LOOP-ERASED PARTITIONING OF CYCLE-FREE GRAPHS

LUCA AVENA†, TWAN KOPERBERG⋆, AND JANNETJE DRIESSEN§

ABSTRACT. We consider random partitions of the vertex set of a given finite graph that can be sampled by means of loop-erased random walks stopped at a random independent exponential time of parameter $q > 0$. The related random blocks tend to cluster nodes visited by the random walk associated to the graph on time scale $1/q$. This random partitioning is induced by a measure of spanning rooted forest of the graph which naturally generalize the classical uniform spanning tree measure and which can be obtained as a "zero-limit" of FK-percolation with an external cemetery state. General properties of this rooted forest measure and related determinantal observables, along with a number of applications in data analysis have been recently explored. We are here interested in understanding the emergent partitioning, referred to as loop-erased partitioning, as the intensity parameter $q$ scales with the graph size. Some first results in this direction have been investigated in the recent [6] for dense geometries. In this work we instead look at very sparse geometries. We start by characterising monotone events in $q$ by deriving a general Russo-like formula for this rooted arboreal gas measure. We then explore the resulting clusters on simple insightful tree-like graphs by looking at the probability that two given vertices do not belong to the same block. In particular, a detailed analysis of the resulting integer partitioning on line segments is pursued. We finally look at simple trees and other almost tree-like geometries, without and with implanted modular structures. For the latter, we characterize the emergence of giants and asymptotic detection of these implanted modules as the scale parameter $q$ varies.

2020 Mathematics Subject Classification. 05A18, 05C05, 05C81, 05C85, 60J10, 60J28.

Key words and phrases. Graph Laplacian, random partitions, loop-erased random walk, spanning rooted forests, determinantal processes.
1. Spanning rooted forests and Loop-erased partitioning.

Consider an arbitrary directed weighted finite graph $G = (V, E, w)$ on $N = |V|$ vertices where $E = \{e = (x, y) : x, y \in V\}$ stands for the edge set and $w : E \to [0, \infty)$ is a given edge-weight function. We call the Random Walk (RW) associated to $G$ the continuous-time Markov chain $X = (X_t)_{t \geq 0}$ with state space $V$ and the discrete Laplacian as infinitesimal generator, i.e., the $N \times N$ matrix:

$$L = A - D,$$

where for any $x, y \in [N] := \{1, 2, \ldots, N\}$, $A(x, y) = w(x, y)1_{\{x \neq y\}}$ is the weighted adjacency matrix and $D(x, y) = 1_{\{x = y\}} \sum_{z \in [N] \setminus \{x\}} w(x, z)$ is the diagonal matrix guaranteeing that the entries of each row in $L$ sum up to 0.

Let $F$ denote the space of spanning rooted forests of $G$, where a spanning rooted forest $F \in F$ of a graph is a collection of rooted trees spanning its vertex set. We consider a rooted tree to be a collection of directed edges pointing towards the root.

Fix a positive parameter $q > 0$ and let $\Phi_q$ be the random variable with values in $F$ with law:

$$\mathbb{P}(\Phi_q = F) = \frac{q^r(F)w(F)}{Z(q)}, \quad F \in F,$$

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 $F \in F$, and $Z(q)$ is a normalizing constant referred to as partition function. We will refer to this measure as random spanning rooted forest or rooted arboreal gas of intensity $q$. In the unitary weight case $w \equiv 1$, when $q = 1$, this measure becomes uniform over the set of rooted spanning forest $F$ and its structure has been partially analyzed in several geometrical setups in relation to random combinatorial models in statistical physics and coalescence theory, see [15, 24, 25, 27, 28, 35, 36]. For any $q > 0$, $\Phi_q$ 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 at different scales. 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 $V$. We refer to this object, defined next, as the Loop-Erased Partitioning (LEP). Its analysis has been initiated on dense graphs in the recent [6], here we instead look at very sparse topologies.

**Definition 1 (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 $q$ of the graph, the random partition, denoted by $\Pi_q$, of $V$, with law:

$$\mathbb{P}(\Pi_q = \pi_m) = \frac{q^m \times \sum_{F \in F, \pi(F) = \pi_m} w(F)}{Z(q)}, \quad \pi_m \in \mathcal{P}(V), m \leq N$$

where the sum runs over the space of spanning rooted forests $F$ 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 $\pi_m$. Equivalently,

$$\Pi_q := \pi(\Phi_q).$$

**Uniform spanning tree, rooted forest and LEP:** The rooted forest $\Phi_q$ is a natural extension of the classical UST (Uniform Spanning Tree) measure which is readily recovered in the constant weight case $w \equiv 1$ by taking the limit of $q$ going to zero in Eq. (1.2). Alternatively, this rooted arboreal gas $\Phi_q$ can also be seen as a measure on weighted spanning trees on the extended weighted graph obtained by adding an extra 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 arboreal gas. For example, edges in $\Phi_q$ form a determinantal process [4], a version of the so-called transfer-current theorem [14] clarifying its status within negatively associated systems, see [18, 26, 32]. Due to the Kirchhoff’s matrix tree theorem, the normalizing constant in Eq. (1.3) can be expressed as the characteristic polynomial of the matrix $L$ evaluated at $q$, i.e.

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

where $A(q) := qA - qI$ is the weighted adjacency matrix and $D(q) := qI$ is the diagonal matrix guaranteeing that the entries of each row in $A(q)$ sum up to $q$. Here the normalizing constant $Z(q)$ is a determinant with entries

where $A(q) := qA - qI$ is the weighted adjacency matrix and $D(q) := qI$ is the diagonal matrix guaranteeing that the entries of each row in $A(q)$ sum up to $q$. Here the normalizing constant $Z(q)$ is a determinant with entries

$$Z(q) := \sum_{F \in F} q^{r(F)}w(F) = \det[qI - L].$$
see e.g. [4, 16]. As far as sampling is concerned, for fixed $q > 0$, one can use the celebrated algorithm due to Wilson [37] 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 FK-percolation “super-model” by properly taking the related “interaction parameter” to zero, see e.g. [20]. Not surprisingly, as expressed in Lemma 1 below, which for simplicity we state in the unitary weight case $w \equiv 1$, the rooted forest in Eq. (1.2) 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 lemma is as in [20], see Thm. 1.23 in Sect 1.5 therein, with the parameters of the FK as specified in the statement below.

**Lemma 1 (Rooted forest as zero-limit of extended FK-percolation).** Given an undirected simple graph $G = (V,E)$, let $G_1 := (V_1, E_1)$ be the extended graph with $V_+ := V \cup \{\dagger\}$ where $\{\dagger\}$ denotes an extra state, $E_1 := E \cup \vec{E}$ with $\vec{E} := \{(x, \dagger) : x \in V\}$. Consider the generalized FK-percolation on $G_1$ with parameter $\lambda > 0$ and vector of weights $\vec{p} = (p_e)_{e \in E_1}$ such that $p_e = p \in (0,1)$ if $e \in E$ and $p_e = \gamma > 0$ for $e \in \vec{E}$, that is, the following measure on subgraphs of $E$ seen as collection of edges in $\Omega := \{0,1\}^{E_1}$:

$$P(FK = \omega) = \frac{\lambda^{k(\omega)} \prod_{e \in E} p_e^{\omega(e)} (1-p)^{\omega(e)} \prod_{\bar{e} \in \vec{E}} \gamma^{\omega(\bar{e})} (1-\gamma)^{\omega(\bar{e})}}{Z(\lambda, \vec{p})}, \quad \omega \in \Omega,$$

(1.6)

with $k(\omega)$ counting the number of connected components of the graph $\omega$ and $Z(\lambda, \vec{p})$ being a normalizing constant. Assume that $\vec{p}$ is a function of $\lambda$ 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 $\lambda$ goes to zero, the law in Eq. (1.6) degenerates into the law of the random rooted forest $\Phi_q$ in Eq. (1.2) with unitary weights.

Yet, if the UST can be seen as the “static global random backbone” of a given network, the forest process $(\Phi_q)_{q>0}$ represents 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 approaches (see [4, Thm.2]), 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, 4, 5] we started to explore these new features. For example, the roots [4, Prop.2.2] in $\Phi_q$ form a determinantal point process with kernel given by the RW Green’s function, that is: for any $A \subset V$

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

(1.7)

with $[K_q]_A$ being the restriction of the matrix $K_q := q(q - L)^{-1}$ to the set of indices in $A$. The number of roots (or trees, or blocks in $\Pi_q$), is distributed as the sum of $N$ independent Bernoulli random variables with success probabilities $\frac{q}{4N}$, for $i \leq N$, with the $\lambda_i$’s being the eigenvalues of $-L$, or their real parts, see e.g. [4, Prop. 2.1], its mean number is monotonically increasing in $q$. Further, these roots turn out to be well-distributed in the given network [4, Thm.1] and, conditional on the induced partition, their law is as the stationary measures of the random walk $X$ restricted to each block [4, Prop.2.3] of the underlying partition. These and other features of the LEP have been recently exploited to build novel algorithms for the following different applications in data science: wavelets basis and filters for signal processing on graphs [3, 33, 34], estimate traces of discrete Laplacians and other diagonally dominant matrices [8], network renormalization [1, 2], centrality measures [16] and statistical learning [7]. These applications give further motivations to explore in more details this LEP. Apart from a few general statements, the main focus in here is to characterize in details the emergent partition on nearly tree-like geometries for which the limited presence of cycles simplify the analysis. Some related results in dense setups have been recently derived in [6]. Let us also stress that on certain geometrical setups 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 dynamical structure [4, see Thm.2 and Sect2.2] within the theory of coalescence-fragmentation processes. The results derived here can form the basis for these future studies.

**Other forest measures:** To conclude this introduction let us clarify that this rooted forest $\Phi_q$ should not be confused with other forest measures that have been receiving great 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. [11, 12, 21, 22, 31] 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 recently investigated in [9, 10].

Results overview and paper structure: The rest of the paper is organized as follows. We start in Section 1.1
by deriving an interesting general characterization of monotone events in $q$, Theorem 1, and introducing the two-
point correlation function which we will analyse in different setups to study the emergent partition. We then state
in Section 1.2 two facts on the LEP potential on arbitrary finite trees: an inclusion-exclusion reduction formula,
Proposition 1, and its monotonicity in $q$, Theorem 2. In Section 1.3, we focus on the LEP on the first $n$ integers where
equipartitions are favoured. Formulas for the partition function are first derived in Thm 3, and extended to a ring,
Corollary 1. Theorem 4 gives a recursive representation of the pairwise correlation in terms of reduced partition
functions and offers bounds in terms of the correspondent RW on the infinite line. The subsequent Corollary 2
shows explicit bulk and boundary asymptotics. Section 1.4 is then devoted to exploration of the emergent blocks
in tree-like structures. In particular, Theorems 5 and 6 look at a star graph without and with a community structure, respectively. Proposition 2 and Theorem 7 show
similar analysis on finite trees with different weighted structures as a function of $q$ in tree-like structures. In particular, Theorems 5 and 6 are those for which the difference
$E[r_q] - E[r_q]$ has a constant sign as $q$ varies. Though it may be difficult to check the sign of this difference
for specific events, still, when possible we believe this statement can be handy and of great help. We also mention
that in [4] 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 it is not useful
to characterize monotone events.

As anticipated, our main interest within this work is to explore the structure of this loop-erased partitioning on
trees and nearly-one-dimensional geometries. To do so, we will mainly analyze two-point correlations associated to
$\Pi_q$, 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 $\Pi_q$. That is, the 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 $\Pi_q$ containing $z \in V$.

\[ \text{This coupling corresponds to an explicit coalescence-fragmentation process (Markovian after proper time-
rescaling) with values in } F \text{ in which coalescence of trees is dominant but whenever the underlying building RW}
\text{produces a loop, a tree gets fragmented into subtrees, see [4, see Thm.2 and Sect.2.2].} \]
Definition 2 (2-point correlations or pairwise LEP-interaction potential). For given $q > 0$ and $G$, and any pair $x, y \in V$, we call pairwise LEP-interaction potential the following probability:

$$U_q(x, y) := \mathbb{P}(B_q(x) \neq B_q(y)) = \sum \mathbb{P}_z^{LE}(\gamma)\mathbb{P}_y(\tau_\gamma > \tau_q)$$

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

The representation in Eq. (1.9) 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 [37].

Furthermore, we notice that, as for any generic random partition of $V$, such an interaction potential 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 $G$. Thus providing a further motivation to analyze it.

1.2. Two-point-correlation on trees. We start here to discuss results specific to trees. Let us notice that in this setup, the analysis is facilitated by the absence of cycles. In particular, a rooted forest induces a unique rooted partition. So that, for example in the constant weight case $\mathcal{E}$, that in this setup, the analysis is facilitated by the absence of cycles. In particular, a rooted forest induces a unique partition. Thus providing a further motivation to analyze it.

Proposition 1 (Inclusion-exclusion for 2-point-correlation on trees). Let $G = (V, E, w)$ be a weighted directed tree. For any $x, y \in V$ with distance $d$ let $(z_i)_{i=0}^d$ be the unique sequence of vertices with $z_0 = x$, $z_d = y$ and such that $z_i$ and $z_{i-1}$ have distance 1 for all $i \in [d]$. 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 $|I| + 1$ (weakly) connected components of $G_I$ by $G_I^1, \ldots, G_I^{|I|+1}$. Then, for every $q > 0$, the following representation is valid

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

(1.10)

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

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

$$U_q(x, y) = \frac{Z_x(q)Z_y(q)}{Z(q)},$$

(1.11)

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 $y$.

Theorem 2 (Monotonicity of 2-point-correlation on trees). For any weighted directed tree $G = (V, E, w)$ with weights $w : E \rightarrow (0, \infty)$, and any distinct $x, y \in V$,

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

is a continuous non-decreasing function.

(1.12)
Theorem 3. emerge neatly from the asymptotic analysis.

The next corollary makes this statement precise and shows that boundary effects or the two points. 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 1 (Partition function of cycle graphs). Let \( PG_n \) and \( CG_n \) denote a path graph and a cycle graph on \( n \) vertices. Then the partition function of \( CG_n \) is given by

\[
Z_{CG_n}(q) = Z_{PG_n}(q) + \frac{q}{q-1} (Z_{PG_n}(q) - Z_{PG_{n-1}}(q)) - 2
\]

Corollary 1 (Partition function of cycle graphs). Let \( PG_n \) and \( CG_n \) denote a path graph and a cycle graph on \( n \) vertices. Then the partition function of \( CG_n \) is given by

\[
Z_{CG_n}(q) = Z_{PG_n}(q) + \frac{q}{q-1} [Z_{PG_n}(q) - Z_{PG_{n-1}}(q)] - 2
\]

Theorem 4 (Correlations on path graph and bounds via random walk on \( \mathbb{Z} \)). Let \( PG_n \) denote a path on \( n \) vertices and let \( x, y \in [n] \) be two vertices at distance \( d := y - x > 0 \). Then, for any \( q > 0 \), the two-point correlation between \( x \) and \( y \) in \( PG_n \) is given by

\[
U_q^{(PG_n)}(x, y) = 1 - \frac{Z_{PG_{n-d}}(q)}{Z_{PG_n}(q)} - \frac{d [Z_{PG_n}(q) - Z_{PG_{n-1}}(q)] [Z_{PG_{n-y+1}}(q) - Z_{PG_{n-y}}(q)]}{q Z_{PG_n}(q)}
\]

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

\[
\left(1 - \left(\frac{2}{q+2}\right)^m\right)^2 \left(2\mathbb{P}(|S_m| < \frac{1}{2}) - 1\right)^2 \leq U_q^{(PG_n)}(x, y) \leq 1 - \mathbb{P}(|S_m| > d) \left(\frac{2}{q+2}\right)^m
\]

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

From the above statement, due to the diffusive behaviour 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 (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

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

In particular, for \( \delta > 0 \) 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 \( q_n \sim \frac{1}{\zeta_n} \), then the following two limits, distinguishing between the bulk and near the boundaries, are possible:

\[
\lim_{n \to \infty} U_{q_n}^{(SG_n)}(x_n, y_n) = \begin{cases} 
1 - \frac{3}{2} e^{-\frac{\alpha}{\sqrt{n}}} & \text{if} \quad \zeta_n = \omega(\sqrt{n}) \quad \text{and} \quad \zeta_n = n - \omega(\sqrt{n}) \\
1 - \frac{3}{2} e^{-\frac{\alpha}{\sqrt{n}}} & \text{if} \quad \zeta_n = \alpha \sqrt{n} + o(\sqrt{n}) \quad \text{or} \quad \zeta_n = n - \alpha \sqrt{n} + o(\sqrt{n}) \quad \text{for some} \quad \alpha \geq \delta.
\end{cases}
\]

Remark 1. 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 [29] 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 \( q_n = o(n^{-2}) \).

1.4. Detecting modular structures in stylized tree-like geometry. We collect here a number of simple statements of different flavour aiming to illustrate the emergence of giants and what sort of modular structures are with high probability detected in tree-like graphs by tuning \( q \). Let us start in the next theorem by making precise how \( q \) scales on a given large star graph.

**Theorem 5 (Potential and its limit 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 that is adjacent to \( n - 1 \) leaves. We equip \( SG_n \) with a uniform weight function that assigns weight \( w \) to all edges. Denote the center vertex by \( c \) and let \( x, y \) be two distinct leaves. Given \( q > 0 \),

\[
U_q(c, x) = \frac{q(q + (n - 1)w)}{(q + w)(q + nw)}
\]

\[
U_q(x, y) = \frac{q(q^2 + (n + 2)qw + 2(n - 1)w^2)}{(q + w)^2(q + nw)},
\]

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

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

\[
\lim_{n \to \infty} U_{q_n}^{(SG_n)}(c, x) = \begin{cases} 
1 - \frac{q}{q + w} & \alpha > \beta \\
0 & \alpha = \beta \\
1 & \alpha < \beta
\end{cases}
\]

and

\[
\lim_{n \to \infty} U_{q_n}^{(SG_n)}(x, y) = \begin{cases} 
1 - \frac{q(q + 2w^*)}{(q + w^*)^2} & \alpha > \beta \\
0 & \alpha = \beta \\
1 & \alpha < \beta
\end{cases}
\]

We see that the critical phase for the appearance of a giant is when \( \alpha = \beta \) for which the resulting connected subtree is thought of as a star whose center has offspring distribution of parameter \( \bar{q}/(\bar{q} + \bar{w}) \), while a unique giant emerges as soon as \( \alpha > \beta \).

The following statement clarifies how \( q \) should be scaled in a non-homogeneous star to detect an implanted sub-module of leaves more densely connected to the center.
Theorem 6 (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. 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

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

(1.26)

and

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

(1.27)

Figure 1 offers a graphical representation of Theorem 6.

Figure 1. Phase diagram for the community star graph, giving a graphical representation of the results in Theorem 6 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 in Theorem 6.
The next two statements show similar detections on finite growing trees of different flavours.

**Proposition 2** (Asymptotic correlation in undirected trees with a bounded number of vertices). Let \( G = (V,E) \) be an undirected tree and let \( w_n : E \to (0,\infty) \) be a sequence of edge weight functions. For each \( n \in \mathbb{N} \) let \( q_n > 0 \) be the intensity parameter and assume that for each edge \( e \in E \) the limit \( \lim_{n \to \infty} \frac{w_n(e)}{q_n} \) exists in \([0,\infty]\). Let \( x,y \in V \) be two adjacent vertices. Then, as \( n \to \infty \) it holds that

\[
U_{q_n}(x,y) \to \begin{cases} 0 & \text{if } q_n = o(w_n(x,y)) \\ 1 & \text{if } q_n = \omega(w_n(x,y)). \end{cases}
\]  

(1.28)

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 an 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:

1. if \( e,e' \in E_n \) are edges in the same generation of the regular tree, then \( w_n(e) = w_n(e') \);  
2. if \( e_i,e_j \in E_n \) are edges in generations \( i \) and \( j \) with \( i < j \), respectively, then \( w_n(e_i) \leq w_n(e_j) \).

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 neighbour 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 the \( d \)-children, then we call the tree \( d \)-regular. The ancestry of a vertex is the unique path from the vertex to the ancestor (including the vertex itself). A depiction of a regular hierarchical tree is given in Fig. 2.

![Figure 2](image)

**Figure 2.** A 2-regular tree of height 3 with hierarchical edge weights. That is, each edge in generation \( i \) has weight \( w_i \) and these weights satisfy \( w_1 \leq w_2 \leq w_3 \).

**Theorem 7** (Asymptotic detection of layers in a regular hierarchical weighted tree). For each \( n \in \mathbb{N} \) let \( G_n = (V_n,E_n,w_n) \) be a undirected \( d \)-regular 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 two-point correlation between \( x_n \) and \( y_n \) that

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

(1.29)

We conclude by showing with an illustrative example how the analysis on trees presented here and those on complete graphs pursued in [6] can be combined to obtain results on mixed geometrical setups. The resulting regimes are summarized in the phase diagram in Figure 3.
Figure 3. Phase diagram for the bottleneck graph, giving a graphical representation of the results in Theorem 8 with \( q = n^\alpha, \ w = n^\beta \) and \( m = \sqrt{n} \). Regions (ii) and (v) correspond to 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) \( \sqrt{n} \) trees, one tree has \( n+1 \) vertices and consists of the \( n \) vertices in the largest clique with the bridge vertex from the small clique, while the other \( \sqrt{n} - 1 \) trees are isolated vertices.

(iv) \( n + \sqrt{n} - 1 \) trees, one consisting of both bridge vertices, while the others are isolated vertices. This regime corresponds with an anti-community model, where the weight of the bridge outgrows the community structure.

(v) Two trees with \( n \) and \( \sqrt{n} \) vertices, while the bridge edge is absent.

(vi) \( \sqrt{n} + 1 \) trees, one tree contains all \( n \) vertices in the largest clique, while the \( \sqrt{n} \) vertices in the smallest clique are isolated.

(vii) \( n + \sqrt{n} \) isolated vertices.

Theorem 8 (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. 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

\[
Z(q) = q(q(q + n)(q + m) + w(q + 1)(2q + n + m))(q + n)^{n+2}(q + m)^{m+2}.
\] (1.30)

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
while retaining all outgoing edges from a clique containing $b'$. Then as $n \to \infty$ it holds for the two-point correlation between these vertices that

$$
\begin{align*}
U_q(x, x') & \to \\
0 & \text{ if } q = o(\sqrt{|C_1|}) \\
1 & \text{ if } q = o(\sqrt{|C_1|})
\end{align*}
$$

(1.31)

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

(1.32)

$$
\begin{align*}
U_q(b, x) & \to \\
\frac{c}{1 + c} & \text{ if } q = o(1), \ q = o(\sqrt{|C_1|}), \ w = o(m), \ |C_1| = n, \ m \sim cn \text{ with } c \in (0, 1] \\
\frac{1}{1 + c} & \text{ if } q = o(1), \ q = o(\sqrt{|C_1|}), \ w = o(m), \ |C_1| = m, \ m \sim cn \text{ with } c \in (0, 1] \\
1 & \text{ if } q = o(\omega(\frac{\omega}{m}))
\end{align*}
$$

(1.33)

$$
\begin{align*}
U_q(x, y) & \to \\
0 & \text{ if } q = o(1), \ q = o(\frac{\omega}{m}) \\
1 & \text{ if } q = o(1) \text{ or } (q = o(1), \ q = o(\frac{\omega}{m}))
\end{align*}
$$

(1.34)

2. Proofs of results on arbitrary graphs

2.1. Monotone events in terms of number of roots: proof of theorem 1. Let $L$ be the graph Laplacian of $G$. Write $n = |V|$ and let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues $-L$. By [4, proposition 2.1] it holds that

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

(2.1)

so that the derivative of the partition function is given by

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

(2.2)

Note that the conditional probability $P(\Phi_q \in \mathcal{H} \mid r_q = k)$ does not depend on $q$. Also, the probability $P(r_q = k)$ can be written as $\frac{ck^2q^k}{Z(q)}$, where $c_k$ is some constant independent of $q$. Hence, we have that

$$
\frac{d}{dq} P(\Phi_q \in \mathcal{H}) = \frac{d}{dq} \sum_{k=1}^{n} P(\Phi_q \in \mathcal{H} \mid r_q = k)P(r_q = k)
$$

$$
= \sum_{k=1}^{n} P(\Phi_q \in \mathcal{H} \mid r_q = k)c_k \frac{d}{dq} \frac{q^k}{Z(q)} = \sum_{k=1}^{n} P(\Phi_q \in \mathcal{H} \mid r_q = k)c_k \frac{kZ(q)q^{k-1} - q^kZ'(q)}{(Z(q))^2}
$$

$$
= \frac{1}{2} \sum_{k=1}^{n} P(\Phi_q \in \mathcal{H} \mid r_q = k)c_k \frac{kq^k - q^k\mathbb{E}[r_q]}{Z(q)} = \frac{1}{2} \sum_{k=1}^{n} P(\Phi_q \in \mathcal{H} \mid r_q = k)P(r_q = k)(k - \mathbb{E}[r_q])
$$

$$
= \frac{1}{2} P(\Phi_q \in \mathcal{H}) \sum_{k=1}^{n} P(r_q = k \mid \Phi_q \in \mathcal{H})(k - \mathbb{E}[r_q]) = \frac{1}{2} P(\Phi_q \in \mathcal{H})(\mathbb{E}[r_q] \mid \Phi_q \in \mathcal{H}) - \mathbb{E}[r_q]),
$$

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

2.2. Some reduction/extension lemmas.

**Definition 3** (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 $G/e$ 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 $y$.

**Remark 2.** 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 $G/B$ to be the graph obtained by performing directed edge contractions on all edges in $B$. 

Lemma 2 (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 $\Phi_q$ be a random rooted forest of $G$ with rooting parameter $q > 0$ and let $R_q$ be the set of root vertices of $\Phi_q$. Let $L$ denote the graph Laplacian of $G$ and $K = (k_{x,y})_{x,y \in V}$ its Green’s kernel given by $K = (qI - L)^{-1}$. Denote by $[qI - L]_{x,y}$ the submatrix of $qI - L$ obtained by removing the row corresponding to $x$ and the column corresponding to $y$. For each directed edge $e$ write $G/e$ to denote the directed e-contraction of $G$. Then it holds that

$$
P(e \in \Phi_q) = \frac{w(e)}{q} \mathbb{P}(x \in R_q, x \leftrightarrow \Phi_q, y) = w(e)(k_{x,x} - k_{y,x})
$$

$$
= w(e) \frac{\det[qI - L]_{x,x}}{\det[qI - L]} - w(e) \frac{(-1)^{x+y} \det[qI - L]_{x,y}}{\det[qI - L]} = w(e) \frac{Z_{G/e}(q)}{Z_G(q)}.
$$

Proof. Let $e = (x, y)$ be an edge from $x$ to $y$. Let $A = \{F \in \mathcal{F}_G; e \in F\}$ denote the set of rooted forests of $G$ that do contain edge $e$. Write $\mathcal{H} = \{F \in \mathcal{F}_G; x \in R(F), B_F(x) \neq B_F(y)\}$ 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 $f : A \rightarrow \mathcal{H}$ 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 $A$. For the second identity we use the Chebotarev-Shamis matrix-forest theorem [16], which states that $q_k_{x,y} = \mathbb{P}(y \in R_q, B_q(x) = B_q(y))$. The third identity follows from the second, since by Cramer’s rule it holds that $k_{y,x} = (\det [qI - L]_{x,x}$. The fourth identity follows by considering the bijection $g : \mathcal{H} \rightarrow \mathcal{F}_{G/e}$ that sends all edges of a forest in $G$ to their corresponding edges in $G/e$. Note that here $G/e$ 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. [23], tailored to the rooted forest measure. In fact the definition of contraction used in this lemma deviates from the usual definition, as it is adapted to the setting of weighted directed graphs.

Lemma 3 (Spatial Markov property). Let $G = (V, E, w)$ be a weighted directed graph and $A, B \subseteq E$ two disjoint sets of edges. Then it holds for all $F \in \mathcal{F}_G$ with $F \cap A = \emptyset$ and $B \subseteq F$ that

$$
P^{(G)}(\Phi_q = F \mid \Phi_q \cap A = \emptyset, B \subseteq \Phi_q) = p^{(G - A)/B}(\Phi_q = F/B).
$$

Moreover, for any edge $e \in E$ the partition function of $G$ satisfies the deletion-contraction identity

$$
Z_G(q) = Z_{G-e}(q) + w(e)Z_{G/e}(q).
$$

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

$$
P^{(G)}(\Phi_q = F \mid e \notin \Phi_q) = \frac{P^{(G)}(\Phi_q = F)}{\mathbb{P}(\Phi_q \notin F)} = \frac{q^{r(F)}w(F)}{\sum_{H \in A} q^{r(H)}w(H)} = \frac{q^{r(F)}w(F)}{\sum_{H \in \mathcal{F}_{G-e}} q^{r(H)}w(H)} = p^{(G-e)}(\Phi_q = F).
$$

Assume instead that $A = \emptyset$ and $B = \{e\}$ for some edge $e \in E$. Then by Lemma 2 we have for all $F \in \mathcal{F}_G$ with $e \in F$ that

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

The proof of Eq. (2.4) is analogous to that of Eq. (2.3).
Lemma 4 (Graph extension lemma (single vertex version)). Let \( G = (V, E, w) \) be a weighted directed graph and \( x \in V \) a vertex. Let \( \nu^{(G)} \) denote the non-normalised measure on the rooted forests of \( G \) given by \( \nu^{(G)}(\Phi_q \in \cdot) = Z(q) \mathbb{B}^{(G)}(\Phi_q \in \cdot) \). Let \( R_q \) be the set of root vertices of \( \Phi_q \).

Let \( H = G[V \setminus \{ x \}] \) denote the induced subgraph of \( G \) obtained by removing vertex \( x \). Let \( \mathcal{H}(F) \) be the partition of \( \mathcal{F}_G \) given by \( \mathcal{H}(F) = \{ F' \in \mathcal{F}_G : F'[V \setminus \{ x \}] = F \} \), 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

\[
\nu^{(G)}(\Phi_q \in \mathcal{H}(F), \ x \in R_q) = q \nu^{(H)}(\Phi_q = F) \prod_{r \in R(F)} \left( 1 + \frac{w(r, x)}{q} \right)
\]

and that

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

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

Proof of Lemma 4. 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, and that can be obtained from \( \mathcal{H}(F_H) \) 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 \( r_y(F_H) \) to \( x \) as this would create a cycle. \( \Box \)

Definition 4. 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 \( A \subseteq \mathcal{F}_{G[A]} \) such that \( \mathcal{H} = \{ F \in \mathcal{F}_G : F[A] \in A \} \).

Lemma 5 (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 exactly one vertex \( b \in A \) 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 \( A, B \subseteq \mathcal{F} \) be sets of rooted forests of \( G \) that are determined by \( A \) and \( B \), respectively, and let \( A' \subseteq \mathcal{F}_{G[A]} \) and \( B' \subseteq \mathcal{F}_{G[B]} \) be such that \( A = \{ F \in \mathcal{F}_G : F[A] \in A' \} \) and \( B = \{ F \in \mathcal{F}_G : F[B] \in B' \} \).

Denote the non-normalised measure on the rooted forests of \( G \) by \( \nu^{(G)}(\Phi_q \in \cdot) = Z(q) \mathbb{B}^{(G)}(\Phi_q \in \cdot) \) and let \( R_q \) be the set of root vertices of \( \Phi_q \). Then it holds that

\[
\nu^{(G)}(\Phi_q \in A \cap B) = \nu^{(G[A])}(\Phi_q \in A') \nu^{(G[B])}(\Phi_q \in B') + \frac{w(b, b')}{q} \nu^{(G[A])}(\Phi_q \in A', b \in R_q) \nu^{(G[B])}(\Phi_q \in B') + \frac{w(b', b)}{q} \nu^{(G[A])}(\Phi_q \in A') \nu^{(G[B])}(\Phi_q \in B', b' \in R_q).
\]

Proof. Let \( F_A \in A' \) and \( F_B \in B' \) be given.

If both \( b \) 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.

(1) The first of these forests consists of the disjoint union of \( F_A \) and \( F_B \) and has non-normalised measure

\[
\nu^{(G)}(\Phi_q = F_1) = \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B).
\]
The second has an additional edge from \( b \) to \( b' \) and has non-normalised measure
\[
\nu^{(G)}(\Phi_q = F_2) = \frac{w(b, b')}{q} \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B),
\]

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

(3) The third forest has an additional edge from \( b' \) to \( b \) and it similarly has non-normalised measure
\[
\nu^{(G)}(\Phi_q = F_3) = \frac{w(b', b)}{q} \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B).
\]

Note that each of these three forests is contained in \( A \cap B \).

If exactly one of the vertices \( b \) 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 \( b \) and \( b' \) are not roots, then only the first forest without an additional edge is a rooted forest of \( G \).

Since each forest \( n A \cap B \) can be obtained in such a manner, summing over all rooted forests in \( A' \) and \( B' \) yields
\[
\nu^{(G)}(A \cap B) = \sum_{F_A \in A'} \sum_{F_B \in B'} \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B)
\]
\[
+ \frac{w(b, b')}{q} \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B) 1_{\{b \in R(F_A)\}}
\]
\[
+ \frac{w(b', b)}{q} \nu^{(G[A])}(\Phi_q = F_A) \nu^{(G[B])}(\Phi_q = F_B) 1_{\{b' \in R(F_B)\}}
\]
\[
= \nu^{(G[A])}(\Phi_q \in A') \nu^{(G[B])}(\Phi_q \in B')
\]
\[
+ \frac{w(b, b')}{q} \nu^{(G[A])}(\Phi_q \in A', b \in R_u) \nu^{(G[B])}(\Phi_q \in B')
\]
\[
+ \frac{w(b', b)}{q} \nu^{(G[A])}(\Phi_q \in A', b' \in R_u).
\]
\[
\square
\]

3. Proofs of inclusion-exclusion and monotonicity of potential on trees.

3.1. Inclusion-exclusion for 2-point correlation on trees.

Proof of Proposition 1. We will prove the statement by induction on \( d \). First assume that \( d = 1 \). Write \( \mathcal{H} = \{ F \in \mathcal{F}_G : x \not\sim F y \} \) 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 (weakly) connected components \( G_x \) and \( G_y \) containing vertex \( x \) and vertex \( y \), respectively. Denote the non-normalised measure on the rooted forests of \( G \) by \( \nu^{(G)}(\Phi_q \in \cdot) := Z_G(q)Z^{(G)}(\Phi_q \in \cdot) \) and note that \( \nu^{(G)}(\Phi_q = F) = \nu^{(G_x)}(\Phi = F[G_x]) \nu^{(G_y)}(\Phi = F[G_y]) \) for all \( F \in \mathcal{H} \), where \( F[G_x] \) and \( F[G_y] \) 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
\[
U^{(G)}_q(x, y) = \sum_{F \in \mathcal{H}} \nu^{(G)}(\Phi_q = F) = \frac{1}{Z_G(q)} \sum_{F \in \mathcal{H}} \nu^{(G_x)}(\Phi_q = F[G_x]) \nu^{(G_y)}(\Phi_q = F[G_y])
\]
\[
= \frac{1}{Z_G(q)} \sum_{F_x \in \mathcal{F}_{G_x}} \sum_{F_y \in \mathcal{F}_{G_y}} \nu^{(G_x)}(\Phi_q = F_x) \nu^{(G_y)}(\Phi_q = F_y) = \frac{Z_G(q)Z_{G_x}(q)Z_{G_y}(q)}{Z_G(q)}.
\]

Now assume that \( d > 1 \). Let \( z = z_{d-1} \) denote the neighbour 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 3 and the induction hypothesis that

\[
U_q^G(x, y) = U_q^G(x, z) + U_q^G(y, z) - \mathbb{P}_q(x \leftrightarrow \Phi_q, y) - \mathbb{P}_q(x \leftrightarrow \Phi_q, z)
\]

\[
= U_q^G(x, y) - U_q^G(y, z)\mathbb{P}_q(y \leftrightarrow \Phi_q, z)
\]

\[
= U_q^G(x, z) + U_q^G(y, z) - U_q^G(y, z)U_q^G(\Phi_q(y, z))
\]

\[
= \frac{1}{Z_G(q)} \left( \sum_{k=1}^{d-1} (-1)^{k+1} \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+1} Z_{G_i}(q) \right) + \frac{Z_{G_q}(q)Z_{G_z}(q)}{Z_G(q)}
\]

\[
- \frac{Z_{G_q}(q)Z_{G_z}(q)}{Z_G(q)} \frac{1}{Z_{G(d)}(q)} \left( \sum_{k=1}^{d-1} (-1)^{k+1} \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+2} Z_{G_i}(q) \right)
\]

\[
= \frac{1}{Z_G(q)} \left( \sum_{k=1}^{d-1} (-1)^{k+1} \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+1} Z_{G_i}(q) \right) + \frac{Z_{G_q}(q)Z_{G_z}(q)}{Z_G(q)}
\]

\[
+ \frac{1}{Z_G(q)} \left( \sum_{k=1}^{d-1} (-1)^k \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+2} Z_{G_i}(q) \right)
\]

\[
= \frac{1}{Z_G(q)} \left( \sum_{k=1}^{d-1} (-1)^{k+1} \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+1} Z_{G_i}(q) \right) + \frac{1}{Z_G(q)} \left( \sum_{k=0}^{d-1} (-1)^k \sum_{i \in \binom{d}{k-1}} \prod_{i=1}^{k+2} Z_{G_i}(q) \right)
\]

\[
= \frac{1}{Z_G(q)} \left( \sum_{k=1}^{d} (-1)^{k+1} \sum_{i \in \binom{d}{k}} \prod_{i=1}^{k+1} Z_{G_i}(q) \right)
\]

\[
\square
\]

3.2. **Monotonicity for 2-point correlation on trees.** We show below the monotonicity of the 2-point correlation restricted to arbitrary trees. We will start by expressing the 2-point correlation via hitting times, Lemma 6, we then show by means of Theorem 1 monotone of one point rooting events, Lemma 8 below, and after a last bound on the derivatives of hitting time events Lemma 9, we derive the main claim using these three lemmas.

**Lemma 6** (Hitting time expression for two-point correlation between adjacent vertices in trees). Let $G = (V, E, w)$ be a weighted directed tree and $x, y$ be vertices in $V$. Let $\mathbb{P}_q$ 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 $\tau_v$ and $\tau_q$ 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 6.** We will reason using the representation in (1.9) coming from Wilson’s sampling construction. We note in particular that in order for the directed edge $(x, y)$ to be present in $\Phi_q$ is equivalent to require that the loop-erased trajectory in (1.9) 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
\[
P((y,x) \in \Phi_q) = (1 - P((x,y) \in \Phi_q))P_y(\tau_x < \tau_q),
\]
where these two factors correspond to (1.9). Therefore, it follows that
\[
U_q(G)(x,y) = 1 - P(x \leftrightarrow \Phi_q y) = 1 - P((x,y) \in \Phi_q) - P((y,x) \in \Phi_q)
\]
\[
= 1 - \frac{P_x(\tau_y < \tau_q)(1 - P_y(\tau_x < \tau_q))}{1 - P_x(\tau_y < \tau_q)P_y(\tau_x < \tau_q)} - \left(1 - \frac{P_x(\tau_y < \tau_q)(1 - P_y(\tau_x < \tau_q))}{1 - P_x(\tau_y < \tau_q)P_y(\tau_x < \tau_q)}\right)P_y(\tau_x < \tau_q)
\]
\[
= 1 - \frac{P_x(\tau_y < \tau_q) - P_y(\tau_x < \tau_q) + P_x(\tau_y < \tau_q)P_y(\tau_x < \tau_q)}{1 - P_x(\tau_y < \tau_q)P_y(\tau_x < \tau_q)}.
\]

\( \square \)

**Lemma 7** (Bound on derivative of hitting probabilities). Let \( G = (V,E,w) \) be a weighted directed graph and \( x, y \in V \) two vertices. Let \( \tilde{P}_x^{(\cdot)} \) denote the law of the random walk \( X \) on \( G \) starting at \( x \). For each \( v \in V \) let \( \tau_v \) denote the hitting time of \( v \) by \( X \) and let \( \tau_q \) be an independent exponential killing time with rate \( q \). Then it holds for the derivative of the function \( q \mapsto P_x(\tau_y < \tau_q) \) that
\[
\frac{1}{q}P_x(\tau_y < \tau_q) - \frac{1}{q} \leq \frac{d}{dq}P_x(\tau_y < \tau_q) \leq 0.
\]

**Proof of Lemma 7.** The upper bound on the derivative in (3.1) is immediate, we therefore show the lower bound.

Let \( L \) be the graph Laplacian of \( G \) and denote by \( \alpha \) the maximal diagonal entry of \( -L \). For convenience we will work with the discrete time skeleton of \( X \) starting at \( x \), that is, the discrete time random walk \( \tilde{X} \) on \( G \) with transition matrix
\[
P = I + \frac{1}{q}L,
\]
and path measure denoted by \( \tilde{P}_x \). Let \( \tilde{\tau}_y \) denote an independent geometric killing time with success probability \( \frac{1}{1 + \alpha} \). Then it holds that \( P_x(\tau_y < \tau_q) = \tilde{P}_x(\tau_y < \tilde{\tau}_y) \), so that
\[
P_x(\tau_y < \tau_q) = \tilde{P}_x(\tau_y < \tilde{\tau}_y) = \sum_{k=1}^{\infty} \tilde{P}_x(\tau_y < k)\tilde{P}(\tilde{\tau}_y = k) = \sum_{k=1}^{\infty} \tilde{P}_x(\tau_y < k)\frac{k}{q + \alpha} \left(1 - \frac{q}{q + \alpha}\right)^{k-1}.
\]
Since \( \tilde{P}_x(\tau_y < k) \) does not depend on \( q \), it follows that
\[
\frac{d}{dq}P_x(\tau_y < \tau_q) = \sum_{k=1}^{\infty} \tilde{P}_x(\tau_y < k)\frac{(1-k)q + \alpha}{(q + \alpha)^2} \left(\frac{\alpha}{q + \alpha}\right)^{k-1} = \frac{1}{q}P_x(\tau_y < \tau_q) - \sum_{k=1}^{\infty} \frac{P_x(\tau_y < k)kq}{(q + \alpha)^2} \left(\frac{\alpha}{q + \alpha}\right)^{k-1}
\]
\[
\geq \frac{1}{q}P_x(\tau_y < \tau_q) - \sum_{k=1}^{\infty} \frac{P_x(\tau_y < k)q}{(q + \alpha)\alpha} \left(\frac{\alpha}{q + \alpha}\right)^{k-1} = \frac{1}{q}P_x(\tau_y < \tau_q) - \frac{1}{\alpha + \alpha} \tilde{E}(\tilde{\tau}_y) = \frac{1}{q}P_x(\tau_y < \tau_q) - \frac{1}{q}.
\]

\( \square \)

**Lemma 8** (Monotonicity of rooting probabilities). Let \( G = (V,E,w) \) be a weighted directed graph and \( x \in V \) a vertex. Let \( \Phi \) denote the law of a random spanning rooted forest \( \Phi_q \) of \( G \) with rooting parameter \( q > 0 \). Let \( R_q \) denote the set of roots of \( \Phi_q \). Then it holds that
\[
0 \leq \frac{d}{dq}\mathbb{P}(x \in R_q) \leq \frac{1}{q}\mathbb{P}(x \in R_q).
\]

**Proof of Lemma 8 via Lemma 7.** Due to the determinantality of the roots in (1.7), we have that \( x \) is a root in \( \Phi_q \) if
\[
\mathbb{P}(x \in R_q) = K_q(x,x) = q(q - L)^{-1}(x,x) = P_x(X_{\tau_q} = x).
\]
Let $N_x$ denote the set of out-neighbours of $x$ in $G$. Let $\sigma = \inf\{t > 0: X_t \neq X_0\}$ be the first jump time of $X$. Then by the Markov property of $X$ we have that

$$P_x(X_{\tau_q} = x) = P_x(\sigma > \tau_q) + \sum_{v \in N_x} P_x(X_\sigma = v)P_v(\tau_x < \tau_q)P_x(X_{\tau_q} = x).$$

Solving this equation gives us that

$$P_x(X_{\tau_q} = x) = \frac{P_x(\sigma > \tau_q)}{1 - \sum_{v \in N_x} P_x(X_\sigma = v)P_v(\tau_x < \tau_q)} = \frac{q}{q + \sum_{v \in N_x} w(x, v)(1 - P_v(\tau_x < \tau_q))}.$$

It follows by Lemma 7 that

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

For the upper bound it holds that

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

**Remark 3.** A simpler proof of Lemma 8 can be given if we restrict the setting to undirected graphs. In that case the result follows from Theorem 1 and Cauchy’s interlacement theorem, since, as mentioned in the introduction, the expected number of roots is distributed as the sum of independent Bernoulli variables whose parameter depends on the Laplacian spectrum.

**Lemma 9** (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

$$\frac{d}{dq}P(x \in R_q \mid x \leftrightarrow y) \leq \frac{1}{q}P(x \in R_q \mid x \leftrightarrow y). \quad (3.4)$$

**Proof of Lemma 9.** Let $d$ denote the distance between $x$ and $y$. We will argue inductively on $d$. For $d = 0$ the statement follows from Lemma 8.

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 5 that

$$P(x \in R_q \mid x \leftrightarrow y) = \frac{w(z, x)\mu(T_x)(x \in R_q)\mu(T_z)(z \in R_q, z \leftrightarrow y)}{w(z, x)\mu(T_x)(x \in R_q)\mu(T_z)(z \in R_q, z \leftrightarrow y) + w(x, z)\mu(T_z)(z \in R_q, z \leftrightarrow y) + w(z, x)\mu(T_z)(z \in R_q, z \leftrightarrow y)}.$$  

It follows by the induction hypothesis and Lemma 8 that

$$\frac{d}{dq}P(x \in R_q \mid x \leftrightarrow y) = \frac{w(z, x)\mu(T_z)(z \in R_q, z \leftrightarrow y)w(x, z)\mu(T_z)(z \in R_q, z \leftrightarrow y)2\frac{d}{dq}\mu(T_z)(z \in R_q)}{w(z, x)\mu(T_z)(z \in R_q, z \leftrightarrow y)\mu(T_z)(z \in R_q) + w(x, z)\mu(T_z)(z \in R_q, z \leftrightarrow y)\mu(T_z)(z \in R_q) + w(z, x)\mu(T_z)(z \in R_q, z \leftrightarrow y)\mu(T_z)(z \in R_q)} \leq \frac{1}{q}.$$

□
Proof of Theorem 2. 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 6 we have that

$$U_q(x, y) = \frac{1 - P_x(\tau_y < \tau_1) - P_y(\tau_x < \tau_1) + P_x(\tau_y < \tau_1)P_y(\tau_x < \tau_1)}{1 - P_x(\tau_y < \tau_1)P_y(\tau_x < \tau_1)}.$$ 

Taking the derivative then gives us that

$$\frac{d}{dq} U_q(x, y) = \frac{1 - P_x(\tau_y < \tau_1)^2 \frac{d}{dq} P_y(\tau_x < \tau_1) + (1 - P_y(\tau_x < \tau_1))^2 \frac{d}{dq} P_x(\tau_y < \tau_1)}{(1 + P_x(\tau_y < \tau_1)P_y(\tau_x < \tau_1))^2}.$$ 

By the upper bound in Lemma 7 it follows that

$$\frac{d}{dq} U_q(x, y) \geq 0.$$ 

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

$$\frac{d}{dq} \mathbb{P}(x \leftrightarrow y) = \frac{d}{dq} (\mathbb{P}(x \leftrightarrow z \mid z \leftrightarrow y)\mathbb{P}(z \leftrightarrow y) + \mathbb{P}(z \leftrightarrow y)) = \mathbb{P}(x \leftrightarrow z \mid z \leftrightarrow y) \frac{d}{dq} \mathbb{P}(z \leftrightarrow y) + \mathbb{P}(z \leftrightarrow y) \frac{d}{dq} \mathbb{P}(x \leftrightarrow z \mid z \leftrightarrow y).$$

By the induction hypothesis, we have that $\frac{d}{dq} \mathbb{P}(x \leftrightarrow z \mid z \leftrightarrow y) \geq 0$. Hence, it remains to show that $\frac{d}{dq} \mathbb{P}(z \leftrightarrow y) \geq 0$.

Removing the edges between $x$ and $z$ from $G$ splits $G$ into two connected components. Let $T_z$ and $T_x$ denote the component containing vertex $x$ and $z$, respectively. It then holds that

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

Taking the derivative and applying Lemmas 8 and 9 gives us that

$$\frac{d}{dq} \mathbb{P}(x \leftrightarrow z \mid z \leftrightarrow y) = \frac{w(x,z)\nu^{(T_z)}(x \in R_q) + w(z,y)\nu^{(T_z)}(z \in R_q \mid z \leftrightarrow y)}{(q + w(x,z)\nu^{(T_z)}(x \in R_q) + w(z,y)\nu^{(T_z)}(z \in R_q \mid z \leftrightarrow y))^2} - \frac{qw(x,z)\frac{d}{dq} \nu^{(T_z)}(x \in R_q) - qw(z,y)\frac{d}{dq} \nu^{(T_z)}(z \in R_q \mid z \leftrightarrow y)}{(q + w(x,z)\nu^{(T_z)}(x \in R_q) + w(z,y)\nu^{(T_z)}(z \in R_q \mid z \leftrightarrow y))^2} \geq 0.$$ 

\[\square\]

4. INTEGER PARTITIONING : PROOFS

4.1. Partition function on the integers and on a ring.

Proof of Theorem 3. 

Eq. (1.13) Let $b$ be a boundary vertex of $PG_n$. Let $\nu^{(n)}$ denote the non-normalised measure on the rooted forests of $PG_n$ defined as $\nu^{(n)}(\Phi_q \in \cdot) = Z_n(q)\mathbb{P}(\Phi_q \in \cdot)$. By Lemma 5 we have that

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

This gives us that

$$Z_n(q) = \nu^{(n)}(b \notin R_q) + \nu^{(n)}(b \notin R_q) = \nu^{(n-1)}(b \in R_q) + (q + 1)Z_{n-1}(q) = (q + 2)Z_{n-1}(q) - \nu^{(n-1)}(b \notin R_q) = (q + 2)Z_{n-1}(q) - Z_{n-2}(q).$$

(4.2)
We will prove Eq. (1.13) 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 Eq. (1.13) holds. Now assume that \( n > 2 \). Then by Eq. (4.2), the induction hypothesis and repeated applications of Pascal’s formula we have that

\[
Z_n(q) = (2 + q)Z_{n-1}(q) - Z_{n-2}(q) = (2 + q)\sum_{k=1}^{n-1} \left(\frac{n+k-2}{2k-1}\right) q^k - \sum_{k=1}^{n-2} \left(\frac{n+k-3}{2k-1}\right) q^k = \sum_{k=1}^{n} \left(\frac{n+k-1}{2k-1}\right) q^k.
\]

**Eq. (1.14)\rightend** Let \( L \) denote the graph Laplacian of \( PG_n \). Since due to (1.5) the partition function is the characteristic polynomial of \( L \), it can be directly obtained from its spectrum, which is given in [30], from which:

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

**Eq. (1.15)** We have shown above that the partition function satisfies the recurrence relation in Eq. (4.2) . Using the initial conditions \( Z_1(q) = q \) and \( Z_2(q) = q^2 + 2q \), this linear recurrence relation has solution

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

**Eq. (1.16)** To verify that the three expressions we found above do indeed coincide, we can use Chebyshev polynomials of the second kind and find that

\[
Z_n(q) = qU_{n-1}\left(\frac{3}{2}\right) + 1.
\]

We next move to the proof of Corollary 1, for which we will first need to express 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 10** (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 \( b \) a boundary vertex. Let \( \nu^{(n)} \) denote the non-normalised measure on the rooted forests of \( PG_n \), defined as \( \nu^{(n)}(\Phi_q \in \cdot) = Z_n(q)\mathbb{P}^{(n)}(\Phi_q \in \cdot) \). Then

\[
\nu^{(n)}(x \in R_q) = \frac{1}{q} \nu^{(d+1)}(b \in R_q) \nu^{(n-d)}(b \in R_q),
\]

with

\[
\nu^{(n)}(b \in R_q) = Z_n(q) - Z_{n-1}(q).
\]

For the non-normalised measure of the event that both boundary vertices \( b \) and \( b' \) are roots it holds that

\[
\nu^{(n)}(b, b' \in R_q) = qZ_{n-1}(q).
\]

**Proof of Lemma 10.**

**Eq. (4.3)** Let \( L_n \) denote the 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 3 that \( \nu^{(n)}(x \in R_q) = q \det[qI - L_n]_x \), from which Eq. (4.3) follows.

**Eq. (4.4)** Since \( \nu^{(n)}(b \in R_q) = Z_n(q) - \nu^{(n)}(b \notin R_q) \), Eq. (4.4) follows directly from Eq. (4.1).

**Eq. (4.5)** By Lemma 5 we have that

\[
\nu^{(n)}(b, b' \in R_q) = q\nu^{(n-1)}(b \in R_q) + \nu^{(n-1)}(b, b' \in R_q), \quad \text{and} \quad \nu^{(n)}(b \in R_q, b' \notin R_q) = \nu^{(n-1)}(b \in R_q).
\]
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 Eqs. (4.2) and (4.4) that

$$\nu^{(n)}(b, b' \in R_q) = (q + 1)\nu^{(n-1)}(b \in R_q) - \nu^{(n-2)}(b \in R_q)$$

$$= (q + 1)Z_{n-1}(q) - (q + 2)Z_{n-2}(q) + Z_{n-3}(q) = qZ_{n-1}(q).$$

\[\Box\]

**Proof of Corollary 1.** We will first prove Eq. (1.17). 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

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

(4.6)

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 4. This gives us by Eq. (4.2) and Lemma 10 that

$$\nu^{(CG_n)}(x \in R_q) = \sum_{F \in PG_{n-1}} q \nu^{(PG_{n-1})}(\Phi = F) (1 + \frac{1}{q})^{|R(F)|\setminus\{y,z\}}$$

$$= (q + 2 + \frac{1}{q})\nu^{(PG_{n-1})}(y, z \in R_q) + 2(q + 1)\nu^{(PG_{n-1})}(y \in R_q, z \notin R_q) + q\nu^{(PG_{n-1})}(y, z \notin R_q)$$

$$= qZ_{PG_{n-1}}(q) + (2 + \frac{1}{q})\nu^{(PG_{n-1})}(y, z \in R_q) + 2\nu^{(PG_{n-1})}(y \in R_q, z \notin R_q)$$

$$= (q + 2)Z_{PG_{n-1}}(q) - 2Z_{PG_{n-2}}(q) + \frac{1}{q}\nu^{(PG_{n-1})}(y, z \in R_q)$$

$$= Z_{PG_{n}}(q) - Z_{PG_{n-2}}(q) + \frac{1}{q}\nu^{(PG_{n-1})}(y, z \in R_q) = Z_{PG_{n}}(q).$$

Let $r_y(F)$ denote the root in the tree of forest $F$ that contains vertex $y$. Again using Lemma 4 and Eq. (4.2), we obtain

$$\nu^{(CG_n)}(x \notin R_q) = \sum_{F \in PG_{n-1}} \nu^{(PG_{n-1})}(\Phi = F) (2(1 + \frac{1}{q}))^{|\Phi \setminus \{x,y,z\}|}$$

$$= 2\nu^{(PG_{n-1})}(\emptyset \notin R_q) + 2(1 + \frac{1}{q})\nu^{(PG_{n-1})}(z \in R_q, r_y(F) \neq z) = (2 + \frac{2}{q})Z_{PG_{n-1}}(q) - \frac{2}{q}Z_{PG_{n-2}}(q) - 2$$

$$= \frac{2}{q}((q + 2)Z_{PG_{n-1}}(q) - Z_{PG_{n-2}}(q) - Z_{PG_{n-1}}(q)) - 2 = \frac{2}{q}(Z_{PG_n}(q) - Z_{PG_{n-1}}(q)) - 2.$$ This proves Eq. (1.17).

Equation (1.18) follows from Eq. (1.17) and the expression for the path graph partition function given in Eq. (1.13), by repeated applications of Pascal’s formula. \[\Box\]

### 4.2. Correlations on path graphs: bounds and asymptotic analysis

**Proof of Theorem 4.**

**Eq. (1.19)** For each $n \in \mathbb{N}$ let $\mathcal{F}_n$ denote the set of rooted forests of $PG_n$ and write

$$\mathcal{F}^k_{n-d} = \{F \in \mathcal{F}_{n-d}: r(F) = k\};$$

$$\mathcal{R}^k_{n-d}(x) = \{F \in \mathcal{F}_{n-d}: r(F) = k, x \in R(F)\};$$

$$\mathcal{C}^k_n(x, y) = \{F \in \mathcal{F}_n: r(F) = k, x \leftrightarrow F y\}.$$ 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 10.

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 $f : \mathcal{C}^k_n(x, y) \rightarrow \mathcal{F}^k_{n-d} \cup (\mathcal{R}^k_{n-d}(x) \times [d])$ 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 $\tilde{X}$ denote the discrete time random walk on $PG_n$ with law $\tilde{P}$ and geometric killing time $\tilde{\tau}_q$, as defined in Eq. (3.2). Since we consider a path graph we have that $\tilde{\tau}_q \sim \text{Geom}(\frac{1}{\tau_q})$.

We will analyse Wilson’s algorithm with the random walk launched in the first step of the algorithm starting at $x$ an the random walk starting in the second step starting at $y$.

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 $\Phi_q$ 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

$$P(x \leftrightarrow \Phi_q, y) \geq \tilde{P}_x(\tilde{\tau}_q < \tau_z) \tilde{P}_y(\tilde{\tau}_q < \tau_z).$$

(4.7)

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

$$\tau_z \overset{d}{=} \min\{\tau_S(\frac{d}{2}), \tau_S(1 - 2x - \frac{d}{2})\},$$

(4.8)

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

By the reflection principle it holds for all $k, n \in \mathbb{N}$ that $P(\tau_S(n) \leq k) = P(S_k \notin [-n, n-1])$. It follows that

$$P(\tilde{\tau}_q < \tau_X(z)) = \sum_{k=1}^{\infty} P(\tilde{\tau}_q = k) \tilde{P}_z(\tau_z > k) \geq \sum_{k=1}^{m} (1 - P(\tau_S(\frac{d}{2}) < k \text{ or } \tau_S(1 - 2x - \frac{d}{2}) < k)) P(\tilde{\tau}_q = k)

\geq \sum_{k=1}^{m} (1 - 2P(\tau_S(\frac{d}{2}) < k)) P(\tilde{\tau}_q = k) = \sum_{k=1}^{m} (2P(S_{k-1} \in [-\frac{d}{2}, \frac{d}{2} - 1]) - 1) P(\tilde{\tau}_q = k)

\geq \sum_{k=1}^{m} (2P(|S_{k-1}| < \frac{d}{2}) - 1) P(\tilde{\tau}_q = k) \geq \sum_{k=1}^{m} (2P(|S_m| < \frac{d}{2}) - 1) P(\tilde{\tau}_q = k)

= (2P(|S_m| < \frac{d}{2}) - 1) P(\tilde{\tau}_q \leq m) = (2P(|S_m| < \frac{d}{2}) - 1) \left(1 - \left(1 - \frac{q}{\tau_q}\right)^m \right)

$$

Hence

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

If $(2P(|S_m| < \frac{d}{2}) - 1)0$ is non-negative, then we also have that

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

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

$$U^{(m)}_q(x, y) \geq (2P(|S_m| < \frac{d}{2}) - 1)^2 \left(1 - \left(1 - \frac{q}{\tau_q}\right)^m \right)^2,$$

which gives the desired lower bound.

**Upper bound** We again analyse 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 $\Phi_q$. Therefore, we have that

$$P(x \leftrightarrow \Phi_q, y) \geq \tilde{P}_y(\tau_x < \tilde{\tau}_q).$$
Using a coupling argument we can show that
\[
\tau_x \overset{d}{=} \min \{ \tau_d(-d), \tau_d(2n + d - 2y + 1) \},
\]
where \( \tau_x \) 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
\[
\tilde{P}_y(\tau_x < \tilde{\tau}_y) = \sum_{k=1}^{\infty} \tilde{P}_y(\tau_x < k) \tilde{P}(\tilde{\tau}_y = k) \geq \sum_{k=m}^{\infty} \tilde{P}_y(\tau_x \leq k) \tilde{P}(\tilde{\tau}_y = k + 1)
\]
\[
= \sum_{k=m}^{\infty} \tilde{P}(\tau_S(-d) \leq k \text{ or } \tau_S(2n + d - 2y + 1) \leq k) \tilde{P}(\tilde{\tau}_y = k + 1)
\]
\[
\geq \sum_{k=m}^{\infty} \tilde{P}(\tau_S(-d) \leq k) \tilde{P}(\tilde{\tau}_y = k + 1) \geq \sum_{k=m}^{\infty} \tilde{P}(\tau_S(d) \leq a) \tilde{P}(\tilde{\tau}_y = k + 1)
\]
\[
= \tilde{P}(\tau_S(d) \leq m) \tilde{P}(\tilde{\tau}_y > m) = \tilde{P}(S_m \notin [-d, d - 1]) \tilde{P}(\tilde{\tau}_y > m)
\]
\[
\geq \tilde{P}(|S_m| > d) \tilde{P}(\tilde{\tau}_y > m) = \tilde{P}(|S_m| > d) \left( 1 - \frac{q_n}{2d} \right)^m.
\]
It follows that
\[
U^{(n)}(x, y) = 1 - \tilde{P}(x \leftrightarrow y) \leq 1 - \tilde{P}(|S_m| > d) \left( 1 - \frac{q_n}{2d} \right)^m.
\]

**Proof of Corollary 2.**

\( q_n = o\left( \frac{1}{\sqrt{S_m}} \right) \)

Set \( m_n = \lceil \frac{dn}{\sqrt{S_m}} \rceil \), i.e. \( m_n \) is the smallest integer that is not smaller than \( \frac{dn}{\sqrt{S_m}} \). We have that \( m_n = \omega(d_n^2) \).

In particular this means that \( m_n \to \infty \) as \( n \to \infty \). So, \( \frac{S_m}{\sqrt{m_n}} \) converges in distribution to a standard normal random variable. Since \( \frac{d_n}{\sqrt{m_n}} \to 0 \), it follows that \( \tilde{P}\left( \frac{|S_m|}{\sqrt{m_n}} > \frac{d_n}{\sqrt{m_n}} \right) \to 1 \). We also have that \( m_n = o\left( \frac{1}{q_n} \right) \), which gives us that \( \left( 1 - \frac{q_n}{2d_n} \right)^m \to 1 \). Therefore, the upper bound from Theorem 4 gives us that
\[
U^{(n)}(x, y) \leq 1 - \tilde{P}\left( \frac{|S_m|}{\sqrt{m_n}} > \frac{d_n}{\sqrt{m_n}} \right) \left( 1 - \frac{q_n}{2d_n} \right)^m = o(1).
\]

\( q_n = \omega\left( \frac{1}{\sqrt{m_n}} \right) \)

Again set \( m_n = \lceil \frac{dn}{\sqrt{m_n}} \rceil \). It holds that \( m_n = \omega\left( \frac{1}{q_n} \right) \), so that \( \left( 1 - \frac{q_n}{2d_n} \right)^m \to 0 \). Furthermore, we have that \( m_n = o(d_n^2) \). This means that \( \frac{d_n}{\sqrt{m_n}} \to \infty \) and thus that \( \tilde{P}\left( \frac{|S_m|}{\sqrt{m_n}} < \frac{d_n}{\sqrt{m_n}} \right) \to 1 \). For large enough \( n \), this gives us that \( \tilde{P}\left( \frac{|S_m|}{\sqrt{m_n}} < \frac{d_n}{\sqrt{m_n}} \right) \geq \frac{1}{2} \), which means that we can apply the lower bound from Theorem 4. This gives us that
\[
U^{(n)}(x, y) \geq \left( 1 - \left( 1 - \frac{q_n}{2d_n} \right)^m \right)^2 \cdot \left( 2 \tilde{P}\left( \frac{|S_m|}{\sqrt{m_n}} < \frac{d_n}{\sqrt{m_n}} \right) - 1 \right)^2 = 1 - o(1).
\]

\( q_n = \frac{\varepsilon}{d_n^2} + o\left( \frac{1}{d_n^2} \right) \)

Now set \( m_n = \lceil \frac{d_n}{\sqrt{m_n}} \rceil \). 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 \( m_n \sim \frac{1}{d_n^2} \). It follows that there exists a \( \varepsilon > 0 \) small enough that \( \varepsilon < \left( 1 - \frac{q_n}{2d_n} \right)^m < 1 - \varepsilon \) for all \( n \in \mathbb{N} \). We also have that \( m_n \sim \frac{1}{d_n^2} \). This gives us that \( \frac{S_m}{\sqrt{m_n}} \) converges in distribution to a standard normal random variable \( Z \) and that \( \frac{d_n}{\sqrt{m_n}} \to 1 \). Since \( 0.6 < \tilde{P}(|Z| < 1) < 0.7 \), we can apply the lower bound from Theorem 4. Using both bounds from Theorem 4, we conclude the non-degeneracy:
\[
\lim_{n \to \infty} U^{(n)}(x, y) \in (0, 1).
\]
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 4 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 \( \mathbb{P}(\tilde{x}_a = 1)^2 \leq \frac{q_n^2}{(2 + q_n)^2} \).

The probability that the \( x_n \) 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

\[
\frac{q_n^2}{(2 + q_n)^2} \leq U^{(n)}(x_n, y_n) \leq 1 - \left( \frac{1}{2 + q_n} \right)^M \frac{q_n}{2 + q_n}.
\]

Since \( q_n \sim \frac{c}{\sqrt{n}} \) and \( d_n \) is bounded, we have that \( q_n \) is bounded away from 0 and away from infinity. Hence, the two-point correlation is also non-degenerate in this case. \( \square \)

**Proof of Eq. (1.21).** We start the proof with three technical limits. Let \( \alpha \in \mathbb{R} \) be a constant. We claim that

\[
\lim_{n \to \infty} \left( \frac{q_n + 2 + \sqrt{q_n^2 + 4q_n}}{2} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = e^{-\alpha}, \tag{4.9}
\]

\[
\lim_{n \to \infty} \left( \frac{q_n + 2 - \sqrt{q_n^2 + 4q_n}}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = e^\alpha \tag{4.10}
\]

and

\[
\lim_{n \to \infty} \left( \frac{q_n + 2 - \sqrt{q_n^2 + 4q_n}}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^{\omega(\sqrt{\pi})} = 0. \tag{4.11}
\]

Each of the three identities will be proven separately.

**Eq. (4.9)** Since \( \frac{\sqrt{1 + 4d_n^2}}{2d_n} - \frac{1}{2d_n} = o\left(\frac{1}{\sqrt{n}}\right) \), we have that

\[
\left( \frac{q_n + 2 + \sqrt{q_n^2 + 4q_n}}{2} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = \left( 1 + \frac{1}{2d_n^2} + o\left(\frac{1}{\sqrt{n}}\right) \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = e^{-\alpha} + o(1).
\]

**Eq. (4.10)** Note that \( \sqrt{q_n^2 + 4q_n} = \frac{1}{\sqrt{n}}(1 + o(1)) \). Hence,

\[
\left( \frac{q_n + 2 - \sqrt{q_n^2 + 4q_n}}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = \left( 1 - \frac{2}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = \left( 1 - \frac{1}{\delta \sqrt{n}} \right)^{\alpha \sqrt{\pi} + o(\sqrt{\pi})} = e^\alpha + o(1).
\]

**Eq. (4.11)** Similar to Eq. (4.10) it holds that

\[
\left( \frac{q_n + 2 - \sqrt{q_n^2 + 4q_n}}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^{\omega(\sqrt{\pi})} = \left( 1 - \frac{1}{\delta \sqrt{n}} \right)^{\omega(\sqrt{\pi})} \to 0.
\]

This concludes the proof of the claim.
Now that we have established these identities we continue with the proof. For brevity write $Z_n(q) = Z_{PG_n}(q)$. Using the expression for the partition function given in Eq. (1.15) we have for each $m \in \mathbb{N}$ that

$$Z_m(q_n) = \frac{1}{\sqrt{1+4q_n^2}} \left( 1 - \left( \frac{q_n + 2 - \sqrt{q_n^2 + 4q_n}}{q_n + 2 + \sqrt{q_n^2 + 4q_n}} \right)^m \right) \left( \frac{q_n + 2 + \sqrt{q_n^2 + 4q_n}}{2} \right)^m. \quad (4.12)$$

By Eq. (1.19) the two-point correlation is given by

$$U_{q_n}^{(PG_n)}(x_n, y_n) = 1 - \frac{Z_{n-d_n}(q_n)}{Z_n(q_n)} - d_n Z_n(q_n) - Z_{n-1}(q_n) [Z_{n-y_n + 1}(q_n) - Z_{n-y_n}(q_n)]. \quad (4.13)$$

The result follows by plugging in Eq. (4.12) into Eq. (4.13) and repeatedly applying the limits in (4.9)–(4.11). □

5. Asymptotic detection of modular structures

5.1. Star graphs: homogeneous case and with implanted communities.

Proof of theorem 5. Let us start by providing an expression for the partition function of a regular tree with homogeneous weights. Let $w \in (0, \infty)$, $h \in \mathbb{N}$, $k \in \mathbb{N}$, and let $L$ be the graph Laplacian of the $k$-regular tree with height $h$ and uniform weight $w$. Define $(\alpha_n)_{n \in \mathbb{N}}$ 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

$$\det[qI - L] = \left( \prod_{i=0}^{k-1} \alpha_i^{h-1} \right) \left( q + kw - \frac{kw^2}{\alpha_{h-1}} \right). \quad (5.1)$$

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 $D^{-1} = \frac{1}{q + w}I$. 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 (5.1).

We’ll continue by checking the validity of the expressions in (1.22) and (1.23). By applying (5.1) to the homogeneous star graph $SG_n$ we obtain that its partition function is given by

$$Z_{SG_n}(q) = q(q + w)^{n-2}(q + nw). \quad (5.2)$$

Since $d(c, x) = 1$, by (1.11) 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 $d(c, y) = 2$, by Proposition 1 we have that

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

Which finishes the proof of (1.22) and (1.23). Then the asymptotics in (1.24) and (1.25) follow immediately. □
Lemma 11. Playing with degrees and hierarchical weights on trees.

5.2. From these explicit formulas, letting $\tau_x$ denote the first hitting time of $x$ with rooting parameter $\Phi_q$, as defined in Eq. (3.2). Let $T_x$ denote the first hitting time of $x$ by $\tilde{X}$. Let $c = (x, y)$

Proof of Lemma 11. The Laplacian $L$ of the community star graph $\text{CSG}_{n,k}$ is given by

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

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

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

which can be found by applying Schur’s determinant identity as we did in the proof of Theorem 5.

The eigenvalues of the graph Laplacian are the zeros of the characteristic polynomial:

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

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 $V_w$, respectively.

Combining Proposition 1 with Eq. (5.3) leads to:

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

and

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

From these explicit formulas, letting $q$ and $w$ be as in the statement, the limits in Theorem 6 follow. \hfill $\square$

5.2. Playing with degrees and hierarchical weights on trees.

Lemma 11 (Upper bound on edge probability). Let $G = (V, E, w)$ be a weighted directed graph. Let $\Phi_q$ be a random rooted forest of $G$ with rooting parameter $q > 0$. Then for each edge $e \in E$ it holds that $P(e \in \Phi_q) \leq \frac{w(e)}{q + w(e)}$.

Proof of Lemma 11. We will analyse Wilson’s algorithm. Let $\tilde{X}$ denote the discrete time random walk on $G$ with law $\tilde{P}$ and geometric killing time $\tau_q$, as defined in Eq. (3.2). Let $\tau_x$ denote the first hitting time of $x$ by $\tilde{X}$. Let $c = (x, y)$
be an edge and let $\sigma_e$ denote the first time that $X$ crosses edge $e$. We then have that
\[
\tilde{P}_x(\sigma_e < \tilde{\tau}_q) = \tilde{P}_x(X_1 = y) + \sum_{v \in V - y} \tilde{P}_x(X_1 = v)\tilde{P}_v(\tau_x < \tilde{\tau}_q)\tilde{P}_x(\sigma_e < \tilde{\tau}_q).
\]
It follows that
\[
\tilde{P}_x(\sigma_e < \tilde{\tau}_q) \leq \tilde{P}_x(X_1 = y) + \tilde{P}_x(X_1 = x)\tilde{P}_x(\sigma_e < \tilde{\tau}_q).
\]
Note that for edge $e$ to be an edge of $\Phi_n$, the random walk $X$ starting at $X$ has to cross $e$ before being killed. Write $w(x)$ to denote the total weight of all outgoing edges of $x$. It then holds that
\[
P(e \in \Phi_n) \leq \tilde{P}_x(\sigma_e < \tilde{\tau}_q) \leq \frac{\tilde{P}_x(X_1 = y)}{1 - \tilde{P}_x(X_1 = x)} = \frac{w(e)}{q + w(x)} \leq \frac{w(e)}{q + w(e)}.
\]

Proof of Proposition 2. If $q_n = \omega(w_n(x, y))$, then by Lemma 11 we have that $U_{q_n}(x, y) \to 1$.

Assume that $q_n = o(w_n(x, y)).$ Let $X$ denote the discrete time random walk on $PG_n$ with For each $n \in \mathbb{N}$ let $\tilde{P}_x(n)$ denote the law of the discrete time random walk $X$ on $G_n$ starting at vertex $x \in V_n$ and let $\tilde{\tau}_q$ be a geometric killing time, as defined in Eq. (3.2). Let $\tau_x$ denote the first hitting time of $x$ by $X$.

Let $K$ denote the number of vertices on the $x$-side of edge $(x, y)$ in $G$. We will show by induction on $K$ that
\[
\tilde{P}_x(n)(\tau_y < \tilde{\tau}_q) = 1 - \Theta\left(\frac{q_n}{w_n(x, y)}\right).
\]
If $K = 1$, then $x$ is a leaf in $G$. It follows that
\[
\tilde{P}_x(n)(\tau_y < \tilde{\tau}_q) = \frac{w_n(x, y)}{q_n + w_n(x, y)} \sim 1 - \frac{q_n}{w_n(x, y)}.
\]
Assume that $K \geq 2$. Let $N_x$ denote the set of neighbours of $x$ in $G$. Since the limit $\lim_{n \to \infty} \frac{w_n(x, y)}{q_n}$ exists for all edges incident to $x$, we can partition $N_x \setminus \{y\}$ into two parts: the first part $N_x^I = \{v \in N_x \setminus \{y\} : w_n(x, v) = O(q_n)\}$ consists of all neighbours for which the weight of the edge between $x$ and that neighbour has no larger order than $q_n$; the second part $N_x^H = \{v \in N_x \setminus \{y\} : w_n(x, v) = \omega(q_n)\}$ consists of the remaining neighbours.

Then for each $v \in N_x^I$ we have that $q_n = o(w_n(x, v))$. For each such $v$ it follows by the induction hypothesis that
\[
\tilde{P}_x(n)(\tau_y < \tilde{\tau}_q) = 1 - \Theta\left(\frac{q_n}{w_n(x, v)}\right).
\]
It follows that
\[
\tilde{P}_x(n)(\tau_y < \tilde{\tau}_q) = \frac{w_n(x, y)}{q_n + w_n(x, y)} = 1 - \Theta\left(\frac{q_n}{w_n(x, y)}\right).
\]

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

Lemma 12 (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 Fig. 2. Denote the weight of an edge at height $i \in [H]$ in $G_n$ by $w_i(n)$ and recall that $w_1(n) \leq \cdots \leq w_H(n)$.

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 sequence of rooting parameters such that $q_n = o\left(\frac{w_n(x, y)}{\ell_n(1)}\right)$.
For each \( n \in \mathbb{N} \) let \( \tilde{\mathbb{P}}^{(n)} \) 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 Eq. (3.2). Let \( \tau_x \) denote the first hitting time of \( x \) by \( \tilde{X} \). Then as \( n \to \infty \) it holds that

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

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

\[
\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{w_H^{(n)}}{w_H^{(n)} + q_n} \sim 1 - \frac{q_n \ell_n(y)}{w_H^{(n)}}.
\]

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 \( \frac{w_h^{(n)}}{\tau_n(y)} \leq \frac{w_h^{(n)}}{\tau_n(z)} \). This means that \( q_n = o\left(\frac{w_h^{(n)}}{\tau_n(y)}\right) \). Thus by the induction hypothesis we then have that

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

Since this holds for all possible choices of sequences of children of \( y_n \), Lemma 13 stated at the end of this section gives us that

\[
\sum_{z \in C_y^{(n)}} \tilde{\mathbb{P}}^{(n)}(\tau_y < \tilde{\tau}_q) \sim \sum_{z \in C_y^{(n)}} 1 - \frac{q_n \ell_n(z)}{w_{h+1}^{(n)}}. \tag{5.4}
\]

Note that for all \( n \) it holds that

\[
\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q) = \tilde{\mathbb{P}}^{(n)}(X_1 = x) + \tilde{\mathbb{P}}^{(n)}(X_1 = y)\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q) + \sum_{z \in C_y^{(n)}} \tilde{\mathbb{P}}^{(n)}(X_1 = z)\tilde{\mathbb{P}}^{(n)}(\tau_y < \tilde{\tau}_q)\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q).
\]

Solving this equation gives us that

\[
\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{\tilde{\mathbb{P}}^{(n)}(X_1 = x)}{1 - \tilde{\mathbb{P}}^{(n)}(X_1 = y)} - \frac{\tilde{\mathbb{P}}^{(n)}(X_1 = y)}{1 - \tilde{\mathbb{P}}^{(n)}(X_1 = y)}\sum_{z \in C_y^{(n)}} \tilde{\mathbb{P}}^{(n)}(\tau_y < \tilde{\tau}_q)\tilde{\mathbb{P}}^{(n)}(\tau_x < \tilde{\tau}_q) = \frac{w_h^{(n)}}{q_n + w_h^{(n)} + w_{h+1}^{(n)}} \sum_{z \in C_y^{(n)}} \left(1 - \tilde{\mathbb{P}}^{(n)}(\tau_y < \tilde{\tau}_q)\right). \tag{5.5}
\]

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

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

\( \square \)

Proof of Theorem 7. If \( d_n \) is bounded, then the result follows from Proposition 2. Hence, we can assume that \( d_n \to \infty \) as \( n \to \infty \). Since \( G_n \) is a regular 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 \( n \to \infty \). Hence, the case \( q_n = o\left(\frac{w_h^{(n)}}{\tau_n(y)}\right) \) follows directly from Lemmas 6 and 12.

Assume that \( q_n = \omega\left(\frac{w_h^{(n)}}{d_n^k}\right) \). By Theorem 2 we can assume without loss of generality that also \( q_n = o\left(\frac{w_h^{(n)}}{d_n^k}\right) \).
For each $n \in \mathbb{N}$, let $\tilde{P}^{(n)}_y$ denote the law of the discrete time random walk $\tilde{X}$ on $G_n$ starting at vertex $x \in V_n$ and let $\tilde{\tau}_y$ be a geometric killing time, as defined in Eq. (3.2). Let $\tau_y$ denote the first hitting time of $y$ by $\tilde{X}$. By Lemma 6 it is sufficient to show that both $\tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) \to 0$ and $\tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) \to 0$ as $n \to \infty$.

First we consider $\tilde{P}^{(n)}_y(\tau_x < \tilde{\tau}_y)$. Let $z_k$ be a child vertex of $y_k$. Then by Lemma 12 we have that

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

So, by using that $G_n$ is a regular tree, we have analogous to Eq. (5.5) that

$$\tilde{P}^{(n)}_y(\tau_x < \tilde{\tau}_y) = \frac{w_n(e_n)}{q_n + w_n(e_n) + d_n(1 - \tilde{P}^{(n)}_x(\tau_y < \tilde{\tau}_y)) w_n(y_n, z_n)}$$

$$\sim \frac{w_n(e_n)}{q_n + w_n(e_n) + d_n q_n \sum_{i=0}^{m-1} d_n^i} = \frac{w_n(e_n)}{q_n + w_n(e_n) + \omega(w_n(e_n))} = o(1).$$

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

$$\tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) = \tilde{P}^{(n)}_y(\tilde{X}_1 = y) + \tilde{P}^{(n)}_y(\tilde{X}_1 = x) = \tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) + \tilde{P}^{(n)}_y(\tilde{X}_1 = u) \tilde{P}^{(n)}(\tau_x < \tilde{\tau}_y) \tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) + (d_n - 1) \tilde{P}^{(n)}_y(\tilde{X}_1 = y) \tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) \tilde{P}^{(n)}(\tau_y < \tilde{\tau}_y).$$

This gives us that

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

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

$$\tilde{P}^{(n)}_y(\tau_y < \tilde{\tau}_y) \leq \frac{\tilde{P}^{(n)}_y(\tilde{X}_1 = y)}{w_n(e_n)} + w_n(e_n)(d_n - 1) \left(1 - \tilde{P}^{(n)}_y(\tau_x < \tilde{\tau}_y) \right) \leq \frac{1}{1 + (d_n - 1)(1 - o(1))} = o(1).$$

The simple lemma below has been used to show Eq. (5.4).

**Lemma 13.** For each $n \in \mathbb{N}$ let $\alpha_i^{(n)} \in \{\ell_i\}$ be real valued sequences of length $\ell_n$. Let $\mathcal{F} = \{f \in \mathbb{N}^\mathbb{N} : f(n) \in \{\ell_i\} \text{ for all } n \in \mathbb{N}\}$ 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 $\alpha_f^{(n)} \sim \beta_f^{(n)}$ 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

$$\left| \frac{\alpha_f^{(n)}}{\beta_f^{(n)}} - 1 \right| < \varepsilon.$$ 

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

$$f^*(n) = \arg\max_{i \in [\ell_n]} \left| \alpha_i^{(n)} - \beta_i^{(n)} \right|.$$
Then for all $\varepsilon > 0$ and all $n \geq N(\varepsilon, f^*)$ it holds that

$$\left| \frac{\sum_{i=1}^{\ell_n} \alpha_i^{(n)}}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} - 1 \right| \leq \frac{1}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \left( \sum_{i=1}^{\ell_n} \alpha_i^{(n)} - \sum_{i=1}^{\ell_n} \beta_i^{(n)} \right) \leq \frac{\sum_{i=1}^{\ell_n} \alpha_i^{(n)} - \sum_{i=1}^{\ell_n} \beta_i^{(n)}}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \leq \frac{1}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \left( \sum_{i=1}^{\ell_n} \alpha_i^{(n)} - 1 \right) \leq \frac{1}{\sum_{i=1}^{\ell_n} \beta_i^{(n)}} \left( \sum_{i=1}^{\ell_n} \alpha_i^{(n)} - 1 \right) \leq \varepsilon.$$  

\[ \square \]

5.3. A two communities bottleneck graph.

**Proof of Theorem 8.**

**Equation (1.30)** For a graph $G$ let $\nu^{(G)}$ denote the non-normalised rooted forest measure on $G$ given by $\nu^{(G)}(\Phi_q \in \cdot) = Z_G(q)E^{(G)}(\Phi_q \in \cdot)$. 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-normalised measure of rooting events in the complete graphs $K_n$ and $K_m$.

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

$$Z_{K_n}(q) = q(q + n)^{n-1}. \quad (5.6)$$

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

$$\nu^{(K_n)}(U \subseteq R_n) = q^r \det[qI - L_n]_U$$

$$= q^r \det[(q + r)I - L_{n-r}]$$

$$= q^r Z_{K_{n-r}}(q + r)$$

$$= q^r (q + r)(q + n)^{n-r-1}. \quad (5.7)$$

For the partition function of $BG_{n,m}$, Lemma 5 gives us that

$$Z_{BG_{n,m}}(q) = Z_{K_n}(q)Z_{K_m}(q) + \frac{w}{q} Z_{K_n}(q)\nu^{(K_m)}(b' \in R_q) + \frac{w}{q} Z_{K_m}(q)\nu^{(K_n)}(b \in R_q)$$

$$= q(q + n)(q + m) + w(q + 1)(2q + n + m)(q + n)^{n-2}(q + m)^{m-2}.$$  

**Equation (1.32)** We can express $U_q(b, b')$ explicitly by using Proposition 1 and Eqs. (1.30) and (5.6)

$$U_q(b, b') = \frac{Z_{K_n}(q)Z_{K_m}(q)}{Z_{BG_{n,m}}(q)} \quad (5.8)$$

$$= \frac{q(q + n)(q + m)}{q(q + n)(q + m) + w(q + 1)(2q + n + m)}. \quad (5.9)$$

The result of Eq. (1.32) follows directly from this expression.

**Equation (1.34)** 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 5 we have that

$$\nu^{(BG_{n,m})}(x \leftrightarrow_{\Phi_q} x') = \nu^{(K_n)}(x \leftrightarrow_{\Phi_q} x')Z_{K_m}(q) + \frac{w}{q} \nu^{(K_n)}(x \leftrightarrow_{\Phi_q} x', b \in R_q)Z_{K_m}(q) + \frac{w}{q} \nu^{(K_m)}(x \leftrightarrow_{\Phi_q} x')\nu^{(K_n)}(b' \in R_q).$$
By Eqs. (5.6) and (5.7) it follows that
\[
U_q^{(BG_{n,m})}(x, x') = 1 - \frac{\nu^{(BG_{n,m})}(x \leftrightarrow q, x')}{Z_{BG_{n,m}}(q)}
= 1 - \frac{\nu^{(K_n)}(x \leftrightarrow q, x') Z_{K_m}(q) + w \nu^{(K_n)}(x \leftrightarrow q, x', b \in R_q) Z_{K_m}(q) + \frac{w(q+1)}{q(q+m)} \nu^{(K_n)}(x \leftrightarrow q, x') Z_{K_m}(q)}{Z_{BG_{n,m}}(q)}
= 1 - U_q^{(BG_{n,m})}(b, b') \left( 1 + \frac{w(q+1)}{q(q+m)} \right) \mathbb{P}_{q}(x \leftrightarrow q, x') + \frac{w(q+1)}{q(q+m)} \mathbb{P}_{q}(x \leftrightarrow q, x' \mid b \in R_q).
\tag{5.10}
\]

Let \( H \) denote the graph obtained by removing all outgoing edges of \( b \) from \( K_n \), while retaining the in-going edges. By Lemma 3 it then holds that \( \mathbb{P}_{q}(x \leftrightarrow q, x' \mid b \in R_q) = \mathbb{P}(H)(x \leftrightarrow q, x') \). Let \( \mathbb{P}_x \) denote the law of the random walk on \( H \) starting at \( x \) and \( \tau_q \) an independent exponential killing time with rate \( q \). Since the hitting time \( \tau_q \) 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
\[
\mathbb{P}_{q}(x \leftrightarrow q, x' \mid b \in R_q) = \mathbb{P}(H)(x \leftrightarrow q, x')
= \mathbb{P}_{q}(K_{n-1})(x \leftrightarrow q, x') + \mathbb{P}_{q}(K_{n-1})(x \leftrightarrow q, x', z) \mathbb{P}_{x}(\tau_q < \tau_q) \mathbb{P}_{x'}(\tau_q < \tau_q)
= \mathbb{P}_{q}(K_{n-1})(x \leftrightarrow q, x') + \frac{1}{(q+1)^2} \mathbb{P}_{q}(K_{n-1})(x \leftrightarrow q, x')
= \frac{1}{(q+1)^2} + \frac{q(q+2)}{(q+1)^2} \mathbb{P}_{q}(K_{n-1})(x \leftrightarrow q, x').
\tag{5.11}
\]

By [6, Theorem 1] we have that
\[
\mathbb{P}_{q}(K_{n})(x \leftrightarrow q, x') \rightarrow \begin{cases} 1 & \text{if } q = o(\sqrt{n}) \\ 0 & \text{if } q = \omega(\sqrt{n}). \end{cases}
\tag{5.12}
\]

which together with Eq. (5.11) gives us that
\[
\mathbb{P}_{q}(K_{n})(x \leftrightarrow q, x' \mid b \in R_q) \rightarrow \begin{cases} 1 & \text{if } q = o(\sqrt{n}) \\ 0 & \text{if } q = \omega(\sqrt{n}). \end{cases}
\]

Assume that \( q = o(\sqrt{n}) \). Fix a small enough \( \varepsilon > 0 \). Then for \( n \) large enough it holds that \( \mathbb{P}_{q}(K_{n})(x \leftrightarrow x') > 1 - \varepsilon \) and that \( \mathbb{P}_{q}(K_{n})(x \leftrightarrow x' \mid b \in R_q) > 1 - \varepsilon \). By Eqs. (5.9) and (5.10), this means that for \( n \) large enough
\[
U_q^{(BG_{n,m})}(x, x') < 1 - U_q^{(BG_{n,m})}(b, b') \left( 1 + \frac{w(q+1)}{q(q+m)} \right) (1 - \varepsilon) + \frac{w(q+1)}{q(q+m)} (1 - \varepsilon) = \varepsilon.
\tag{5.13}
\]

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 (1.33) Assume that \( x, x' \) and \( b \) belong to the clique of size \( n \). By again considering the random walk on \( H \), we find that
\[
\mathbb{P}_{q}(x \leftrightarrow b \mid b \in R_q) = \mathbb{P}_{x}(\tau_q < \tau_q) = \frac{1}{q+1}.
\]

So, since \( \frac{1}{q+1} \to 0 \) for \( q = \omega(\sqrt{n}) \), the case \( q = \omega(\sqrt{n}) \) follows analogous to Eq. (5.13).
Now assume that $q = o(\sqrt{n})$. Then we have that $\mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} b) \to 1$, so that

$$
U_q^{(BG_{n,m})}(x, b) = 1 - U_q^{(BG_{n,m})}(b, b') \left( \left( 1 + \frac{w(q+1)}{q(q+m)} \right) \mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} b) + \frac{w(q+1)}{q(q+m)} \frac{1}{r^{q+1}} \right)
$$

$$
\sim 1 - U_q^{(BG_{n,m})}(b, b') \left( \left( 1 + \frac{w(q+1)}{q(q+m)} \right) + \frac{w(q+1)}{q(q+m)} \frac{1}{r^{q+1}} \right)
$$

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

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

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

Performing the same computation for $U_q(y, b')$ yields the result of Eq. (1.33).

**Equation (1.34)** By Lemma 2 and Eqs. (5.8) and (5.11) it holds that

$$
U_q^{(BG_{n,m})}(x, y) = 1 - \frac{\nu^{(BG_{n,m})}(x \leftrightarrow_{\Phi_q} y, (b, b') \in \Phi_g)}{Z_{BG_{n,m}}(q)} - \frac{\nu^{(BG_{n,m})}(x \leftrightarrow_{\Phi_q} y, (b', b) \in \Phi_g)}{Z_{BG_{n,m}}(q)}
$$

$$
= 1 - \frac{w}{q} \mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} b, b \in R_q) \mathbb{P}^{(K_m)}(b' \leftrightarrow_{\Phi_q} y) - \frac{w}{q} \mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} b) \mathbb{P}^{(K_m)}(b' \leftrightarrow_{\Phi_q} y, b' \in R_q)
$$

$$
+ \frac{w}{q} \mathbb{P}^{(K_n)}(x \leftrightarrow_{\Phi_q} b) \mathbb{P}^{(K_m)}(b' \leftrightarrow_{\Phi_q} y, b' \in R_q)
$$

$$
= 1 - \frac{w}{q} \mathbb{P}^{(K_n)}(b' \leftrightarrow_{\Phi_q} y) + \frac{w}{q} \mathbb{P}^{(K_m)}(x \leftrightarrow_{\Phi_q} b)
$$

$$
= 1 - \frac{w(q+m)}{q(q+n)(q+m) + w(q+1)(2q + n + m)}
$$

from which the limits in Eq. (1.34) follow.

**Acknowledgments**

L. Avena was supported by NWO Gravitation Grant 024.002.003-NETWORKS. Parts of this work were originally initiated in the Bachelor and Master theses of J.E.P. Driesen [17] and V.T. Koperberg [29], respectively.

**References**

[1] L. Avena, F. Castell, A. Gaudillière, and C. Mélot. “Random Forests and Networks Analysis”. In: Journal of Statistical Physics 173 (2018), 985–1027.

[2] L. Avena, F. Castell, A. Gaudillière, and C. Mélot. “Approximate and exact solutions of intertwining equations through random spanning forests,” in: In and Out of Equilibrium 3. Celebrating Vladas Sidoravicius. (to appear). Birkhäuser Basel, 2021. arXiv: 1702.05992.

[3] L. Avena, F. Castell, A. Gaudillière, and C. Mélot. “Intertwining wavelets or multiresolution analysis on graphs through random forests”. In: Applied and Computational Harmonic Analysis 48.3 (2020), pp. 949–992.

[4] L. Avena and A. Gaudillière. “Two Applications of Random Spanning Forests”. In: Journal of Theoretical Probability 31.4 (2018), pp. 1975–2004.
[5] L. Avena and A. Gaudillière. “A proof of the transfer-current theorem in absence of reversibility”. In: Statistics & Probability Letters 142 (2018), pp. 17–22.

[6] L. Avena, A. Gaudillière, P. Milanesi, and M. Quattropani. Loop-erased partitioning of a graph: mean-field analysis. 2019. arXiv: 1906.03858.

[7] K Avrachenkov, P Chebotarev, and A Mishenin. “Semi-supervised learning with regularized Laplacian”. In: Optimization methods & software 32.2 (2017), pp. 222–236.

[8] S. Barthelmé, N. Tremblay, A. Gaudillière, L. Avena, and P.-O. Amblard. “Estimating the inverse trace using random forests on graphs”. In: XXVIIème colloque GRETSI. 2019. arXiv: 1811.11685.

[9] R. Bauerschmidt, N. Crawford, T. Helmuth, and A. Swan. “Random Spanning Forests and Hyperbolic Symmetry”. In: Communications in mathematical physics 381.3 (2021), pp. 1223–1261.

[10] A. Bedini, S. Caracciolo, and A. Sportiello. “Phase transition in the spanning-hyperforest model on complete hypergraphs”. In: Nuclear physics. B 822.3 (2009), pp. 493–516.

[11] I. Benjamini, H. Kesten, Y. Peres, and O. Schramm. “Geometry of the uniform spanning forest: Transitions in dimensions 4, 8, 12”. In: Annals of mathematics 160.2 (2004), pp. 465–491.

[12] I. Benjamini, R. Lyons, Y. Peres, and O. Schramm. “Uniform Spanning Forests”. In: The Annals of probability 29.1 (2001), pp. 1–65.

[13] D. F. de Bernardini and S. Popov. “Russo’s Formula for Random Interlacements”. In: Journal of statistical physics 160.2 (2015), pp. 321–335.

[14] R. Burton and R. Pemantle. “Local Characteristics, Entropy and Limit Theorems for Spanning Trees and Domino Tilings Via Transfer-Impedances”. In: The Annals of Probability 21.3 (1993), pp. 1329–1371.

[15] P. Chebotarev. “Spanning forests and the golden ratio”. In: Discrete Applied Mathematics 156.5 (2008), pp. 813–821.

[16] P. Chebotarev and E. Shamis. “The Matrix-Forest Theorem and Measuring Relations in Small Social Groups”. In: Automation and Remote Control 58.9 (1997), pp. 1505–1514.

[17] J. Driessen. “Loop-Erased Partitions on Tree Structures”. Leiden University, Bachelor’s thesis. 2019.

[18] G. R Grimmett and S. N Winkler. “Negative association in uniform forests and connected graphs”. In: Random structures & algorithms 24.4 (2004), pp. 444–460.

[19] G. Grimmett. Percolation. Berlin / Heidelberg: Springer, 1999.

[20] G. Grimmett. The Random-Cluster Model. Berlin / Heidelberg: Springer, 2006.

[21] T. Hutchcroft. “Interlacements and the wired uniform spanning forest”. In: The Annals of probability 46.2 (2018), p. 1170.

[22] T. Hutchcroft and A. Nachmias. “Indistinguishability of trees in uniform spanning forests”. In: Probability theory and related fields 168.1-2 (2017), pp. 113–152.

[23] T. Hutchcroft and A. Nachmias. “Uniform Spanning Forests of Planar Graphs”. In: Forum of mathematics. Sigma 7 (2019).

[24] B. D. Jones, B. G. Pittel, and J. S. Verducci. “Tree and Forest Weights and Their Application to Nonuniform Random Graphs”. In: The Annals of applied probability 9.1 (1999), pp. 197–215.
REFERENCES

[25] A. A. Járai, F. Redig, and E. Saada. “Approaching Criticality via the Zero Dissipation Limit in the Abelian Avalanche Model”. In: Journal of statistical physics 159.6 (2015), pp. 1369–1407.

[26] J Kahn and M Neiman. “Negative correlation and log-concavity”. In: Random structures & algorithms 37.3 (2010), pp. 367–388.

[27] R. Kenyon. “Spanning forests and the vector bundle Laplacian”. In: The Annals of probability 39.5 (2011), pp. 1983–2017.

[28] R. Kenyon. “Determinantal spanning forests on planar graphs”. In: The Annals of probability 47.2 (2019).

[29] T. Koperberg. “Loop-erased partitioning of sparse graphs”. Leiden University, Master’s thesis. 2020.

[30] P. van Mieghem. Graph Spectra for Complex Networks. Cambridge University Press, 2010.

[31] R. Pemantle. “Choosing a Spanning Tree for the Integer Lattice Uniformly”. In: The Annals of Probability 19.4 (1991), pp. 1559–1574.

[32] R. Pemantle. “Towards a theory of negative dependence”. In: Journal of mathematical physics 41.3 (2000), pp. 1371–1390.

[33] Y. Y. Pilavci, P.-O. Amblard, S. Barthelmé, and N. Tremblay. Smoothing graph signals via random spanning forests. 2020. arXiv: 1910.07963.

[34] Y. Pilavci, P.-O. Amblard, S. Barthelme, and N. Tremblay. Graph Tikhonov Regularization and Interpolation via Random Spanning Forests. 2020. arXiv: 2011.10450.

[35] J. Pitman. “Coalescent Random Forests”. In: Journal of combinatorial theory. Series A 85.2 (1999), pp. 165–193.

[36] J. Pitman. Combinatorial Stochastic Processes. Springer, 2006.

[37] D. B. Wilson. “Generating random spanning trees more quickly than the cover time”. In: Proceedings of the Twenty-Eight Annual ACM Symposium on the Theory of Computing. Vol. 96. 1996, pp. 296–303.

† Leiden University, Mathematical Institute, Niels Bohrweg 1 2333 CA, Leiden. The Netherlands. Email address: l.avena@math.leidenuniv.nl

⋆ Leiden University, Mathematical Institute, Niels Bohrweg 1 2333 CA, Leiden. The Netherlands. Email address: v.t.koperberg@math.leidenuniv.nl

§ Leiden University, Mathematical Institute, Niels Bohrweg 1 2333 CA, Leiden. The Netherlands. Email address: jannetjedriessen@gmail.com