n}\big) = o(1)$.

Since $np/2 \geq 10$ and $n/a \geq 4$, the probability that there is a subset of size a which violates the assertion (ii) is at most

$$\mathbb{P}_a \leq \binom{n}{a}\binom{a^2/2}{anp/2}p^{anp/2} \leq \left(\frac{en}{a}\Big(\frac{ea}{np}\Big)^{np/2}p^{np/2}\right)^a = \left(\frac{e^{np/2+1}}{(n/a)^{np/2-1}}\right)^a \leq \left(\frac{e^{11}}{(n/a)^9}\right)^a.$$

It is easy to see that $\mathbb{P}_a \ll n^{-1}$ for all $a \leq n/4$ and so $\sum_a \mathbb{P}_a = o(1)$. $\square$

Proof of Lemma 3.1. Let $G = G(n,p)$ be a random graph with $p = \frac{4D}{n}$ and let X be the set of $\theta n/2$ vertices of largest degrees in G. By Part (ii) of Proposition 3.2, a.s. this set spans less than $|X|np/2 = 2D|X|$ edges. Also, since $\frac{4D}{n}|X|(n-|X|) \geq 2D\theta(n/2) \geq 50n$, Part (i) of this proposition implies that a.s. the number of edges between X and $V(G) - X$ is at most $3|X|np/2 = 6D|X|$. Therefore the sum of the degrees of the vertices in X is bounded by $10D|X|$ and hence there is a vertex in X with degree at most $10D$. By definition of X, this implies that there are at most $\theta n/2$ vertices in G with degree larger than $10D$. Delete these vertices and denote the remaining graph by G'. Next as long as G' contains a vertex v of degree less than D, delete it. If we deleted more than $\theta n/2$ vertices, then the original random graph contains two sets Y and $V(G') - Y$ such that $|Y| = \theta n/2$, $|V(G') - Y| \geq (1-\theta)n \geq n/2$ and there are less than $D|Y| \leq p|Y||V(G') - Y|/2$ edges between them. Since $\frac{4D}{n}|Y||V(G') - Y| \geq \theta D n \geq 50n$, again by Part (i) of the previous statement this a.s. does not happen. Denote the resulting graph by G^*. Then it satisfies the first two conditions of the lemma and it remains to verify the third condition.

Suppose to the contrary that G^* contains a subset of vertices U such that the induced subgraph $G_0 = G^*[U]$ has minimum degree at least $D_0 = 100d\log D$ and is not a $\left(\frac{1}{2d+2}, d+1\right)$-expander. Then there exists a set $X \subset U$ of size $|X| = t$ such that the set $C = N_{G_0}(X)$ has size at most $(d+1)t$ and there are at least $D_0|X|/2 = 50dt\log D$ edges with an end in X and another end in C. If $t \leq \frac{\log D}{D}n$, then the probability that $G(n,p)$ contains such sets is at most

$$
\begin{aligned}
\mathbb{P}_t &\leq \binom{n}{t}\binom{n}{(d+1)t}\binom{t(d+1)t}{50dt\log D}p^{50dt\log D} \\
&\leq \left[\left(\frac{en}{t}\right)\left(\frac{en}{(d+1)t}\right)^{d+1}\left(\frac{e(d+1)tp}{50d\log D}\right)^{50d\log D}\right]^t \\
&\leq \left[e\left(\frac{n}{t}\right)^{2d}\left(\frac{e}{8}\cdot\frac{Dt}{n\log D}\right)^{50d\log D}\right]^t \\
&= \left[e\left(\frac{e}{8}\right)^{50d\log D}\left(\frac{D}{\log D}\right)^{2d}\left(\frac{Dt}{n\log D}\right)^{50d\log D - 2d}\right]^t \\
&< \left[e^{-25d\log D + 2d\log D + 1}\left(\frac{t}{n\log D/D}\right)^{40d\log D}\right]^t \\
&\leq \left[D^{-20d}\left(\frac{t}{n\log D/D}\right)^{40d\log D}\right]^t.
\end{aligned}
$$

Checking separately two cases $t < \log n$ and $\log n \leq t < \frac{\log D}{D}n$ it is easy to see that in both $\mathbb{P}_t \ll n^{-1}$. If $t \geq \frac{\log D}{D}n$ we apply a different argument. Note that there are no edges of $G(n,p)$ from X to $C = U - (X \cup N_{G_0}(X))$ since G_0 is an induced subgraph. From $t = |X| \leq |U|/(2d+2)$ and $|N_{G_0}(X)| \leq (d+1)t$ it follows that $|C| \geq dt$ and therefore the probability of such event in $G(n,p)$ is at most

$$
\begin{aligned}
\mathbb{P}_t &\le \binom{n}{t}\binom{n}{dt}(1-p)^{dt^2} \le \left[\frac{en}{t}\cdot\left(\frac{en}{dt}\right)^d e^{-pdt}\right]^t \\
&\le \left[\left(\frac{en}{t}\right)^{2d} e^{-pdt}\right]^t = \left[\left(\frac{en}{t}\right)^2 \cdot e^{-pt}\right]^{dt} \\
&\le \left[\left(\frac{en}{n\log D/D}\right)^2 \cdot e^{-\frac{4D}{n}\cdot\frac{\log D}{D}n}\right]^{dt} \\
&\le \left(D^2 D^{-4}\right)^{dt} = o(n^{-1}) .
\end{aligned}
$$

Thus the probability that G^* fails to satisfy the third condition is at most $\sum_{t=1}^n \mathbb{P}_t = o(1)$. □

Proof of Theorem 1.1. Let $d \ge 2$, $0 < \varepsilon < 1/2$ and c satisfy the assumption of Theorem 1.1. Set $\theta = 0.01\varepsilon$, $D = c/4$ and let $\varepsilon_1 = \frac{\varepsilon-\theta}{1-\theta} \ge 0.99\varepsilon$. Then by Lemma 3.1 $G(n, c/n)$ almost surely contains a subgraph G^* of order $n_1 \ge (1-\theta)n$ such that $D \le \delta(G^*) \le \Delta(G^*) \le 10D$ and every induced subgraph of G^* with minimum degree at least $100d\log D$ is an $\left(\frac{1}{2d+2}, d+1\right)$-expander. Using that $\Delta(G^*) \le 10\delta(G^*)$ and

$$
n_1 \ge \delta(G^*) \ge D \ge \frac{10^6 d^3 \log d \log^2(2/\varepsilon)}{4\varepsilon} > \frac{480 d^3 \log(2/\varepsilon_1)}{\varepsilon_1}
$$

we conclude that G^* satisfies Conditions 1 and 2 of Theorem 1.4 (with ε_1). To verify the third condition it is enough to check that the assumptions in Theorem 1.1 imply that

$$
100d\log D \le \frac{\varepsilon_1 D}{40d^2\log(2/\varepsilon_1)}.
$$

Note that one can simply substitute the lower bound for D in the above expression, since $x/\log x$ is an increasing function for $x > 3$. Therefore by Theorem 1.4, G^* contains every tree of size $(1-\varepsilon_1)n_1 \ge (1-\varepsilon_1)(1-\theta)n = (1-\varepsilon)n$ with maximum degree at most d. □

4 Embedding plan and main tools

To prove Theorem 1.4 we will use the following framework. Given a tree T, we first cut it into subtrees $T_1, T_2, \dots, T_s$ of carefully chosen sizes, so that the number s of these subtrees satisfies $s \le 10d^2\log(2/\varepsilon)$, and each subtree T_i is connected by a unique edge to the union of all previous subtrees. The subtrees T_i will be embedded sequentially in order, starting from T_1.

We then choose s pairwise disjoint sets of vertices $S_1, S_2, \dots, S_s$ whose total size is at most $\varepsilon n/2$, such that each vertex of the graph has many neighbors in each set S_i. The set S_i will be used only when embedding the subtree T_i, and will not be touched before that step. (It can be used later, but we will not do it here, as it complicates matters and does not improve the estimates in any essential way). During the embedding process we maintain a set R of at most s vertices, which will consist of all roots of the trees T_i that still have to be embedded, and will not contain any vertex of the sets S_i.

At the i-th step we are to embed the tree T_i starting from a given root $x_i \in R$. (At the first step a root is chosen arbitrarily.) Suppose that the current set of unused vertices of G is V_{i-1}. We take an arbitrary subset U_i of size $|U_i| = \Theta(|V(T_i)|d)$ which contains the vertex x_i that will be the root of T_i (but contains no other members of R), contains the set S_i, and contains no member of S_j for $j \neq i$. Note that as each vertex has many neighbors in S_i, the minimum degree in the induced subgraph of G on U_i is large.

Since the minimum degree of $G[U_i]$ is large enough, we can use the result of Friedman and Pippenger to embed a copy of T_i in U_i, rooting it at x_i. (We actually need a slightly modified, rooted version, of their result). All vertices of U_i unused when embedding T_i are recycled, and we thus get V_i by deleting from V_{i-1} only the vertices used for embedding T_i.

The final step of embedding T_i is to embed the edges crossing from T_i to yet unembedded pieces T_j, with $j > i$ (using vertices of U_i). We then add the endpoints of those edges outside T_i to the list R of special vertices, and delete x_i from R. Each of the newly added special vertices will serve as a root for embedding the corresponding piece T_j. Observe that the number of special vertices is at most s at any stage of the embedding.

The precise technical details are described in what follows.

4.1 The result of Friedman and Pippenger

The cornerstone of our proof is the embedding result of Friedman and Pippenger. In fact, we need a slightly stronger version of it – they showed the existence of a tree T, while we need to embed a rooted version of T in G starting from a fixed vertex $v \in V(G)$ as its root. Luckily, a careful reading of [15] reveals that the following holds as well.

Theorem 4.1 ([15]) *Let T be a tree on k vertices of maximum degree at most d rooted at r. Let $H = (V, E)$ be a non-empty graph such that, for every $X \subset V(H)$ with $|X| \leq 2k - 2$,*

$$|N_H(X)| \geq (d+1)|X| \ .$$

Let further $v \in V(H)$ be an arbitrary vertex of H. Then H contains a copy of T, rooted at v.

4.2 Cutting the tree into pieces

Proposition 4.2 *Let $d \geq 2$ and k be positive integers. Let T be a tree on at least $k+1$ vertices with maximum degree at most d. Then there exists an edge $e \in E(T)$ such that at least one of the two trees obtained from T by deleting e has at least k and at most $(d-1)(k-1)+1$ vertices.*

Proof. Choose a leaf r of T arbitrarily and root T at r. For $i \geq 0$ denote by L_i the set of vertices of T at distance i from r. For a vertex $v \in V(T)$ let $t(v)$ be the number of vertices in the subtree of T rooted at v. Now, let

$$i_0 = \max\{i : L_i \text{ contains a vertex } v \text{ with } t(v) \geq k\}\ .$$

As L_1 has only one vertex v with $t(v) = |V(T)| - 1 \geq k$, it follows that $i_0 \geq 1$. Choose a vertex $u \in L_{i_0}$ such that $t(u) \geq k$. Then by the definition of i_0 all sons w of u in T satisfy: $t(w) \leq k-1$, the number of sons does not exceed $d-1$, and therefore $t(u) \leq (d-1)(k-1)+1$.

Let now x be the father of u in T. Then $e = (x, u)$ is the required edge. □

Corollary 4.3 *Suppose $0 < \varepsilon < 1/2$, and let T be an arbitrary tree on $(1-\varepsilon)n$ vertices, with maximum degree at most d. Then one can cut T into s subtrees $T_1, T_2, \ldots, T_s$, so that each tree T_i is connected by a unique edge to the union of all trees T_j with $j < i$, and such that for every $i > 1$,*

$$\frac{\varepsilon n/2 + \sum_{j>i} |V(T_j)|}{8d^2} \leq |V(T_i)| \leq \frac{\varepsilon n/2 + \sum_{j>i} |V(T_j)|}{8d}.$$

For $i = 1$, the upper bound holds, but the lower bound may fail. Moreover, $s \leq 10d^2 \log(2/\varepsilon)$.

Proof. We choose the trees T_i one by one, starting from the *last* one. By Proposition 4.2 we first find a tree T_1' of size at least $\frac{\varepsilon n}{16d^2}$ and at most $\frac{\varepsilon n}{16d}$, and omit it from T. Suppose we have already chosen $T_1', T_2', \ldots, T_{i-1}'$ such that for every $j < i$

$$\frac{\varepsilon n/2 + \sum_{r<j} |V(T_r')|}{8d^2} \leq |V(T_j')| \leq \frac{\varepsilon n/2 + \sum_{r<j} |V(T_r')|}{8d},$$

and each T_j' has a unique edge joining it to $V(T) \setminus \cup_{r<j} V(T_r')$. Let T' be the tree obtained from T by omitting all the vertices of all subtrees T_j', $j < i$. If the number of vertices of T' is at most $\frac{\varepsilon n/2 + \sum_{r<i} |V(T_r')|}{8d}$, then define $T_i' = T'$ and $s = i$. Else, apply Proposition 4.2 to find a tree T_i' in T' whose size is at least $\frac{\varepsilon n/2 + \sum_{r<i} |V(T_r')|}{8d^2}$ and at most $\frac{\varepsilon n/2 + \sum_{r<i} |V(T_r')|}{8d}$, and continue. To estimate the number of steps in this process define $a_i = \varepsilon n/2 + \sum_{j \leq i} |V(T_j')|$. Observe that $a_0 = \varepsilon n/2, a_i \leq \varepsilon n/2 + |V(T)| \leq n$ and $a_i = a_{i-1} + |V(T_i')| \geq \left(1 + \frac{1}{8d^2}\right) a_{i-1}$. From this it follows that

$$\frac{2}{\varepsilon} \geq \frac{a_i}{a_0} \geq \left(1 + \frac{1}{8d^2}\right)^i,$$

and hence this process terminates after at most $10d^2 \log(2/\varepsilon)$ steps. Finally, for $1 \leq i \leq s$ define $T_i = T_{s-i+1}'$. □

4.3 Splitting vertex degrees

Lemma 4.4 *Let numbers K, δ, Δ satisfy*

$$K\Delta^2 e^{-(\delta/8K)+1} < 1.$$

Then the following holds. Let $H = (V, E)$ be a graph in which $\delta \leq d(v) \leq \Delta$ for each $v \in V$. Then H contains K pairwise disjoint sets of vertices $S_1, S_2, \ldots, S_K$ such that every vertex of H has at least $\frac{\delta}{2K}$ neighbors in each set S_i.

Proof. This is a simple consequence of the Lovász Local Lemma (c.f., e.g., [3], Chapter 5). Color the vertices of H randomly and independently by K colors. For each vertex v and color i, $1 \leq i \leq K$, let $A_{v,i}$ be the event that v has less than $\frac{\delta}{2K}$ neighbors of color i. By Chernoff's Inequality the probability of each event $A_{v,i}$ is at most $e^{-\delta/(8K)}$. In addition, each event $A_{v,i}$ is mutually independent of all events but the events $A_{u,j}$ where either $u = v$ or u and v have common neighbors in H. As there are less than $K\big(\Delta(\Delta-1)+1\big) \leq K\Delta^2$ such events, it follows by the Local Lemma that with positive probability none of the events $A_{v,i}$ holds. The desired result follows, by letting S_i denote the set of all vertices of color i. $\square$

5 Proof of Theorem 1.4

Let $G = (V, E)$ be a graph satisfying the assumptions of the theorem. Let T be a tree on at most $(1-\varepsilon)n$ vertices with maximum degree at most d. By Corollary 4.3 the tree T can be partitioned into subtrees $T_1, T_2, \ldots, T_s$ satisfying the conditions of the Corollary, where $s \leq 10d^2 \log(2/\varepsilon)$. Choose an arbitrary root for T_1. For $i > 1$, the root of T_i is the vertex incident with the unique edge that connects T_i to the union of the previous trees. Put $K = 2s/\varepsilon$. By Condition 2 in the theorem, and Lemma 4.4 there are K pairwise disjoint sets of vertices S_i of G such that every vertex of G has at least $\delta/(2K) \geq \frac{\varepsilon\delta}{40d^2\log(2/\varepsilon)}$ neighbors in each set S_i. Take the s smallest sets S_i and renumber them so that they are denoted by $S_1, S_2, \ldots, S_s$. Obviously, their total size is at most $\frac{ns}{K} = \frac{\varepsilon n}{2}$. We will not use all the other sets $S_i, i > s$ in the rest of the proof.

Let x_1 be an arbitrary vertex of G that does not lie in any of the sets S_i. Define $R = \{x_1\}$ and let U_1 denote the set of all vertices of G besides those in $\cup_{j\neq 1} S_j$. As U_1 contains S_1, every vertex in the induced subgraph $G[U_1]$ of G on U_1 has degree at least $\frac{\varepsilon\delta}{40d^2\log(2/\varepsilon)}$. Therefore, by Condition 3 in Theorem 1.4, and by Theorem 4.1 there is a copy of T_1 in $G[U_1]$ rooted at x_1. (Note that by Corollary 4.3, the size of T_1 is at most $\frac{|U_1|}{8d}$ and hence indeed one can apply here Theorem 4.1.) Moreover, we can in fact embed in U_1 the required tree T_1 together with the edges connecting it to the trees T_j with $j > 1$. Add the endpoints of these edges to the list R of planned roots for the trees T_j, and delete x_1 from R. This completes the embedding of T_1.

Assume that we have already embedded the first $i-1$ trees T_r, where each T_r has been rooted in the vertex of R specified as its root, and where in step number r the tree T_r has been embedded using no vertices of R besides its root, and no vertices of $\cup_{j\neq r} S_j$, we proceed to the i-th step, in which we are to embed the tree T_i starting from a given root $x_i \in R$. Let U_i be the set of all vertices of G that have not been used for embedding the part of T embedded so far, besides the vertices in $R - \{x_i\}$ and besides the vertices in $\cup_{j\neq i} S_j$. As before, since U_i contains S_i, every vertex in the induced subgraph $G[U_i]$ of G on U_i has degree at least $\frac{\varepsilon\delta}{40d^2\log(2/\varepsilon)}$. Therefore, by Condition 3 in Theorem 1.4, it is a $(\frac{1}{2d+2}, d+1)$-expander. By Corollary 4.3, the size of T_i is at most $|U_i|/8d$ and the number of edges connecting T_i to the trees T_j for $j > i$ is bounded by

$$s \leq 10d^2\log(2/\varepsilon) \leq \frac{\varepsilon n}{48d} \leq \frac{|U_i|}{24d}.$$

Therefore the size of T_i together with the vertices in T_j for $j > i$ that are connected to it is less than a fraction $\frac{1}{6d} \leq \frac{1}{2(2d+2)}$ of the size of U_i. Therefore, by Theorem 4.1 we can embed T_i including the edges connecting it to the trees T_j with $j > i$ in U_i, rooting it at x_i, and add the endpoints of the edges from T_i to future T_j's to R.

As this process can be carried out until we finish the embedding of T_s, the assertion of the theorem follows. □

6 Concluding remarks

- Our lower bound on the edge probability of a random graph in Theorem 1.1 seems far from being best possible, and the correct order of magnitude should probably be more similar to the case of a longest path. Hence it is likely that already when $c = O(d\log(1/\epsilon))$ the random graph $G(n, c/n)$ contains a.s. every tree on $(1-\epsilon)n$ vertices with maximum degree at most d.

- Embedding *spanning* trees of bounded degree in sparse random graphs is an intriguing question which is completely open. In case of the path, this question is very well understood (see, e.g., Chapter 8 of [6]) and it is known that for $p = O(\log n/n)$ the random graph $G(n,p)$ a.s. contains a Hamiltonian path. We believe that a more general result should be true, i.e., such a random graph should already contain a.s. every tree on n vertices with maximum degree at most d. Our methods are insufficient to attack this problem. Using Theorem 1.1 we can only prove the following much weaker result. Let T be a tree on n vertices with at least ϵn leaves, then there exists a constant $a(\epsilon, d)$ such that the random graph $G\big(n, \frac{a\log n}{n}\big)$ a.s. contains T. Here is a brief sketch of the proof: we split $G(n,p)$ into two random graphs $G(n,p_1)$ and $G(n,p_2)$, where $1-p = (1-p_1)(1-p_2)$, $p_1 = \Theta(1/n)$, $p_2 = \Theta(\log n/n)$. Let T' be the tree obtained from T by deleting its leaves. We use Theorem 1.1 to embed a copy of T' in $G(n,p_1)$. Then we expose the edges of $G(n,p_2)$ between the set of vertices V_0 of G, not occupied by a copy of T', and the rest of the graph, and embed the leaves of T in V_0 using matching-type results.

- Besides the model $G(n,p)$, another model of random graphs, drawing a lot of attention is the model of random regular graphs. A random regular graph $G_{n,D}$ is obtained by sampling uniformly at random over the set of all simple D-regular graphs on a fixed set of n vertices. By the result of Friedman, Kahn and Szemerédi [14] the second eigenvalue of $G_{n,D}$ is almost surely at most $O(\sqrt{D})$ (see [13] for a more precise result). Therefore our Theorem 1.2 immediately implies that if $D = D(d, \epsilon)$ is sufficiently large then $G_{n,D}$ a.s. contains every tree on $(1-\epsilon)n$ vertices with maximum degree d.

This result as well as the result of Theorem 1.2 are probably not optimal. We suspect that sufficiently large spectral gap (as a function of d only) already suffices to guarantee the embedding of every spanning tree of bounded degree in a graph G of order n. This is not known even for the Hamiltonian path, and the best result in this case, obtained in [17], requires spectral gap of order roughly $\log n$.

Acknowledgment. A major part of this work was carried out when the authors were visiting Microsoft Research at Redmond, WA. We would like to thank the members of the Theory Group at Microsoft Research for their hospitality and for creating a stimulating research environment.

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. Capalbo, Y. Kohayakawa, V.Rödl, A. Ruciński and E. Szemerédi, Near-optimal universal graphs for graphs with bounded degrees, *Proc.* 5^{th} *Int. Workshop on Randomization and Approximation techniques in Computer Science (RANDOM-APPROX 2001)*, Berkeley 2001, 170–180.

[3] N. Alon and J. H. Spencer, **The probabilistic method**, 2^{nd} Ed., Wiley, New York, 2000.

[4] S. N. Bhatt, F. Chung, F. T. Leighton and A. Rosenberg, Universal graphs for bounded-degree trees and planar graphs, *SIAM J. Disc. Math.* 2 (1989), 145–155.

[5] B. Bollobás, Long paths in sparse random graphs, *Combinatorica* 2 (1982), 223–228.

[6] B. Bollobás, **Random graphs**, 2^{nd} ed. Cambridge Studies in Advanced Mathematics, 73. Cambridge University Press, Cambridge, 2001.

[7] F. R. K. Chung and R. L. Graham, On graphs which contain all small trees, *J. Combin. Th. Ser. B* 24 (1978), 14–23.

[8] F. R. K. Chung and R. L. Graham, On universal graphs, *Ann. New York Acad. Sci.* 319 (1979), 136–140.

[9] F. R. K. Chung and R. L. Graham, On universal graphs for spanning trees, *Proc. London Math. Soc.* 27 (1983), 203–211.

[10] F. R. K. Chung, R. L. Graham and N. Pippenger, On graphs which contain all small trees II, *Proc. 1976 Hungarian Colloquium on Combinatorics* 1978, 213–223.

[11] W. Fernandez de la Vega, Long paths in random graphs, *Studia Sci. Math. Hungar.* 14 (1979), 335–340.

[12] W. Fernandez de la Vega, Trees in sparse random graphs, *J. Combin. Theory Ser. B* 45 (1988), 77–85.

[13] J. Friedman, On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica* 11 (1991), 331–362.

[14] J. Friedman, J. Kahn and E. Szemerédi, On the second eigenvalue in random regular graphs, *Proc. of* 21^{th} *ACM STOC* (1989), 587–598.

[15] J. Friedman and N. Pippenger, Expanding graphs contain all small trees, *Combinatorica* 7 (1987), 71–76.

[16] A. Frieze, On large matchings and cycles in sparse random graphs, *Discrete Math.* 59 (1986), 243–256.

[17] M. Krivelevich and B. Sudakov, Sparse pseudo-random graphs are Hamiltonian, *J. Graph Theory* 42 (2003), 17–33.

[18] L. Pósa, Hamiltonian circuits in random graphs, *Discrete Math.* 14 (1976), 359–364.