NOTES ON GROWING A TREE IN A GRAPH[☉]

Luc Devroye,[☿] *Vida Dujmović,*[♀] *Alan Frieze,*[♁] *Abbas Mehrabian,*[♂] *Pat Morin,*[♃] *and Bruce Reed*[♄]

Monday 26$^{\text{th}}$ June, 2017, 10:23

Abstract. We study the height of a spanning tree T of a graph G obtained by starting with a single vertex of G and repeatedly selecting, uniformly at random, an edge of G with exactly one endpoint in T and adding this edge to T.

[☉]This research was partly supported by NSERC and by NSF GRANT DMS1661063. Some of this work was carried out while Vida Dujmović, Pat Morin, and Bruce Reed were visiting IMPA in Rio de Janeiro. They thank the Institute for its hospitality.

[☿]School of Computer Science, McGill University, Montréal, Canada

[♀]Department of Computer Science and Electrical Engineering, University of Ottawa, Ottawa, Canada

[♁]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA

[♂]Department of Computer Science, University of British Columbia, Vancouver, Canada. This author was supported by an NSERC Postdoctoral Fellowship and a Simons-Berkeley Research Fellowship. Part of this work was done while he was visiting the Simons Institute for the Theory of Computing at UC Berkeley.

[♃]School of Computer Science, Carleton University, Ottawa, Canada

[♄]CNRS, Projet COATI, I3S (CNRS and UNS) UMR7271 and INRIA, Sophia Antipolis, France. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Brasil. Visiting Research Professor, ERATO Kawarabayashi Large Graph Project, Japan.

Contents

1 Introduction

Let s be a vertex of a simple connected graph G on n vertices. We build a sequence $T_1, T_2, \ldots, T_n$ of random subtrees of G as follows. The tree T_1 has a single vertex, s. For each $1 < i \leq n$, tree T_i is obtained by choosing a uniformly random edge of G with exactly one endpoint in T_{i-1}, and adding the edge to T_{i-1}. Note that T_n is a (not necessarily uniform) random spanning tree of G rooted at s, which we denote by $\mathcal{T}(G,s)$. In this paper we study the height (maximum length of a root-to-leaf path) of $\mathcal{T}(G,s)$ and give several bounds for it in terms of parameters of G.

In the special case when G is the complete graph, each tree T_i is obtained from T_{i-1} by choosing a uniformly random node of T_{i-1} and joining a new leaf to that node. This is the well studied *random recursive tree* process, and Devroye [5] and Pittel [12] have shown that the height of $T_n = \mathcal{T}(K_n, s)$ is $(e + o(1))\ln n$ with probability $1 - o(1)$.

Our results. Let $D = D(G)$ and $\Delta = \Delta(G)$ denote the diameter and maximum degree of G, respectively, and let us denote the height of a tree T by $h(T)$. An obvious lower bound for $h(\mathcal{T}(G,s))$ is $D/2$. We prove the following bounds hold with probability $1 - o_n(1)$ for any n-vertex graph G and any $s \in V(G)$. (The notation $o_k(1)$ denotes the set of functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(k) \to 0$ as $k \to \infty$.)

1. In Theorem 6 we show $h(\mathcal{T}(G,s)) \in O(\Delta(D + \log n))$. For $D \in \Omega(\log \Delta)$ this is tight: in Theorem 19 we show that for every $\Delta \geq 2$ and every $D \geq e^6 \ln \Delta$, there exist G and s with $h(\mathcal{T}(G,s)) \in \Omega(\Delta(D + \log n))$.

2. If G is d-degenerate (that is, every subgraph of G has a vertex of degree at most d), then in Theorem 8 we show $h(\mathcal{T}(G,s)) \in O(\sqrt{d\Delta}(D + \log n))$. The class of $O(1)$-degenerate graphs is quite rich and includes every minor-closed graph family. This upper bound is tight, even for planar graphs ($d = 5$), graphs of thickness t ($d = 5t$), and graphs of treewidth k ($d = k$). (The concepts of Euler genus, thickness, and treewidth are defined in Section 4).

   For $D \in \Omega(\log \Delta)$ and planar graphs (which are 5-degenerate) this is tight: in Theorem 23 we show for any $\Delta > 2$ and $D > 10^6 \ln \Delta$ there exists a planar graph G and vertex s with $h(\mathcal{T}(G,s)) \in \Omega(\sqrt{\Delta}(D + \log n))$.

   Also, for $D \in \Omega(\log \Delta)$ and $d \leq \Delta$ this is tight: in Theorem 24 we show for any $\Delta > 1$, $D > 10^6 \ln \Delta$ and $d \leq \Delta$ there exist a d-degenerate graph G and vertex s with $h(\mathcal{T}(G,s)) \in \Omega(\sqrt{d\Delta}(D + \log n))$

3. If G has Euler genus less than $C\sqrt{\Delta}D/\log \Delta$, then $h(\mathcal{T}(G,s)) \in O(\sqrt{\Delta}(D + \log n))$ (see Theorem 10) . For $D \in \Omega(\log \Delta)$ and zero Euler genus this is tight: in Theorem 23 we show for any $\Delta > 2$ and $D > 10^6 \ln \Delta$ there exist a planar graph G and vertex s with $h(\mathcal{T}(G,s)) \in \Omega(\sqrt{\Delta}(D + \log n))$

4. For any $d, k \geq 1$, if G is the d-dimensional grid of side-length k (which has $n = (k+1)^d$ vertices), we have $h(\mathcal{T}(G,s)) \in O(dk + d^{5/3}\ln(k+1))$. If $k = 2$ or $k/\ln(k+1) = \Omega(d^{2/3})$, we have $h(\mathcal{T}(G,s)) \in \Theta(D) = \Theta(dk)$ (see Theorem 13 and Corollary 15).

5. If G has edge-expansion factor[1] (i.e., Cheeger constant) Φ, then $h(\mathcal{T}(G,s)) \in O(\Phi^{-1}\Delta \log n)$ (see Theorem 11). This implies, for example, that $h(\mathcal{T}(G,s)) \in O(\log n)$ if G is the complete graph or if G is a random Δ-regular graph (since a random Δ-regular graph has $\Phi \in \Omega(\Delta)$, see [2]).

Our main tool for proving upper bounds, Lemma 4, bounds $h(\mathcal{T}(G,s))$ in terms of the first-passage percolation cover time and the number of paths of a given length starting at s. To prove our results using this tool, we prove several new bounds on first-passage percolation cover times as well as the number of simple paths in various families of graphs, which are of independent interest.

Our results on first-passage percolation cover time. Suppose independent exponential(1) random variables $\{\tau_e\}$ are assigned to edges of G. Let $\Gamma(s,v)$ denote the set of all (s,v)-paths in the graph. Then the *first-passage percolation cover time* is defined as

$$\tau(G,s) = \max_{v \in V(G)} \min_{\gamma \in \Gamma(s,v)} \sum_{e \in \gamma} \tau(e)$$

In Lemma 5 we show a general upper bound of $O(\ln n + D)$ for $\tau(G,s)$. (This and the following results hold with probability $1 - o_n(1)$.)

In the special case when G is the d-dimensional grid with side length k (and diameter dk), we prove the improved bound $\tau(G,s) = O(k)$. The special case of $k = 1$, namely the d-cube graph, was studied by Fill and Pemantle [7], who showed $1.414 \le \tau(G,s) \le 14.041$. The upper bound was subsequently improved to 1.694 by Bollobás and Kohayakawa [3] and recently to 1.575 by Martinsson [9].

The remainder of the paper is organized as follows: Section 2 presents some preliminaries and useful facts about sums of independent random variables, In Section 3 we present the connection with first-passage percolation and prove a general upper bound. Section 4–Section 7 present our upper bounds on $h(T)$. Section 8 and Section 9 present families of graphs with matching lower bounds.

We use the following notational conventions: $\log x$ denotes the binary logarithm of x and $\ln x$ denotes the natural logarithm of x. Every graph, G, that we consider is finite, simple, undirected and connected, and n denotes its number of vertices.

2 Preliminaries

Recall that an exponential(λ) random variable, X, has a distribution defined by

$$\Pr\{X > x\} = e^{-\lambda x} \ , \ \ x \ge 0 \ ,$$

and mean $\mathrm{E}[X] = \int_0^\infty \Pr\{X > x\}\,\mathrm{d}x = 1/\lambda$. We make extensive use of the fact that exponential random variables are *memoryless*:

$$\Pr\{X > t + x \mid X > t\} = \frac{\Pr\{X > t + x\}}{\Pr\{X > t\}} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = \Pr\{X > x\} \ .$$

[1] The edge expansion factor and related quantities are defined in Section 6.

We will also often take the minimum of δ independent exponential(λ) random variables and use the fact that this is distributed like an exponential($\lambda\delta$) random variable:

$$\Pr\{\min\{X_1,\ldots,X_\delta\} > x\} = (\Pr\{X_1 > x\})^\delta = e^{-\delta\lambda x} \sim \text{exponential}(\lambda\delta) \ .$$

We will make use of two concentration inequalities for sums of exponential random variables, both of which can be obtained using Chernoff's bounding method (see, e.g., [8, Theorem 5.1]). If $Z_1,\ldots,Z_k$ are independent exponential(λ) random variables (so that they each have mean $\mu = 1/\lambda$), then for all $d > 1$,

$$\Pr\left\{\sum_{i=1}^{k} Z_i \le \mu k/d\right\} \le \exp(-k(\ln d - 1 + 1/d)) \le \left(\frac{e}{d}\right)^k \tag{1}$$

and for all $t > 1$,

$$\Pr\left\{\sum_{i=1}^{k} Z_i \ge \mu k t\right\} \le \exp(k - kt/2) \ . \tag{2}$$

The sum of k independent exponential(λ) random variables is called an Erlang(k,λ) random variable.

For positive integers a, b and c, we define two random variables $Y_{a,b}$ and $Y_{a,b,c}$ as follows. Consider a tree in which the root has a children, each of the root's children have b children, and each of root's grandchildren have c children. This is a tree of height 3 with $1 + a + ab + abc$ nodes. Put an independent exponential(1) weight on each edge. Then $Y_{a,b}$ is defined as the minimum weight of a path from the root to a node at level 2, and $Y_{a,b,c}$ is defined as the minimum weight of a root-to-leaf path.

The following auxiliary lemmas are proved in the appendix.

Lemma 1. *Let $X_1,\ldots,X_m$ be i.i.d. distributed as $Y_{a,b}$ for some a,b. Then we have* $\mathrm{E}X_1 = O(1/a + 1/\sqrt{ab})$ *and moreover,*

$$\Pr\left\{\sum_{i=1}^{m} X_i \ge 3m(64/a + 1024/\sqrt{ab})\right\} \le \exp(-m/9)\,.$$

Let $\Gamma(t) := \int_0^\infty x^{t-1}\mathrm{e}^{-x}\mathrm{d}x$. If t is a positive integer, then it is well known that $\Gamma(t) = (t-1)!$.

Lemma 2. *For any positive integer p we have*

$$\mathrm{E}Y_{a,b,c}^p \le p(a/64)^{-p}\Gamma(p) + p(ab/1024)^{-p/2}\Gamma(p/2)/2 + p(abc/16384)^{-p/3}\Gamma(p/3)/3$$

and, in particular, $\mathrm{E}Y_{a,b,c} \le 64/a + 1024/\sqrt{ab} + 16384/\sqrt[3]{abc}$.

Finally, we will also use the following version of Bernstein's inequality (see Theorem 2.10 and Corollary 2.11 in [4]).

Theorem 3 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_m$ be non-negative independent random variables for which there exist v, c satisfying*

$$\sum_{i=1}^{m} \mathrm{E}[X_i^p] \leq v p! c^{p-2}/2$$

for all positive integers $p \geq 2$. Then for any $t > 0$ we have

$$\Pr\left\{\sum_{i=1}^{m}(X_i - \mathrm{E}X_i) \geq ct + \sqrt{2vt}\right\} \leq e^{-t},$$

and

$$\Pr\left\{\sum_{i=1}^{m}(X_i - \mathrm{E}X_i) \geq t\right\} \leq \exp\left(-\frac{t^2}{2v + 2ct}\right),$$

3 Connection with first-passage percolation and a generic upper bound

In this section, we establish the connection with first-passage percolation, and prove an upper bound for $\tau(G,s)$ in general graphs, which results in an upper bound for $h(\mathcal{T}(G,s))$. This connection will be used in subsequent sections to provide tighter bounds for $h(\mathcal{T}(G,s))$ in several graph classes.

Recall the generation process for $\mathcal{T}(G,s)$: we start with a tree containing only vertex s initially; in each round, we choose an edge uniformly at random among edges with exactly one endpoint in the existing tree, and add it to the existing tree.

We may view this as an *infection process*: at round 0 only vertex s is infected. In each round, suppose the set of infected vertices is S. We choose a uniformly random edge between S and its compliment, and let the disease spread along that edge, hence increasing the number of infected vertices by one.

Now consider the following continuous time view of this infection process, which is known as Richardson's model [6] or first-passage percolation [1]. At time 0 we infect vertex s. For each edge uv, whenever one of u and v gets infected, we put an exponential(1) timer on edge uv. When the timer rings, the disease spreads along that edge and both u and v get infected (it might be the case that both u and v are already infected by that time). Suppose at some moment in this process, the subset S of vertices are infected. Then, by memorylessness of the exponential distribution, the disease is equally like to spread along any of the edges existing between S and its complement. Therefore, the tree along which the disease spreads has the same distribution as $\mathcal{T}(G,s)$.

This viewpoint induces weights on the edges: to each edge e we assign weight $\tau(e)$, which is the ringing time for the timer on this edge. Note that the weights are i.i.d. exponential (1) random variables. The weight of a path P, denoted $\tau(P)$, is simply the sum of weights of its edges. The *first-passage percolation hitting time* (or simply, the *hitting time*) for v is the weight of the lightest path from s to v:

$$\tau(G,s,v) = \min_{\gamma \in \Gamma(s,v)} \tau(\gamma) \ .$$

The *first-passage percolation cover time* (or simply, the *cover time*) is the first time that all vertices are infected, which can be written as

$$\tau(G,s) = \max_{v \in V(G)} \tau(G,s,v) \ .$$

Note that this is also the maximum *weight* of a root-to-leaf path in the infection tree $\mathcal{T}(G,s)$, which we will use to bound the height of $\mathcal{T}(G,s)$, the maximum *length* of a path in (the unweighted version of) $\mathcal{T}(G,s)$ (in general, the longest path and the heaviest path may be different).

For a positive integer L and a vertex s of graph G, let $\Pi(G,s,L)$ denote the number of simple paths of length L in G that start from s. We now prove a lemma that upper bounds $h(\mathcal{T}(G,s))$ in terms of $\tau(G,s)$ and $\Pi(G,s,L)$.

Lemma 4. *Let $s \in V(G)$, $0 \le p < 1 \le a$, $c > 0$, and $L = \lceil ceaK \rceil$ be such that $\Pr\{\tau(G,s) > K\} \le p$ and $\Pi(G,s,L) \le a^L$. Then $h(\mathcal{T}(G,s)) \le L$ with probability at least $1 - p - c^{-L}$.*

Proof. Let $T = \mathcal{T}(G,s)$. If $h(T) > L$, then at least one of the following two events occurred:

1. T contains a root-to-leaf path of weight greater than K.
2. G contains a path starting at s of length L whose weight is less than K.

By assumption, the probability of the first event is at most p. The weight of a single path of length L is the sum of L exponential(1) random variables so, by (1) and the union bound over all a^L paths, the probability of the second event is at most

$$a^L \left(\frac{eK}{L}\right)^L \le c^{-L} \ . \qquad \square$$

In light of Lemma 4, we can upper bound $h(\mathcal{T}(G,s))$ if we have upper bounds on the cover time and on the number of paths of length L originating at s. An obvious upper bound for the latter is Δ^L. The following lemma gives a general upper bound for the former, which results in a general upper bound for $h(\mathcal{T}(G,s))$. In the following sections we obtain better bounds for these two quantities in special graph classes, resulting in sharper bounds on $h(\mathcal{T}(G,s))$.

Lemma 5. *For any $s \in V(G)$, we have $\tau(G,s) \le 4 \ln n + 2D$ with probability at least $1 - 1/n$.*

Proof. For each vertex $v \in V(G)$, we show the probability that it is not infected by time $4 \ln n + 2D$ is at most n^{-2}, and then apply the union bound over all vertices. Let P be a shortest (s,v)-path in G. Let $k \le D$ denote the length of P, so $\tau(P) \sim \mathrm{Erlang}(k,1)$. Note that for any t, $\tau(P) \le t$ implies v is infected by time t. Thus, using (2), the probability that v is not infected by time $4 \ln n + 2D$ is bounded by

$$\Pr\{\tau(P) > 4\ln n + 2D\} = \Pr\{\mathrm{Erlang}(k,1) > 4 \ln n + 2D\} \le \exp(k - 2\ln n - D) \le n^{-2} \ . \qquad \square$$

We immediately get a general upper bound for $h(\mathcal{T}(G,s))$.

Theorem 6. *Let G be an n-vertex graph with diameter D and maximum degree $\Delta > 1$, and let s be an arbitrary vertex. Then, with probability at least $1 - O(1/n)$ we have*

$$\frac{D}{2} \le h(\mathcal{T}(G,s)) \le 2e\Delta(4\ln n + 2D) \le (4e\Delta + 8e\Delta\ln\Delta)D + 16e\Delta\,.$$

Note that this gives an asymptotically tight bound of $h(\mathcal{T}(G,s)) = \Theta(D)$ for graphs with bounded maximum degree.

Proof. The first inequality is trivial. The second inequality is an application of Lemma 4 with $a = \Delta$, $p = 1/n$, $K = 4\ln n + 2D$ and $c = 2$, using the bound of Lemma 5 for the cover time. The last inequality follows from the crude bound $\Delta^D \ge n/3$, which holds for any n-vertex graph with maximum degree Δ and diameter D. □

4 An upper bound in terms of graph degeneracy

Recall that a graph is *d-degenerate* if each of its subgraphs has a vertex of degree at most d. The following lemma shows that, for large L, d-degenerate graphs have considerably less than Δ^L walks of length L.

Lemma 7. *Let G be an n-vertex d-degenerate graph with maximum degree Δ. Then the number of walks in G of length L is bounded by $2n2^L(d\Delta)^{L/2}$.*

Proof. Enumerate the vertices of G as $v_1,\dots,v_n$ so that v_i has at most d edges in the subgraph induced by $v_i,\dots,v_n$ (this ordering may be obtained by repeatedly removing a vertex of degree at most d).

We give a way to encode the walks in a one-to-one way, and then bound the total number of possible generated codes. Let $W = v_{i_0},\dots,v_{i_L}$ be a walk of length L in G and let $k = k(W)$ denote the number of indices $\ell \in \{1,\dots,L\}$ such that $i_{\ell-1} < i_\ell$. If $k \ge L/2$ then we say that W is *easy*; note that at least one of W and its reverse is easy, hence the total number of L-walks is at most twice the number of easy L-walks. We encode an easy walk W in the following way:

1. We first specify the starting vertex v_{i_0}. There are n ways to do this.
2. Next we specify whether $i_{\ell-1} < i_\ell$ for each $\ell \in \{1,\dots,L\}$. There are at most 2^L ways to do this.
3. Next, we specify each edge of W. For each $\ell \in \{1,\dots,L-1\}$, if $i_\ell < i_{\ell+1}$, then there are at most d ways to do this, otherwise there are at most Δ ways to do this. Therefore, the total number of ways to specify all edges of the walk is at most

   $$d^k\Delta^{L-k} \le (d\Delta)^{L/2}\ ,$$

   since $d \le \Delta$ and $k \ge L/2$.

Therefore, the number of easy L-walks is bounded by $n2^L(d\Delta)^{L/2}$, as required. □

Theorem 8. *Let G be an n-vertex d-degenerate graph with diameter D and maximum degree Δ, and let s be an arbitrary vertex. Then, with probability at least $1 - O(1/n)$ we have $h(\mathcal{T}(G,s)) \le 8e\sqrt{d\Delta}(2D + 4\ln n)$.*

Proof. Let $c = 2$, $K = 4\ln n + 2D$, $p = 1/n$, $a = 4\sqrt{d\Delta}$, and $L = \lceil ceaK \rceil > 8\ln n$. Lemma 5 guarantees $\tau(G,s) \le 4\ln n + 2D$ with probability at least $1 - 1/n$, and Lemma 7 guarantees $\Pi(G,s,L) \le 2n2^L(d\Delta)^{L/2} \le a^L$. Applying Lemma 4 completes the proof. □

Note that Theorem 8 actually implies Theorem 6 up to constant factors, since all graphs of maximum degree Δ are Δ-degenerate, so $\sqrt{d\Delta} \le \Delta$ in all cases. However, Theorem 8 provides sharper bounds for many important graph classes:

- Planar graphs are 5-degenerate. (This is a consequence of Euler's formula and the fact that planarity is preserved under taking subgraphs).
- The *thickness* of a graph is the minimum number of planar graphs into which the edges of G can be partitioned. Graphs of thickness t are $5t$-degenerate. (This follows from definitions and the 5-degeneracy of each individual planar graph in the partition.)
- The *Euler genus* of a graph is the minimum Euler genus of a surface on which the graph can be drawn without crossing edges. Graphs of Euler genus g are $O(\sqrt{g})$-degenerate.[2]
- A *tree decomposition* of a graph G is a tree T' whose vertex set B is a collection of subsets of $V(G)$ called *bags* with the following properties:

  1. For each edge vw of G, there is at least one bag $b \in B$ with $\{v,w\} \subseteq B$.
  2. For each a vertex v of G, the subgraph of T' induced by the set of bags that contain v is connected.

  The *width* of a tree-decomposition is one less than the size of its largest bag. The *treewidth* of G is the minimum width of any tree decomposition of G. Graphs of treewidth k are k-degenerate. (This is a consequence of the fact that k-trees are edge-maximal graphs of treewidth k.)

Therefore, Theorem 8 implies that, when the relevant parameter, g, t or k, is bounded, $h(T) \in O(\sqrt{\Delta}(D + \log n))$ with high probability.

5 An upper bound in terms of Euler genus

Since graphs of Euler genus g are $O(\sqrt{g})$-degenerate, Theorem 8 implies that if G has Euler genus g, then $h(\mathcal{T}(G,s)) \in O(g^{1/4}\Delta^{1/2}(D + \log n))$. In this section we show that the dependence on the genus g can be eliminated when the diameter is large compared to the genus.

[2] This follows from the facts in every n-vertex Euler-genus g graph, $n \in \Omega(\sqrt{g})$ and there exists a vertex of degree at most $6 + O(g/n)$. (See, e.g., [13, Lemma 7 and Theorem 2].)

We begin with a upper-bound on path counts that is better (for graphs of small genus) than Lemma 7.

Lemma 9. *Let G be a simple n-vertex graph of Euler genus g, diameter D, and maximum degree $\Delta \geq 6$. Then the number of simple paths in G of length L is at most $2n2^L 6^{L/2-3g}\Delta^{L/2+3g}$.*

Proof. The following proof makes use of some basic notions related to graphs on surfaces; see Mohar and Thomassen [11] for basic definitions and results. Since G has Euler genus g, it has a 2-cell embedding in a surface of Euler genus g. Euler's formula then states that

$$m = n + f - 2 + g \quad , \tag{3}$$

where n and m are the numbers of vertices and edges of G and f is the number of faces in the embedding of G. Every edge is on the boundary of at most 2 faces of the embedding and, since G is simple, every face is bounded by at least 3 edges. Therefore, $f \leq 2m/3$, so (3) implies

$$m \leq 3n - 6 + 3g \quad .$$

Therefore, the average degree of an n-vertex Euler genus g graph is at most $6+(6g-12)/n$. In particular, if $n \geq 6g$, then g has average degree less than 7, so G contains a vertex of degree at most 6.

When we remove a vertex from G we obtain a graph whose Euler genus is not more than that of G. Therefore, by repeatedly removing a degree 6 vertex, we can order the vertices of G as $v_1,\ldots,v_n$ so that, for each $i \in \{1,\ldots,n-6g\}$, v_i has at most 6 neighbours among $v_{i+1},\ldots,v_n$. We call $v_{n-6g+1},\ldots,v_n$ *annoying vertices* and edges between them are *annoying edges*.

Let $P = v_{i_0},\ldots,v_{i_L}$ be a path of length L in G. For each $i \in \{1,\ldots,L\}$, the edge $v_{i_{\ell-1}}v_{i_\ell}$ is called *bad* if it is annoying or if $i_{\ell-1} > i_\ell$; otherwise it is called good. Let k denote the number of good edges in P. Say P is good if $k \geq L/2-3g$. Note that the number of annoying edges of P is bounded by $6g-1$, hence at least one of P and its reverse is good. We bound the number of good L-paths; the total number of L-paths is at most twice this bound. We encode a good L-path P as follows:

1. We first specify the starting vertex v_{i_0}. There are n ways to do this.
2. Next we specify whether each edge of P is good or bad. There are 2^L ways to do this.
3. Next, we specify each edge of P. For each good edge, there are at most 6 ways to do this. For each bad edge there are at most Δ ways to do this. Therefore, the total number of ways to specify the edges of P is at most

$$6^k\Delta^{L-k} \leq 6^{L/2-3g}\Delta^{L/2+3g} \quad ,$$

since $k \geq L/2-3g$ and $\Delta \geq 6$.

Therefore, the number of good L-paths is at most $n2^L 6^{L/2-3g}\Delta^{L/2+3g}$, as required. □

Theorem 10. *Let G be an n-vertex Euler-genus g graph with diameter D, maximum degree Δ and let $s \in V(G)$ be an arbitrary vertex. If $g \ln \Delta \le 36\sqrt{\Delta}(D + \ln n)$ then, with probability at least $1 - O(1/n)$, $h(\mathcal{T}(G,s)) \le 107\sqrt{\Delta}(2D + 4\ln n)$.*

Proof. The conclusion follows from Theorem 6 for $\Delta \le 6$, so we will assume $\Delta > 6$. Let $c = 2$, $K = 4\ln n + 2D$, $p = 1/n$, $a = 8\sqrt{6\Delta}$, and $L = \lceil ceaK \rceil > 8\ln n$. Lemma 5 guarantees $\tau(G,s) \le 4\ln n + 2D$ with probability at least $1 - 1/n$, and Lemma 7 guarantees

$$\Pi(G,s,L) \le 2n \times 2^L \times (6\Delta)^{L/2} \times \Delta^{3g} \le (2\times 2\times\sqrt{6\Delta})^L \exp\left(108\sqrt{\Delta}(D + \ln n)\right) \le (2\times 2\times\sqrt{6\Delta}\times 2)^L = a^L .$$

Applying Lemma 4 completes the proof. $\square$

6 An upper bound for edge expanders

All of the preceding upper bounds on $h(T)$ have a (linear or rootish) dependence on Δ, the maximum degree of a vertex in G. This seems somewhat counterintuitive, since high degree vertices in G should produce high degree vertices in T and therefore decrease $h(T)$. In this section we show that indeed large edge expansion (also called isoperimetric number or Cheeger constant) results in low-height trees.

For an n-vertex graph G and a subset $A \subseteq V(G)$, define $e(A) = |\{vw \in E(G) : v \in A, w \notin A\}|$, and for any $k \in \{1,\dots,n-1\}$, define

$$e_k(G) = \min\{e(A) : A \subseteq V(G), |A| = k\} \ .$$

Observe that $e_k(G)$ is symmetric in the sense that $e_k(G) = e_{n-k}(G)$. We define the *edge expansion* of G is

$$\Phi(G) = \min\{e_k(G)/k : k \in \{1,\dots,\lfloor n/2 \rfloor\}\}$$

We will express the height of T in terms of the *total inverse perimeter size* Ψ, which is closely related to the edge expansion:

$$\Psi(G) = \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{e_k(G)} \le \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k\Phi(G)} = \frac{\ln n + O(1)}{\Phi(G)} \ .$$

Theorem 11. *Let G be an n-vertex graph with with maximum degree Δ, edge-expansion Φ, total inverse perimeter size Ψ, and let s be an arbitrary vertex. Then, with probability at least $1 - \exp(-\Omega(\Psi\Delta))$ we have $h(\mathcal{T}(G,s)) \in O(\Psi\Delta) \subseteq O(\Phi^{-1}\Delta \log n)$.*

Before proving Theorem 11, we consider the example of the complete graph $G = K_n$. In this graph, the minimum degree is $n-1$, so all preceding theorems (at best) imply an upper bound of $O(n)$ on $h(\mathcal{T}(K_n,s))$. However, $e_k(K_n) = k(n-k)$, so $\Phi(K_n) = \lceil n/2 \rceil$, and $\Psi(K_n) = O(\log n/n)$. Then Theorem 11 implies that $h(\mathcal{T}(K_n,s)) \in O(\log n)$ with high probability. This upper bound is of the right order of magnitude, since it matches the (tight) results of Devroye and Pittel for the height of the random recursive tree [5, 12].

Proof. Fix some path $P = (s = v_0), v_1, \ldots, v_L$ in G and suppose that P appears as a path in T. Then there are times $1 \le k_1 < \cdots < k_L < n$ such that for each $i \in \{1, \ldots, L\}$, v_i joins T when T has size k_i. For a fixed P and fixed $1 \le k_1 < \ldots < k_L < n$, the probability that this happens is at most

$$\prod_{i=1}^{L} \frac{1}{e_{k_i}(G)},$$

and the probability that P appears in T (without fixing $k_1, \ldots, k_L$) is at most

$$\sum_{1 \le k_1 < \cdots < k_L < n} \left(\prod_{i=1}^{L} \frac{1}{e_{k_i}(G)} \right) < \frac{1}{L!} \left(\sum_{(k_1, \ldots, k_L) \in \{1, \ldots, n-1\}^L} \left(\prod_{i=1}^{L} \frac{1}{e_{k_i}(G)} \right) \right) = \frac{1}{L!} \left(\sum_{k=1}^{n-1} \frac{1}{e_k(G)} \right)^L \le \frac{(2\Psi)^L}{L!}$$

Finally, since G contains at most Δ^L paths of length L,

$$\Pr\{h(\mathcal{T}(G, s)) \ge L\} \le \Delta^L \times \frac{(2\Psi)^L}{L!} \le \left(\frac{2e\Psi\Delta}{L} \right)^L \le \left(\frac{1}{2} \right)^L ,$$

for $L \ge 4e\Psi\Delta$. □

Observe that the last step in the proof of Theorem 11 is to use the union bound over all paths of length L. If we have a better upper-bound than Δ^L on the number of such paths, then we obtain a better upper bound on $h(T)$. Lemma 7 gives a better upper bound for d-degenerate graphs, using which we immediately obtain the following corollary.

Corollary 12. *Let G be an n-vertex d-degenerate graph with diameter D and maximum degree Δ, total inverse perimeter size Ψ, and let s be an arbitrary vertex. Then, with probability at least $1 - O(1/n)$, $h(\mathcal{T}(G, s)) \in O(\Psi\sqrt{d\Delta} + \log n) \in O(\log n(1 + \sqrt{d\Delta}/\Phi))$.*

Proof. As in the proof of Theorem 11, and using the upper bound $2n2^L(d\Delta)^{L/2}$ for the number of paths of length L, given by Lemma 7, we have

$$\Pr\{h(\mathcal{T}(G, s)) \ge L\} \le 2n2^L(d\Delta)^{L/2} \times \frac{(2\Psi)^L}{L!} \le 2n\left(4e\Psi\sqrt{d\Delta}/L\right)^L \le \left(8e\Psi\sqrt{d\Delta}n^{1/L}/L\right)^L ,$$

which is smaller than $1/n$ for $L \ge 8e^3\Psi\sqrt{d\Delta} + \ln n$, as required. □

7 Upper bounds for high dimensional grids and hypercubes

The *d-cube* is the graph having vertex set $\{0, 1\}^d$ in which two vertices are adjacent if and only if they differ in exactly one coordinate. Every vertex in the d-cube has degree d and the d-cube has diameter d. The d-cube is an interesting example in which the path count is high, but this is counteracted by a low first-passage percolation time.

Theorem 13. *Let $n = 2^d$, let G be the d-cube and let $s \in V(G)$ be arbitrary. Then, with probability at least $1 - o_n(1)$, $h(\mathcal{T}(G, s)) \in \Theta(d)$.*

Proof. Fill and Pemantle [7] showed that the first-passage percolation cover time for the d-cube is at most 14.05 with probability $1-o_n(1)$. Every vertex of the hypercube has degree d, so the number of paths of length L starting at s is less than d^L. The result then follows by applying Lemma 4 with $p = o_n(1)$, $c = 2$, $K = 14.05$, and $a = d$. $\square$

A natural generalization of the d-cube is the (d,k)-*grid*, which has vertex set $\{0,\dots,k\}^d$ and has an edge between two vertices if and only if the (Euclidean or ℓ_1) distance between them is 1. The (d,k)-grid has diameter $D = dk$ and maximum degree $\Delta = 2d$.

Note that in the case $k = 1$, the $(d,1)$-grid is the d-cube, for which Theorem 13 gives the optimal bound and this bound can be extended to $k \in O(1)$. Theorem 6 gives an upper bound of $O(d^2k)$ on $h(\mathcal{T}(G,s))$, which is optimal for $d \in O(1)$. The rest of this section is devoted to proving the following result on the first-passage-percolation cover time of grids, which gives a bound on the height of $\mathcal{T}(G,s)$.

Theorem 14. *Let G be the (d,k)-grid and $n = (k+1)^d$. Then, for any vertex $s \in V(G)$, we have that $\tau(G,s) = O(k)$ with probability $1-o_n(1)$.*

Applying Lemma 4 gives the following corollary of Theorem 14.

Corollary 15. *Let G be the (d,k)-grid and $n = (k+1)^d$. For any vertex $s \in V(G)$ we have that with probability $1-o_n(1)$, $h(\mathcal{T}(G,s)) = \Theta(dk)$*

The proof of Theorem 14 is broken up into two parts, one which holds when k is large with respect to d and one which holds when k is not too large. Together these two results cover all ranges of k and d.

7.1 Large side length

The following lemma allows us to prove Theorem 14 for the case whe k is relatively large.

Lemma 16. *Let G be the (d,k)-grid. There exists a constant C such that for any vertex $s \in V(G)$ we have*

$$\Pr\{\tau(G,s) \ge (4+C)k + Cd^{2/3}\ln(k+1)\} \le 3\exp(-d\ln(k+1)) = 3/n\,;$$

in particular, if $k/\ln(k+1) \ge d^{2/3}$, then with probability $1-O(1/n)$ we have $\tau(G,s) = O(k)$.

For the rest of this section, G is a fixed (d,k)-grid. Lemma 16 follows from the following lemma by choosing $\Delta = 4k + Cd^{2/3}\ln(k+1)$.

Lemma 17. *There exists an absolute positive constant C such that for any two vertices s and v and any $\Delta \ge 4k$ we have*

$$\Pr\{\tau(G,s,v) > Ck + \Delta\} \le 3\exp\left(-\Delta d^{1/3}/C\right) \tag{4}$$

Proof. We identify the vertices of the graph with points in $\mathbb{Z}^d$, and without loss of generality, we may assume that s is the origin. For a grid vertex $u = (u_1,\dots,u_d)$, its ℓ_1 norm $\|u\|_1$, or

ℓ_1 distance to the origin, is defined as $\sum_{i=1}^d u_i$, and its Hamming norm $\|u\|_H$, or Hamming distance from the origin, is its number of non-zero entries.

We give an algorithm to find a path from v to the origin. This is an iterative algorithm, starting from v, in each step gets closer (in ℓ_1 distance) to the origin. Suppose the algorithm has find a path from v to u so far, and then wants to find the next part of the path. Suppose $h = \|u\|_H$ and $\ell = \|u\|_1$. The algorithm does one of the following, depending on the value of h:

1. If $h > d/2$, then the algorithm chooses the lightest edge that brings it to a vertex closer to the origin. Let X_ℓ denotes the weight of this edge. Note that X_ℓ is an exponential(h) random variable and is stochastically dominated by an exponential($d/2$) random variable.

2. Otherwise, we have $h \le d/2$. Define $L_i := \{x \in V(G) : \|x\|_1 = i\}$. The algorithm chooses the lightest path $uxyz$ among all paths with $x \in L_{\ell+1}$, $y \in L_\ell \setminus \{u\}$, and $z \in L_{\ell-1}$. Note that z is one step closer to the origin (in ℓ_1 distance). Let Z_ℓ denote the weight of the chosen path $uxyz$. Then observe that Z_ℓ is a $Y_{a,b,c}$ random variable for some $a \ge d-h \ge d/2$, $b \ge \max\{h-1,1\}$ and $c \ge \max\{h-1,1\}$. Also note that we have $h \ge \ell/k$, so Z_ℓ is stochastically dominated by a $Y_{d/2,\max\{\ell/2k,1\},\max\{\ell/2k,1\}}$ random variable.

The weight of the path generated by L steps of the preceding algorithm is stochastically dominated by

$$\mathcal{S} := \sum_{\ell=1}^{L} X_\ell + \sum_{\ell=1}^{L} Z_\ell\,,$$

where X_ℓ is an exponential($d/2$) random variable, and Z_ℓ is a $Y_{d/2,\max\{\ell/2k,1\},\max\{\ell/2k,1\}}$ random variable. Moreover, since Z_ℓ only depends on edges between $L_{\ell-1}$, L_ℓ and $L_{\ell+1}$, each of the families $\{X_\ell\}$, $\{Z_{2i-1}\}$, and $\{Z_{2i}\}$ is mutually independent.

We first bound the expected value of $\mathcal{S}$, and then use Bernstein's inequality to prove with high probability it does not exceed its expected value by much. By Lemma 2, for some absolute constant C we have

$$\begin{aligned}
\mathrm{E}\mathcal{S} &\le 2L/d + C\sum_{\ell=1}^{L}\left(1/d + \sqrt{k/d\ell} + \sqrt[3]{k^2/d\ell^2}\right)\\
&\le (C+2)L/d + C\sqrt{k/d}\int_0^L x^{-1/2}dx + C\sqrt[3]{k^2/d}\int_0^L x^{-2/3}dx\\
&= (C+2)L/d + 2C\sqrt{kL/d} + 3C\sqrt[3]{Lk^2/d} \le (6C+2)k\,,
\end{aligned}$$

where in the last line we used $L \le kd$.

For any integer $p > 1$, we have $p\Gamma(p/2)/2 \le p!$ and $p\Gamma(p/3)/3 \le p!$, so Lemma 2 gives that for some absolute constant C we have

$$\begin{aligned}
\mathrm{E}Z_\ell^p &\le p(d/C)^{-p}\Gamma(p) + p(d\max\{\ell/k,1\}/C)^{-p/2}\Gamma(p/2)/2 + p(d\max\{\ell/k,1\}^2/C)^{-p/3}\Gamma(p/3)/3\\
&\le p!(C/d)^p + p!(C/d\max\{\ell/k,1\})^{p/2} + p!(C/d\max\{\ell/k,1\}^2)^{p/3}
\end{aligned}$$

Hence, we have

$$\sum_{\ell=1}^{L} \mathrm{E}Z_\ell^p/p! \le L(C/d)^p + \sum_{\ell=1}^{k}\Big((C/d)^{p/2} + (C/d)^{p/3}\Big) + \sum_{\ell=k+1}^{L}\Big((kC/d\ell)^{p/2} + (k^2C/d\ell^2)^{p/3}\Big)$$

$$\le L(C/d)^p + k(C/d)^{p/2} + k(C/d)^{p/3} + (kC/d)^{p/2}\sum_{\ell=k+1}^{L}\ell^{-p/2} + (k^2C/d)^{p/3}\sum_{\ell=k+1}^{L}\ell^{-2p/3}$$

Note that

$$k^{p/2}\sum_{\ell=k+1}^{L}\ell^{-p/2} \le k^{p/2}\int_{\ell=k}^{L}\ell^{-p/2} \le \begin{cases} k\ln(L/k) \le 2k\ln d & \text{if } p=2 \\ 2k \le 2k\ln d & \text{if } p>2 \end{cases}$$

and $\sum_{\ell=k+1}^{L}\ell^{-2p/3} \le \int_k^\infty x^{-2p/3}dx \le k^{1-2p/3}$, so

$$\sum_{\ell=1}^{L}\mathrm{E}Z_\ell^p/p! \le L(C/d)^p + k(C/d)^{p/2} + k(C/d)^{p/3} + 2k(C/d)^{p/2}\ln d + k(C/d)^{p/3} \le vc^{p-2}/2\,,$$

for $v = C'kd^{-2/3}/2$ and $c = C'd^{-1/3}/2$, for a suitably large constant C'. Since $\{Z_{2i-1}\}$ are independent, using Bernstein's inequality we have, for any $t > 1$,

$$\Pr\left\{\sum_{i=1}^{\ell/2}(Z_{2i-1} - \mathrm{E}Z_{2i-1}) \ge tk\right\} \le \exp\left(-\frac{t^2k^2}{C'kd^{-2/3} + C'tkd^{-1/3}}\right) \le \exp\left(-\frac{t^2k^2}{2C'tkd^{-1/3}}\right)$$

similarly,

$$\Pr\left\{\sum_{i=1}^{\ell/2}(Z_{2i} - \mathrm{E}Z_{2i}) \ge tk\right\} \le \exp\left(-\frac{t^2k^2}{2C'tkd^{-1/3}}\right).$$

For the variables X_ℓ, using (2) we have

$$\Pr\left\{\sum_{i=1}^{\ell}(X_i - \mathrm{E}X_i) \ge tk\right\} \le \exp(L/2 - dtk/2) \le \exp(kd(1-t)/2)\,.$$

Finally, we have

$$\begin{aligned}
&\Pr\{\mathcal{S} > (6C+2)k + 3tk\} \le \Pr\{\mathcal{S} > \mathrm{E}\mathcal{S} + 3tk\} \\
&\le \Pr\left\{\sum_{i=1}^{\ell/2}(Z_{2i-1} - \mathrm{E}Z_{2i-1}) \ge tk\right\} + \Pr\left\{\sum_{i=1}^{\ell/2}(Z_{2i} - \mathrm{E}Z_{2i}) \ge tk\right\} + \Pr\left\{\sum_{i=1}^{\ell}(X_i - \mathrm{E}X_i) \ge tk\right\} \\
&\le 2\exp\left(-\frac{t^2k^2}{2C'tkd^{-1/3}}\right) + \exp(kd(1-t)/2) = 2\exp\left(-\frac{tkd^{1/3}}{2C'}\right) + \exp(kd(1-t)/2)\,,
\end{aligned}$$

and choosing $t = \Delta/3k$ gives (4). □

7.2 Large dimension

Thus far, we have established Theorem 14 for the case where the side-length k is not too small compared to the dimension d. In particular, we are done if $k/\ln(k+1) \geq d^{2/3}$. To finish the proof of Theorem 14, we will make use of a concentration result about the diameter of the d-cube whose strength increases with d. For a graph G, define

$$\mathrm{diam}(G) = \max\{\tau(G,s) : s \in V(G)\} \enspace .$$

Lemma 18. *Let G be the d-cube. Then there exists universal constants $c > 0$ and $x_0 > 0$ such that,*

$$\Pr\{\mathrm{diam}(G) > x\} \leq e^{-cxd} \enspace ,$$

for all $x_0 \leq x \leq 10d^2$.

Proof. We assume that d is greater than some sufficiently large constant, d_0. Otherwise the result follows trivially from the union bound: With probability at least $1 - d2^{d-1}e^{-x/d}$, every edge of the d-cube has weight at most x/d. For $d \leq d_0$, this satisifies the statement of the lemma with $x_0 = 3d_0 \ln d_0$ and $c = 1/(3d_0^2)$.

Fix some vertex s and, for each $i \in \{0,1,\ldots,d\}$, let L_i denote the subset of vertices whose distance to s is i. Balister et al. [**?**, Lemma 4] show that there are constants $a, c > 0$ such that, if we sample each edge of Q independently with probability a/d then, with probability at least $1 - e^{-cd^2}$, there is a path of length $d-4$ consisting entirely of sampled edges and having one endpoint in L_2 and one endpoint in L_{d-2}.

In our setting, this means that, if we only consider edges of weight at most $\ln(d/(d-a)) = O(1/d)$, then, with probability at least $1 - e^{-cd^2}$, there is a path of weight at most $d\ln(d/(d-a)) = O(1)$ joining a vertex u in L_2 to a vertex w in L_{d-2}.

Now, there are d edge-disjoint paths of length 2 joining s to u and with probability at least $1 - 2e^{-xd/6}$ at least one of these paths has weight no more than $x/3$. Similarly, with probability at least $1 - 2e^{-xd/6}$ there is a path of length 2 and weight at most $x/3$ joining w to the unique vertex $\bar{s}$ in L_d.

Therefore, with probability at least

$$1 - 4e^{-xd/6} - e^{-cd^2}$$

there is a path of weight at most

$$2x/3 + d\ln(d/(d-a)) \leq x$$

provided that $x \geq 3d\ln(d/(d-a))$. □

We can now finish the proof of Theorem 14.

Proof of Theorem 14. The idea of this proof is that, for any vertex v, there is a path from s to v that visits at most k d-cubes. Therefore, there is a path from s to v whose length is at most the sum of diameters of these cubes. We upper-bound this sum using Lemma 18 and Chernoff bounding.

For each vertex $u = (u_1, \dots, u_d)$ with $u_i \in \{0, \dots, k-1\}$ for each $i \in \{1, \dots, d\}$, define the subgraph $Q_u = G[V_u]$ of G induced by the vertex set

$$V_u = \{(u_1 + x_1, \dots, u_d + x_d) : x_i \in \{0,1\} \text{ for each } i \in \{1, \dots, d\}\} \ .$$

Each Q_u is a $(d,1)$-grid, i.e, a d-cube. We wish to apply Lemma 18 to these d-cubes, but Lemma 18 has a limited range of applicability, so we first rule out an unusual case. Let $\mathcal{E}$ denote the event

$$\max\{\mathrm{diam}(Q_u) : u \in \{0, \dots, k-1\}^d\} > 10d^2 \ .$$

By Lemma 18 and the union bound,

$$\Pr\{\mathcal{E}\} \le k^d e^{-10cd^2} = e^{d \ln k - 10cd^2}$$

For any graph Q, define $\mathrm{diam}'(Q) = \min\{10cd^2, \mathrm{diam}(Q)\}$ and observe that if the event $\mathcal{E}$ does not occur then $\mathrm{diam}'(Q_u) = \mathrm{diam}(Q_u)$ for all $u \in V(G)$.

For any vertex $v \in G$, there is a sequence $v_1, v_2, \dots, v_{k'}$ of vertices in G with $k' \le k$ such that

1. $s \in Q_{v_1}$ and $v \in Q_{v_{k'}}$;
2. for each $i \in \{1, \dots, k'-1\}$, Q_{v_i} and $Q_{v_{i+1}}$ have at least one vertex in common.
3. for each $i \in \{1, \dots, k\}$ and each $j \in \{1, \dots, k\} \setminus \{i-1, i, i+1\}$, Q_{v_i} and Q_{v_j} have no edges (or vertices) in common.

The sequence $v_1, \dots, v_{k'}$ can be found with a greedy algorithm: Define $v'_0 = s$. Now, if v'_{i-1} and v differ in r coordinates, then there is some vertex v_i such that Q_{v_i} contains v'_{i-1} as well as some vertex v'_i whose distance to v is r less than the distance from v'_{i-1} to v. It is straightforward to verify that the resulting sequence of vertices $v_1, \dots, v_{k'}$ satisfies the three properties described above.

Observe that $\tau(G, s, v)$ does not exceed the sum of the weighted diameters of $Q_{v_1}, \dots, Q_{v'_k}$:

$$\tau(G, s, v) \le \sum_{i=1}^{k'} \mathrm{diam}(Q_{v_i}) = \sum_{i=1}^{k'} \mathrm{diam}'(Q_{v_i}) \ ,$$

where the second equality holds provided that $\mathcal{E}$ does not occur. Point 3, above, ensures that the random variables $\mathrm{diam}(Q_{v_1}), \dots, \mathrm{diam}(Q_{v_{k'}})$ can be partitioned into two sets of size $\lfloor k'/2 \rfloor$ and $\lceil k'/2 \rceil$ where the variables within each set are independent.

Let $X_1, \dots, X_k$ and $Y_1, \dots, Y_k$ be random variables each distributed like $\mathrm{diam}'(Q_s)$ and such that $X_1, \dots, X_k$ are independent and $Y_1, \dots, Y_k$ are independent. It follows from the preceding discussion that $\sum_{i=1}^{k'} \mathrm{diam}'(Q_{v_i})$ is stochastically dominated by $\sum_{i=1}^{k} (X_i + Y_i)$. We now use Chernoff's bounding technique to bound the tails of the two sums $X = \sum_{i=1}^{k} X_i$ and $Y = \sum_{i=1}^{k} Y_i$.

For any $t > 0$,

$$\mathrm{E}\left[e^{tX}\right] = \mathrm{E}\left[e^{t\left(\sum_{i=1}^{k} X_i\right)}\right] = \mathrm{E}\left[\prod_{i=1}^{k} e^{tX_i}\right] = \prod_{i=1}^{k} \mathrm{E}\left[e^{tX_i}\right] = \mathrm{E}\left[e^{tX_1}\right]^k \quad .$$

Now,

$$\begin{aligned}
\mathrm{E}\left[e^{tX_1}\right] &= \int_0^\infty \Pr\left\{e^{tX_1} > x\right\} \mathrm{d}x \\
&= \int_0^\infty \Pr\left\{X_1 > \ln x/t\right\} \mathrm{d}x \\
&\le e^{tx_0} + \int_{e^{tx_0}}^\infty \Pr\{X_1 > \ln x/t\} \mathrm{d}x \\
&= e^{tx_0} + \int_{e^{tx_0}}^{e^{10td^2}} \Pr\{X_1 > \ln x/t\} \mathrm{d}x && \text{(since } X_1 \le 10d^2\text{)} \\
&\le e^{tx_0} + \int_{e^{tx_0}}^{e^{10td^2}} e^{-c \ln x d/t} \mathrm{d}x && \text{(by Lemma 18)} \\
&= e^{tx_0} + \int_{e^{tx_0}}^{e^{10td^2}} x^{-cd/t} \mathrm{d}x \\
&\le e^{tx_0} + \int_1^\infty x^{-cd/t} \mathrm{d}x \\
&= e^{tx_0} + \frac{t}{cd(1-t/cd)} \\
&= e^{tx_0} + 1 && \text{(for } t = cd/2\text{)} \\
&\le e^{tx_0+1} \quad .
\end{aligned}$$

Now we finish by applying Markov's Inequality. For any $a > 0$

$$\begin{aligned}
\Pr\{X > ak\} &= \Pr\{e^{tX} > e^{atk}\} \\
&\le \frac{\mathrm{E}[e^{tX}]}{e^{atk}} \\
&\le \frac{(e^{tx_0+1})^k}{e^{atk}} \\
&= e^{(tx_0+1)k - atk} \\
&= e^{(cdx_0/2+1)k - acdk/2} && \text{(recalling that } t = cd/2\text{)} \quad .
\end{aligned}$$

Letting $\tau'(G,s,v) = \max\{10kd^2, \tau(G,s,v)\}$ and $\tau'(G,s) = \max\{\tau(G,s,v) : v \in V(G)\}$ we now have

$$\Pr\{\tau'(G,s,v) > ak\} \le \Pr\{X > ak\} + \Pr\{Y > ak\} \le 2e^{(cdx_0/2+1)k - acdk/2} \quad .$$

Applying the union bound over all $(k+1)^d$ choices of v completes the proof:

$$\begin{aligned}\Pr\{\tau'(G,s) > ak\} &\le \sum_{v\in V(G)} \Pr\{\tau'(G,s,v) > ak\} \\ &\le 2(k+1)^d e^{(cdx_0/2+1)k - acdk/2} \\ &= e^{d\ln(2(k+1)) + (cdx_0/2+1)k - acdk/2} \\ &= o_n(1) \ ,\end{aligned}$$

for

$$a > \frac{2(d\ln(2(k+1) + (cdx_0/2+1)k)}{cdk} \ .$$

To summarize, there is a constant C such that

$$\begin{aligned}\Pr\{\tau(G,s) > Ck\} &\le \Pr\{\tau'(G,s) > ak\} + \Pr\{\tau'(G,s) \ne \tau(G,s)\} \\ &\le \Pr\{\tau'(G,s) > ak\} + \Pr\{\mathcal{E}\} \\ &\le o_n(1) + e^{\ln d + d\ln(k+1) - 10cd^2} \\ &\le o_n(1)\end{aligned}$$

for $\ln(k+1) < 5d$. This last condition is certainly true if $k/\ln(k+1) < d^{2/3}$. □

8 Lower Bounds for General Graphs

Next, we describe a series of lower bound constructions that match the upper bounds obtained in Theorems 6–10. In particular, these constructions show that the dependence on Δ in the upper bounds in the previous sections can not be asymptotically reduced.

In this section we prove the following theorem.

Theorem 19. *There exists a positive constant c such that for any given positive integers* $1 < \Delta, D$ *satisfying* $D \ge 16e^3 \ln\Delta$, *there exists an n-vertex graph G with maximum degree* $\le \Delta$, *diameter* $\le D$, *and a vertex s satisfying* $\Pr\{h(\mathcal{T}(G,s)) \ge c(\Delta\ln n + \Delta D)\} \ge 1 - o_n(1)$.

The graph G is obtained by gluing together two graphs H and I. The graph H has large diameter and high connectivity. The graph I has low connectivity and small diameter. By gluing them we obtain a graph of low diameter (because of I) but for which the infection is more likely to spread via H (because of its high connectivity), and hence will have a large height. We begin by defining and studying H and I individually.

8.1 The Ladder Graph H

Let L, δ, a be positive integers. The graph H is shown in Figure 1. The vertices of H are partitioned into L groups $V_1, \ldots, V_L$, each of size δ. The edge set of H is

$$E(H) = \bigcup_{i=1}^{L-1} \{vw : v \in V_i,\, w \in V_{i+1}\} \ .$$

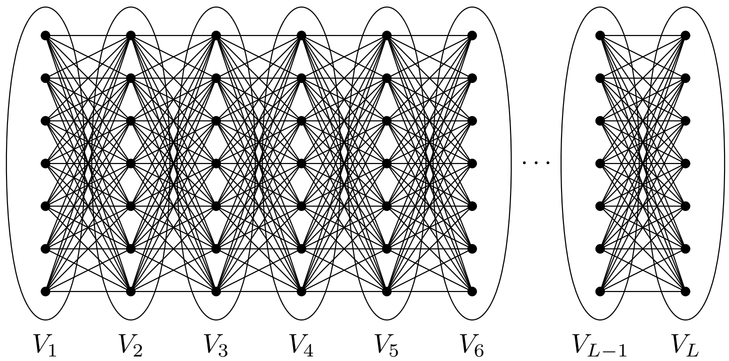

Figure 1: The graph H.

First we show that the infection spreads rather quickly in H, namely we prove upper bounds for $\tau(H, v, w)$.

Lemma 20. *Let $a > e^2$. Then for any $1 \le i < j \le n$ and any $v \in V_i$, $w \in V_j$ we have*

$$\Pr\{\tau(H, v, w) > 2aL/(e^2\delta)\} \le \exp(L - aL/(2e^2)) + \exp(-aL/(e^2\delta)) \ .$$

Proof. Consider the following greedy algorithm for finding a path from v to w: The path starts at v (which is in V_i). When the path has reached some vertex $x \in V_k$, for $k < j-1$, the algorithm extends the path by taking the minimum-weight edge joining x to some vertex in V_{k+1}. When the algorithm reaches some $x \in V_{j-1}$, it takes the edge xw.

Let $m = j - i$. Each of the first $m-1$ edges in the resulting path has a weight that is the minimum of δ exponential(1) random variables. Thus, the sum of weights of these edges is the sum of $m-1$ exponential(δ) random variables, i.e. an Erlang($m-1, \delta$) random variable. The weight of the final edge is an exponential(1) random variable. Thus we find

$$\begin{aligned}
\Pr\left\{\tau(H, v, w) > 2aL/(e^2\delta)\right\} &\le \Pr\left\{\text{Erlang}(m-1, \delta) + \text{exponential}(1) > 2aL/(e^2\delta)\right\} \\
&\le \Pr\left\{\text{Erlang}(m-1, \delta) > aL/(e^2\delta)\right\} + \Pr\left\{\text{exponential}(1) > aL/(e^2\delta)\right\} \\
&\le \Pr\left\{\text{Erlang}(L, \delta) > aL/(e^2\delta)\right\} + \exp(-aL/(e^2\delta)) \\
&\le \exp(L - aL/(2e^2)) + \exp(-aL/(e^2\delta)) \,.
\end{aligned}$$

The first inequality follows from the discussion above. The second inequality follows from the union bound. The third inequality is because an Erlang(L, δ) random variable stochastically dominates an Erlang($m-1, \delta$) random variable, and the definition of the exponential distribution. The final equality follows from the tail bound (2). $\square$

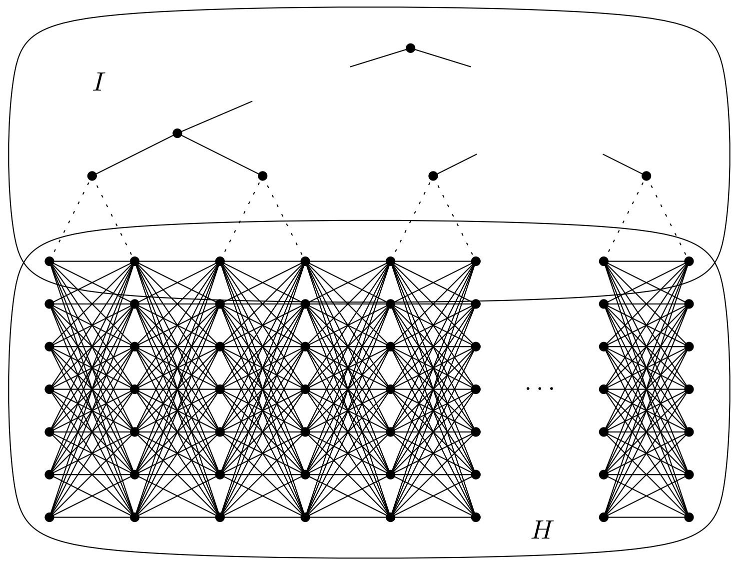

Figure 2: The lower bound graph G. Dotted segments denote subdivided edges (paths of length $\lceil aL/\delta \rceil$).

8.2 The Subdivided Tree I

Next, we consider a tree I that is obtained by starting with a perfect binary tree[3] having L leaves and then subdividing each edge incident to a leaf $\lceil aL/\delta \rceil - 1$ times so that each leaf-incident edge becomes a path of length $\lceil aL/\delta \rceil$. Note that I has height $\lceil aL/\delta \rceil + \log_2 L - 1$ (we assume L is a power of 2).

We next show that the infection spreads rather slowly in I, namely we prove lower bounds for $\tau(I, v, w)$.

Lemma 21. *For any distinct leaves v and w we have* $\Pr\{\tau(I, v, w) \le 2aL/(e^2\delta)\} \le \exp(-2aL/\delta)$.

Proof. The path from v to w in I contains at least $2\lceil aL/\delta \rceil$ edges. Therefore, the weight of this path is lower-bounded by the sum of $2\lceil aL/\delta \rceil$ independent exponential(1) random variables. The lemma then follows by applying (1) to this sum. □

8.3 Putting it Together

The lower-bound graph G is now constructed by taking a tree I with L leaves and a graph H with L groups $V_1, \ldots, V_L$ each of size $\delta = \lfloor (\Delta - 1)/2 \rfloor$. Next, we consider the leaves of I in the order they are encountered in a depth first-traversal of I and, for each $i \in \{1, \ldots, L\}$ we identify the ith leaf of I with some vertex in V_i. See Figure 2.

[3] A perfect binary tree, sometimes called a complete binary tree, is a binary tree in which all vertices have 0 or 2 children, and all leaves have the same depth: `https://xlinux.nist.gov/dads/HTML/perfectBinaryTree.html`

Lemma 22. *For any vertex $s \in V_1$ in the graph G described above, we have*

$$\Pr\{h(\mathcal{T}(G,s)) < L-1\} \le L^2\left(\exp((1-a/2e^2)L) + \exp(-aL/(e^2\delta)) + \exp(-2aL/\delta)\right)$$

Proof. Recall that $\mathcal{T}(G,s)$ is the shortest-path tree rooted at s for the first-passage percolation in G. If this tree contains no edge of I, its height is at least $L-1$. If it does use some edge of I, then there must be two leaves v and w of I such that $\tau(I,v,w) \le \tau(H,v,w)$. Since there are $\binom{L}{2} < L^2$ choices for the pair $\{v,w\}$, using Lemma 20 and Lemma 21, we can bound the probability of this event by

$$\begin{aligned}
&L^2\left(\Pr\{d_H(v,w) > 2aL/\delta\} + \Pr\{d_I(v,w) < 2aL/\delta\}\right)\\
&\le L^2\left(\exp((1-a/2e^2)L) + \exp(-aL/(e^2\delta)) + \exp(-2aL/\delta)\right),
\end{aligned}$$

which proves the lemma. □

We now have all the ingredients to prove the main theorem of this section, Theorem 19.

Proof of Theorem 19. Let $a = 4e^2$, $\delta = (\Delta-1)/2$, and let L be the largest power of 2 that is not larger than $D\Delta/8a$. Let G be the graph described above. The maximum degree of G is $2\delta+1 = \Delta$, and the diameter of G is bounded by

$$2(aL/\delta + \log_2 L) \le 2(a \times (D\Delta/8a)/(\Delta/2) + \log_2(D\Delta/8a)) \le D\,,$$

and its number of vertices is

$$n = L\delta + (2L-1) + L(aL/\delta - 1) < L(\delta + 1 + aL/\delta)\,.$$

We have

$$L \ge D\Delta/4a = \Omega(D\Delta + \Delta\ln L + \Delta\ln(\delta+1+aL/\delta)) = \Omega(\Delta\ln n + \Delta D)\,.$$

By Lemma 22, there exists a vertex s such that

$$\begin{aligned}
&\Pr\{h(\mathcal{T}(G,s) \ge \Omega(\Delta\ln n + \Delta D))\} \ge \Pr\{h(\mathcal{T}(G,s) \ge L-1)\}\\
&\ge 1 - L^2\left(\exp((1-a/2e^2)L) - \exp(-aL/(e^2\delta)) - \exp(-2aL/\delta)\right)\\
&= 1 - \left(\exp(-L + 2\ln L) - \exp(-8L/\Delta + 2\ln L) - \exp(-16e^2L/\Delta + 2\ln L)\right) = 1 - o_L(1) = 1 - o_n(1)\,,
\end{aligned}$$

completing the proof. □

9 Lower Bounds for Degenerate Graphs

Theorem 19 shows that Theorem 6 cannot be strengthened without knowing more about G than its number of vertices, maximum degree, and diameter. Theorem 8 provides a stronger upper bound under the assumption that G is d-degenerate. In this section we show that Theorem 8 is also tight, even when restricted to very special subclasses of d-degenerate graphs.

First we show that the bound given by Theorem 8 for $O(1)$-degenerate graphs is tight, even when we restrict our attention to planar graphs, which are 5-degenerate. Since planar graphs have genus 0, this lower bound also shows that Theorem 10, which applies to bounded genus graphs, is tight.

Theorem 23. *There exists an absolute constant $c > 0$ such that for any $\Delta > 1$ and $D \geq 10^6 \ln \Delta$ there exists a planar graph with diameter $\leq D$, maximum degree $\leq \Delta$, and a vertex s such that with probability $1 - o_n(1)$ we have $h(\mathcal{T}(G,s)) > c\sqrt{\Delta}(D + \ln n)$.*

Proof. Let $C = 10^5$, $a = e^2 C$, $\delta = \Delta/2$, and $L = D\sqrt{\delta}/3a$, and Let H be the graph shown in Figure 3, where each V_i has δ vertices. Let I be the perfect binary tree with L leaves, with each leaf-incident edge subdivided $aL/\sqrt{\delta} - 1$ times. Let G be the graph obtained from identifying the ith leaf of I with an arbitrary vertex from V_i. Note that G is a planar graph with maximum degree $2\delta = \Delta$, diameter $2(aL/\sqrt{\delta} + 1 + \log_2 L) \leq D$, and $n = \delta L + L - 1 + (2L-1) + L(aL/\delta - 1) = O(\delta L + L^2/\delta)$ vertices. Let s be an arbitrary vertex in V_1. Since $L = \Omega(\sqrt{\Delta}(D + \ln n))$, to complete the proof, we need only show that with probability $1 - o_n(1)$ we have $h(\mathcal{T}(G,s)) \geq 2L - 2$.

Choose an arbitrary vertex $t \in V_L$. Let $\mathcal{A}$ denote the event $\tau(H,s,t) \leq CL/\sqrt{\delta}$, and let $\mathcal{B}$ denote the event "for all pairs v and w of leaves of I we have $\tau(I,v,w) > CL/\sqrt{\delta}$. Note that if both $\mathcal{A}$ and $\mathcal{B}$ happen, then the path in $\mathcal{T}(G,s)$ from s to t uses edges from H only, which implies the height of this tree is at least $2L-2$. To complete the proof via the union bound, we need only show that each of $\mathcal{A}$ and $\mathcal{B}$ happen with probability $1 - o_L(1) = 1 - o_n(1)$.

We start with $\mathcal{A}$. In H, one can go from the vertex in-between V_i and V_{i+1} to the vertex in-between V_{i+1} and V_{i+2} by taking a path whose weight is distributed as a $Y_{\delta,1}$ random variable (recall the definition of a $Y_{a,b}$ random variable from Section 2). Therefore, we have

$$\tau(H,s,t) = X_1 + X_2 + \sum_{i=1}^{L-2} Z_i\,,$$

where X_1, X_2 are independent exponential(1) random variables (weights of the first and last edges), and Z_i's are independent $Y_{\delta,1}$ random variables. Since $C/3 \geq 3 \times (64 + 1024)$, Using Lemma 1 (concentration for the sum of $Y_{a,b}$ random variables) we have

$$1 - \Pr\{\mathcal{A}\} \leq 2\Pr\left\{X_1 > CL/3\sqrt{\delta}\right\} + \Pr\left\{\sum_{i=1}^{L-1} Z_i > CL/3\sqrt{\delta}\right\} \leq 2\exp\left(-CL/3\sqrt{\delta}\right) + \exp(-(L-2)/9) = o_L(1)$$

We now prove $\mathcal{B}$ happens with high probability. The path connecting any pair of leaves of I contains at least $2aL/\sqrt{\delta}$ edges, each of them having an independent exponential(1) weight. Therefore, using union bound over all pairs and using (1) we get

$$1 - \Pr\{\mathcal{B}\} \leq \binom{L}{2} \times \Pr\{\mathrm{Erlang}(2aL/\sqrt{\delta}, 1) \leq CL/\sqrt{\delta}\} \leq L^2 \times (eC/2a)^{2aL/\sqrt{\delta}} = o_L(1)\,,$$

completing the proof. □

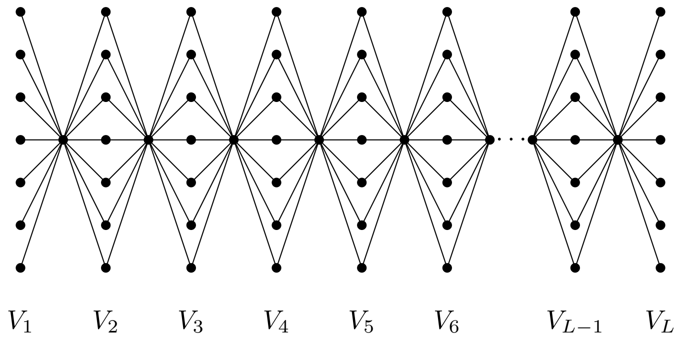

Figure 3: The graph H in the proof of Theorem 23.

Next we describe a lower-bound construction that is d-degenerate, has thickness d and treewidth $O(d)$. This construction shows that Theorem 8 is asymptotically tight for all values $d \le \Delta$ (with certain restrictions etc.).

Theorem 24. *There exists an absolute constant $c > 0$ such that for any $\Delta > 1$ with $D \ge 10^6 \ln \Delta$ and $d < \Delta$, there exists a graph G with diameter $\le D$, maximum degree $\le \Delta$, and the following properties:*

(i) G is d-degenerate, has thickness $\le d$ and treewidth $\le 2d+1$.

(ii) G has a vertex s such that with probability $1 - o_n(1)$ we have $h(\mathcal{T}(G,s)) > c\sqrt{d\Delta}(D + \ln n)$.

Proof. Let $C = 10^5$, $a = e^2 C$, $\delta = \Delta/2$, and $L = D\sqrt{d\Delta}/8a$, and Let H be the graph shown in Figure 4, where each V_i has δ vertices and each V_i' has d vertices, and each of the pairs (V_1, V_1'), (V_1', V_2), (V_2, V_2'), etc. form a complete bipartite graph. Let I be the perfect binary tree with L leaves, with each leaf-incident edge subdivided $aL/\sqrt{d\delta} - 1$ times. Consider the leaves of I in the order they are encountered in a depth first-traversal, for each $i \in \{1, \dots, L\}$ identify the ith leaf of I with some vertex in V_i. Let G be the resulting graph. Note that G has maximum degree $2\delta = \Delta$, diameter $\le 2(1 + aL/\sqrt{d\delta} + \log_2 L) \le D$, and $n = (\delta + D)L + 2L - 1 + L(aL/d\delta - 1) = O(\Delta L + L^2/d\Delta)$ vertices.

(i) Graph G is d-degenerate because the vertices of degree greater than d form an independent set. Therefore, every induced subgraph of G is either an independent set (so has a vertex of degree 0) or contains a vertex of degree at most d.

To see that G has thickness d, for each $i = 1, \dots, L$, assign to each vertex of V_i' a distinct colour from one of d colour classes. Now partition the edges incident to these vertices among d subgraphs depending on the color of the vertex they are incident to. Edges not incident to these vertices can be assigned to any subgraph. With this partition of edges, each subgraph becomes a subgraph of the planar graph used in the proof of Theorem 23.

To see that G has treewidth $2d+1$, we build a tree decomposition of G with bags of maximum size $2d+2$. For convenience, we define $V_0 = V_{L+1} = \emptyset$.

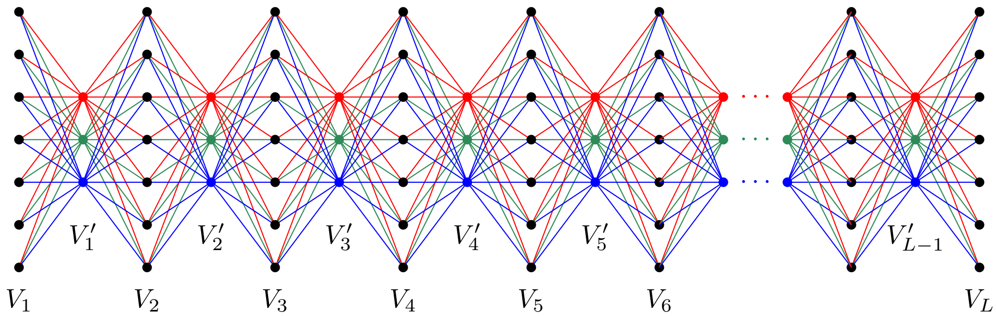

Figure 4: The d-degenerate graph H used in the proof of Theorem 24. In this example, $\delta = 7$ and $d = 3$.

We begin with a tree T' of empty bags that has the same shape as I. For each vertex v of I, let B_v denote the bag of v.

1. Assign each vertex of v of I to B_v and to the (up to 2) children of B_v in T'.
2. Let $v_1, \dots, v_L$ be the leaves of I ordered so that each $v_i \in V_i$. In the leaf bag B_{v_i} of T' we add all vertices in V'_{i-1} and V'_i.

   Now each vertex in V'_i appears in B_{v_i} and $B_{v_{i+1}}$; so we add all vertices of V'_i to each of the bags on the path in T' from B_{v_i} to $B_{v_{i+1}}$.
3. Finally, to each B_{v_i} we attach $\delta - 1$ bags as leaves of T'; in each bag we put all the vertices in $V'_i \cup V'_{i+1}$, and a distinct vertex of $V_i \setminus \{v_i\}$. We call each such bag B_v, where v is the unique vertex of $V_i \setminus \{v_i\}$ contained in the bag.

No bag contains more than $2d + 2$ vertices: for a leaf v_i, B_{v_i} contains v_i and its parent, as well as vertices in $V'_{i-1} \cup V'_i$. For a non-leaf vertex v of I, observe that (in any binary tree) there are at most two distinct indices i, j such that v lies on the (v_i, v_{i+1})-path in I and on the (v_j, v_{j+1})-path, hence B_v contains v and its parent, as well as possibly V'_i and V'_j. For each $v \in V_i \setminus \{v_i\}$, B_v contains at most $2d + 1$ vertices; v and the vertices in $V'_{i-1} \cup V'_i$.

For each edge vw of G, there is some bag that contains both v and w: If vw is an edge of T with v a child of w then B_v contains both v and w. Otherwise, $v \in V_i$ and $w \in V'_{i-1}$ or $w \in V'_i$, in which case v and w appear in B_v.

Finally, for each vertex v of G, the subgraph of T' induced by bags containing v is connected: For a vertex $v \in I$ this subgraph is either an edge or a single vertex. For a vertex $v \in V_i$ this subgraph is a single vertex. For a vertex $v \in V'_i$ this subgraph is a path joining two vertices of T'.

Therefore, T' is a tree-decomposition of G whose largest bag has size $2d + 2$, and thus treewidth of G is at most $2d + 1$.

(ii) Let s be an arbitrary vertex in V_1. Since $L = \Omega(\sqrt{d\Delta}(D + \ln n))$, to prove part (ii) we need only show that with probability $1 - o_n(1)$ we have $h(\mathcal{T}(G, s)) \geq 2L - 2$.

Choose an arbitrary vertex $t \in V_L$. Let $\mathcal{A}$ denote the event $\tau(H,s,t) \le CL/\sqrt{d\delta}$, and let $\mathcal{B}$ denote the event "for all pairs v and w of leaves of I we have $\tau(I,v,w) > CL/\sqrt{d\delta}$. Note that if both $\mathcal{A}$ and $\mathcal{B}$ happen, then the path in $\mathcal{T}(G,s)$ from s to t uses edges from H only, which implies the height of this tree is at least $2L-2$. To complete the proof via the union bound, we need only show that each of $\mathcal{A}$ and $\mathcal{B}$ happen with probability $1-o_L(1) = 1-o_n(1)$.

We start with $\mathcal{A}$. In H, one can go from a given vertex in V_i' to some vertex in V_{i+1}' by taking a path whose weight is distributed as a $Y_{\delta,d}$ random variable. Therefore, $\tau(H,s,t)$ is stochastically dominated by

$$X_1 + X_2 + \sum_{i=1}^{L-2} Z_i \,,$$

where X_1, X_2 are independent exponential(1) random variables (weights of the first and last edges), and Z_i's are independent $Y_{\delta,d}$ random variables. Since $C/3 \ge 3 \times (64+1024)$, Using Lemma 1 (concentration for the sum of $Y_{a,b}$ random variables) we have

$$1-\Pr\{\mathcal{A}\} \le 2\Pr\{X_1 > CL/3\sqrt{d\delta}\} + \Pr\{\sum_{i=1}^{L-2} Z_i > CL/3\sqrt{d\delta}\} \le 2\exp(-CL/3\sqrt{d\delta}) + \exp(-(L-2)/9) = o_L(1)$$

We now prove $\mathcal{B}$ happens with high probability. The path connecting any pair of leaves of I contains at least $2aL/\sqrt{d\delta}$ edges, each of them having an independent exponential(1) weight. Therefore, using union bound over all pairs and using (1) we get

$$1-\Pr\{\mathcal{B}\} \le \binom{L}{2} \times \Pr\{\mathrm{Erlang}(2aL/\sqrt{d\delta}, 1) \le CL/\sqrt{d\delta}\} \le L^2 \times (eC/2a)^{2aL/\sqrt{d\delta}} = o_L(1)\,,$$

completing the proof. □

Acknowledgements

Some of this research took place at the Workshop on Random Geometric Graphs and Their Applications to Complex Networks, at the Banff International Research Station, November 6–11, 2016. More of this research took place at the Rio Workshop on Geometry and Graphs, at IMPA, February 12–18, 2017. In both cases, we are grateful to the workshop organizers and other participants for providing stimulating working environments. We are especially grateful to Gábor Lugosi for helpful discussions on many aspects of this work and to Tasos Sidiropoulos for asking us about planar graphs.

References

[1] A. Auffinger, M. Damron, and J. Hanson. 50 years of first passage percolation. *arXiv*, 1511.03262 [math.PR], 2016.

[2] Béla Bollobás. The isoperimetric number of random regular graphs. *Eur. J. Comb.*, 9(3):241–244, 1988.

[3] Béla Bollobás and Yoshiharu Kohayakawa. On Richardson's model on the hypercube. In *Combinatorics, Geometry, and Probability (Cambridge 1993)*, pages 129–137. Cambridge University Press, 1997.

[4] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[5] Luc Devroye. Branching processes in the analysis of the heights of trees. *Acta Informatica*, 24(3):277–298, 1987.

[6] R. Durrett. Stochastic growth models: recent results and open problems. In *Mathematical approaches to problems in resource management and epidemiology (Ithaca, NY, 1987)*, volume 81 of *Lecture Notes in Biomath.*, pages 308–312. Springer, Berlin, 1989.

[7] James Allen Fill and Robin Pemantle. Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *The Annals of Applied Probability*, 3(2):593–629, 1993.

[8] S. Janson. Tail bounds for sums of geometric and exponential variables. available in http://www2.math.uu.se/~svante/papers/sjN14.pdf.

[9] Anders Martinsson. Unoriented first-passage percolation on the n-cube. *The Annals of Applied Probability*, 26(5):2597–2625, 2016.

[10] Colin McDiarmid. *Concentration*, pages 195–248. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. available at http://www.stats.ox.ac.uk/people/academic_staff/colin_mcdiarmid/?a=4139.

[11] Bojan Mohar and Carsten Thomassen. *Graphs on Surfaces*. Johns Hopkins University Press, Baltimore, Maryland, 2001.

[12] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Struct. Algorithms*, 5(2):337–348, 1994.

[13] T Wolle and A. M. . C. A. Koster adn H. L. Bodlaender. A note on contraction degeneracy. Technical Report UU-CS-2004-042, University of Utrecht, 2004.

A Proof of Lemma 1

We will use the following inequality, which holds for any positive integer k and any real number λ (see [8, Theorem 5.1(ii)]):

$$\Pr\{\mathrm{Erlang}(k,1) \geq \lambda k\} \leq \exp(1-\lambda)\,. \tag{5}$$

We will also use the following inequality, which holds for any binomial random variable X, and any $M \leq \mathrm{E}X$ (see [10, Theorem 2.3(c)]):

$$\Pr\{X < M/2\} \leq \exp(-M/8)\,. \tag{6}$$

Lemma 25. *For any t we have*

$$\Pr\{Y_{a,b} > t\} \le \exp(-at/64) + \exp(-abt^2/1024)\,.$$

Proof. First, consider the case $t > 4$. Note that there exist a independent root-to-leaf paths, the weight of each is Erlang$(2,1)$. Hence, using (5) and since $t \ge 4$,

$$\Pr\{Y_{a,b} > t\} \le \Pr\{\text{Erlang}(2,1) > t\}^a \le (\exp(1-t/2))^a \le (\exp(-t/4))^a = \exp(-at/4) \le \exp(-at/64).$$

The case $t \le 0$ is trivial, so we consider the case $0 \le t \le 4$. Note that for such t we have $1 - \exp(-t/2) \ge t/8$. We say a node in the tree *survives* if each of the edges on its path to the root have weight at most $t/2$. Note that $Y_{a,b} > t$ implies no node at level 2 survives. The probability that a node at level 1 (children of the root) survives is $1 - \exp(-t/2)$, so the number of surviving nodes at level 1, S_1, is a binomial random variable with mean $a(1-\exp(-t/2)) \ge at/8$. From (6) we have

$$\Pr\{S_1 < at/16\} \le \Pr\{S_1 < \mathrm{E}S_1/2\} \le \exp(-\mathrm{E}S_1/8) \le \exp(-at/64)\,.$$

Conditioned on $S_1 \ge at/16$, the number of surviving nodes at level 2, S_2, is a binomial random variable with mean $S_1 b(1-\exp(-t/2)) \ge abt^2/128$, so using(6) again we have

$$\begin{aligned}\Pr\{Y_{a,b} > t | S_1 \ge at/16\} &\le \Pr\{S_2 = 0 | S_1 \ge at/16\} \le \Pr\{S_2 < abt^2/256 | S_1 \ge at/16\} \\ &\le \exp(-abt^2/1024)\,,\end{aligned}$$

completing the proof. □

We are now ready to prove Lemma 1. Let $X_1, \ldots, X_m$ be i.i.d. distributed as $Y_{a,b}$ for some a, b. Then we want to prove $\mathrm{E}X_1 = O(1/a + 1/\sqrt{ab})$ and moreover,

$$\Pr\left\{\sum_{i=1}^{m} X_i \ge 3m(64/a + 1024/\sqrt{ab})\right\} \le \exp(-m/9)\,.$$

Let $d_1 = a/64$ and $d_2 = ab/1024$. For any positive integer p, by Lemma 25 we have

$$\mathrm{E}X_1^p = \int_0^\infty \Pr\{X_1 > t^{1/p}\}dt \le \int_0^\infty \exp(-d_1 t^{1/p}) + \int_0^\infty \exp(-d_2 t^{2/p})$$

For any positive numbers c, α, we have

$$\int_0^\infty \exp(-ct^\alpha)dt = \int_0^\infty \exp(-x)\frac{x^{1/\alpha-1}}{\alpha c^{1/\alpha}}dx = \frac{c^{-1/\alpha}}{\alpha}\int_0^\infty e^{-x}x^{1/\alpha-1} = \frac{c^{-1/\alpha}\Gamma(1/\alpha)}{\alpha}, \tag{7}$$

whence,

$$\mathrm{E}X_1^p \le pd_1^{-p}\Gamma(p) + pd_2^{-p/2}\Gamma(p/2)/2$$

In particular, setting $p = 1$ gives $\mathrm{E}X_1 \le 64/a + 1024/\sqrt{ab} =: c$. Let $v = 4c^2 m$. For $p \ge 2$, we have

$$\sum_{i=1}^{m} \mathrm{E}[X_i^p] \le mpd_1^{-p}\Gamma(p) + mpd_2^{-p/2}\Gamma(p/2)/2 \le mp!d_1^{-p} + mp!d_2^{-p/2}/2 \le vp!c^{p-2}/2\,.$$

Bernstein's inequality (Theorem 3) gives that for all t,

$$\Pr\left\{\sum_{i=1}^{m} X_i \geq m\left(64/a + 1024/\sqrt{ab}\right) + ct + 3c\sqrt{mt}\right\} \leq e^{-t},$$

and choosing $t = m/9$ completes the proof of the lemma.

B Proof of Lemma 2

In this section, we prove that for any positive integer p we have

$$\mathrm{E}Y_{a,b,c}^{p} \leq p(a/64)^{-p}\Gamma(p) + p(ab/1024)^{-p/2}\Gamma(p/2)/2 + p(abc/16384)^{-p/3}\Gamma(p/3)/3$$

and, in particular, $\mathrm{E}Y_{a,b,c} \leq 64/a + 1024/\sqrt{ab} + 16384/\sqrt[3]{abc}$.

We will first show, using an argument similar to the proof of Lemma 25, that for any t we have

$$\Pr\{Y_{a,b,c} > t\} \leq \exp(-at/64) + \exp(-abt^2/1024) + \exp(-abct^3/16384). \tag{8}$$

First, consider the case $t \geq 6$. Note that there exist a independent root-to-leaf paths, the weight of each is Erlang$(3,1)$. Hence, using (5) and since $t \geq 6$,

$$\Pr\{Y_{a,b,c} > t\} \leq \Pr\{\text{Erlang}(3,1) > t\}^a \leq (\exp(1-t/3))^a \leq (\exp(-t/6))^a = \exp(-at/6) \leq \exp(-at/64)$$

The case $t \leq 0$ is trivial, so we consider the case $0 < t < 6$. Note that for such t we have $1 - \exp(-t/3) \geq t/8$. We say a node in the tree *survives* if each of the edges on its path to the root have weight at most $t/3$. Note that $Y_{a,b,c} > t$ implies no node at level 3 survives. The probability that a node at level 1 (children of the root) survives is $1 - \exp(-t/3)$, so the number of surviving nodes at level 1, S_1, is a binomial random variable with mean $a(1 - \exp(-t/3)) \geq at/8$. From (6) we have

$$\Pr\{S_1 < at/16\} \leq \exp(-at/64).$$

Conditioned on $S_1 \geq at/16$, the number of surviving nodes at level 2, S_2, is a binomial random variable with mean $S_1 b(1 - \exp(-t/3)) \geq abt^2/128$, so using(6) again we have

$$\Pr\{S_2 < abt^2/256 | S_1 \geq at/16\} \leq \exp(-abt^2/1024).$$

Finally, conditioned on $S_2 \geq abt^2/256$, the number of surviving nodes at level 3, S_3, is a binomial random variable with mean $S_2 c(1-\exp(-t/3)) \geq abct^3/2048$, so using(6) again we have

$$\begin{aligned}\Pr\{Y_{a,b,c} > t | S_2 \geq abt^2/256\} &\leq \Pr\{S_3 = 0 | S_2 \geq abt^2/256\} \leq \Pr\{S_3 < abct^3/4096 | S_2 \geq abt^2/256\} \\ &\leq \exp(-abct^3/16384),\end{aligned}$$

completing the proof of (8).

Now let $d_1 = a/64$, $d_2 = ab/1024$, and $d_3 = abc/16384$. For any positive integer p we have, using (8) and (7),

$$\begin{aligned} \mathrm{E}Y_{a,b,c}^p &= \int_0^\infty \Pr\{Y_{a,b,c} > t^{1/p}\} dt \leq \int_0^\infty \exp(-d_1 t^{1/p}) + \int_0^\infty \exp(-d_2 t^{2/p}) + \int_0^\infty \exp(-d_3 t^{3/p}) \\ &= p d_1^{-p} \Gamma(p) + p d_2^{-p/2} \Gamma(p/2)/2 + p d_3^{-p/3} \Gamma(p/3)/3 \end{aligned}$$

In particular, setting $p = 1$ gives $\mathrm{E}Y_{a,b,c} \leq 64/a + 1024/\sqrt{ab} + 16384/\sqrt[3]{abc}$, completing the proof of the lemma.