

The Local Limit of the Uniform Spanning Tree on Dense Graphs

Jan Hladký¹  · Asaf Nachmias² · Tuan Tran³

Received: 21 October 2017 / Accepted: 28 November 2017 / Published online: 10 January 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Let G be a connected graph in which almost all vertices have linear degrees and let \mathcal{T} be a uniform spanning tree of G . For any fixed rooted tree F of height r we compute the asymptotic density of vertices v for which the r -ball around v in \mathcal{T} is isomorphic to F . We deduce from this that if $\{G_n\}$ is a sequence of such graphs converging to a graphon W , then the uniform spanning tree of G_n locally converges to a multi-type branching process defined in terms of W . As an application, we prove that in a graph with linear minimum degree, with high probability, the density of leaves in a uniform spanning tree is at least $e^{-1} - o(1)$, the density of vertices of degree 2 is at most $e^{-1} + o(1)$ and the density of vertices of degree $k \geq 3$ is at most $\frac{(k-2)^{k-2}}{(k-1)!e^{k-2}} + o(1)$. These bounds are sharp.

Keywords Uniform spanning tree · Graph limits · Benjamini-Schramm convergence · Graphon · Branching process

1 Introduction

It is a classical fact [9, 11] that a uniformly chosen tree from the set of n^{n-2} trees on n vertices, viewed from an independently chosen uniform random vertex, is locally distributed as a Poisson(1) Galton–Watson branching process conditioned to live forever when n is large. One can use this result to find the distribution of a certain “local” structures. For

✉ Asaf Nachmias
asafnach@post.tau.ac.il

Jan Hladký
honzahladky@gmail.com

Tuan Tran
manh.tran@math.ethz.ch

¹ Institut für Geometrie, TU Dresden, 01062 Dresden, Germany

² Department of Mathematical Sciences, Tel Aviv University, 69978 Tel Aviv, Israel

³ Department of Mathematics, ETH, 8092 Zurich, Switzerland

instance, it follows that the degree distribution of a uniformly chosen vertex of a uniformly chosen tree on n vertices converges to the law of a $\text{Poisson}(1) + 1$ random variable.

A uniformly chosen tree on n vertices is a uniform spanning tree (UST) of the complete graph on n vertices. Our goal in this paper is to explicitly describe the local structure of the UST of any dense graph or, equivalently, of a sequence of dense graphs converging to a given graphon. Let us first present this result.

Given a connected graph G we write \mathcal{T} for a uniformly drawn spanning tree of G and $B_{\mathcal{T}}(v, r)$ for the graph-distance ball of radius r in \mathcal{T} around the vertex $v \in V(G)$. Our goal is to describe the asymptotic distribution of $B_{\mathcal{T}}(X, r)$, viewed up to graph isomorphism, where X is a uniformly chosen random vertex of G . To that aim, let $\Omega = [0, 1]$ and μ be the Lebesgue measure on Ω . For a given graphon $W : \Omega^2 \rightarrow [0, 1]$ (see Sect. 1.1 for a brief introduction to graphons) and $\omega \in \Omega$ we write $\text{deg}(\omega) = \int_{y \in \Omega} W(\omega, y) dy$, and call this number the **degree** of ω . A graphon is called **nondegenerate** if for almost every $\omega \in \Omega$ we have $\text{deg}(\omega) > 0$.

Let T be a fixed rooted tree with $\ell \geq 2$ vertices and of height $r \geq 1$. We write by Stab_T the set of graph automorphisms of T that preserve the root. In what follows we denote the vertices of T by the numbers $\{1, \dots, \ell\}$ such that the vertices p, \dots, ℓ are the vertices at height r of T and $p \in \{2, \dots, \ell\}$. Given a nondegenerate graphon W and a tree T as above we define

$$\begin{aligned} \text{Freq}(T; W) := & \frac{1}{|\text{Stab}_T|} \int_{\omega_1, \dots, \omega_\ell} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \frac{\sum_{j=p}^{\ell} \text{deg}(\omega_j)}{\prod_{j=1}^{\ell} \text{deg}(\omega_j)} \\ & \times \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) d\omega_1 \cdots d\omega_\ell, \end{aligned} \tag{1}$$

where

$$b_W(\omega) = \int_{y \in \Omega} \frac{W(\omega, y)}{\text{deg}(y)} d\mu. \tag{2}$$

Theorem 1.1 *Let T be a fixed rooted tree as above, and $W : \Omega^2 \rightarrow [0, 1]$ be a nondegenerate graphon. Then for any $\varepsilon > 0$ there exists $\xi = \xi(\varepsilon, W, T) > 0$ such that if G is a connected simple graph on at least ξ^{-1} vertices that is ξ -close to W in the cut-distance, then with probability at least $1 - \varepsilon$ a uniformly chosen spanning tree \mathcal{T} of G satisfies*

$$|\mathbb{P}(B_{\mathcal{T}}(X, r) \cong T) - \text{Freq}(T; W)| \leq \varepsilon,$$

where X is an independently and uniformly chosen vertex of G , and by $B_{\mathcal{T}}(X, r) \cong T$ we mean that between the two rooted trees there is a graph-isomorphism preserving the root.

Theorem 1.1 is a natural statement in the context of limits of graph sequences. It asserts that if $\{G_n\}$ is a sequence of connected graphs converging to a graphon W , then the UST of G_n converges locally to a certain multi-type branching process that is defined in terms of W . This interpretation involve two graph limit procedures: a dense graph limit [7, 16] to describe the limit of the dense graph sequence G_n , and sparse graph limit to describe the random limit of the UST of G_n , known as Benjamini–Schramm convergence [4]. We refer the reader to Sects. 1.1 and 1.2 for an introduction to these limiting procedure, and begin by describing the limiting branching process.

Definition 1.2 Given a nondegenerate graphon $W : \Omega^2 \rightarrow [0, 1]$ we define a multi-type branching process κ_W . The process has continuum many types that are either (**anc**, ω) or (**oth**, ω), where $\omega \in \Omega$. Here, “**anc**” stands for “ancestral” and “**oth**” stands for “other”.

- (1) The initial particle (i.e., the root) has type (anc, ω) , where the distribution of ω is μ .
- (2) If a particle has type (oth, ω) , then its progeny is $\{(\text{oth}, \omega_1), \dots, (\text{oth}, \omega_k)\}$ where $\{\omega_1, \dots, \omega_k\}$ are a Poisson point process on Ω with intensity $\frac{W(\omega, \omega')}{\text{deg}(\omega')}$ at $\omega' \in \Omega$. In particular, k has distribution $\text{Poisson}(b_W(\omega))$, where $b_W(\cdot)$ is defined in (2).
- (3) If a particle has type (anc, ω) the progeny is $\{(\text{anc}, \omega_0), (\text{oth}, \omega_1), \dots, (\text{oth}, \omega_k)\}$ where $\{\omega_1, \dots, \omega_k\}$ are a Poisson point process on Ω with the same intensity as above and ω_0 is an independent new particle of Ω which is distributed according to the probability measure on Ω that has density $\frac{W(\omega, \cdot)}{\text{deg}(\omega)}$.

We note that an ancestral vertex (i.e., of type anc) always has at least 1 progeny and since the initial particle is ancestral the process κ_W survives forever with probability 1. The following theorem is quickly deduced from Theorem 1.1 at Sect. 4.2.

Theorem 1.3 *Suppose that $\{G_n\}$ is a sequence of simple connected graphs of growing orders that converge to a nondegenerate graphon W in the cut-distance and let T_n be a UST of G_n . Then the sequence $\{T_n\}$ almost surely converge in the Benjamini–Schramm sense to κ_W .*

Since the L^1 -norm (corresponding to the edit distance in graph theory) is finer than the cut-norm, we in particular obtain the following. Suppose that G is a large connected simple n -vertex graph with minimum degree $\Omega(n)$. Suppose that we add and/or delete $o(n^2)$ edges in a way that the resulting graph G' stays connected. Then the structure of a typical UST of G and of G' is very similar in the Benjamini–Schramm sense. Even this statement seems to be new.

When G_n are a sequence of dense regular graphs, then the function $b_W : \Omega \rightarrow (0, \infty)$ defined in (2) is identically 1 and we obtain the following.

Corollary 1.4 *Suppose that $\{G_n\}$ is a sequence of simple connected regular graphs of growing orders that converge to a nondegenerate graphon W in the cut-distance and let T_n be a UST of G_n . Then the sequence $\{T_n\}$ almost surely converge in the Benjamini–Schramm sense to a Poisson(1) Galton–Watson branching process conditioned to live forever.*

Theorem 1.3 allows us to deduce several extremal properties of the UST on dense graphs. The number of vertices of degree $k \geq 1$ in the UST of the complete graph K_n is $(e^{-1}/(k - 1)! + o(1))n$. In fact, using Prüfer codes one can establish that the degree of a vertex in the UST of K_n has distribution $1 + \text{Bin}(n - 2, \frac{1}{n}) \approx 1 + \text{Poisson}(1)$.¹ Using Theorem 1.1 we are able to find the extremal values of the number of vertices of degree k in a general dense graph. In the following theorem we show that the complete graph (or any other regular dense graph) is the minimizer of the number of leaves and the maximizer of the number of vertices of degrees 2 and 3 among the class of dense graphs. Somewhat surprisingly, the maximizer for the number of vertices of degree $k \geq 4$ is a different dense graph (see Sect. 5).

Theorem 1.5 *For any $k \geq 1$ we denote by L_k the random variable counting the number of vertices of degree k in a UST of a simple connected graph G . For every $\varepsilon, \delta > 0$ there exist numbers $n_0 \in \mathbb{N}$ and $\gamma > 0$ such that the following holds: Whenever G is a graph on $n \geq n_0$ vertices with at least $(1 - \gamma)n$ vertices of degrees at least δn then*

$$\mathbb{P}(L_1 \leq (e^{-1} - \varepsilon)n) \leq \varepsilon, \tag{3}$$

$$\mathbb{P}(L_2 \geq (e^{-1} + \varepsilon)n) \leq \varepsilon, \tag{4}$$

¹ Prüfer codes, see e.g. [19, p. 245], provide a standard bijection between spanning trees of K_n and words of length $n - 2$ over the alphabet $V(K_n)$. A quick look at this bijection shows the number of occurrences of any letter $v \in V(K_n)$ in a Prüfer code is the degree of the vertex v in the corresponding spanning tree decreased by 1. Therefore, the degree of v in a uniform spanning tree has indeed distribution $1 + \text{Bin}(n - 2, \frac{1}{n})$.

and for any $k \geq 3$ we have

$$\mathbb{P} \left(L_k \geq \left(\frac{1}{(k-1)!} \frac{(k-2)^{k-2}}{e^{k-2}} + \varepsilon \right) n \right) \leq \varepsilon. \tag{5}$$

We derive Theorem 1.5 in Sect. 5. The rest of this section is organized as follows. We first give the formal definitions and background of dense and sparse graph limits necessary for the reader to parse the statement of Theorems 1.1 and 1.3. Next we discuss some aspects of the theorem, such as the necessity of its assumptions. We end this section with a discussion of related results in the literature.

1.1 Dense Graph Limits

Graphons were introduced by Borgs et al. [7, 16] as limit objects to sequences of dense graphs. Here, we review basic facts and we refer the reader to [15, Part II] for a thorough treatment of the subject. A **graphon** is a symmetric Lebesgue measurable function $W : \Omega^2 \rightarrow [0, 1]$, where Ω is any standard atomless probability space.² The underlying measure on Ω will be always denoted by μ .

Given a measurable function $U : \Omega^2 \rightarrow [-1, 1]$ we define its **cut-norm** by

$$\|U\|_{\square} = \sup_{S, T} \left| \int_{x \in S} \int_{y \in T} U(x, y) \right|, \tag{6}$$

where S and T range over all measurable subsets of Ω . Now, we can define the key notion of **cut-distance**. For two graphons $W_1, W_2 : \Omega^2 \rightarrow [0, 1]$, we define

$$\delta_{\square}(W_1, W_2) = \inf_{\varphi} \|W_1^{\varphi} - W_2\|_{\square}, \tag{7}$$

where $\varphi : \Omega \rightarrow \Omega$ ranges through all measure preserving automorphisms of Ω , and W_1^{φ} stands for a graphon defined by $W_1^{\varphi}(x, y) = W_1(\varphi(x), \varphi(y))$. Note that the definition of $\delta_{\square}(W_1, W_2)$ extends in a straightforward way if W_2 lives on some other standard atomless probability space Λ . In that case, φ ranges of all measure preserving bijections from Λ to Ω .

Suppose that G is an n -vertex graph. Then we can consider a **graphon representation** of G . To this end, partition an atomless standard probability space Ω into n sets, each set of measure $\frac{1}{n}$, $\Omega = \bigsqcup_{v \in V(G)} \Omega_v$. Then define a graphon W_G to be 1 on $\Omega_u \times \Omega_v$ for each edge $uv \in E(G)$, and 0 otherwise. Note that W_G is not unique as it depends on the choice of the partition $\{\Omega_v\}$. With the notion of a graphon representation, we can define distance between a graph and a graphon. Namely, if $W : \Omega^2 \rightarrow [0, 1]$ is a graphon and G is a graph, we define $\delta_{\square}(W, G) := \delta_{\square}(W, W_G)$. This definition does not depend on the choice of the representation W_G . We say that G is α -**close** to W if $\delta_{\square}(W, G) \leq \alpha$. Though through much of the paper, we shall work with loopless multigraphs (introduced in Sect. 2.1), we always restrict ourselves to graphs when representing as graphons, or when referring to the cut-distance.

Throughout the paper, all sets and functions are tacitly assumed to be measurable. Conversely all our constructions of auxiliary sets and functions are measurable, too.

We say that a sequence $(G_n)_n$ of simple graphs **converges** to a graphon W if and only if $\delta_{\square}(W, G_n) \rightarrow 0$. Now, it is possible to understand the first sentence of Theorem 1.3. The following statement, proved first in [16], is the core of the theory of graphons.

Theorem 1.6 *For every sequence of simple graphs of increasing orders there exists a subsequence which converges to a graphon.*

² Recall that all such probability spaces are isomorphic.

1.2 Sparse Graph Limits

Let \mathcal{G}_\bullet denote the space of rooted locally finite graphs viewed up to root-preserving graph isomorphisms. That is, each element of \mathcal{G}_\bullet is (G, ϱ) where G is a graph and ϱ is a vertex of it, and two such elements (G_1, ϱ_1) and (G_2, ϱ_2) are considered equivalent if and only if there exists a graph automorphism $\varphi : G_1 \rightarrow G_2$ such that $\varphi(\varrho_1) = \varrho_2$. Given a rooted graph (G, ϱ) and an integer $r \geq 1$ we write $B_G(\varrho, r)$ for the graph-distance ball of radius r around ϱ in G , that is, $B_G(\varrho, r) \in \mathcal{G}_\bullet$ is a finite graph rooted at ϱ on the set of vertices of graphs distance at most r from ϱ in G together with all the edges induced from G . There is a natural notion of a metric on \mathcal{G}_\bullet . The distance between two elements $(G_1, \varrho_1), (G_2, \varrho_2) \in \mathcal{G}_\bullet$ is defined to be 2^{-R} where $R \geq 0$ is the largest number such that there is a root-preserving graph isomorphism between $B_{G_1}(\varrho_1, R)$ and $B_{G_2}(\varrho_2, R)$. Having defined the metric we can consider probability spaces on \mathcal{G}_\bullet with respect to the Borel σ -algebra.

We say that the law of a random element $(G, \varrho) \in \mathcal{G}_\bullet$ is the **Benjamini–Schramm limit** of a (possibly random) sequence of finite graphs G_n , if and only if, for any fixed integer $r \geq 1$ the random variable $B_{G_n}(\varrho_n, r)$ converges in distribution to $B_G(\varrho, r)$ where ϱ_n is an independently chosen uniform random vertex of G_n . In this case we say that the sequence G_n Benjamini–Schramm converges to (G, ϱ) . Note that by putting $r = 1$ in the definition we deduce that in this convergence the degree of the random root ϱ_n must converge to the degree of ϱ which explains why this limiting procedure is best suited for sparse graphs. See further discussion in [4].

We have finished defining all the needed terminology required to parse Theorems 1.1 and 1.3.

1.3 Necessity of the Assumptions

The assumptions in Theorem 1.3 are the minimal necessary. First, we obviously need the assumption that G_n are connected in order for spanning trees to exist.

Next we claim that the local structure of a uniform spanning tree cannot be determined from a degenerate graphon W . Indeed, suppose that a graphon W is given, and let Ω^0 be the elements of Ω that have zero degree W and $\Omega^+ = \Omega \setminus \Omega^0$. Assume that W is degenerate so that $\mu(\Omega^0) = \delta > 0$.

We can now construct two graph sequences that converge to W . We start with dense graphs G_n of size $(1 - \delta)n$ that converge³ to $W^+ = W|_{\Omega^+}$ and in the first sequence we attach to G_n a path of length δn at an arbitrary vertex and in the second sequence we attach δn edges arbitrary to an vertex of G_n creating δn new vertices of degree 1. It is clear that both sequences converge to W . However, the USTs on the two sequences have different Benjamini–Schramm limits. Indeed, let p_1 denote the probability that in κ_{W^+} the root is a leaf. Then the probability that a randomly chosen vertex is a leaf in the first sequence tends to $(1 - \delta)p_1$ and in the second sequence this probability tends to $(1 - \delta)p_1 + \delta$.

1.4 Discussion

Theorem 1.3 shows that the local structure of the UST is continuous on the space of dense graphs with the cut-metric (7) and describes this local structure explicitly. As mentioned earlier, the only instance in the literature of Theorem 1.3 that we are aware of is the case of the UST of the complete graph K_n . In this case Grimmett [9] showed that the limiting object

³ Such a sequence can be obtained for example by taking typical inhomogeneous random graphs $\mathbb{G}(n, W^+)$, see [15, Lemma 10.16].

is an infinite path upon which we to each vertex an independent $\text{Poisson}(1)$ branching process—that is, a $\text{Poisson}(1)$ branching process conditioned to survive forever. This is precisely κ_W where corresponding graphon W is just $W \equiv 1$.

The analogous continuity result for sparse graphs is also true, though in this case it is typically harder to describe explicitly the limiting object.

Theorem 1.7 [1, Proposition 7.1] *Suppose that $\{G_n\}$ is a Benjamini–Schramm convergent sequence of connected graphs and let T_n be a UST of G_n . Then there exists a random rooted tree (T, ϱ) such that T_n Benjamini–Schramm converges to (T, ϱ) .*

The limiting object in the above theorem (T, ϱ) is the *wired uniform spanning forest* of the Benjamini–Schramm limit (G, ϱ) of the graphs $\{G_n\}$, see [18].

One can also ask whether the normalized number of spanning trees is continuous with respect to taking graph limits. Given a graph G , let $t(G)$ be the number of spanning trees in G . In the bounded-degree model, Lyons [17, Theorem 3.2] proved that $n^{-1} \log t(G)$ is continuous in the Benjamini–Schramm topology. In the dense model, the natural normalization of $t(G)$ is $n^{-1}t(G)^{1/n}$. For example when $G = K_n$ then by Cayley’s formula, $n^{-1}t(G)^{1/n}$ tends to 1, and when G is a typical Erdős–Rényi random graph $\mathbb{G}(n, p)$ for $p \in (0, 1)$ fixed, then $n^{-1}t(G)^{1/n}$ tends to p . However, one cannot expect that with no further assumptions continuity of this parameter with respect to the cut-metric will hold. Indeed, let G_n formed by a clique of order $n - \frac{n}{\sqrt{\log n}}$ and a path of length $\frac{n}{\sqrt{\log n}}$ attached to it at an arbitrary vertex. The sequence $\{G_n\}$ converges to the complete graphon $W \equiv 1$ but the number of spanning is substantially lower than for complete graphs; indeed, in this case it can easily be seen using Cayley’s formula that $n^{-1}t(G_n)^{1/n}$ tends to 0.

However, when we impose a minimum degree condition on the graphs, we can infer the asymptotic normalized number of spanning trees from the limit graphon.

Theorem 1.8 *Let $\delta > 0$. Suppose that G_1, G_2, \dots is a sequence of simple connected graphs that converge to a graphon W . Suppose that the order of G_n is n and the minimum degree is at least δn . Then the number of spanning trees satisfies*

$$\lim_{n \rightarrow \infty} \frac{\sqrt[n]{t(G_n)}}{n} = \exp \left(\int_x \log(\deg_W(x)) \right).$$

□

Theorem 1.8 follows almost immediately from a result due to Kostochka [12] which states that if $1 < d_1 \leq d_2 \leq \dots \leq d_n$ is the degree sequence of a simple connected graph G then for some absolute constant $C > 0$ we have

$$\frac{\prod_i d_i}{d_1^{(Cn \log d_1)/d_1}} \leq t(G) \leq \frac{\prod_i d_i}{n - 1}. \tag{8}$$

To see that (8) yields Theorem 1.8, it is enough to recall that the degree distribution of a limit graphon is inherited from degree distribution of graphs that converge to it (see Lemma 5.1(1)). In the case of regular graphs, Theorem 1.8 can also be derived from [2, 8].

Lastly, let us mention a result of a similar flavor to ours in the context of percolation [5]. There, the authors show that the critical percolation probability of a dense graph is $\frac{1}{\lambda_n}$ where λ_n is the largest eigenvalue of the adjacency matrix. In particular it follows that if two dense graphs are close in the cut-metric, then this threshold is also close. They also describe the limiting local structure of bond percolation on dense graph in terms of a branching process on the limiting graphon. While there is some resemblance to the branching process of Theorem 1.3, they are quite different.

1.5 About the Proof of Theorem 1.1

Our proof combines, for the first time to the best of our knowledge, two seemingly unrelated mathematical areas, Kirchhoff's electric network theory and Szemerédi's regularity lemma-like graph partitioning techniques. These two are shown here to work seamlessly together.

Kirchhoff's theory of electric networks [10] allows to compute the probability that a given edge $e = xy$ is in a UST of a connected graph. This probability is precisely the *effective electric resistance* between x and y , when we consider the graph as an electric network and let current flow from x to y , see Sect. 3.1.1 and (24). Since there is an edge connecting x to y , this quantity is always a number in $[0, 1]$. This is the starting point of our proof.

Next we use partition theory (Sect. 2) to decompose our graph G into a bounded number of dense expanders so that different expanders of the decomposition are connected by $\mathcal{O}(n^2)$ edges. Heuristically, the UST of G is close the union of independent USTs on each of these dense expanders.

Thus, it is natural to study electric theory on dense expanders. It is intuitive (and easy to prove) that if x and y are two vertices in graph, then the effective resistance between x and y is at least $\frac{1}{\deg(x)} + \frac{1}{\deg(y)}$ since at the most efficient scenario the electric current splits equally from x to all its neighbors and arrives to y equally from all of y 's neighbors. Of course this lower bound is not sharp—the graph could be the disjoint union of two large stars around x and y and an edge connecting x and y , in which case the resistance between x and y is 1.

However, when the graph is a dense expander, one can use random walks estimates and employ the fascinating and classical connection between random walks and electric networks, to deduce a corresponding approximate upper bound, see Corollary 3.3. The random walk estimate we prove (Lemma 3.2) states that if one starts a random walk on a dense expander from some vertex that is not x or y , then the probability that x is visited before y is $\frac{(1+\mathcal{O}(1))\deg(x)}{\deg(x)+\deg(y)}$.

It is now quite pleasant to observe that Rayleigh's monotonicity (21), which states that the electric resistance can only decrease by enlarging a network, shows that this upper bound on the resistance holds in each expander in the decomposition of G , and the matching lower bound holds for *most* edges of G since there are $\mathcal{O}(n^2)$ edges between components, see Lemma 3.3.

This explains why for most edges in G the probability that they are exhibited in the UST is the sum of the degree reciprocals. An iterative argument is presented in Sect. 3.3, employing the spatial Markov property of the UST (Proposition 3.1), to control the probability of events such as $B_{\mathcal{T}}(v, r) \cong T$. There are some delicate technicalities to overcome involving the "outside" effects of the decomposition. Once these are overcome, one reaches the discrete version $\text{Freq}(T; G)$ of the parameter $\text{Freq}(T; W)$ of Theorem 1.1 and we show that this parameter approximates the desired probability (Lemma 3.12). In Sect. 4 it is shown that $\text{Freq}(T; G)$ is close to $\text{Freq}(T; W)$, its continuous counterpart, if G is close to W in the cut-distance, concluding the proof.

1.6 Organisation of the Paper

We tried to write the paper so that it can be read by probabilists and graph theorists alike. For this reason we recall even concepts relatively well known to one of the communities in a pedestrian manner. Also, at places we try to convey an idea of a proof even when this idea is standard, but in only one of the two communities.

In Sect. 2 we introduce a suitable decomposition of dense graphs, into so-called linear expanders. In Sect. 3 we prove the discrete estimate approximating the probability of the

event $\mathbb{P}(B_{\mathcal{T}}(X, r) = T)$ by the discrete parameter $\text{Freq}(T; G)$. In Sect. 4 we show that $\text{Freq}(T; G)$ and $\text{Freq}(T; W)$ are close whenever G and W are close in the cut-metric. Lastly, in Sect. 5 we derive Theorem 1.5.

2 Decomposing Dense Graphs into Linear Expanders

2.1 Dense Expanders

Informally, a graph is a dense expander if whenever a vertex set and its complement are of linear size (in the order of the graph) then there are quadratically many edges between these two parts. So, a primal example of a dense graph that is *not* an expander is a disjoint union of two cliques of order $\frac{n}{2}$ with a perfect matching connecting them.

We give our definition of expansion for **loopless multigraphs**. That is, self-loops are not allowed, and two vertices may be connected by several edges. The quantity $e(A, B)$ counts all ordered pairs ab that form an edge, $a \in A$ and $b \in B$, including multiplicities. Note that each edge with both endvertices in $A \cap B$ contributes twice to $e(A, B)$. For a vertex v and a vertex set A , we write $\text{deg}(v, A) := e(\{v\}, A)$. We write $\text{deg}(v) := \text{deg}(v, V)$, where V is the vertex set of the (multi)graph. Note that with these conventions, we have $2e(G) = \sum_{v \in V} \text{deg}(v)$, and $\sum_{a \in A} \text{deg}(a, B) = e(A, B) = \sum_{b \in B} \text{deg}(b, A)$, for each vertex sets A and B .

Definition 2.1 We say that a loopless multigraph H is a γ -**expander** if for each $U \subseteq V(H)$, we have $e(U, V(H) \setminus U) \geq \gamma|U|(v(H) - |U|)$.

We will later use a simple observation. Removing edges from an expander can obviously render its expansion properties. However, if one removes edges touching only one vertex while leaving the degree of the vertex high, the expansion properties are not damaged by too much as the following simple proposition states.

Proposition 2.2 *Let G be a γ -expander loopless multigraph on m vertices and let $v \in V(G)$ be a vertex. Assume that the maximal number of edges between any two vertices in $V(G) \setminus \{v\}$ is at most $\ell \in \mathbb{N}$. Consider the graph G' obtained from G by erasing some set of edges emanating from v of size at most ℓm so that the degree of v in G' is at least γm . Assume also that $8\ell^2\gamma^{-2} \leq m$. Then G' is a $\frac{\gamma}{2}$ -expander.*

Proof We need to prove that for any $U \subseteq V(G)$,

$$e_{G'}(U, V(G) \setminus U) \geq \frac{\gamma}{2}|U|(m - |U|). \tag{9}$$

By symmetry it is enough to prove it when $|U| \leq m/2$. We proceed by considering two cases. In the first case we assume $|U| \geq 4\gamma^{-1}\ell$. Then, since we erased at most ℓm edges, we have

$$e_{G'}(U, V(G) \setminus U) \geq \gamma|U|(m - |U|) - \ell m \geq \frac{\gamma}{2}|U|(m - |U|),$$

since in this case $\gamma|U|(m - |U|) \geq 2\ell m$.

In the second case we assume that $|U| \leq 4\gamma^{-1}\ell$. If $U = \{v\}$, then (9) follows since the degree of v in G' is at least γm . If $v \notin U$, then the maximum number of edges that could have been erased from $e_G(U, V(G) \setminus U)$ is at most $\ell|U|$, hence

$$e_{G'}(U, V(G) \setminus U) \geq \gamma|U|(m - |U|) - 4\ell^2\gamma^{-1} \geq \frac{\gamma}{2}|U|(m - |U|),$$

since $\gamma|U|(m - |U|) \geq \gamma m$ and $8\ell^2\gamma^{-2} \leq m$. Lastly, if $v \in U$ and $|U| > 1$, then

$$e_{G'}(U, V(G)\setminus U) \geq e_G(U\setminus\{v\}, V(G)\setminus U \cup \{v\}) - \ell|U| \geq \frac{\gamma}{2}|U|(m - |U|),$$

by the same logic as above, concluding the proof. □

2.2 Expander Decomposition of Dense Graphs

The main result of this section, Theorem 2.7, asserts that each graphon that is close to a non-degenerate graphon can be decomposed into a bounded number expanders that are almost isolated from each other. In Definition 2.6 below we describe the expander decomposition that we actually use. Let us recall that the need of such a decomposition (rather than a single expander) stems from examples such as that of a disjoint union of two cliques of order $\frac{n}{2}$ with a perfect matching connecting them mentioned at the beginning of Sect. 2.1. Passing to a limit we see that in the graphon perspective, the perfect matching vanishes and we are left with two components. Therefore, we now introduce graphon counterparts to the notion of graph connectivity and components, and give their basic properties.

Definition 2.3 A graphon W on a ground space (Ω, μ) is **disconnected** if either $W = 0$ a.e. or there exists a subset $A \subseteq \Omega$ with $0 < \mu(A) < 1$ such that $W = 0$ a.e. on $A \times A^c$; otherwise W is **connected**.

We shall require a result of Bollobás, Janson and Riordan [6, Lemma 5.17] which enables us to decompose a graphon into (at most) countably many connected components.

Lemma 2.4 *Let $W : \Omega \times \Omega \rightarrow [0, 1]$ be a graphon. Then there exists a partition $\Omega = \bigcup_{i=0}^N \Omega_i$ into measurable subsets with $0 \leq N \leq \infty$ such that $\mu(\Omega_i) > 0$ for $i \geq 1$, the restriction of W to $\Omega_i \times \Omega_i$ is connected for each $i \geq 1$, and $W = 0$ a.e. on $(\Omega \times \Omega) \setminus \bigcup_{i=1}^N (\Omega_i \times \Omega_i)$.*

Bollobás et al. [5, Lemma 7] showed that connectivity implies an apparently stronger statement.

Lemma 2.5 *Let $W : \Omega^2 \rightarrow [0, 1]$ be a connected graphon, and let $0 < \alpha < \frac{1}{2}$ be given. There is some constant $\beta = \beta(W, \alpha) > 0$ such that $\int_{A \times A^c} W \geq \beta$ for every measurable subset $A \subseteq \Omega$ with $\alpha \leq \mu(A) \leq \frac{1}{2}$.*

We can now give our definition of expander decomposition.

Definition 2.6 Suppose that G is a loopless multigraph of order n . We say that $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$ is a $(\gamma, \eta, \varepsilon)$ -**expander decomposition** if

- (G1) $|V_0| \leq \varepsilon n$,
- (G2) for each $i \in [k]$ we have that $e(V_i, V \setminus V_i) \leq \eta|V_i|n$,
- (G3) for each $i \in [k]$ and each $U \subseteq V_i$, we have $e(U, V_i \setminus U) \geq \gamma|U||V_i \setminus U|$.

Theorem 2.7 *Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a nondegenerate graphon. Then for every $\varepsilon, \eta > 0$ there exist positive constants $\gamma = \gamma(W, \varepsilon, \eta)$, $\xi = \xi(W, \varepsilon, \eta)$ and $n_0 = n_0(W, \varepsilon, \eta)$ such that if G is a graph with $v(G) > n_0$ and $\delta_\varepsilon(G, W) < \xi$ then G admits a $(\gamma, \eta, \varepsilon)$ -expander decomposition.*

For the proof of Theorem 2.7, we shall need the following result. While we were not able to find an explicit reference, we consider this result folklore. To this end, we need the notion

of subgraphons which we introduce now. Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a graphon on a probability space (μ, Ω) . Similarly to the graph case, for a set $A \subseteq \Omega$ we have the notion of a **subgraphon of W induced by A** . This is the restricted function $W[A] := W|_{A \times A}$. In order for $W[A]$ to be a graphon, we always have to consider it together with the renormalized probability space $(\frac{\mu(\cdot)}{\mu(A)}, A)$.

Lemma 2.8 *Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a nondegenerate graphon. Suppose that we have a partition $\Omega = \Omega^* \sqcup \bigsqcup_{i=1}^k \Omega_i$ such that for each $i \in [k]$ we have that W is zero almost everywhere on $\Omega_i \times (\Omega \setminus \Omega_i)$. Then for every $\lambda > 0$ there exists a number $\xi > 0$ so that we have the following. If G is an n -vertex graph with $\delta_{\square}(G, W) < \xi$ then there exists a partition $V(G) = \bigsqcup_{i=0}^k V_i$ such that for each $i \in [k]$,*

- (a) $|V_0| = \mu(\Omega^*)n \pm \lambda n$, and $|V_i| = \mu(\Omega_i)n \pm \lambda n$,
- (b) $e_G(V_i, V(G) \setminus V_i) < \lambda n^2$,
- (c) $\delta_{\square}(G[V_i], W[\Omega_i]) < \lambda$.

Proof of Theorem 2.7 Suppose that we are given a graphon $W : \Omega^2 \rightarrow [0, 1]$ and two parameters $\varepsilon, \eta > 0$. By Lemma 2.4, there exists a partition $\Omega = \bigsqcup_{i=1}^K \Omega_i$ (with $K \leq \infty$) into components of W . Let $k \in \mathbb{N}$ be such that $\mu(\bigcup_{i=0}^k \Omega_i) \geq 1 - \frac{\varepsilon}{4}$. Set $\Omega^* := \bigcup_{i=k+1}^K \Omega_i$.

Let $\alpha = \min \left\{ \frac{\varepsilon}{6k}, \min_{i \in [k]} \frac{\mu(\Omega_i)}{40} \right\}$. Lemma 2.5 shows the existence of a positive constant $\beta = \beta(W, \alpha)$ such that

$$\iint_{A \times (\Omega_i \setminus A)} W \geq \beta \text{ for all } i \in [k] \text{ and all } A \subseteq \Omega_i \text{ with } \alpha \mu(\Omega_i) \leq \mu(A) \leq \mu(\Omega_i)/2. \tag{10}$$

Let γ, ξ and λ satisfy

$$\gamma = \min \left\{ \frac{\mu(\Omega_i)\eta}{48}, \frac{\mu(\Omega_i)\beta}{500} \right\} \text{ and } 0 < \xi \ll \lambda \ll \min \left\{ \beta, \min_{i \in [k]} \mu(\Omega_i)\eta \right\}. \tag{11}$$

Suppose that G is a graph given at the input of the proposition.

Let $V(G) = V'_0 \sqcup V'_1 \sqcup \dots \sqcup V'_k$ be a partition satisfying properties of Lemma 2.8 for the graphon W and its partition $\Omega = \Omega^* \sqcup \bigsqcup_{i=1}^k \Omega_i$, together with input error parameter ξ and output error parameter λ . We will modify this partition to obtain a $(\gamma, \eta, \varepsilon)$ -expander decomposition of G .

Lemma 2.8(a) gives that

$$|V'_0| \leq \frac{1}{2}\varepsilon n. \tag{12}$$

For each $i \in [k]$, we perform the following cleaning procedure. Let $U : \Omega_i^2 \rightarrow [0, 1]$ be a graphon representation of $G[V'_i]$ on the (renormalized) probability space Ω_i such that we have $\|W[\Omega_i] - U\|_{\square} < \lambda$. Such a representation exists by Lemma 2.8(c).

Let $P_i^0 := V'_i$ and $Q_i^0 := \emptyset$. Now, for $j = 1, 2, 3, \dots$ we proceed as follows. If there exists at least one set $X_i^j \subseteq P_i^{j-1}$ of size at most $\frac{3}{5}|P_i^{j-1}|$ with $e(X_i^j, P_i^{j-1} \setminus X_i^j) < \gamma|X_i^j|n$, then we take this set, and let $Q_i^j := X_i^1 \cup X_i^2 \cup \dots \cup X_i^j$, $P_i^j := V'_i \setminus Q_i^j$, and proceed with $j + 1$. If no set X_i^j exists, then we set $j(i) := j - 1$, $V_i := P_i^{j(i)}$, and terminate. Since the sets X_i^j ($j = 1, 2, \dots, j(i) - 1$) are nonempty, we will stop eventually.

For every $i \in [k]$ and every $j \in [j(i)]$, we have

$$e(Q_i^j, P_i^j) = \sum_{\ell=1}^j e(X_i^\ell, P_i^j) \leq \sum_{\ell=1}^j e(X_i^\ell, P_i^{\ell-1} \setminus X_i^\ell) \leq \gamma n \sum_{\ell=1}^j |X_i^\ell| = \gamma |Q_i^j| n. \tag{13}$$

Claim 2.7.1 For each $i \in [k]$,

$$|V'_i \setminus V_i| \leq 3\alpha n. \tag{14}$$

Proof Suppose to the contrary that $|V'_i \setminus V_i| \geq 3\alpha n$. Let $j \in \{0, 1, 2, \dots, j(i)\}$ be the largest index for which

$$|P_i^j| \geq \frac{1}{4} |V'_i|. \tag{15}$$

Now, there are two cases to consider. Either $|P_i^{j+1}| < \frac{1}{4} |V'_i|$ and then we have $|P_i^j| \leq (\frac{1}{4} + \frac{3}{5}) |V'_i|$ by the way we chose the set X_i^{j+1} . Another case is that $j = j(i)$, that is, we terminated in the step j . Then, by our assumption, $|V'_i \setminus P_i^j| = |V'_i \setminus V_i| \geq 3\alpha n$. Put together,

$$|Q_i^j| = |V'_i \setminus P_i^j| \geq \min \left\{ \left(1 - \frac{1}{4} - \frac{3}{5}\right) |V'_i|, 3\alpha n \right\} = 3\alpha n, \tag{16}$$

as $|V'_i| \geq \frac{1}{2} \mu(\Omega_i) n$ (by Lemma 2.8(a)) and $\alpha \leq \frac{1}{40} \mu(\Omega_i)$ (by (11)).

We learn from (13) that

$$e(Q_i^j, P_i^j) \leq \gamma |Q_i^j| n \stackrel{(15)}{\leq} 4\gamma |Q_i^j| |P_i^j| \cdot \frac{n}{|V'_i|} \leq 5\gamma |Q_i^j| |P_i^j| \cdot \frac{1}{\mu(\Omega_i)} \stackrel{(11)}{<} \frac{\beta}{100} |Q_i^j| |P_i^j|. \tag{17}$$

Let $\Lambda \subseteq \Omega_i$ represent the vertices of P_i^j . We have $\mu(\Lambda) \geq \frac{1}{4} \mu(\Omega_i) - \lambda \geq \frac{1}{5} \mu(\Omega_i)$, due to Lemma 2.8(a) and (15). Similarly, (16) gives $\mu(\Omega_i \setminus \Lambda) \geq \alpha \mu(\Omega_i)$. Thus, (10) applies. We have

$$\int_{\Lambda} \int_{\Omega_i \setminus \Lambda} U \geq \int_{\Lambda} \int_{\Omega_i \setminus \Lambda} W - \|W[\Omega_i] - U\|_{\square} \stackrel{(10)}{\geq} \beta - \lambda,$$

which contradicts (17). □

We have defined the sets V_1, V_2, \dots, V_k . Set $V_0 := V'_0 \cup \bigcup_{i=1}^k (V'_i \setminus V_i)$. Let us now check that $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$ is indeed a desired expander decomposition.

As for property (G1), we have

$$|V_0| = |V'_0| + \sum_{i=1}^k |V'_i \setminus V_i| \stackrel{(12), \text{cl.2.7.1}}{\leq} \frac{1}{2} \varepsilon n + 3\alpha kn \leq \varepsilon n.$$

For (G2), we first notice that

$$|V_i| \stackrel{(14)}{\geq} |V'_i| - 3\alpha n \geq (\mu(\Omega_i) - \lambda)n - 3\alpha n \geq \frac{1}{2} \mu(\Omega_i) n,$$

as $\alpha \leq \frac{1}{12} \mu(\Omega_i)$. Thus we find, as required,

$$\begin{aligned} e(V_i, V \setminus V_i) &\leq e(V'_i, V \setminus V'_i) + e(V'_i \setminus V_i, V_i) \stackrel{(13)}{\leq} \lambda n^2 + 4\gamma |V'_i \setminus V_i| n \stackrel{\text{cl.2.7.1}}{\leq} \lambda n^2 \\ &\quad + 12\gamma \alpha n^2 \leq \eta |V_i| n, \end{aligned}$$

where the last inequality holds since $|V_i| \geq \frac{1}{2} \mu(\Omega_i) n$, $\lambda \ll \mu(\Omega_i) \eta$, and $\gamma \leq \frac{\mu(\Omega_i) \eta}{48}$.

Finally, property (G3) follows immediately from the stopping condition. This completes our proof of Theorem 2.7. □

2.3 Properties of the Expander Decomposition

For the proof of Theorem 1.1 we argue that the majority of vertices do not “see” much beyond the component in the expander decomposition they belong to. This is formalized in the following definitions.

Definition 2.9 Suppose that G is a loopless multigraph of order n . Assume that $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$ is some expander decomposition of G . For a vertex v we write $i(v)$ for the unique $i \in \{0, 1, \dots, k\}$ such that $v \in V_i$. Given $\alpha > 0$ and $\varepsilon > 0$ we say that $v \in V \setminus V_0$ is (α, ε) -**good** with respect to the decomposition if the following hold:

- (a) $\deg(v) \geq \Omega(\varepsilon n)$,
- (b) $\deg(v; V_{i(v)}) \geq (1 - O(\varepsilon^2)) \deg(v)$,
- (c) $\sum_{u \in V_{i(v)}, u \sim v} \left(\frac{1}{\deg(u; V_{i(v)})} - \frac{1}{\deg(u)} \right) \leq O(\alpha^{1/2})$,
- (d) $\sum_{u \in V_{i(v)}, u \sim v} \frac{1}{\deg(u; V_{i(v)})} \leq O(\alpha^{-1/4})$.

Definition 2.10 Suppose that G is a loopless multigraph of order n . Given numbers $\beta, \alpha, \gamma > 0$ and $\varepsilon \in (0, \alpha)$, we say that G has an $(\beta, \alpha, \gamma, \varepsilon)$ -**good-decomposition** if

- (1) G admits a $(\gamma, \varepsilon^5, \varepsilon^5)$ -expander decomposition $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$, and
- (2) At least $(1 - O(\alpha^{1/4}))n$ vertices of G are (α, ε) -good.
- (3) At least $(1 - O(\beta))n$ vertices of G have degree at least $\Omega(\alpha^{1/10}n)$.

Next we refine Theorem 2.7.

Lemma 2.11 For any $\beta > 0$ and any nondegenerate graphon $W : \Omega^2 \rightarrow [0, 1]$, there exist $\alpha, \varepsilon, \gamma, \xi > 0$ with $\beta \gg \alpha \gg \varepsilon \gg \gamma \gg \xi$ such that if G is a simple graph on $n \geq \xi^{-1}$ vertices with $d_{\square}(G, W) \leq \xi$, then G has a $(\beta, \alpha, \varepsilon, \gamma)$ -good-decomposition.

Proof Let β and W be given. Since W is nondegenerate, there exists $\alpha > 0$ such that any m -vertex graph (m is arbitrary) that is ξ_1 -close (for $\xi_1 > 0$ sufficiently small) to W has at least $(1 - \beta)m$ of degrees at least $\alpha^{1/10}m$, so requirement (3) of Definition 2.10 holds. Similarly, we can find constants $\varepsilon \in (0, \alpha^{20})$ and $\xi_2 > 0$ such that any m -vertex graph (m is arbitrary) which is ξ_2 -close to W has at most αm vertices of degrees at most εm . We apply Theorem 2.7 with input ε^5 and $\eta = \varepsilon^5$ and retrieve $\gamma > 0$ and $\xi_3 > 0$. We set $\xi := \min(\xi_1, \xi_2, \xi_3)$. Suppose now that $G = (V, E)$ is a graph satisfying the assumptions of the lemma. Theorem 2.7 readily gives item (1) of Definition 2.10.

To show item (2), we first note that by property (G2) of the expander decomposition we have that $e(V_i, V \setminus V_i) \leq \varepsilon^5 n |V_i|$ for all $i \in [k]$. By summing over i we deduce that

$$\sum_{i=1}^k e(V_i, V \setminus V_i) \leq \varepsilon^5 n^2. \tag{18}$$

Denote by S the set of vertices of $V \setminus V_0$ violating (b) of Definition 2.9 using 1 as the implicit constant in the term $O(\varepsilon^2)$,⁴ that is,

$$S = \{v \in V \setminus V_0 : \deg(v; V \setminus V_{i(v)}) \geq \varepsilon^2 \deg(v)\}.$$

⁴ Note that later, in the proof of Lemma 3.10, we shall be forced to use larger implicit constants in (b) of Definition 2.9.

Then by (18) we have that

$$\varepsilon^5 n^2 \geq \sum_{v \in S} \deg(v; V \setminus V_{i(u)}) \geq \varepsilon^2 \sum_{v \in S} \deg(v).$$

Therefore $\sum_{v \in S} \deg(v) \leq \varepsilon^3 n^2$, from which we learn that

$$|S \cap \{v : \deg(v) \geq \varepsilon n\}| \leq \varepsilon^2 n. \tag{19}$$

We deduce that

$$|\{v \in V \setminus V_0 : \deg(v) \leq \varepsilon n \text{ or } \deg(v; V \setminus V_{i(v)}) \geq \varepsilon^2 \deg(v)\}| \leq (\alpha + \varepsilon^2)n. \tag{20}$$

Next, for $i \in [k]$ we write

$$\sum_{v \in V_i} \sum_{u \in V_i, u \sim v} \left(\frac{1}{\deg(u; V_i)} - \frac{1}{\deg(u)} \right) = |V_i| - \sum_{u \in V_i} \frac{\deg(u; V_i)}{\deg(u)} = \sum_{u \in V_i} \frac{\deg(u; V \setminus V_i)}{\deg(u)}.$$

We sum this over $i \in [k]$ and get that

$$\begin{aligned} & \sum_{v \in V \setminus V_0} \sum_{u \in V_{i(v)}, u \sim v} \left(\frac{1}{\deg(u; V_i)} - \frac{1}{\deg(u)} \right) \\ &= \sum_{u \in V \setminus V_0} \frac{\deg(u; V \setminus V_{i(u)})}{\deg(u)} \leq (\alpha + \varepsilon^2)n + \varepsilon^2 n = O(\alpha n), \end{aligned}$$

where we bounded the ratio by 1 for those vertices counted in (20), and by ε^2 for the vertices that were not. From the last inequality we deduce that there cannot be more than $\Omega(\alpha^{1/2}n)$ vertices $v \in V \setminus V_0$ such that requirement (c) in the definition of (α, ε) -good is not satisfied.

Lastly, to show (d), we have that

$$\sum_{i=1}^k \sum_{v \in V_i} \sum_{u \in V_i; u \sim v} \frac{1}{\deg(u; V_i)} = \sum_{i=1}^k |V_i| \leq n,$$

therefore there cannot be more than $\Omega(\alpha^{1/4}n)$ vertices $v \in V \setminus V_0$ such that (d) is violated. This concludes our proof. \square

Part (2) of Definition 2.10 asserts that there are many (α, ε) -good vertices in the graph, yet there could still be components of the decomposition V_i in which the majority of their vertices are not (α, ε) -good. These cannot occupy too much of the mass. Indeed, for some $i \in [k]$, we say the set V_i is (α, ε) -**big** if at least $(1 - O(\alpha^{1/8}))|V_i|$ of its vertices are (α, ε) -good, and $e(G[V_i]) \geq \Omega(\alpha^{1/9}|V_i|n)$.

Proposition 2.12 *Suppose that G is a graph with n vertices that has a $(\beta, \alpha, \gamma, \varepsilon)$ -good-decomposition (as in Definition 2.10). Then*

$$\sum_{i \in [k]: V_i \text{ is } (\alpha, \varepsilon)\text{-big}} |V_i| \geq (1 - O(\beta^{1/8}))n.$$

Proof Let I_1 be the indices $i \in [k]$ such that V_i has $\Omega(\alpha^{1/8}n)$ vertices that are not (α, ε) -good. Since this is a $(\beta, \alpha, \gamma, \varepsilon)$ -good-decomposition we have that the total number of not (α, ε) -good vertices is $O(\alpha^{1/4}n)$. Hence,

$$\sum_{i \in I_1} |V_i| \leq O(\alpha^{1/8}n) = O(\beta^{1/8}n).$$

Next, let I_2 be the indices $i \in [k]$ such that $e(G[V_i]) = O(\alpha^{1/9}|V_i|n)$. Put $V' = \bigcup_{i \in I_2} V_i$. We have

$$\begin{aligned} \sum_{v \in V'} \deg(v) &\leq 2 \sum_{i \in I_2} e(G[V_i]) + \sum_{i \in I_2} e(V_i, V \setminus V_i) \leq O(\alpha^{1/9}|V'|n) + \varepsilon^5|V'|n \\ &= O(\alpha^{1/9}|V'|n). \end{aligned}$$

If $|V'| = \Omega(\beta^{1/8}n)$, then by property (3) in Definition 2.10, we may bound $\sum_{v \in V'} \deg(v)$ from below by $\Omega(\alpha^{1/10}|V'|n)$, giving a contradiction to the last estimate. The proof is concluded since $\sum_{i \in [k]} |V_i| \geq (1 - \varepsilon^5)n$. \square

3 Local Neighborhoods of the UST via Electric Networks

3.1 Preliminaries

Suppose that $G = (V, E)$ is a loopless multigraph. We denote by $\{X_t\}_{t \geq 0}$ the simple random walk starting from some (possibly random) vertex X_0 . That is, $\{X_t\}_{t \geq 0}$ is a Markov chain with state space V and transition matrix $p(x, y) = \frac{e(\{x\}, \{y\})}{\deg(x)}$. We denote by \mathbb{P}_v the probability measure of the simple random walk started at a vertex v . We will frequently use two stopping times: the hitting time of a vertex v is the random variable $\tau_v := \min\{t \geq 0 : X(t) = v\}$ and the hitting time after zero of a vertex v is the random variable $\tau_v^+ := \min\{t > 0 : X(t) = v\}$. Clearly when $X_0 \neq v$ these two random times are equal.

3.1.1 Effective Resistance

Our analysis relies on the relation between random walks, USTs and the theory of electrical networks. We briefly recall here the basic theory we will use and refer the reader to [18, Chapter 2] for a comprehensive study. Given a graph $G = (V, E)$ we write \vec{E} for the set of directed edges of size $2|E|$ which contain each edge of E in both direction. Given two distinct vertices u, v we say that an antisymmetric function $f : \vec{E} \rightarrow \mathbb{R}$ is a **flow** from u to v if for each vertex $w \notin \{u, v\}$ the sum of f over edges outgoing from w is zero. A flow is called **unit** if the sum of f over edges outgoing from u is 1. The **effective resistance** $\mathcal{R}_{\text{eff}}(u \leftrightarrow v; G)$ between u and v is defined as the minimum energy $\mathcal{E}(f) = \sum_{e \in E} f(e)^2$ of any unit flow f from u to v . If u and v are not in the same connected component, we define effective resistance between them to be ∞ . When it is clear what the underlying graph G is we simply write $\mathcal{R}_{\text{eff}}(u \leftrightarrow v)$. From this definition it is immediate that if G' is a subgraph of G , then

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v; G) \leq \mathcal{R}_{\text{eff}}(u \leftrightarrow v; G'). \tag{21}$$

The latter inequality is also known as Rayleigh’s monotonicity law. The *discrete Dirichlet’s principle* gives a dual definition of the effective resistance in terms of functions on the vertices. It states that

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v)^{-1} = \inf \left\{ \sum_{(x,y) \in E} (h(x) - h(y))^2 : h : V \rightarrow \mathbb{R}, h(u) = 0, h(v) = 1 \right\}, \tag{22}$$

see [18, Exercise 2.13]. We will also a basic probabilistic interpretation of the effective resistance which can be found in [18, Chapter 2]:

$$\mathbb{P}_u [\tau_v < \tau_u^+] = \frac{1}{\deg(u)\mathcal{R}_{\text{eff}}(u \leftrightarrow v)}. \tag{23}$$

3.1.2 Uniform Spanning Trees

There is a fundamental connection between the uniform spanning tree and electric networks due to Kirchhoff [10]. Let G be connected loopless multigraph and $e = xy$ be an edge of the graph. As before we denote by \mathcal{T} a UST of G . Kirchhoff’s formula [10] (see also [18, Chapter 4]) states that for any edge $e = (x, y)$ of G we have

$$\mathbb{P}(e \in \mathcal{T}) = \mathcal{R}_{\text{eff}}(x \leftrightarrow y). \tag{24}$$

Let S be a subset of $E(G)$. We would like to condition on events of the form $S \subseteq \mathcal{T}$ or $S \cap \mathcal{T} = \emptyset$. We denote by G/S the loopless multigraph obtained from G by contracting the edges of S and erasing any loops that has been formed, and by $G - S$ the graph G with the edges of S erased. The following is an easy and classical observation, see [18, Chapter 4].

Proposition 3.1 *Let G be a connected loopless multigraph and S a subset of edges of G .*

- (1) *If $G - S$ is connected, then the UST \mathcal{T} of G conditioned on $S \cap \mathcal{T} = \emptyset$ has the distribution of the UST on $G - S$.*
- (2) *If S does not contain a cycle, then the UST \mathcal{T} of G conditioned on $S \subseteq \mathcal{T}$ has the distribution of $S \cup \mathcal{T}_{G/S}$ where $\mathcal{T}_{G/S}$ is a UST of G/S .*

3.1.3 Mixing Time

Let $G = (V, E)$ be a finite connected loopless multigraph and consider the lazy simple random walk on it, that is, the Markov chain on the vertex set V with transition probability $p(x, y) = \frac{e(\{x\}, \{y\})}{2 \deg(x)}$ whenever $x \neq y$ and $p(x, x) = 1/2$ for any vertex x . Let π be the stationary distribution $\pi(x) = \deg(x)/2|E|$ and for each two disjoint subsets of vertices A, B we write

$$Q(A, B) = \sum_{x \in A, y \in B} \pi(x)p(x, y) = e(A, B)/4|E|.$$

The Cheeger constant Φ_* is defined as

$$\Phi_* = \min_{S: \pi(S) \leq \frac{1}{2}} \frac{Q(S, V \setminus S)}{\pi(S)},$$

where $\pi(S) = \sum_{x \in S} \pi(x)$. This “bottleneck” ratio is frequently used to control the spectral gap of the lazy random walk from which we may bound its mixing time. Let $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n \geq 0$ be the eigenvalues of the transition matrix p (we have $\lambda_1 > \lambda_2$ since G is connected, and $\lambda_n \geq 0$ since the chain is lazy, see [14]). A result by Jerrum and Sinclair [20], Lawler and Sokal [13] and Alon and Milman [3] states that

$$\Phi_*^2/2 \leq 1 - \lambda_2 \leq 2\Phi_*. \tag{25}$$

Assume now that G is a γ -expander as in Definition 2.1 and that the number of parallel edges between any two vertices is at most $f \geq 1$. Then the degree of each vertex is at most

fn , hence $\pi(S) \leq f|S|n/2|E|$ for any $S \subseteq V$. Similarly, for any $S \subseteq V$ we have that $\pi(V \setminus S) \leq fn|V \setminus S|/2|E|$, so if $\pi(V \setminus S) \geq 1/2$ we get that $|V \setminus S| \geq |E|/fn$. Since the minimum degree in a γ -expander is at least $\gamma(n - 1)$ we get that if $\pi(V \setminus S) \geq 1/2$, then $|V \setminus S| \geq \gamma(n - 1)/2f$. Putting all this together we get that if G is a γ -expander and $n \geq 2$, then

$$\Phi_* \geq \frac{\gamma|S||V \setminus S|}{2|S|fn} \geq \frac{\gamma^2}{8f^2},$$

from which we get by (25) that

$$1 - \lambda_2 \geq \frac{c\gamma^4}{f^4}, \tag{26}$$

where $c = 1/128$. Recall that the **total variation** distance $\|\mu - \nu\|_{TV}$ between two probability measures μ and ν on the same probability space is defined to be $\sup_A |\mu(A) - \nu(A)|$, where the sup is ranging over all events A . For $\varepsilon > 0$ the ε -mixing-time $T_{\text{mix}}(\varepsilon)$ of the chain is defined as

$$T_{\text{mix}}(\varepsilon) = \min \{t : \|p^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq \varepsilon \text{ for all } x \in V\}.$$

The mixing time and the spectral gap are related via the following statement, see [14, Theorem 12.4],

$$T_{\text{mix}}(\varepsilon) \leq \frac{1}{1 - \lambda_2} \left[\frac{1}{2} \log \frac{1}{\min_{x \in V} \pi(x)} + \log(1/2\varepsilon) \right].$$

Since $f \geq 1$ bounds the maximal number of parallel edges between any two vertices, we get that $|E| \leq fn^2$ and hence $\pi(x) \geq \gamma(n - 1)/2|E| \geq c\gamma/(fn)$ for some universal constant $c > 0$. We deduce from this, the above bound on $T_{\text{mix}}(\varepsilon)$ and (26) that

$$T_{\text{mix}}(\varepsilon) \leq Cf^4\gamma^{-4} [\log n + \log f/\gamma + \log \varepsilon^{-1}], \tag{27}$$

for some universal constant $C > 0$.

3.2 Random Walks on Dense Expanders

Lemma 3.2 *Suppose that H is a loopless multigraph on n vertices that is a γ -expander and that the number of parallel edges between any two vertices is at most $f \geq 1$. Then for any two distinct nodes u and v and a node $w \neq v$ we have that*

$$\mathbb{P}_w [\tau_v < \tau_u^+] = \frac{\text{deg}(v)}{\text{deg}(u) + \text{deg}(v)} + \mathcal{O}(\gamma^{-7} f^7 n^{-1} \log(n)).$$

Proof Let us assume that there are no edges between u and v —this can only matter for the assertion of the statement when $w = u$ and in this case affect the estimate by the probability that this edge is traversed on the first step of the random walk; the latter probability is at most $\mathcal{O}(\gamma^{-1} f/n)$ the assumption, since $\text{deg}(u) \geq \gamma n$. This error is swallowed in the error estimate of the lemma.

We denote by \mathbb{P}^{lazy} the lazy random walk on the graph, that is, the random walk that with probability $1/2$ stays put and otherwise jumps to a uniformly chosen neighbor. It is clear that if w is a vertex such that $w \notin \{u, v\}$ then $\mathbb{P}_w^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_w(\tau_v < \tau_u^+)$. We will first show, via a coupling argument, that for any two vertices $w_1, w_2 \notin \{u, v\}$ we have

$$\mathbb{P}_{w_1}^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{w_2}^{\text{lazy}}(\tau_v < \tau_u^+) + \mathcal{O}(\gamma^{-6} f^6 n^{-1} \log(n)). \tag{28}$$

Indeed, let $\{X_t\}_{t \geq 0}$ and $\{Y_t\}_{t \geq 0}$ be two lazy simple random walks starting at w_1 and w_2 , respectively. We put $\varepsilon = n^{-1}$ in (27) and bound $\log(f/\gamma)$ by f/γ and get that if $T = C\gamma^{-5}f^5 \log(n)$, then $\|X_T - \pi\|_{TV} \leq n^{-1}$ and the same estimate holds for Y_T , hence $\|X_T - Y_T\|_{TV} \leq 2n^{-1}$ (where by $\|X_T - Y_T\|_{TV}$ we mean the total variation distance between the laws of X_T and Y_T).

By [14, Proposition 4.7] we deduce that we can couple the walks $\{X_t\}$ and $\{Y_t\}$ so that $X_t = Y_t$ with probability at least $1 - 2n^{-1}$. If this occurs we continue the coupling so that $X_t = Y_t$ for all $t \geq T$ by using the same random neighbor at each step of the walk. Thus, if both walks have not visited u or v between time 1 and T , then the event $\{\tau_v < \tau_u^+\}$ occurs for $\{X_t\}$ if and only if it occurs for $\{Y_t\}$. Since the minimal degree of H is at least $\gamma(n - 1)$ and the maximal number of parallel edges between any two vertices is at most f we learn that the probability of visiting u or v between time 1 and T is at most $Tf/\gamma(n - 1) \leq 2C\gamma^{-6}f^6n^{-1} \log(n)$, concluding the proof of (28).

We now continue the proof of the lemma for the case that $w = u$. If the walker starts at u and $\tau_v < \tau_u^+$, then it cannot be lazy in the first step. Hence

$$\mathbb{P}_u^{\text{lazy}}(\tau_v < \tau_u^+) = \frac{1}{2}\mathbb{P}_u(\tau_v < \tau_u^+). \tag{29}$$

Consider now the Markov chain on the two states $\{u, v\}$ with transition probabilities

$$p(u, v) = \mathbb{P}_u^{\text{lazy}}(\tau_v < \tau_u^+), \quad p(v, u) = \mathbb{P}_v^{\text{lazy}}(\tau_u < \tau_v^+),$$

with $p(u, u) = 1 - p(u, v)$ and $p(v, v) = 1 - p(v, u)$. This is the lazy random walk “watched” on the vertices u and v . By summing over paths it is immediate that

$$\text{deg}(u)p(u, v) = \text{deg}(v)p(v, u). \tag{30}$$

In order to visit v before returning to u , the lazy walker must walk to a random neighbor in the first step, so

$$p(u, v) = \frac{1}{2}\mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+), \tag{31}$$

where $\mathbb{P}_{N(u)}^{\text{lazy}}$ indicates a uniform starting position from the set $N(u)$ of neighbors of u . Similarly, when starting from v , in order to return to v before visiting u , the lazy walker can either stay put on the first step, or jump to a uniform neighbor of v and from there visit v before u , thus

$$p(v, v) = \frac{1}{2} + \frac{1}{2}\mathbb{P}_{N(v)}^{\text{lazy}}(\tau_v < \tau_u^+). \tag{32}$$

Since we assumed there are no edges between u and v , by (28) we have that

$$\mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+) + \mathcal{O}(\gamma^{-6}f^6n^{-1} \log(n)).$$

This together with (31) and (32) gives that $p(u, v) + p(v, u) = 1/2 + \mathcal{O}(\gamma^{-6}f^6n^{-1} \log(n))$. Together with (30) and the fact that all degrees are at least $\gamma(n - 1)$ and at most fn gives that

$$p(u, v) = \frac{\text{deg}(v)}{2(\text{deg}(u) + \text{deg}(v))} + \mathcal{O}(\gamma^{-7}f^7n^{-1} \log(n)),$$

concluding the proof of lemma when $w = u$ by (29). The proof for any $w \notin \{u, v\}$ can now be completed easily. By (28) and our assumption that there are no edges between u and v shows that

$$\mathbb{P}_w^{\text{lazy}}(\tau_v < \tau_u^+) = \mathbb{P}_{N(u)}^{\text{lazy}}(\tau_v < \tau_u^+) + \mathcal{O}(\gamma^{-6} f^6 n^{-1} \log n),$$

and so the lemma follows by (31). □

Corollary 3.3 *Suppose that H is a loopless multigraph on n vertices that is a γ -expander and that the number of parallel edges between any two vertices is at most $f \geq 1$. Then for any two vertices $u \neq v$*

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) = (1 + \mathcal{O}(\gamma^{-8} f^8 n^{-1} \log(n))) \left(\frac{1}{\deg(u)} + \frac{1}{\deg(v)} \right).$$

Proof Follows immediately by (23) and Lemma 3.2 together with the fact that all degrees are at least $\gamma(n - 1)$ and at most fn . □

We now extend Corollary 3.3 to the setting of a general dense graph.

Lemma 3.4 *Suppose that G is a loopless multigraph with n vertices given together with a $(\gamma, \varepsilon^5, \varepsilon^5)$ -expander decomposition $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$. Assume further that the maximal number of parallel edges among any two pairs of vertices is at most $f \geq 1$. Assume that $\gamma^{-8} f^8 n^{-1} \log n \leq \varepsilon$. Let $i \in [k]$ and $u \neq v$ be two distinct vertices of V_i . Then*

$$(1 - \mathcal{O}(\varepsilon)) \left(\frac{1}{\deg(u)} + \frac{1}{\deg(v)} \right) \leq \mathcal{R}_{\text{eff}}(u \leftrightarrow v) \leq (1 + \mathcal{O}(\varepsilon)) \left(\frac{1}{\deg(u; V_i)} + \frac{1}{\deg(v; V_i)} \right),$$

and if in addition u and v are (α, ε) -good, then

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) = (1 + \mathcal{O}(\varepsilon)) \left(\frac{1}{\deg(u)} + \frac{1}{\deg(v)} \right).$$

Proof Since $G[V_i]$ is a γ -expander on at least γn vertices, by Corollary 3.3 together with Rayleigh’s monotonicity (21) we have

$$\mathcal{R}_{\text{eff}}(u \leftrightarrow v) \leq (1 + \mathcal{O}(\gamma^{-8} f^8 n^{-1} \log(n))) \left(\frac{1}{\deg(u; V_i)} + \frac{1}{\deg(v; V_i)} \right),$$

giving the upper bound of the first assertion of the lemma. The upper bound of the second assertion immediately follows using the part (b) of Definition 2.9.

For the lower bound we will use Dirichlet’s principle (22) and let $h : V(G) \rightarrow [0, 1]$ be the function assigning $h(v) = 1$, $h(u) = 0$ and for any vertex $x \notin \{u, v\}$ we put $h(x) = \deg(v)/(\deg(u) + \deg(v))$. By our assumption there are at most f edges (x, y) such that $x = u$ and $y = v$ in which $h(y) - h(x) = 1$. Next, there are at most $\deg(u)$ edges (x, y) for which $x = u$ and $h(y) - h(x) = \deg(v)/(\deg(u) + \deg(v))$. Similarly, there are at most $\deg(v)$ edges (x, y) for which $y = v$ and $h(y) - h(x) = \deg(u)/(\deg(u) + \deg(v))$. All other edges (x, y) of the graph have $h(x) - h(y) = 0$. Hence by (22) we have

$$\begin{aligned} \mathcal{R}_{\text{eff}}(u \leftrightarrow v)^{-1} &\leq f + \frac{\deg(u) \deg(v)^2 + \deg(v) \deg(u)^2}{(\deg(u) + \deg(v))^2} \\ &\leq \frac{\deg(u) \deg(v)}{\deg(u) + \deg(v)} (1 + \gamma^{-4} f^2 n^{-1}), \end{aligned}$$

where we used the fact that since each V_i is a γ -expander, its cardinality must be $\Omega(\gamma n)$ and hence $\deg(u)$ and $\deg(v)$ can be bounded below by $\Omega(\gamma^2 n)$ and above by fn . This gives the required lower bound. □

3.3 The Density of Fixed Trees in the UST

The main result of this section, Lemma 3.12, allows express the frequency of a given fixed rooted tree T in a graph using a discrete analogue of the parameter Freq , see Definition 3.6 below.

Let us introduce some definitions and a setting that will be used throughout this section. In what follows we are always given an arbitrary $\beta > 0$ and a nondegenerate graphon W . We then apply Lemma 2.11 and extract the corresponding $\alpha, \varepsilon, \gamma$ and ξ so that if G is a connected graph on n vertices and n is sufficiently large (as a function of α, ε and γ) and $\delta_{\square}(G, W) \leq \xi$, then G has a $(\beta, \alpha, \gamma, \varepsilon)$ -good decomposition as in Definition 2.10. We denote the given expander decomposition of G by $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$. In light of the quantification of Lemma 2.11 we may assume that

$$\beta \gg \alpha \gg \varepsilon \gg \gamma \gg \xi \gg n^{-1}.$$

Next, let T be a finite rooted tree of height r and ℓ vertices denoted by $1, \dots, \ell$ so that 1 is the root of T . Given a graph G we say that ℓ distinct vertices (v_1, \dots, v_ℓ) of G are **compatible with T** if the pairs

$$T(v_1, \dots, v_\ell) := \{(v_q, v_t) : (q, t) \in E(T)\},$$

are all edges of G . Without loss of generality we may assume that the numbering $\{1, \dots, \ell\}$ of the vertices of T is such that there exists some $p \in \{2, \dots, \ell\}$ such that the vertices of distance r from the root (which all must be leaves) are p, \dots, ℓ .

Definition 3.5 Assume the setting as introduced at the beginning of Sect. 3.3. Given some fixed $\ell \geq 1$ and $i \in [k]$ we say that an ℓ -tuple of distinct vertices of G are **i -pure** if each vertex in the ℓ -tuple belongs to V_i and is (α, ε) -good.

Next we define $\text{Freq}(T; G)$ which is the discrete analogue of $\text{Freq}(T; W)$. Note the notation $\text{Freq}(T; G)$ does not reflect the fact that this parameter depends on the partition, and not just on the graph G .

Definition 3.6 Assume the setting as introduced at the beginning of Sect. 3.3. For each $i \in [k]$ we define

$$\text{Freq}(T; G, i) := |\text{Stab}_T|^{-1} \sum_{\substack{(v_1, \dots, v_\ell) \\ i\text{-pure} \\ \text{compatible with } T}} |V_i|^{-1} \exp\left(-\sum_{j=1}^{p-1} b(v_j)\right) \frac{\sum_{j=p}^{\ell} \text{deg}(v_j)}{\prod_{j=1}^{\ell} \text{deg}(v_j)},$$

where

$$b(v) = \sum_{u \in V_i(v), u \sim v} \frac{1}{\text{deg}(u)}. \tag{33}$$

Note that (33) is a graph counterpart to the graphon quantity defined in (2).

Finally, let

$$\text{Freq}(T; G) := \sum_{i \in [k]: V_i \text{ is } (\alpha, \varepsilon)\text{-big}} \frac{|V_i|}{n} \cdot \text{Freq}(T; G, i). \tag{34}$$

We denote by \mathcal{T} a sample of the UST of G and by $B_{\mathcal{T}}(v, r)$ the graph-distance ball in \mathcal{T} of radius r around $v \in V(G)$ and we think about it as a subset of edges of \mathcal{T} . We will begin our proof with estimating the probability that $B_{\mathcal{T}}(v, r)$ is manifested on an i -pure ℓ -tuple of vertices (as in Definition 3.5) that are compatible with T ; later we will see that that all other manifestations of $B_{\mathcal{T}}(v, r)$ are negligible. For an i -pure ℓ -tuple that is compatible with T we write $B_{\mathcal{T}}(v_1, r) \stackrel{\text{pure}}{\cong} T(v_1, \dots, v_{\ell})$ for the event

- the edges $T(v_1, \dots, v_{\ell})$ are in \mathcal{T} , and,
- for each $1 \leq j \leq p - 1$, the edges of emanating from v_j that are not in $T(v_1, \dots, v_{\ell})$ and have both endpoints in V_i are not in \mathcal{T} .

Lemma 3.7 *Assume the setting as introduced at the beginning of Sect. 3.3, and let \mathcal{T} be a UST of G . Let T be a fixed rooted tree with $\ell \geq 2$ vertices $1, \dots, \ell$ and height $r \geq 1$ (as usual T is rooted at 1). Assume that the vertices at height r of T are $\{p, \dots, \ell\}$ for some $2 \leq p \leq \ell$. Then for any $i \in [k]$ and any i -pure ℓ -tuple (v_1, \dots, v_{ℓ}) that is compatible with T we have that*

$$\mathbb{P}(B_{\mathcal{T}}(v_1, r) \stackrel{\text{pure}}{\cong} T(v_1, \dots, v_{\ell})) = (1 + \mathcal{O}(\ell\alpha^{1/4})) \exp\left(-\sum_{j=1}^{p-1} b(v_j)\right) \cdot \frac{\sum_{j=p}^{\ell} \text{deg}(v_j)}{\prod_{j=1}^{\ell} \text{deg}(v_j)},$$

where $b(v)$ is defined in (33).

Proof Assume that n is large enough as in Lemma 3.4. We first show by induction that the probability that $T(v_1, \dots, v_{\ell}) \subseteq E(\mathcal{T})$ is

$$(1 + \mathcal{O}(\ell\varepsilon)) \frac{\sum_{j=1}^{\ell} \text{deg}(v_j)}{\prod_{j=1}^{\ell} \text{deg}(v_j)}. \tag{35}$$

Indeed, when T contains only one edge this statement follows immediately from Kirchoff’s formula (24) and Lemma 3.4. If T has more than one edge, assume without loss of generality that ℓ is a leaf in T at distance r from the root and that (q, ℓ) is an edge of T . We then use the induction hypothesis on the tree $T \setminus \{\ell\}$, which has $\ell - 1$ vertices. We condition on the event $T(v_1, \dots, v_{\ell}) \setminus \{(v_q, v_{\ell})\} \subseteq E(\mathcal{T})$ and contract these $\ell - 1$ vertices to a single vertex of degree $\text{deg}(v_1) + \dots + \text{deg}(v_{\ell-1})$. We shall make use of Proposition 3.1 when working with the contracted graph. Since the contracted multigraph is still a γ -expander with at most ℓ parallel edges between any two vertices, and the vertices (v_1, \dots, v_{ℓ}) are all (α, ε) -good, we may apply Lemma 3.4 and Kirchoff’s formula (24) to get that the probability that the edge (v_q, v_{ℓ}) of G is in \mathcal{T} is

$$(1 - \mathcal{O}(\varepsilon)) \left(\frac{1}{\text{deg}(v_{\ell})} + \frac{1}{\text{deg}(v_1) + \dots + \text{deg}(v_{\ell-1})} \right).$$

By our induction hypothesis we get that the required probability is

$$(1 - \mathcal{O}((\ell - 1)\varepsilon))(1 - \mathcal{O}(\varepsilon)) \frac{\sum_{q=1}^{\ell-1} \text{deg}(v_q)}{\prod_{q=1}^{\ell-1} \text{deg}(v_q)} \left(\frac{1}{\text{deg}(v_{\ell})} + \frac{1}{\text{deg}(v_1) + \dots + \text{deg}(v_{\ell-1})} \right),$$

concluding the proof of (35).

We now condition on the event $T(v_1, \dots, v_{\ell}) \subseteq E(\mathcal{T})$ and turn to compute the probability that all other edges of G emanating from v_1, \dots, v_{p-1} which have both endpoints in V_i are

not in \mathcal{T} . That is, the event that $E_{v_j} \cap \mathcal{T} = \emptyset$ for all $1 \leq j \leq p - 1$, where E_{v_j} are all the edges of G between v_j and $V_i \setminus \{v_1, \dots, v_\ell\}$.

Let j be an integer $1 \leq j \leq p - 1$ and condition additionally on all the edges of $(E_{v_1} \cup \dots \cup E_{v_{j-1}}) \cap \mathcal{T} = \emptyset$ (if $j = 1$ there is no further conditioning). We will prove that under this conditioning the probability that $E_{v_j} \cap \mathcal{T} = \emptyset$ is

$$(1 + O(\alpha^{1/4})) \exp(-b(v_j)) \cdot \frac{\sum_{q=j+1}^{\ell} \deg(v_q)}{\sum_{q=j}^{\ell} \deg(v_q)}. \tag{36}$$

Thus, once (36) is proved, by multiplying it over $j = 1, \dots, p - 1$, we get that conditioned on \mathcal{O} and on $E_G(\mathcal{T}) \subseteq E(\mathcal{T})$, the probability that all the required edges are not in \mathcal{T} is

$$(1 + O(\ell\alpha^{1/4})) \exp\left(-\sum_{j=1}^{p-1} b(v_j)\right) \cdot \frac{\deg(v_p) + \dots + \deg(v_\ell)}{\deg(v_1) + \dots + \deg(v_\ell)}.$$

We multiply (35) by this and get the required assertion of the lemma.

We are left to prove (36). Enumerate the neighbors of v_j which are not in $\{v_1, \dots, v_\ell\}$ by $u_1, \dots, u_{\deg(v_j; V_i)}$. For each $1 \leq m \leq \deg(v_j; V_i)$ we condition on the first $m - 1$ edges being closed, by Proposition 3.1 we remove these edges from the graph, and after this removal the graph remains an $\frac{\gamma}{2}$ -expander by Proposition 2.2. Thus, by Lemma 3.4 we have that in this conditioned graph the resistance on the m -th edge is

$$(1 + O(\varepsilon)) \left(r_m + \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m - 1)} \right), \tag{37}$$

where r_m satisfies

$$\frac{1}{\deg(u_m)} \leq r_m \leq \frac{1}{\deg(u_m; V_i)}. \tag{38}$$

We do not necessarily know if u_m is (α, ε) -good itself, and that is why it is not necessarily the case that the two bounds on r_m are up to $(1 + O(\varepsilon))$ apart from each other. However, we will use the fact that v_j is (α, ε) -good and use property (c) of this definition.

By (37), the probability that all other edges emanating from v_j are closed, conditioned on this already occurring for v_1, \dots, v_{j-1} , is

$$\prod_{m=1}^{\deg(v_j)} \left[1 - (1 + O(\varepsilon)) \left(r_m + \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m - 1)} \right) \right],$$

which equals

$$\exp\left(- (1 + O(\varepsilon)) \sum_m r_m\right) \cdot \exp\left(- (1 + O(\varepsilon)) \sum_m \frac{1}{\sum_{q=j}^{\ell} \deg(v_q) - (m - 1)}\right). \tag{39}$$

Since v_j is (α, ε) -good, by property (c) of the definition we learn that

$$\sum_m r_m = (1 + O(\alpha^{1/2})) \sum_m \frac{1}{\deg(u_m)} = (1 + O(\alpha^{1/2})) b(v_j).$$

Property (d) of the same definition asserts that $b(v_j) \leq \alpha^{-1/4}$, hence the first term in (39) is

$$\exp\left(- (1 + O(\varepsilon)) \sum_m r_m\right) = \left(1 + O(\alpha^{1/4})\right) e^{-b(v_j)}.$$

To handle the second term of (39) we note that

$$\sum_m \frac{1}{\sum_{q=j}^\ell \deg(v_q) - (m - 1)} = \log \left(\frac{\sum_{q=j}^\ell \deg(v_q)}{\sum_{q=j+1}^\ell \deg(v_q)} \right) + O((\varepsilon n)^{-1}).$$

Since $\{v_1, \dots, v_\ell\}$ are (α, ε) -good we have that the ratio inside the logarithm is at most $1 + \ell\varepsilon^{-1}$ and so the second term of (39) equals

$$\left(1 + O(\varepsilon \log(\varepsilon^{-1}))\right) \frac{\sum_{q=j+1}^\ell \deg(v_q)}{\sum_{q=j}^\ell \deg(v_q)}.$$

We multiply these two terms of (39) and get that (36) holds. □

Using Lemma 3.7 we may estimate the probability that $B_{\mathcal{T}}(X, r) \cong T$ “purely”, that is, that $E(B_{\mathcal{T}}(X, r)) \cap E(G[V_i])$ is a rooted tree that is isomorphic to T . We denote this event by $B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T$. Note that $E(B_{\mathcal{T}}(X, r)) \cap E(G[V_i])$ need not even be a tree, so the events $B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T$ and $B_{\mathcal{T}}(X, r) \cong T$ can be very different.

Corollary 3.8 *Assume the setting as introduced at the beginning of Sect. 3.3, and let \mathcal{T} be a UST of G . Fix $i \in [k]$ and let X be a uniformly chosen random vertex of V_i . Then*

$$\mathbb{P}\left(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T\right) = (1 + O(\ell\alpha^{1/4}))\text{Freq}(T; G, i),$$

where $\text{Freq}(T; G, i)$ is defined in Definition 3.6.

Proof This is immediate since the event $E(B_{\mathcal{T}}(X, r)) \cap E(G[V_i]) \cong T$ is the union over all i -pure tuples (v_1, \dots, v_ℓ) of the event $B_{\mathcal{T}}(v_1, r) \stackrel{\text{pure}}{\cong} T(v_1, \dots, v_\ell)$. These events are not disjoint, since for each automorphism of τ of T that fixes the root, we can permute (v_1, \dots, v_ℓ) according to τ and get the identical event. However, up to this invariance, the events are disjoint (which explains the factor of $|\text{Stab}_{\mathcal{T}}|^{-1}$ in the definition of $\text{Freq}(T; G, i)$) and so Lemma 3.7 gives the proof. □

Corollary 3.8 is an annealed statement, that is, the probability space is the product of the UST probability measure and an independent uniform vertex X . A quenched statement follows by a second moment argument. We will first need the following assertion.

Lemma 3.9 *Suppose that G is a connected graph on n vertices and $S \subseteq V(G)$. Let \mathcal{T} be a spanning tree of G and let X be a uniformly chosen vertex of some set $A \subseteq V(G)$. Then for any fixed rooted tree T of height r we have that*

$$\mathbb{P}(B_{\mathcal{T}}(X, r) \cong T \text{ and } V(B_{\mathcal{T}}(X, r)) \cap S \neq \emptyset) \leq |T|^2 |A|^{-1} |S|.$$

Furthermore, if in addition G has some decomposition $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$, then for each $i \in [k]$,

$$\mathbb{P}\left(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T \text{ and } V(B_{\mathcal{T}}(X, r)) \cap V_i \cap S \neq \emptyset\right) \leq |T|^2 |A|^{-1} |S|.$$

Proof We prove only the first statement; the second follows by the same logic. Fix $v \in S$. For each vertex q of T , if \mathcal{T} contains an induced copy of T such that v takes the place of q , then the probability that $B_{\mathcal{T}}(X, r) = T$ such that v takes the place of q is at most $\deg_{\mathcal{T}}(q)/|A|$ since once we choose which edge of \mathcal{T} that touches v corresponds to the edge of T that

touches q towards the root of T , then the corresponding root in \mathcal{T} is chosen and X has to be chosen precisely to be that root. We bound this probability by $|T|/|A|$ and use the union bound over the vertices q of T and v of S . \square

We are now ready to prove the quenched version of Corollary 3.8. Observe that without the assumption that V_i is (α, ε) -big $\text{Freq}(T; G, i)$ can be very small or even 0 (for instance, V_i could have size less than $\varepsilon^5 n$ and if all the vertices of V_0 are connected to each vertex of V_i , then V_i has no (α, ε) -good vertices at all). Thus, we cannot hope to have the concentration required for the following quenched statement without this assumption.

Lemma 3.10 *Assume the setting as introduced at the beginning of Sect. 3.3. Let \mathcal{T} be a UST of G . Fix $i \in [k]$ and assume that V_i is (α, ε) -big. Let X be a uniformly chosen vertex of V_i . Then with probability at least $1 - O(\ell\alpha^{1/8})$ the random tree \mathcal{T} is such that*

$$\mathbb{P}\left(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T\right) = (1 + O(\alpha^{1/16}))\text{Freq}(T; G, i).$$

Proof We write by $Z = Z(T)$ the \mathcal{T} -measurable random variable counting the number of vertices v of V_i such that $B_{\mathcal{T}}(v, r) \stackrel{\text{pure}}{\cong} T$. Corollary 3.8 is equivalent to the assertion that

$$\mathbb{E}Z = (1 + O(\ell\alpha^{1/4}))\text{Freq}(T; G, i) \cdot |V_i|. \tag{40}$$

The second moment of Z is

$$\mathbb{E}Z^2 = \sum_{v, v' \in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v, r) \stackrel{\text{pure}}{\cong} T \text{ and } B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T\right). \tag{41}$$

We will show that for any $v \in V_i$,

$$\sum_{v' \in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T \mid B_{\mathcal{T}}(v, r) \stackrel{\text{pure}}{\cong} T\right) \leq (1 + O(\ell\alpha^{1/4}))\text{Freq}(T; G, i)|V_i|. \tag{42}$$

If we have this, then since Corollary 3.8 gives that

$$\sum_{v \in V_i} \mathbb{P}\left(B_{\mathcal{T}}(v, r) \stackrel{\text{pure}}{\cong} T\right) = (1 + O(\ell\alpha^{1/4}))\text{Freq}(T; G, i) \cdot |V_i|,$$

by putting both in (41) we get that

$$\mathbb{E}Z^2 = (1 + O(\ell\alpha^{1/4}))[\text{Freq}(T; G, i) \cdot |V_i|]^2.$$

Hence

$$\text{Var}(Z) = O(\ell\alpha^{1/4})[\text{Freq}(T; G, i) \cdot |V_i|]^2,$$

and so by Chebychev’s inequality we learn that

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq \alpha^{1/16}\text{Freq}(T; G, i) \cdot |V_i|) \leq O(\ell\alpha^{1/8}),$$

concluding the proof.

To prove (42) we fix $v \in V_i$ and condition on the event $B_{\mathcal{T}}(v, r) \stackrel{\text{pure}}{\cong} T$ and on the vertices (v_1, \dots, v_ℓ) which form $B_{\mathcal{T}}(v, r)$. By Proposition 3.1 we contract the edges of \mathcal{T} in $B_{\mathcal{T}}(v, r)$ and erase the edges we conditioned that are not in \mathcal{T} . Denote the resulting multigraph by G^*

and by v^* the vertex to which the vertices v_1, \dots, v_ℓ have been identified. We consider the loopless multigraph G^* with the partition

$$V(G^*) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_{i-1} \sqcup V_i^* \sqcup V_{i+1} \sqcup \dots \sqcup V_k,$$

where V_i^* is V_i with the vertices v_1, \dots, v_ℓ replaced by v^* . We now claim that this partition is a $(\beta, \alpha, \gamma/2, \varepsilon)$ -good-decomposition (as in Definition 2.10). First, by Proposition 2.2 the graph $G^*[V_i^*]$ is still a $\gamma/2$ -expander and so the partition $V_0 \sqcup V_1 \sqcup \dots \sqcup V_{i-1} \sqcup V_i^* \sqcup V_{i+1} \sqcup \dots \sqcup V_k$ is a $(\gamma/2, \varepsilon^5, \varepsilon^5)$ -expander decomposition of G^* . Next we verify that the number of good vertices has not dropped by too much. Indeed, from each vertex $u \in V_i \setminus \{v_1, \dots, v_\ell\}$ at most ℓ edges touching it were erased and hence it is immediate to check in Definition 2.9 that, as long as n is large enough (in terms of ε and α), if u was (α, ε) -good in G , then it is (α, ε) -good in G^* —the constants may have changed, but they are swallowed in the $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ notation of Definition 2.9.

Thus we may apply Corollary 3.8 and obtain that

$$\sum_{v' \in V_i^*} \mathbb{P} \left(B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T \mid G^* \right) = (1 + \mathcal{O}(\ell\alpha^{1/4})) \text{Freq}(T; G^*, i) \cdot |V_i^*|. \tag{43}$$

We wish to bound the sum in (42) by the above sum, however, there are two important differences between the sums. The first is the difference in the input to the Freq function. \square

Claim 3.10.1 *If V_i is (α, ε) -big, then*

$$c\gamma^{\ell+1} e^{-\ell/\gamma} \leq \text{Freq}(T; G, i) \leq \ell\gamma^{-\ell-1},$$

for some universal constant $c > 0$.

Proof Since V_i is (α, ε) -big we have at least $(1 - \mathcal{O}(\alpha^{1/8}))|V_i|$ vertices which are (α, ε) -good and these must span at least $\Omega(\alpha^{1/9}|V_i|n)$ edges, since the number of edges of $G[V_i]$ touching non (α, ε) -good vertices of V_i is at most $\mathcal{O}(\alpha^{1/8}|V_i|n)$ and we also have $e(G[V_i]) = \Omega(\alpha^{1/9}|V_i|n)$. Thus the average degree of this graph is $d = \Omega(\alpha^{1/9}n)$, and, by greedily removing vertices of degree $d/4$ we can obtain a subgraph of $G[V_i]$ of minimal degree $\Omega(\alpha^{1/9}n) = \Omega(\gamma n)$ such that all of its vertices are (α, ε) -good. Thus it is immediate to find $\Omega(\gamma^\ell n^\ell)$ copies of the tree T . We deduce that the number of ℓ -tuples counted in $\text{Freq}(T; G, i)$ is at least $\Omega(\gamma^\ell n^\ell)$ and at most n^ℓ . Since each vertex degree in V_i is at most n and at least γn we learn that each ℓ -tuple contributes to $\text{Freq}(T; G, i)$ at most $\ell\gamma^{-\ell-1}n^{-\ell}$ and at least $\gamma e^{-\ell/\gamma}n^{-\ell}$ and the claim follows. \square

By this claim, since V_i and V_i^* differ by $\ell - 1$ vertices, and ℓ -tuples of V_i which one of vertices is in $\{v_1, \dots, v_\ell\}$ can contribute at most $\mathcal{O}(\ell\gamma^{-\ell-1}n^{-1})$ to $\text{Freq}(T; G, i)$ we learn that

$$\text{Freq}(T; G, i) = (1 + \mathcal{O}(\ell\gamma^{-\ell-1}n^{-1}))\text{Freq}(T; G^*, i), \tag{44}$$

and n can be taken large enough to that the error in the \mathcal{O} -notation above is $\mathcal{O}(\ell\alpha^{1/4})$. This handles the first difference.

The second difference is that $B_{\mathcal{T}}(v', r)$ may be isomorphic to T in G by using some of the edges in $T(v_1, \dots, v_\ell)$ that were contracted to v^* in G^* . In this case, it does *not* necessarily hold that $B_{\mathcal{T}}(v', r) \cong T$ in G^* , so it is possible that this contribution to (42) is large and not counted for in (43). We will show that this is not the case.

We bound the LHS of (42) as follows. The terms corresponding to $v' \in \{v_1, \dots, v_\ell\}$ we bound by 1 and get a negligible contribution of ℓ . For any other v' we split the event

$B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T$ according to whether $V(B_{\mathcal{T}}(v', r)) \cap V(B_{\mathcal{T}}(v, r)) = \emptyset$. This intersection is empty if and only if $v^* \notin V(B_{\mathcal{T}}(v', r))$ in the graph G^* , and if it is empty, then it holds that $B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T$ in G^* . Thus the LHS of (42) is at most

$$\ell + \sum_{v' \in V_i^*} \mathbb{P} \left(B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T \mid G^* \right) + \sum_{v' \in V_i^*} \mathbb{P} \left(B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T \text{ and } v^* \in V(B_{\mathcal{T}}(v', r)) \mid G^* \right).$$

The first term in the above is negligible, the second we bound using (43) and (44) by the RHS of (42). The third term we bound using Lemma 3.9 applied on the graph G^* with $S = \{v^*\}$ and $A = V_i^*$, giving the bound

$$\sum_{v' \in V_i^*} \mathbb{P} \left(B_{\mathcal{T}}(v', r) \stackrel{\text{pure}}{\cong} T \text{ and } v^* \in V(B_{\mathcal{T}}(v', r)) \mid G^* \right) \leq \ell^2,$$

which is also negligible. This completes the proof of (42) and concludes the proof of the lemma. \square

We now turn to estimating the event $B_{\mathcal{T}}(X, r) \cong T$ rather than the “pure” version of this event. To that aim we define

$$\mathcal{O} = \{(x, y) \in \mathcal{T} : i(x) \neq i(y) \text{ or } x \in V_0 \text{ or } y \in V_0\}, \tag{45}$$

that is, \mathcal{O} is the set of edges of \mathcal{T} that are between components or contained in V_0 . We first assert that this set cannot be too large.

Lemma 3.11 *Assume the setting as introduced at the beginning of Sect. 3.3. Let \mathcal{T} be a UST of G and $\mathcal{O} \subseteq \mathcal{T}$ be defined in (45). Then*

$$\mathbb{E}|\mathcal{O}| = \mathcal{O}(\alpha^{1/4}n).$$

Proof Let us assume that n is large enough as in Lemma 3.4. By Lemma 3.4 and Kirchoff’s formula (24) we have that

$$\begin{aligned} \mathbb{E}|\mathcal{T} \cap E(G[V_i])| &\geq (1 - \mathcal{O}(\varepsilon)) \frac{1}{2} \sum_{u \in V_i} \sum_{v \in V_i : v \sim u} \left(\frac{1}{\deg(u)} + \frac{1}{\deg(v)} \right) \\ &= (1 - \mathcal{O}(\varepsilon)) \sum_{u \in V_i} \frac{\deg(u; V_i)}{\deg(u)}. \end{aligned}$$

Hence

$$\mathbb{E} \left| \mathcal{T} \cap \bigcup_{i=1}^k E(G[V_i]) \right| \geq (1 - \mathcal{O}(\varepsilon)) \sum_{u \in V \setminus V_0} \frac{\deg(u; V_{i(u)})}{\deg(u)}.$$

For any u that is (α, ε) -good we have by part (b) of Definition 2.9 that $\deg(u; V_{i(u)}) \geq (1 - \varepsilon^2) \deg(u)$. Since at least $(1 - \mathcal{O}(\alpha^{1/4}))n$ vertices are (α, ε) -good we deduce that

$$\mathbb{E} \left| \mathcal{T} \cap \bigcup_{i=1}^k E(G[V_i]) \right| \geq (1 - \mathcal{O}(\alpha^{1/4}))n.$$

The proof is now completed since $|\mathcal{T}| = n - 1$ with probability 1. \square

We now reach our final destination.

Lemma 3.12 *Assume the setting as introduced at the beginning of Sect. 3.3. Let \mathcal{T} be a UST of G and let X be a uniformly chosen vertex of G . Then with probability at least $1 - O(\ell\alpha^{1/8})$ for the random tree \mathcal{T} we have*

$$\mathbb{P}(B_{\mathcal{T}}(X, r) \cong T) = (1 + O(\ell^2\alpha^{1/16}))\text{Freq}(T; G) + O(\ell^2\beta^{1/8}).$$

Proof Let \mathcal{O} be defined in (45). We apply Lemma 3.10, together with Lemma 3.11 and Markov’s inequality, and get that with probability at least $1 - O(\ell\alpha^{1/8})$ the random tree \mathcal{T} satisfies

- (1) $|\mathcal{O}| \leq \alpha^{1/8}n$, and
- (2) For each $i \in [k]$ for which V_i is (α, ε) -big we have that

$$\mathbb{P}\left(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T\right) = (1 + O(\ell\alpha^{1/16}))\text{Freq}(T; G, i),$$

where X is a uniformly chosen vertex of V_i .

Let X be a uniformly chosen vertex of G . By Proposition 2.12 the probability that X is in either V_0 or in some V_i that is *not* (α, ε) -big is at most $O(\beta^{1/8})$. Conditioned on $X \in V_i$ we have that X is uniform random vertex of V_i . Hence we may sum item (2) above over these i ’s and get that with probability at least $1 - O(\ell\alpha^{1/8})$ the random tree \mathcal{T} satisfies

$$\mathbb{P}(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T) = (1 + O(\ell\alpha^{1/16}))\text{Freq}(T; G) + O(\beta^{1/8}),$$

by definition of $\text{Freq}(T; G)$.

Assume \mathcal{T} satisfies these. If $B_{\mathcal{T}}(X, r) \cong T$ but not $B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T$, then we must have that $B_{\mathcal{T}}(X, r) \cong T$ and $B_{\mathcal{T}}(X, r) \cap V(\mathcal{O}) \neq \emptyset$. Similarly, if $B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T$ but not $B_{\mathcal{T}}(X, r) \cong T$, then we similarly must have that $B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T$ and $B_{\mathcal{T}}(X, r) \cap V(\mathcal{O}) \neq \emptyset$. Since $|\mathbb{P}(A) - \mathbb{P}(B)| \leq \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A)$ we have that

$$\left| \mathbb{P}(B_{\mathcal{T}}(X, r) \cong T) - \mathbb{P}\left(B_{\mathcal{T}}(X, r) \stackrel{\text{pure}}{\cong} T\right) \right| = O(\ell^2\beta^{1/8}),$$

where the last inequality is due to Lemma 3.9 with $A = V(G)$ and $S = V(\mathcal{O})$. □

4 Proof of Theorems 1.1 and 1.3

4.1 Deriving Theorem 1.1 from Lemma 3.12 via Anatomies

In this section, we deduce Theorem 1.1 from Lemma 3.12. The main technical result of this section, Lemma 4.3, says that the quantity Freq is continuous in a certain robust sense. To prove Lemma 4.3, we need to group the elements of Ω for a given graphon $W : \Omega^2 \rightarrow [0, 1]$ to groups with similar degrees. Actually, we need a recursive refinement of this, as follows. We call the above partition of Ω according to the degrees *anatomy of depth 1*. Now, having defined an anatomy of depth d , *anatomy of depth $d + 1$* is a decomposition of Ω into groups in which elements have approximately similar degrees into every individual cell of the anatomy of depth d . Let us make this precise.

Suppose that $h \in \mathbb{N}$ and S is a finite set. Let $\mathfrak{C}(h, S)$ be the partition of $[0, 1]^S$ into Voronoi cells generated by points $p \in \{0, \frac{1}{h}, \dots, \frac{h-1}{h}, 1\}^S$ (we assign each boundary point to one arbitrary neighboring cell to break ties). Note that these cells form $(h + 1)^{|S|}$ cubes of the form $\prod_{i \in S} \left[\max\{0, \frac{2r_i-1}{2h}\}, \min\{1, \frac{2r_i+1}{2h}\} \right]$, for some $\{r_i \in \{0, \dots, h\}\}_{i \in S}$.⁵ In the degenerate case $S = \emptyset$, we define $\mathfrak{C}(h, S) := \{\emptyset\}$. We call the points p the **centers of the cells**. When $C \in \mathfrak{C}(h, S)$ is a Voronoi cell with center $p = (p_i)_{i \in S}$, for $i \in S$ we write $center_i(C) := p_i$.

Let us write \mathfrak{b} for an (abstract) element. Below, the only purpose of \mathfrak{b} will be to refer to a coordinate that will have to do with the function $b_W(\cdot)$.

Now, suppose that $d \in \mathbb{N}$ and $\mathbf{h} \in \mathbb{N}^d$. For $t \in [d]$, we write \mathbf{h}_t for the t -coordinate of \mathbf{h} , and $\mathbf{h}[\![t]\!]$ for the t -dimensional vector obtained from \mathbf{h} by removing the last $d - t$ components. For $h \in \mathbb{N}$, let $\mathfrak{D}_{h,0} := \mathfrak{C}(h, \emptyset)$ and for $d \in \mathbb{N}$ and $\mathbf{h} \in \mathbb{N}^d$ let $\mathfrak{D}_{\mathbf{h}} := \mathfrak{C}(\mathbf{h}_d, \mathfrak{D}_{\mathbf{h}[\![d-1]\!]}) \sqcup \{\mathfrak{b}\}$. We have

$$\sum_{t=0}^d |\mathfrak{D}_{\mathbf{h}[\![t]\!]})| = \mathfrak{h}(\mathbf{h}) , \tag{46}$$

for a suitable tower-function $\mathfrak{h}(\cdot)$.

Suppose $W : \Omega^2 \rightarrow [0, 1]$ is a graphon and $h \in \mathbb{N}$ is arbitrary. Let $\mathcal{A}_{\emptyset} := \Omega$. We call $\{\mathcal{A}_{\emptyset}\} = \{\mathcal{A}_C\}_{C \in \mathfrak{D}_{h,0}}$ **anatomy of W of depth 0**. Suppose that $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$ and that we already know the anatomy $\{\mathcal{A}_C\}_{C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]}}$ of W of depth $d - 1$, which is a partition of Ω . Now, for each $\omega \in \Omega$ we consider the $|\mathfrak{D}_{\mathbf{h}[\![d-1]\!]})|$ -tuple of degrees

$$\mathbf{deg}_d(\omega) := (\deg_W(\omega, \mathcal{A}_C))_{C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]}} .$$

Then for each $F \in \mathfrak{D}_{\mathbf{h}}$ we define $\mathcal{A}_F := (\mathbf{deg}_d)^{(-1)}(F) \cap (\exp(-b_W(\cdot)))^{(-1)}(F)$. In words, each cell $\mathcal{A}_F \subseteq \Omega$ has the property that for each $C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]})$ and for each $\omega \in \mathcal{A}_F$ we have that

$$\deg_W(\omega, \mathcal{A}_C) = center_C(F) \pm \frac{1}{2\mathbf{h}_d} \quad \text{and} \quad \exp(-b_W(\omega)) = center_{\mathfrak{b}}(F) \pm \frac{1}{2\mathbf{h}_d} . \tag{47}$$

We call $\{\mathcal{A}_F\}_{F \in \mathfrak{D}_{\mathbf{h}}}$ the **anatomy of W of depth d** . Obviously, $\{\mathcal{A}_F\}_{F \in \mathfrak{D}_{\mathbf{h}}}$ is a partition of Ω . Consider an arbitrary $\omega \in \mathcal{A}_F$. Summing (47) over all $C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]})$ (for which (46) tells us that there are $\mathfrak{h}(\mathbf{h}[\![d-1]\!]})$ summands), we get

$$\deg_W(\omega) = \sum_{C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]})} center_C(F) \pm \frac{\mathfrak{h}(\mathbf{h}[\![d-1]\!]})}{2\mathbf{h}_d} . \tag{48}$$

For this reason, we shall call the number $\sum_{C \in \mathfrak{D}_{\mathbf{h}[\![d-1]\!]})} center_C(F)$ the **degree of \mathcal{A}_F** (in W), and denote it by $\mathbf{deg}_W^{\text{anat}}(\mathcal{A}_F)$.

Suppose $U : \Omega^2 \rightarrow [0, 1]$ and $X \subseteq \Omega$. Suppose that $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$, and $\kappa \geq 0$ are given. Let $\mathcal{B}_{\emptyset} := \Omega$ and for each $t \in [d]$ let $\{\mathcal{B}_C\}_{C \in \mathfrak{D}_{\mathbf{h}[\![t]\!]}}$ be a partition of Ω . Suppose that for each $t \in [d]$, each $F \in \mathfrak{D}_{\mathbf{h}[\![t]\!]})$, each $C \in \mathfrak{D}_{\mathbf{h}[\![t-1]\!]})$ and each $\omega \in \mathcal{B}_F \setminus X$ we have

$$|\deg_U(\omega, \mathcal{B}_C) - center_C(F)| \leq \frac{1}{2\mathbf{h}_t} + \kappa \quad \text{and} \quad |\exp(-b_U(\omega)) - center_{\mathfrak{b}}(F)| \leq \frac{1}{2\mathbf{h}_d} + \kappa . \tag{49}$$

⁵ Strictly speaking, when some but not all coordinates r_i are 0 or h , these are not cubes but rectangular prisms. This is however not important.

We then say that $\left\{ \{ \mathcal{B}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[t]}} \right\}_{t=0}^d$ are κ -**approximate anatomies of U up to depth d with exceptional set X** .

Note that when we take $\kappa = 0$ and $X = \emptyset$ then we recover the notion of anatomies.

Our next lemma says that two graphons that are close in the cut-distance have similar anatomies.

Lemma 4.1 *Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a nondegenerate graphon and $d \in \mathbb{N}$, $\mathbf{h} \in \mathbb{N}^d$ are arbitrary. Let $\left\{ \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[t]}} \right\}_{t=0}^d$ be the anatomies of W up to depth d . For every $\kappa > 0$ there exists a number $\delta > 0$ such that the following holds for every graphon $U : \Omega^2 \rightarrow [0, 1]$ with $\|W - U\|_{\square} < \delta$. There exists a set $X \subseteq \Omega$ of measure at most κ so that $\left\{ \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[t]}} \right\}_{t=0}^d$ are κ -approximate anatomies of U up to depth d with exceptional set X .*

Proof Suppose that we are given W , d , \mathbf{h} , and κ as above.

Since W is nondegenerate, there exists $\beta \in (0, 10^{-6})$ such that the measure of the set

$$S := \left\{ \omega \in \Omega : \deg(\omega) < 16\sqrt[4]{\beta} \right\}$$

is less than $\frac{\kappa^2}{100}$. Let $\delta := \min\left\{ \frac{\beta\kappa^4}{400}, \frac{\kappa^2}{4h(\mathbf{h})^2} \right\}$. Suppose that U is given with $\|W - U\|_{\square} < \delta$. It is enough to prove that there exists a set $X_b \subseteq \Omega$ of measure at most $\frac{\kappa}{4}$ such that for each $t \in [d]$ and each $F \in \mathfrak{D}_{\mathbf{h}[t]}$, we have for each $\omega \in F \setminus X_b$ that

$$|\exp(-b_U(\omega)) - center_b(F)| \leq \frac{1}{2h_t} + \kappa, \tag{50}$$

and that (with t and F as above) for each $C \in \mathfrak{D}_{\mathbf{h}[t-1]}$ we have that all but at most $\frac{\kappa}{4h(\mathbf{h})^2}$ measure of elements $\omega \in \mathcal{A}_F$ satisfy

$$\deg_U(\omega, \mathcal{A}_C) \geq center_C(F) - \frac{1}{2h_t} - \kappa, \tag{51}$$

and all but at most $\frac{\kappa}{4h(\mathbf{h})^2}$ measure of elements $\omega \in \mathcal{A}_F$ satisfy

$$\deg_U(\omega, \mathcal{A}_C) \leq center_C(F) + \frac{1}{2h_t} + \kappa. \tag{52}$$

The lemma will then follow from (46) by taking X to be the union of X_b together with the exceptional sets from (51), (52) over all t , C , and F . □

The following claim clearly implies (50).

Claim 4.1.1 *There exists a set X_b of measure at most $\frac{\kappa}{4}$ such that for all $\omega \in \Omega \setminus X_b$ we have $|b_W(\omega) - b_U(\omega)| < \kappa$.*

For the proof of Claim 4.1.1, we need Claims 4.1.2–4.1.4 below.

Claim 4.1.2 *Suppose that $\Gamma : \Omega^2 \rightarrow [0, 1]$ is a graphon and that $A \subseteq \Omega$. Then*

$$\int_{x \in \Omega} \int_{\omega \in A} \frac{\Gamma(x, \omega)}{\deg_{\Gamma}(\omega)} \leq \mu(A).$$

Proof By Fubini’s Theorem, we have

$$\int_{x \in \Omega} \int_{\omega \in A} \frac{\Gamma(x, \omega)}{\deg_{\Gamma}(\omega)} = \int_{\omega \in A} \frac{1}{\deg_{\Gamma}(\omega)} \int_{x \in \Omega} \Gamma(x, \omega) = \int_{\omega \in A} \frac{1}{\deg_{\Gamma}(\omega)} \deg_{\Gamma}(\omega, A) \leq \mu(A).$$

□

Claim 4.1.3 Suppose $\Gamma_1, \Gamma_2 : \Omega^2 \rightarrow [0, 1]$ are two graphons and that $f : \Omega \rightarrow [0, C]$ is an arbitrary function. Then

$$\int_{x \in \Omega} \left| \int_{\omega \in \Omega} f(\omega)(\Gamma_1(x, \omega) - \Gamma_2(x, \omega)) \right| \leq 2C \cdot \|\Gamma_1 - \Gamma_2\|_{\square}.$$

Proof By [15, (8.20)] we can alternatively express as $\|\Gamma_1 - \Gamma_2\|_{\square}$

$$\begin{aligned} \|\Gamma_1 - \Gamma_2\|_{\square} &= \sup_{F, G: \Omega \rightarrow [0, 1]} \left| \int_x \int_y G(x)F(y)(\Gamma_1(x, y) - \Gamma_2(x, y)) \right| = \\ &= \frac{1}{C} \cdot \sup_{F: \Omega \rightarrow [0, C], G: \Omega \rightarrow [0, 1]} \left| \int_x \int_y G(x)F(y)(\Gamma_1(x, y) - \Gamma_2(x, y)) \right| \cdot \\ &\geq \frac{1}{2C} \cdot \sup_{F: \Omega \rightarrow [0, C], G: \Omega \rightarrow [-1, 1]} \left| \int_x \int_y G(x)F(y)(\Gamma_1(x, y) - \Gamma_2(x, y)) \right|. \end{aligned}$$

The claim follows by considering in the latter supremum functions

$$F(\cdot) := f(\cdot) \quad \text{and} \quad G(\cdot) := \operatorname{sgn} \left(\int_y f(y)(\Gamma_1(\cdot, y) - \Gamma_2(\cdot, y)) \right).$$

□

For the last auxiliary claim, we shall introduce an auxiliary notion. Below we shall work with not necessarily symmetric L^1 -functions $K : \Omega^2 \rightarrow [0, +\infty)$. Note that the notions of cut-norm and cut-distance extend to this setting.⁶ The following claim is then obvious.

Claim 4.1.4 Suppose that $\varepsilon > 0$ and $\Gamma, \Gamma' : \Omega^2 \rightarrow [0, 1]$ are two L^1 -functions such that for each $x, y \in \Omega$, $\max(\Gamma(x, y), \Gamma'(x, y)) < (1 + \varepsilon) \min(\Gamma(x, y), \Gamma'(x, y))$. Then $\|\Gamma - \Gamma'\|_{\square} \leq \varepsilon$. □

Proof of Claim 4.1.1 Let

$$D := \left\{ \omega \in \Omega \setminus S : \deg_W(\omega) \neq \left(1 \pm \frac{\beta^{0.3} \kappa^2}{4} \right) \deg_U(\omega) \right\}.$$

Since for each $\omega \in D$ we have $|\deg_W(\omega) - \deg_U(\omega)| \geq 3\beta^{0.55} \kappa^2$ we get from $\|W - U\|_{\square} < \frac{\beta \kappa^4}{400}$ that $\mu(D) \leq \frac{\kappa^2}{100}$.

Put $A := S \cup D$. We have $\mu(A) \leq \frac{\kappa^2}{50}$.

We have

$$\begin{aligned} \int_{x \in \Omega} |b_W(x) - b_U(x)| &\leq \int_{x \in \Omega} \left| \int_{\omega \in \Omega \setminus A} \frac{W(x, \omega)}{\deg_W(\omega)} - \frac{U(x, \omega)}{\deg_U(\omega)} \right| \\ &\quad + \int_{x \in \Omega} \left| \int_{\omega \in A} \frac{W(x, \omega)}{\deg_W(\omega)} - \frac{U(x, \omega)}{\deg_U(\omega)} \right| \end{aligned} \tag{53}$$

The second term on the right-hand side of (53) is bounded by Claim 4.1.2 by at most $2\mu(A) \leq \kappa^2/50$. Now, let us consider the L^1 -function \tilde{U} , defined by

$$\tilde{U}(x, y) := \begin{cases} \frac{\deg_W(y)}{\deg_U(y)} \cdot U(x, y) & \text{if } y \notin A \\ U(x, y) & \text{if } y \in A \end{cases}.$$

⁶ Note that in this case, the cut-norm really has to be defined over all rectangles in (6), and not only over all squares.

Observe that U and \tilde{U} satisfy the assumptions of Claim 4.1.4 with error parameter $\left(\frac{\beta^{0.3}\kappa^2}{4}\right)$. Then the first term of the right-hand side of (53) can be rewritten as

$$\int_{x \in \Omega} \left| \int_{\omega \in \Omega} \frac{\mathbf{1}_{\Omega \setminus A}(\omega)}{\text{deg}_W(\omega)} \cdot \left(W(x, \omega) - \tilde{U}(x, \omega) \right) \right|. \tag{54}$$

Observe that the function $\frac{\mathbf{1}_{\Omega \setminus A}(\cdot)}{\text{deg}_W(\cdot)}$ is bounded by above by $\frac{1}{16\sqrt[3]{\beta}}$. Claim 4.1.3 tells us that (54) is at most

$$\frac{1}{8\sqrt[4]{\beta}} \cdot \left(\|W - U\|_{\square} + \|U - \tilde{U}\|_{\square} \right) \stackrel{\text{C4.1.4}}{\leq} \frac{1}{3200} \beta^{3/4} \kappa^2 + \frac{\beta^{0.3} \kappa^2}{4 \cdot 8\beta^{1/4}} < \frac{\kappa^2}{25}.$$

Plugging this back into (53) we obtain that $\int_{x \in \Omega} |b_W(x) - b_U(x)| < \frac{\kappa^2}{4}$. The claim follows. \square

It now remains to prove (51) and (52). We will only prove (51) since the proof of (52) is verbatim. So, suppose that (51) fails, i.e., the set $X := \{\omega \in \mathcal{A}_F : \text{deg}_U(\omega, \mathcal{A}_C) > \text{center}_C(F) + \frac{1}{2\mathbf{h}_t} + \kappa\}$ satisfies $\mu(X) > \frac{\kappa}{2\mathbf{h}(\mathbf{h})}$. By the definition of \mathcal{A}_F we have for every $\omega \in X$ that $\text{deg}_W(\omega, \mathcal{A}_C) \leq \text{center}_C(F) + \frac{1}{2\mathbf{h}_t}$. Therefore,

$$\begin{aligned} \int_{X \times \mathcal{A}_F} W &\leq \left(\text{center}_C(F) + \frac{1}{2\mathbf{h}_t} \right) \mu(X) \quad \text{but} \\ \int_{X \times \mathcal{A}_F} U &> \left(\text{center}_C(F) + \frac{1}{2\mathbf{h}_t} + \kappa \right) \mu(X) \geq \left(\text{center}_C(F) + \frac{1}{2\mathbf{h}_t} \right) \mu(X) + \frac{\kappa^2}{2\mathbf{h}(\mathbf{h})}. \end{aligned}$$

\square

This is a contradiction to the fact that $\|U - W\|_{\square} < \delta$.

For the proof of Lemma 4.3 we need the following result.

Proposition 4.2 *Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a nondegenerate graphon. Then*

$$\mathbb{E}_{x \in \Omega} [b_W(x)] = 1.$$

Proof By Fubini’s Theorem,

$$\mathbb{E}_{x \in \Omega} [b_W(x)] = \int_x \left(\int_y \frac{W(x, y)}{\text{deg}(y)} dy \right) dx = \int_y \left(\int_x \frac{W(x, y)}{\text{deg}(y)} dx \right) dy = 1. \quad \square$$

Lemma 4.3 puts a relation between quantities $\text{Freq}(T; W)$ and $\text{Freq}^-(T; G, V_0, E_0)$, where G is a graph, $V_0 \subseteq V(G)$ and $E_0 \subseteq E(G)$, and $\text{Freq}^-(T; G, V_0, E_0)$ is defined as

$$\begin{aligned} \text{Freq}^-(T; G, V_0, E_0) &:= |\text{Stab}_T|^{-1} \\ &\times \sum_{\substack{v_1, \dots, v_{\ell} \in V(G) \setminus V_0 \\ v_i j \in E(T) : v_i v_j \in E(G) \setminus E_0}} \frac{1}{v(G)} \exp \left(- \sum_{j=1}^{p-1} b_G(v_j) \right) \frac{\sum_{j=p}^{\ell} \text{deg}_G(v_j)}{\prod_{j=1}^{\ell} \text{deg}_G(v_j)}. \end{aligned}$$

We are now ready to state Lemma 4.3. It says that the parameter $\text{Freq}(T; \cdot)$ is continuous in a certain sense. While it would be possible to prove continuity of $\text{Freq}(T; \cdot)$ for nondegenerate graphons with respect to the cut-distance, here we need to put a relation between $\text{Freq}(T; \cdot)$ of a graphon and the parameter $\text{Freq}^-(T; \cdot)$ of a graph that is close to that graphon. Let us note that while our proof of continuity is long, the statement itself is natural. Indeed, the

definition (1) is an integration involving products of values of the graphon over the edges of T (similar to the way that is used to define the density of T in the graphon), degrees in the graphon and the function $b(\cdot)$ from (2) which is also defined using degrees in the graphon. As subgraph densities are continuous with respect to the cut-metric, and so is the degree sequence (c.f. Lemma 5.1), it is actually plausible to have continuity of many graph(on) parameters obtained by combinations thereof.

Lemma 4.3 *Suppose that $W : \Omega^2 \rightarrow [0, 1]$ is a nondegenerate graphon and T is a fixed tree. For every $\varepsilon > 0$ there exists $a > 0$ such that for every $\chi > 0$ there exists $n_0 \in \mathbb{N}$ and $\delta > 0$ such that if G is a graph and $V_0 \subseteq V(G)$, $E_0 \subseteq E(G)$ satisfy*

- (a) $n > n_0$,
- (b) $\delta_{\square}(W, G) < \delta$,
- (c) $|V_0| \leq an$, $|E_0| \leq an^2$, and
- (d) for each $v \in V(G) \setminus V_0$ we have $\deg_G(v) \geq \chi n$,

then we have

$$|\text{Freq}(T; W) - \text{Freq}^-(T; G, V_0, E_0)| < \varepsilon. \tag{55}$$

Proof Suppose that W, T , and ε are given.

Let L be the height of T . Suppose that the vertices of T are $[\ell]$. Suppose that 1 is the root of T , suppose that the height of each vertex i is denoted g_i and that the vertices of T are enumerated so that we have $0 = g_1 < g_2 \leq g_3 \leq \dots \leq g_{p-1} < g_p = g_{p+1} = \dots = g_{\ell} = L$.

In the course of deriving Theorem 1.3 from Theorem 1.1 in Sect. 4.2 we prove that $\text{Freq}(T; W) \leq 1$.⁷ In particular, the function $f_W : \Omega^{\ell} \rightarrow [0, +\infty)$,

$$f_W(\omega_1, \dots, \omega_{\ell}) := \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \cdot \frac{\sum_{j=p}^{\ell} \deg_W(\omega_j)}{\prod_{j=1}^{\ell} \deg_W(\omega_j)} \cdot \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \tag{56}$$

is integrable. Let $\delta_0 > 0$ be such that for any set $A \subseteq \Omega^{\ell}$, $\mu^{\otimes \ell}(A) < \delta_0$ we have

$$\int_A f_W \leq \frac{\varepsilon}{10}. \tag{57}$$

Since W is nondegenerate, there exists a number $\Delta \in (0, \frac{\delta_0}{4\ell})$ such that the set

$$\Omega_{\text{small}} := \{\omega \in \Omega : \deg_W(\omega) < 2\Delta\}$$

has measure less than $\frac{\delta_0}{2\ell}$.

Let $a := \min\{\frac{\delta_0}{10\ell}, \frac{\varepsilon\Delta^{\ell}}{400\ell^{\ell}}\}$. Now, suppose that χ is given.

Let $\mathbf{h} \in \mathbb{N}^{L+1}$ and $d', d, \tau, \kappa > 0$ satisfy

$$\min(\Delta, \varepsilon, \chi, \delta_0) \gg d' \gg d \gg \tau \gg \frac{1}{\mathbf{h}_1} \gg \frac{1}{\mathbf{h}_2} \gg \dots \gg \frac{1}{\mathbf{h}_L} \gg \frac{1}{\mathbf{h}_{L+1}} \gg \kappa > 0. \tag{58}$$

We note that the above dependencies are tower-type, i.e.,

$$\frac{1}{\mathbf{h}_i} \gg \frac{\hbar(\mathbf{h}[[i]])}{\mathbf{h}_{i+1}}. \tag{59}$$

⁷ At this moment, we have not established Theorem 1.3 nor Theorem 1.1. However, the fact $\text{Freq}(T; W) \leq 1$ did not rely on the validity of either of these theorems, but rather followed from making the connection to the branching process κ_W .

Let $\left\{ \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}} \right\}_{\ell=0}^{L+1}$ be the anatomies of W up to depth $L + 1$. Let δ be given by Lemma 4.1 for input graphon W and parameters \mathbf{h} and κ .

Suppose now that the graph G is given. Take a graphon representation U of G such that $\|W - U\|_{\square} < \delta$. Lemma 4.1 tells us that there exists a set $X \subseteq \Omega$ of measure at most κ such that $\left\{ \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}} \right\}_{\ell=0}^{L+1}$ are a κ -approximate anatomies of U up to depth $L + 1$ with exceptional set X .

Given a vertex $v \in V(G)$, we write $\Omega_v \subseteq \Omega$ for the set representing v in U . We write $\Lambda_{\text{vert}} := \bigcup_{v \in V_0} \Omega_v$, and $\Lambda_{\text{edge}} = \bigcup_{uv \in E_0} (\Omega_u \times \Omega_v \cup \Omega_v \times \Omega_u)$. By assumption (c) we have

$$\mu(\Lambda_{\text{vert}}) \leq a \quad \text{and} \quad \mu^{\otimes 2}(\Lambda_{\text{edge}}) \leq 2a. \tag{60}$$

To express $\text{Freq}^-(T; G, V_0, E_0)$, we introduce a counterpart to (56) that reflects U ,

$$f_U(\omega_1, \dots, \omega_\ell) := \exp \left(- \sum_{j=1}^{p-1} b_U(\omega_j) \right) \cdot \frac{\sum_{j=p}^{\ell} \text{deg}_U(\omega_j)}{\prod_{j=1}^{\ell} \text{deg}_U(\omega_j)} \cdot \prod_{(i,j) \in E(T)} U(\omega_i, \omega_j)$$

and a version f_U^- , which takes into the account the “deleted” vertices V_0 and edges E_0 ,

$$f_U^-(\omega) := \begin{cases} 0 & \text{if } \exists i \in [\ell] : \omega_i \in \bigcup_{v \in V_0} \Omega_v \text{ or } \exists ij \in E(T) : (\omega_i, \omega_j) \in \bigcup_{vw \in E_0} \Omega_v \times \Omega_w, \\ f_U(\omega) & \text{otherwise.} \end{cases}$$

Note that the assumption (d) implies that

$$f_U^- \leq \frac{\ell}{\chi^\ell} \quad \text{on the whole domain } \Omega^\ell. \tag{61}$$

We say that a map $\pi : [\ell] \rightarrow \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}; F \in \mathfrak{D}_{\mathbf{h}[\ell]}}$ is a **depth preserving anatomy assignment** if for each $i \in [\ell]$ we have $\pi(i) = \mathcal{A}_F$ for some $F \in \mathfrak{D}_{\mathbf{h}[\ell+1-g_i]}$. This means that the root 1 is assigned an anatomy of depth $L + 1$ while vertices further from the root are assigned anatomies of smaller depths. Note that depth preserving anatomy assignment partition the set Ω^ℓ in the sense that

$$\Omega^\ell = \bigsqcup_{\substack{\pi: [\ell] \rightarrow \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}; \\ \text{depth preserving anatomy assignment}}} \prod_{j=1}^{\ell} \pi(j). \tag{62}$$

In particular,

$$1 = \sum_{\substack{\pi: [\ell] \rightarrow \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}; \\ \text{depth preserving anatomy assignment}}} \prod_{j=1}^{\ell} \mu(\pi(j)) \quad , \text{ and} \tag{63}$$

$$1 = \sum_{\substack{\pi: [\ell] \rightarrow \{ \mathcal{A}_F \}_{F \in \mathfrak{D}_{\mathbf{h}[\ell]}; \\ \text{depth preserving anatomy assignment}}} \prod_{j \in [\ell] \setminus \{j^*\}}^{\ell} \mu(\pi(j)) \quad \text{for each } j^* \in [\ell]. \tag{64}$$

We need to define three additional classes of depth preserving anatomy assignments.

- We say that a depth preserving anatomy assignment π is **singular** if there exists $i \in [\ell]$ such that a positive measure of the elements $\omega \in \pi(i)$ satisfy $\text{deg}_W(\omega) \leq \Delta$, or $b_W(\omega) > \Delta^{-1}$.

- We say that a depth preserving anatomy assignment π is **dense** if for every vertex $j \in [\ell]$ and every child j^* of j we have that $center_{\pi(j^*)}(\pi(j)) > d \cdot \mu(\pi(j^*))$. If this fails for some pair jj^* , then we call j^* a **witness**. (Of course, several witnesses may exist.)
- We say that a depth preserving anatomy assignment π is **non-problematic** if it is dense, and it is not singular.

For a depth preserving anatomy assignment π , we write $\Omega_{\pi} = \prod_{i=1}^{\ell} (\pi(i) \setminus X)$.

The next claim establishes some basic properties of singular depth preserving anatomy assignment. □

Claim 4.3.1 *Suppose that π is a singular depth preserving anatomy assignment, and that $i \in \ell$ is a witness of singularity. Then we have that $\pi(i) \subseteq \Omega_{\text{small}}$ or that $b_W(\omega') > \frac{1}{2\Delta}$ for all $\omega' \in \pi(i)$.*

Proof The definition of singular depth preserving anatomy assignment says that there exists $\omega \in \pi(i)$ such that $\deg_W(\omega) \leq \Delta$, or $b_W(\omega) > \Delta^{-1}$.

Let us first deal with the former case. A double application of (48) gives that for each $\omega' \in \pi(i)$ we have

$$\begin{aligned} \deg_W(\omega') &\leq \deg_W^{\text{anat}}(\pi(i)) + \frac{\hbar(\mathbf{h}[L-1-g_i])}{2\mathbf{h}_{L-g_i}} \\ &\leq \deg_W(\omega) + \frac{2\hbar(\mathbf{h}[L-1-g_i])}{2\mathbf{h}_{L-g_i}} \stackrel{(58),(59)}{<} 2\Delta. \end{aligned}$$

Therefore, in this case, $\pi(i) \subseteq \Omega_{\text{small}}$.

We can deal with the latter case, but using (47) instead of (48). We have

$$b_W(\omega') > \Delta^{-1} - \frac{4}{2\mathbf{h}_{L-g_i}} \geq \frac{1}{2\Delta} \quad \text{for all } \omega' \in \pi(i). \quad \square$$

The next key claim says that the integrals of f_W and f_U^- over sets corresponding to non-problematic depth preserving anatomy assignment are almost the same.

Claim 4.3.2 *Suppose that $\pi : [\ell] \rightarrow \{\mathcal{A}_F\}_{F=1, \dots, L+1; F \in \mathfrak{D}_{\mathbf{h}[i]}}$ is a non-problematic depth preserving anatomy assignment. Let*

$$\begin{aligned} e_{\text{vert}}^{\pi} &:= \sum_{i=1}^{\ell} \mu(\pi(i) \cap \Lambda_{\text{vert}}) \cdot \prod_{k \in V(T) \setminus \{i\}} \mu(\pi(k)) \quad , \text{ and} \\ e_{\text{edge}}^{\pi} &:= \sum_{ij \in E(T)} \mu^{\otimes 2} \left(\Lambda_{\text{edge}} \cap \bigcup_{ij \in E(T)} \pi(i) \times \pi(j) \right) \cdot \prod_{k \in V(T) \setminus \{i, j\}} \mu(\pi(k)). \end{aligned}$$

Then for the quantities

$$\begin{aligned} Q_1 &:= \int_{\omega \in \Omega_{\pi}} f_W(\omega) \text{ and} \\ Q_2 &:= \int_{\omega \in \Omega_{\pi}} f_U^-(\omega) \end{aligned}$$

we have $Q_1 = Q_2 \cdot (1 \pm d) \pm \frac{(e_{\text{vert}}^{\pi} + e_{\text{edge}}^{\pi})^{\ell}}{\Delta^{\ell}}$.

Proof Suppose that $j \in V(T)$. If $j \neq 1$, then we write $\text{Par}(j)$ for the parent of j in T . Let us write F_j for the tree obtained from T by erasing the edge from j to $\text{Par}(j)$ and taking the component containing j . When $j = 1$, we take $F_j := T$. Also, for $j = 1$ we define $\text{center}_{\pi(j)}(\pi(\text{Par}(j))) := 1$.

For $j \in V(T)$, write

$$R_j := \prod_{k \in V(F_j) \setminus \{j\}} \text{center}_{\pi(k)}(\pi(\text{Par}(k))).$$

Here, recall the convention that a product over the empty set is 1. This in particular applies when j is a leaf and $j \neq 1$.

We first want to show that the quantities

$$\int_{(\omega_1, \dots, \omega_\ell) \in \Omega_\pi} \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \quad \text{and} \quad \int_{(\omega_1, \dots, \omega_\ell) \in \Omega_\pi} \prod_{(i,j) \in E(T)} U(\omega_i, \omega_j)$$

are very close. To this end, for $j \in V(T)$ we define $A_j(\omega_j) = B_j(\omega_j) := 1$ if j is a leaf (different than the root) and otherwise we define

$$\begin{aligned} A_j(\omega_j) &:= \int_{\{\omega_k \in \pi(k) \setminus X\}_{k \in V(F_j) \setminus \{j\}}} \prod_{(a,b) \in E(F_j)} W(\omega_a, \omega_b), \\ B_j(\omega_j) &:= \int_{\{\omega_k \in \pi(k) \setminus X\}_{k \in V(F_j) \setminus \{j\}}} \prod_{(a,b) \in E(F_j)} U(\omega_a, \omega_b). \end{aligned} \tag{65}$$

Inductively for $t = L + 1, L, \dots, 0$, assume that for each vertex $j \in V(T)$ at height $g_j = t$,⁸ for each $\omega_j \in \pi(j)$ we have

$$A_j(\omega_j) = R_j \cdot \exp\left(\pm \frac{v(F_j)}{h_1}\right) \tag{66}$$

and for every $\omega_j \in \pi(j) \setminus X$ we have

$$B_j(\omega_j) = R_j \cdot \exp\left(\pm \frac{v(F_j)}{h_1}\right). \tag{67}$$

Now, let suppose that the statement is true for all vertices at height $t + 1$. Let j be an arbitrary vertex at height t , and let $\omega_j \in \pi(j)$ be arbitrary. When j is a leaf then (66) and (67) hold trivially. So, suppose that j has some children j_1, j_2, \dots, j_q . Since these children are at height $t + 1$, applying the induction hypothesis, we disintegrate (65) with respect to these children, and get

$$\begin{aligned} A_j(\omega_j) &= \prod_{c=1}^q \left(\int_{\omega_{j_c} \in \pi(j_c) \setminus X} W(\omega_j, \omega_{j_c}) \int_{\{\omega_k \in \pi(k) \setminus X\}_{k \in V(F_{j_c}) \setminus \{j_c\}}} \prod_{(a,b) \in E(F_{j_c})} W(\omega_a, \omega_b) \right) \\ &= \prod_{c=1}^q \left(\int_{\omega_{j_c} \in \pi(j_c) \setminus X} W(\omega_j, \omega_{j_c}) \cdot A_{j_c}(\omega_{j_c}) \right) \\ &\stackrel{(66)}{=} \prod_{c=1}^q \left(\int_{\omega_{j_c} \in \pi(j_c) \setminus X} W(\omega_j, \omega_{j_c}) \cdot R_{j_c} \exp\left(\pm \frac{v(F_{j_c})}{h_1}\right) \right) \end{aligned}$$

⁸ Note that the first step of the induction is satisfied trivially, as the height of T is L .

$$\begin{aligned} &= \prod_{c=1}^q \deg_W(\omega_j, \pi(j_c) \setminus X) \cdot \prod_{c=1}^q \left(R_{j_c} \exp\left(\pm \frac{v(F_{j_c})}{\mathbf{h}_1}\right) \right) \\ &= \prod_{c=1}^q (\deg_W(\omega_j, \pi(j_c)) \pm \kappa) \cdot \prod_{c=1}^q \left(R_{j_c} \exp\left(\pm \frac{v(F_{j_c})}{\mathbf{h}_1}\right) \right). \end{aligned}$$

Now, each of the factors in the first product equals to $center_{\pi(j_c)}(\pi(j)) \pm \frac{1}{\mathbf{h}_{L+1-g_j}}$ by using the property (47). Since π is a dense depth preserving anatomy assignment and since $\mu(\pi(j_c))$ is a positive number which depends only on $\mathbf{h}\llbracket L - g_j \rrbracket$, we have

$$\deg_W(\omega_j, \pi(j_c)) \pm \kappa = center_{\pi(j_c)}(\pi(j)) \cdot (1 \pm \sqrt{\kappa}).$$

Since $(1 \pm \sqrt{\kappa})^q = \exp(\pm \frac{1}{\mathbf{h}_1})$, we just verified (66) for j and ω_j .

We can get exactly the same calculations for $B_j(\omega)$, where $\omega \in \pi(j) \setminus X$ and get

$$B_j(\omega_j) = \prod_{c=1}^q (\deg_U(\omega_j, \pi(j_c)) \pm \kappa) \cdot \prod_{c=1}^q \left(R_{j_c} \exp\left(\pm \frac{v(F_{j_c})}{\mathbf{h}_1}\right) \right)$$

Now, the fact ω_j is a non-exceptional element of the cell $\pi(j)$ of the κ -approximate anatomy $\{\mathcal{A}_F\}_{F \in \mathcal{D}_{\mathbf{h}}\llbracket L+1-j \rrbracket}}$ (for U) implies (67).

Clearly, the term $\int_{\omega_1 \in \pi(1) \setminus X} A_1(\omega_1)$ corresponds to the term $\int \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j)$ in the definition of Q_1 . We shall now control all the remaining terms. The idea is that these all are almost constant since we are working inside one anatomy cell. Let us first deal with the terms in the definition of Q_1 .

As $\pi(j)$ is non-singular for each $j \in [\ell]$, the denominator of $\frac{\sum_{j=p}^{\ell} \deg_W(\omega_j)}{\prod_{j=1}^{\ell} \deg_W(\omega_j)}$ is at least Δ^{ℓ} . Further, the fluctuations in the denominator are at most $\sum_{j=1}^{\ell} \frac{\hbar(\mathbf{h}\llbracket L-g_j \rrbracket)}{\mathbf{h}_{L+1-g_j}} \leq \tau$ around $C_1 := \prod_{j=1}^{\ell} \deg_W^{\text{anat}}(\pi(j))$ by (48). Similarly, the nominator is bounded from below by Δ , and the fluctuations in the nominator are at most $\sum_{j=p}^{\ell} \frac{\hbar(\mathbf{h}\llbracket L-g_j \rrbracket)}{\mathbf{h}_{L+1-g_j}} \leq \tau$ around $C_2 := \sum_{j=p}^{\ell} \sum_{F \in \mathcal{D}_{\mathbf{h}}\llbracket L-g_j \rrbracket}} center_F(\pi(j))$. The fluctuations in the term $\exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right)$ are at most τ around $C_3 := \prod_{j=1}^{p-1} center_{\mathfrak{b}}(\pi(j))$. Also, since $\pi(j)$ is non-singular for each $j \in [p-1]$, we have $C_3 > \exp(-\frac{\ell}{\Delta})$. The two upper-lower- bounds we have derived using from the non-singularity will be used below to transform an additive error of the type $value \pm error$ to a multiplicative one, $value \cdot (1 \pm \frac{error}{lower\ bound\ on\ the\ value})$. Therefore,

$$\begin{aligned} Q_1 &= \int_{(\omega_1, \dots, \omega_{\ell}) \in \Omega_{\pi}} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \cdot \frac{\sum_{j=p}^{\ell} \deg_W(\omega_j)}{\prod_{j=1}^{\ell} \deg_W(\omega_j)} \cdot \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \\ &= \int_{(\omega_1, \dots, \omega_{\ell}) \in \Omega_{\pi}} (C_3 \pm \tau) \cdot \frac{C_2 \pm \tau}{C_1 \pm \tau} \cdot \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \\ &= C_3 \cdot \frac{C_2}{C_1} \cdot (1 \pm 6\tau \cdot \exp(\frac{\ell}{\Delta})) \cdot \int_{(\omega_1, \dots, \omega_{\ell}) \in \Omega_{\pi}} \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) \\ &= C_3 \cdot \frac{C_2}{C_1} \cdot (1 \pm 6\tau \cdot \exp(\frac{\ell}{\Delta})) \cdot \int_{\omega_1 \in \pi(1) \setminus X} A_1(\omega_1) \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(66)}{=} C_3 \cdot \frac{C_2}{C_1} \cdot \left(1 \pm 6\tau \cdot \exp\left(\frac{\ell}{\Delta}\right)\right) \cdot \mu(\pi(1) \setminus X) \cdot R_1 \cdot \exp(\pm \frac{\ell}{\mathbf{h}_1}) \\
 &= C_3 \cdot \frac{C_2}{C_1} \cdot \mu(\pi(1) \setminus X) \cdot R_1 \cdot \left(1 \pm \frac{d}{3}\right). \tag{68}
 \end{aligned}$$

We now attempt to show that this quantity is close to Q_2 . First, let us introduce a modified quantity in which even integration over V_0 and E_0 is considered, $Q_2^+ := \int_{\omega \in \Omega_\pi} f_U(\omega)$. Now, we claim that we get exactly the same bound as in (68) even for Q_2^+ . Let us explain that there are no traps in doing so. First, we can use $\int_{\omega_1 \in \pi(1) \setminus X} B_1(\omega_1)$ to control $\int \prod_{(i,j) \in E(T)} U(\omega_i, \omega_j)$, exactly in the same way as we used $\int_{\omega_1 \in \pi(1) \setminus X} A_1(\omega_1)$ to control $\int \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j)$, because (67) is a perfect counterpart to (66). The remaining quantities appearing in the definition of Q_2 are again centered around the constants C_1 , C_2 , and C_3 as above, except the fluctuations can be bigger by κ since $\left\{ \{A_F\}_{F \in \mathcal{D}_{\mathbf{h}[\ell]}} \right\}_{l=0}^{L+1}$ are only κ -approximate anatomies for U (as opposed to exact anatomies for W). However, κ is the smallest constant in (58) and thus this additional error can be easily swallowed by other error terms. Third, the upper- and lower- bounds based on non-nonsingularity that we used to transform an additive errors to multiplicative ones are still valid when the additional approximation term κ is taken into account. Therefore, we conclude that $Q_2^+ = C_3 \cdot \frac{C_2}{C_1} \cdot \mu(\pi(1) \setminus X) \cdot R_1 \cdot \left(1 \pm \frac{d}{3}\right)$. This finishes the proof.

Now, let us show that we have $|Q_2^+ - Q_2| < \frac{(e^{\pi_{\text{vert}}} + e^{\pi_{\text{edge}}})^\ell}{\Delta^\ell}$. Clearly, we have $Q_2^+ \geq Q_2$. On the other hand, if for some $(\omega_1, \dots, \omega_\ell) \in \Omega_\pi$ we have that $f_U^-(\omega_1, \dots, \omega_\ell) = 0$ but $f_U(\omega_1, \dots, \omega_\ell) > 0$ then for some $i \in [\ell]$ we have $\omega_i \in \pi(i) \cap \Lambda_{\text{vert}}$ or for some $ij \in E(T)$ we have $(\omega_i, \omega_j) \in \pi(i) \times \pi(j) \cap \Lambda_{\text{edge}}$. \square

Claim 4.3.3 *Suppose that $\pi : [\ell] \rightarrow \{A_F\}_{F \in \mathcal{D}_{\mathbf{h}[\ell]}}$ is a depth preserving anatomy assignment that is not singular. Suppose that $j^* \in [\ell]$ is a witness that it is not dense either. Then*

$$\int_{\prod_{j=1}^\ell \pi(j)} f_W < \frac{d\ell}{\Delta^\ell} \cdot \prod_{j \in [\ell]} \mu(\pi(j)) + \frac{\ell}{\mathbf{h}_{L+1-g_{j^*}} \cdot \Delta^\ell} \cdot \prod_{j \in [\ell] \setminus \{j^*\}} \mu(\pi(j)). \tag{69}$$

Proof Let us bound all the terms in (56). More precisely, let $kj^* \in E(T)$ be any edge that witnesses that π is not a dense depth preserving anatomy assignment. To bound f_W , we shall use that

$$f_W(\omega_1, \dots, \omega_\ell) \leq \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \cdot \frac{\sum_{j=p}^\ell \deg_W(\omega_j)}{\prod_{j=1}^\ell \deg_W(\omega_j)} \cdot W(\omega_k, \omega_{j^*}).$$

The term $\exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right)$ is trivially at most 1. The nominator in the term $\frac{\sum_{j=p}^\ell \deg_W(\omega_j)}{\prod_{j=1}^\ell \deg_W(\omega_j)}$ is at most ℓ while the denominator is at least Δ^ℓ by non-singularity. Finally, because $kj^* \in E(T)$ is an edge that witnesses that π is not a dense depth preserving anatomy assignment. Then for every choice of $\omega_k \in \pi(k)$, we have that $\int_{\omega_{j^*} \in \pi(j^*)} W(\omega_k, \omega_{j^*}) = \deg_W(\omega_k, \pi(\omega_{j^*})) \leq d \cdot \mu(\pi(j^*)) + \frac{1}{2\mathbf{h}_{L+1-g_{j^*}}}$. The claim now follows by integrating over the remaining dimensions. \square

We are now ready to express $\text{Freq}(T; W)$ and $\text{Freq}^-(T; G, V_0, E_0)$ in a way that most terms will approximately cancel. To express $\text{Freq}(T; W)$, we work with the function f_W

defined in (56). Also, we partition the integration over Ω^ℓ in (1) into terms corresponding to individual depth preserving anatomy assignments as in (62).

$$\begin{aligned}
 |\text{Stab}_T| \cdot \text{Freq}(T; W) &= \sum_{\pi \text{ non-problematic}} \int_{\Omega_\pi} f_W \\
 &+ \sum_{\pi \text{ non-problematic}} \int_{\prod_{j=1}^\ell \pi(j) \setminus \Omega_\pi} f_W \\
 &+ \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_W \\
 &+ \sum_{\pi \text{ singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_W.
 \end{aligned} \tag{70}$$

Now, we have

$$\begin{aligned}
 |\text{Stab}_T| \cdot \text{Freq}^-(T; G, V_0, E_0) &= \sum_{\pi \text{ non-problematic}} \int_{\Omega_\pi} f_U^- \\
 &+ \sum_{\pi \text{ non-problematic}} \int_{\prod_{j=1}^\ell \pi(j) \setminus \Omega_\pi} f_U^- \\
 &+ \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_U^- \\
 &+ \sum_{\pi \text{ singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_U^-.
 \end{aligned} \tag{71}$$

We will show that the first sums in (70) and (71) cancel almost perfectly, and that all the remaining terms are negligible.

Claim 4.3.2 tells us that for the first sums in (70) and (71) we have

$$\sum_{\pi \text{ non-probl.}} \int_{\Omega_\pi} f_W = \sum_{\pi \text{ non-probl.}} \int_{\Omega_\pi} f_U^- \cdot (1 \pm d) \pm \sum_{\pi \text{ non-probl.}} \frac{(e_{\text{vert}}^\pi + e_{\text{edge}}^\pi)^\ell}{\Delta^\ell}. \tag{72}$$

We have

$$\begin{aligned}
 \sum_{\pi \text{ non-probl.}} e_{\text{vert}}^\pi &= \sum_{i=1}^\ell \sum_{\pi \text{ non-probl.}} \mu(\pi(i) \cap \Lambda_{\text{vert}}) \cdot \prod_{k \in V(T) \setminus \{i\}} \mu(\pi(k)) \\
 &\stackrel{(64)}{\leq} \sum_{i=1}^\ell \mu(\Lambda_{\text{vert}}) \stackrel{(60)}{\leq} \ell \cdot a, \text{ and} \\
 \sum_{\pi \text{ non-probl.}} e_{\text{edge}}^\pi &= \sum_{ij \in E(T)} \sum_{\pi \text{ non-probl.}} \mu^{\otimes 2} \left(\Lambda_{\text{edge}} \cap \bigcup_{ij \in E(T)} \pi(i) \times \pi(j) \right) \\
 &\cdot \prod_{k \in V(T) \setminus \{i, j\}} \mu(\pi(k)) \\
 &\leq |E(T)| \mu^{\otimes 2}(\Lambda_{\text{edge}}) \stackrel{(60)}{\leq} \ell \cdot 2a.
 \end{aligned}$$

Recall that in the course of proving Theorems 1.3 from 1.1 we showed that $\int_{\omega \in \Omega} f_W(\omega) \leq |\text{Stab}_T| \leq \ell^\ell$. In particular, using the above bounds, we can turn the multiplicative error in (72) into an additive error,

$$\begin{aligned} \sum_{\pi \text{ non-probl.}} \int_{\Omega_\pi} f_W &= \sum_{\pi \text{ non-probl.}} \int_{\Omega_\pi} f_U^- \pm 2d\ell^\ell \pm \frac{3a\ell^2}{\Delta^\ell} \\ &= \sum_{\pi \text{ non-probl.}} \int_{\Omega_\pi} f_U^- \pm \frac{\varepsilon}{100}. \end{aligned} \tag{73}$$

Let us now focus on the second sum in (70). Observe that the set $\bigcup_{\pi \text{ non-problematic}} \prod_{j=1}^\ell \pi(j) \setminus \Omega_\pi$ has measure at most $\ell \cdot \mu(X) \leq \delta_0$. Therefore, (57) applies, and we get that

$$\sum_{\pi \text{ non-problematic}} \int_{\prod_{j=1}^\ell \pi(j) \setminus \Omega_\pi} f_W \leq \frac{\varepsilon}{10}.$$

Let us now turn to the third sum in (70). We write

$$\sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_W \leq \sum_{g=0}^L \sum_{\pi: \text{witness at height } g} \int_{\prod_{j=1}^\ell \pi(j)} f_W,$$

where the last sum ranges over all non-singular depth preserving anatomy assignments π with a witness for not being dense at height g . By Claim 4.3.3 and by (63) and (64) for each $g \in \{0, \dots, L\}$ we have

$$\sum_{\pi: \text{witness at height } g} \int_{\prod_{j=1}^\ell \pi(j)} f_W \leq \frac{d\ell}{\Delta^\ell} + \frac{\ell^2}{\mathbf{h}_{L+1-g} \cdot \Delta^\ell} \leq \frac{d'}{L+1}.$$

Therefore,

$$\sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_W < d'. \tag{74}$$

The fourth sum of (70) can be bounded from above as follows. If π is singular then by Claim 4.3.1 there exists $i \in [\ell]$ such that we have $b_W(\omega) > 1/(2\Delta)$ for all $\omega \in \pi(i)$, or we have $\pi(i) \subseteq \Omega_{\text{small}}$. Since the average value of b_W is 1 by Proposition 4.2, the measure of the elements that satisfy the former condition is at most 2Δ . Thus, the measure of the set $\bigcup_{\pi} \prod_{j=1}^\ell \pi(j)$, where the union ranges over all depth preserving anatomy assignments π which are singular at coordinate i , is at most $2\Delta + \mu(\Omega_{\text{small}}) < \frac{\delta_0}{\ell}$. We conclude that the measure of $\bigcup_{\pi \text{ singular}} \prod_{j=1}^\ell \pi(j)$ is less than δ_0 . Therefore, (57) applies, and $\sum_{\pi \text{ singular}} \int_{\prod_{j=1}^\ell \pi(j)} f_W \leq \frac{\varepsilon}{10}$. It now remains to bound the second, third, and the fourth term (71). Recall that when bounding the corresponding terms in (70) above, we argued that the domains of integration of the second and the fourth term have measures at most δ_0 each. Combining this with the upper bound (61), we get that the second and the fourth term are at most $\delta_0 \cdot \frac{\ell}{\chi} < \frac{\varepsilon}{100}$, using (58). So, the only term we have to control now is

$$\begin{aligned}
 & \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j)} f_{\bar{U}} \\
 &= \sum_{\pi \text{ not dense and not singular}} \int_{\Omega_{\pi}} f_{\bar{U}} + \sum_{\pi \text{ not dense and not singular}} \int_{\prod_{j=1}^{\ell} \pi(j) \setminus \Omega_{\pi}} f_{\bar{U}}.
 \end{aligned} \tag{75}$$

The first summand can be bounded by $2d'$ in exactly the same way as we derived (74). To see this, recall that on Ω_{π} the degrees into anatomies are similar for W and for U and thus we can make use of the “non-denseness” and “non-singularity” assumption even for $f_{\bar{U}}$ (with slightly worse constants). The second summand of (75) corresponds represents integration over a set of measure at δ_0 . In other words, we can bound the second summand of (75) by $\frac{\varepsilon}{100}$ in the same way we bounded the second and the fourth term of (71).

The lemma now follows by expanding $\text{Freq}(T; W) - \text{Freq}^{-}(T; G, V_0, E_0)$ using (70) and (71) and using the bounds above. □

Theorem 1.1 now follows in a straightforward way by combining Lemmas 3.12 and 4.3.

Proof of Theorem 1.1 Suppose that a graphon W , an ℓ -vertex tree T and $\varepsilon > 0$ are given. Let $a > 0$ be given by Lemma 4.3 for W, T , and $\frac{\varepsilon}{2}$. Lemma 2.11 with parameter $\beta := c_1 \cdot \min\{a^8, \varepsilon^{16}\}$ (for a sufficiently small constant c_1) yields positive constants $\alpha, \varepsilon', \gamma$ and ξ' with $\beta \gg \alpha \gg \varepsilon' \gg \gamma \gg \xi'$. Now, let n_0 and δ' be numbers given by Lemma 4.3 for input parameters as above together with $\chi := \gamma$. Set $\xi := \min(\xi', \delta, \frac{1}{n_0})$.

Suppose that G is a graph with $d_{\circ}(G, W) \leq \xi$. Owing to Lemma 2.11, we know that G has a $(\beta, \alpha, \varepsilon', \gamma)$ -good decomposition $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$. Lemma 3.12 now tells us that with probability at least $1 - \mathcal{O}(\ell\alpha^{1/8})$ the uniform spanning tree \mathcal{T} of G satisfies

$$\mathbb{P}_X(B_{\mathcal{T}}(X, r) \cong T) = (1 + \mathcal{O}(\ell^2\alpha^{1/16}))\text{Freq}(T; G) + \mathcal{O}(\ell^2\beta^{1/8}). \tag{76}$$

Recall that the quantity $\text{Freq}(T; G)$ is defined in (34), and depends on the decomposition $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$. Observe that $\text{Freq}(T; G) = \text{Freq}^{-}(T; G, U_0, E_0)$, where U_0 is the union of V_0 and sets V_i ($i \geq 1$) that are not (α, ε') -big, and E_0 consists of all edges running across two distinct (α, ε') -big sets $\{V_i\}_i$. Proposition 2.12 tells us that $|U_0| = \mathcal{O}(\beta^{1/8}n) < an$. The fact that $V(G) = V_0 \sqcup V_1 \sqcup \dots \sqcup V_k$ is a $(\gamma, \varepsilon'^5, \varepsilon'^5)$ -expander decomposition (in particular, (G2) of Definition 2.6 applies) tells us that $|E_0| \leq \varepsilon'^5 n^2 < an^2$. Thus, Lemma 4.3 gives $\text{Freq}(T; G) = \text{Freq}(T; W) \pm \frac{\varepsilon}{2}$. Plugging this back to (76) and using that the term $(1 + \mathcal{O}(\ell^2\alpha^{1/16}))$ can be bounded by $(1 \pm \frac{\varepsilon}{4})$ and the term $\mathcal{O}(\ell^2\beta^{1/8})$ is smaller than $\frac{\varepsilon}{4}$, we get the theorem. □

4.2 Proof of Theorem 1.3 from Theorem 1.1

Let T be a fixed rooted tree with $\ell \geq 2$ vertices and height r . We denote the vertices of T , as before, by $\{1, \dots, \ell\}$ so that 1 is the root and the vertices of distance r from the root are $\{p, \dots, \ell\}$ for some $2 \leq p \leq \ell$. Our goal is to show that the probability that the first r generations of the branching process yield a tree which is root-isomorphic to T is $\text{Freq}(T; W)$.

First, the factor of $|\text{Stab}_T|^{-1}$ comes from the $|\text{Stab}_T|$ different vertex labellings of T which yield the same event. Secondly, we may split this event to $\ell - p + 1$ disjoint events, indexed by the vertices $q \in \{p, \dots, \ell\}$ at height r of T , and each of these is the event that the path from 1 to q is the first r steps of the ancestral path defined in κ_W . We will show that the probability of each of these events is precisely

$$\int_{\omega_1, \dots, \omega_\ell} \exp\left(-\sum_{j=1}^{p-1} b_W(\omega_j)\right) \cdot \frac{\deg(\omega_q)}{\prod_{j=1}^\ell \deg(\omega_j)} \cdot \prod_{(i,j) \in E(T)} W(\omega_i, \omega_j) d\omega_1 \cdots d\omega_\ell, \tag{77}$$

which is the q -th term in (1) when we expand the sum over j in the numerator. We denote the path from 1 to q in T by x_1, \dots, x_r where $x_1 = 1$ and $x_r = q$. By definition of κ_W , the density function of the first r ancestral particles $\omega_{x_1}, \dots, \omega_{x_r}$ is

$$\prod_{j=1}^{q-1} \frac{W(\omega_{x_j}, \omega_{x_{j+1}})}{\deg(x_j)}.$$

Thirdly, by independence of the branching process, conditioned on this path, the rest of the branching process emanating from the particles on this path is independent and all new particles (if there are any) are ‘‘other’’ particles. Given a particle of type ω (either ancestral or other), the number of its other progeny is distributed as $\text{Poisson}(b_W(\omega))$ random variable, so the probability that it equals $k \geq 0$ is $\frac{e^{-b_W(\omega)} b_W(\omega)^k}{k!}$. Given the the number of its ‘‘other’’ progeny is k , these k points of Ω are distributed as k i.i.d. points drawn from the probability density function $\frac{W(\omega, \omega')}{b_W(\omega) \deg(\omega')}$. Thus, conditioned on having k progeny, the density of these k -tuple $\{\omega_1, \dots, \omega_k\}$ of points is

$$\frac{k!}{b(\omega)^k} \prod_{i=1}^k \frac{W(\omega, \omega_i)}{\deg(\omega_i)}.$$

The $k!/b_W(\omega)^k$ cancels with the $b_W(\omega)^k/k!$ in the probability of having k points and we are left with last term times $e^{-b_W(\omega)}$. We now apply this throughout all the branching points of the tree T (even for $k = 0$) and this concludes the proof (note that we do not have a factor for $e^{-b_W(\omega)}$ for tree vertices at level r since we do not care if they have progeny or not). \square

5 Bounds on the Number of Vertices of a Given Degree in the UST

In this section we prove Theorem 1.5. Let us make two remarks beforehand. Firstly, we cannot hope for converse bounds to the theorem. Indeed, let us take a small $\alpha > 0$ and let us consider the complete bipartite graph $K_{\alpha n, (1-\alpha)n}$ with color classes A and B , $|A| = \alpha n$, $|B| = (1 - \alpha)n$. The handshaking lemma tells us that for *any* spanning tree T of $K_{\alpha n, (1-\alpha)n}$ we have

$$n - 1 = \sum_{b \in B} \deg_T(b) = (1 - \alpha)n + \sum_{b \in B, \deg_T(b) > 1} (\deg_T(b) - 1).$$

In particular, the last sum can have at most $\alpha n - 1$ summands. We conclude that T has more than $(1 - 2\alpha)n$ leaves, and less than $2\alpha n$ vertices of degrees more than 1.

Secondly, as we mentioned in the beginning of this paper, the degree distribution of a UST in a complete graph is approximately $\text{Poisson}(1) + 1$. Considering Theorem 1.5, we see that the complete graph is an asymptotic minimizer for the number of leaves (density e^{-1}), the asymptotic maximizer for the number of vertices of degree 2 (density e^{-1}) and 3 (density $e^{-1}/2$). However, the density of vertices degree $k \geq 4$ in the UST of the complete graph is $e^{-(k-1)}/(k-1)!$ which is smaller than the $(k-2)^{k-2} e^{-(k-2)}/(k-1)!$ given by Theorem 1.5. Examining the proof shows that the bounds in Theorem 1.5 for $k \geq 4$ are optimal in the following sense. Let $\alpha > 0$ be arbitrarily small, and let $G_{n,k,\alpha}$ be the graph obtained by

taking a complete graph on $n/(k - 2)$ vertices and additional $n(k - 3)/(k - 2)$ vertices such that each of the $n^2(k - 3)$ possible edges between the two parts is retained with probability α and erased otherwise, independently of all other edges. It is a straightforward computation using Theorem 1.3 (see also the proof of Theorem 1.5 below) to show that for n large, with high probability the proportion of vertices of degree k in a UST will be arbitrarily close, as $\alpha \rightarrow 0$, to $\frac{1}{(k-1)!}((k - 2)/e)^{k-2}$.

We are almost ready to prove the theorem. We will need the following folklore statement that provides a relation between degrees in a graphon and finite graphs that converge to it. Lemma 5.1 is easy to prove directly, but let us note that is a straightforward consequence of Lemma 4.1 for anatomies of depth 1.

Lemma 5.1 *Suppose that G_1, G_2, \dots are finite graphs converging to a graphon $W : \Omega^2 \rightarrow [0, 1]$.*

- (1) *Suppose that $0 \leq a < b \leq 1$ are given. and let d be the measure of points of Ω whose degree is in the interval (a, b) . Then for an arbitrary $\varepsilon > 0$ there exists an n_0 such that for every $n > n_0$ we have that in G_n there are at least $(d - \varepsilon)v(G_n)$ many vertices with degrees in the interval $((a - \varepsilon)v(G_n), (b + \varepsilon)v(G_n))$ and there are at most $(d + \varepsilon)v(G_n)$ many vertices with degrees in the interval $((a + \varepsilon)v(G_n), (b - \varepsilon)v(G_n))$.*
- (2) *Suppose that there exists $\alpha > 0$ such that in each G_n , all but at most $\mathbf{o}_n(1)v(G_n)$ vertices have degrees at least $\alpha v(G_i)$. Then W is nondegenerate.*

Proof of Theorem 1.5 We start by proving (3). Assume by contradiction that there are $\varepsilon_0, \delta_0 > 0$ such that there exists a sequence of graphs G_n (which we assume have n vertices) such that

- (A) at least $(1 - \mathbf{o}(1))n$ vertices of G_n are of degree at least $\delta_0 n$, and,
- (B) $\mathbb{P}(L_1(n) \leq (e^{-1} - \varepsilon_0)n) > \varepsilon_0$,

where $L_1(n)$ is the number of leaves in a UST of G_n . By Theorem 1.6 we may assume without loss of generality that G_n is a converging sequence and let W be the corresponding limit graphon. By Lemma 5.1(2) and assumption (A) we deduce that W is nondegenerate. Let \mathcal{T}_n be a UST of G_n . By Theorem 1.3 the sequence \mathcal{T}_n almost surely satisfies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1) = \mathbb{E}_{x \in \Omega} [\mathbb{P}[1 + \text{Poisson}(b_W(x)) = 1]] = \mathbb{E}_{x \in \Omega} [e^{-b_W(x)}],$$

where X is a uniformly chosen vertex of G_n . By Proposition 4.2 we have that $\mathbb{E}_{x \in \Omega}[b_W(x)] = 1$. Hence Jensen’s inequality implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1) \geq e^{-1}.$$

Since $L_1(n) = n\mathbb{P}(\deg_{\mathcal{T}_n}(X) = 1)$ we arrive at a contradiction to assumption (B), showing (3).

The proof of (4) and (5) proceeds similarly. We fix $k \geq 2$ and make the an analogous contradictory assumption as in (A) and (B). We have that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\deg_{\mathcal{T}_n}(X) = k) &= \mathbb{E}_{x \in \Omega} [\mathbb{P}[1 + \text{Poisson}(b_W(x)) = k]] \\ &= \frac{1}{(k - 1)!} \mathbb{E}_{x \in \Omega} [e^{-b_W(x)} b_W(x)^{k-1}]. \end{aligned}$$

When $k = 2$, since $ye^{-y} \leq e^{-1}$ for all $y \geq 0$ we learn that the last quantity is at most e^{-1} , proving (4) by contradiction. When $k \geq 3$, it follows by Lemma 5.2 (provided immediately below) that the last quantity is at most $\frac{1}{(k-1)!}((k - 2)/e)^{k-2}$, similarly proving (5). \square

The last missing piece is to state and prove the optimization result that was key for our proof of Theorem 1.5 above.

Lemma 5.2 *Let $b : [0, 1] \rightarrow [0, \infty)$ satisfy $\int_{[0,1]} b(x)dx = 1$. Then for any $k \geq 2$,*

$$\int_{[0,1]} e^{-b(x)} b(x)^k dx \leq \left(\frac{k-1}{e}\right)^{k-1}.$$

Proof For ease of notation, let $f(x) = e^{-x} x^k$. Fix $\kappa > 0$. Approximating $b(x)$ by simple functions, we can find $n \in \mathbb{N}$ and n non-negative numbers b_1, \dots, b_n such that $\frac{1}{n} \sum_{i=1}^n b_i = 1$, and $\int_{[0,1]} f(b(x))dx \leq \frac{1}{n} \sum_{i=1}^n f(b_i) + \kappa$. Hence

$$\int_{[0,1]} e^{-b(x)} b(x)^k dx \leq \max_{(b_1, \dots, b_n) \in \mathcal{C}} \Phi(b_1, \dots, b_n) + \kappa,$$

where $\Phi(b_1, \dots, b_n) := \frac{1}{n} \sum_{i=1}^n f(b_i)$ and $\mathcal{C} := \{(b_1, \dots, b_n) \in [0, +\infty)^n : \frac{1}{n} \sum_{i=1}^n b_i \leq 1\}$. To prove the lemma, it thus suffices to show that $\Phi|_{\mathcal{C}} \leq \left(\frac{k-1}{e}\right)^{k-1}$.

By abuse of notation, we use the same symbol $\vec{b} = (b_1, \dots, b_n)$ to denote the maximizer of $\Phi|_{\mathcal{C}}$. As $f'(x) = e^{-x} x^{k-1} (k - x) < 0$ for every $x > k$, $f(x)$ is decreasing on (k, ∞) , and consequently

$$0 \leq b_i \leq k \quad \text{for all } i \in [n]. \tag{78}$$

There are two possibilities.

Case 1 There exists a non-negative number y so that $b_i \in \{0, y\}$ for every $i \in [n]$.

Let $\lambda \in [0, 1]$ denote the proportion of $i \in [n]$ with $b_i = y$. Then $\lambda y = \frac{1}{n} \sum_{i=1}^n b_i \leq 1$. Hence

$$\Phi(\vec{b}) = \lambda e^{-y} y^k \leq e^{-y} y^{k-1} \leq \left(\frac{k-1}{e}\right)^{k-1},$$

as required.

Case 2 b_i receives at least two positive values.

Let i and j be two arbitrary indices such that $1 \leq i < j \leq n$ and $\min\{b_i, b_j\} > 0$. For every small $\varepsilon > 0$, two vectors $\vec{b}' := (b_1, \dots, b_i + \varepsilon, \dots, b_j - \varepsilon, \dots, b_n)$ and $\vec{b}'' := (b_1, \dots, b_i - \varepsilon, \dots, b_j + \varepsilon, \dots, b_n)$ clearly belong to the domain \mathcal{C} . By Taylor's formula and the optimality of \vec{b} , we thus obtain

$$\begin{aligned} 0 \leq \Phi(\vec{b}) - \Phi(\vec{b}') &= (f'(b_j) - f'(b_i))\varepsilon + O_{b_i, b_j}(\varepsilon^2), \\ 0 \leq \Phi(\vec{b}) - \Phi(\vec{b}'') &= -(f'(b_j) - f'(b_i))\varepsilon + O_{b_i, b_j}(\varepsilon^2). \end{aligned}$$

This forces $f'(b_i) = f'(b_j)$.

As $f''(x) = e^{-x} x^{k-2} ((x - k)^2 - k)$, we see that $f''(x) > 0$ for $x \in (0, k - \sqrt{k})$, and $f''(x) < 0$ for $x \in (k - \sqrt{k}, k)$. It follows that $f'(x)$ is increasing in $[0, k - \sqrt{k}]$, and decreasing in $[k - \sqrt{k}, k]$. Hence for each $\alpha \in \mathbb{R}$, the equation $f'(x) = \alpha$ has at most two solutions in $[0, k]$.

From the discussion above and (78), we learn that there are two numbers y and z satisfying the following properties:

- (i) $b_i \in \{0, y, z\}$ for every $i \in [n]$,
- (ii) $f'(y) = f'(z)$, that is, $e^{-y} y^{k-1} (k - y) = e^{-z} z^{k-1} (k - z)$,
- (iii) $0 < y < z \leq k$.

Let λ and μ be the proportions of $i \in [n]$ such that $b_i = y$ and $b_i = z$, respectively. Then $\lambda + \mu \leq 1$, $\lambda y + \mu z \stackrel{(i)}{=} \frac{1}{n} \sum_{i=1}^n b_i \leq 1$, and $\Phi(\vec{b}) \stackrel{(i)}{=} \Psi(\lambda, \mu)$, where $\Psi(\lambda, \mu) := \lambda f(y) + \mu f(z)$.

We will view y and z as constants, and seek to maximize $\Psi(\lambda, \mu)$ under the constraints: $\lambda, \mu \geq 0$, $\lambda + \mu \leq 1$, and $\lambda y + \mu z \leq 1$. Let (λ_0, μ_0) be a maximizer of Ψ . We claim that $\min\{\lambda_0, \mu_0\} = 0$. Before proving the claim, let us show how it implies the lemma. Indeed, if one of λ_0 and μ_0 is zero, say μ_0 , then

$$\Phi(\vec{b}) = \Psi(\lambda, \mu) \leq \Psi(\lambda_0, \mu_0) = \lambda_0 e^{-y} y^k \leq e^{-y} y^{k-1} \leq \left(\frac{k-1}{e}\right)^{k-1},$$

as desired.

It remains to prove that $\min\{\lambda_0, \mu_0\} = 0$. Suppose to the contrary that $\lambda_0, \mu_0 > 0$. Let $\lambda_1 = \lambda_0 - \varepsilon$ and $\mu_1 = \mu_0 + \varepsilon y/z$, where $\varepsilon > 0$ is a small constant. It is not difficult to verify that $\lambda_1, \mu_1 \geq 0$, $\lambda_1 + \mu_1 \leq 1$, and $\lambda_1 y + \mu_1 z \leq 1$. Thus, by the optimality of (λ_0, μ_0) , we get $\Psi(\lambda_1, \mu_1) \leq \Psi(\lambda_0, \mu_0)$, giving $e^{-z} z^{k-1} \leq e^{-y} y^{k-1}$. Combined with (iii), we obtain $e^{-z} z^{k-1} (k-z) < e^{-y} y^{k-1} (k-y)$, contradicting (ii). This completes our proof of the lemma. \square

Remark 5.3 Our proof of Theorem 1.5 crucially relied on our running assumption that almost all the degrees in the graph G are linear. It seems natural to investigate similar extremal questions with a weakened form of this assumption. The minimal meaningful assumption seems to be that G contains no vertices of degree 2; this is to avoid the case of paths. This general setting seems much more complicated. For example, the proportion of leaves in a uniform spanning tree on an $n \times n$ torus is asymptotically almost surely $(1 - 2/\pi) \cdot \frac{8}{\pi^2} + o(1) \approx 0.294$ (see for example [18, p. 112]), which is less than the lower bound of $e^{-1} \approx 0.368$ in Theorem 1.5.

Acknowledgements Hladký and Tran were supported by the Czech Science Foundation, Grant Number GJ16-07822Y and by institutional support RVO:67985807. Hladký is supported by the Alexander von Humboldt Foundation. Nachmias is supported by ISF Grant 1207/15, and ERC starting Grant 676970 RANDGEOM. Part of this work was done while Hladký and Tran were visiting Tel Aviv University. They thank the department of mathematical sciences at TAU for their hospitality. The trip was funded by the Czech Academy of Sciences. We thank Omer Angel and Tom Hutchcroft for their assistance with the proof of Lemma 3.2 and to Michael Krivelevich and Wojciech Samotij for useful discussions.

References

1. Aldous, D., Lyons, R.: Processes on unimodular random networks. *Electron. J. Probab.* **12**(54), 1454–1508 (2007)
2. Alon, N.: The number of spanning trees in regular graphs. *Random Struct. Algorithms* **1**(2), 175–181 (1990)
3. Alon, N., Milman, V.D.: λ_1 , isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory Ser. B* **38**(1), 73–88 (1985)
4. Benjamini, I., Schramm, O.: Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6**(23), 1–13 (2001)
5. Bollobás, B., Borgs, C., Chayes, J., Riordan, O.: Percolation on dense graph sequences. *Ann. Probab.* **38**(1), 150–183 (2010)
6. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms* **31**(1), 3–122 (2007)
7. Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztegombi, K.: Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219**(6), 1801–1851 (2008)

8. Frieze, A.: On the number of perfect matchings and Hamilton cycles in ε -regular non-bipartite graphs. *Electron. J. Combin.* **57**, 11 (2000)
9. Grimmett, G.R.: Random labelled trees and their branching networks. *J. Aust. Math. Soc. Ser. A* **30**(2), 229–237 (1980/81)
10. Kirchhoff, G.: Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Ann. Phys. Chem.* 72(12)
11. Kolchin, V.F.: Branching processes, random trees, and a generalized scheme of arrangements of particles. *Math. Notes Acad. Sci. USSR* **21**(5), 386–394 (1977)
12. Kostochka, A.V.: The number of spanning trees in graphs with a given degree sequence. *Random Struct. Algorithms* **6**(2–3), 269–274 (1995)
13. Lawler, G., Sokal, A.: Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality. *Trans. Am. Math. Soc.* **309**, 557–580 (1988)
14. Levin, D., Peres, Y., with contributions by E. Wilmer: *Markov Chains and Mixing Times*. American Mathematical Society, 2nd edn (2017). <http://darkwing.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf>
15. Lovász, L.: *Large Networks and Graph Limits*. American Mathematical Society, Providence, RI (2012)
16. Lovász, L., Szegedy, B.: Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96**(6), 933–957 (2006)
17. Lyons, R.: Asymptotic enumeration of spanning trees. *Combin. Prob. Comput.* **14**(4), 491–522 (2005)
18. Lyons, R., Peres, Y.: *Probability on Trees and Networks*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42. Cambridge University Press, New York (2016)
19. Matoušek, J., Nešetřil, J.: *Invitation to Discrete Mathematics*, 2nd edn. Oxford University Press, Oxford (2009)
20. Sinclair, A., Jerrum, M.: Approximate counting, uniform generation and rapidly mixing Markov chains. *Inf. Comput.* **82**, 93–133 (1989)