Probabilistic graph inspections through forests

Koperberg, V.T.

PART I

KIRCHHOFF FORESTS

CHAPTER 2

Loop-erased partitioning via parametric spanning trees

This chapter is based on the following paper: L. Avena, J. E. P. Driessen, and V. T. Koperberg. "Loop-erased partitioning via parametric spanning trees: Monotonicities & 1D-scaling". In: *Stochastic processes and their applications* 176 (2024), p. 104436.

Abstract

We consider a parametric version of the UST (Uniform Spanning Tree) measure on arbitrary directed weighted finite graphs with tuning (killing) parameter $q > 0$. This is obtained by considering the standard random weighted spanning tree on the extended graph built by adding a ghost state † and directed edges to it, of constant weights q, from any vertex of the original graph. The resulting measure corresponds to a random spanning rooted forest of the graph where the parameter q tunes the intensity of the number of trees as follows: partitions with many trees are favoured for $q > 1$, while as $q \to 0$, the standard UST of the graph is recovered. We are interested in the behaviour of the induced random partition, referred to as loop-erased partitioning, which gives a correlated cluster model, as the multiscale parameter $q \in [0, \infty)$ varies.

Emergence of giant clusters in this correlated percolation model as a function of q has been recently explored on certain dense growing graphs [7]. Herein we derive two types of results. First, we explore monotonicity properties in q of this forest measure on general graphs showing in particular some counter-intuitive subtleties in non-reversible settings where the electrical-network interpretation of the UST observables gets partially lost. Second, by analyzing two-point correlations on trees and various very sparse growing graph models, we characterize emerging macroscopic clusters, as q scales with the graph size, and derive related phase diagrams.

2.1 Rooted spanning forests, loop-erased partitioning and weighted spanning tree measures

Consider an arbitrary directed weighted finite graph $G = (V, E, w)$ on $n = |V|$ vertices where $E \subseteq \{e = (x, y) : x, y \in V\}$ stands for the edge set and $w : E \to [0, \infty)$ is a given edge-weight function. We call Random Walk (RW) associated to G the continuous-time Markov chain $X = (X_t)_{t\geq 0}$ with state space V and infinitesimal generator L given by the negative of the graph Laplacian, i.e. L is the $n \times n$ matrix with off-diagonal entries $L_{xy} = w(x, y)$ and diagonal entries $L_{xx} = -\sum_{z\in V\setminus\{x\}} w(x, z)$ guaranteeing that the entries of each row in L sum up to 0.

A *spanning rooted forest* of a graph is a union of vertex disjoint rooted trees spanning its vertex set, where we consider a rooted tree to be a collection of directed edges pointing towards the root. That is, a rooted forest F is a subset of E such that:

(i) each vertex has at most one outgoing edge in F;

(ii) if there exists a directed path in F from vertex x to vertex y, then no such path exists from y to x.

The *roots* of F are those vertices without an outgoing edge. Let $\mathcal{F}$ denote the collection of all spanning rooted forests of G.

Definition 2.1.1 (Rooted Spanning Forest of intensity q). Fix a positive parameter $q > 0$ and let Φ_q be the random variable with values in $\mathcal{F}$ with law:

$$\mathbb{P}(\Phi_q = F) = \frac{q^{r(F)} w(F)}{Z(q)}, \qquad F \in \mathcal{F}, \tag{2.1}$$

where $w(F) := \prod_{e\in F} w(e)$ stands for the forest weight, $r(F)$ denotes the number of trees (or equivalently the number of roots) in $F \in \mathcal{F}$, and $Z(q)$ is a normalizing constant referred to as the partition function. We will refer to this measure as *random spanning rooted forest* of intensity q. ■

In the unitary weight case $w \equiv 1$, when $q = 1$, this measure becomes uniform over the set of spanning rooted forests $\mathcal{F}$ and its structure has been partially analyzed in several geometrical setups in relation to random combinatorial models in statistical physics and coalescence theory, see [18, 40, 41, 44, 45, 56, 71, 72]. For any $q > 0$, Φ_q induces a randomized decomposition of a given network into blocks (corresponding to its trees) and for each block it identifies a representative node (the root of a tree). The presence of the tuning parameter q makes this object natural for exploring a network architecture in a multiscale fashion. The goal of this paper is to understand the structure of the resulting unrooted random blocks on the set of partitions $\mathcal{P}(V)$ of the vertex set V as the scaling parameter q varies. We refer to this object, defined below, as the Loop-Erased Partitioning (LEP). Its analysis has been initiated on dense graphs in the recent [7]. In this work we derive general results on the monotonicity properties of this measure (see Section 2.2.1) and then, by means of these and other properties, we perform a systematic analysis of the emergent partition on various very sparse simple growing topologies (see Section 2.2.4).

Definition 2.1.2 (Loop-Erased Partitioning (LEP) of intensity q). Given $G = (V, E, w)$, fix a positive parameter $q > 0$. We call *loop-erased partitioning of intensity* q, the random unrooted partition, denoted by Π_q, of V, with law:

$$\mathbb{P}(\Pi_q = \pi_m) = \frac{q^m \sum_{F \in \mathcal{F}: \pi(F) = \pi_m} w(F)}{Z(q)}, \qquad \pi_m \in \mathcal{P}(V),\ m \leq |V|, \tag{2.2}$$

where the sum runs over the space of spanning rooted forests $\mathcal{F}$ of G and $\pi(F)$ stands for the partition of V induced by a given spanning rooted forest F where each block is determined by vertices belonging to the same tree, and m counts the number of blocks in the partition π_m. Equivalently,

$$\Pi_q := \pi(\Phi_q). \tag{2.3}$$

■

Rooted forests and spanning tree measure with uniform killing.

The rooted forest Φ_q is a natural extension of the classical UST (Uniform Spanning Tree) measure which on strongly connected graphs is readily recovered in the constant weight case $w \equiv 1$ by taking the limit of q going to zero in Eq. (2.1). Alternatively, this rooted forest Φ_q can also be seen as a measure on weighted spanning trees on the extended weighted graph obtained by adding an extra cemetery state accessible from any vertex via an edge with weight q. Under this perspective, it is clear that most results known for the UST do have a generalized analogue in the context of this rooted forest measure. For example, edges in Φ_q form a determinantal process [6] due to a version of the so-called transfer-current theorem [17], clarifying its status within negatively associated systems, see [31, 42, 68]. Due to the Kirchhoff's matrix tree theorem, the normalizing constant in Eq. (2.2) can be expressed as the characteristic polynomial of the matrix L evaluated at q, i.e.

$$Z(q) := \sum_{F \in \mathcal{F}} q^{r(F)} w(F) = \det[qI - L], \tag{2.4}$$

see e.g. [6, 19]. As far as sampling is concerned, for fixed $q > 0$, one can use the celebrated algorithm due to Wilson [77] based on loop-erased random walks. The latter is in fact a classical efficient procedure allowing to sample a rooted tree of a graph with probability proportional to its weight. Further, it is well known that the UST can be obtained from the unifying Fortuin-Kasteleyn-percolation 'super-model' by properly taking the related interaction parameter to zero, see e.g. [30]. Not surprisingly, as expressed in Lemma 2.1 below, which for simplicity we state in the unitary weight case $w \equiv 1$, the rooted forest in Eq. (2.1) can also be obtained via a similar zero-limit but by considering a proper FK-percolation with an additional cemetery state. The proof of this proposition is as in [30], see Thm. 1.23 in Sect 1.5 therein, with the parameters of the FK as specified in the statement below.

Lemma 2.1 (Rooted forest as zero-limit of extended FK-percolation). *Given an undirected simple graph $G = (V, E)$, let $G_\dagger := (V_\dagger, E_\dagger)$ be the extended graph with*

$V_\dagger := V \cup \{\dagger\}$ where $\dagger$ denotes an extra state, $E_\dagger := E \cup \bar{E}$ with $\bar{E} := \{(x,\dagger) : x \in V\}$. Consider the generalized FK-percolation on $G_\dagger$ with parameter $\lambda > 0$ and vector of weights $\vec{p} = (p_e)_{e\in E_\dagger}$ such that $p_e = p \in (0,1)$ if $e \in E$ and $p_e = \gamma > 0$ for $e \in \bar{E}$, that is, the following measure on spanning subgraphs of $G_\dagger$ seen as collection of edges in $\mathcal{G} := 2^{E_\dagger}$:

$$\mathbb{P}(FK = H) = \frac{\lambda^{k(H)} \prod_{e\in E} p^{\mathbf{1}\{e\in H\}}(1-p)^{1-\mathbf{1}\{e\in H\}} \prod_{e\in \bar{E}} \gamma^{\mathbf{1}\{e\in H\}}(1-\gamma)^{1-\mathbf{1}\{e\in H\}}}{Z(\lambda,\vec{p})}, \tag{2.5}$$

for $H \in \mathcal{G}$, with $k(H)$ counting the number of connected components of the graph $(V_\dagger, H)$ and $Z(\lambda,\vec{p})$ being a normalizing constant. Assume that $\vec{p}$ is a function of λ such that, as $\lambda \to 0$, $\gamma = \gamma(\lambda) \to 0$, $p = p(\lambda) \to 0$ and $\gamma(\lambda)/p(\lambda) \to q \in (0,\infty)$. Then as λ goes to zero, the law in Eq. (2.5) *(projected onto subgraphs of G) degenerates into the law of the random rooted forest Φ_q in Eq.* (2.1) *with unitary weights.*

If the UST represents the "static global random backbone" of a given network, the forest process $(\Phi_q)_{q>0}$ can be seen as its "mesoscopic and dynamic" analogue where the notion of locality is captured parametrically by what the RW sees on time-scale $1/q$. As such, it naturally leads to dynamic multiscale approaches (see [6, Thm.2] for an associated coalescence-fragmentation process), and new structures and questions which do not make sense within the more restrictive global and static UST context.

Applications of rooted forest measure and LEP:

In a series of recent works [1, 5, 6] some general properties of the rooted forest measure have been explored. For example, the roots [6, Prop.2.2] in Φ_q form a determinantal point process with kernel related to the RW Green's function, that is: for any $A \subset V$

$$\mathbb{P}(A \text{ is in the set of roots}) = \det[K_q]_A, \tag{2.6}$$

with $[K_q]_A$ being the restriction of the matrix $K_q := q(qI - L)^{-1}$ to the set of indices in A. The number of roots (or trees, or blocks in Π_q) is distributed as the sum of n Bernoulli random variables with success probabilities $\frac{q}{q+\lambda_i}$, for $i \leq n$, with the λ_i's being the eigenvalues of $-L$, or their real parts, see e.g. [6, Prop. 2.1]. Its mean number is monotonically increasing in q. Further, these roots turn out to be well-distributed in the given network [6, Thm.1] and, conditional on the induced partition, their joint law is determined by the stationary measures of the random walk X restricted to each block of the underlying partition [6, Prop.2.3]. These and other features of the LEP have been recently exploited to build novel algorithms for the following different applications in data science: multiresolution scheme, wavelets basis and filters for signal processing on graphs [2, 69, 70], estimate traces of discrete Laplacians and other diagonally dominant matrices [10], network renormalization [1, 3], centrality measures [19] and statistical learning [8]. These applications give further motivations to explore this LEP in more detail. Let us also stress that on general graphs or certain specific settings such as the integers, it would be of interest to study the LEP in connection to other random partitions and a natural line of investigation would be to study its intriguing dynamical structure [6, see Thm.2 and Sect 2.2] within the theory of coalescence-fragmentation processes.

Other forest measures:

To conclude this introduction let us clarify that this *rooted* forest Φ_q should not be confused with other forest measures that have been receiving a large amount of attention in the literature in relation to universality classes in statistical physics and to negatively correlated systems. In particular, when taking the (weak) infinite volume limit of the UST on d-dimensional lattices for $d > 4$ (and other transitive settings), depending on the boundary condition procedure when approaching the limit, the resulting measure concentrates on *unrooted* forests referred to as wired or free spanning forests, see e.g. [13, 14, 37, 38, 67] and references therein. On finite graphs another natural extension of the UST is obtained when considering the uniform measure on unrooted spanning forests. Properties of this other fascinating forest measure have been investigated in [12] and recently in [11].

Results overview and paper structure:

The statements of our main results are organized in Section 2.2 divided into the following four subsections.

Monotonicities on arbitrary graphs: We start in Subsection 2.2.1 by presenting results regarding the monotonicity properties of the rooted forest measure in the intensity parameter q. In particular: in Lemma 2.2 we state a general formula to characterize q-monotone events; in Theorem 2.3 we establish monotonicities on *undirected graphs* for both the edge process and the 2-point correlation function, which will later be analyzed in different graph settings to study the emergent partition; and in Remark 2.4 we then discuss subtle issues and counterexamples when trying to extend the latter results to a general non-reversible setting.

Connectivity function on trees: In Section 2.2.2 we look at general weighted directed trees. We state that the connectivity function is still monotone on this directed sub-class of graphs, Theorem 2.5, and we present a related inclusion-exclusion reduction formula, see Theorem 2.6.

$q-$**scaling on growing segments:** Section 2.2.3 is devoted to a detailed analysis of the LEP on the first n integers where equipartitions are favored. Formulas for the partition function are first derived in Theorem 2.7, and extended to a ring, Corollary 2.7.1. Theorem 2.8 gives a recursive representation of the pairwise correlation in terms of reduced partition functions and offers bounds in terms of the corresponding RW on the infinite line. The subsequent Corollary 2.8.1 shows explicit bulk and boundary asymptotics.

Emergence of giants on modular growing tree-like structures: Section 2.2.4 is then devoted to the exploration of the emergent blocks and detection of simple modular structures in certain insightful models for the proper $q-$scales. In particular, Proposition 2.10 and Theorem 2.11 look at a star graph without and with a community structure, respectively. Proposition 2.12 and Theorem 2.13 show similar analysis on finite trees with different weighted structures in which for different magnitudes of q different layers are detected. Finally, in Theorem 2.9

we consider asymptotic detection in a bottleneck graph with two variable-size connected complete subgraphs by combining the results on the segment, after suitable contraction, and those for the mean-field case obtained in [7].

All proofs are given in Section 2.3. Technical short lemmas are provided in the appendices.

2.2 Results: monotonicities & emergent partition

2.2.1 Monotonicity & connectivity function

A notoriously difficult issue for most of the measures that can be obtained from FK-percolation, is to establish monotonicity properties as a function of the involved parameters. The following lemma, which is reminiscent of Russo's pivotality formula in percolation models [15, 29], offers a general characterization of monotone events w.r.t. Φ_q as a function of q.

Lemma 2.2 (Monotone events for the rooted forest on arbitrary networks). *Let $G = (V, E, w)$ be a weighted directed graph, and let $r_q := r(\Phi_q)$ be the number of roots of the random rooted forest Φ_q. Then, for any set of rooted forests $\mathcal{H} \subseteq \mathcal{F}$, it holds that the derivative w.r.t q of the probability of the event $\Phi_q \in \mathcal{H}$ is given by*

$$\frac{d}{dq}\mathbb{P}(\Phi_q \in \mathcal{H}) = \tfrac{1}{q}\mathbb{P}(\Phi_q \in \mathcal{H})\big(\mathbb{E}[r_q \mid \Phi_q \in \mathcal{H}] - \mathbb{E}[r_q]\big). \tag{2.7}$$

This statement, which is proven in Section 2.3.1, is a special case of the more general fact that for Gibbsian measures μ_β with Hamiltonian H and inverse-temperature β it holds that $\frac{\mathrm{d}}{\mathrm{d}\beta}\int X \,\mathrm{d}\mu_\beta = -\int HX \,\mathrm{d}\mu_\beta$. The main advantage of Lemma 2.2 compared to this more general statement is the geometrical interpretation of the right hand side in terms of root numbers. It shows that monotone events in q are those for which the difference $\mathbb{E}[r_q \mid \Phi_q \in \mathcal{H}] - \mathbb{E}[r_q]$ has a constant sign as q varies. In practice it might be not straightforward to check the sign of this difference, since it requires control on the conditional distribution of r_q. Still, for specific events we believe this statement can be of great help, of which we give an example in the proof of Theorem 2.3. We also mention that in [6] a coupled version of the forest[1] is constructed by means of an algorithm allowing to sample an entire forest trajectory $(\Phi_q)_{q\in[0,\infty)}$. Yet, this coupling is monotone only in mean, but not trajectory-wise, hence this coupling is not useful to characterize monotone events.

As anticipated, our main interest within this work is to explore monotonicity properties of this loop-erased partitioning and its detailed structure on trees and nearly-one-dimensional geometries. To do so, we will mainly analyze 2-point correlations associated to Π_q, which we introduce next. For a pair of distinct vertices $x, y \in V$, consider the event that these vertices belong to different blocks in Π_q. That is, the

[1] This coupling corresponds to an explicit Markovian coalescence-fragmentation process with values in $\mathcal{F}$ in which coalescence of trees is dominant but whenever the underlying building RW produces a loop, a tree gets fragmented into subtrees, see [6, see Thm.2 and Sect2.2].

event

$$\{B_q(x) \neq B_q(y)\} := \{x \text{ and } y \text{ are in different blocks of } \Pi_q\},$$

where $B_q(z)$ stands for the block in Π_q containing $z \in V$.

Definition 2.2.1 (2-point correlation or connectivity function). For given $q > 0$ and G, and any pair $x, y \in V$, we call *connectivity function* the following probability:

$$\begin{aligned} U_q(x,y) :=& \mathbb{P}(B_q(x) \neq B_q(y)) \\ =& \sum_{\gamma} \mathbb{P}_x^{LE_q}(\gamma)\mathbb{P}_y(\tau_\gamma > \tau_q) \end{aligned} \tag{2.8}$$

where τ_q denotes an independent exponential random variable of rate q, $\mathbb{P}_z$ and $\mathbb{P}_z^{LE_q}$ stand for the laws of the RW X and the corresponding loop-erased RW killed at rate q, respectively, starting from $z \in V$. Further, the above sum runs over all possible self-avoiding paths γ starting at x and $\tau_\gamma := \inf\{t \geq 0 : X_t \cap \gamma \neq \varnothing\}$ is the random walk hitting time of the set of vertices in γ. ■

The representation in Eq. (2.8) is a consequence of Wilson's sampling procedure and it holds true since, remarkably, this algorithm is exchangeable with respect to the starting point of each loop-erased random walk launched along the algorithm steps [77]. Furthermore, we notice that, as for any generic random partition of V, such a connectivity function defines a distance on the vertex set. This specific metric $U_q(x,y)$ can be interpreted as an affinity measure capturing how densely connected vertices x and y are in the graph G.

Our second general result, Theorem 2.3, further explores monotonicities in q when considering undirected networks. Since spanning rooted forests impose a directionality on its edges, it is convenient to interpret an undirected graph as a symmetric directed graph with a symmetric weight function, $w(x,y) = w(y,x)$ for $(x,y) \in E$. For these symmetric graphs Theorem 2.3 states that the "unoriented" edge process, see (2.9), as well as the 2-point correlations, see (2.10), are both monontone in q. To state the result about the edge process, we will use the following notation. For a directed edge $e = (x,y)$ write $e^- = (y,x)$ to denote its reversed edge, and let $\{\pm A \subseteq \Phi_q\} = \bigcap_{e\in A}(\{e \in \Phi_q\} \cup \{e^- \in \Phi_q\})$ denote the event that for each edge $e \in A$ either e or e^- is present in the random rooted forest Φ_q.

Theorem 2.3 (Monotonicity of edges and 2-point correlations on undirected networks). *Consider a symmetric weighted directed graph $G = (V,E,w)$ and the rooted forest Φ_q on G for $q \in [0,\infty)$. Let $A \subseteq E$ be a set of directed edges, then the function*

$$q \mapsto \mathbb{P}(\pm A \subseteq \Phi_q) \tag{2.9}$$

is monotone non-increasing. Furthermore, for any distinct $x, y \in V$, the function

$$q \mapsto U_q(x,y) \tag{2.10}$$

is continuous and non-decreasing with $U_0(x,y) = 0$ and $\lim_{q\to\infty} U_q(x,y) = 1$.

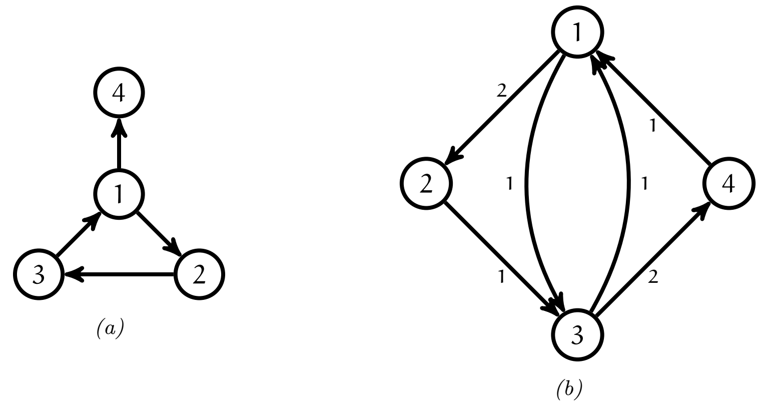


Figure 2.1: Two directed graphs for which the edge monotonicity (2.9) *does* not *hold.*

The monotonicities presented in Theorem 2.3 on undirected graphs are not too surprising, as this result could be derived from the well-known negative association result of Feder and Mihail for balanced matroids, see [60, Theorem 4.6 and Lemma 10.3]. This becomes apparent when we interpret the killing rate as edge weights on an extended graph (as in Lemma 2.1), and in turn interpret all edge weights as multi-edges, which is straightforward for integer edge weights, but can also be achieved for real valued edge weights via rational approximation and up-scaling of rational edge weights into integers.

Remark 2.4 (Beyond reversibility & main open problem). The proof of Theorem 2.3 presented in Section 2.3.1 will not use the result of Feder and Mihail, and will instead showcase Lemma 2.2, which could in principle be applied in a directed setting. The Feder and Mihail result, in contrast, fails in directed setting, as rooted trees do not exhibit the same matroid basis structure of undirected trees. Our proof will exploit the undirectedness assumption, but we believe such monotonicity to be valid in greater generality, though this remains a delicate open problem.

The edge monotonicity of (2.9) should be extendable to reversible graphs, i.e. strongly connected graphs on which the RW is reversible, since reversible graphs can be identified with symmetric graphs with inhomogeneous killing, to which the proofs presented in Section 2.3.1 are still applicable. Moreover, (2.10) might very well hold for all weighted directed graphs. As will become clear in the proof of Theorem 2.3, the monotonicity of the (unoriented) edge process in (2.9) implies in particular the monotonicity of the connectivity function in (2.10). However, these two monotonicities are not equivalent. As the counterexamples depicted in Figure 2.1 in fact show, it is not difficult to find non-reversible graphs for which the monotonicity of the edges in (2.9) fails whereas the one in (2.10) does still hold true. Indeed, for the unweighted graph in Figure 2.1a it holds that

$$\mathbb{P}(\pm\{(1,2)\} \subseteq \Phi_q) = \mathbb{P}((1,2) \in \Phi_q) = \frac{2q^2 + q^3}{q + 5q^2 + 4q^3 + q^4},$$

which is increasing for $q < \sqrt{3} - 1$. This failure of monotonicity can be explained heuristically with the help of Lemma 2.2. For q very small the expected number of roots is close to one. Conditioning on edge $(1, 2)$ being present forces the random forest to have at least two roots, so then the conditioned expectation of the number of roots will be larger then the unconditional expectation. The above heuristic uses that the graph is not strongly connected. In strongly connected graphs, conditioning on unoriented edges being present cannot increase the minimal number of roots in a random forest. Still, Figure 2.1b shows a strongly connected weighted graph with

$$\mathbb{P}(\pm\{(1,3)\} \subseteq \Phi_q) = \frac{6q + 8q^2 + 2q^3}{18q + 21q^2 + 8q^3 + q^4},$$

which is increasing for $q < \sqrt{2} - 1$.

Unlike the subtle edge monotonicity in (2.9), we could not find examples where the one for the 2-point correlation (2.10) fails and in fact, we conjecture the latter to hold true in general but proving it remains an open problem. A first careful attempt to settle this conjecture in a non-reversible setting is offered in Theorem 2.5, where it is shown that (2.10) also holds at least on arbitrary weighted directed trees. In particular, on trees that are not strongly connected and hence non-reversible. It should be noted that even trees that are not strongly connected do satisfy the cycle condition for reversibility [62, p. 307], by virtue of having no non-trivial cycles. So, in some sense such trees are still close to be reversible.

2.2.2 Two-point correlation on trees

We start here to discuss results specific to trees, which are weighted directed graphs in which for any pair of vertices there exists a unique undirected path between them. Let us notice that in this setup, the analysis is facilitated by the absence of cycles. In general, the mapping from $\mathcal{F}$ to rooted partitions is not injective, while on trees this is the case.

In the constant weight case $w \equiv 1$, for a partition into $m \leq |V|$ blocks $\pi_m = \{B_1, B_2, \ldots, B_m\} \in \mathcal{P}(V)$, the measure in Equation (2.2) reads as

$$\mathbb{P}(\Pi_q = \pi_m) = \frac{q^m \prod_{i=1}^{m} |B_i|}{Z(q)},$$

from which we see that, for a given q, it concentrates on partitions where the block sizes tend to be of the same order. In this sense equipartitions are favored.

The first result in this tree specific setting extends the monotonicity of the 2-point correlation, as expressed in Theorem 2.3, to a specific weighted directed setting.

Theorem 2.5 (Monotonicity of 2-point correlations on trees). *If $G = (V, E, w)$ is a weighted directed tree, then for all $x, y \in V$ the function*

$$q \mapsto U_q(x, y)$$

is monotone non-decreasing.

Next we derive a representation of the 2-point correlation on arbitrary trees, in terms of reduced partition functions over subtrees.[2] To avoid confusion, in each statement in the sequel we will add proper indices to the partition and connectivity functions specifying the considered graph. The distance $d(x, y)$ between two vertices x and y will refer to the unweighted shortest path distance, i.e. the minimum number of edges on an undirected path between the two vertices.

Theorem 2.6 (Inclusion-exclusion for 2-point-correlation on trees). *Let $G = (V, E, w)$ be a weighted directed tree. Fix $x, y \in V$ with $d(x, y) = d$ and let $(z_i)_{i=0}^d$ be the unique undirected path with $z_0 = x$ and $z_d = y$. For a subset $I \subseteq [d]$ let G_I denote the graph obtained by removing all edges between z_{i-1} and z_i from G for all $i \in I$. Denote the $|I| + 1$ connected components of G_I by $G_I^1, \dots, G_I^{|I|+1}$. Then, for every $q > 0$, the following representation is valid*

$$U_q^{(G)}(x, y) = \frac{1}{Z_G(q)} \left(\sum_{k=1}^{d} (-1)^{k+1} \sum_{I \in \binom{[d]}{k}} \prod_{i=1}^{k+1} Z_{G_I^i}(q) \right). \tag{2.11}$$

Here $\binom{[d]}{k}$ denotes the collection of k-element subsets of $[d]$.

In particular for x, y such that $d(x, y) = 1$:

$$U_q(x, y) = \frac{Z_x(q) Z_y(q)}{Z_G(q)}, \tag{2.12}$$

where $Z_x(q)$ and $Z_y(q)$ denote the partition functions of the two connected components of the graph obtained by removing the edges between x and y.

2.2.3 Integer partitioning: analysis on lines and rings

In what follows we denote by $PG_n := \mathbb{Z} \cap [1, n]$ the (undirected and unweighted) path-graph constituted by the first n integers and by CG_n the cycle-graph on n vertices (i.e. the one dimensional discrete torus).

Theorem 2.7 (Partition function of path-graphs). *The partition function in (2.4) of PG_n can be expressed in the following ways:*

$$Z_{PG_n}(q) = \sum_{k=1}^{n} \binom{n+k-1}{2k-1} q^k \tag{2.13}$$

$$= \prod_{k=1}^{n} \left(q + 2 - 2\cos\left(\tfrac{\pi(n-k)}{n}\right)\right) \tag{2.14}$$

$$= \frac{q\left(q + 2 + \sqrt{q^2+4q}\right)^n - q\left(q + 2 - \sqrt{q^2+4q}\right)^n}{2^n \sqrt{q^2+4q}} \tag{2.15}$$

$$= q T_{n-1}(\tfrac{q}{2} + 1). \tag{2.16}$$

Here T_{n-1} denotes the $n-1$-th degree Chebyshev polynomial of the second kind.

[2] This type of reduction is due to the well-known spatial Markov property for UST-like measures (see Proposition A.2).

As can be appreciated in the proof, the above different representations reflect different computational methods suited for the random forest. We notice that for $q = 1$ evaluating this partition function corresponds to counting the number of rooted forests of the path-graph, as previously derived in [18].

One of the messages of this paper is that having an explicit characterization of a simple given geometry can be useful to derive information on some more involved geometry. The next corollary shows one such very simple instance by expressing the partition function on the torus in terms of partition functions of the simpler path-graph.

Corollary 2.7.1 (Partition function of cycle-graphs). *The partition function of CG_n is given by*

$$Z_{CG_n}(q) = Z_{PG_n}(q) + \tfrac{2}{q}\left[Z_{PG_n}(q) - Z_{PG_{n-1}}(q)\right] - 2 \tag{2.17}$$

$$= \sum_{k=1}^{n}\left(\binom{n+k}{2k} + \binom{n+k-1}{2k}\right) q^k. \tag{2.18}$$

The following result uses the notations $[n] = \{1, 2, \ldots, n\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Theorem 2.8 (Bounds for correlations on path-graph via random walk on $\mathbb{Z}$). *Let $x, y \in [n]$ be two vertices in PG_n at distance $d := y - x > 0$. Then, for any $q > 0$, the 2-point correlation between x and y is given by*

$$U_q^{(PG_n)}(x,y) = 1 - \frac{Z_{PG_{n-d}}(q)}{Z_{PG_n}(q)} - \frac{d\left[Z_{PG_x}(q) - Z_{PG_{x-1}}(q)\right]\left[Z_{PG_{n-y+1}}(q) - Z_{PG_{n-y}}(q)\right]}{qZ_{PG_n}(q)}. \tag{2.19}$$

Moreover, by denoting with $S = (S_m)_{m\in\mathbb{N}_0}$ the discrete-time simple random walk on $\mathbb{Z}$ starting at 0, the following bounds are satisfied

$$\left(1 - \left(\tfrac{2}{2+q}\right)^m\right)^2 \left(2\mathbb{P}(|S_m| < \tfrac{d}{2}) - 1\right)^2 \le U_q^{(PG_n)}(x,y) \le 1 - \mathbb{P}(|S_m| > d)\left(\tfrac{2}{2+q}\right)^m, \tag{2.20}$$

where the upper bound is valid for any $m \in \mathbb{N}$, while the lower bound holds for m such that $\mathbb{P}(|S_m| < \frac{d}{2}) \ge \frac{1}{2}$.

From the above statement, due to the diffusive behavior of the simple random walk S, it is clear that the correlation function between two points in a segment is non-degenerate when q_n scales with the inverse square distance between the two points. The next corollary makes this statement precise and shows that boundary effects emerge neatly from the asymptotic analysis.

Corollary 2.8.1 (Non-degenerate scaling and asymptotic boundary effects). *For each $n \in \mathbb{N}$ let x_n and y_n be vertices in PG_n. Let d_n denote the distance between these vertices and let $(q_n)_{n\in\mathbb{N}}$ be a monotone sequence of positive parameters. Then, if the limit $\lim_{n\to\infty} U_{q_n}^{(PG_n)}(x_n, y_n)$ exists, it holds that*

$$\lim_{n\to\infty} U_{q_n}^{(PG_n)}(x_n, y_n) \in (0,1) \text{ if and only if } q_n = \tfrac{c}{d_n^2} + o(\tfrac{1}{d_n^2}) \text{ for some constant } c > 0.$$

In particular, fix $\delta > 0$ and let $(\zeta_n)_{n\in\mathbb{N}}$ be a sequence such that $\zeta_n \in [\delta\sqrt{n}, n-\delta\sqrt{n}]$ for large enough n. Set $x_n = \zeta_n - \delta\sqrt{n} + o(\sqrt{n})$, $y_n = \zeta_n + \delta\sqrt{n} + o(\sqrt{n})$ and $q_n \sim \frac{1}{d_n^2}$, then the following two limits, distinguishing between the bulk and near the boundaries, are possible:

$$\lim_{n\to\infty} U_q^{(PG_n)}(x_n, y_n) = \begin{cases} 1 - \frac{3}{2e} & \text{if } \zeta_n = \omega(\sqrt{n}) \text{ and } \zeta_n = n - \omega(\sqrt{n}) \\ 1 - \frac{3}{2e} - \frac{1}{2}e^{-\frac{\alpha}{\delta}} & \text{if } \exists \alpha \geq \delta \text{ s.t. } \begin{cases} \zeta_n & = \alpha\sqrt{n} + o(\sqrt{n}) \\ & \text{or} \\ \zeta_n & = n - \alpha\sqrt{n} + o(\sqrt{n}) \end{cases} \end{cases} \tag{2.21}$$

In the above statement we computed the exact asymptotics only when the distance of the two vertices scales as the square root of n. Similar exact computations can be derived for other choices of the magnitude of this distance. We refer the interested reader to [52] for analogous statements in the cases when d_n stays of order one or diverges linearly. In particular, we note that *giants* (i.e. blocks of order $|V|$) appear at scale $q_n \sim d_n^{-2}$ and a unique giant emerges as soon as $q_n = o(n^{-2})$.

2.2.4 Detecting modular structures

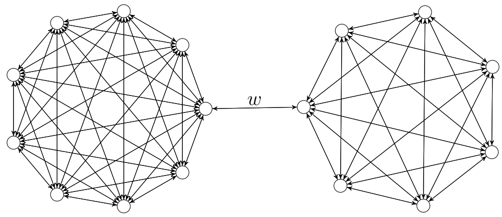
We collect here a number of simple statements of similar flavour aiming to illustrate that in tree-like graphs the emergence of giants and other modular structures can be detected with high probability by properly tuning q. Figures 2.2, 2.3 and 2.4 give a graphical overview of the main results in this section, which are given in Theorems 2.9, 2.11 and 2.13.

We start by showing with an illustrative example how the analysis on trees presented here and that on complete graphs, pursued in [7], can be combined to obtain results on mixed geometrical setups. The resulting regimes are summarized in the phase diagram in Figure 2.2b.

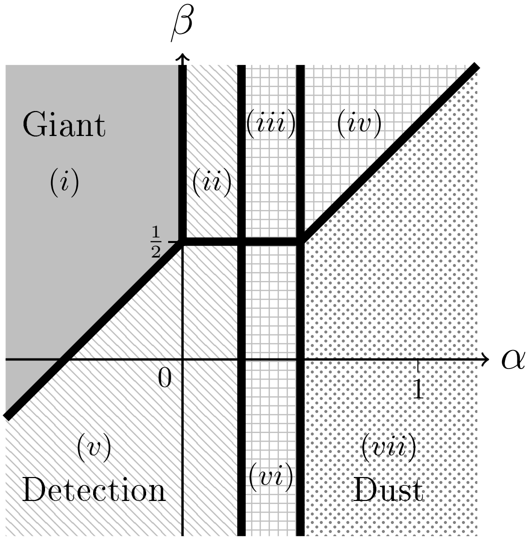
Theorem 2.9 (Detection of cliques in a bottleneck graph). *Let $BG_{n,m}$ be a bottleneck (two-cluster) graph. That is, an undirected graph consisting of two disjoint cliques C_1, C_2 on n and m vertices, respectively, that are connected via a single bridge edge, as depicted in Figure 2.2a. Equip $BG_{n,m}$ with a weight function that assigns weight w to the bridge and weight 1 to all other edges. Then its partition function is given by*

$$Z(q) = q\left(q(q+n)(q+m) + w(q+1)(2q+n+m)\right)(q+n)^{n-2}(q+m)^{m-2}. \tag{2.22}$$

Further, set $q = q_n > 0$ and let $w = w_n$ and $m = m_n$ depend on n where $n \geq m$. Denote by b, b' the two vertices incident to the bridge, by x, x' two vertices that both belong to the clique C_i containing b, and by y a vertex in the clique containing b'.



(a) A bottleneck graph with bridge weight w and two cliques of size $n = 9$ and $m = 7$.



(b) Phase diagram for the bottleneck graph, with $q = n^\alpha$, $w = n^\beta$ and $m = \sqrt{n}$. Regions (ii) and (v) are the regimes where the LEP detects the community structure. For each of the regions the following event occurs with high probability: (i) One single tree; (ii) Two trees on $n+1$ and $\sqrt{n}$ vertices, with the large tree containing both bridge vertices; (iii) One tree consists of the n vertices in the largest clique with the bridge vertex from the small clique, while the other vertices in the small clique are isolated; (iv) Both bridge vertices are connected, and all others are isolated; (v) Two trees with n and $\sqrt{n}$ vertices, while the bridge edge is absent; (vi) One tree with all n vertices in the largest clique, while the $\sqrt{n}$ vertices in the small clique are isolated; (vii) $n + \sqrt{n}$ isolated vertices.

Figure 2.2: Summary of the results in Theorem 2.9.

Then as $n \to \infty$ it holds for the 2-point correlation between these vertices that

$$U_q(x,x') \to \begin{cases} 0 & \text{if } q = o(\sqrt{|C_i|}) \\ 1 & \text{if } q = \omega(\sqrt{|C_i|}) \end{cases} \tag{2.23}$$

$$U_q(b,b') \to \begin{cases} 0 & \text{if } q = o(\frac{w}{m}) \text{ or } (q = o(w),\ w = \omega(m)) \\ 1 & \text{if } q = \omega(w) \text{ or } (q = \omega(\frac{w}{m}),\ w = o(m)) \end{cases} \tag{2.24}$$

$$U_q(b,x) \to \begin{cases} 0 & \text{if } \begin{cases} q = o(1) \text{ or } (q = o(\sqrt{|C_i|}),\ w = o(m)) \\ \text{or } (q = o(\sqrt{|C_i|}),\ m = o(n)), \end{cases} \\ \frac{c}{1+c} & \text{if } \begin{cases} q = \omega(1),\ q = o(\sqrt{|C_i|}),\ w = \omega(m), \\ |C_i| = n,\ m \sim cn \text{ with } c \in (0,1] \end{cases} \\ \frac{1}{1+c} & \text{if } \begin{cases} q = \omega(1),\ q = o(\sqrt{|C_i|}),\ w = \omega(m), \\ |C_i| = m,\ m \sim cn \text{ with } c \in (0,1] \end{cases} \\ 1 & \text{if } q = \omega(\sqrt{|C_i|}) \end{cases} \tag{2.25}$$

$$U_q(x,y) \to \begin{cases} 0 & \text{if } q = o(1),\ q = o(\frac{w}{m}) \\ 1 & \text{if } q = \omega(1) \text{ or } (q = o(1),\ q = \omega(\frac{w}{m})). \end{cases} \tag{2.26}$$

In Proposition 2.10 we make precise how q scales on a given large star graph with homogeneous edge weights.

Proposition 2.10 (Connectivity function on a homogeneous star graph). *Let SG_n denote the star graph on n vertices, i.e. SG_n is an undirected tree consisting of a single center vertex c that is adjacent to $n-1$ leaves. Let x, y be two distinct leaves and equip SG_n with a uniform weight function that assigns weight w to all edges. Given $q > 0$,*

$$U_q(c,x) = \frac{q(q+(n-1)w)}{(q+w)(q+nw)} \tag{2.27}$$

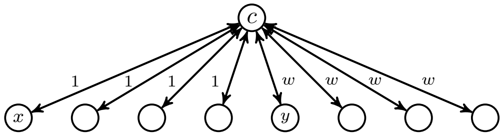
$$U_q(x,y) = \frac{q(q^2+(n+2)wq+2(n-1)w^2)}{(q+w)^2(q+nw)}, \tag{2.28}$$

which implies that $q \mapsto U_q(c,x)$ and $q \mapsto U_q(x,y)$ are strictly concave.

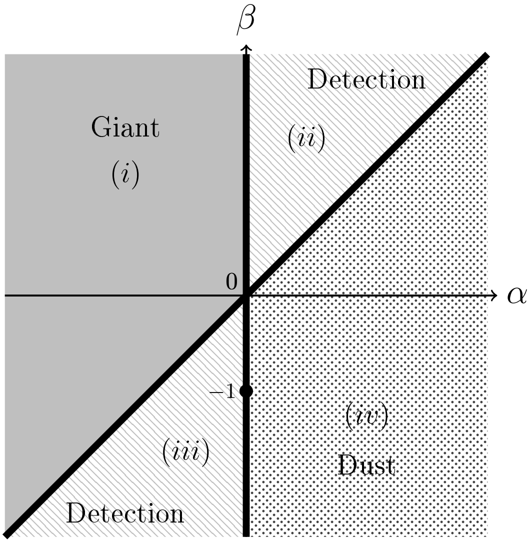
Let $q_n = \bar{q}n^\alpha$ and $w_n = \bar{w}n^\beta$ with $\alpha, \beta \in \mathbb{R}$ and $\bar{q}, \bar{w} \in (0,\infty)$. Then

$$\begin{aligned} \lim_{n\to\infty} U_{q_n}^{(SG_n)}(c,x) &= \begin{cases} 1 & \alpha > \beta \\ \frac{\bar{q}}{\bar{q}+\bar{w}} & \alpha = \beta \\ 0 & \alpha < \beta \end{cases}, \\ \lim_{n\to\infty} U_{q_n}^{(SG_n)}(x,y) &= \begin{cases} 1 & \alpha > \beta \\ \frac{\bar{q}(\bar{q}+2\bar{w}^2)}{(\bar{q}+\bar{w})^2} & \alpha = \beta \\ 0 & \alpha < \beta \end{cases} \end{aligned} \tag{2.29}$$

We see that the critical phase for the appearance of a giant is when $\alpha = \beta$. In this critical case the resulting connected subtree can be thought of as a star whose center has offspring distribution of parameter $\bar{q}/(\bar{q}+\bar{w})$.



(a) A community star graph with $n = 9$ vertices, with center vertex c and $k = 4$ vertices in the weight-1 community. The remaining four vertices belong to the weight-w community.



(b) Phase diagram for the community star graph, with $q = n^{\alpha}$ and $w = n^{\beta}$. For each of the regions the following event occurs with high probability: (i) One single tree; (ii) $k+1$ trees, all k vertices incident to a weight 1 edge are isolated, while the remaining vertices form a single tree; (iii) $n-k$ trees, all $n-k-1$ vertices incident to a weight w edge are isolated, while the remaining vertices form a single tree; (iv) n isolated vertices. The exact limit values of the correlations along the bold lines, i.e. in the non-degenerate regimes, can be found in Theorem 2.11.

Figure 2.3: Summary of the results in Theorem 2.11.

The following statement clarifies how q should be scaled in a non-homogeneous star, to detect an implanted sub-module of leaves that are more densely connected to the center. Figure 2.3b offers a graphical representation of Theorem 2.11.

Theorem 2.11 (Asymptotic detection in a star graph with two communities). *Let $CSG_{n,k}$ denote the community star graph on n vertices, which is a star graph on n vertices equipped with an inhomogeneous weight function, that assigns weight 1 to k edges and weight w to the remaining $n-k-1$ edges, as depicted in Figure 2.3a. Let c denote the center vertex, x, y vertices incident to an edge with weight 1 and w, respectively. For $\alpha, \beta \in \mathbb{R}$ take $q_n = n^\alpha$, $w_n = n^\beta$ and k constant. Then*

$$\lim_{n\to\infty} U_{q_n}^{(CSG_{n,k})}(c,y) = \begin{cases} 0 & \text{if } \alpha < \beta \\ \frac{1}{2} & \text{if } \alpha = \beta \\ 1 & \text{if } \alpha > \beta \end{cases},$$

$$\lim_{n\to\infty} U_{q_n}^{(CSG_{n,k})}(c,x) = \begin{cases} 0 & \text{if } \alpha < 0 \\ \frac{1}{2} & \text{if } \alpha = 0,\ \beta > -1 \\ \frac{k+3}{2k+8} & \text{if } \alpha = 0,\ \beta = -1 \\ \frac{k+1}{2k+4} & \text{if } \alpha = 0,\ \beta < -1 \\ 1 & \text{if } \alpha > 0 \end{cases} \tag{2.30}$$

The next two statements show similar detections on trees of different flavours.

Proposition 2.12 (Asymptotic correlation in undirected trees with bounded number of vertices). *Let $G = (V,E)$ be an undirected tree. For each $k \in \mathbb{N}$ let $w_k : E \to (0,\infty)$ be a symmetric edge weight function and $q_k > 0$ an intensity parameter. Write $G_k = (V, E, w_k)$ to denote the weighted graph obtained by equiping G with w_k. Assume that for each edge $e \in E$ the limit $\lim_{k\to\infty} \frac{w_k(e)}{q_k}$ exists in $[0,\infty]$. Fix two adjacent vertices $x, y \in V$. Then, as $k \to \infty$ it holds that*

$$U_{q_k}^{(G_k)}(x,y) \to \begin{cases} 0 & \text{if } q_k = o(w_k(x,y)) \\ 1 & \text{if } q_k = \omega(w_k(x,y)). \end{cases} \tag{2.31}$$

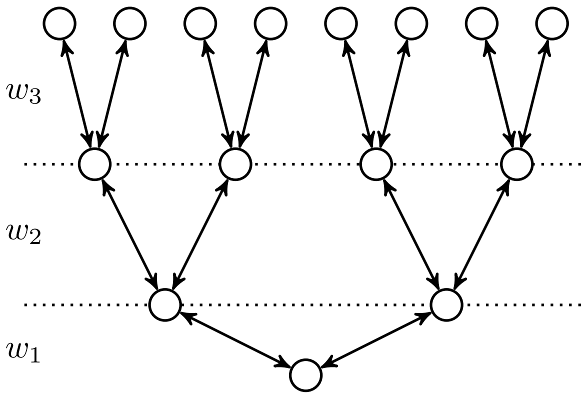
The following theorem holds for a specific class of undirected weighted trees that will be called 'hierarchical trees'. In these trees one vertex is specified as *ancestor* vertex. The *height* or *generation* of a vertex or edge is its distance to the ancestor. A *hierarchical tree* is a tree with edge weights $w : E \to [0,\infty)$ satisfying the following two properties:

(*i*) if $e, e' \in E$ are edges in the same generation of the tree, then $w(e) = w(e')$;

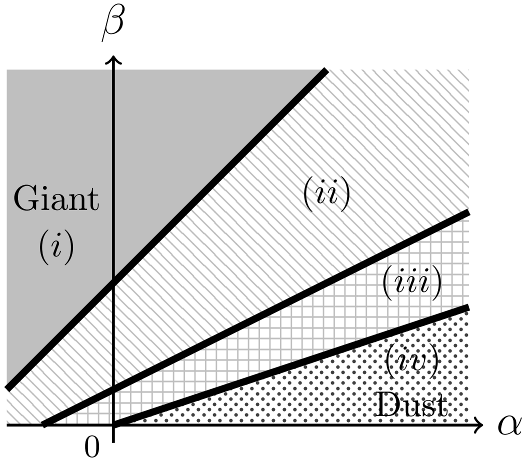
(*ii*) if $e_i, e_j \in E$ are edges in generations i and j with $i < j$, respectively, then $w(e_i) \leq w(e_j)$.

So, edges further from the ancestor of the hierarchical tree have more weight.

The *height of the tree* is the maximal height of its vertices. If x is a vertex at height h and y is a neighbor of x at height $k-1$, then we call x a *child* of y and y the *parent* of x. If each vertex with height less than the height of the tree has exactly d children, then we call the tree *complete d-ary*. The *ancestry* of a vertex is the unique path from the vertex to the ancestor (including the vertex itself). A depiction of a complete d-ary hierarchical tree is given in Figure 2.4a.



(a) A complete binary tree of height $h = 3$ with hierarchical edge weights. Each edge in generation i has weight w_i and these weights satisfy $w_1 \leq w_2 \leq w_3$.



(b) Phase diagram for the complete d-ary hierarchical tree of height $h = 3$, with $d = n$, $q = n^{\alpha}$ and j-th generation edge weights $w_j = n^{j\beta}$ for $\beta \geq 0$. For each of the regions the following event occurs with high probability: (i) One single tree; (ii) All 2nd and 3rd generation edges are present, while all 1st generation edges are absent; (iii) All 3rd generation edges are present, while all 1st and 2nd generation edges are absent; (iv) All $1 + n + n^2 + n^3$ vertices are isolated.

Figure 2.4: Summary of the results in Theorem 2.13.

Theorem 2.13 (Asymptotic detection of layers in a hierarchical weighted tree). *For each $n \in \mathbb{N}$ let $G_n = (V_n, E_n, w_n)$ be an undirected complete d_n-ary tree with hierarchical edge weights. For each $n \in \mathbb{N}$ let $x_n, y_n \in V_n$ be vertices such that x_n is the parent of y_n and such that the minimal distance between y_n and a leaf of G_n is constant in n. Denote this constant distance by k. Let e_n denote the edge between*

x_n and y_n. For each $n \in \mathbb{N}$ let $q_n > 0$ be the intensity parameter. Then as $n \to \infty$ it holds for the 2-point correlation between x_n and y_n that

$$U_{q_n}^{(G_n)}(x_n, y_n) \to \begin{cases} 0 & \text{if } q_n = o\left(d_n^{-k} w_n(e_n)\right) \\ 1 & \text{if } q_n = \omega\left(d_n^{-k} w_n(e_n)\right). \end{cases} \tag{2.32}$$

2.3 Proofs

2.3.1 Proofs of results on general graphs

2.3.1.1 Monotone events in terms of number of roots

Proof of Lemma 2.2. Let L be the negative graph Laplacian of G. Write $n = |V|$ and let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues $-L$. By [6, proposition 2.1] it holds that

$$\mathbb{E}[r_q] = \sum_{i=1}^{n} \frac{q}{q + \lambda_i} = \frac{qZ'(q)}{Z(q)}, \tag{2.33}$$

so that the derivative of the partition function is given by

$$Z'(q) = \tfrac{1}{q}\mathbb{E}[r_q]Z(q). \tag{2.34}$$

Note that the conditional probability $\mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k)$ does not depend on q. Also, the probability $\mathbb{P}(r_q = k)$ can be written as $\frac{c_k q^k}{Z(q)}$, where c_k is some constant independent of q, corresponding to the coefficent of degree k of the characteristic polynomial in (2.4). Hence, we have that

$$\begin{aligned}
\frac{d}{dq}\mathbb{P}(\Phi_q \in \mathcal{H}) &= \frac{d}{dq}\sum_{k=1}^{n} \mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k)\mathbb{P}(r_q = k) \\
&= \sum_{k=1}^{n} \mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k) c_k \frac{d}{dq}\frac{q^k}{Z(q)} \\
&= \sum_{k=1}^{n} \mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k) c_k \frac{kZ(q)q^{k-1} - q^k Z'(q)}{Z(q)^2} \\
&= \tfrac{1}{q}\sum_{k=1}^{n} \mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k) c_k \frac{kq^k - q^k \mathbb{E}[r_q]}{Z(q)} \\
&= \tfrac{1}{q}\sum_{k=1}^{n} \mathbb{P}(\Phi_q \in \mathcal{H} \mid r_q = k)\mathbb{P}(r_q = k)\left(k - \mathbb{E}[r_q]\right) \\
&= \tfrac{1}{q}\mathbb{P}(\Phi_q \in \mathcal{H})\sum_{k=1}^{n} \mathbb{P}(r_q = k \mid \Phi_q \in \mathcal{H})\left(k - \mathbb{E}[r_q]\right) \\
&= \tfrac{1}{q}\mathbb{P}(\Phi_q \in \mathcal{H})\left(\mathbb{E}[r_q \mid \Phi_q \in \mathcal{H}] - \mathbb{E}[r_q]\right),
\end{aligned}$$

where in the last step we use that $\sum_{k=1}^{n} \mathbb{P}(r_q = k \mid \Phi_q \in \mathcal{H}) = 1$. □

2.3.1.2 Monotonicities on undirected networks: proof of Theorem 2.3

Lemma 2.14. *Let $G = (V, E, w)$ be a weighted symmetric graph and let $B \subseteq E$ be a set of directed edges. Then for all $q > 0$ it holds that*

$$\mathbb{E}[r_q \mid \pm B \cap \Phi_q = \varnothing] \geq \mathbb{E}[r_q].$$

Proof. Let $H = G - B$ denote the subgraph of G obtained by removing all edges in B. Let $L^{(G)}$ and $L^{(H)}$ denote the negative graph Laplacians of G and H, respectively. Since these Laplacians are symmetric, $-L^{(G)}$ and $-L^{(H)}$ have real eigenvalues $\lambda_n \geq \ldots \geq \lambda_1$ and $\mu_n \geq \ldots \geq \mu_1$, respectively. By Weyl's monotonicity principle, these eigenvalues satisfy $\lambda_i \geq \mu_i$ for all $i \in [n]$. It follows that $\mathrm{Tr}((qI - L^{(G)})^{-1}) \leq \mathrm{Tr}((qI - L^{(H)})^{-1})$. By the spatial Markov property and [6, prop 2.1] it then holds that

$$\mathbb{E}^{(G)}[r_q \mid B \cap \Phi_q = \varnothing] = \mathbb{E}^{(H)}[r_q] = q\mathrm{Tr}((qI - L^{(H)})^{-1}) \geq q\mathrm{Tr}((qI - L^{(G)})^{-1}) = \mathbb{E}^{(G)}[r_q].$$

□

Lemma 2.15. *Let $G = (V, E, w)$ be a be a weighted symmetric graph and $A \subseteq E$. If $\mathbb{P}(\pm A \subseteq \Phi_q) > 0$, then for all $q > 0$ it holds that*

$$\mathbb{E}[r_q \mid \pm A \subseteq \Phi_q] \leq \mathbb{E}[r_q].$$

Proof. Let $e = (x, y) \in A$ be given. By Lemma A.2 and Lemma 2.14 it holds that

$$\begin{aligned}
&\mathbb{E}[r_q \mid \pm A \subseteq \Phi_q] \\
&= \frac{\mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q] - \mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q, \pm e \notin \Phi_q]\mathbb{P}(\pm e \notin \Phi_q \mid \pm(A-e) \subseteq \Phi_q)}{\mathbb{P}(\pm e \in \Phi_q \mid \pm(A-e) \subseteq \Phi_q)} \\
&= \frac{\mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q] - \mathbb{E}^{(G/(A-e))}[r_q \mid \pm e \notin \Phi_q]\mathbb{P}(\pm e \notin \Phi_q \mid \pm(A-e) \subseteq \Phi_q)}{\mathbb{P}(\pm e \in \Phi_q \mid \pm(A-e) \subseteq \Phi_q)} \\
&\leq \frac{\mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q] - \mathbb{E}^{(G/(A-e))}[r_q]\mathbb{P}(\pm e \notin \Phi_q \mid \pm(A-e) \subseteq \Phi_q)}{\mathbb{P}(\pm e \in \Phi_q \mid \pm(A-e) \subseteq \Phi_q)} \\
&= \frac{\mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q] - \mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q]\mathbb{P}(\pm e \notin \Phi_q \mid \pm(A-e) \subseteq \Phi_q)}{\mathbb{P}(\pm e \in \Phi_q \mid \pm(A-e) \subseteq \Phi_q)} \\
&= \mathbb{E}[r_q \mid \pm(A-e) \subseteq \Phi_q].
\end{aligned}$$

Hence, the result follows by induction on $|A|$. □

Proof of Theorem 2.3. The proof of (2.9) follows directly from Lemmas 2.2 and 2.15.

For the statement about the pairwise connectivity function in (2.10) we argue as follows. Fix $q > 0$. By Lemma 2.2 it is sufficient to show that $\mathbb{E}[r_q] \geq \mathbb{E}[r_q \mid B_q(x) = B_q(y)]$.

Let $\mathcal{P}$ denote the set of undirected paths from x to y, where we interpret a path as a set of directed edges. Then the event $\{B_q(x) = B_q(y)\}$ can be written as the disjoint union

$$\{B_q(x) = B_q(y)\} = \bigcup_{\pi \in \mathcal{P}} \{\pm\pi \subseteq \Phi_q\}.$$

It follows by Lemma 2.15 that

$$\mathbb{E}[r_q \mid B_q(x) = B_q(y)] = \sum_{\pi\in\mathcal{P}} \mathbb{E}[r_q \mid \pm\pi \subseteq \Phi_q]\mathbb{P}(\pm\pi \subseteq \Phi_q \mid B_q(x) = B_q(y))$$
$$\leq \sum_{\pi\in\mathcal{P}} \mathbb{E}[r_q]\mathbb{P}(\pm\pi \subseteq \Phi_q \mid B_q(x) = B_q(y)) = \mathbb{E}[r_q].$$

□

2.3.2 Two-point correlations on trees

2.3.2.1 Monotonicity of correlations on general trees: proof of Theorem 2.5

Below we show the monotonicity of the 2-point correlation restricted to arbitrary trees. We will start by expressing the 2-point correlation via hitting times in Lemma 2.16. Then in Lemma A.6 we show the monotocity of one point rooting events, by means of Lemma 2.2. After a last bound on the derivatives of hitting time events, given by Lemma A.7, we derive the main claim using these three lemmas.

Lemma 2.16 (Hitting time expression for 2-point correlation between adjacent vertices in trees). *Let $G = (V, E, w)$ be a weighted directed tree and $x, y \in V$ two adjacent vertices. Let $\mathbb{P}_v$ denote the law of the random walk X starting at vertex $v \in V$. The hitting time of vertex v by X is denoted by τ_v and τ_q is an independent exponential killing time with rate q. Then it holds that*

$$U_q^{(G)}(x, y) = \frac{1 - \mathbb{P}_x(\tau_y < \tau_q) - \mathbb{P}_y(\tau_x < \tau_q) + \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}{1 - \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}.$$

Proof of Lemma 2.16. We will reason using the representation in (2.8) coming from Wilson's sampling construction. We note in particular that in order for the directed edge (x, y) to be present in Φ_q, it is equivalent to require that the loop-erased trajectory in (2.8) includes y, which can be expressed in terms of hitting times of the random walk as

$$\mathbb{P}((x, y) \in \Phi_q) = \mathbb{P}_x(\tau_y < \tau_q) \sum_{k=0}^{\infty} (\mathbb{P}_y(\tau_x < \tau_q)\mathbb{P}_x(\tau_y < \tau_q))^k \, \mathbb{P}_y(\tau_q < \tau_x)$$
$$= \frac{\mathbb{P}_x(\tau_y < \tau_q)\,(1 - \mathbb{P}_y(\tau_x < \tau_q))}{1 - \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)},$$

where the index k in the above sum represents the number of times that the random walk reaches y and then does return to x. We notice in particular that the above step is equivalent to use the forest transfer-current kernel in [5]. For the reversed edge (y, x), we can write

$$\mathbb{P}((y, x) \in \Phi_q) = (1 - \mathbb{P}((x, y) \in \Phi_q))\, \mathbb{P}_y(\tau_x < \tau_q),$$

where these two factors correspond to (2.8). Therefore, it follows that

$$\begin{aligned} U_q^{(G)}(x,y) &= 1 - \mathbb{P}((x,y) \in \Phi_q) - \mathbb{P}((y,x) \in \Phi_q) \\ &= 1 - \tfrac{\mathbb{P}_x(\tau_y<\tau_q)(1-\mathbb{P}_y(\tau_x<\tau_q))}{1-\mathbb{P}_x(\tau_y<\tau_q)\mathbb{P}_y(\tau_x<\tau_q)} - \left(1 - \tfrac{\mathbb{P}_x(\tau_y<\tau_q)(1-\mathbb{P}_y(\tau_x<\tau_q))}{1-\mathbb{P}_x(\tau_y<\tau_q)\mathbb{P}_y(\tau_x<\tau_q)}\right)\mathbb{P}_y(\tau_x < \tau_q) \\ &= \frac{1 - \mathbb{P}_x(\tau_y < \tau_q) - \mathbb{P}_y(\tau_x < \tau_q) + \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}{1 - \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}. \end{aligned}$$

□

For brevity the proofs below use the notation $\{x \leftrightarrow y\} := \{B_q(x) = B_q(y)\}$ to denote the event that x and y are connected in Φ_q. We write $\{x \not\leftrightarrow y\}$ to denote the complementary event.

Proof of Theorem 2.5. Let $d = d(x,y)$ denote the distance between x and y in G and let z be the vertex adjacent to x with distance $d-1$ to y. We proceed by induction on d.

If $d = 1$, then by Lemma 2.16 we have that

$$U_q(x,y) = \frac{1 - \mathbb{P}_x(\tau_y < \tau_q) - \mathbb{P}_y(\tau_x < \tau_q) + \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}{1 - \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q)}.$$

Taking the derivative gives us that

$$\frac{d}{dq}U_q(x,y) = -\frac{(1-\mathbb{P}_x(\tau_y < \tau_q))^2 \frac{d}{dq}\mathbb{P}_y(\tau_x < \tau_q) + (1-\mathbb{P}_y(\tau_x < \tau_q))^2 \frac{d}{dq}\mathbb{P}_x(\tau_y < \tau_q)}{(1 + \mathbb{P}_x(\tau_y < \tau_q)\mathbb{P}_y(\tau_x < \tau_q))^2},$$

which is non-negative by the upper bound in Lemma A.5.

Now assume that $d \geq 2$. We then have that

$$\begin{aligned} &\frac{d}{dq}U_q(x,y) \\ &= \frac{d}{dq}\left(\mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y)\mathbb{P}(z \leftrightarrow y) + \mathbb{P}(z \not\leftrightarrow y)\right) \\ &= \mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y)\tfrac{d}{dq}\mathbb{P}(z \leftrightarrow y) + \mathbb{P}(z \leftrightarrow y)\tfrac{d}{dq}\mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y) + \tfrac{d}{dq}\mathbb{P}(z \not\leftrightarrow y) \\ &= (1 - \mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y))\tfrac{d}{dq}\mathbb{P}(z \not\leftrightarrow y) + \mathbb{P}(z \leftrightarrow y)\tfrac{d}{dq}\mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y). \end{aligned}$$

By the induction hypothesis, we have that $\frac{d}{dq}\mathbb{P}(z \not\leftrightarrow y) \geq 0$. Hence, it remains to show that

$$\tfrac{d}{dq}\mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y) \geq 0.$$

Removing the edges between x and z splits G into two connected components. Let T_x and T_z denote the components containing vertex x and z, respectively. By Lemma A.4 it then holds that

$$\mathbb{P}(x \not\leftrightarrow z \mid z \leftrightarrow y) = \frac{q}{q + w(x,z)\mathbb{P}^{(T_x)}(x \in R_q) + w(z,y)\mathbb{P}^{(T_z)}(z \in R_q \mid z \leftrightarrow y)}.$$

Taking the derivative and applying Lemmas A.6 and A.7 gives us that

$$\begin{aligned}\frac{d}{dq}\mathbb{P}(x \nleftrightarrow z \mid z \leftrightarrow y) &= \frac{w(x,z)\mathbb{P}^{(T_x)}(x \in R_q) + w(z,y)\mathbb{P}^{(T_z)}(z \in R_q \mid z \leftrightarrow y)}{\left(q + w(x,z)\mathbb{P}^{(T_x)}(x \in R_q) + w(z,y)\mathbb{P}^{(T_z)}(z \in R_q \mid z \leftrightarrow y)\right)^2} \\ &- \frac{qw(x,z)\frac{d}{dq}\mathbb{P}^{(T_x)}(x \in R_q) + qw(z,y)\frac{d}{dq}\mathbb{P}^{(T_z)}(z \in R_q \mid z \leftrightarrow y)}{\left(q + w(x,z)\mathbb{P}^{(T_x)}(x \in R_q) + w(z,y)\mathbb{P}^{(T_z)}(z \in R_q \mid z \leftrightarrow y)\right)^2} \\ &\geq 0.\end{aligned}$$

□

2.3.2.2 Inclusion-exclusion for connectivity function on general trees

Proof of Theorem 2.6. We will prove the statement by induction on d. First assume that $d = 1$. Write $\mathcal{H} = \{F \in \mathcal{F}_G \colon\ x \nleftrightarrow_F y\}$ to denote the set of rooted forests not containing an edge between x and y. Since G is a tree, removing the edges between x and y yields two connected components G_x and G_y containing vertex x and vertex y, respectively. Note that for the non-normalized measure on G it holds that $\nu^{(G)}(\Phi_q = F) = \nu_q^{(G_x)}(\Phi = F[G_x])\nu_q^{(G_y)}(\Phi = F[G_y])$ for all $F \in \mathcal{H}$, where $F[G_x]$ and $F[G_x]$ denote the induced subgraphs of F on the vertices of G_x and G_y, respectively. For all $F_x \in \mathcal{F}_{G_x}$ and $F_y \in \mathcal{F}_{G_y}$ there is exactly one $F \in \mathcal{H}$ with $F[G_x] = F_x$ and $F[G_y] = F_y$, namely the disjoint graph union of F_x and F_y. Hence, it holds that

$$\begin{aligned}U_q^{(G)}(x,y) &= \sum_{F \in \mathcal{H}} \mathbb{P}^{(G)}(\Phi_q = F) = \frac{1}{Z_G(q)} \sum_{F \in \mathcal{H}} \nu^{(G_x)}(\Phi_q = F[G_x])\nu^{(G_y)}(\Phi_q = F[G_y]) \\ &= \frac{1}{Z_G(q)} \sum_{F_x \in \mathcal{F}_{G_x}} \sum_{F_y \in \mathcal{F}_{G_y}} \nu^{(G_x)}(\Phi_q = F_x)\nu^{(G_y)}(\Phi_q = F_y) = \frac{Z_{G_x}(q)Z_{G_y}(q)}{Z_G(q)}.\end{aligned}$$

Now assume that $d > 1$. Let $z = z_{d-1}$ denote the neighbor of y with distance $d-1$ to x. Let $G_{\{d\}}$ denote the graph obtained from G by removing the edges between y and z. Then $G_{\{d\}}$ consists of two components G_y and G_z containing vertex y and z

respectively. It then holds by Lemma A.2 and the induction hypothesis that

$$\begin{aligned}
U_q^{(G)}(x,y) &= U_q^{(G)}(x,z) + U_q^{(G)}(y,z) - \mathbb{P}^{(G)}(x \nleftrightarrow_{\Phi_q} z,\ y \nleftrightarrow_{\Phi_q} z) \\
&= U_q^{(G)}(x,z) + U_q^{(G)}(y,z) - U_q^{(G)}(y,z)\mathbb{P}^{(G)}(x \nleftrightarrow_{\Phi_q} z \mid y \nleftrightarrow_{\Phi_q} z) \\
&= U_q^{(G)}(x,z) + U_q^{(G)}(y,z) - U_q^{(G)}(y,z)U_q^{(G_{\{d\}})}(x,z) \\
&= \frac{1}{Z_G(q)}\left(\sum_{k=1}^{d-1}(-1)^{k+1}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+1} Z_{G_I^i}(q)\right) + \frac{Z_{G_y}(q)Z_{G_z}(q)}{Z_G(q)} \\
&\quad - \frac{Z_{G_y}(q)Z_{G_z}(q)}{Z_G(q)}\frac{1}{Z_{G_{\{d\}}}(q)}\left(\sum_{k=1}^{d-1}(-1)^{k+1}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+2} Z_{G^i_{I\cup\{d\}}}(q)\right) \\
&= \frac{1}{Z_G(q)}\left(\sum_{k=1}^{d-1}(-1)^{k+1}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+1} Z_{G_I^i}(q)\right) + \frac{Z_{G_y}(q)Z_{G_z}(q)}{Z_G(q)} \\
&\quad + \frac{1}{Z_G(q)}\left(\sum_{k=1}^{d-1}(-1)^{k}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+2} Z_{G^i_{I\cup\{d\}}}(q)\right) \\
&= \frac{1}{Z_G(q)}\left(\sum_{k=1}^{d-1}(-1)^{k+1}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+1} Z_{G_I^i}(q)\right) \\
&\quad + \frac{1}{Z_G(q)}\left(\sum_{k=0}^{d-1}(-1)^{k}\sum_{I\in\binom{[d-1]}{k}}\prod_{i=1}^{k+2} Z_{G^i_{I\cup\{d\}}}(q)\right) \\
&= \frac{1}{Z_G(q)}\left(\sum_{k=1}^{d}(-1)^{k+1}\sum_{I\in\binom{[d]}{k}}\prod_{i=1}^{k+1} Z_{G_I^i}(q)\right).
\end{aligned}$$

□

2.3.3 Partition function on segments and rings

Proof of Theorem 2.7.

Equation (2.13) Let b be a boundary vertex of PG_n. Let $\nu^{(n)}$ and Z_n denote the non-normalized measure and partition function of PG_n, respectively. By Lemma A.4 we have that

$$\nu^{(n)}(b \notin R_q) = Z_{n-1}(q), \text{ and } \nu^{(n)}(b \in R_q) = \nu^{(n-1)}(b \in R_q) + qZ_{n-1}(q). \tag{2.35}$$

This gives us that

$$\begin{aligned}
Z_n(q) &= \nu^{(n)}(b \in R_q) + \nu^{(n)}(b \notin R_q) = \nu^{(n-1)}(b \in R_q) + (q+1)Z_{n-1}(q) \\
&= (q+2)Z_{n-1}(q) - \nu^{(n-1)}(b \notin R_q) = (q+2)Z_{n-1}(q) - Z_{n-2}(q).
\end{aligned} \tag{2.36}$$

We will prove Equation (2.13) by induction on n. Note that for $n = 1$ we have $Z_1(q) = q$ and for $n = 2$ we have $Z_2(q) = q^2 + 2q$, so in both these cases Equation (2.13) holds. Now assume that $n > 2$. Then by Equation (2.36), the induction hypothesis and repeated applications of Pascal's formula we have that

$$Z_n(q) = (2+q)Z_{n-1}(q) - Z_{n-2}(q) = (2+q)\sum_{k=1}^{n-1}\binom{n+k-2}{2k-1}q^k - \sum_{k=1}^{n-2}\binom{n+k-3}{2k-1}q^k$$
$$= \sum_{k=1}^{n}\binom{n+k-1}{2k-1}q^k.$$

Equation (2.14) Let L denote the negative graph Laplacian of PG_n, since due to (2.4) the partition function is the characteristic polynomial of L, it can be directly obtained from its spectrum, which is given in [64], from which:

$$Z_n(q) = \prod_{k=1}^{n}\left(q + 2 - 2\cos\left(\tfrac{\pi(n-k)}{n}\right)\right).$$

Equation (2.15) We have shown above that the partition function satisfies the recurrence relation in Equation (2.36) . Using the initial conditions $Z_1(q) = q$ and $Z_2(q) = q^2 + 2q$, this linear recurrence relation has solution

$$Z_n(q) = \frac{q\left(q+2+\sqrt{q^2+4q}\right)^n - q\left(q+2-\sqrt{q^2+4q}\right)^n}{2^n\sqrt{q^2+4q}}.$$

Equation (2.16) To verify that the three expressions above do indeed coincide, we can use Chebyshev polynomials of the second kind and find that

$$Z_n(q) = qT_{n-1}(\tfrac{q}{2}+1).$$

□

We next move to the proof of Corollary 2.7.1, for which we will first need to expresses in the next lemma the probability of a boundary point in the path-graph being a root in terms of differences of the partition function.

Lemma 2.17 (Rooting events in path-graphs). *Let PG_n be the path-graph on n vertices and $Z_n(q)$ its partition function. Let $x \in V$ be a vertex with distance $d \in \mathbb{N}_0$ from the boundary and $b \in V$ a boundary vertex. Let $\nu^{(n)}$ denote the non-normalized measure on PG_n and R_q the set of roots of Φ_q. Then*

$$\nu^{(n)}(x \in R_q) = \tfrac{1}{q}\nu^{(d+1)}(b \in R_q)\ \nu^{(n-d)}(b \in R_q), \tag{2.37}$$

with

$$\nu^{(n)}(b \in R_q) = Z_n(q) - Z_{n-1}(q). \tag{2.38}$$

For the non-normalized measure of the event that both boundary vertices b and b' are roots it holds that

$$\nu^{(n)}(b, b' \in R_q) = qZ_{n-1}(q). \tag{2.39}$$

Proof of Lemma 2.17.
Equation (2.37) Let L_n denote the negative graph Laplacian of the path-graph on n vertices. Inspection of the Laplacian and using the symmetry of the path-graph shows that

$$\det[qI - L_n]_x = \det[qI - L_{d+1}]_b \det[qI - L_{n-d}]_b,$$

as removing a row and column from $qI - L_n$ results in a matrix comprised of two blocks. Since the event that vertex x is a root equals the event that none of the outgoing edges of x are present, it holds by Lemma A.2 that $\nu^{(n)}(x \in R_q) = q\ \det[qI - L_n]_x$, from which Equation (2.37) follows.
Equation (2.38) Since $\nu^{(n)}(b \in R_q) = Z_n(q) - \nu^{(n)}(b \notin R_q)$, Equation (2.38) follows directly from Equation (2.35).
Equation (2.39) By Lemma A.4 we have that

$$\begin{aligned}\nu^{(n)}(b, b' \in R_q) &= q\nu^{(n-1)}(b \in R_q) + \nu^{(n-1)}(b, b' \in R_q),\\ \nu^{(n)}(b \in R_q,\ b' \notin R_q) &= \nu^{(n-1)}(b \in R_q).\end{aligned}$$

Since $\nu^{(n-1)}(b, b' \in R_q) = \nu^{(n-1)}(b \in R_q) - \nu^{(n-1)}(b \in R_q,\ b' \notin R_q)$, it follows from Equations (2.36) and (2.38) that

$$\begin{aligned}\nu^{(n)}(b, b' \in R_q) &= (q+1)\nu^{(n-1)}(b \in R_q) - \nu^{(n-2)}(b \in R_q)\\ &= (q+1)Z_{n-1}(q) - (q+2)Z_{n-2}(q) + Z_{n-3}(q) = qZ_{n-1}(q).\end{aligned}$$

□

Proof of Corollary 2.7.1. We will first prove Equation (2.17). Let V denote the vertex set of CG_n and let $x \in V$ be a vertex. The partition function can be split into two terms

$$Z_{CG_n}(q) = \nu^{(CG_n)}(x \in R_q) + \nu^{(CG_n)}(x \notin R_q). \tag{2.40}$$

Note that the induced subgraph $CG_n[V \setminus \{x\}]$ obtained by removing vertex x, is a path-graph on $n-1$ vertices. Let y and z denote the two vertices adjacent to x in CG_n. So, these are the boundary vertices of PG_{n-1}. We will use Lemma A.3. This gives us by Equation (2.36) and lemma 2.17 that

$$\begin{aligned}\nu^{(CG_n)}(x \in R_q) &= \sum_{F \in \mathcal{F}_{PG_{n-1}}} q\ \nu^{(PG_{n-1})}(\Phi = F)\ (1 + \tfrac{1}{q})^{|R(F)\cap\{y,z\}|}\\ &= (q + 2 + \tfrac{1}{q})\nu^{(PG_{n-1})}(y, z \in R_q) + 2(q+1)\nu^{(PG_{n-1})}(y \in R_q, z \notin R_q)\\ &\quad + q\nu^{(PG_{n-1})}(y, z \notin R_q)\\ &= qZ_{PG_{n-1}}(q) + (2 + \tfrac{1}{q})\nu^{(PG_{n-1})}(y, z \in R_q) + 2\nu^{(PG_{n-1})}(y \in R_q, z \notin R_q)\\ &= (q+2)Z_{PG_{n-1}}(q) - 2Z_{PG_{n-2}}(q) + \tfrac{1}{q}\nu^{(PG_{n-1})}(y, z \in R_q)\\ &= Z_{PG_n}(q) - Z_{PG_{n-2}}(q) + \tfrac{1}{q}\nu^{(PG_{n-1})}(y, z \in R_q) = Z_{PG_n}(q).\end{aligned}$$

Let $r_y(F)$ denote the root in the tree of forest F that contains vertex y. Again using Lemma A.3 and Equation (2.36), we obtain

$$\begin{aligned}\nu^{(CG_n)}(x \notin R_q) &= \sum_{F \in \mathcal{F}_{PG_{n-1}}} \nu^{(PG_{n-1})}(\Phi = F)\ 2(1+\tfrac{1}{q})^{\mathbf{1}\{z \in R(F),\ r_y(F) \neq z\}} \\ &= 2\nu^{(PG_{n-1})}(z \notin R_q) + 2(1+\tfrac{1}{q})\nu^{(PG_{n-1})}(z \in R_q, r_y(F) \neq z) \\ &= (2+\tfrac{2}{q})Z_{PG_{n-1}}(q) - \tfrac{2}{q}Z_{PG_{n-2}}(q) - 2 \\ &= \tfrac{2}{q}((q+2)Z_{PG_{n-1}}(q) - Z_{PG_{n-2}}(q) - Z_{PG_{n-1}}(q)) - 2 \\ &= \tfrac{2}{q}(Z_{PG_n}(q) - Z_{PG_{n-1}}(q)) - 2.\end{aligned}$$

This proves Equation (2.17).

Equation (2.18) follows from Equation (2.17) and the expression for the path-graph partition function given in Equation (2.13), by repeated applications of Pascal's formula. ☐

2.3.3.1 Asymptotic analysis of path-graphs

Proof of Theorem 2.8.

Equation (2.19) Let $\mathcal{F}_n$ denote the set of rooted forests of PG_n and write

$$\begin{aligned}\mathcal{F}^k_{n-d} &= \{F \in \mathcal{F}_{n-d} \colon\ r(F) = k\}; \\ \mathcal{R}^k_{n-d}(x) &= \{F \in \mathcal{F}_{n-d} \colon\ r(F) = k,\ x \in R(F)\}; \\ \mathcal{C}^k_n(x,y) &= \{F \in \mathcal{F}_n \colon\ r(F) = k,\ x \leftrightarrow_F y\}.\end{aligned}$$

It is sufficient to show that for all $k \in [n-d]$ it holds that $|\mathcal{C}^k_n(x,y)| = |\mathcal{F}^k_{n-d}| + d|\mathcal{R}^k_{n-d}(x)|$. The result then follows from Lemma 2.17.

We will construct a bijection between the set $\mathcal{C}^k_n(x,y)$ and the set $\mathcal{F}^k_{n-d} \cup (\mathcal{R}^k_{n-d}(x) \times [d])$. Let $F \in \mathcal{C}^k_n(x,y)$ be given. Let $r \in [n]$ denote the vertex of F that is the root in the component of x and y. Let $B = [y] \setminus [x-1]$ denote the set of all vertices from x to y and let $F_B \in \mathcal{F}^k_{n-d}$ denote the B-vertex contraction of F. Then we have that $F_B \in \mathcal{R}^k_{n-d}(x)$ if and only if $r \in [y] \setminus [x-1]$. Define the function $f : \mathcal{C}^k_n(x,y) \to \mathcal{F}^k_{n-d} \cup (\mathcal{R}^k_{n-d}(x) \times [d])$ by

$$f(F) = \begin{cases} F_B & \text{if } r \notin [y] \setminus [x] \\ (F_B, r - x) & \text{if } r \in [y] \setminus [x]. \end{cases}$$

It is easily verified that this gives a bijection.

Lower bound Let $\tilde{\mathbb{P}}_x$ denote the law of the discrete-time random walk $\tilde{X}$ on PG_n starting on x, as defined in Equation (A.6). Since in this case we consider a path-graph, we have that $\tilde{\tau}_q \sim \mathrm{Geom}(\frac{q}{q+2})$.

We will analyze the expression in Equation (2.8). Let z denote a vertex halfway between x and y. For notational simplicity we assume that d is even, so that $z = x + \frac{d}{2}$. The argument in the case where d is odd is similar. Note that the vertices x and y

are disconnected in Φ_q if both the random walks starting at x and the random walk starting at y are killed before reaching vertex z. So, we have that

$$\mathbb{P}(x \nleftrightarrow_{\Phi_q} y) \geq \tilde{\mathbb{P}}_x(\tilde{\tau}_q \leq \tau_z)\tilde{\mathbb{P}}_y(\tilde{\tau}_q \leq \tau_z). \tag{2.41}$$

Let $\tau_S(k)$ denote the hitting time of $k \in \mathbb{Z}$ by S. A coupling of $\tilde{X}$ and S can be used to show that

$$\tau_z \overset{d}{=} \min\{\tau_S\left(\tfrac{d}{2}\right), \tau_S\left(1 - 2x - \tfrac{d}{2}\right)\}, \tag{2.42}$$

where $\overset{d}{=}$ denotes equality in distribution.

By the reflection principle it holds for all $k, n \in \mathbb{N}$ that

$$\mathbb{P}(\tau_S(n) \leq k) = \mathbb{P}(S_k \notin [-n, n-1]).$$

For $u \in \{x, y\}$ it follows that

$$\begin{aligned}
&\tilde{\mathbb{P}}_u(\tilde{\tau}_q \leq \tau_z) = \sum_{k=1}^{\infty} \mathbb{P}(\tilde{\tau}_q = k)\tilde{\mathbb{P}}_u(\tau_z \geq k) \geq \sum_{k=1}^{m} \left(1 - \tilde{\mathbb{P}}_u(\tau_z < k)\right) \mathbb{P}(\tilde{\tau}_q = k) \\
&= \sum_{k=1}^{m} \left(1 - \mathbb{P}(\tau_S(\tfrac{d}{2}) < k \text{ or } \tau_S(1 - 2x - \tfrac{d}{2}) < k)\right) \mathbb{P}(\tilde{\tau}_q = k) \\
&\geq \sum_{k=1}^{m} \left(1 - 2\mathbb{P}(\tau_S(\tfrac{d}{2}) < k)\right) \mathbb{P}(\tilde{\tau}_q = k) = \sum_{k=1}^{m} \left(2\mathbb{P}(S_{k-1} \in [-\tfrac{d}{2}, \tfrac{d}{2} - 1]) - 1\right) \mathbb{P}(\tilde{\tau}_q = k) \\
&\geq \sum_{k=1}^{m} \left(2\mathbb{P}(|S_{k-1} < \tfrac{d}{2}) - 1\right) \mathbb{P}(\tilde{\tau}_q = k) \geq \sum_{k=1}^{m} \left(2\mathbb{P}(|S_m| < \tfrac{d}{2}) - 1\right) \mathbb{P}(\tilde{\tau}_q = k) \\
&= \left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right) \mathbb{P}(\tilde{\tau}_q \leq m) = \left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right) \left(1 - \left(1 - \tfrac{q}{2+q}\right)^m\right).
\end{aligned}$$

Hence,

$$\min_{u \in \{x,y\}} \tilde{\mathbb{P}}_u(\tilde{\tau}_q \leq \tau_z) \geq \left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right) \left(1 - \left(1 - \tfrac{q}{2+q}\right)^m\right).$$

If $\left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right)$ is non-negative, then we also have that

$$\tilde{\mathbb{P}}_x(\tilde{\tau}_q \leq \tau_z)\tilde{\mathbb{P}}_y(\tilde{\tau}_q \leq \tau_z) \geq \left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right)^2 \left(1 - \left(1 - \tfrac{q}{2+q}\right)^m\right)^2.$$

Therefore, we have for all $m \in \mathbb{N}$ with $\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) \geq \tfrac{1}{2}$ that

$$U_q^{(n)}(x, y) \geq \left(2\mathbb{P}\left(|S_m| < \tfrac{d}{2}\right) - 1\right)^2 \left(1 - \left(1 - \tfrac{q}{2+q}\right)^m\right)^2,$$

which gives the desired lower bound.

Upper bound We again analyze by means of Wilson's algorithm with the first random walk starting at x and the second one starting at y. Note that the trajectory of the loop-erasure of the first random walk will always contain its starting vertex x. Thus if the second random walk hits x before being killed, then x and y are connected in Φ_q. Therefore, we have that

$$\mathbb{P}(x \leftrightarrow_{\Phi_q} y) \geq \tilde{\mathbb{P}}_y(\tau_x < \tilde{\tau}_q).$$

Using a coupling argument we can show that

$$\tau_x \stackrel{d}{=} \min\{\tau_S(-d), \tau_S(2n + d - 2y + 1)\},$$

where τ_x denotes the first hitting time vertex x by the random walk $\tilde{X}$ starting at y. So, in a manner similar to that used for the lower bound, we find for all $m \in \mathbb{N}$ that

$$\begin{aligned}
\tilde{\mathbb{P}}_y(\tau_x < \tilde{\tau}_q) &= \sum_{k=1}^{\infty} \tilde{\mathbb{P}}_y(\tau_x < k)\mathbb{P}(\tilde{\tau}_q = k) \geq \sum_{k=m}^{\infty} \tilde{\mathbb{P}}_y(\tau_x \leq k)\mathbb{P}(\tilde{\tau}_q = k+1) \\
&= \sum_{k=m}^{\infty} \mathbb{P}(\tau_S(-d) \leq k \text{ or } \tau_S\,(2n + d - 2y + 1) \leq k)\mathbb{P}(\tilde{\tau}_q = k+1) \\
&\geq \sum_{k=m}^{\infty} \mathbb{P}(\tau_S(-d) \leq k)\mathbb{P}(\tilde{\tau}_q = k+1) \geq \sum_{k=m}^{\infty} \mathbb{P}(\tau_S(d) \leq m)\mathbb{P}(\tilde{\tau}_q = k+1) \\
&= \mathbb{P}(\tau_S(d) \leq m)\mathbb{P}(\tilde{\tau}_q > m) = \mathbb{P}(S_m \notin [-d, d-1])\mathbb{P}(\tilde{\tau}_q > m) \\
&\geq \mathbb{P}(|S_m| > d)\mathbb{P}(\tilde{\tau}_q > m) = \mathbb{P}\left(|S_m| > d\right)\left(1 - \tfrac{q}{2+q}\right)^m.
\end{aligned}$$

It follows that

$$U_q^{(n)}(x, y) = 1 - \mathbb{P}(x \leftrightarrow_{\Phi_q} y) \leq 1 - \mathbb{P}\left(|S_m| > d\right)\left(1 - \tfrac{q}{2+q}\right)^m.$$

□

Proof of Corollary 2.8.1.

$q_n = o(\frac{1}{d_n^2})$ Set $m_n = \left\lceil \frac{d_n}{\sqrt{q_n}} \right\rceil$, i.e. m_n is the smallest integer that is not smaller than $\frac{d_n}{\sqrt{q_n}}$. We have that $m_n = \omega(d_n^2)$. In particular this means that $m_n \to \infty$ as $n \to \infty$. So, $\frac{S_{m_n}}{\sqrt{m_n}}$ converges in distribution to a standard normal random variable. Since $\frac{d_n}{\sqrt{m_n}} \to 0$, it follows that $\mathbb{P}\left(\frac{|S_{m_n}|}{\sqrt{m_n}} > \frac{d_n}{\sqrt{m_n}}\right) \to 1$. We also have that $m_n = o(\frac{1}{q_n})$, which gives us that $\left(1 - \frac{q_n}{2+q_n}\right)^{m_n} \to 1$. Therefore, the upper bound from Theorem 2.8 gives us that

$$U_{q_n}^{(n)}(x_n, y_n) \leq 1 - \mathbb{P}\left(\tfrac{|S_{m_n}|}{\sqrt{m_n}} > \tfrac{d_n}{\sqrt{m_n}}\right)\left(1 - \frac{q_n}{2 + q_n}\right)^{m_n} = o(1).$$

$q_n = \omega(\frac{1}{d_n^2})$ Again set $m_n = \left\lceil \frac{d_n}{\sqrt{q_n}} \right\rceil$. It holds that $m_n = \omega(\frac{1}{q_n})$, and hence that $\left(1 - \frac{q_n}{2+q_n}\right)^{m_n} \to 0$. Furthermore, we have that $m_n = o(d_n^2)$. This means that $\frac{d_n}{2\sqrt{m_n}} \to \infty$ and thus that $\mathbb{P}\left(\frac{|S_{m_n}|}{\sqrt{m_n}} < \frac{d_n}{2\sqrt{m_n}}\right) \to 1$. For large enough n, this gives us that $\mathbb{P}\left(\frac{|S_{m_n}|}{\sqrt{m_n}} < \frac{d_n}{2\sqrt{m_n}}\right) \geq \frac{1}{2}$, which means that we can apply the lower bound from Theorem 2.8. This gives us that

$$U_{q_n}^{(n)}(x_n, y_n) \geq \left(1 - \left(1 - \frac{q_n}{2+q_n}\right)^{m_n}\right)^2 \left(2\mathbb{P}\left(\frac{|S_{m_n}|}{\sqrt{m_n}} < \frac{d_n}{2\sqrt{m_n}}\right) - 1\right)^2 = 1 - o(1).$$

$\boxed{q_n = \frac{c}{d_n^2} + o(\frac{1}{d_n^2})}$ Now set $m_n = \left\lceil \frac{d_n}{4\sqrt{cq_n}} \right\rceil$. We will distinguish between the case where d_n diverges and the case where d_n is bounded.

First assume that $d_n = \omega(1)$. Then we find that $m_n \sim \frac{1}{4q_n}$. It follows that there exists an $\varepsilon > 0$ small enough that $\varepsilon < \left(1 - \frac{q_n}{2+q_n}\right)^{m_n} < 1 - \varepsilon$ for all $n \in \mathbb{N}$. We also have that $m_n \sim \frac{1}{4}d_n^2$. This gives us that $\frac{S_{m_n}}{\sqrt{m_n}}$ converges in distribution to a standard normal random variable Z and that $\frac{d_n}{2\sqrt{m_n}} \to 1$. Since $0.6 < \mathbb{P}(|Z| < 1) < 0.7$, we can apply the lower bound from Theorem 2.8. Using both bounds from Theorem 2.8, we conclude the non-degeneracy:

$$\lim_{n\to\infty} U_{q_n}^{(n)}(x_n, y_n) \in (0, 1).$$

Now instead assume that d_n is bounded, i.e. there exists an $M \in \mathbb{N}$ with $M \geq d_n$ for all $n \in \mathbb{N}$. Then the lower bound from Theorem 2.8 can not necessarily be applied. However, we can lower bound the probability that x_n and y_n are disconnected by the probability that the discrete-time random walks on PG_n starting at x and y are both killed at time 1, while still at their starting points. This probability equals $\mathbb{P}(\tilde{\tau}_q = 1)^2 = \frac{q_n^2}{(2+q_n)^2}$.

The probability that x_n and y_n are connected can be lower bounded by the probability that the discrete-time random walk on PG_n starting at x jumps M times in the direction of y and then is then killed at time $M+1$. This probability equals $\left(\frac{1}{2+q_n}\right)^M \frac{q_n}{2+q_n}$. So, we have for all $n \in \mathbb{N}$ that

$$\frac{q_n^2}{(2+q_n)^2} \leq U_{q_n}^{(n)}(x_n, y_n) \leq 1 - \left(\tfrac{1}{2+q_n}\right)^M \tfrac{q_n}{2+q_n}.$$

Since $q_n \sim \frac{c}{d_n^2}$ and d_n is bounded, we have that q_n is bounded away from 0 and away from infinity. Hence, the 2-point correlation is also non-degenerate in this case. □

Proof of Equation (2.21). For brevity write $Z_n(q) = Z_{PG_n}(q)$ and set $a_n = \sqrt{q_n^2 + 4q_n}$. Using the expression for the partition function given in Equation (2.15) we have for each $m \in \mathbb{N}$ that

$$Z_m(q_n) = \frac{1}{\sqrt{1+4d_n^2}} \left(1 - \left(\frac{q_n + 2 - a_n}{q_n + 2 + a_n}\right)^m\right) \left(\frac{q_n + 2 + a_n}{2}\right)^m. \tag{2.43}$$

By Equation (2.19) the 2-point correlation is given by

$$U_{q_n}^{(G_n)}(x_n, y_n) = 1 - \frac{Z_{n-d_n}(q_n)}{Z_n(q_n)} - \frac{d_n\left[Z_{x_n}(q_n) - Z_{x_n-1}(q_n)\right]\left[Z_{n-y_n+1}(q_n) - Z_{n-y_n}(q_n)\right]}{q_n Z_n(q_n)}. \tag{2.44}$$

The result follows by plugging in Equation (2.43) into Equation (2.44) and repeatedly applying the following limits:

$$\lim_{n\to\infty}\left(\frac{2}{q_n+2+a_n}\right)^{\alpha\sqrt{n}+o(\sqrt{n})}=e^{-\frac{\alpha}{2\delta}},\qquad \lim_{n\to\infty}\left(\frac{q_n+2-a_n}{q_n+2+a_n}\right)^{\alpha\sqrt{n}+o(\sqrt{n})}=e^{-\frac{\alpha}{\delta}},$$

$$\text{and}\qquad \lim_{n\to\infty}\left(\frac{q_n+2-a_n}{q_n+2+a_n}\right)^{\omega(\sqrt{n})}=0,$$

with $\alpha\in\mathbb{R}$ being a constant. To conclude, let us check the correctness of these last three simple limits.

For the first one, since $\frac{\sqrt{1+4d_n^2}}{2d_n^2}-\frac{1}{2\delta\sqrt{n}}=o(\frac{1}{\sqrt{n}})$, we have that

$$\left(\frac{q_n+2+a_n}{2}\right)^{\hat{\alpha}_n}=\left(1+\frac{\sqrt{1+4d_n^2}}{2d_n^2}+\frac{1}{2d_n^2}\right)^{\hat{\alpha}_n}=\left(1+\frac{1}{2\delta\sqrt{n}}+o(\tfrac{1}{\sqrt{n}})\right)^{\hat{\alpha}_n}$$
$$=e^{\frac{\alpha}{2\delta}}+o(1).$$

where we shortened $\hat{\alpha}_n:=\alpha\sqrt{n}+o(\sqrt{n})$.

For the second limit, note that $a_n=\sqrt{q_n^2+4q_n}=\frac{1}{\delta\sqrt{n}}(1+o(1))$. Hence,

$$\left(\frac{q_n+2-a_n}{q_n+2+a_n}\right)^{\hat{\alpha}_n}=\left(1-\frac{a_n}{1+o(1)}\right)^{\hat{\alpha}_n}=\left(1-\frac{1}{\delta\sqrt{n}}(1+o(1))\right)^{\hat{\alpha}_n}=e^{-\frac{\alpha}{\delta}}+o(1).$$

Similarly for the last limit, we have that

$$\left(\frac{q_n+2-a_n}{q_n+2+a_n}\right)^{\omega(\sqrt{n})}=\left(1-\frac{1}{\delta\sqrt{n}}(1+o(1))\right)^{\omega(\sqrt{n})}\to 0,$$

which concludes the claims.

□

2.3.4 Asymptotic detection of modular structures

2.3.4.1 A two communities bottleneck graph

Proof of Theorem 2.9. Equation (2.22) Let $\nu^{(G)}$ denote the non-normalized measure on a graph G. Let K_n be the complete graph on n vertices. We can express the partition function of $BG_{n,m}$ in terms of the partition functions and the non-normalized measure of rooting events in the complete graphs K_n and K_m.

Let L_n denote the negative graph Laplacian of K_n. The partition function of K_n is given by

$$Z_{K_n}(q)=q(q+n)^{n-1}. \tag{2.45}$$

Let U be a set of vertices of K_n with $|U|=r$ and write $[qI-L_n]_U$ to denote the submatrix of $qI-L_n$ otained by removing all rows and columns corresponding to vertices in U. Then non-normalized measure of the event that at least all vertices in U are roots in a random rooted forest of K_n is given by

$$\begin{aligned}\nu^{(K_n)}(U \subseteq R_q) &= q^r \det[qI - L_n]_U = q^r \det[(q+r)I - L_{n-r}] \\ &= q^r Z_{K_{n-r}}(q+r) = q^r(q+r)(q+n)^{n-r-1}.\end{aligned} \tag{2.46}$$

For the partition function of $BG_{n,m}$, Lemma A.4 gives us that

$$\begin{aligned}Z_{BG_{n,m}}(q) &= Z_{K_n}(q)Z_{K_m}(q) + \tfrac{w}{q} Z_{K_n}(q)\nu^{(K_m)}(b' \in R_q) + \tfrac{w}{q} Z_{K_m}(q)\nu^{(K_n)}(b \in R_q) \\ &= q\left(q(q+n)(q+m) + w(q+1)(2q+n+m)\right)(q+n)^{n-2}(q+m)^{m-2}.\end{aligned}$$

Equation (2.24)

We can express $U_q(b, b')$ explicitly by using Theorem 2.6 and eqs. (2.22) and (2.45)

$$U_q(b, b') = \frac{Z_{K_n}(q)Z_{K_m}(q)}{Z_{BG_{n,m}}(q)} \tag{2.47}$$

$$= \frac{q(q+n)(q+m)}{q(q+n)(q+m) + w(q+1)(2q+n+m)}. \tag{2.48}$$

The result of Equation (2.24) follows directly from this expression.

Equation (2.23) We will assume that x and x' both belong to the clique of size n, as the other case can be proven similarly. By Lemma A.4 we have that

$$\begin{aligned}\nu^{(BG_{n,m})}(x \leftrightarrow_q x') &= \nu^{(K_n)}(x \leftrightarrow_q x')Z_{K_m}(q) + \tfrac{w}{q}\nu^{(K_n)}(x \leftrightarrow_q x',\ b \in R_q)Z_{K_m}(q) \\ &\quad + \tfrac{w}{q}\nu^{(K_n)}(x \leftrightarrow_q x')\nu^{(K_m)}(b' \in R_q).\end{aligned}$$

By Equations (2.45) and (2.46) it follows that

$$\begin{aligned}U_q^{(BG_{n,m})}(x, x') &= 1 - \frac{\nu^{(K_n)}(x \leftrightarrow_q x')Z_{K_m}(q) + \frac{w}{q}\nu^{(K_n)}(x \leftrightarrow_q x',\ b \in R_q)Z_{K_m}(q)}{\frac{Z_{K_n}(q)Z_{K_m}(q)}{U_q^{(BG_{n,m})}(b,b')}} \\ &\quad + \frac{\frac{w(q+1)}{q(q+m)}\nu^{(K_n)}(x \leftrightarrow_q x')Z_{K_m}(q)}{\frac{Z_{K_n}(q)Z_{K_m}(q)}{U_q^{(BG_{n,m})}(b,b')}} \\ &= 1 - U_q^{(BG_{n,m})}(b, b')\left(\left(1 + \frac{w(q+1)}{q(q+m)}\right)\mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} x')\frac{w(q+1)}{q(q+n)}\mathbb{P}^{(K_n)}(x \leftrightarrow_q x' \mid b \in R_q)\right).\end{aligned} \tag{2.49}$$

Let H denote the graph obtained by removing all outgoing edges of b from K_n, while retaining the ingoing edges. By Lemma A.2 it then holds that $\mathbb{P}^{(K_n)}(x \leftrightarrow_q x' \mid b \in R_q) = \mathbb{P}^{(H)}(x \leftrightarrow_q x')$. Let $\mathbb{P}_x$ denote the law of the random walk on H starting at x and τ_q an independent exponential killing time with rate q. Since the hitting time τ_b has an exponential distribution with rate 1, we can identify the random walk on H killed at rate q with a random walk on K_{n-1} killed at rate $q+1$, by killing the random walk when it hits b. By analyzing Wilson's algorithm on H with the first two random walks starting at x and x', this gives us that

$$\begin{aligned}
&\mathbb{P}^{(K_n)}(x \leftrightarrow_q x' \mid b \in R_q) = \mathbb{P}^{(H)}(x \leftrightarrow_q x') \\
&= \mathbb{P}^{(K_{n-1})}(x \leftrightarrow_{q+1} x') + \mathbb{P}^{(K_{n-1})}(x \not\leftrightarrow_{q+1} x')\, \mathbb{P}_x(\tau_b < \tau_q)\, \mathbb{P}_{x'}(\tau_b < \tau_q) \\
&= \mathbb{P}^{(K_{n-1})}(x \leftrightarrow_{q+1} x') + \tfrac{1}{(q+1)^2}\mathbb{P}^{(K_{n-1})}(x \not\leftrightarrow_{q+1} x') \\
&= \tfrac{1}{(q+1)^2} + \tfrac{q(q+2)}{(q+1)^2}\mathbb{P}^{(K_{n-1})}(x \leftrightarrow_{q+1} x').
\end{aligned} \tag{2.50}$$

By [7, Theorem 1] we have that

$$\mathbb{P}^{(K_n)}(x \leftrightarrow_q x') \to \begin{cases} 1 \text{ if } q = o(\sqrt{n}) \\ 0 \text{ if } q = \omega(\sqrt{n}), \end{cases} \tag{2.51}$$

which together with Equation (2.50) gives us that

$$\mathbb{P}^{(K_n)}(x \leftrightarrow_q x' \mid b \in R_q) \to \begin{cases} 1 \text{ if } q = o(\sqrt{n}) \\ 0 \text{ if } q = \omega(\sqrt{n}). \end{cases}$$

Assume that $q = o(\sqrt{n})$. Fix a small enough $\varepsilon > 0$. Then for n large enough it holds that $\mathbb{P}^{(K_n)}(x \leftrightarrow x') > 1 - \varepsilon$ and that $\mathbb{P}^{(K_n)}(x \leftrightarrow x' \mid b \in R_q) > 1 - \varepsilon$. By Equations (2.48) and (2.49), this means that for n large enough

$$U_q^{(BG_{n,m})}(x, x') < 1 - U_q^{(BG_{n,m})}(b, b')\left(\left(1 + \tfrac{w(q+1)}{q(q+m)}\right)(1-\varepsilon) + \tfrac{w(q+1)}{q(q+n)}(1-\varepsilon)\right) = \varepsilon. \tag{2.52}$$

If instead $q = \omega(\sqrt{n})$, then analogously we find for large enough n that

$$U_q^{(BG_{n,m})}(x, x') > 1 - \varepsilon.$$

Equation (2.25) Assume that x, x' and b belong to the clique of size n. By again considering the random walk on H, we find that

$$\mathbb{P}^{(K_n)}(x \leftrightarrow b \mid b \in R_q) = \mathbb{P}_x(\tau_b < \tau_q) = \frac{1}{q+1}.$$

So, since $\frac{1}{q+1} \to 0$ for $q = \omega(\sqrt{n})$, the case $q = \omega(\sqrt{n})$ follows analogous to Equation (2.52).

Now assume that $q = o(\sqrt{n})$. Then we have that $\mathbb{P}^{(K_n)}(x \leftrightarrow_q b) \to 1$, so that

$$\begin{aligned}
U_q^{(BG_{n,m})}(x, b) &= 1 - U_q^{(BG_{n,m})}(b, b')\left(\left(1 + \tfrac{w(q+1)}{q(q+m)}\right)\mathbb{P}^{(K_n)}(x \leftrightarrow_q b) + \tfrac{w(q+1)}{q(q+n)}\ \tfrac{1}{q+1}\right) \\
&\sim 1 - U_q^{(BG_{n,m})}(b, b')\left(\left(1 + \tfrac{w(q+1)}{q(q+m)}\right) + \tfrac{w(q+1)}{q(q+n)}\ \tfrac{1}{q+1}\right) \\
&= \frac{wq(q+m)}{q(q+n)(q+m) + w(q+1)(2q+n+m)}.
\end{aligned}$$

This asymptotic expression for $U_q^{(BG_{n,m})}(x, b)$ gives us that

$$U_q^{(BG_{n,m})}(x, b) \to \begin{cases} 0 & \text{if } \begin{cases} q = o(1) \text{ or } (q = o(\sqrt{|C_i|}),\ w = o(m)) \\ \text{ or } (q = o(\sqrt{|C_i|}),\ m = o(n)), \end{cases} \\ \frac{c}{1+c} & \text{if } q = \omega(1),\ q = o(\sqrt{n}),\ w = \omega(m),\ m \sim cn \text{ with } c \in (0, 1] \\ 1 & \text{if } q = \omega(\sqrt{n}) \end{cases}$$

Performing the same computation for $U_q(y,b')$ yields the result of Equation (2.25).

Equation (2.26) By Lemma A.1 and eqs. (2.47) and (2.50) it holds that

$$\begin{aligned}
&U_q^{(BG_{n,m})}(x,y)\\
&= 1 - \frac{\nu^{(BG_{n,m})}(x \leftrightarrow_q y,\ (b,b') \in \Phi_q)}{Z_{BG_{n,m}}(q)} - \frac{\nu^{(BG_{n,m})}(x \leftrightarrow_q y,\ (b',b) \in \Phi_q)}{Z_{BG_{n,m}}(q)}\\
&= 1 - \frac{\frac{w}{q}\nu^{(K_n)}(x \leftrightarrow_q b,\ b \in R_q)\ \nu^{(K_m)}(b' \leftrightarrow_q y)}{Z_{BG_{n,m}}(q)}\\
&\quad - \frac{\frac{w}{q}\nu^{(K_n)}(x \leftrightarrow_q b)\ \nu^{(K_m)}(b' \leftrightarrow_q y,\ b' \in R_q)}{Z_{BG_{n,m}}(q)}\\
&= 1 - U_q^{(BG_{n,m})}(b,b')\Big(\tfrac{w}{q}\mathbb{P}^{(K_n)}(x \leftrightarrow_q b,\ b \in R_q)\ \mathbb{P}^{(K_m)}(b' \leftrightarrow_q y)\\
&\quad + \tfrac{w}{q}\mathbb{P}^{(K_n)}(x \leftrightarrow_q b)\ \mathbb{P}^{(K_m)}(b' \leftrightarrow_q y,\ b' \in R_q)\Big)\\
&= 1 - U_q^{(BG_{n,m})}(b,b')\left(\frac{w}{q(q+n)}\mathbb{P}^{(K_m)}(b' \leftrightarrow_q y) + \frac{w}{q(q+m)}\mathbb{P}^{(K_n)}(x \leftrightarrow_q b)\right)\\
&= 1 - \frac{w(q+m)\mathbb{P}^{(K_m)}(b' \leftrightarrow_q y) + w(q+n)\mathbb{P}^{(K_n)}(x \leftrightarrow_q b)}{q(q+n)(q+m) + w(q+1)(2q+n+m)},
\end{aligned}$$

from which the limits in Equation (2.26) follow. □

2.3.4.2 Star graphs: homogeneous case and with implanted communities

Proof of Proposition 2.10. Let us start by providing an expression for the partition function of a complete k-ary tree with homogeneous weights. Let $w \in (0,\infty)$, $h \in \mathbb{N}$, $k \in \mathbb{N}$, and let L be the negative graph Laplacian of the complete k-ary tree with height h and uniform weight w. Define $(\alpha_n)_{n\in\mathbb{N}_0}$ such that $\alpha_0 = q + w$ and $\alpha_{n+1} = q + (k+1)w - \frac{kw^2}{\alpha_n}$ for $n \in \mathbb{N}$. Then the characteristic polynomial of L is given by

$$\det[qI - L] = \left(\prod_{i=0}^{h-1} \alpha_i^{k^{h-i}}\right)\left(q + kw - \frac{kw^2}{\alpha_{h-1}}\right) \tag{2.53}$$

In fact, observe that in the matrix $[qI - L]$ there is a $k^h \times k^h$ diagonal matrix with entries $q + w$ since the leaves are not connected with each other. Call this right lower diagonal matrix D and call the corresponding left upper matrix A, right upper matrix B and left lower matrix C. By Schur's determinant identity, we get $\det[qI - L] = \det[D]\det[A - BD^{-1}C]$. Here, $\det[D] = (q+w)^{k^h}$ since $D = (q+w)I$. This also gives us $D^{-1} = \frac{1}{q+w}I$. Thus, $BD^{-1}C = \frac{1}{q+w}BC$ is a diagonal matrix with lower entries $\frac{kw^2}{q+w}$, on the places of the parents of the leaves, and upper entries 0, on the places of the nodes that are not parents of the leaves (if there are any). If $h = 1$ we see that $A - BD^{-1}C = q + kw - \frac{kw^2}{q+w}$ and we are done. If $h > 1$ we see that $A - BD^{-1}C$ is again a matrix with a right lower diagonal matrix. This time, the entries of the diagonal matrix are $q + (k+1)w - \frac{kw^2}{q+w}$. By iteration of Schur's determinant identity we get the formula in (2.53).

We'll continue by checking the validity of the expressions in (2.27) and (2.28). By applying (2.53) to the homogeneous star graph SG_n we obtain that its partition function is given by

$$Z_{SG_n}(q) = q(q+w)^{n-2}(q+nw) \tag{2.54}$$

Since $d(c,x) = 1$, by (2.12) we have that

$$U_q(c,x) = \frac{q\ Z_{SG_{n-1}}(q)}{Z_{SG_n}(q)} = \frac{q(q+(n-1)w)}{(q+w)(q+nw)}.$$

Similarly, since $d(x,y) = 2$, by Theorem 2.6 we have that

$$U_q(x,y) = \frac{2q\ Z_{SG_{n-1}}(q)\ -\ q^2\ Z_{SG_{n-2}}(q)}{Z_{SG_n}(q)} = \frac{q(q^2+(n+2)wq+2(n-1)w^2)}{(q+w)^2(q+nw)},$$

which finishes the proof of (2.27) and (2.28). The asymptotics in (2.29) follow immediately. □

Proof of Theorem 2.11. The generator matrix L of the community star graph $CSG_{n,k}$ is given by

$$L = \begin{pmatrix} -k-(n-k-1)w & 1 & 1 & \cdots & w & w \\ 1 & -1 & & & & \\ 1 & & -1 & & & \\ \vdots & & & \ddots & & \\ w & & & & -w & \\ w & & & & & -w \end{pmatrix}$$

where the empty places are to be filled with zeros. The characteristic polynomial of this matrix is

$$\det[qI - L] = q(q+w)^{n-k-2}(q+1)^{k-1}[q^2 + ((n-k)w+k+1)q + nq] \tag{2.55}$$

which can be found by applying Schur's determinant identity as we did in the proof of Proposition 2.10. Hence, the eigenvalues of L are:

$$\lambda_i = \begin{cases} 0 & \text{if } \ i = 1 \\ -w & \text{if } \ i = 2,\dots,n-k-1 \\ -1 & \text{if } \ i = n-k,\dots,n-2 \\ -\frac{1}{2}\mu + \frac{1}{2}\delta & \text{if } \ i = n-1 \\ -\frac{1}{2}\mu - \frac{1}{2}\delta & \text{if } \ i = n \end{cases}$$

where

$$\mu = (n-k)w + k + 1$$
$$\delta = \sqrt{((n-1)^2 - 2nk + k^2 + 2n - 1)w^2 + k^2 + 2((n-1)k - k^2 - n)w + 2k + 1}.$$

Denote the sets of vertices that are connected to the center vertex c with a weight 1 and w by V_1 and V_w, respectivley. Combining Theorem 2.6 with Equation (2.55) and writing $c = c(n,k,w,q) = ((n-k)w+k+1)q + nw$ leads to:

$$U_q(c,x) = \begin{cases} \frac{q(q^2-q-w+c)}{(q+1)(q^2+c)} & x \in V_1 \\ \\ \frac{q(q^2-wq-w+c)}{(q+w)(q^2+c)} & x \in V_w \end{cases}$$

and

$$U_q(x,y) = \begin{cases} \frac{q(q^3+2q^2+(c-2)q+2c-2w)}{(q+1)^2(q^2+c)} & x,y \in V_1 \\ \frac{q(q^3+(w+1)q^2+c(w+q+1)-w(w+1))}{(q+1)(q+w)(q^2+c)} & x \in V_1, y \in V_w \\ \frac{q(q^3+2wq^2+(c-2w-4wk)q+2w(c-w))}{(q+w)^2(q^2+c)} & x,y \in V_w \end{cases}$$

From these explicit formulas, letting q and w be as in the statement, the limits in Theorem 2.11 follow. □

2.3.4.3 Playing with degrees and hierarchical weights on trees

Proof of Proposition 2.12. Note that $\mathbb{P}(e \in \Phi_q) \leq \frac{w(e)}{q+w(e)}$, since by Lemma A.1 it holds that

$$1 \geq \mathbb{P}(e \in \Phi_q) + \mathbb{P}(x \in R_q,\ x \not\leftrightarrow_q y) = \mathbb{P}(e \in \Phi_q)(1 + \tfrac{q}{w(e)}).$$

Hence, if $q_k = \omega(w_k(x,y))$, then we have that $U_{q_k}(x,y) \to 1$.

Assume that $q_k = o(w_k(x,y))$. Let $\tilde{\mathbb{P}}_x^{(k)}$ denote the law of the discrete-time random walk $\tilde{X}$ on G_k starting at vertex $x \in V_k$ and let $\tilde{\tau}_q$ be a geometric killing time, as defined in Equation (A.6). Let τ_x denote the first hitting time of x by $\tilde{X}$. Let m denote the number of vertices on the x-side of edge (x,y) in G. We will show by induction on m that

$$\tilde{\mathbb{P}}_x^{(k)}(\tau_y < \tilde{\tau}_q) = 1 - \Theta\left(\tfrac{q_k}{w_k(x,y)}\right).$$

If $m = 1$, then x is a leaf in G. It follows that

$$\tilde{\mathbb{P}}_x^{(k)}(\tau_y < \tilde{\tau}_q) = \frac{w_k(x,y)}{q_k + w_k(x,y)} \sim 1 - \tfrac{q_k}{w_k(x,y)}.$$

Assume that $m \geq 2$. Let N_x denote the set of neighbors of x in G. Since the limit $\lim_{k\to\infty} \frac{w_k(e)}{q_k}$ exists for all edges incident to x, we can partition $N_x \setminus \{y\}$ into two parts: the first part $N_x^{\leq} = \{v \in N_x \setminus \{y\}\colon\ w_k(x,v) = \mathcal{O}(q_k)\}$ consists of all neighbors for which the weight of the edge between x and that neighbor has no larger order than q_k; the second part $N_x^{>} = \{v \in N_x \setminus \{y\}\colon\ w_k(x,v) = \omega(q_k)\}$ consists of the remaining neighbors. Then for each $v \in N_x^{>}$ we have that $q_k = o(w_k(x,v))$. For each such v it follows by the induction hypothesis that $\mathbb{P}_v^{(k)}(\tau_x < \tilde{\tau}_q) = 1 - \Theta\left(\frac{q_k}{w_k(x,v)}\right)$. It follows that

$$\begin{aligned} \tilde{\mathbb{P}}_x^{(k)}(\tau_y < \tilde{\tau}_q) &= \frac{w_k(x,y)}{q_k + w_k(x,y) + \sum_{v \in N_x \setminus \{y\}} w_k(x,v)(1 - \tilde{\mathbb{P}}_v^{(k)}(\tau_x < \tilde{\tau}_q))} \\ &= \frac{w_k(x,y)}{w_k(x,y) + \Theta(q_k) + \sum_{v \in N_x^{\leq}} w_k(x_n,v)(1 - \tilde{\mathbb{P}}_v^{(k)}(\tau_x < \tilde{\tau}_q))} \\ &= \frac{w_k(x,y)}{w_k(x,y) + \Theta(q_k)} = 1 - \Theta\left(\frac{q_k}{w_k(x,y)}\right). \end{aligned}$$

Thus we have that $\tilde{\mathbb{P}}_x^{(k)}(\tau_y < \tilde{\tau}_q) = 1 - o(1)$, from which it follows that $U_{q_k}^{(G_k)}(x,y) \to 0$. □

Lemma 2.18 (Parent hitting asymptotics with small q in hierarchical trees of bounded height). *For each $n \in \mathbb{N}$ let $G_n = (V_n, E_n, w_n)$ be a hierarchical tree of height $H = H_n$, see Figure 2.4a. Denote the weight of an edge at height $i \in [H]$ in G_n by $w_i^{(n)}$ and recall that $w_1^{(n)} \leq \ldots \leq w_H^{(n)}$.*

For each $n \in \mathbb{N}$ let y_n be a vertex in G_n at height $h = h_n$ such that $H_n - h_n$ is constant in n. Let x_n denote the parent of y_n. For each vertex v in G_n let $\ell_n(v)$ denote the number of vertices in G_n that have v in their ancestry. Let $(q_n)_{n\in\mathbb{N}}$ be a sequence of rooting parameters such that $q_n = o\left(\frac{w_h^{(n)}}{\ell_n(y)}\right)$.

For each $n \in \mathbb{N}$ let $\tilde{\mathbb{P}}_x^{(n)}$ denote the law of the discrete-time random walk $\tilde{X}$ on G_n starting at vertex $x \in V_n$ and let $\tilde{\tau}_q$ be a geometric killing time, as defined in Equation (A.6)*. Let τ_x denote the first hitting time of x by $\tilde{X}$. Then as $n \to \infty$ it holds that*

$$\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) \sim 1 - \frac{q_n \ell_n(y)}{w_h^{(n)}}.$$

Proof. Write $k = H_n - h_n$, which is independent of n. We proceed by induction on k.

For $k = 0$ we have that all vertices y_n are leaves. We then have that $\ell_n(y) = 1$, so that $q_n = o(w_H^{(n)})$. It follows that

$$\mathbb{P}_y^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{w_H^{(n)}}{w_H^{(n)} + q_n} \sim 1 - \frac{q_n}{w_H^{(n)}}.$$

Now assume that $k > 0$. Let $C_y^{(n)} \subseteq V_n$ denote the set of child vertices of y_n in G_n. Note that since $k > 0$, we have for all n that $C_y^{(n)}$ is non-empty. For each $n \in \mathbb{N}$ let z_n be a child of y_n. Note that $\frac{w_h^{(n)}}{\ell_n(y)} \leq \frac{w_{h+1}^{(n)}}{\ell_n(z)}$. This means that $q_n = o\left(\frac{w_{h+1}^{(n)}}{\ell_n(z)}\right)$. Thus by the induction hypothesis we then have that

$$\mathbb{P}_z^{(n)}(\tau_y < \tilde{\tau}_q) \sim 1 - \frac{q_n \ell_n(z)}{w_{h+1}^{(n)}}.$$

Since this holds for all possible choices of sequences of children of y_n, Lemma A.8 stated at the end of this section gives us that

$$\sum_{z \in C_y^{(n)}} \mathbb{P}_z^{(n)}(\tau_y < \tilde{\tau}_q) \sim \sum_{z \in C_y^{(n)}} 1 - \frac{q_n \ell_n(z)}{w_{h+1}^{(n)}}. \tag{2.56}$$

Note that for all n it holds that

$$\begin{aligned}\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) = {} & \tilde{\mathbb{P}}_y^{(n)}(X_1 = x) + \tilde{\mathbb{P}}_y^{(n)}(X_1 = y)\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) \\ & + \sum_{z \in C_y^{(n)}} \tilde{\mathbb{P}}_y^{(n)}(X_1 = z)\tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q)\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q).\end{aligned}$$

Solving this equation gives us that

$$\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{\tilde{\mathbb{P}}_y^{(n)}(\tilde{X}_1 = x)}{1 - \tilde{\mathbb{P}}_y^{(n)}(\tilde{X}_1 = y) - \sum_{z \in C_y^{(n)}} \tilde{\mathbb{P}}_y^{(n)}(\tilde{X}_1 = z)\tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q)}$$
$$= \frac{w_h^{(n)}}{q_n + w_h^{(n)} + w_{h+1}^{(n)} \sum_{z \in C_y^{(n)}} \left(1 - \tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q)\right)}. \tag{2.57}$$

Since $q_n = o\left(\frac{w_h^{(n)}}{\ell_n(y)}\right)$, we then have that

$$\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{w_h^{(n)}}{q_n + w_h^{(n)} + w_{h+1}^{(n)} \sum_{z \in C_y^{(n)}} \left(1 - \tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q)\right)}$$
$$\sim \frac{w_h^{(n)}}{q_n + w_h^{(n)} + w_{h+1}^{(n)} \sum_{z \in C_y^{(n)}} \frac{q_n \ell_n(z)}{w_{h+1}^{(n)}}}$$
$$= \frac{w_h^{(n)}}{q_n + w_h^{(n)} + q_n \sum_{z \in C_y^{(n)}} \ell_n(z)} = \frac{w_h^{(n)}}{w_h^{(n)} + q_n \ell_n(y)} \sim 1 - \frac{q_n \ell_n(y)}{w_h^{(n)}}.$$

□

Proof of Theorem 2.13. If d_n is bounded, then the result follows from Proposition 2.12. Hence, we can assume that $d_n \to \infty$ as $n \to \infty$. Since G_n is a complete d_n-ary tree, the number of vertices with y_n in their ancestry is given by $\ell_n(y) = \sum_{i=0}^{k} d_n^i$. This means that $\ell_n(y) \sim d_n^k$ as as $n \to \infty$. Hence, the case $q_n = o\left(\frac{w_n(e_n)}{d_n^k}\right)$ follows directly from Lemmas 2.16 and 2.18.

Assume that $q_k = \omega\left(\frac{w_k(e_k)}{d_k^m}\right)$. By Theorem 2.3 we can assume without loss of generality that also $q_n = o\left(\frac{w_n(e_n)}{d_n^{k-1}}\right)$.

For ech $n \in \mathbb{N}$ let $\tilde{\mathbb{P}}_x^{(n)}$ denote the law of the discrete-time random walk $\tilde{X}$ on G_n starting at vertex $x \in V_n$ and let $\tilde{\tau}_q$ be a geometric killing time, as defined in Equation (A.6). Let τ_x denote the first hitting time of x by $\tilde{X}$. By Lemma 2.16 it is sufficient to show that both $\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) \to 0$ and $\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) \to 0$ as $n \to \infty$.

First we consider $\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q)$. Let z_k be a child vertex of y_k. Then by Lemma 2.18 we have that

$$\tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q) \sim 1 - \frac{q_n \sum_{i=0}^{m-1} d_n^i}{w_n(y_n, z_n)}$$

So, by using that G_n is a complete d-ary tree, we have analogous to Equation (2.57) that

$$\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{w_n(e_n)}{q_n + w_n(e_n) + d_n \left(1 - \tilde{\mathbb{P}}_z^{(n)}(\tau_y < \tilde{\tau}_q)\right) w_n(y_n, z_n)}$$
$$\sim \frac{w_n(e_n)}{q_n + w_n(e_n) + d_n q_n \sum_{i=0}^{m-1} d_n^i} = \frac{w_n(e_n)}{q_n + w_n(e_n) + \omega(w_n(e_n))} = o(1).$$

It remains to show that $\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) \to 0$. Let u denote the parent of x. Then it holds that

$$\begin{aligned}\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) = &\tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = y) + \tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = x)\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) \\ &+ \tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = u)\tilde{\mathbb{P}}_u^{(n)}(\tau_x < \tilde{\tau}_q)\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) \\ &+ (d_n - 1)\tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = y)\mathbb{P}_y^{(k,q)}(\tau_x < \tilde{\tau}_q)\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q).\end{aligned}$$

This gives us that $\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q)$ is equal to

$$\begin{aligned}&\frac{\tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = y)}{1 - \tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = x) - \tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = u)\tilde{\mathbb{P}}_u^{(n)}(\tau_x < \tilde{\tau}_q) - (d_n - 1)\tilde{\mathbb{P}}_x^{(n)}(\tilde{X}_1 = y)\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q)\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q)} \\ &= \frac{w_n(e_n)}{q_n + (1 - \tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q))w_n(x,u) + w_n(e_n) + w_n(e_n)(d_n - 1)\left(1 - \tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q)\right)}.\end{aligned}$$

Since we have already shown that $\tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q) \to 0$, it follows that

$$\begin{aligned}\tilde{\mathbb{P}}_x^{(n)}(\tau_y < \tilde{\tau}_q) &\leq \frac{w_n(e_n)}{w_n(e_n) + w_n(e_n)(d_n - 1)\left(1 - \tilde{\mathbb{P}}_y^{(n)}(\tau_x < \tilde{\tau}_q)\right)} \\ &= \frac{1}{1 + (d_n - 1)\left(1 - o(1)\right)} = o(1).\end{aligned}$$

□

APPENDIX A

Appendix: Chapter 2

A.1 Graph reduction/extension lemmas

We introduce here some rather classical contraction tools. Though, we stress that the following definition of contraction is slightly different from what is often encountered in the UST literature, as it is adapted to the setting of weighted directed graphs.

Definition A.1.1 (Directed edge contraction). Let $G = (V, E, w)$ be a weighted directed graph and $e \in E$ a directed edge from vertex x to y, i.e. $e = (x, y)$. The graph $G\vec{/}e$ obtained by performing *the directed edge contraction* in G over edge e is the graph obtained by first removing all outgoing edges of x and then contracting x and y into a single vertex, while retaining all outgoing edges from y and all ingoing edges to both x and y.

If B is a set of edges that constitutes a rooted forest of G, then the operations of performing a directed edge contraction on different edges in B commute. Thus for such a B we can define the graph $G\vec{/}B$ to be the graph obtained by performing directed edge contractions on all edges in B. ■

Besides this notation for directed edge contractions, we will also use the standard notation $G - e$ to denote the graph obtained by removing the directed edge e (without removing the reversed edge), and G/e to denote a regular edge contraction over edge e, i.e. G/e is the graph obtained by identifying the two endpoints of e as a single vertex.

Lemma A.1 (Various expressions for edge probabilities). *Let $G = (V, E, w)$ be a weighted directed graph and $e = (x, y)$ a directed edge from vertex x to y. Let R_q be the set of root vertices of Φ_q. Let L denote the negative graph Laplacian of G and for $q > 0$ write $K_q = q(qI - L)^{-1}$. For each directed edge e write $G\vec{/}e$ to denote the directed e-contraction of G. Then it holds that*

$$\mathbb{P}(e \in \Phi_q) = \tfrac{w(e)}{q}\mathbb{P}(x \in R_q,\ x \not\leftrightarrow_{\Phi_q} y) = \tfrac{w(e)}{q}(K_q(x, x) - K_q(y, x)) = w(e)\frac{Z_{G\vec{/}e}(q)}{Z_G(q)}.$$

Proof. Let $e = (x, y)$ be an edge from x to y. Let $\mathcal{A} = \{F \in \mathcal{F}_G : e \in F\}$ denote the set of rooted forests of G that do contain edge e. Write $\mathcal{H} = \{F \in \mathcal{F}_G : x \in R(F),\ x \not\leftrightarrow_F y\}$ to denote the set of forests in which x is a root that is not connected to y. Note that there is a one-to-one correspondence $f : \mathcal{A} \to \mathcal{H}$ given by $f(F) = F - e$. Moreover, it holds that $w(F) = w(e)w(f(F))$ and that $r(F) = r(f(F)) - 1$, where $r(F)$ denotes the number of roots of F. The first identity follows by summation over all forests in $\mathcal{A}$. For the second identity we use the Chebotarev-Shamis matrix-forest theorem [19], which states that $K_q(y,x) = \mathbb{P}(x \in R_q,\ x \leftrightarrow_{\Phi_q} y)$. The third identity follows by considering the bijection $g : \mathcal{H} \to \mathcal{F}_{G\vec{/}e}$ that sends all edges of a forest in $\mathcal{H}$ to their corresponding edges in $G\vec{/}e$. Note that here $G\vec{/}e$ could be a multigraph. This bijection satisfies $w(F) = w(g(F))$ and $r(F) = r(g(F)) + 1$, so that summation over all forests in $\mathcal{H}$ yields the result. □

The following lemma shows the well-known spatial Markov property for the UST, see e.g. [39], tailored to the rooted forest measure Φ_q.

Lemma A.2 (Directed Spatial Markov property). *Let $G = (V, E, w)$ be a weighted directed graph and $A, B \subseteq E$ two disjoint sets of directed edges. Then it holds for all $F \in \mathcal{F}_G$ with $F \cap A = \varnothing$ and $B \subseteq F$ that*

$$\mathbb{P}^{(G)}(\Phi_q = F \mid \Phi_q \cap A = \varnothing,\ B \subseteq \Phi_q) = \mathbb{P}^{((G-A)\vec{/}B)}(\Phi_q = F\vec{/}B). \tag{A.1}$$

For any edge $e \in E$ the partition function of G satisfies the deletion-contraction identity [76]

$$Z_G(q) = Z_{G-e}(q) + w(e)Z_{G\vec{/}e}(q). \tag{A.2}$$

Moreover, if G is a symmetric graph, then it holds that

$$\mathbb{P}^{(G)}(\Phi_q = F \mid \Phi_q \cap A = \varnothing, \pm B \subseteq \Phi_q) = \mathbb{P}^{((G-A)/B)}(\Phi_q = F/B), \tag{A.3}$$

where G/B denotes the regular edge contraction of all edges in B.

Proof. It is sufficient to show that the statement holds when $|A \cup B| = 1$, since the general statement then follows by induction. First assume that $B = \varnothing$ and $A = \{e\}$ for some edge $e \in E$. Let $\mathcal{A} = \{F \in \mathcal{F}_G \colon e \notin F\}$ denote the set of rooted forests of G that do not contain edge e. Write $r(F)$ to denotes the number of roots of the rooted forest F. There is a natural one-to-one correspondence $f : \mathcal{A} \to \mathcal{F}_{G-e}$ given by $f(F) = F$. Hence, we have for all $F \in \mathcal{A}$ that

$$\begin{aligned}
\mathbb{P}^{(G)}(\Phi_q = F \mid e \notin \Phi_q) &= \frac{\mathbb{P}^{(G)}(\Phi_q = F)}{\mathbb{P}^{(G)}(e \notin \Phi_q)} = \frac{q^{r(F)}w(F)}{\sum_{H\in\mathcal{A}} q^{r(H)}w(H)} \\
&= \frac{q^{r(F)}w(F)}{\sum_{H\in\mathcal{F}_{G-e}} q^{r(f^{-1}(H))}w(f^{-1}(H))} \\
&= \frac{q^{r(F)}w(F)}{\sum_{H\in\mathcal{F}_{G-e}} q^{r(H)}w(H)} = \frac{q^{r(F)}w(F)}{Z_{G-e}(q)} = \mathbb{P}^{(G-e)}(\Phi_q = F).
\end{aligned}$$

Assume instead that $A = \varnothing$ and $B = \{e\}$ for some edge $e \in E$. Then by lemma A.1 we have for all $F \in \mathcal{F}_G$ with $e \in F$ that

$$\mathbb{P}^{(G)}(\Phi_q = F \mid e \in \Phi_q) = \frac{\mathbb{P}^{(G)}(\Phi_q = F)}{\mathbb{P}^{(G)}(e \in \Phi_q)} = \frac{q^{r(F)}w(F)}{w(e)Z_{G\vec{/}e}(q)} = \mathbb{P}^{(G\vec{/}e)}(\Phi_q = F/e).$$

The proof of eq. (A.2) is analogous to that of eq. (A.1), while eq. (A.3) follows directly from the spatial Markov property for the UST. □

Lemmas A.3 and A.4 both represent the same simple combinatorial manipulation, but in two slightly different settings. The same manipulation can be extended beyond the simple setups of these lemmas, but for notational simplicity we stick to these versions, which are tailored to sparse geometries.

These lemmas are phrased in terms of the *non-normalized* rooted forest measure defined as

$$\nu^{(G)}(\Phi_q \in \cdot) = Z_G(q)\mathbb{P}^{(G)}(\Phi_q \in \cdot). \tag{A.4}$$

This measure has the benefit that the measure of a rooted forest depends on the geometry of the underlying graph only through the total number of vertices. That is, for any rooted forest $F \in \mathcal{F}_H$ of a subgraph H of G it holds that $q^m \nu^{(H)}(\Phi_q = F) = \nu^{(G)}(\Phi_q = F)$, where m is the difference between the number of vertices in G and H. This simplifies the notation required for various combinatorial manipulations.

Lemma A.3 (Graph extension lemma (single vertex version)). *Let $G = (V, E, w)$ be a weighted directed graph and $x \in V$ a vertex. Let R_q be the set of root vertices of Φ_q. Let $H = G[V \setminus \{x\}]$ denote the induced subgraph of G obtained by removing vertex x. Let $\{\mathcal{H}(F)\colon F \in \mathcal{F}_H\}$ be the partition of $\mathcal{F}_G$ given by $\mathcal{H}(F) = \{F' \in \mathcal{F}_G\colon F'[V \setminus \{x\}] = F\}$, i.e. $\mathcal{H}(F)$ denotes the set of spanning rooted forests of G for which the induced subgraph obtained by removing x equals F. For each vertex $y \in V \setminus \{x\}$ let $r_y(F)$ denote the unique root in F that is connected to y. Then it holds for all $F \in \mathcal{F}_H$ that*

$$\nu^{(G)}(\Phi_q \in \mathcal{H}(F),\ x \in R_q) = q\ \nu^{(H)}(\Phi_q = F) \prod_{r \in R(F)} (1 + \tfrac{w(r,x)}{q})$$

and that

$$\nu^{(G)}(\Phi_q \in \mathcal{H}(F),\ x \notin R_q) = \nu^{(H)}(\Phi_q = F) \sum_{y \in V\setminus\{x\}} w(x,y) \prod_{r \in R(F)\setminus\{r_y(F)\}} (1 + \tfrac{w(r,x)}{q}).$$

Here we take $w(e) = 0$ when $e \notin E$.

Proof of lemma A.3. We will first prove the first equality. Let $F_H \in \mathcal{F}_H$ be given. Each forest in $F \in \mathcal{H}(F_H)$ with $x \in R(F)$ can be obtained from F_H by adding any number of edges from roots of F_H to x. So, for each root we can choose either to add this edge or not to add this edge. For each edge we do add, there will be one less component, since the root from which that edge originated will cease to be a root in the new forest. This contributes a factor $\frac{1}{q}$. We then also have an additional edge, which contributes a factor equal to the weight of that edge. This gives us the product over the roots r, where the 1 term is chosen if no edge is added from r to x and the $\frac{w(r,x)}{q}$ term is chosen if we do add such an edge. If we don't add any such edges, then the obtained forest will have one more root than F_H, so this gives us the additional factor q.

The second equality is proven similarly. Each forest in $F \in \mathcal{H}(F_H)$ with $x \notin R(F)$ can be obtained from F_H by first adding a single edge from x to any other vertex y. We then add any number of edges from roots of F_H to x, but we cannot add an edge from r_y to x as this would create a cycle. □

Definition A.1.2. Let $G = (V, E)$ be a directed graph. Let $A \subseteq V$ be a set of vertices and denote by $G[A]$ the induced subgraph of G on the vertices in A. A set $\mathcal{H} \subseteq \mathcal{F}$ of rooted forests of G is said to be *determined* by A if there exists an $\mathcal{A} \subseteq \mathcal{F}_{G[A]}$ such that $\mathcal{H} = \{F \in \mathcal{F}_G\colon\ F[A] \in \mathcal{A}\}$. ■

Lemma A.4 (Graph extension lemma (single edge version)). *Let $G = (V, E, w)$ be a weighted directed graph. Let $\{A, B\}$ be a partition of V and assume that there exists exists exactly one vertex $a \in A$ that is adjacent to any vertices in B and exactly one vertex $b \in B$ adjacent to any vertices in A. Write $G[A]$ and $G[B]$ to denote the induced subgraphs on A and B. Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$ be sets of rooted forests of G that are determined by A and B, respectively, and let $\mathcal{A}' \subseteq \mathcal{F}_{G[A]}$ and $\mathcal{B}' \subseteq \mathcal{F}_{G[B]}$ be such that $\mathcal{A} = \{F \in \mathcal{F}_G\colon\ F[A] \in \mathcal{A}'\}$ and $\mathcal{B} = \{F \in \mathcal{F}_G\colon\ F[B] \in \mathcal{B}'\}$. Denote by R_q the set of root vertices of Φ_q. Then it holds that*

$$\begin{aligned}\nu^{(G)}(\Phi_q \in \mathcal{A} \cap \mathcal{B}) = {}& \nu^{(G[A])}(\Phi_q \in \mathcal{A}')\ \nu^{(G[B])}(\Phi_q \in \mathcal{B}') \\ &+ \tfrac{w(a,b)}{q} \nu^{(G[B])}(\Phi_q \in \mathcal{B}')\ \nu^{(G[A])}(\Phi_q \in \mathcal{A}',\ a \in R_q) \\ &+ \tfrac{w(b,a)}{q} \nu^{(G[A])}(\Phi_q \in \mathcal{A}')\ \nu^{(G[B])}(\Phi_q \in \mathcal{B}',\ b \in R_q).\end{aligned}$$

Proof. Let $F_A \in \mathcal{A}'$ and $F_B \in \mathcal{B}'$ be given.

If both a is a root in F_A and b is a root in F_B, then there are exactly three forests $F_1, F_2, F_3 \in \mathcal{F}_G$ for which the induced subgraphs on A and B correspond to F_A and F_B, respectively.

(a) The first of these forests consists of the disjoint graph union of F_A and F_B and has non-normalized measure

$$\nu^{(G)}(\Phi_q = F_1) = \nu^{(G[A])}(\Phi_q = F_A)\ \nu^{(G[B])}(\Phi_q = F_B).$$

(b) The second has an additional edge from a to b and has non-normalized measure

$$\nu^{(G)}(\Phi_q = F_2) = \tfrac{w(a,b)}{q} \nu^{(G[A])}(\Phi_q = F_A)\ \nu^{(G[B])}(\Phi_q = F_B),$$

since it contains one less root than the sum of the roots in F_A and F_B and one additional edge with weight $w(a, b)$.

(c) The third forest has an additional edge from b to a and it similarly has non-normalized measure

$$\nu^{(G)}(\Phi_q = F_3) = \tfrac{w(b,a)}{q} \nu^{(G[A])}(\Phi_q = F_A)\ \nu^{(G[B])}(\Phi_q = F_B).$$

Note that each of these three forests is contained in $\mathcal{A} \cap \mathcal{B}$.

If exactly one of the vertices a and b is a root in F_A and F_B, then only two of the above mentioned forests are rooted forest of G, since adding an outgoing edge to a non-root vertex does not yield a rooted forest.

If both a and b are not roots, then only the first forest without an additional edge is a rooted forest of G.

Since each forest in $\mathcal{A} \cap \mathcal{B}$ can be obtained in such a manner, summing over all rooted forests in $\mathcal{A}'$ and $\mathcal{B}'$ yields

$$\begin{aligned}
\nu_q^{(G)}(\mathcal{A} \cap \mathcal{B}) = \sum_{F_A \in \mathcal{A}'} \sum_{F_B \in \mathcal{B}'} & \nu^{(G[A])}(\Phi_q = F_A)\nu^{(G[B])}(\Phi_q = F_B) \\
& + \tfrac{w(a,b)}{q}\nu^{(G[A])}(\Phi_q = F_A)\nu^{(G[B])}(\Phi_q = F_B)\mathbf{1}_{\{b \in R(F_A)\}} \\
& + \tfrac{w(b,a)}{q}\nu^{(G[A])}(\Phi_q = F_A)\nu^{(G[B])}(\Phi_q = F_B)\mathbf{1}_{\{b' \in R(F_B)\}} \\
= \nu^{(G[A])}(\Phi_q \in \mathcal{A}')\ & \nu^{(G[B])}(\Phi_q \in \mathcal{B}') \\
+ \tfrac{w(a,b)}{q}\nu^{(G[A])}(\Phi_q \in \mathcal{A}',\ a \in R_q)\ & \nu^{(G[B])}(\Phi_q \in \mathcal{B}') \\
+ \tfrac{w(b,a)}{q}\nu^{(G[A])}(\Phi_q \in \mathcal{A}')\ & \nu^{(G[B])}(\Phi_q \in \mathcal{B}',\ b \in R_q).
\end{aligned}$$

□

A.2 Technical lemmas

The next three lemmas are simple statements used in the proof of theorem 2.5 in section 2.3.2.1.

Lemma A.5 (Bound on derivative of hitting probabilities). *Let $G = (V, E, w)$ be a weighted directed graph and $x, y \in V$ two vertices. Let $\mathbb{P}_x$ denote the law of the random walk X on G starting at x. For each $v \in V$ let τ_v denote the hitting time of v by X and let τ_q be an independent exponential killing time with rate q. Then it holds for the derivative of the function $q \mapsto \mathbb{P}_x(\tau_y < \tau_q)$ that*

$$\tfrac{1}{q}\mathbb{P}_x(\tau_y < \tau_q) - \tfrac{1}{q} \leq \frac{d}{dq}\mathbb{P}_x(\tau_y < \tau_q) \leq 0. \tag{A.5}$$

In the subsequent proofs it will be convenient to work with the discrete-time skeleton of the random walk X, that is, the discrete-time random walk $\tilde{X}$ on G with transition matrix

$$P = I + \tfrac{1}{\alpha}L, \tag{A.6}$$

with α the maximal diagonal entry of the graph Laplacian $-L$. The path measure of $\tilde{X}$ starting at x is denoted by $\tilde{\mathbb{P}}_x$. For $\tilde{\tau}_q$ an independent ($\mathbb{N}$-valued) geometric killing time with success probability $\frac{q}{q+\alpha}$, it then holds that $\mathbb{P}_x(\tau_y < \tau_q) = \tilde{\mathbb{P}}_x(\tau_y < \tilde{\tau}_q)$. Since the law of the loop-erased trajectory of $\tilde{X}$ corresponds to that of X, we can also use this discrete-time random walk to analyze eq. (2.8).

Proof of lemma A.5. The upper bound on the derivative in (A.5) is immediate, we therefore show the lower bound.

Let $\tilde{\mathbb{P}}_x$ denote the law of the discrete-time random walk $\tilde{X}$, as defined in eq. (A.6).

Then it holds that

$$\mathbb{P}_x(\tau_y < \tau_q) = \tilde{\mathbb{P}}_x(\tau_y < \tilde{\tau}_q) = \sum_{k=1}^{\infty} \tilde{\mathbb{P}}_x(\tau_y < k)\mathbb{P}(\tilde{\tau}_q = k)$$

$$= \sum_{k=1}^{\infty} \tilde{\mathbb{P}}_x(\tau_y < k)\tfrac{q}{q+\alpha}\left(1 - \tfrac{q}{q+\alpha}\right)^{k-1}.$$

Since $\tilde{\mathbb{P}}_x(\tau_y < k)$ does not depend on q, it follows that

$$\frac{d}{dq}\mathbb{P}_x(\tau_y < \tau_q) = \sum_{k=1}^{\infty} \tilde{\mathbb{P}}_x(\tau_y < k)\frac{(1-k)q+\alpha}{(q+\alpha)^2}\left(\frac{\alpha}{q+\alpha}\right)^{k-1}$$

$$= \tfrac{1}{q}\mathbb{P}_x(\tau_y < \tau_q) - \sum_{k=1}^{\infty} \tilde{\mathbb{P}}_x(\tau_y < k)\frac{kq}{(q+\alpha)^2}\left(\frac{\alpha}{q+\alpha}\right)^{k-1}$$

$$\geq \tfrac{1}{q}\mathbb{P}_x(\tau_y < \tau_q) - \sum_{k=1}^{\infty} \frac{kq}{(q+\alpha)^2}\left(\frac{\alpha}{q+\alpha}\right)^{k-1}$$

$$= \tfrac{1}{q}\mathbb{P}_x(\tau_y < \tau_q) - \tfrac{1}{q+\alpha}\mathbb{E}(\tilde{\tau}_q) = \tfrac{1}{q}\mathbb{P}_x(\tau_y < \tau_q) - \tfrac{1}{q}.$$

□

Lemma A.6 (Monotonicity of rooting probabilities). *Let $G = (V, E, w)$ be a weighted directed graph and $x \in V$ a vertex. Let $\mathbb{P}$ denote the law of a random spanning rooted forest Φ_q of G with rooting parameter $q > 0$. Let R_q denote the set of roots of Φ_q. Then it holds that*

$$0 \leq \frac{d}{dq}\mathbb{P}(x \in R_q) \leq \tfrac{1}{q}\mathbb{P}(x \in R_q). \tag{A.7}$$

Proof of lemma A.6 via lemma A.5. By (2.6) we have that x is a root in Φ_q if

$$\mathbb{P}(x \in R_q) = K_q(x,x) = q(qI - L)^{-1}(x,x) = \mathbb{P}_x(X_{\tau_q} = x).$$

Let N_x denote the set of out-neighbours of x in G. Let $\sigma = \inf\{t > 0\colon\ X_t \neq X_0\}$ be the first jump time of X. Then by the Markov property of X we have that

$$\mathbb{P}_x(X_{\tau_q} = x) = \mathbb{P}_x(\sigma > \tau_q) + \sum_{v \in N_x} \mathbb{P}_x(X_\sigma = v)\mathbb{P}_v(\tau_x < \tau_q)\mathbb{P}_x(X_{\tau_q} = x).$$

Solving this equation gives us that

$$\mathbb{P}_x(X_{\tau_q} = x) = \frac{\mathbb{P}_x(\sigma > \tau_q)}{1 - \sum_{v \in N_x} \mathbb{P}_x(X_\sigma = v)\mathbb{P}_v(\tau_x < \tau_q)}$$

$$= \frac{q}{q + \sum_{v \in N_x} w(x,v)(1 - \mathbb{P}_v(\tau_x < \tau_q))}.$$

It follows by lemma A.5 that

$$\frac{d}{dq}\mathbb{P}(x \in R_q) = \frac{d}{dq}\mathbb{P}_x(X_{\tau_q} = x)$$

$$= \frac{\sum_{v \in N_x} w(x,v)\left(1 - \mathbb{P}_v(\tau_x < \tau_q) + q\frac{d}{dq}\mathbb{P}_v(\tau_x < \tau_q)\right)}{\left(q + \sum_{v \in N_x} w(x,v)(1 - \mathbb{P}_v(\tau_x < \tau_q))\right)^2} \geq 0,$$

which proves the lower bound. For the upper bound it holds that

$$\frac{\frac{d}{dq}\mathbb{P}(x\in R_q)}{\mathbb{P}(x\in R_q)} = \frac{\sum_{v\in N_x} w(x,v)\left(1-\mathbb{P}_v(\tau_x<\tau_q)+q\frac{d}{dq}\mathbb{P}_v(\tau_x<\tau_q)\right)}{q\left(q+\sum_{v\in N_x} w(x,v)(1-\mathbb{P}_v(\tau_x<\tau_q))\right)}$$
$$\leq \frac{\sum_{v\in N_x} w(x,v)\left(1-\mathbb{P}_v(\tau_x<\tau_q)\right)}{q\left(q+\sum_{v\in N_x} w(x,v)(1-\mathbb{P}_v(\tau_x<\tau_q))\right)} \leq \frac{1}{q}.$$

□

Lemma A.7 (Bound on conditional rooting derivative in trees). *Let $G = (V, E, w)$ be a weighted directed tree and $x, y \in V$ two vertices. Then it holds that*

$$\tfrac{d}{dq}\mathbb{P}(x\in R_q \mid x\leftrightarrow y) \leq \tfrac{1}{q}\mathbb{P}(x\in R_q \mid x\leftrightarrow y). \tag{A.8}$$

Proof of lemma A.7. Let d denote the distance between x and y. We will argue inductively on d. For $d = 0$ the statement follows from lemma A.6.

Now assume that $d \geq 1$. Let z denote the vertex adjacent to x with distance $d-1$ to y. Note that we possibly have that $z = y$. Since G is a tree, removing the edges between x and z splits the graph into two components T_x and T_z, where T_x and T_z denote the component containing vertex x and z, respectively. It then holds by lemma A.4 that

$$\begin{aligned}
&\mathbb{P}(x\in R_q \mid x\leftrightarrow y)\\
&= \frac{w(z,x)\nu^{(T_x)}(x\in R_q)\nu^{(T_z)}(z\in R_q, z\leftrightarrow y)}{w(z,x)Z_{T_x}(q)\nu^{(T_z)}(z\in R_q, z\leftrightarrow y)+w(x,z)\nu^{(T_x)}(x\in R_q)\nu^{(T_z)}(z\leftrightarrow y)}\\
&= \frac{w(z,x)\mathbb{P}^{(T_z)}(z\in R_q \mid z\leftrightarrow y)\mathbb{P}^{(T_x)}(x\in R_q)}{w(z,x)\mathbb{P}^{(T_z)}(z\in R_q \mid z\leftrightarrow y)+w(x,z)\mathbb{P}^{(T_x)}(x\in R_q)}.
\end{aligned}$$

It follows by the induction hypothesis and lemma A.6 that

$$\begin{aligned}
&\frac{d}{dq}\mathbb{P}(x\in R_q \mid x\leftrightarrow y)/\mathbb{P}(x\in R_q \mid x\leftrightarrow y)\\
&= \frac{w(z,x)\mathbb{P}^{(T_x)}(x\in R_q)^2\frac{d}{dq}\mathbb{P}^{(T_z)}(z\in R_q|z\leftrightarrow y)+w(x,z)\mathbb{P}^{(T_z)}(z\in R_q|z\leftrightarrow y)^2\frac{d}{dq}\mathbb{P}^{(T_x)}(x\in R_q)}{w(z,x)\mathbb{P}^{(T_z)}(z\in R_q|z\leftrightarrow y)\mathbb{P}^{(T_x)}(x\in R_q)^2+w(x,z)\mathbb{P}^{(T_z)}(z\in R_q|z\leftrightarrow y)^2\mathbb{P}^{(T_x)}(x\in R_q)}\\
&\leq \frac{1}{q}.
\end{aligned}$$

□

The simple lemma below has been used to show eq. (2.56).

Lemma A.8. *For each $n \in \mathbb{N}$ let $\ell_n \in \mathbb{N}$ be given and let $(\alpha_i^{(n)})_{i\in[\ell_n]}$ and $(\beta_i^{(n)})_{i\in[\ell_n]}$ be real valued sequences of length ℓ_n. Let $\mathcal{F} = \{f \in \mathbb{N}^{\mathbb{N}} : f(n) \in [\ell_n] \text{ for all } n \in \mathbb{N}\}$ denote the set of choice functions on the collection $\{[\ell_1], [\ell_2], \ldots\}$. Assume that for each $f \in \mathcal{F}$ it holds that $\alpha_{f(n)}^{(n)} \sim \beta_{f(n)}^{(n)}$ as $n \to \infty$. Then as $n \to \infty$ it holds that*

$$\sum_{i=1}^{\ell_n}\alpha_i^{(n)} \sim \sum_{i=1}^{\ell_n}\beta_i^{(n)}.$$

Proof. For all $\varepsilon > 0$ and each $f \in \mathcal{F}$, there exists an $N(\varepsilon, f) \in \mathbb{N}$ such that for all $n \geq N(\varepsilon, f)$ it holds that

$$\left|\frac{\alpha^{(n)}_{f(n)}}{\beta^{(n)}_{f(n)}} - 1\right| < \varepsilon.$$

Define the function $f^* \in \mathcal{F}$ by

$$f^*(n) = \operatorname{argmax}_{i \in [\ell_n]} \left|\alpha_i^{(n)} - \beta_i^{(n)}\right|.$$

Then for all $\varepsilon > 0$ and all $n \geq N(\varepsilon, f^*)$ it holds that

$$\begin{aligned}\left|\frac{\sum_{i=1}^{\ell_n} \alpha_i^{(n)}}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} - 1\right| &= \frac{1}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \left|\sum_{i=1}^{\ell_n} \alpha_i^{(n)} - \sum_{i=1}^{\ell_n} \beta_i^{(n)}\right| \leq \frac{\sum_{i=1}^{\ell_n} \left|\alpha_i^{(n)} - \beta_i^{(n)}\right|}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \\ &= \frac{\sum_{i=1}^{\ell_n} \beta_i^{(n)} \left|\frac{\alpha_i^{(n)}}{\beta_i^{(n)}} - 1\right|}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \leq \left|\frac{\alpha^{(n)}_{f^*(n)}}{\beta^{(n)}_{f^*(n)}} - 1\right| \frac{\sum_{i=1}^{\ell_n} \beta_i^{(n)}}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} = \left|\frac{\alpha^{(n)}_{f^*(n)}}{\beta^{(n)}_{f^*(n)}} - 1\right| < \varepsilon.\end{aligned}$$

□