Title

Aspects of Local Time Processes with Applications

ISBN

9798189270994

Author

Ding, Qinghua

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Aspects of Local Time Processes with Applications

by

Qinghua Ding

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Venkat Anantharam, Chair
Professor Alistair Sinclair
Associate Professor Thomas Courtade
Associate Professor Shirshendu Ganguly

Spring 2026

Aspects of Local Time Processes with Applications

Abstract

Aspects of Local Time Processes with Applications

by

Qinghua Ding

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Venkat Anantharam, Chair

This dissertation studies the role of local time in the analysis of random walks and self-interacting random processes on finite graphs. Local time, which records the time a trajectory spends at a state, along an edge, or in a prescribed region, appears as a central object across four interconnected problems: non-reversible extensions of isomorphism theorems connecting local times to Gaussian and permanental fields; the time needed to achieve a second cover of a graph after the first; statistical estimation and information-theoretic analysis of edge-reinforced random walks; and acceleration of Markov chain Monte Carlo via self-repellent dynamics.

In each setting, local time plays a different but related role. For non-reversible Markov chains, we develop a density-formula approach that unifies proofs of several isomorphism theorems and yields a comparison inequalities for permanental processes and a symmetrization bound on cover times. For the second-cover problem, we combine the Ray-Knight isomorphism theorem with excursion decompositions to bound the marginal time to the second cover in terms of the geometry of the graph. For edge-reinforced random walks, we exploit the random-environment representation to give the first sample-complexity guarantees for parameter estimation, and to derive explicit formulas for entropy rate and Kullback-Leibler divergences between reinforced walk laws. For Markov chain Monte Carlo, we show that a true-self-avoiding-walk mechanism keeps empirical occupation counts close to stationarity at a rate nearly of order one over the number of steps, strictly improving on the diffusive rate of standard methods.

Dedicated to my late mother, Qin Zhou.

Contents

List of Figures

List of Tables

Acknowledgments

This research was supported by the grants CCF-1901004 and CIF-2007965 from the U.S. National Science Foundation.

Chapter 1

Overview

The common theme of this dissertation is the role of local time in the analysis of random walks and self-interacting random processes. Local time records the amount of time a trajectory spends at a state, along an edge, or in a prescribed part of the state space. In different chapters, this object appears in different but closely related forms: vertex occupation counts, continuous-time local times, edge traversal counts, inverse local times, and reinforced edge weights.

Let V be a finite state space, and let $(X_t)_{t\geq 0}$ be a discrete-time random process on V. For $i \in V$, define the discrete-time vertex local time by

$$L_t(i) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}.$$

Equivalently, $L_t(i)$ is the number of visits to i before time t, and $L_t = (L_t(i))_{i\in V}$ records the empirical occupation profile of the trajectory.

When the process is a Markov chain with transition kernel P and stationary distribution π, the vector L_t measures how the trajectory explores the state space. For the continuous-time simple random walk considered below, one instead uses degree-normalized continuous-time local times $(\mathcal{L}_t^v)_{v\in V}$, defined as integrals of indicator functions of the path with a degree normalization; see Equation (1.2) for the precise convention used in this dissertation. In both settings, local times provide a pathwise way to study occupation, recurrence, hitting, and covering phenomena.

The same basic idea reappears in self-interacting processes. In edge-reinforced random walks, edge traversal counts modify future transition probabilities. In true self-avoiding walks, occupation and transition counts are used to discourage overuse of states or transitions. Thus local time is not merely a passive statistic of a trajectory; it can also be part of the mechanism that drives the future evolution of the process.

The rest of this chapter is organized around the central topic of local time and random processes. Section 1.1 discusses local times for Markov chains, isomorphism theorems, permanental processes, and cover-time questions. Section 1.2 discusses edge-reinforced random walks and their random-environment representations. Finally, Section 1.3 discusses true self-avoiding walks and self-repellent dynamics.

1.1 Local time and Markov chains

Since the discovery of the Dynkin-BFS isomorphism theorem [43, 19], isomorphism theorems for reversible Markov chains have been an important tool to study stochastic processes related to Markov chains. Examples of applications of these isomorphism theorems include the study of large deviations, continuity, and boundedness of local time processes of Markov processes [46, 78], tight bounds for and concentration of the cover time of random walks on graphs [36, 110], percolation and level sets of the loop soup ensemble [71], and related questions in mathematical physics such as random walk loop soups and Brownian loop soups [63]. The book [78] contains a rather diverse list of such applications.

In the reversible setting, these isomorphism theorems relate local times of Markov processes to Gaussian fields. This connection is especially useful because geometric information about the underlying graph can often be encoded in the covariance structure of the associated Gaussian free field. Thus questions about random-walk occupation can be transformed into questions about Gaussian processes, where comparison inequalities and concentration tools are available.

The continuous-time simple random walk and its local times

We fix the graph and local-time notation used throughout this section. Let $G = (V, E)$ be a finite connected graph. We write $d(v)$ for the degree of v and $u \sim v$ when $\{u, v\} \in E$. We consider the *continuous-time simple random walk* (CT-SRW) $(X_t)_{t\geq 0}$ with i.i.d. Exp(1) holding times: when the chain is at $u \in V$, it waits an Exp(1) time and then jumps to a uniformly chosen neighbor of u. The generator of the CT-SRW acts on functions $f : V \to \mathbb{R}$ by

$$(\mathcal{A}f)(u) \;:=\; \frac{1}{d(u)} \sum_{v\sim u} \big(f(v) - f(u)\big),$$

and the stationary distribution is

$$\pi(u) = \frac{d(u)}{2|E|}, \qquad u \in V.$$

We write $\mathbb{P}_{v_0}$ and $\mathbb{E}_{v_0}$ for probability and expectation when $X_0 = v_0$. Let $0 = T_0 < T_1 < T_2 < \cdots$ denote the jump times of the chain, and set $Y_k := X_{T_k}$ for the embedded discrete-time chain. For $v \in V$ and $t \geq 0$, the *visit count* is

$$N_t(v) \;:=\; \sum_{k\geq 0} \mathbf{1}\{T_k \leq t,\, Y_k = v\}; \tag{1.1}$$

this counts the number of jump-time visits to v by continuous time t, including the initial visit if $v = v_0$.

Since holding times have unit mean, the discrete-time and continuous-time versions of the random walk differ only by constant factors in asymptotic expressions. We therefore work throughout in continuous time, for compatibility with local-time methods.

For $v \in V$ and $t \geq 0$, the *degree-normalized local time* is

$$\mathcal{L}_t(v) := \frac{1}{d(v)} \int_0^t \mathbf{1}\{X_s = v\}\, ds. \tag{1.2}$$

This normalization gives the identity

$$\sum_{v \in V} d(v)\, \mathcal{L}_t(v) = t. \tag{1.3}$$

We also consider stationary-normalized local time in this thesis; for details, see Remark 1.1.2.1. Fix a distinguished vertex $v_0 \in V$. The *inverse local time at* v_0 is

$$\tau_{\mathrm{inv}}(t) := \inf\{s \geq 0 : \mathcal{L}_s(v_0) \geq t\}, \qquad t \geq 0. \tag{1.4}$$

The stopping time $\tau_{\mathrm{inv}}(t)$ is the canonical time-change at which the generalized second Ray-Knight isomorphism theorem is naturally stated.

We remark that distinct uses of the base letter L appear in this dissertation. In this Markov-chain section, $\mathcal{L}_t(v)$ denotes the degree-normalized continuous-time local time of Equation (1.2). In the TSAW section, $L_t(i)$ denotes the discrete-time occupation count $\sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}$. In the ERRW section, $L_t(e)$ denotes the edge local time and dynamic edge weight Equation (1.14). The superscript type (vertex label vs. edge label) and the surrounding context always disambiguate these uses.

Green kernel, Gaussian free field, and cover times

We now recall the standard geometric objects attached to G.

Let $\tau_{\mathrm{hit}}(v) := \inf\{t \geq 0 : X_t = v\}$ denote the first hitting time of v. Associate $G = (V, E)$ with the electrical network in which each edge carries unit resistance. For $u, v \in V$, the *effective resistance* $r(u, v)$ is defined via Thomson's principle; see [41, 73]. Set

$$R := \max_{u,v \in V} r(u, v).$$

For the embedded discrete-time simple random walk $Y = (Y_k)_{k \geq 0}$, let

$$\tau^Y_{\mathrm{hit}}(v) := \inf\{k \geq 0 : Y_k = v\}.$$

The *commute-time identity* states that for all $u, v \in V$,

$$\mathbb{E}^Y_u[\tau^Y_{\mathrm{hit}}(v)] + \mathbb{E}^Y_v[\tau^Y_{\mathrm{hit}}(u)] = 2|E|\, r(u, v); \tag{1.5}$$

see [23, 68]. Since the CT-SRW has unit-mean holding times, the continuous-time hitting times have the same order as their embedded discrete-time analogues. In particular,

$$t_{\mathrm{hit}} := \max_{u,v \in V} \mathbb{E}_u[\tau_{\mathrm{hit}}(v)] = \Theta(|E|R).$$

Fix $v_0 \in V$. The *Green kernel killed at* v_0 is

$$G^{v_0}(u,v) \;:=\; \mathbb{E}_u\big[\mathcal{L}_{\tau_{\mathrm{hit}}(v_0)}(v)\big] \;=\; \frac{1}{d(v)}\,\mathbb{E}_u\left[\int_0^{\tau_{\mathrm{hit}}(v_0)} \mathbf{1}\{X_s = v\}\,ds\right], \qquad u,v \in V. \tag{1.6}$$

If $\mathsf{L}^{(v_0)}$ denotes the principal submatrix, indexed by $V \setminus \{v_0\}$, of the unit-conductance graph Laplacian, then, with the degree-normalized local-time convention Equation (1.2),

$$\big(G^{v_0}(u,v)\big)_{u,v \in V\setminus\{v_0\}} = \big(\mathsf{L}^{(v_0)}\big)^{-1};$$

see [73, 77].

The *Gaussian free field (GFF) pinned at* v_0 is the centered Gaussian process $\eta = (\eta(v), v \in V)$ with $\eta(v_0) = 0$ a.s. and covariance

$$\mathbb{E}[\eta(u)\eta(v)] \;=\; G^{v_0}(u,v), \qquad u,v \in V. \tag{1.7}$$

The GFF is naturally tied to the resistance metric via $r(u,v) = \mathbb{E}[(\eta(u) - \eta(v))^2]$, and the explicit formula

$$G^{v_0}(u,v) \;=\; \frac{1}{2}\big(r(v_0,u) + r(v_0,v) - r(u,v)\big) \tag{1.8}$$

holds; see [34, 73]. Set $M := \mathbb{E}[\max_{v\in V} \eta(v)]$; this quantity does not depend on the choice of the pinned vertex v_0.

The central result of Ding-Lee-Peres [37] (sharpened and extended in [34], with connections to majorizing measures [98]) states that the cover time satisfies

$$t_{\mathrm{cov}} \;=\; \Theta\big(|E|\,M^2\big), \tag{1.9}$$

with absolute implied constants. A useful comparison between R and M is

$$R \;\le\; 2\pi M^2; \tag{1.10}$$

a proof is included in Appendix B.3 for completeness.

The quantities R and M should be viewed as two complementary measurements of the geometry of G. The maximal effective resistance R controls worst-case hitting-time behavior, while the expected GFF maximum M captures the majorizing-measure scale that governs cover-time behavior.

Tables 1.1 and 1.2 summarize the known orders of magnitude for several classical graph families.

Graph	t_{mix}	t_{hit}	t_{cov}
Path / cycle on n vertices [68]	$\Theta(n^2)$	$\Theta(n^2)$	$\Theta(n^2)$
2D torus/box ($n = L^2$) [68, 30, 35]	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n(\log n)^2)$
d-dim torus ($d \geq 3$ fixed, $n = L^d$) [6, 13]	$\Theta(n^{2/d})$	$\Theta(n)$	$\Theta(n \log n)$
d-dim hypercube ($n = 2^d$) [68]	$\Theta(\log n \cdot \log\log n)$	$\Theta(n)$	$\Theta(n \log n)$
Full b-ary tree ($b \geq 2$ fixed, $n \asymp b^h$) [38, 68]	$\Theta(n)$	$\Theta(n \log n)$	$\Theta(n(\log n)^2)$
Star on n leaves ($\lvert V\rvert = n+1$) [68]	$\Theta(\log n)$	$\Theta(n)$	$\Theta(n \log n)$
d-regular expander (d constant) [68]	$\Theta(\log n)$	$\Theta(n)$	$\Theta(n \log n)$
Barbell [4, 72]	$\Theta(n^3)$	$\Theta(n^3)$	$\Theta(n^3)$
Lollipop [68]	$\Theta(n^3)$	$\Theta(n^3)$	$\Theta(n^3)$

Table 1.1: Order-of-magnitude summary for the CT-SRW on classical graphs. Here t_{mix} is the total-variation mixing time to threshold $\frac{1}{4}$, $t_{\text{hit}} := \max_{u,v} \mathbb{E}_u[\tau_{\text{hit}}(v)]$, and $t_{\text{cov}} := \max_{v_0} \mathbb{E}_{v_0}[\tau_{\text{cov}}]$.

Graph	$\lvert E\rvert$	R	M
n-path / n-cycle	$\Theta(n)$	$\Theta(n)$	$\Theta(\sqrt{n})$
2D grid/torus ($n = L^2$)	$\Theta(n)$	$\Theta(\log n)$	$\Theta(\log n)$
Full b-ary tree ($b \geq 2$ fixed, height $h \asymp \log n$)	$\Theta(n)$	$\Theta(\log n)$	$\Theta(\log n)$
d-dim grid/torus ($d \geq 3$ fixed, $n = L^d$)	$\Theta(n)$	$\Theta(1)$	$\Theta(\sqrt{\log n})$
Hypercube	$\Theta(n \log n)$	$\Theta(1/\log n)$	$\Theta(1)$
Star ($\lvert V\rvert = n+1$)	$\Theta(n)$	$\Theta(1)$	$\Theta(\sqrt{\log n})$
d-regular expander (d fixed)	$\Theta(n)$	$\Theta(1)$	$\Theta(\sqrt{\log n})$
Barbell	$\Theta(n^2)$	$\Theta(n)$	$\Theta(\sqrt{n})$
Lollipop	$\Theta(n^2)$	$\Theta(n)$	$\Theta(\sqrt{n})$

Table 1.2: Typical orders of the resistance diameter R and the pinned GFF maximum M for classical graphs. These are inferred from Table 1.1 together with $t_{\text{hit}} = \Theta(|E|R)$ and $t_{\text{cov}} = \Theta(|E|M^2)$.

The generalized second Ray-Knight isomorphism theorem

The following isomorphism theorem relates local times at inverse local time to the pinned GFF.

Theorem 1.1.1 (Generalized second Ray-Knight isomorphism [49, 77]). *Let $G = (V, E)$*

be a finite connected graph and let $(X_t)_{t\geq 0}$ *be the CT-SRW with* $\mathrm{Exp}(1)$ *holding times. Fix* $v_0 \in V$*, let* $\mathcal{L}_t(v)$ *be the degree-normalized local times Equation* (1.2)*, and let* $\tau_{\mathrm{inv}}(\cdot)$ *be the inverse local time Equation* (1.4)*. Let* $(\eta(v), v \in V)$ *be the GFF pinned at* v_0*. Then for every* $t > 0$*,*

$$\left\{\tfrac{1}{2}\,\eta(v)^2 + \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) \,:\, v \in V\right\} \overset{d}{=} \left\{\tfrac{1}{2}\left(\eta(v) + \sqrt{2t}\right)^2 \,:\, v \in V\right\}. \tag{1.11}$$

Theorem 1.1.1 is one of the main bridges between Gaussian free field and cover-time estimates. It allows one to study fluctuations of occupation fields at inverse local time through the pinned GFF, and hence through the effective-resistance geometry of the graph.

Taking the weighted sum $\sum_{v\in V} d(v)(\cdot)$ on both sides of Equation (1.11) and using identity Equation (1.3), one obtains

$$\tau_{\mathrm{inv}}(t) + \frac{1}{2}\sum_{v\in V} d(v)\,\eta(v)^2 \overset{d}{=} \frac{1}{2}\sum_{v\in V} d(v)\left(\eta(v) + \sqrt{2t}\right)^2. \tag{1.12}$$

Identity Equation (1.12) is the starting point for the following concentration estimate on $\tau_{\mathrm{inv}}(t)$, due to Ding [34].

Lemma 1.1.2 (Concentration of inverse local time [34, Lemma 2.1])**.** *Let* $G = (V, E)$ *be a finite connected graph with a distinguished vertex* v_0*, and let* $R := \max_{u,v\in V} r(u, v)$*. For any* $t > 0$ *and any* $\lambda \geq 1$*,*

$$\mathbb{P}\Big(\big|\tau_{\mathrm{inv}}(t) - 2t|E|\big| \,\geq\, \big(\sqrt{\lambda t R} + \lambda R\big)|E|\Big) \,\leq\, 6e^{-\lambda/16}.$$

Remark 1.1.2.1 (CT-SRW rate)**.** *We remark that, importantly, there are two continuous-time local-time normalizations used in the dissertation. In Chapter 2, local time is normalized by the stationary distribution:*

$$\mathcal{L}^{\pi}_t(v) := \frac{1}{\pi(v)}\int_0^t \mathbf{1}\{X_s = v\}\,ds.$$

In Chapter 3, local time is normalized by degree:

$$\mathcal{L}^{d}_t(v) := \frac{1}{d(v)}\int_0^t \mathbf{1}\{X_s = v\}\,ds.$$

For the CT-SRW on an undirected graph, $\pi(v) = d(v)/(2|E|)$*, and hence*

$$\mathcal{L}^{\pi}_t(v) = 2|E|\,\mathcal{L}^{d}_t(v).$$

Consequently, the associated killed Green kernels and pinned Gaussian fields differ by the same deterministic scaling:

$$G^{v_0,\pi} = 2|E|\,G^{v_0,d}, \qquad \eta^{\pi} \overset{d}{=} \sqrt{2|E|}\,\eta^{d}.$$

The isomorphism theorem holds for each specific normalization and corresponding GFF. When the normalization is unambiguous, either from context or because it has been specified explicitly, we write $\mathcal{L}_t$*,* η*, etc., without a superscript for simplicity.*

Cover time and the double Dixie cup problem

We now define the cover-time objects used throughout the dissertation. The *cover time* is

$$\tau_{\mathrm{cov}} := \inf\{t \geq 0 : \min_{v \in V} N_t(v) \geq 1\},$$

the *second cover time* is

$$\tau_{\mathrm{cov}}^{(2)} := \inf\{t \geq 0 : \min_{v \in V} N_t(v) \geq 2\},$$

and fixing some $v_0 \in V$, the *marginal time to the second cover* is

$$\Delta_{\mathrm{cov}} := \tau_{\mathrm{cov}}^{(2)} - \tau_{\mathrm{cov}}.$$

We study both Δ_{cov} and its expectation $\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}]$; when the distinction is clear from context or has been specified explicitly, we refer to either simply as the marginal time to the second cover. is

$$\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] := \mathbb{E}_{v_0}[\tau_{\mathrm{cov}}^{(2)} - \tau_{\mathrm{cov}}].$$

We also define the worst-case expected cover time, worst-case expected hitting time, and mixing time:

$$t_{\mathrm{cov}} := \max_{v_0 \in V} \mathbb{E}_{v_0}[\tau_{\mathrm{cov}}], \qquad t_{\mathrm{hit}} := \max_{u,v \in V} \mathbb{E}_u[\tau_{\mathrm{hit}}(v)],$$

$$t_{\mathrm{mix}} := \inf\left\{ t \geq 0 : \max_{u \in V} d_{\mathrm{TV}}\big(P_t(u,\cdot), \pi\big) \leq \tfrac{1}{4} \right\}.$$

Here $P_t(u,\cdot)$ denotes the time-t transition law of the CT-SRW started at u, and $\pi(u) = d(u)/(2|E|)$ is its stationary distribution. See [68] for background on these notions.

The classical *coupon collector problem* studies the number of i.i.d. uniform draws from $[n] := \{1, \dots, n\}$ required to observe each element at least once. The expectation of this duration is $n \log n + O(n)$, with logarithm to the natural base. A natural extension is the *double Dixie cup problem*, which studies the number of draws needed to collect at least two copies of each kind; it is known that the expectation equals $n \log n + n \log \log n + O(n)$ [86]. The corresponding random-walk problem asks for an analogous “second-collection” phenomenon on a graph: after the walk has visited every vertex at least once, how much additional time is needed until every vertex has been visited at least twice? When the underlying graph is an n-star, this reduces exactly to the double Dixie cup problem, and the expected marginal time to the second cover is of order $(1 + o(1))n \log \log n$.

The marginal second-cover problem, therefore, lies naturally at the interface between hitting-time geometry, cover-time geometry, and local-time fluctuations. Using Theorem 1.1.1 and Lemma 1.1.2 together with an excursion decomposition of the random walk, one can show that for any $v_0 \in V$,

$$\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] = O\big(|E|\, M\sqrt{R}\big) = O\big(\sqrt{t_{\mathrm{hit}}\, t_{\mathrm{cov}}}\big), \tag{1.13}$$

where the second equality uses $t_{\mathrm{hit}} = \Theta(|E|R)$ and $t_{\mathrm{cov}} = \Theta(|E|M^2)$. This bound is the central estimate developed in Chapter 3.

Non-reversible isomorphism theorems and permanental processes

Non-reversible versions of Dynkin's isomorphism theorem and the generalized second Ray-Knight isomorphism theorem were discovered by Le Jan, Eisenbaum, and Kaspi [64, 46, 44]. These isomorphism theorems have been used to study properties of the local time process for non-reversible Markov chains, such as continuity [75] and boundedness [76], and also to study the large deviations of local time processes [21].

In the non-reversible setting, the relevant random fields are no longer ordinary Gaussian fields. Instead, one is led to permanental vectors and permanental processes. For $\alpha > 0$, a random vector $\ell \in \mathbb{R}_+^n$ is called an *α-permanental vector* with kernel $K \in \mathbb{R}^{n\times n}$ if, for every diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ with $\lambda_i \geq 0$,

$$\mathbb{E}\left[\exp\left\{-\sum_{i=1}^{n}\lambda_i\ell_i\right\}\right] = \det(I + \Lambda K)^{-\alpha}.$$

Vere-Jones [105] has established necessary and sufficient conditions on the pair (K, α) for K to be the kernel of an α-permanental vector; see also [45, Sec. 2]. When $\alpha = 1/2$ and K is symmetric positive semidefinite, this reduces to the square of a Gaussian process. In general, the kernel K need not be symmetric, and permanental processes provide a natural replacement for Gaussian fields in non-reversible isomorphism theorems.

A particularly important class of kernels arises from Markov chains. Let P be an irreducible discrete-time transition kernel on $[n]$ with stationary distribution π, and write $\Pi = \mathrm{diag}(\pi)$. The Laplacian is $Q = \Pi(P - I)$. Introducing a killing rate vector $\boldsymbol{h} = (h_1, \dots, h_n) \geq 0$ with $\sum_i h_i > 0$, the killed rate matrix is $Q_{\boldsymbol{h}} = Q - \mathrm{diag}(\boldsymbol{h})$, and the *Markovian kernel* is

$$G_{\boldsymbol{h}} := -Q_{\boldsymbol{h}}^{-1}.$$

For such Markovian kernels, the corresponding 1-permanental vector exists; see [105].

This allows us to generalize Ray-Knight isomorphism theorems to the non-reversible case, with permanental vectors being the non-reversible extension of Gaussian squares in the local-time isomorphism theory: permanental processes connect non-reversible Markov chains and local-time fields in precisely the same way that squared Gaussian fields connect reversible chains to local times.

The density-formula approach developed in Chapter 2 starts from this permanental-process viewpoint. The goal is to derive density formulas for 1-permanental processes and conditioned 1-permanental processes, and then combine these formulas with Ward identities for twisted Gaussian measures, as in [12], to give succinct proofs of the non-reversible isomorphism theorems, extending the results of [12]. This method also yields comparison inequalities for permanental processes, generalizing inequalities of Eisenbaum [44], and cover-time bounds for non-reversible Markov chains.

1.2 Local time and edge-reinforced random walks

An edge-reinforced random walk (ERRW) is a random walk on a graph whose transition probabilities along an edge from a vertex are proportional to the weights of those edges, and where the weight of an edge increases by 1 whenever that edge is traversed in either direction. Introduced in the work of Coppersmith and Diaconis [25], the model has since been studied from several points of view, including partial exchangeability, Bayesian statistics for reversible Markov chains, and random-environment representations; see, for example, [62, 91, 33, 84, 80]. On finite graphs, a basic structural fact is that ERRW has the same law as a mixture of reversible Markov chains, or equivalently as a random walk in a random environment, with explicit mixing measure given by the so-called magic formula [62, 80]. This representation places ERRW in the broader framework of de Finetti-type results for Markov chains and reversible processes [31, 33]. The random-environment point of view has also led to further developments relating ERRW to the vertex-reinforced jump process (VRJP) and to random Schrödinger operators and the supersymmetric hyperbolic sigma model [92, 93]. We refer the reader to the survey article of Merkl and Rolles [84] for background on edge-reinforced random walk and its principal structural properties.

In this section, we introduce the two families of reinforced random walks studied in the dissertation, vertex-reinforced random walks and edge-reinforced random walks, and summarize the main statistical and information-theoretic properties of ERRW that underlie Chapters 4 and 5.

Vertex-reinforced and edge-reinforced random walks

Throughout this section let $G = (V, E)$ be a finite connected simple undirected graph, and fix a starting vertex $v_0 \in V$.

Vertex-reinforced random walk. Let $\Lambda_0 = (\Lambda_0^i)_{i\in V} \in (0,\infty)^V$ be a vector of initial vertex weights. The *vertex local time* at state $i \in V$ and time $t \in \mathbb{Z}_+$ is

$$\Lambda_t^i := \Lambda_0^i + \sum_{s=1}^{t} \mathbf{1}(X_s = i).$$

The process $(\Lambda_t)_{t\geq 0}$, where $\Lambda_t := (\Lambda_t^i)_{i\in V}$, is the vertex local time process. The *vertex-reinforced random walk* (VRRW) starting at v_0 with initial weights Λ_0 is defined by

$$\mathbb{P}^{v_0,\Lambda_0}_{\text{vrrw}}(X_{t+1} = i \mid X_0, \dots, X_t) = \frac{\Lambda_t^i}{\sum_{i'\sim X_t} \Lambda_t^{i'}}, \qquad i \sim X_t, \quad t \in \mathbb{Z}_+.$$

VRRW can exhibit localization: on $\mathbb{Z}$, given $\Lambda_0 = \mathbf{1}$ and $v_0 = 0$, for any $k \in \mathbb{Z}$ there is positive probability that the walk is eventually confined to the five-vertex induced subgraph $\{k-2, k-1, k, k+1, k+2\}$ [99]. This localization phenomenon generalizes to other graphs [106], and is in stark contrast with the behavior of ERRW.

Edge-reinforced random walk. Let $A = (a_e)_{e\in E} \in (0,\infty)^E$ be a vector of initial edge weights, and write $a_v := \sum_{e\ni v} a_e$. The *edge local time* at edge $e \in E$ and time $t \in \mathbb{Z}_+$ is

$$L_t(e) := a_e + \sum_{s=0}^{t-1} \mathbf{1}\big(\{X_s, X_{s+1}\} = e\big) = a_e + N_e(X_0^t), \tag{1.14}$$

where $N_e(X_0^t) := N_{uv}(X_0^t) + N_{vu}(X_0^t)$ counts undirected traversals of $e = \{u, v\}$. The process $(L_t)_{t\ge 0}$, where $L_t := (L_t(e))_{e\in E}$, is the edge local time process, or equivalently the dynamic edge weight process. The *edge-reinforced random walk* (ERRW) starting at v_0 with initial weights A is the process $(X_t)_{t\ge 0}$ with $X_0 = v_0$ and

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(X_{t+1} = i \mid X_0, \dots, X_t) = \frac{L_t^{\{X_t,i\}}}{\sum_{i'\sim X_t} L_t^{\{X_t,i'\}}}, \qquad i \sim X_t, \quad t \in \mathbb{Z}_+. \tag{1.15}$$

Thus, the edge traversal counts directly determine the future transition probabilities. In this sense, local time is not only a statistic of the trajectory but part of the mechanism defining the process.

More results are available for ERRW on infinite graphs: on $\mathbb{Z}^d$, ERRW is recurrent when $d \le 2$, and transient when $d \ge 3$ [92, 12]. These results exploit the connection between ERRW and VRJP [92] and the relation of VRJP to the Anderson localization model [12, 92].

Random-environment representation and the magic formula

The key structural property of ERRW is partial exchangeability: for any t, v_0, and $A, B \in (0,\infty)^E$, conditioned on $X_0 = v_0$, $L_0 = A$, and $L_t = B$, the distribution of the trajectory $(X_0, \dots, X_t)$ is uniform among all trajectories consistent with these constraints. By de Finetti's theorem for reversible Markov chains [33, 31], this partial exchangeability implies that ERRW is a mixture of reversible Markov chains.

The explicit mixing measure is given by the magic formula. Fix a reference edge $e_0 \in E$ incident to v_0, and work in the pinned space $\mathcal{W}_{e_0} := \{w \in (0,\infty)^E : w_{e_0} = 1\}$. Write $w_v := \sum_{e\ni v} w_e$ for the vertex sum of conductances. Let $\mathcal{T}$ denote the set of spanning trees of G, and define the spanning-tree polynomial

$$\tau(w) := \sum_{T\in\mathcal{T}} \prod_{e\in T} w_e.$$

For $v \in V$, define

$$b_v := \frac{a_v + 1 - \mathbf{1}(v = v_0)}{2}. \tag{1.16}$$

Definition 1.2.0.1 (Magic-formula mixing measure [80]). *The* mixing measure $\mu_{v_0,A}$ *is the probability measure on* $\mathcal{W}_{e_0}$ *with density*

$$d\mu_{v_0,A}(w_{-e_0}) \;=\; Z_{v_0,A}^{-1}\,\frac{\prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{b_v}}\sqrt{\tau(w)}\,dw_{-e_0}, \tag{1.17}$$

where $w_{-e_0} := (w_e)_{e\neq e_0}$ *and the normalizing constant is*

$$Z_{v_0,A} \;:=\; \frac{\pi^{(|V|-1)/2}}{2^{1-|V|+\sum_{e\in E} a_e}}\cdot\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(b_v)}. \tag{1.18}$$

Theorem 1.2.1 (Random-environment representation [62, 80]). *Let* $(X_t)_{t\geq 0}$ *be the ERRW on* (G, v_0, A). *Then there exists a random environment* $W \sim \mu_{v_0,A}$ *such that, conditional on* $W = w$, *the process* $(X_t)_{t\geq 0}$ *is the reversible Markov chain on* V *with transition kernel*

$$p_{ij}(w) = \frac{w_{\{i,j\}}}{w_i}, \qquad i \sim j.$$

Equivalently,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\,\cdot\,) = \int \mathbb{P}^{v_0,w}(\,\cdot\,)\,\mu_{v_0,A}(dw),$$

where $\mathbb{P}^{v_0,w}$ *denotes the law of the reversible Markov chain with conductance environment* w, *started from* v_0.

The quenched chain $P^w = (p_{ij}(w))$ is reversible with stationary distribution

$$\pi_i^w := \frac{w_i}{\sum_{v\in V} w_v}, \qquad i \in V.$$

Since G is finite and connected and all conductances are positive, the quenched chain P^w is finite and irreducible, and hence visits every state infinitely often. Therefore, by the random-environment representation, ERRW on a finite connected graph visits every state infinitely often almost surely.

The quenched likelihood of a trajectory $x_0^T = (x_0, \ldots, x_T)$ with $x_0 = v_0$ factors through the directed transition counts $N_{uv}(x_0^T) := |\{t : x_{t-1} = u, x_t = v\}|$ and the departure counts $N_u(x_0^T) := \sum_{v\sim u} N_{uv}(x_0^T)$:

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T) = \prod_{u\in V}\prod_{v\sim u}\left(\frac{w_{\{u,v\}}}{w_u}\right)^{N_{uv}(x_0^T)}. \tag{1.19}$$

In particular, for fixed w, the quenched law of X_0^T depends on the trajectory only through the directed transition count vector $N(X_0^T) := (N_{uv}(X_0^T))_{u\sim v}$. This count-sufficiency property underlies both the statistical estimation theory of Chapter 4 and the information-theoretic analysis of Chapter 5.

The random-environment representation has two levels at which local time appears. At the path level, the edge traversal counts $N_e(X_0^T)$ update the reinforced weights $L_t(e)$ and determine the future transitions of the walk. At the environment level, the same counts determine the posterior update of $\mu_{v_0,A}$ after observing a path: by the conjugacy property of the magic formula [33], observing a trajectory of length T ending at some vertex v_T updates the mixing measure from $\mu_{v_0,A}$ to $\mu_{v_T,\,A+N(X_0^T)}$, where $A + N(X_0^T)$ denotes the componentwise sum of initial weights and undirected edge counts. This posterior-update structure connects ERRW to Bayesian inference for reversible Markov chains and is a key input in the information-theoretic analysis.

The Sabot-Tarrès-Zeng field representation

A deeper structural property of ERRW is revealed by the random-field representation developed by Sabot, Tarrès, and Zeng [92, 93]. In this representation, one begins with independent edge variables $\beta = (\beta_e)_{e\in E}$ with $\beta_e \sim \mathrm{Gamma}(a_e, 1)$, and, conditional on β, samples a field $\phi = (\phi_i)_{i\in V}$ in the pinned gauge $\phi_{v_0} = 0$ according to a hyperbolic Gaussian density depending on β:

$$p_\beta(\phi) = \frac{1}{(2\pi)^{(n-1)/2}}\, e^{-\sum_{e=\{i,j\}\in E}\beta_e(\cosh(\phi_i-\phi_j)-1)} e^{-\sum_{i\in V\setminus\{v_0\}}\phi_i} \sqrt{\sum_{T\in\mathcal{T}}\prod_{e=\{i,j\}\in T}\beta_e e^{\phi_i+\phi_j}}.$$

The associated conductances and vertex variables are then

$$W_{ij} := \beta_{ij}\, e^{\phi_i+\phi_j}, \qquad \tilde\psi_i := \tfrac{1}{2}\sum_{j\sim i}\beta_{ij}\, e^{\phi_j-\phi_i}, \qquad \{i,j\}\in E,\ i\in V. \tag{1.20}$$

Since $W_i := \sum_{j\sim i} W_{ij} = 2\tilde\psi_i e^{2\phi_i}$, it is natural to introduce the *normalized edge field*

$$\tilde\beta_{ij} := \frac{\beta_{ij}}{\sqrt{\tilde\psi_i\tilde\psi_j}} = \frac{2W_{ij}}{\sqrt{W_iW_j}}, \qquad \{i,j\}\in E. \tag{1.21}$$

In this parametrization, the vertex variables separate from the environment. Specifically, the coordinates of $\tilde\psi = (\tilde\psi_v)_{v\in V}$ are independent with $\tilde\psi_v \sim \mathrm{Gamma}(b_v, 1)$, where b_v is as in Equation (1.16), and the joint law factorizes as

$$(W, \tilde\psi) \sim \mu_{v_0,A} \otimes \bigotimes_{v\in V}\mathrm{Gamma}(b_v, 1). \tag{1.22}$$

The independence and the Gamma marginals are established in Theorem 5.3.3. Equivalently, if $\tilde\nu_{v_0,A}$ denotes the law of the normalized field $\tilde\beta = (\tilde\beta_e)_{e\in E}$, then $(\tilde\beta, \tilde\psi) \sim \tilde\nu_{v_0,A} \otimes \bigotimes_{v\in V}\mathrm{Gamma}(b_v, 1)$.

For the purposes of this dissertation, the STZ field representation is useful for two related reasons. First, in the $(W, \tilde\psi)$ parametrization, the ERRW environment W and the independent vertex field $\tilde\psi$ separate completely. This separation provides coordinates in which the random environment can be controlled quantitatively, which is the key technical input for the sample-complexity bounds in Chapter 4. Second, the same factorization reveals cancellations in information-theoretic quantities: the KL divergence between two ERRW environment laws decomposes as a difference of edge-Gamma and vertex-Gamma KL divergences, as described in the next subsection.

Information-theoretic quantities associated with ERRW

The random-environment representation gives ERRW two layers of randomness: the hidden environment $W \sim \mu_{v_0,A}$ and the Markov chain trajectory generated conditional on W. The information-theoretic quantities studied in Chapter 5 are the entropy rate of the reinforced walk, the KL divergence between two ERRW environment laws, and the KL divergence between the corresponding finite-trajectory laws.

Entropy rate. For a finite irreducible Markov chain P with stationary distribution π, the entropy rate is

$$r(P) := -\sum_{i\in V} \pi_i \sum_{j\in V} p_{ij} \log p_{ij}.$$

For the ERRW on (G, v_0, A) the entropy rate exists and satisfies

$$r(v_0, A) := \lim_{T\to\infty} \frac{1}{T} \mathrm{H}(X_0^T) = \mathbb{E}\big[r(P^W)\big], \tag{1.23}$$

where the expectation is over $W \sim \mu_{v_0,A}$. This *annealed representation* follows from the count-sufficiency property Equation (1.19) together with a mutual-information bound: the mutual information $I(W; X_0^T) \le (2|E| - 1)\log(T + 1)$, which is $o(T)$, so the environment contributes a negligible fraction of the total entropy per step.

KL divergence between environment laws. Fix two initial weight vectors $A^{(0)}, A^{(1)} \in (0,\infty)^E$, and let $\mu_s := \mu_{v_0,A^{(s)}}$. Once the graph and starting vertex are fixed, the family $\{\mu_{v_0,A}\}_{A\in(0,\infty)^E}$ forms a canonical exponential family with natural parameter A and sufficient statistic $T_e(w) := \log(w_e/\sqrt{w_u w_v})$ for $e = \{u, v\}$. Consequently, the KL divergence between two environment laws is the Bregman divergence of the log-partition function $\Phi(A) := \log Z_{v_0,A}$, and evaluates explicitly to

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = \sum_{e\in E} \Lambda\big(a_e^{(0)}, a_e^{(1)}\big) - \sum_{v\in V} \Lambda\big(b_v^{(0)}, b_v^{(1)}\big), \tag{1.24}$$

where $b_v^{(s)} := b_v(A^{(s)})$ and $\Lambda(p, q) := \log\Gamma(q) - \log\Gamma(p) - (q - p)\Psi(p)$ is the KL divergence between $\mathrm{Gamma}(p, 1)$ and $\mathrm{Gamma}(q, 1)$, with $\Psi = (\log\Gamma)'$ the digamma function. Through the STZ factorization Equation (1.22), this formula admits a probabilistic interpretation: the environment-level KL divergence equals the edge-Gamma KL divergence minus the vertex-Gamma KL divergence,

$$D_{\mathrm{KL}}(\mu_0 \| \mu_1) = \sum_{e \in E} D_{\mathrm{KL}}\big(\beta^{(0)} \big\| \beta^{(1)}\big) - \sum_{v \in V} D_{\mathrm{KL}}\big(\tilde{\psi}^{(0)} \big\| \tilde{\psi}^{(1)}\big).$$

We prove this probabilistic interpretation of the KL divergence in Chapter 5.

KL divergence between trajectory laws. Let $P_s^{(T)} := \mathbb{P}_{\mathrm{errw}}^{v_0, A^{(s)}}(X_0^T \in \cdot)$ denote the law of the first T steps of ERRW with weights $A^{(s)}$. By the data-processing inequality,

$$D_{\mathrm{KL}}\big(P_0^{(T)} \big\| P_1^{(T)}\big) \le D_{\mathrm{KL}}(\mu_0 \| \mu_1).$$

The trajectory-level divergence converges to the environment-level divergence as $T \to \infty$, and Chapter 5 derives quantitative bounds on the rate of this convergence. The statistical interpretation is provided by Stein's lemma: $D_{\mathrm{KL}}(P_0^{(T)} \| P_1^{(T)})$ is the optimal type-II error exponent in testing $A^{(0)}$ against $A^{(1)}$ from a single trajectory of length T at a fixed type-I error level.

1.3 Local time and true self-avoiding walks

MCMC estimation and occupation imbalance

Let V be a finite state space, let P be an irreducible Markov kernel on V with stationary distribution π, and let $f : V \to \mathbb{R}$ be bounded. Writing the vertex local time $L_t(i) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}$, the empirical integral error decomposes as

$$\frac{1}{t} \sum_{s=0}^{t-1} f(X_s) - \pi(f) = \frac{1}{t} \sum_{i \in V} f(i)\big(L_t(i) - t\pi_i\big).$$

The finite-time quality of the estimator is therefore governed by the discrepancy $L_t(i) - t\pi_i$ between empirical occupation counts and their stationary targets.

A substantial literature addresses this phenomenon by modifying the base chain to avoid backtracking or to break reversibility; see, e.g., [32, 85, 7, 65, 104, 24, 74, 100]. Nonlinear and self-interacting Markov chains [28, 29, 9, 10, 27, 55, 54] provide a general framework in which the transition rule depends on the occupation history. The closest prior work is Doshi, Hu, and Eun [40], who prove that a self-repellent random walk achieves a CLT with improved asymptotic variance, with empirical integral error $O(t^{-1/2})$.

Chapter 6 studies true self-avoiding walk (TSAW) as a mechanism for reducing occupation imbalance and achieving a strictly better rate.

The TSAW transition rule and main result

Fix a repulsion parameter $\lambda > 0$. Define the *directed transition-count discrepancy*

$$\epsilon_t(i,j) := N_t(i,j) - L_t(i)\, P_{ij}, \tag{1.25}$$

where $N_t(i,j) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i,\ X_{s+1} = j\}$ counts directed transitions. The *TSAW transition rule* is the time-inhomogeneous kernel

$$\mathbb{P}(X_{t+1} = j \mid \mathcal{F}_t,\ X_t = i) = Q_t(i,j) := \frac{P_{ij} \exp\{-\lambda\, \epsilon_t(i,j)\}}{\sum_{k \in V} P_{ik} \exp\{-\lambda\, \epsilon_t(i,k)\}}. \tag{1.26}$$

The softmax weight $e^{-\lambda \epsilon_t(i,j)}$ suppresses overused transitions ($\epsilon_t(i,j) > 0$) and amplifies underused ones ($\epsilon_t(i,j) < 0$), performing local row-wise balancing at each vertex i.

The main result of Chapter 6 is the following almost-sure bound: for every state $i \in V$ and every edge (i,j) with $P_{ij} > 0$,

$$L_t(i) - t\pi_i \;=\; O(\sqrt{\log t}) \qquad \text{and} \qquad N_t(i,j) - t\pi_i P_{ij} \;=\; O(\sqrt{\log t}) \qquad \text{a.s.} \tag{1.27}$$

Consequently, for every bounded $f : V \to \mathbb{R}$,

$$\left| \frac{1}{t} \sum_{s=0}^{t-1} f(X_s) - \pi(f) \right| = O\left(\frac{\sqrt{\log t}}{t} \right) \qquad \text{almost surely,}$$

improving the $O(t^{-1/2})$ rate of [40] by a factor of $O(\sqrt{t/\log t})$.

Local time appears in Chapter 6 in its most direct role: as the primary object being controlled. The TSAW transition rule is designed to keep the empirical occupation counts $L_t(i)$ close to their stationary targets $t\pi_i$, and Equation (1.27) characterizes precisely how well this is achieved. Unlike the ERRW chapters, where local time is a tool entering indirectly through cover times and random environments, here local time is the direct output: the chapter provides almost-sure finite-time pathwise bounds on occupation discrepancy with an explicit rate that is nearly $1/t$, up to the logarithmic factor.

Notation

The five chapters of this dissertation draw on several related but distinct uses of local time: vertex occupation counts, directed transition counts, undirected edge traversal counts, inverse local times, and reinforced edge weights. To avoid ambiguity, we fix the following notation throughout the dissertation. When a chapter requires a different convention, the local definition is stated explicitly.

Table 1.3: Notation used throughout the dissertation.

Notation	Meaning
Graphs and state spaces	
$G = (V, E)$	A finite graph with vertex set V and edge set E.
$\lvert V\rvert, \lvert E\rvert$	Number of vertices and edges of G.
$u \sim v$	Vertices u and v are adjacent.
$d(v)$	Degree of vertex v in an undirected graph.
$\mathcal{N}(i)$	Out-neighborhood $\{j : P_{ij} > 0\}$ for a Markov kernel P.
d_i	Out-degree $\lvert\mathcal{N}(i)\rvert$.
Random walks and Markov chains	
X_t	Random walk or self-interacting process at time t.
$P = (P_{ij})$	Transition kernel of a Markov chain.
π	Stationary distribution of P.
$\mathcal{F}_t$	Natural filtration generated by the process up to time t.
$\mathbb{P}_{v_0}$, $\mathbb{E}_{v_0}$	Probability and expectation for the process started at $X_0 = v_0$.
Local times and empirical counts	
$L_t(i)$	Discrete-time occupation count $\sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}$; mainly used in Chapter 6.
L_t	Occupation vector $(L_t(i))_{i\in V}$ in discrete time, or edge-weight vector $(L_t(e))_{e\in E}$ in the ERRW section.
$\mathcal{L}_t(v)$	Degree-normalized continuous-time local time $\frac{1}{d(v)} \int_0^t \mathbf{1}\{X_s = v\}\, ds$; used in the CT-SRW cover-time sections.
$L_t(e)$	ERRW edge local time and dynamic edge weight $a_e + N_e(X_0^t)$; used in the ERRW sections.
$\tau_{\mathrm{inv}}(t)$	Inverse local time at v_0, $\inf\{s \ge 0 : \mathcal{L}_s(v_0) \ge t\}$.
$N_t(i, j)$	Directed transition count $\sum_{s=0}^{t-1} \mathbf{1}\{X_s = i,\ X_{s+1} = j\}$.
$N_e(t)$	Undirected traversal count of edge e, $N_{uv}(X_0^t) + N_{vu}(X_0^t)$ for $e = \{u, v\}$.
$\Delta_i(t)$	TSAW vertex discrepancy $L_t(i) - t\pi_i$.
$\epsilon_t(i, j)$	TSAW directed transition-count discrepancy $N_t(i, j) - L_t(i)P_{ij}$.
T_k	Jump times of CT-SRW, $0 = T_0 < T_1 < T_2 < \cdots$.
Y_k	Embedded discrete-time chain, $Y_k := X_{T_k}$.
$N_t(v)$	CT-SRW visit count $\sum_{k\ge0} \mathbf{1}\{T_k \le t,\ Y_k = v\}$; see Equation (1.1).
Cover, hitting, and mixing times	

Continued on next page

Notation	Meaning
$\tau_{\text{hit}}(v)$	Hitting time of vertex v.
t_{hit}	Maximal expected hitting time.
τ_{cov}	First cover time.
$\tau^{(2)}_{\text{cov}}$	Second cover time.
Δ_{cov}	Marginal time to second cover $\tau^{(2)}_{\text{cov}} - \tau_{\text{cov}}$.
t_{mix}	Total-variation mixing time.
t_{cov}	Worst-case expected cover time $\max_{v_0 \in V} \mathbb{E}_{v_0}[\tau_{\text{cov}}]$.
Electrical networks and Gaussian fields	
$r(u, v)$	Effective resistance between vertices u and v.
R	Maximal effective resistance $\max_{u,v \in V} r(u, v)$.
η	Gaussian free field, usually pinned at v_0.
M	Expected GFF maximum, $M = \mathbb{E}[\max_{v \in V} \eta(v)]$.
ERRW and random environments	
$A = (a_e)_{e \in E}$	Initial ERRW edge-weight vector.
a_v	Initial vertex weight $a_v = \sum_{e \ni v} a_e$.
$W = (W_e)_{e \in E}$	Random conductance environment, we use w for its realization.
W_v	Environment vertex weight $W_v = \sum_{e \ni v} W_e$.
$\mu_{v_0,A}$	Mixing/environment law for ERRW started at v_0 with initial weights A.
$\mathcal{T}$	Set of all spanning trees of G.
$\tau(w)$	Spanning-tree polynomial $\sum_{T \in \mathcal{T}} \prod_{e \in T} w_e$.
b_v	Vertex weight parameter $b_v := (a_v + 1 - \mathbf{1}(v = v_0))/2$ in the ERRW mixing measure.
$\mathbb{P}^{v_0,A}_{\text{errw}}$	ERRW law started at v_0 with initial weights A.
$\mathbb{P}^{v_0,w}_{\text{srw}}$	Quenched reversible chain law, conductance w, started at v_0.
Information quantities	
$\mathrm{H}(\cdot)$	Entropy.
$D_{\text{KL}}(P \| Q)$	Kullback-Leibler divergence.
$d_{\text{TV}}(P, Q)$	Total-variation distance.
LLR	Log-likelihood ratio.
Linear algebra	
$\mathbf{1}$	All-ones vector; also used as an indicator when clear from context.
$\mathbf{e}_i$	Coordinate basis vector at state i.

Continued on next page

Notation	Meaning
$\mathrm{diag}(x)$	Diagonal matrix with diagonal vector x.
$\mathrm{tr}(A)$	Trace of a matrix A.
$\mathrm{rank}(A)$	Rank of a matrix A.
$\mathrm{supp}(\mu)$	Support of a measure μ.
STZ field (see Section 4.4 and Section 5.3 for detailed definition)	
$\beta = (\beta_e)_{e\in E}$	Edge field; $\beta_e \sim \mathrm{Gamma}(a_e, 1)$ independently.
$\phi = (\phi_i)_{i\in V}$	Hyperbolic gaussian field; sampled from the hyperbolic Gaussian density $p_\beta(\phi)$ conditional on β.
$Q = (Q_e)_{e\in E}$	Random environment $Q_e := \beta_e e^{\phi_i+\phi_j}$.
$\nu_{v_0,A}$	Law of the field Q.
$\psi = (\psi_i)_{i\in V}$	Vertex field in the original STZ parametrization, with an extra independent $\mathrm{Gamma}(\frac{1}{2}, 1)$ term added at v_0.
$\tilde{\psi} = (\tilde{\psi}_v)_{v\in V}$	Deghosted vertex field; equals ψ with the extra $\mathrm{Gamma}(\frac{1}{2}, 1)$ term at v_0 removed.
$\tilde{\beta} = (\tilde{\beta}_e)_{e\in E}$	Normalized β field; $\tilde{\beta}_{ij} := \beta_{ij}/\sqrt{\psi_i\psi_j}$.
$\tilde{\nu}_{v_0,A}$	Law of the normalized field $\tilde{\beta}$.

Chapter 2

Density Formula Approach for Non-reversible Isomorphism Theorems

This chapter develops the density-formula approach to non-reversible isomorphism theorems. The main technical input is an explicit density formula for a broad class of 1-permanental vectors, together with its conditioned version. We then combine these density formulas with Ward identities for twisted Gaussian measures to prove the non-reversible Dynkin, generalized second Ray-Knight, and Eisenbaum-type isomorphism theorems. The same method also yields comparison inequalities for 1-permanental processes and a symmetrization bound for cover times of non-reversible Markov chains.

2.1 Introduction

As reviewed in the Chapter 1, isomorphism theorems for reversible Markov chains connect local times with Gaussian fields and have many applications, including large deviations, continuity and boundedness of local-time processes, cover-time estimates, and loop soups. This chapter focuses on the non-reversible analogue, where Gaussian squares are replaced by 1-permanental processes.

Non-reversible versions of Dynkin's isomorphism theorem and the generalized second Ray-Knight theorem were proved by Le Jan, Eisenbaum, and Kaspi. Our aim here is not to re-survey this literature, but to give a unified density-formula-based proof. We first derive density formulas for 1-permanental vectors and conditioned 1-permanental vectors, and then combine them with Ward identities for twisted Gaussian measures.

We then use the density formula to extend comparison inequalities for permanental processes, in the spirit of [44]. Finally, we derive a symmetrization upper bound for the cover time of non-reversible Markov chains using the non-reversible generalized second Ray–Knight theorem together with the conditional density formula.

The chapter is organized as follows. In Section 2.2, we prove the density formula for 1-permanental vectors and its conditioned version. Section 2.3 derives Ward identities and uses them to prove the non-reversible Dynkin, Ray–Knight, and Eisenbaum isomorphism theorems. Section 2.4 develops comparison inequalities for 1-permanental vectors, including Kahane-type and Slepian-type inequalities. Finally, Section 2.5 applies the conditioned density formula and the non-reversible Ray–Knight theorem to obtain a symmetrization bound for cover times.

2.2 Density of 1-permanental vectors

We start with some basic notational conventions. For a positive integer n, let $[n]$ denote $\{1, \dots, n\}$. Let $M_n(\mathbb{R})$ denote the set of $n \times n$ matrices with real entries and $M_n(\mathbb{R}_+)$ the set of $n \times n$ matrices with nonnegative real entries. We write $:=$ for equality by definition.

We say that $G \in M_n(\mathbb{R})$ is diagonally equivalent to $\tilde{G} \in M_n(\mathbb{R})$ if there exists a diagonal matrix $D \in M_n(\mathbb{R})$ with non-zero diagonal entries such that $\tilde{G} = D^{-1}GD$. Given $A \in M_n(\mathbb{R})$, its spectral radius is defined as $\rho(A) := \lim_{k\to\infty} \|A^k\|^{\frac{1}{k}}$, where $\|\cdot\|$ is any matrix norm (the choice of the norm does not matter). We say that $A \in M_n(\mathbb{R})$ is diagonally dominant if $|A_{ii}| \geq \sum_{j\neq i} |A_{ij}|$ for all $i \in [n]$. It is called strictly diagonally dominant if these inequalities are strict for all $i \in [n]$. $A \in M_n(\mathbb{R})$ is called an M-matrix if it can be written as $sI - B$ for matrix B with nonnegative entries, and where $s \geq \rho(B)$, see [90, pg. 176]. It is known that A is a non-singular M-matrix iff:

1. $A_{ij} \leq 0$ for $i \neq j$;
2. A is non-singular and A^{-1} is entry-wise nonnegative.

See [90, Criterion F15, pg. 180].

For any $\alpha > 0$, a random vector $\ell \in \mathbb{R}_+^n$ is called an α-permanental vector with kernel $G \in M_n(\mathbb{R})$ if, for every $\lambda \in \mathbb{R}_+^n$, we have the Laplace transform

$$\mathbb{E}[e^{-\langle \lambda, \ell\rangle}] = |I + \Lambda G|^{-\alpha}.$$

Here, on the right-hand side, $\Lambda := \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, and $|I + \Lambda G|$ denotes the determinant of $I + \Lambda G$. We remark that a 1-permanental vector can correspond to different kernels (e.g., arising from diagonal equivalence). Vere-Jones [105] has established necessary and sufficient conditions on the pair (G, α) for G to be the kernel of an α-permanental vector, see also [45, Sec. 2].

1-permanental vectors are closely related to non-reversible isomorphism theorems [64, 46], and will be the main object of our study. Our first contribution is a density formula for a broad range of 1-permanental vectors. The proof is based on characteristic functions, following an approach similar to that of Brydges-Hofstad-König [20].

In order to state the result, we need to introduce some conventions regarding complex numbers and vectors. We write i for $\sqrt{-1}$, to distinguish it from a running index i. For a complex number $\phi \in \mathbb{C}$, we let $\Re(\phi)$ denote the real part, and $\Im(\phi)$ the imaginary part. We write $\phi = u + \mathrm{i}v$, so that $u = \Re(\phi)$ and $v = \Im(\phi)$. Then $\bar{\phi} = u - \mathrm{i}v$ denotes the complex conjugate of ϕ. For $\phi, \psi \in \mathbb{C}^n$, let $\langle \phi, \psi \rangle := \sum_{i=1}^n \phi_i \psi_i$ denote the (indefinite) real inner product on $\mathbb{C}^n$. Note that we do not conjugate either argument in computing this inner product. It will also be convenient to think of $\phi \in \mathbb{C}$ as $\sqrt{l}e^{\mathrm{i}\theta}$, so that $\sqrt{l} = |\phi|$ denotes the absolute value of ϕ and θ denotes its phase. Note that, in one complex dimension, we have $dudv = 2^{-1} dld\theta$. Also observe that we have, in one complex dimension, the formula

$$d\bar{\phi} \wedge d\phi = (du - \mathrm{i}dv) \wedge (du + \mathrm{i}dv) = 2\mathrm{i}du \wedge dv = \mathrm{i}dl \wedge d\theta, \tag{2.1}$$

as an equality of differential forms.

We can now state our density formula for the density of a class of 1-permanental vectors. Here, for any matrix $M \in M_n(\mathbb{R})$, its symmetric part is defined to be $\frac{1}{2}(M + M^T)$. The point of this formula is that it replaces a probabilistic construction of the permanental vector by an explicit oscillatory integral. This representation is particularly useful because integration by parts with respect to the twisted Gaussian weight produces Ward identities. These identities are the main algebraic mechanism behind the isomorphism theorems proved later in the chapter.

Lemma 2.2.1 (Density of 1-permanental vectors)**.** *Let $G \in M_n(\mathbb{R})$ have a positive definite symmetric part. If there exists a 1-permanental process ℓ with kernel G, then the density $\rho(l)$ can be represented as follows:*

$$\rho(l) = \int_{(0,2\pi)^n} \frac{1}{(2\pi)^n |G|} \, e^{\sum_{i,j} Q_{ij} \sqrt{l_i l_j} e^{\mathrm{i}(\theta_i - \theta_j)}} d\theta. \tag{2.2}$$

Here $Q := -G^{-1}$. Also l_i denotes the i-th component of the vector l, and we write $d\theta$ for $\prod_i d\theta_i$. □

Remark 2.2.1.1. *It is straightforward to check that if $G \in M_n(\mathbb{R})$ has a positive definite symmetric part, then G is invertible.* □

Remark 2.2.1.2. *Lemma 2.2.1 is not completely new. It appeared in [1], but only in the Markovian setting (defined in Remark 2.2.1.4 below). Using the results derived by Brydges-Hofstad-König [20], our contribution is to generalize the density formula of [1] to any kernel with a positive definite symmetric part.* □

Remark 2.2.1.3 (Infinitely divisible kernels)**.** *By the definition of infinite divisibility, a 1-permanental vector ℓ is infinitely divisible if and only if, for every positive integer n, there exist n i.i.d. random vectors ℓ_1, ..., ℓ_n such that $\tilde{\ell} := \ell_1 + \cdots + \ell_n$ is equal in distribution to ℓ.*

We say that a kernel G of a 1-permanental vector is infinitely divisible if the corresponding 1-permanental vector ℓ is infinitely divisible. In [46], several characterizations are given of the family of kernels of infinitely divisible 1-permanental vectors. One such characterization is [46, Lemma 4.2], from which one can conclude that any kernel G of an infinitely divisible 1-permanental vector is diagonally equivalent to an inverse M-matrix, i.e., the inverse of an invertible M-matrix A. [1].

We will now prove that any infinitely divisible kernel G of a 1-permanental vector has an equivalent kernel $\tilde{G}$ with a positive definite symmetric part. To see this, fix some infinitely divisible kernel G of a 1-permanental vector. By [46, Lemma 4.2], it is diagonally equivalent to the inverse of a non-singular M-matrix A. Then, using [90, Condition I_{26}] for A, we know that there exists a diagonal matrix D with positive diagonal terms, such that $D^{-1}AD$ has a positive definite symmetric part. Therefore, using Lemma A.4.1 the inverse of $D^{-1}AD$ as $D^{-1}(A^{-1})D$ should also have a positive definite symmetric part. Since A^{-1} is diagonally equivalent to the kernel $D^{-1}(A^{-1})D$, and G is diagonally equivalent to A^{-1}, by transitivity of diagonal equivalence, G should also be diagonally equivalent to the matrix $D^{-1}(A^{-1})D$, which has a positive definite symmetric part. $\square$

Remark 2.2.1.4 (Markovian kernels)**.** *A sub-family of kernels of infinitely divisible 1-permanental vectors, which we call the Markovian kernels, is closely related to Markov chains and plays an important role in the study of local-time isomorphism theorems. Given an irreducible discrete-time Markov chain on the state space $[n]$, let us denote its transition kernel and stationary distribution by P and π, respectively, where π is thought of as a column vector. Let $\Pi = \mathrm{diag}(\pi)$ denote the diagonal matrix with diagonal terms given by π. The Laplacian associated with this Markov chain is defined as $Q = \Pi(P - I)$. Consider the continuous-time Markov process with rate matrix Q (whose stationary distribution would then be the uniform distribution). We introduce a cemetery state Δ, and a nonnegative killing rates given by $\boldsymbol{h} = (h_1, \cdots, h_n) \geq 0$. Namely, when in state $i \in [n]$, there is now a rate h_i of jumping to the cemetery state. The rate matrix associated with the corresponding continuous-time killed Markov process is defined to be $Q_{\boldsymbol{h}} = Q - \mathrm{diag}(\boldsymbol{h})$. It can be checked that $Q_{\boldsymbol{h}}$ is invertible when $\sum_i h_i > 0$, and we can define the Green's kernel associated with it as $G_{\boldsymbol{h}} := -Q_{\boldsymbol{h}}^{-1}$. Any matrix $G_{\boldsymbol{h}}$ obtained from such a construction is called a Markovian kernel.*

We claim that $G_{\boldsymbol{h}}$ will have a positive definite symmetric part. To see this, let $P^* := \Pi^{-1}P^T\Pi$ be the time reversal of P. Note that P^* is a stochastic matrix. The additive-reversibilization of P is defined to be the Markov chain with transition matrix $M := \frac{1}{2}(P + P^*)$, see [5, Lemma 9.20]. Note that π is the stationary distribution of M, since $\pi^T \frac{P+P^*}{2} =$

[1] If G is a kernel of an infinitely divisible 1-permanental vector then there should be a $\frac{1}{2}$-permanental vector ℓ with kernel G. This random vector ℓ must be infinitely divisible. On the other hand, if ℓ is a $\frac{1}{2}$-permanental vector with kernel G such that ℓ is infinitely divisible, then the corresponding 1-permanental vector should also be infinitely divisible, and hence G is an infinitely divisible kernel. According to [46, Lemma 4.2], if a $\frac{1}{2}$-permanental vector G is infinitely divisible, then its kernel is diagonally equivalent to an inverse M-matrix. Therefore, if G is an infinitely divisible kernel for a 1-permanental vector, it should be diagonally equivalent to an inverse M-matrix.

$\frac{1}{2}\pi^T + \frac{1}{2}\pi^T\Pi^{-1}P^T\Pi = \pi^T$. *Further, M is reversible, since* $\Pi M = \frac{1}{2}(\Pi P + P^T\Pi) = M^T\Pi$. *This Markov chain is clearly irreducible, and its Laplacian is $\Pi(M-I)$, which we denote by S. Since S is the Laplacian of a reversible Markov chain, $-S$ is nonnegative definite. This can be seen by observing that S has the same eigenvalues as $\Pi^{\frac{1}{2}}S\Pi^{-\frac{1}{2}}$, which is symmetric, and so has real eigenvalues, while S, as the rate matrix of an irreducible continuous-time Markov chain, has 0 as a simple eigenvalue and all other eigenvalues having strictly negative real parts.*

By the Perron-Frobenius theorem, we know that $\langle v, -Sv\rangle = 0$ if $v \in \mathrm{span}(\mathbf{1})$, and $\langle v, -Sv\rangle > 0$ otherwise. But if $v \in \mathrm{span}(\mathbf{1})$ and $v \neq 0$, then $v^T\mathrm{diag}(\boldsymbol{h})v > 0$. Hence, for any v, we have $\langle v, (\mathrm{diag}(\boldsymbol{h}) - S)v\rangle > 0$ if $v \neq 0$, i.e. $\mathrm{diag}(\boldsymbol{h}) - S$ is positive definite, i.e. $-\frac{1}{2}(Q_{\boldsymbol{h}} + Q_{\boldsymbol{h}}^T)$ is positive definite. By Lemma A.4.1, we know that $-Q_{\boldsymbol{h}}$ and hence $G_{\boldsymbol{h}}$ has a positive definite symmetric part.

It is apparent that $G_{\boldsymbol{h}}^{-1} = -Q_{\boldsymbol{h}} = \mathrm{diag}(\boldsymbol{h}) + \Pi - \Pi P$ has non-positive off-diagonal terms. We also showed that it has a positive definite symmetric part. Then we can use [90, Condition I_{26}] to see that $-Q_{\boldsymbol{h}}$ is indeed a non-singular M-matrix. Therefore, $G_{\boldsymbol{h}}$ is an inverse M-matrix and hence infinitely divisible by [46, Lemma 4.2]. □

Remark 2.2.1.5 (Positive definite kernels)**.** *Let $G \in M_n(\mathbb{R})$ be a positive definite matrix. Let $\eta \sim N(0, G)$. One can see from the moment generating function that $\frac{1}{2}\eta^2$ is $\frac{1}{2}$-permanental with kernel G. Since the sum of two independent copies of a $\frac{1}{2}$-permanental random vector with a given kernel is a 1-permanental vector with that kernel, it follows that for every positive definite matrix $G \in M_n(\mathbb{R})$, there is a 1-permanental vector with kernel G.* □

Remark 2.2.1.6. *One might wonder if the condition that G should have a positive definite symmetric part is too stringent. It is obvious that any positive definite and symmetric (PSD) matrix satisfies this condition. Less obvious is that infinitely divisible kernels also fall into this category up to diagonal equivalence (see Remark Remark 2.2.1.3).*

In his classic 1967 paper, Vere-Jones [105] gave necessary and sufficient conditions on the kernel for the corresponding 1-permanental vector to exist. However, these conditions are not easy to check. These conditions are known to hold for matrices that are either diagonally equivalent to a PSD matrix or are infinitely divisible [45]. We already showed in Remark 2.2.1.3 that any infinitely divisible kernel G is diagonally equivalent to another kernel $\tilde{G}$ with a positive definite symmetric part. Therefore, both scenarios are included in Lemma 2.2.1 up to diagonal equivalence, in the sense that given any 1-permanental vector with such a kernel G, we can find an equivalent kernel $\tilde{G}$ that has a positive definite symmetric part and apply Lemma 2.2.1 to that kernel. We also remark that these two families of kernels are not inclusive of one another.

To the best of our knowledge, the only known instances of 1-permanental vectors whose kernel is neither diagonally equivalent to a PSD matrix nor an infinitely divisible one are the examples constructed in [48, Theorem 7.3], but their construction is very delicate, hence this *family of kernels is rather restricted. We remark that it is an important open problem to give an algorithmically checkable criterion that characterizes the set of 1-permanental kernels.* □

In the discussion following the proof of Lemma 2.2.1 below, and in the rest of the chapter, we say that a function $f:\mathbb{C}^n\to\mathbb{C}$ has subexponential growth rate if $\lim_{\|\phi\|\to\infty}\frac{\log(1+|f(\phi)|)}{\|\phi\|}=0$. We say that such a function is sub-Gaussian if there exists some $c,c'>0$, such that $|f(\phi)|\le ce^{-c'\|\phi\|^2}$. Similar definitions apply to functions $f:\mathbb{R}_+^n\to\mathbb{C}$. We now proceed to prove Lemma 2.2.1.

Proof of Lemma 2.2.1. This proof mimics the procedure of Brydges-Hofstad-König [20]. By the definition of a 1-permanental vector, for any $\lambda\in\mathbb{R}_+^n$, we have the Laplace transform

$$\mathbb{E}\big[e^{-\langle\lambda,\ell\rangle}\big]=\frac{1}{|I+\Lambda G|}=\frac{1}{|G||\Lambda-Q|}.$$

Here we used $I+\Lambda G=(G^{-1}+\Lambda)G=(\Lambda-Q)G$. Note that $|G|\neq 0$ by Remark 2.2.1.1.

Under the conditions assumed in Lemma 2.2.1, we can write $\frac{1}{2}(G+G^T)=-\frac{1}{2}(Q^{-1}+Q^{-T})=-\frac{1}{2}Q^{-1}(Q+Q^T)Q^{-T}$, from which we see that, since G is assumed to have a positive definite symmetric part, so does $-Q$, and hence so does $\Lambda-Q$. Now, consider n complex numbers $\phi_i=u_i+\mathrm{i}v_i$, $\forall i\in[n]$. Using the formula in [20, Lemma 2.2] for $\int e^{\langle\phi,M\bar\phi\rangle}du\,dv$ for the matrix $M=\Lambda-Q$, we have

$$\mathbb{E}\big[e^{-\langle\lambda,\ell\rangle}\big]=\int_{\mathbb{R}^{2n}}\frac{1}{\pi^n|G|}\,e^{-\langle\phi,(\Lambda-Q)\bar\phi\rangle}du\,dv=\int_{\mathbb{R}^{2n}}\frac{1}{\pi^n|G|}\,e^{-\langle\phi,\Lambda\bar\phi\rangle+\langle\phi,Q\bar\phi\rangle}du\,dv.$$

Using the change-of-variable $\phi_i=\sqrt{l_i}\,e^{\mathrm{i}\theta_i}$ where $l_i\in(0,\infty),\theta_i\in[0,2\pi)$ for all $i\in[n]$ and $\phi_i\neq 0$ gives

$$\begin{aligned}\mathbb{E}\big[e^{-\langle\lambda,\ell\rangle}\big]&=\int_{\mathbb{R}_+^n}\int_{[0,2\pi)^n}\frac{1}{(2\pi)^n|G|}\,e^{-\langle\lambda,l\rangle+\sum_{i,j}Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i-\theta_j)}}d\theta\,dl\\&=\int_{\mathbb{R}_+^n}e^{-\langle\lambda,l\rangle}\bigg(\int_{[0,2\pi)^n}\frac{1}{(2\pi)^n|G|}\,e^{\sum_{i,j}Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i-\theta_j)}}d\theta\bigg)dl,\end{aligned}$$

where we have used the fact that $du\,dv=2^{-n}d\theta\,dl$. Let us define the measure $\rho(l)$ as

$$\rho(l)=\int_{[0,2\pi)^n}\frac{1}{(2\pi)^n|G|}\,e^{\sum_{i,j}Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i-\theta_j)}}d\theta.$$

We claim that $\rho(l)$ is finite. To see this, by the triangle inequality, we have

$$\begin{aligned}|\rho(l)|&\le\int_{[0,2\pi)^n}\frac{1}{(2\pi)^n|G|}\Big|e^{\sum_{i,j}Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i-\theta_j)}}\Big|d\theta\\&=\int_{[0,2\pi)^n}\frac{1}{(2\pi)^n|G|}\,e^{\sum_{i,j}S_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i-\theta_j)}}d\theta,\end{aligned}\tag{2.3}$$

where $S := \frac{1}{2}(Q + Q^T)$ denotes the symmetric part of Q. Here, we used the fact that $|G| > 0$ is necessary for the existence of a 1-permanental vector whose kernel has a positive definite symmetric part. (To see this, let $\Lambda = tI$ and take $t > 0$ large enough; then the positivity of the Laplace transform $\mathbb{E}\big[e^{-\langle\lambda,\ell\rangle}\big]$ implies that $|G| > 0$.) We have observed earlier that $-S$ has a positive definite symmetric part. Let $\lambda_{\min}(-S) > 0$ be the minimum eigenvalue of $-S$. Then we know by spectral bounds that

$$|\rho(l)| \leq \int_{[0,2\pi)^n} \frac{1}{(2\pi)^n|G|}\, e^{-\lambda_{\min}(-S)\left(\sum_i l_i\right)} d\theta = \frac{1}{|G|}\, e^{-\lambda_{\min}(-S)\left(\sum_i l_i\right)} < \infty.$$

This shows that $\rho(l)$ is indeed finite. Moreover, since

$$\int_{\mathbb{R}_+^n} |\rho(l)| dl \leq \int_{\mathbb{R}_+^n} \frac{1}{|G|}\, e^{-\lambda_{\min}(-S)\left(\sum_i l_i\right)} dl = \frac{1}{|G|(\lambda_{\min}(-S))^n} < \infty,$$

we have $\rho \in L^1(\mathbb{R}_+^n)$. Finally, to see that $\rho(l)$ is in fact real, note that

$$\begin{aligned}
\overline{\rho(l)} &= \int_{[0,2\pi)^n} \frac{1}{(2\pi)^n|G|}\, e^{\sum_{i,j} Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{-\mathrm{i}(\theta_i-\theta_j)}} d\theta \\
&= \int_{[0,2\pi)^n} \frac{1}{(2\pi)^n|G|}\, e^{\sum_{i,j} Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}((-\theta_i)-(-\theta_j))}} (-1)^n d(-\theta_1)\cdots d(-\theta_n) \\
&= \int_{(-2\pi,0]^n} \frac{1}{(2\pi)^n|G|}\, e^{\sum_{i,j} Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i'-\theta_j')}} d\theta_1'\cdots d\theta_n' \\
&= \int_{[0,2\pi)^n} \frac{1}{(2\pi)^n|G|}\, e^{\sum_{i,j} Q_{ij}\sqrt{l_i}\sqrt{l_j}e^{\mathrm{i}(\theta_i'-\theta_j')}} d\theta_1'\cdots d\theta_n' \\
&= \rho(l),
\end{aligned}$$

where in the third equality we used the substitution $\theta' = -\theta$. By the invertibility of the Laplace transform given the region of convergence, $\rho(l)$ is exactly the density of a 1-permanental random vector with kernel G, if such a vector exists. □

From the proof above, we know that for any $G \in M_n(\mathbb{R})$ with a positive definite symmetric part, $\rho(l)$ is a well-defined real density in $L^1(\mathbb{R}_+^n)$. Moreover, from its definition, we see that $\rho(l)$ is infinitely differentiable in the interior of $\mathbb{R}_+^n$. [2] Therefore, the existence of a 1-permanental vector with kernel G is equivalent to the nonnegativity of the density $\rho(l)$.

We remark that Lemma 2.2.1 is implicitly in [1] and [52], but only in the Markovian setting. To clarify this, we need to define the Markovian loop soups.

For a transient Markov process $\{X_t : t \geq 0\}$ on discrete state space V, and any $t > 0$, let $\Omega_{r,t}$ be the space of right-continuous functions $[0,t] \to V$, with finitely many jumps and the same value at 0 and t. Also, let $\Omega_r := \cup_{t>0}\Omega_{r,t}$. We call this the space of rooted loops.

[2] The integrand in Eq. (2.2), after being differentiated for a finite number of times, would still be integrable, due to essentially the same reasoning as used in the proof of Lemma 2.2.1.

For any $\phi \in \Omega_r$, let $\zeta(\phi)$ be the unique $t > 0$ such that $\phi \in \Omega_{r,t}$. We identify Ω_r with $\Omega_{r,1} \times (0, \infty)$ via the map $(w, t) \in \Omega_{r,1} \times (0, \infty) \to \phi(\cdot) = w(\frac{\cdot}{t}) \in \Omega_r$, and endow $\Omega_{r,1} \times (0, \infty)$ with the canonical product σ-algebra (where $\Omega_{r,1}$ is endowed with the σ-algebra generated by the maps $X_s, 0 \leq s \leq 1$, from $\Omega_{r,1}$ into V). For any $x \in V$, define $P^t_{x,x}$ on $\Omega_{r,t}$ such that for any measurable subset $B_t \subset \Omega_{r,t}$,

$$P^t_{x,x}(B_t) = P(B_t \cap \{X_t = x\} \mid X_0 = x).$$

We then define the rooted loop measure μ_r on Ω_r such that for any measurable subset $B \subset \Omega_r$,

$$\mu_r(B) := \sum_{x \in V} \int_0^\infty P^t_{x,x}(B) \frac{dt}{t}.$$

The shift operator θ_s acts on a rooted loop $\phi \in L_{r,t}$ by cyclically rotating the loop by time s. The set of unrooted loops, Ω_*, is the set of equivalence classes of rooted loops Ω_r under the equivalence relationship defined by the shifts $\{\theta_s : s > 0\}$. The unrooted loop measure μ_* on Ω_* is then induced by the rooted loop measure under this equivalence relationship.

An unrooted loop soup (or Poisson gas of Markovian loops), at level $\alpha > 0$, is then a Poisson point process on the space Ω_* of unrooted loops, with intensity measure $\alpha\mu_*$. We will be mostly interested in loop soups at level $\alpha = 1$. Therefore, we will always assume the level is at 1 unless stated otherwise. Let $\Phi = \{\phi^*_i : i \in I\}$ be a countable collection of loops generated via such Poisson point process. Then we can define the loop soup occupation times (or occupation field) as $\ell^x(\Phi) := \sum_{i \in I} T^x(\phi_i)$ for any $x \in V$. Here $T^x(\phi)$ represents the local time the loop ϕ spends at state $x \in V$.

It is shown in Fitzsimmons and Rosen [52] that the Markovian loop soup local time ℓ is equal in distribution to a 1-permanental vector with the corresponding Green's kernel, by using the generalized Le Jan's isomorphism theorem (non-symmetric version, Theorem 3.1 in [52]). Moreover, Theorem 2.15(a) in [1] states that for any bounded and continuous function $F : \mathbb{R}^n_+ \to \mathbb{C}$, we have $\int \frac{1}{(2\pi \mathrm{i})^n |G|} F(\phi\bar{\phi}) e^{\langle \phi, Q\bar{\phi}\rangle} d\bar{\phi} d\phi = \mathbb{E}F(\ell)$, where ℓ is the Markovian loop soup local time with intensity $\alpha = 1$. This provides a connection between the twisted Gaussian density and the Markovian loop soup local time: using the result by Fitzsimmons and Rosen [52], we know that $\int \frac{1}{(2\pi \mathrm{i})^n |G|} F(\phi\bar{\phi}) e^{\langle \phi, Q\bar{\phi}\rangle} d\bar{\phi} d\phi = \mathbb{E}F(\ell)$ for ℓ being a 1-permanental vector with a Markovian kernel G as well. A polar coordinate change then reveals the density of the 1-permanental vector, in the Markovian case, in the form given in Lemma 2.2.1.

In contrast, our density formula also works for non-Markovian kernels with a positive definite symmetric part, and for this we critically used the results derived by Brydges-Hofstad-König[20] in the proof above.

Remark 2.2.1.7. *A natural consequence of the above density formula is a symmetrization bound for the density. Let $\iota := \frac{|2G|}{|G+G^T|}$, and denote by $\tilde{\rho}(\bar{l})$ the density of a 1-permanental vector $\bar{\ell}$ with the symmetric kernel $(-S)^{-1}$, where $S = \frac{1}{2}(Q + Q^T)$, with $Q := -G^{-1}$. In the same setting as the previous lemma, it is immediate from its proof that*

$$\rho(l) \leq \iota \bar{\rho}(l),$$

see equation (2.3)*.*

This symmetrization bound will be used several times in the later sections. Similarly, one concludes that for any $\mathbb{C}$*-valued function* f *with a subexponential growth rate, we have* $|\mathbb{E}[f(\ell)]| \leq \iota\mathbb{E}[|f(\bar{\ell})|]$*:*

$$|\mathbb{E}[f(\ell)]| = \left| \int_{\mathbb{R}_+^n} f(l)\rho(l)dl \right| \leq \int_{\mathbb{R}_+^n} |f(l)|\rho(l)dl \leq \iota \int_{\mathbb{R}_+^n} |f(l)|\bar{\rho}(l)dl = \iota\mathbb{E}[|f(\bar{\ell})|].$$

The integrand above is absolutely integrable for any function of subexponential growth since $\bar{\rho}(l)$ *decays exponentially fast in* $\|l\|$*. Since* $\rho(l) \leq \iota\bar{\rho}(l)$ *and both* ρ *and* $\bar{\rho}$ *integrate to* 1*, integrating gives* $1 \leq \iota$*.*

In linear algebra language, for any matrix G *with a positive definite symmetric part, we always have* $|\frac{1}{2}(G + G^T)| \leq |G|$*. This is a well-known fact called the Ostrowski-Taussky inequality in the literature [87].*

Another consequence of the proof above is the subexponentiality of 1*-permanental vectors. In the same setting as Lemma 2.2.1, for the* 1*-permanental vector* ℓ*, we know that*

$$\mathbb{P}(\ell \geq t) \leq \frac{1}{|G|} \int_{\prod_i [t_i,\infty)} e^{-\lambda_{\min}(-S)\|l\|_1} dl = \frac{1}{|G|(\lambda_{\min}(-S))^n} e^{-\lambda_{\min}(-S)\|t\|_1}.$$

Since for any $k \in \mathbb{Z}_+$*, a* k*-permanental process with kernel* G *is equal in distribution to* $\sum_{i=1}^k \ell_i$*, where* ℓ_i*'s are i.i.d. 1-permanental vectors with kernel* G*, we also have*

$$\mathbb{P}\Big(\sum_{i=1}^k \ell_i \geq t \Big) \leq k\,\mathbb{P}\Big(\ell \geq \frac{1}{k} t \Big) \leq c\, e^{-\frac{1}{k}\lambda_{\min}(-S)\|t\|_1}.$$

Here, $c > 0$ *does not depend on* t*. The subexponentiality of* k*-permanental vectors will also be used in later sections.* □

Density of conditioned 1-permanental vectors

The density formula allows us to study the density of a 1-permanental vector ℓ conditioned on some coordinate being equal to a fixed value r. Without loss of generality, we will consider $(\ell_2, \cdots, \ell_n \,|\, \ell_1 = r)$. Let us also assume for now that $r > 0$, the case $r = 0$ being trivial, see Remark 2.2.2.1 below.

To derive the conditional density $\rho_r(l_2, \cdots, l_n)$ of the conditioned process $(\ell_2, \cdots, \ell_n \,|\, \ell_1 = r)$, let us decompose the Laplacian as

$$Q = \begin{pmatrix} Q_{11} & Q_{1*}^T \\ Q_{*1} & Q_{**} \end{pmatrix}. \tag{2.4}$$

Similarly, we let $\phi_* := (\phi_2, \cdots, \phi_n)$, and, in general, whenever we use the underscript $*$, we mean the original vector or matrix restricted to $[n]\backslash\{1\}$. We then have the following density formula, for which the rotation symmetry in θ_i's plays an important role in the proof.

Lemma 2.2.2 (Density of conditioned 1-permanental vector)**.** *In the same setting as Lemma 2.2.1, the conditional density $\rho_r(l_2, \cdots, l_n)$ of the conditioned process $(\ell_2, \cdots, \ell_n \,|\, \ell_1 = r)$ is*

$$\rho_r(l_2, \cdots, l_n) = \int_{(0,2\pi)^{n-1}} \frac{|-Q_{**}|}{(2\pi)^{n-1}} \exp\Big(\langle(\phi_* + \sqrt{r}Q_{**}^{-T}Q_{1*}), Q_{**}(\bar{\phi}_* + \sqrt{r}Q_{**}^{-1}Q_{*1})\rangle\Big) d\theta_*.$$

Proof. First, plug in $l_1 = r$ in the density formula for ℓ. We know $\rho(r, l_2, \cdots, l_n)$ equals to

$$\int_{(0,2\pi)^n} \frac{e^{Q_{11}r}}{(2\pi)^n|G|} \exp\Bigg(\sum_{i,j\neq 1} Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)} + \sum_{i\neq 1}\big(Q_{i1}\sqrt{l_ir}e^{\mathrm{i}(\theta_i-\theta_1)} + Q_{1i}\sqrt{l_ir}e^{\mathrm{i}(\theta_1-\theta_i)}\big)\Bigg) d\theta.$$

Now use the change of measure $\theta_i' = \theta_i - \theta_1, \forall i \neq 1$ and $\theta_1' = \theta_1$. We know it's further equal to

$$\int_{(0,2\pi)}\int_{(-\theta_1',-\theta_1'+2\pi)^{n-1}} \frac{e^{Q_{11}r}}{(2\pi)^n|G|} \exp\Bigg(\sum_{i,j\neq 1} Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i'-\theta_j')} + \sum_{i\neq 1}\big(Q_{i1}\sqrt{l_ir}e^{\mathrm{i}\theta_i'} + Q_{1i}\sqrt{l_ir}e^{-\mathrm{i}\theta_i'}\big)\Bigg) d\theta_*' d\theta_1'.$$

But, given any θ_1', the inner integral is invariant whether it's integrated over $(-\theta_1', -\theta_1' + 2\pi)^{n-1}$ or over $(0, 2\pi)^{n-1}$, due to rotation symmetry. Hence, we conclude that

$$\begin{aligned}
&\rho(r, l_2, \cdots, l_n)\\
&\quad= \int_{(0,2\pi)}\int_{(0,2\pi)^{n-1}} \frac{e^{Q_{11}r}}{(2\pi)^n|G|} \exp\Bigg(\sum_{i,j\neq 1} Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i'-\theta_j')} + \sum_{i\neq 1}\big(Q_{i1}\sqrt{l_ir}e^{\mathrm{i}\theta_i'} + Q_{1i}\sqrt{l_ir}e^{-\mathrm{i}\theta_i'}\big)\Bigg) d\theta_*' d\theta_1'\\
&\quad= \int_{(0,2\pi)^{n-1}} \frac{e^{Q_{11}r}}{(2\pi)^{n-1}|G|} \exp\Bigg(\sum_{i,j\neq 1} Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i'-\theta_j')} + \sum_{i\neq 1}\big(Q_{i1}\sqrt{l_ir}e^{\mathrm{i}\theta_i'} + Q_{1i}\sqrt{l_ir}e^{-\mathrm{i}\theta_i'}\big)\Bigg) d\theta_*'.
\end{aligned}$$

Marginally, ℓ_1 is distributed as an exponential random variable with mean G_{11} as is shown in [47]. Hence, we conclude that

$$\begin{aligned}
\rho_r(l_2, \cdots, l_n) &= G_{11}e^{G_{11}^{-1}r}\rho(r, l_2, \cdots, l_n)\\
&= \int_{(0,2\pi)^{n-1}} \frac{G_{11}e^{(Q_{11}+G_{11}^{-1})r}}{(2\pi)^{n-1}|G|} \exp\Bigg(\sum_{i,j\neq 1} Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)} + \sum_{i\neq 1}\big(Q_{i1}\sqrt{l_ir}e^{\mathrm{i}\theta_i} + Q_{1i}\sqrt{l_ir}e^{-\mathrm{i}\theta_i}\big)\Bigg) d\theta_*.
\end{aligned}$$

By the decomposition of the Laplacian Q, we have

$$\rho_r(l_2,\cdots,l_n)=\int_{(0,2\pi)^{n-1}}\frac{G_{11}e^{(Q_{11}+G_{11}^{-1})r}}{(2\pi)^{n-1}|G|}\exp\Big(\langle\phi_*,Q_{**}\bar{\phi}_*\rangle+\sqrt{r}\big(\langle Q_{*1},\phi_*\rangle+\langle Q_{1*},\bar{\phi}_*\rangle\big)\Big)d\theta_*.$$

We claim that Q_{**} is invertible. To see this, note that by the adjugate formula $G=|-Q|^{-1}\mathrm{adj}(-Q)$, we have $G_{11}=\frac{|-Q_{**}|}{|-Q|}=|G||-Q_{**}|$. But $G_{11}>0$ and $|G|>0$ since $|G|$ is 1-permanental, so we conclude that $|-Q_{**}|>0$ and Q_{**} is indeed invertible. Therefore we have

$$\begin{aligned}&\rho_r(l_2,\cdots,l_n)\\&=\int_{(0,2\pi)^{n-1}}\frac{G_{11}e^{(Q_{11}+G_{11}^{-1})r}}{(2\pi)^{n-1}|G|}\exp\Big(\langle\phi_*,Q_{**}\bar{\phi}_*\rangle+\big(\langle Q_{**}^T\phi_*,\sqrt{r}Q_{**}^{-1}Q_{*1}\rangle\\&\qquad\qquad\qquad\qquad\qquad\qquad+\langle\sqrt{r}Q_{**}^{-T}Q_{1*},Q_{**}\bar{\phi}_*\rangle\big)\Big)d\theta_*\\&=\int_{(0,2\pi)^{n-1}}\frac{G_{11}e^{(Q_{11}+G_{11}^{-1}-Q_{1*}^TQ_{**}^{-1}Q_{*1})r}}{(2\pi)^{n-1}|G|}\exp\big(\langle(\phi_*+\sqrt{r}Q_{**}^{-T}Q_{1*}),Q_{**}(\bar{\phi}_*+\sqrt{r}Q_{**}^{-1}Q_{*1})\rangle\big)d\theta_*.\end{aligned}$$

Now, by the block matrix inverse formula, if Q_{**} is invertible we have

$$G_{11}=(-Q_{11}+Q_{1*}^TQ_{**}^{-1}Q_{*1})^{-1}.$$

Therefore, $Q_{11}+G_{11}^{-1}-Q_{1*}^TQ_{**}^{-1}Q_{*1}=0$ and we conclude that

$$\rho_r(l_2,\cdots,l_n)=\int_{(0,2\pi)^{n-1}}\frac{|-Q_{**}|}{(2\pi)^{n-1}}\exp\big(\langle(\phi_*+\sqrt{r}Q_{**}^{-T}Q_{1*}),Q_{**}(\bar{\phi}_*+\sqrt{r}Q_{**}^{-1}Q_{*1})\rangle\big)d\theta_*.$$

One can apply the same arguments as those in the proof of Lemma 2.2.1 to see that the integrand is absolutely integrable. Hence, the lemma is proved. □

Remark 2.2.2.1. *We remark that when $r=0$, we have*

$$\rho_0(l_2,\cdots,l_n)=\int_{(0,2\pi)^{n-1}}\frac{|-Q_{**}|}{(2\pi)^{n-1}}\exp\big(\langle\phi_*,Q_{**}\bar{\phi}_*\rangle\big)d\theta_*.$$

Computing the Laplace transform, we see that the conditioned 1-permanental process

$$\{\ell_2,\cdots,\ell_n\,|\,\ell_1=0\}$$

should again be a 1-permanental process with kernel $(-Q_{**})^{-1}$, which has positive definite symmetric part.

In the Markovian setting, this property has the following loop soup interpretation. A loop soup conditioned on not visiting some node a is just an unconditioned loop soup on the loops that do not visit a, by the independence property of the Poissonian point process. This observation is mentioned in several previous works on loop soup (for example, [97, 52]). However, it is not a priori obvious that every 1-permanental G with a positive definite symmetric part will also satisfy a similar property, which is what we have established. □

Another interesting observation is that when G is positive definite we have the following conditional law for squared Gaussian vectors.

Corollary 2.2.3 (Conditioned χ-squared process)**.** *For any non-degenerate Gaussian vector $\tilde{U} \sim \mathcal{N}(0, G)$, let $U \sim \mathcal{N}(-\sqrt{r}Q_{**}^{-1}Q_{*1}, (-Q_{**})^{-1})$. We have*

$$(\tilde{U}_*^2 \,|\, \tilde{U}_1^2 = r) \overset{d.}{=} U^2.$$

Proof. By Lemma 2.2.2, in the Gaussian case, we have

$$\begin{aligned}
&\rho_r(l_2, \cdots, l_n) \\
&= \int_{(0,2\pi)^{n-1}} \frac{|-Q_{**}|}{(2\pi)^{n-1}} \exp\left(\left[\left((u_* + \sqrt{r}Q_{**}^{-1}Q_{*1}) + \mathrm{i}v_* \right), Q_{**}\left((u_* + \sqrt{r}Q_{**}^{-1}Q_{*1}) - \mathrm{i}v_* \right) \right]_\mu \right) d\theta_* \\
&= \int_{(0,2\pi)^{n-1}} \frac{|-Q_{**}|}{(2\pi)^{n-1}} \exp\left(\left[(u_* + \sqrt{r}Q_{**}^{-1}Q_{*1}), Q_{**}(u_* + \sqrt{r}Q_{**}^{-1}Q_{*1}) \right]_\mu + \langle v_*, Q_{**}v_* \rangle \right) d\theta_*.
\end{aligned}$$

This is exactly the law of $\left\{(U - \sqrt{r}Q_{**}^{-1}Q_{*1})^2 + V^2\right\}$, for U, V being independent $\mathcal{N}(0, \frac{1}{2}(-Q_{**})^{-1})$ vectors. Since we know that $\ell \overset{d.}{=} \{\tilde{U}^2 + \tilde{V}^2 : i \in [n]\}$, where $\tilde{U}, \tilde{V}$ are independent $\mathcal{N}(0, \frac{1}{2}G)$ vectors, we have

$$\{\ell_* \,|\, \ell_1 = r\} \overset{d.}{=} \{\tilde{U}_*^2 + \tilde{V}_*^2 \,|\, \tilde{U}_1^2 + \tilde{V}_1^2 = r\} \overset{d.}{=} \left\{(U - \sqrt{r}Q_{**}^{-1}Q_{*1})^2 + V^2\right\}.$$

We claim that

$$\{\tilde{U}_*^2 + \tilde{V}_*^2 \,|\, \tilde{U}_1^2 + \tilde{V}_1^2 = r\} \overset{d.}{=} \{\tilde{U}_*^2 + \tilde{V}_*^2 \,|\, \tilde{U}_1^2 = r, \tilde{V}_1^2 = 0\}.$$

To see this, it suffices to show that, conditioned on $\ell_1 = r$, ℓ_* is independent of θ_1, and therefore independent of (U_1, V_1). This is readily concluded from the proof of Lemma 2.2.2. Indeed, fixing any θ_1, the inner integral there does not depend on θ_1. Now note that $\{\tilde{V}_*^2 \,|\, \tilde{V}_1^2 = 0\} \overset{d.}{=} V^2$. It is immediate that

$$\{\tilde{U}_*^2 \,|\, \tilde{U}_1^2 = r\} \overset{d.}{=} \left\{(U - \sqrt{r}Q_{**}^{-1}Q_{*1})^2\right\}.$$

□

A special version of this corollary, applying to a certain family of Gaussian free fields, was used earlier in studying isomorphism theorems [97]. Proving this theorem by directly working with the χ-squared process seems to be hard. Here, the rotation symmetry trick (i.e., Lemma 2.2.2) helps us circumvent this issue.

2.3 Non-reversible isomorphism theorems

In the Markovian setting, the 1-permanental process is equal in distribution to the loop soup local time when the intensity of the soup is given by the loop measure (instead of 1/2 of the loop measure). In light of this, the following theorem can be seen as a generalization of Le Jan's isomorphism theorem that connects the loop soup local time with a certain Gaussian integral in the scenario when the intensity is 1/2 of the loop measure (see, e.g., Theorem 4.5 in [97]).

Theorem 2.3.1 (Non-reversible Le Jan's isomorphism theorem)**.** *In the same setting as Lemma 2.2.1, for any real function on $\mathbb{R}_+^n$ with a subexponential growth rate, we have*

$$\mathbb{E}[f(\ell)] = \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^n |G|} f(\phi\bar{\phi})\, e^{\langle \phi, Q\bar{\phi}\rangle} d\bar{\phi} d\phi. \tag{2.5}$$

Here $\phi\bar{\phi} := (\phi_i\bar{\phi}_i : i \in V)$. Moreover, for any $k \in \mathbb{N}_+$, the k-permanental vector $\ell^{(k)}$ satisfies

$$\mathbb{E}\big[f(\ell^{(k)})\big] = \int_{\mathbb{C}^{n\times k}} \frac{1}{(2\pi \mathrm{i})^{nk}|G|^k} f\big(\phi^{(1)}\overline{\phi^{(1)}} + \cdots + \phi^{(k)}\overline{\phi^{(k)}}\big) e^{\sum_{i=1}^k \langle \phi^{(i)}, Q\overline{\phi^{(i)}}\rangle} \prod_i d\overline{\phi^{(i)}} d\phi^{(i)}.$$

Proof. We know that

$$\begin{aligned}\mathbb{E}[f(\ell)] &= \int_{\mathbb{R}_+^n} \int_{(0,2\pi)^n} \frac{1}{(2\pi)^n |G|} f(l)\, e^{\sum_{i,j} Q_{ij}\sqrt{l_i}\sqrt{l_j} e^{\mathrm{i}(\theta_i - \theta_j)}} d\theta dl \\ &= \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^n |G|} f(\phi\bar{\phi})\, e^{\langle \phi, Q\bar{\phi}\rangle} d\bar{\phi} d\phi.\end{aligned}$$

Here we used change of variable $\phi_i = \sqrt{l_i} e^{\mathrm{i}\theta_i}$. Now, using the formula repeatedly and the fact that $\ell^{(k)}$ is distributed as the sum of k independent 1-permanental vectors proves the theorem. ∎

Using the distributional equivalence between a 1-permanental vector and loop soup local time when the intensity of the soup is given by the loop measure, the probabilistic content of Theorem 2.3.1 can be equivalently stated in terms of loop soup local times. When the Markov chain is reversible, the right-hand side of equation (2.5) becomes exactly complex Gaussian density. To see this connection, observe that we have $Q = Q^\top$ in the reversible case, and so $(\pi)^{-n} e^{\langle \phi, Q\bar{\phi}\rangle}$ can be written as a product of two Gaussians, as

$$\frac{1}{(\pi)^{\frac{n}{2}}} e^{\langle u, Qu\rangle} \frac{1}{(\pi)^{\frac{n}{2}}} e^{\langle v, Qv\rangle},$$

where $\phi = u + \mathrm{i}v$. The theorem therefore implies that $\phi\bar{\phi}$ where $\phi \sim \mathcal{CN}(0, G)$[3] and ℓ, the 1-permanental vector, are equal in distribution in this case. This special case is now seen to

[3]It is standard notation for a complex Gaussian that $\phi := u + \mathrm{i}v \sim \mathcal{CN}(0, \Gamma)$ iff

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \Re(\Gamma) & -\Im(\Gamma) \\ -\Im(\Gamma) & \Re(\Gamma) \end{pmatrix} \right).$$

be a consequence of the reversible Le Jan's isomorphism theorem, i.e., Theorem 4.5 in [97]. It is in this sense that Theorem 2.3.1 can be seen as a generalization to the non-reversible case of Theorem 4.5 in [97].

In the rest of this section, we are going to give a unified proof of several non-reversible isomorphism theorems. There will be two ingredients important in our proof: one is a pair of Ward identities for twisted Gaussian measure along the lines of those in [12], and the other is the previously derived density formula in equation (2.2). In this way, we obtain non-reversible versions of Dynkin's isomorphism theorem [44, 64], the generalized second Ray-Knight isomorphism theorem [46], and a non-reversible analogue of Eisenbaum's isomorphism theorem.

Ward identities for non-reversible Markov chains

Let P denote an irreducible transition probability matrix for a Markov chain on the state space $[n]$, and let the stationary distribution be denoted by π. Let $Q = \Pi(P - I)$ denote the normalized Laplacian, where Π denotes the diagonal matrix with diagonal entries given by π. We think of Q as the rate matrix of $\{X_s : s \geq 0\}$, a continuous-time irreducible Markov chain on the finite state space $[n]$, starting at $X_0 = a$. Define the local time of the Markov chain as

$$\mathcal{L}_t(i) = \int_0^t \mathbf{1}\{X_s = i\} ds. \tag{2.6}$$

For any $r \geq 0$, let the inverse local time be $\tau_{\text{inv}}(r) := \inf\{t : \mathcal{L}_t(a) \geq r\}$. We will consider the following unnormalized (since $-Q$ is only positive semi-definite here) twisted Gaussian measure

$$d\mu(\phi) = \pi^{-n} e^{\langle \phi, Q\bar{\phi} \rangle} dudv = (2\pi \mathrm{i})^{-n} e^{\langle \phi, Q\bar{\phi} \rangle} d\bar{\phi} d\phi.$$

Although the measure is unnormalized, when $g(\phi) : \mathbb{C}^n \to \mathbb{C}$ is sub-Gaussian, it is still integrable with respect to the measure $d\mu(\phi)$. To see this, note that

$$\left| [g(\phi)]_\mu \right| \leq \int_{\mathbb{R}^{2n}} \pi^{-n} e^{-c(\|u\|^2 + \|v\|^2)} e^{\langle \phi, S\bar{\phi} \rangle} dudv \leq \int_{\mathbb{R}^{2n}} \pi^{-n} e^{-c(\|u\|^2 + \|v\|^2)} dudv < \infty.$$

Here $[\cdot]_\mu$ denotes integration with respect to $d\mu(\phi)$ and

$$S = \frac{1}{2}(Q + Q^T) \tag{2.7}$$

denotes the symmetric part of Q, which we note satisfies $S \preceq 0$ in the Loewner partial order.

A function $f : \mathbb{C}^n \to \mathbb{C}$ or $f : \mathbb{R}^n_+ \to \mathbb{C}$ is said to have sub-polynomial tails if for every $d > 0$, there exists some constant $c_d > 0$ such that $|f(l)| \leq c_d \|l\|^{-d}$. We will dub the functions in $C^\infty(\mathbb{R}^n_+)$ or $C^\infty(\mathbb{C}^n)$ with sub-polynomial tails for all orders of derivatives as having property $\mathfrak{F}$. Similarly we say that a function $g : \mathbb{C}^n \to \mathbb{C}$ or $g : \mathbb{R}^n_+ \to \mathbb{C}$ has polynomial growth if there exist some $d > 0$, $b > 0$, and $c \in (0, \infty)$ such that $|g(\phi)| \leq b\|\phi\|^d$ for all ϕ that is large enough, i.e., $\|\phi\| > c$. We say that $g(\phi)$ in $C^\infty(\mathbb{R}^n_+)$ or $C^\infty(\mathbb{C}^n)$ has property $\mathfrak{G}$ if for any $k \in \mathbb{N}$, there exists some $d_k > 0$, $b_k > 0$ and $c_k \in (0, \infty)$ such that $|g^{(k)}(\phi)| \leq b_k\|\phi\|^{d_k}$ for all $\|\phi\| > c_k$, i.e. if every derivative of g has polynomial growth.

In the scenario where one has a killing rate $h > 0$ at each node, the Laplacian is replaced by $Q_h := Q - hI$. One can then consider the Green's function $G_h := (hI - Q)^{-1}$ since Q_h is invertible. Note that there will be a 1-permanental random variable with such a kernel [46].

Define the differential operator T_i as $T_i g(\phi) = \frac{\partial}{\partial \bar{\phi}_i} g(\phi)$. Similarly, define P_i as $P_i g(\phi) = \frac{\partial}{\partial \phi_i} g(\phi)$.[4] We also define $\partial_i f(l) := \frac{\partial}{\partial l_i} f(l)$ as differentiation in $l_i \in \mathbb{R}_+$. We use $\mathbb{E}_{i,l}$ to denote the expectation with respect to a Markov chain starting at state i with initial local time l. Similarly, let $\mathbb{E}^*_{i,l}$ denote expectation for the reversed chain starting at state i with initial local time l. We say that a function $f(i, l) : \mathcal{X} \times \mathbb{R}^n_+ \to \mathbb{C}$ has property $\mathfrak{F}$ if it has property $\mathfrak{F}$ in l for each fixed first component $i \in \mathcal{X}$.

The next lemma is the analytic engine of the chapter. It is an integration by parts identity for the twisted Gaussian measure. When combined with the density formula, it converts analytic identities in the ϕ-variables into distributional identities for permanental vectors and Markov-chain local times.

Lemma 2.3.2 (Ward identities for twisted Gaussian measure)**.** *Consider an irreducible continuous-time Markov chain on discrete space $[n]$ with Laplacian Q. For $f(i, l) : \mathcal{X} \times \mathbb{R}^n_+ \mapsto \mathbb{C}$ with property $\mathfrak{F}$ and $g(\phi) : \mathbb{C}^n \to \mathbb{C}$ with property $\mathfrak{G}$, we have*

$$\sum_i [\phi_i g(\phi) f(i, \phi\bar{\phi})]_\mu = \sum_i \left[(T_i g)(\phi) \int_0^\infty \mathbb{E}_{i,\phi\bar{\phi}} f(X_t, \mathcal{L}_t) dt \right]_\mu,$$

and

$$\sum_i [\bar{\phi}_i g(\phi) f(i, \phi\bar{\phi})]_\mu = \sum_i \left[(P_i g)(\phi) \int_0^\infty \mathbb{E}^*_{i,\phi\bar{\phi}} f(X_t, \mathcal{L}_t) dt \right]_\mu.$$

Proof. First note that[5]

$$\begin{aligned} 0 &= \int_{\mathbb{R}^{2n}} (2\pi \mathrm{i})^{-n} T_i (g(\phi) f(i, \phi\bar{\phi}) e^{\langle \phi, Q\bar{\phi} \rangle}) d\bar{\phi} d\phi \\ &= [(T_i g)(\phi) f(i, \phi\bar{\phi})]_\mu + [\phi_i g(\phi) \partial_i f(i, \phi\bar{\phi})]_\mu + [g(\phi)(Q^T\phi)_i f(i, \phi\bar{\phi})]_\mu. \end{aligned}$$

Summing over all $i \in [n]$. we have

$$\begin{aligned} \sum_i [(T_i g)(\phi) f(i, \phi\bar{\phi})]_\mu &= -\sum_i [\phi_i g(\phi) \partial_i f(i, \phi\bar{\phi})]_\mu - \sum_i [g(\phi)(Q^T\phi)_i f(i, \phi\bar{\phi})]_\mu \\ &= -\sum_i [\phi_i g(\phi) \partial_i f(i, \phi\bar{\phi})]_\mu - \sum_i [\phi_i g(\phi)(Qf)(i, \phi\bar{\phi})]_\mu. \end{aligned}$$

---

[4] We can take $\frac{\partial}{\partial \phi_i} := \frac{1}{2}\left(\frac{\partial}{\partial u_i} - \mathrm{i}\frac{\partial}{\partial v_i}\right)$ and $\frac{\partial}{\partial \bar{\phi}} := \frac{1}{2}\left(\frac{\partial}{\partial u_i} + \mathrm{i}\frac{\partial}{\partial v_i}\right)$.

[5] Note that the integrals on the right-hand side exist because of absolute integrability, since property $\mathfrak{F}$ of the function f and property $\mathfrak{G}$ of the function g guarantees that given any $d \in \mathbb{N}$, terms like $|(T_i g)(\phi) f(i, \phi\bar{\phi})|$ will be bounded by $r_d(1 + \|\phi\|)^{-d}$ for some constant $r_d > 0$.

Let $\mathcal{Q}$ be the generator of the joint process $(X_t, \mathcal{L}_t)$, which is a Markov process itself. It's not hard to see that we have $(\mathcal{Q}f)(i,l) = (\partial_i f + Qf)(i,l)$, and hence

$$\begin{aligned}\sum_i [(T_i g)(\phi) f(i, \phi\bar{\phi})]_\mu &= -\sum_i [\phi_i g(\phi)\,(\partial_i f + Qf)(i, \phi\bar{\phi}))]_\mu \\ &= -\sum_i [\phi_i g(\phi)\,(\mathcal{Q}f)(i, \phi\bar{\phi})]_\mu.\end{aligned}$$

Now we can use Komogorov's backward equation for the joint process $(X_t, \mathcal{L}_t)$, and deduce that $\mathcal{Q}f_t = \partial_t f_t$ for $f_t(i,l) := \mathbb{E}_{i,l}(f(X_t, \mathcal{L}_t))$. Hence letting $f = f_t$, we have

$$\begin{aligned}\int_0^\infty \sum_i [(T_i g)(\phi) f_t(i, \phi\bar{\phi})]_\mu dt &= -\int_0^\infty \sum_i [\phi_i g(\phi)\,(\mathcal{Q}f_t)(i, \phi\bar{\phi})]_\mu dt \\ &= -\int_0^\infty \sum_i [\phi_i g(\phi)\,(\partial_t f_t)(i, \phi\bar{\phi})]_\mu dt = -\sum_i [\phi_i g(\phi)\, f_t(i, \phi\bar{\phi})]_\mu \Big|_0^\infty.\end{aligned}$$

Now we have $\phi_i g(\phi) \mathbb{E}_{i,\phi\bar{\phi}}(f(X_t, \mathcal{L}_t)) \to 0$ almost surely. Since the event $\{\forall i \in \mathcal{X},\ \mathcal{L}_t(i) \to \infty\}$ happens almost surely, we know the r.h.s. is further equal to $\sum_i [\phi_i g(\phi)\, f(i, \phi\bar{\phi})]_\mu$. On the other hand, we know the l.h.s. is indeed $\sum_i \left[(T_i g)(\phi) \int_0^\infty \mathbb{E}_{i,\phi\bar{\phi}} f(X_t, \mathcal{L}_t) dt\right]_\mu$ by absolute integrability. The detailed proof of absolute integrability is deferred to the appendix (Section A.1).

If one starts with the following equation:

$$\begin{aligned}0 &= \int_{\mathbb{R}^{2n}} (2\pi \mathrm{i})^{-n} P_i (g(\phi) f(i, \phi\bar{\phi}) e^{\langle \phi, Q\bar{\phi}\rangle}) d\bar{\phi} d\phi \\ &= [(P_i g)(\phi) f(i, \phi\bar{\phi})]_\mu + [\bar{\phi}_i g(\phi) \partial_i f(i, \phi\bar{\phi})]_\mu + [g(\phi)(Q\bar{\phi})_i f(i, \phi\bar{\phi})]_\mu,\end{aligned}$$

then one would arrive at the second equation. Note that the reversed chain will be irreducible if the original chain is. □

Non-reversible Dynkin's isomorphism theorem

We have the following theorem, previously proved by Eisenbaum[44] and Le Jan[64], extending Dynkin's isomorphism theorem from the reversible case[97]. Here $\mathbb{E}_a$ is the expectation operator associated with the Markov chain started at state $a \in \mathcal{X}$.

Theorem 2.3.3 (non-reversible Dynkin's isomorphism theorem)**.** *Consider an irreducible continuous-time Markov chain on discrete space $[n]$ with Laplacian Q and killing rate $\boldsymbol{h} = (h_1, \cdots, h_n) \geq 0$ such that $\sum_i h_i > 0$. Let ℓ be a 1-permanental process with kernel $G_{\boldsymbol{h}} = (\mathrm{diag}(h) - Q)^{-1}$. For $u : \mathbb{R}_+^n \to \mathbb{C}$ with property $\mathfrak{F}$, we have for any $a \in [n]$,*

$$\mathbb{E}\left[\ell_a u(\ell)\right] = \int_0^\infty \mathbb{E}_a u(\mathcal{L}_t + \ell) \mathbf{1}_{X_t = a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt = \int_0^\infty \mathbb{E}_a^* u(\mathcal{L}_t + \ell) \mathbf{1}_{X_t = a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt.$$

We remark that in the equation above, the operator $\mathbb{E}_a$ also takes the expectation with respect to the 1-permanental process ℓ. Note that $G_{\boldsymbol{h}} := (\mathrm{diag}(\boldsymbol{h}) - Q)^{-1}$ is positive definite for any $\boldsymbol{h} \geq 0$, s.t. $\sum_i h_i > 0$. Hence, the 1-permanental vector indeed exists [46]. According to Theorem 2.3.1, for any real function on $\mathbb{R}_+^n$ with a subexponential growth rate, we have

$$\mathbb{E}[u(\ell)] = [u(\phi\bar{\phi})]_{\mu_{\boldsymbol{h}}}.$$

Here $[\cdot]_{\mu_{\boldsymbol{h}}}$ denotes integration with respect to $d\mu_{\boldsymbol{h}}(\phi) := \frac{1}{(2\pi\mathrm{i})^n |G_h|} e^{\langle \phi, Q_{\boldsymbol{h}}\bar{\phi}\rangle} d\bar{\phi} d\phi$, where $Q_{\boldsymbol{h}} := Q - \mathrm{diag}(\boldsymbol{h})$. Now we prove Theorem 2.3.3 above.

Proof. Take $g(\phi) = \bar{\phi}_b$, and $f(i,l) := v(l)\mathbf{1}_{i=a}$ for some function $v : \mathbb{R}_+^n \to \mathbb{C}$ that will be determined later in the first equation of Lemma 2.3.2, we obtain

$$\sum_i [\phi_i \bar{\phi}_b v(\phi\bar{\phi})\mathbf{1}_{i=a}]_\mu = \sum_i \left[(T_i g)(\phi) \int_0^\infty \mathbb{E}_{i,\phi\bar{\phi}} v(\mathcal{L}_t)\mathbf{1}_{X_t=a} dt \right]_\mu.$$

Note that $T_i g(\phi) = 1$ if $i = b$, and $T_i g(\phi) = 0$ otherwise. Also note $\mathbb{E}_{b,\phi\bar{\phi}} v(\mathcal{L}_t)\mathbf{1}_{X_t=a} = \mathbb{E}_b v(\mathcal{L}_t + \phi\bar{\phi})\mathbf{1}_{X_t=a}$. Simplifying the equation above, we have

$$[\phi_a \bar{\phi}_b v(\phi\bar{\phi})]_\mu = \left[\int_0^\infty \mathbb{E}_b v(\mathcal{L}_t + \phi\bar{\phi})\mathbf{1}_{X_t=a} dt \right]_\mu.$$

Now let $v(l) := u(l)e^{-\sum_i h_i l_i}$ in the equation above and get

$$[\phi_a \bar{\phi}_b u(\phi\bar{\phi}) e^{-\sum_i h_i |\phi_i|^2}]_\mu = \left[\left(\int_0^\infty \mathbb{E}_b u(\mathcal{L}_t + \phi\bar{\phi})\mathbf{1}_{X_t=a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt \right) e^{-\sum_i h_i |\phi_i|^2} \right]_\mu.$$

Equivalently, after dividing both sides by $|G_{\boldsymbol{h}}|$, we have

$$[\phi_a \bar{\phi}_b u(\phi\bar{\phi})]_{\mu_{\boldsymbol{h}}} = \left[\int_0^\infty \mathbb{E}_b u(\mathcal{L}_t + \phi\bar{\phi})\mathbf{1}_{X_t=a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt \right]_{\mu_{\boldsymbol{h}}}.$$

We remark that the measure $\mu_{\boldsymbol{h}}$ is normalizable in this case since $Q_{\boldsymbol{h}}$ has a positive definite symmetric part. When $a = b$, we can use Theorem 2.3.1 on both sides of the integration, and obtain the first equality in the theorem due to the following equations.

$$[\phi_a \bar{\phi}_a u(\phi\bar{\phi})]_{\mu_{\boldsymbol{h}}} = \mathbb{E}\left[\ell_a u(\ell)\right],$$
$$\left[\int_0^\infty \mathbb{E}_b u(\mathcal{L}_t + \phi\bar{\phi})\mathbf{1}_{X_t=a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt \right]_{\mu_{\boldsymbol{h}}} = \int_0^\infty \mathbb{E}_a u(\mathcal{L}_t + \ell)\mathbf{1}_{X_t=a} e^{-\sum_i h_i \mathcal{L}_t(i)} dt.$$

For the second equality, we use the second equation in Lemma 2.3.2. The rest of the computation is essentially the same as above. □

The general case, when $a, b \in [n]$ are not necessarily equal, corresponds to the isomorphism theorem as stated in Proposition 1 in [64]. However, when $a \neq b$, and Q is non-reversible, the l.h.s does not have a direct probabilistic meaning.

non-reversible Ray-Knight's isomorphism

The following theorem, first proved by Eisenbaum-Kaspi[46], extends the generalized second Ray-Knight's isomorphism theorem to the non-reversible case (the reversible version of generalized Ray-Knight isomorphism theorem can be found in, e.g., [97]).

Theorem 2.3.4 (non-reversible generalized second Ray-Knight's isomorphism theorem). *Given a Laplacian Q for an irreducible finite Markov chain, let ℓ be a 1-permanental process with kernel $G^h = (he_a e_a^T - Q)^{-1}$ for some $h > 0$ and some $a \in [n]$. For $u : \mathbb{R}_+^n \to \mathbb{C}$ a smooth compactly supported function and any $r \geq 0$,*

$$\mathbb{E}_a\big[u(\ell + \mathcal{L}_{\tau_{\text{inv}}(r)}) \,|\, \ell_a = 0\big] = \mathbb{E}\big[u(\ell) \,|\, \ell_a = r\big].$$

Here $\tau_{\text{inv}}(r) := \inf\{s \geq 0 : \mathcal{L}_s(a) \geq r\}$.

To see that the symmetric part of G^h is positive definite, refer to the discussion in Remark 2.2.1.4.

Proof. Directly using the choice of g similar to that in Bauerschmidt's proof[12] in the Ward identities will incur some problems, because we are dealing with complex numbers. For any $\epsilon > 0$, let $\delta_{0,\epsilon}(u)$ be a smooth approximation of the delta function with support on $\{u \in \mathbb{R} : |u| < \epsilon\}$.

The case $r = 0$ is trivial since under $\mathbb{P}_a$, we have $\mathcal{L}_{\tau_{\text{inv}}(0)} = 0$ and the theorem holds. Hence, we consider from now on $r > 0$, and any $\epsilon < \frac{r}{4}$. In the first equation of Lemma 2.3.2, we take $g_\epsilon(\phi) = \int_{2\epsilon}^{r-2\epsilon} \frac{1}{\phi_a}\delta_{0,\epsilon}(\phi_a\bar{\phi}_a - s)ds$. Note this equation is meaningless when $\phi_a = 0$, so we simply define $g_\epsilon(\phi) = 0$ when $\phi_a = 0$. We also take $f(i,l) = u(l)\delta_{0,\epsilon}(l_a - r)\mathbf{1}_{i=a}$. In this case, we have the integration formula[6]

$$[\phi_a g_\epsilon(\phi)u(\phi\bar{\phi})\delta_{0,\epsilon}(\phi_a\bar{\phi}_a - r)]_\mu = \left[(T_a g_\epsilon)(\phi)\int_0^\infty \mathbb{E}_a u(\phi\bar{\phi} + \mathcal{L}_t)\delta_{0,\epsilon}(\phi_a\bar{\phi}_a + \mathcal{L}_t(a) - r)\mathbf{1}_{X_t = a}dt\right]_\mu.$$

Note that $g_\epsilon(\phi)$ is nonzero only if $\phi_a\bar{\phi}_a < r - \epsilon$, which is disjoint from the support of $\delta_{0,\epsilon}(\phi_a\bar{\phi}_a - r)$ since it's non-zero only if $\phi_a\bar{\phi}_a > r - \epsilon$. Hence, the left-hand side is always zero.

Now consider the right-hand side. We have

$$T_a g_\epsilon(\phi) = \int_{2\epsilon}^{r-2\epsilon} \frac{1}{\phi_a} T_a \delta_{0,\epsilon}(\phi_a\bar{\phi}_a - s)ds = \int_{2\epsilon}^{r-2\epsilon} \delta'_{0,\epsilon}(\phi_a\bar{\phi}_a - s)ds.$$

Here the validity of the interchange of integration and differentiation is due to the continuity of $g_\epsilon(\phi)$ and $T_a g_\epsilon(\phi)$, which satisfies the Leibniz rule. So in (l,θ)-coordinates, we have

$$T_a g_\epsilon(\phi) = \int_{2\epsilon}^{r-2\epsilon} \delta'_{0,\epsilon}(l_a - s)ds = \delta_{0,\epsilon}(l_a - 2\epsilon) - \delta_{0,\epsilon}(l_a - r + 2\epsilon).$$

[6] We remark that integral for both sides exists since the integrand inside each bracket is continuous with bounded support.

Therefore, we conclude for any sufficiently small $\epsilon > 0$,

$$\left[\delta_{0,\epsilon}(l_a - 2\epsilon)\int_0^\infty \mathbb{E}_a u(l+\mathcal{L}_t)\delta_{0,\epsilon}(l_a+\mathcal{L}_t(a)-r)\mathbf{1}_{X_t=a}dt\right]_\mu$$
$$= \left[\delta_{0,\epsilon}(l_a - r + 2\epsilon)\int_0^\infty \mathbb{E}_a u(l+\mathcal{L}_t)\delta_{0,\epsilon}(l_a+\mathcal{L}_t(a)-r)\mathbf{1}_{X_t=a}dt\right]_\mu.$$

Fix any $l \in \mathbb{R}_+^n$ s.t. $l_a \leq r$. We will perform time change using $dL^a = \mathbf{1}_{X_t=a}dt$, we have

$$\begin{aligned}
&\int_0^\infty \mathbb{E}_a u(l+\mathcal{L}_t)\delta_{0,\epsilon}(l_a+L^a-r)\mathbf{1}_{X_t=a}dt \\
&\qquad = \mathbb{E}_a\int_0^\infty u(l+\mathcal{L}_t)\delta_{0,\epsilon}(l_a+L^a-r)\mathbf{1}_{X_t=a}dt \\
&\qquad = \mathbb{E}_a\int_0^\infty u(l+\mathcal{L}_{\tau_{\text{inv}}(L^a)})\delta_{0,\epsilon}(l_a+L^a-r)dL^a \\
&\qquad = \int_0^\infty \mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(\gamma)})\delta_{0,\epsilon}(l_a+\gamma-r)d\gamma
\end{aligned}$$

Now taking the limit $\epsilon \downarrow 0$, and using the continuity of $\gamma \mapsto \mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(\gamma)})$[7], we get

$$\lim_{\epsilon\downarrow 0}\int_0^\infty \mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(\gamma)})\delta_{0,\epsilon}(l_a+\gamma-r)d\gamma = \mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(r-l_a)}).$$

Therefore, taking the limit $\epsilon \downarrow 0$ in the previous equation, we have

$$\left[\delta_0(l_a)\mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(r-l_a)})\right]_\mu = \left[\delta_0(l_a-r)\mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(r-l_a)})\right]_\mu.$$

Since $\tau_{\text{inv}}(r-l_a) = \tau_{\text{inv}}(0) = 0$ when $l_a = r$, and $\tau_{\text{inv}}(r-l_a) = \tau_{\text{inv}}(r)$ when $l_a = 0$, we conclude

$$\left[\delta_0(l_a)\mathbb{E}_a u(l+\mathcal{L}_{\tau_{\text{inv}}(r)})\right]_\mu = \left[\delta_r(l_a)u(l)\right]_\mu.$$

Now the left-hand side corresponds to expectation with respect to a conditioned 1-permanental vector ℓ with kernel $G^h = (-Q^h)^{-1}$, where $Q^h := Q - he_ae_a^T$, conditioned on $\ell_a = 0$. The right-hand side, however, corresponds to expectation with respect to ℓ conditioned on $\ell_a = r$. To see this, let us decompose the Laplacian Q as stated in Eq. (2.4)(here

[7]The continuity follows from the same analysis as [12, footnote 1, pg. 9], which we include for completeness. By assumption, u is compactly supported, so it suffices to show that for a sufficiently large T, $\mathbb{E}_{a,l}u(\mathcal{L}_{\tau_{\text{inv}}(\gamma)\wedge T})$ is continuous. Since u is Lipschitz, it suffices to show that $\mathbb{E}_{a,l}\|\mathcal{L}_{\tau_{\text{inv}}(\gamma-\delta)\wedge T} - \mathcal{L}_{\tau_{\text{inv}}(\gamma+\delta)\wedge T}\|_1 \to 0$ as $\delta \downarrow 0$, with $\|\cdot\|_1$ being the 1-norm. Let J_δ be the event that a jump occurs during the time interval $[\gamma-\delta, \gamma+\delta]$ at a. Then, using the total expectation rule with respect to the event of J_δ and J_δ^c, we have

$$\mathbb{E}_{a,l}\|\mathcal{L}_{\tau_{\text{inv}}(\gamma-\delta)\wedge T} - \mathcal{L}_{\tau_{\text{inv}}(\gamma+\delta)\wedge T}\|_1 \leq \delta\mathbb{P}_{a,l}(J_\delta^c) + T\mathbb{P}_{a,l}(J_\delta) \leq \delta + T\cdot O(\delta) = O(\delta).$$

we identify node 1 in the decomposition of Eq. (2.4) with node a mentioned above without loss of generality). Recall that we have Lemma 2.2.2, with $l_* := (l_i : i \neq a)$ here,

$$\rho(l_*|l_a = r) =: \rho_r(l_*) = \int_{(0,2\pi)^{n-1}} \frac{|-Q_{**}|}{(2\pi)^{n-1}} \exp\left(\langle(\phi_* + \sqrt{r}Q_{**}^{-T}Q_{a*}^h), Q_{**}(\bar{\phi}_* + \sqrt{r}Q_{**}^{-1}Q_{*a}^h)\rangle\right) d\theta_*.$$

Here we used $Q_{**}^h = Q_{**}$. To apply this formula, note that

$$\begin{aligned}\left[\delta_0(l_a)\mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)})\right]_\mu &= \frac{1}{(2\pi)^n} \int_{(0,2\pi)^n} \int_{\mathbb{R}_+^n} \delta_0(l_a)\mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) e^{\langle\phi, Q\bar{\phi}\rangle} dl d\theta \\ &= \frac{1}{(2\pi)^n} \int_{(0,2\pi)^{n-1}} \int_{\mathbb{R}_+^{n-1}} \left(\int_{(0,2\pi)} \int_{\mathbb{R}_+} \delta_0(l_a)\mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) e^{\langle\phi, Q\bar{\phi}\rangle} dl_a d\theta_a \right) dl_* d\theta_* \\ &= \frac{1}{(2\pi)^{n-1}} \int_{(0,2\pi)^{n-1}} \int_{\mathbb{R}_+^{n-1}} \mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) e^{\langle\phi_*, Q_{**}\bar{\phi}_*\rangle} dl_* d\theta_*.\end{aligned}$$

Here we integrated out l_a and θ_a using the fact that

$$\int_{\mathbb{R}_+} \delta_0(l_a)\mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) e^{\langle\phi, Q\bar{\phi}\rangle} dl_a = \mathbb{E}_a u(\tilde{l} + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) e^{\langle\phi_*, Q_{**}\bar{\phi}_*\rangle}.$$

Here $\tilde{l}$ is l under the restriction $l_a = 0$. We can use Lemma 2.2.2 with $r = 0$ as

$$\begin{aligned}\left[\delta_0(l_a)\mathbb{E}_a u(l + \mathcal{L}_{\tau_{\mathrm{inv}}(r)})\right]_\mu &= \frac{1}{|-Q_{**}|} \int_{\mathbb{R}_+^{n-1}} \mathbb{E}_a u(\tilde{l} + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) \rho_0(l_*) dl_* \\ &= \frac{1}{|-Q_{**}|} \mathbb{E}_a\left[u(\ell + \mathcal{L}_{\tau_{\mathrm{inv}}(r)}) \,|\, \ell_a = 0\right].\end{aligned}$$

Similarly, we know

$$\begin{aligned}\left[\delta_r(l_a)u(l)\right]_\mu &= \frac{1}{(2\pi)^n} \int_{(0,2\pi)^n} \int_{\mathbb{R}_+^n} \delta_r(l_a)u(l) e^{\langle\phi, Q\bar{\phi}\rangle} dl d\theta \\ &= \frac{1}{(2\pi)^n} \int_{(0,2\pi)^{n-1}} \int_{\mathbb{R}_+^{n-1}} \left(\int_{(0,2\pi)} \int_{\mathbb{R}_+} \delta_r(l_a)u(l) e^{\langle\phi, Q\bar{\phi}\rangle} dl_a d\theta_a \right) dl_* d\theta_* \\ &= \frac{1}{(2\pi)^n} \int_{(0,2\pi)^{n-1}} \int_{\mathbb{R}_+^{n-1}} \left(\int_{(0,2\pi)} u(l') e^{\langle\phi', Q\bar{\phi}'\rangle} d\theta_a \right) dl_* d\theta_*.\end{aligned}$$

Here l' is l restricted to $l_a = r$. Similarly for ϕ' and $\bar{\phi}'$. We now continue to compute that

$$\begin{aligned}
&\frac{1}{(2\pi)^n}\int_{(0,2\pi)^n} u(l')e^{\langle\phi',Q\bar{\phi}'\rangle}d\theta_a d\theta_* \\
&= \frac{u(l')e^{Q_{aa}r}}{(2\pi)^n}\int_{(0,2\pi)^n}\exp\Big(\sum_{i,j\neq a}Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)}+\sum_{i\neq a}\big(Q_{ia}\sqrt{l_ir}e^{\mathrm{i}(\theta_i-\theta_a)}+Q_{ai}\sqrt{l_ir}e^{\mathrm{i}(\theta_a-\theta_i)}\big)\Big) \\
&\qquad d\theta_a d\theta_* \\
&= \frac{u(l')e^{Q_{aa}r}}{(2\pi)^n}\int_{(0,2\pi)^n}\exp\Big(\sum_{i,j\neq a}Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta'_i-\theta'_j)}+\sum_{i\neq a}\big(Q_{ia}\sqrt{l_ir}e^{\mathrm{i}\theta'_i}+Q_{ai}\sqrt{l_ir}e^{-\mathrm{i}\theta'_i}\big)\Big)d\theta'_a d\theta'_* \\
&= \frac{u(l')e^{Q_{aa}r}}{(2\pi)^{n-1}}\int_{(0,2\pi)^{n-1}}\exp\Big(\sum_{i,j\neq a}Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta'_i-\theta'_j)}+\sum_{i\neq a}\big(Q_{ia}\sqrt{l_ir}e^{\mathrm{i}\theta'_i}+Q_{ai}\sqrt{l_ir}e^{-\mathrm{i}\theta'_i}\big)\Big)d\theta'_* \\
&= \frac{u(l')e^{Q_{aa}r}}{|-Q_{**}|e^{(Q^h_{aa}+(G^h_{aa})^{-1})r}}\rho_r(l_*).
\end{aligned}$$

In the second step, we used the rotation trick to construct θ' as in the proof of Lemma 2.2.2. In the third step, we used $Q^h_{ij} = Q_{ij}$ unless $i = j = a$. We also used $G^h_{aa} = |G^h||-Q^h_{**}| = |G^h||-Q_{**}|$ and the following intermediate step in Lemma 2.2.2:

$$\begin{aligned}
\rho_r(l_*) = \int_{(0,2\pi)^{n-1}}\frac{G^h_{aa}e^{(Q^h_{aa}+(G^h_{aa})^{-1})r}}{(2\pi)^{n-1}|G^h|}\exp\Big(&\sum_{i,j\neq a}Q_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta'_i-\theta'_j)}+\sum_{i\neq a}\big(Q_{ia}\sqrt{l_ir}e^{\mathrm{i}\theta'_i} \\
&+Q_{ai}\sqrt{l_ir}e^{-\mathrm{i}\theta'_i}\big)\Big)d\theta'_*.
\end{aligned}$$

Note that we have $G^h_{aa} = \frac{1}{h}$ since we kill the chain at rate h at state a. Therefore, $Q_{aa} - Q^h_{aa} - (G^h_{aa})^{-1} = h - (G^h_{aa})^{-1} = 0$. This allows us to conclude that

$$\big[\delta_r(l_a)u(l)\big]_\mu = \frac{1}{|-Q_{**}|}\int_{\mathbb{R}^n_+}u(l')\rho_r(l_*)dl_* = \frac{1}{|-Q_{**}|}\mathbb{E}\big[u(\ell)\,|\,\ell_a = r\big]$$

This proves the theorem. □

non-reversible version of Eisenbaum's isomorphism theorem

To the best of our knowledge, the non-reversible version of Eisenbaum's isomorphism theorem hasn't been discovered previously, probably because one cannot interpret one side of the equation probabilistically in this case.

Theorem 2.3.5 (non-reversible Eisenbaum's isomorphism)**.** *Consider an irreducible continuous-time Markov chain X_t on discrete space $[n]$ with Laplacian Q. Let ζ be an independent exponential random variable with mean h^{-1} for some $h > 0$. For any $u : \mathbb{R}^n_+ \to \mathbb{C}$*

with property $\mathfrak{F}$ and any $a \in [n], r > 0$, we have

$$\left[\left(\frac{\phi_a + \sqrt{r}}{\sqrt{r}}\right) u\big(|\phi + \sqrt{r}\mathbf{1}|^2\big)\right]_{\mu_{h\mathbf{1}}} = \left[\mathbb{E}_a^*\left[u\big(|\phi + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_\zeta\big)\right]\right]_{\mu_{h\mathbf{1}}}.$$

Proof. In Lemma 2.3.2, we take $g(\phi) = e^{h\sum_i |\phi_i|^2 - h\sum_i |\phi_i - \sqrt{r}|^2} = e^{\sqrt{r}h\sum_i(\phi_i + \bar{\phi}_i) - nrh}$, and $f(i,l) = u(l)e^{-h\sum_j l_j}\mathbf{1}_{i=a}$.[8] This yields

$$\left[\phi_a u(\phi\bar{\phi})e^{-h\sum_i |\phi_i - \sqrt{r}|^2}\right]_\mu = \sqrt{r}h\sum_i \left[g(\phi)\int_0^\infty \mathbb{E}_{i,\phi\bar{\phi}}\left[u(\mathcal{L}_t)e^{-ht}\mathbf{1}_{X_t=a}\right]dt\right]_\mu.$$

Now note that $g(\phi)\mathbb{E}_{i,\phi\bar{\phi}}[u(\mathcal{L}_t)e^{-ht}\mathbf{1}_{X_t=a}] = e^{-h\sum_i |\phi_i - \sqrt{r}|^2}\mathbb{E}_i[u(\phi\bar{\phi} + \mathcal{L}_t)e^{-ht}\mathbf{1}_{X_t=a}]$. Hence,

$$\left[\phi_a u(\phi\bar{\phi})e^{-h\sum_i |\phi_i - \sqrt{r}|^2}\right]_\mu = \sqrt{r}h\sum_i \left[e^{-h\sum_i |\phi_i - \sqrt{r}|^2}\int_0^\infty \mathbb{E}_i\left[u(\phi\bar{\phi} + \mathcal{L}_t)e^{-ht}\mathbf{1}_{X_t=a}\right]dt\right]_\mu.$$

Make the change of variable $\phi' = \phi - \sqrt{r}\mathbf{1}$, and notice that $\langle \phi', Q\bar{\phi}'\rangle = \langle \phi, Q\bar{\phi}\rangle$. Dividing both sides by $|-Q_{h\mathbf{1}}| = |hI - Q|$ gives

$$\left[\left(\frac{\phi'_a + \sqrt{r}}{\sqrt{r}}\right) u\big(|\phi' + \sqrt{r}\mathbf{1}|^2\big)\right]_{\mu_{h\mathbf{1}}} = h\sum_i \left[\int_0^\infty \mathbb{E}_i\left[u\big(|\phi' + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_t\big)e^{-ht}\mathbf{1}_{X_t=a}\right]dt\right]_{\mu_{h\mathbf{1}}}.$$

Reverse the chain on the right-hand side[9], and get

$$\begin{aligned}
\left[\left(\frac{\phi'_a + \sqrt{r}}{\sqrt{r}}\right) u\big(|\phi' + \sqrt{r}\mathbf{1}|^2\big)\right]_{\mu_{h\mathbf{1}}} &= h\sum_i \left[\int_0^\infty \mathbb{E}_a^*\left[u\big(|\phi' + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_t\big)e^{-ht}\mathbf{1}_{X_t=i}\right]dt\right]_{\mu_{h\mathbf{1}}} \\
&= h\left[\int_0^\infty \mathbb{E}_a^*\left[u\big(|\phi' + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_t\big)e^{-ht}\right]dt\right]_{\mu_{h\mathbf{1}}}.
\end{aligned}$$

The change of summation and integral can be justified by the bounded convergence theorem, due to the sub-Gaussianity of the density of $[\cdot]_{\mu_{h\mathbf{1}}}$ and property $\mathfrak{F}$ of the function u.

Let ζ be an exponential r.v. with mean h^{-1}, independent of the recurrent random walk. Since we have

$$\mathbb{E}_a^*\left[u\big(|\phi' + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_\zeta\big)\right] = h\int_0^\infty \mathbb{E}_a^*\left[u\big(|\phi' + \sqrt{r}\mathbf{1}|^2 + \mathcal{L}_t\big)e^{-ht}\right]dt,$$

[8]The function g does not decay exponentially fast as required in Lemma 2.3.2, but one can check that the sub-Gaussianity of f will kill any exponential growth in g.

[9]Recall that the generator for the reverse process is Q^T, and the transition kernel is $P_t^*(i,j) = (e^{tQ^T})_{i,j}$. Since the original Markov process has transition kernel $P_t(j,i) = (e^{tL})_{j,i}$, we know that $P_t^*(i,j) = P_t(j,i)$ and they should yield exactly the same bridge measure on paths of length t under path reversal.

we conclude that

$$\left[\left(\frac{\phi'_a+\sqrt{r}}{\sqrt{r}}\right)u\big(|\phi'+\sqrt{r}\mathbf{1}|^2\big)\right]_{\mu_{h\mathbf{1}}} = \big[\mathbb{E}^*_a\big[u\big(|\phi'+\sqrt{r}\mathbf{1}|^2+\mathcal{L}_\zeta\big)\big]\big]_{\mu_{h\mathbf{1}}}.$$

$\square$

Remark 2.3.5.1. *There is a generalized version of Theorem 2.3.5 above. Consider any $h_1,\cdots,h_n\geq 0$ such that $\sum_{i=1}^n h_i>0$. Take*

$$g(\phi)=e^{\sum_i h_i|\phi_i|^2-\sum_i h_i|\phi_i-\sqrt{r}|^2}=e^{\sqrt{r}\sum_i h_i(\phi_i+\bar{\phi}_i)-r\sum_i h_i},$$

and $f(i,l)=u(l)e^{-\sum_j h_j l_j}\mathbf{1}_{i=a}$ in Lemma 2.3.2. This yields

$$\big[\phi_a u(\phi\bar{\phi})e^{-\sum_i h_i|\phi_i-\sqrt{r}|^2}\big]_\mu=\sqrt{r}\sum_i h_i\left[g(\phi)\int_0^\infty \mathbb{E}_{i,\phi\bar{\phi}}\big[u(\mathcal{L}_t)e^{-\sum_j h_j\mathcal{L}_t(j)}\mathbf{1}_{X_t=a}\big]dt\right]_\mu.$$

Note that $g(\phi)\mathbb{E}_{i,\phi\bar{\phi}}[u(\mathcal{L}_t)e^{-\sum_j h_j\mathcal{L}_t(j)}\mathbf{1}_{X_t=a}]=e^{-\sum_i h_i|\phi_i-\sqrt{r}|^2}\mathbb{E}_i[u(\phi\bar{\phi}+\mathcal{L}_t)e^{-\sum_j h_j\mathcal{L}_t(j)}\mathbf{1}_{X_t=a}]$. Hence,

$$\begin{aligned}&\big[\phi_a u(\phi\bar{\phi})e^{-\sum_i h_i|\phi_i-\sqrt{r}|^2}\big]_\mu\\ &\qquad=\sqrt{r}\sum_i h_i\left[e^{-\sum_i h_i|\phi_i-\sqrt{r}|^2}\int_0^\infty \mathbb{E}_i\big[u(\phi\bar{\phi}+\mathcal{L}_t)e^{-\sum_j h_j\mathcal{L}_t(j)}\mathbf{1}_{X_t=a}\big]dt\right]_\mu.\end{aligned}$$

Making the change of variable $\phi'=\phi-\sqrt{r}\mathbf{1}$, and dividing both sides by $|-Q_{\boldsymbol{h}}|$ gives

$$\left[\left(\frac{\phi'_a+\sqrt{r}}{\sqrt{r}}\right)u\big(|\phi'+\sqrt{r}\mathbf{1}|^2\big)\right]_{\mu_{\boldsymbol{h}}}=\sum_i h_i\left[\int_0^\infty \mathbb{E}_i\big[u\big(|\phi'+\sqrt{r}\mathbf{1}|^2+\mathcal{L}_t\big)e^{-\sum_j h_j\mathcal{L}_t(j)}\mathbf{1}_{X_t=a}\big]dt\right]_{\mu_{\boldsymbol{h}}}.$$

2.4 Comparison inequalities for permanental processes

Kahane-type inequalities are inequalities between expectations of functions of centered multivariate Gaussian random variables when one assumes that the covariance matrices of the Gaussians can be compared in some sense, see e.g. [15, Lecture 5]. Slepian's lemma [15, 95] is a famous special case. In [44], a Kahane-type inequality and a version of Slepian's lemma for 4-permanental processes has been given. However, 4-permanental processes are not directly related to the isomorphism theorems, which limits the applications of these results. We use the density-formula approach to derive analogous results for 1-permanental processes. We remark that Eisenbaum's method used a lemma of [94], which is not satisfied by 1-permanental vectors.

For any function $f : \mathbb{R}_+^n \to \mathbb{R}$, we say it has subexponential growth for $k \leq 2$ orders of partial derivatives if (i.)f has subexponential growth; (ii.)both $\{\partial_i f := \frac{\partial}{\partial l_i} f, i \in [n]\}$ and $\{\partial_{ij} f := \frac{\partial^2}{\partial l_i \partial l_j} f, i, j \in [n]\}$ exist and have subexponential growth. Recall that any infinitely divisible kernel G of a 1-permanental random vector has an equivalent kernel $\tilde{G}$ that is an inverse M-matrix with a positive definite symmetric part according to Remark 2.2.1.3. We have the following comparison inequality.

Lemma 2.4.1 (Kahane-type inequality for 1-permanental vectors)**.** *Consider two kernels for 1-permanental vectors $\{G_i, i = 0, 1\}$ and an interpolation $\{G_\alpha, \alpha \in [0, 1]\}$ between them such that G_α is an inverse M-matrix with a positive definite symmetric part for any $\alpha \in [0, 1]$, and $\alpha \mapsto G_{\alpha,ij}$ is of class $C^1([0, 1])$, with $\frac{d}{d\alpha} G_{\alpha,ij}$ having a constant sign for any $i, j \in [n]$. Let $\{\ell_\alpha, \alpha \in [0, 1]\}$ be the corresponding 1-permanental vectors. Let $f : \mathbb{R}_+^n \to \mathbb{R}$ be any function with subexponential growth for $k \leq 2$ orders of partial derivatives, such that for any $i \in [n]$, either of the following is true:*

1. $\frac{d}{d\alpha} G_{\alpha,ii} \geq 0$, $\forall \alpha \in [0, 1]$, *and* $\partial_i f(l) + l_i \, \partial_{ii} f(l) \geq 0$, $\forall l \in \mathbb{R}_+^n$;
2. $\frac{d}{d\alpha} G_{\alpha,ii} \leq 0$, $\forall \alpha \in [0, 1]$, *and* $\partial_i f(l) + l_i \, \partial_{ii} f(l) \leq 0$, $\forall l \in \mathbb{R}_+^n$;

and for any $i \neq j$, either of the following is true:

1. $\frac{d}{d\alpha} G_{\alpha,ij} \geq 0$, $\forall \alpha \in [0, 1]$, *and* $\partial_{ij} f(l) \geq 0$, $\forall l \in \mathbb{R}_+^n$;
2. $\frac{d}{d\alpha} G_{\alpha,ij} \leq 0$, $\forall \alpha \in [0, 1]$, *and* $\partial_{ij} f(l) \leq 0$, $\forall l \in \mathbb{R}_+^n$.

Then we have $\mathbb{E}[f(\ell_1)] \geq \mathbb{E}[f(\ell_0)]$.

Proof. Let

$$Q_\alpha := -(G_\alpha)^{-1}. \tag{2.8}$$

By Theorem 2.3.1, we know that

$$\mathbb{E}[f(\ell_\alpha)] = \frac{1}{(2\pi \mathrm{i})^n |G_\alpha|} \int_{\mathbb{C}^n} f(\phi\bar{\phi}) e^{\langle \phi, Q_\alpha \bar{\phi} \rangle} d\bar{\phi} d\phi.$$

Therefore, we have

$$\mathbb{E}[f(\ell_1)] - \mathbb{E}[f(\ell_0)] = \int_0^1 \Big(\frac{\partial}{\partial \alpha} \mathbb{E}[f(\ell_\alpha)] \Big) d\alpha = \int_0^1 \frac{\partial}{\partial \alpha} \left(\int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^n |G_\alpha|} f(\phi\bar{\phi}) e^{\langle \phi, Q_\alpha \bar{\phi} \rangle} d\bar{\phi} d\phi \right) d\alpha.$$

Computing the derivative as[10]

$$\frac{\partial}{\partial \alpha} \left(\frac{1}{|G_\alpha|} e^{\langle \phi, Q_\alpha \bar{\phi} \rangle} \right) = \mathrm{tr} \left(\frac{d}{dG_\alpha} \left(\frac{1}{|G_\alpha|} e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} \right) \frac{dG_\alpha}{d\alpha} \right),$$

[10]Recall that for a

we know that

$$\mathbb{E}[f(\ell_1)] - \mathbb{E}[f(\ell_0)] = \frac{1}{(2\pi\mathrm{i})^n} \int_0^1 \int_{\mathbb{C}^n} f(\phi\bar{\phi}) \operatorname{tr}\left(\frac{d}{dG_\alpha} \left(\frac{1}{|G_\alpha|} e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} \right) \frac{dG_\alpha}{d\alpha} \right) d\bar{\phi} d\phi d\alpha.$$

We remark that interchanging the derivative and the integral can be justified by the absolute integrability of the partial derivative. The detailed justification of this fact can be found in the appendix (Section A.2).

The Jacobi formula (see, e.g., Section 0.8.2 of [59]) states that for invertible matrix A, $\frac{\partial}{\partial A_{ij}}|A| = (adj(A))_{ji}$, where the adjugate is $adj(A) := |A|A^{-1}$. It is then easy to verify that

$$\frac{d}{dG_\alpha}\left(\frac{1}{|G_\alpha|}\right) = -\frac{1}{|G_\alpha|^2} \frac{d|G_\alpha|}{dG_\alpha} = -\frac{1}{|G_\alpha|^2} |G_\alpha| (G_\alpha^{-1}) = \frac{1}{|G_\alpha|} Q_\alpha.$$

Using the fact that $\frac{d}{dA}\operatorname{tr}(A^{-1}B) = -A^{-1}BA^{-1}$, we know that

$$\begin{aligned} \frac{d}{dG_\alpha} e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} &= -e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} \frac{d}{dG_\alpha} \operatorname{tr}\big(G_\alpha^{-1} \bar{\phi} \phi^T\big) \\ &= e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} G_\alpha^{-1} \bar{\phi} \phi^T G_\alpha^{-1} = e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} Q_\alpha \bar{\phi} \phi^T Q_\alpha. \end{aligned}$$

Let

$$d\mu_\alpha(\phi) := (2\pi\mathrm{i})^{-n} |G_\alpha|^{-1} e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} d\bar{\phi} d\phi \tag{2.9}$$

denote the normalized twisted Gaussian density. We know that

$$\begin{aligned} \mathbb{E}[f(\ell_1)] - \mathbb{E}[f(\ell_0)] &= \int_0^1 \int_{\mathbb{C}^n} f(\phi\bar{\phi}) \operatorname{tr}\big((Q_\alpha + Q_\alpha \bar{\phi} \phi^T Q_\alpha) \frac{dG_\alpha}{d\alpha}\big) d\mu_\alpha(\phi) d\alpha. \\ &= \int_0^1 \int_{\mathbb{C}^n} f(\phi\bar{\phi}) \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha} \left(Q_{\alpha,ji} + \sum_{i',j'} Q_{\alpha,jj'} Q_{\alpha,i'i} \phi_{i'} \bar{\phi}_{j'} \right) d\mu_\alpha(\phi) d\alpha \\ &= \int_0^1 \int_{\mathbb{C}^n} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha} \left(Q_{\alpha,ji} f(\phi\bar{\phi}) + \sum_{i'} Q_{\alpha,i'i} \Big(-\delta_{i'j} f(\phi\bar{\phi}) - \phi_{i'} \bar{\phi}_j \partial_j f(\phi\bar{\phi}) \Big) \right) d\mu_\alpha(\phi) d\alpha. \end{aligned}$$

In the third step, we used the following integration by parts for any $i, j \in [n]$:

$$\int_{\mathbb{C}^n} \frac{\partial}{\partial \phi_j} \big(\phi_{i'} f(\phi\bar{\phi}) e^{-\langle \phi, G_\alpha^{-1} \bar{\phi} \rangle} \big) d\bar{\phi} d\phi = 0 \quad \Rightarrow$$

$$0 = \int_{\mathbb{C}^n} \delta_{i'j} f(\phi\bar{\phi}) d\mu_\alpha(\phi) + \int_{\mathbb{C}^n} \phi_{i'} \bar{\phi}_j \partial_j f(\phi\bar{\phi}) d\mu_\alpha(\phi) + \int_{\mathbb{C}^n} \phi_{i'} f(\phi\bar{\phi}) \left(\sum_{j'} Q_{\alpha,jj'} \bar{\phi}_{j'} \right) d\mu_\alpha(\phi).$$

We can simplify by canceling terms and doing another integration-by-parts, and get

$$\begin{aligned} \mathbb{E}[f(\ell_1)] - \mathbb{E}[f(\ell_0)] &= \int_0^1 \int_{\mathbb{C}^n} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha} \left(-\sum_{i'} Q_{\alpha,i'i} \phi_{i'} \bar{\phi}_j \partial_j f(\phi\bar{\phi}) \right) d\mu_\alpha(\phi) d\alpha \\ &= \int_0^1 \int_{\mathbb{C}^n} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha} \left(\delta_{ij} \partial_j f(\phi\bar{\phi}) + \phi_i \bar{\phi}_j \partial_{ij} f(\phi\bar{\phi}) \right) d\mu_\alpha(\phi) d\alpha. \end{aligned}$$

The integration-by-parts formula used here is that for any $i, j \in [n]$,

$$\int_{\mathbb{C}^n} \frac{\partial}{\partial \bar{\phi}_i}\left(\bar{\phi}_j \partial_j f(\phi\bar{\phi}) e^{-\langle \phi, G_\alpha^{-1}\bar{\phi}\rangle}\right) d\bar{\phi} d\phi = 0 \quad \Rightarrow$$
$$0 = \int_{\mathbb{C}^n} \delta_{ij}\partial_j f(\phi\bar{\phi}) d\mu_\alpha(\phi) + \int_{\mathbb{C}^n} \phi_i\bar{\phi}_j \partial_{ij} f(\phi\bar{\phi}) d\mu_\alpha(\phi) + \int_{\mathbb{C}^n} \bar{\phi}_j \partial_j f(\phi\bar{\phi}) \left(\sum_{i'} Q_{\alpha,i'i}\phi_{i'}\right) d\mu_\alpha(\phi).$$

We remark that the validity of these integration-by-parts procedures is guaranteed by the subexponential growth conditions. Given that G_α is an infinitely divisible kernel of a 1-permanental vector with a positive definite symmetric part, and $\partial_{ij} f(l)$ is subexponential for any i, j, using Lemma A.3.1, we know that for any $i \neq j$,

$$\int_{\mathbb{C}^n} \frac{dG_{\alpha,ij}}{d\alpha} \phi_i\bar{\phi}_j \partial_{ij} f(\phi\bar{\phi}) d\mu_\alpha(\phi) = \frac{dG_{\alpha,ij}}{d\alpha} \int_{\mathbb{C}^n} \phi_i\bar{\phi}_j \partial_{ij} f(\phi\bar{\phi}) d\mu_\alpha(\phi) \geq 0.$$

On the other hand, for $i \in [n]$, by Theorem 2.3.1, we have

$$\int_{\mathbb{C}^n} \frac{dG_{\alpha,ii}}{d\alpha}\Big(\partial_i f(\phi\bar{\phi}) + \phi_i\bar{\phi}_i \partial_{ii} f(\phi\bar{\phi})\Big) d\mu_\alpha(\phi) = \mathbb{E}\left[\frac{dG_{\alpha,ii}}{d\alpha}\big(\partial_i f(\ell_\alpha) + \ell_i\, \partial_{ii} f(\ell_\alpha)\big)\right] \geq 0.$$

It is then straightforward to check that the conditions on f stated in the lemma imply that $\mathbb{E}[f(\ell_\alpha)]$ is increasing with α. Therefore, $\mathbb{E}[f(\ell_1)] \geq \mathbb{E}[f(\ell_0)]$ and the claim is proved. □

Remark 2.4.1.1 (Sign patterns I)**.** *Given any sign pattern $\{s_{ij} \in \{-1, 1\}, i, j \in V\}$, we know there indeed exist nontrivial functions that satisfy the equations $s_{ii}(\partial_i f(l) + l_i\, \partial_{ii} f(l)) > 0$, and $s_{ij}\partial_{ij} f(l) > 0$, $\forall l \in \mathbb{R}^n_+$, for any $i \neq j \in V$. Consider the function*

$$f(l) = \sum_{(i,j): i\neq j} -s_{ij} \log(1 + l_i + l_j) + \sum_i s_{ii}(2ne^{l_i}).$$

Then we know that $\forall i \in V$,

$$\begin{aligned} s_{ii}(\partial_i f(l) + l_i\, \partial_{ii} f(l)) &= \sum_{j\neq i} -\frac{s_{ii}s_{ij}}{(1 + l_i + l_j)} + 2ne^{l_i} + \sum_{j\neq i} \frac{s_{ii}s_{ij}l_i}{(1 + l_i + l_j)^2} + 2nl_ie^{l_i} \\ &\geq -(n-1) + 2n - (n-1) > 0. \end{aligned}$$

In the first step, we used $s_{ii}^2 = 1$. In the second step, we used the fact that $\frac{1}{(1+l_i+l_j)} \in (0, 1]$, $\frac{l_i}{(1+l_i+l_j)^2} \in (0, 1]$, $e^{l_i} \geq 1$, and $l_ie^{l_i} \geq 0$. Also, note that $\forall i \neq j \in V$,

$$s_{ij}\partial_{ij} f(l) = \frac{1}{(1 + l_i + l_j)^2} > 0, \forall l \in \mathbb{R}^n_+.$$

Remark 2.4.1.2. *Let us consider the following instance for Lemma 2.4.1. Let $P_0, P_1 \in M_n(\mathbb{R}_+)$ be two sub-stochastic matrices(i.e., every row sum ≤ 1) such that $P_{0,ij} \leq P_{1,ij} \forall i,j \in [n]$. For any $s > 1$, consider $G_0 := (sI - P_0)^{-1}$ and $G_1 := (sI - P_1)^{-1}$. They are both inverse M-matrices according to the definition of an M-matrix [90], since $G_0^{-1} = sI - P_0$ where $\rho(P_0) \leq \|P_0\|_\infty := \max_i \sum_j |P_{0,ij}| \leq 1 < s$ and the same for G_1^{-1}. Moreover, we know that $\{P_\alpha := \alpha P_1 + (1-\alpha)P_0,\ \alpha \in [0,1]\}$ is a family of sub-stochastic matrices, and $\{G_\alpha := (sI - P_\alpha)^{-1},\ \alpha \in [0,1]\}$ is a family of inverse M-matrices. Using the Neumann series expansion, we have*

$$\frac{d}{d\alpha} G_\alpha = \frac{d}{d\alpha} \sum_{i=0}^{\infty} s^{-(i+1)} P_\alpha^i = \sum_{i=0}^{\infty} s^{-(i+1)} \frac{d}{d\alpha} P_\alpha^i = \sum_{i=1}^{\infty} s^{-(i+1)} \sum_{j=1}^{i-1} P_\alpha^j \Big(\frac{d}{d\alpha} P_\alpha\Big) P_\alpha^{i-1-j}.$$

Since $\frac{d}{d\alpha} P_\alpha = P_1 - P_0$ is entry-wise nonnegative, we know that $\frac{d}{d\alpha} G_{\alpha,ij} \geq 0, \forall i,j \in [n]$. One can then plug in any function of interest that satisfies for any $i \in [n]$, $\partial_i f(l) + l_i \partial_{ii} f(l) \geq 0, \forall l \in \mathbb{R}_+^n$ and for any $i \neq j \in [n]$, $\partial_{ij} f(l) \geq 0, \forall l \in \mathbb{R}_+^n$ to obtain non-trivial comparison inequalities.

Remark 2.4.1.3 (Sign patterns II)**.** *Consider the following family of inverse M-matrices in $M_n(\mathbb{R}_+)$, where $n \geq 3$. Given $0 < x < 1$, let $y := \sqrt{\frac{1}{n-2}\, x + \frac{(n-3)}{(n-2)}\, x^2}$. Consider the family of matrices $\mathcal{B} := \{B : B_{ii} = 1, \forall i \in [n]; 0 < x < B_{ij} < y < 1, \forall i \neq j\}$. This family of matrices is proved to be inverses of strictly diagonally dominant (both by rows and columns) M-matrices in [108, Theorem 2]. We claim that any M-matrix A that is strictly diagonally dominant both by rows and columns has a positive definite symmetric part, since for any $x \in \mathbb{R}^n$ such that $x \neq 0$,*

$$\begin{aligned}
\langle x, Ax \rangle &= \sum_i |A_{ii}| x_i^2 - \sum_{i,j:i<j} (|A_{ij}| + |A_{ji}|) x_i x_j \\
&> \sum_i \sum_{j \neq i} \frac{1}{2} (|A_{ij}| + |A_{ji}|) x_i^2 - \sum_{i,j:i<j} (|A_{ij}| + |A_{ji}|) x_i x_j \\
&= \sum_{i,j:i<j} \frac{1}{2} (|A_{ij}| + |A_{ji}|)(x_i^2 + x_j^2) - \sum_{i,j:i<j} (|A_{ij}| + |A_{ji}|) x_i x_j \\
&= \sum_{i,j:i<j} \frac{1}{2} (|A_{ij}| + |A_{ji}|)(x_i^2 - 2x_i x_j + x_j^2) \\
&= \sum_{i,j:i<j} \frac{1}{2} (|A_{ij}| + |A_{ji}|)(x_i - x_j)^2 \geq 0.
\end{aligned}$$

In the first step, we used the sign pattern of A as an M-matrix. In the second step, we used strict diagonal dominance in rows and columns, together with the fact that $x \neq 0$.

Therefore, we know that any matrix in the family $\mathcal{B}$, as an inverse of a strictly diagonally dominant (both by rows and columns) M-matrix, should have a positive definite symmetric part.

Moreover, for any two matrices $G_0, G_1 \in \mathcal{B}$, we can construct a linear path between them as $G_\alpha := \alpha G_1 + (1-\alpha)G_0, \forall \alpha \in (0,1)$. The whole path will be within family $\mathcal{B}$ since this family forms a convex set. Moreover, the sign of $\frac{d}{d\alpha}G_\alpha = G_1 - G_0$ will be constant entry-wise along the path. We can then vary G_0 and G_1 to obtain whatever arbitrary off-diagonal sign pattern we want.

Note that Eisenbaum derived the comparison lemma (Lemma 3.1, [44]) for 4-permanental vectors by directly working with the Fourier transform, which is very different from our approach. Also, their lemma requires the function f to be bounded. With the density formula, we substantially relax this condition to functions with subexponential growth here.

Convolutional tricks can be used to generalize the above lemma to k-permanental vectors for $k \in \mathbb{N}_+$ under the condition

$$G_{1,ii} > (<) G_{0,ii} \Rightarrow k\,\partial_i f + l_i\,\partial_{ii} f \geq (\leq) 0; \qquad G_{1,ij} > (<) G_{0,ij} \Rightarrow \partial_{ij} f \geq (\leq) 0.$$

To see this, let us consider the case $k = 2$. The general case follows the same idea. Let ℓ_i and ℓ_i' be independent 1-permanental vectors with kernel G_i, for $i = 0, 1$. Then $\ell_i + \ell_i'$ will be a 2-permanental vector with kernel G_i, which can be seen by calculating the moment generating function. Using the Theorem 2.3.1 for both ℓ_1, ℓ_1' and ℓ_0, ℓ_0', we have

$$\begin{aligned}
\mathbb{E}\big[f(\ell_1+\ell_1')\big] - &\mathbb{E}\big[f(\ell_0+\ell_0')\big] \\
&= \int_0^1 \int_{\mathbb{C}^{2n}} \frac{1}{(2\pi \mathrm{i})^{2n}} f(\phi\bar{\phi}+\phi'\bar{\phi}') \frac{\partial}{\partial\alpha}\left(\frac{1}{|G_\alpha|^2} e^{\langle \phi, Q_\alpha \bar{\phi}\rangle + \langle \phi', Q_\alpha \bar{\phi}'\rangle}\right) d\bar{\phi}\, d\phi\, d\phi'\, d\bar{\phi}'\, d\alpha \\
&= \int_0^1 \int_{\mathbb{C}^n} \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^{n}} f(\phi\bar{\phi}+\phi'\bar{\phi}') \frac{\partial}{\partial\alpha}\left(\frac{1}{|G_\alpha|} e^{\langle \phi, Q_\alpha \bar{\phi}\rangle}\right) d\mu_\alpha(\phi')\, d\bar{\phi}\, d\phi\, d\alpha + \\
&\qquad \int_0^1 \int_{\mathbb{C}^n} \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^{n}} f(\phi\bar{\phi}+\phi'\bar{\phi}') \frac{\partial}{\partial\alpha}\left(\frac{1}{|G_\alpha|} e^{\langle \phi', Q_\alpha \bar{\phi}'\rangle}\right) d\mu_\alpha(\phi)\, d\phi'\, d\bar{\phi}'\, d\alpha.
\end{aligned}$$

Now note that the proof of Lemma 2.4.1 shows that for any function $g : \mathbb{R}_+^n \to \mathbb{R}$ with subexponential growth for $k \leq 2$ orders of partial derivative, and any $\alpha \in (0,1)$, we have

$$\int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^n} \frac{\partial}{\partial\alpha}\Big(\frac{1}{|G_\alpha|} e^{\langle \phi, Q_\alpha \bar{\phi}\rangle}\Big) g(\phi\bar{\phi})\, d\bar{\phi}\, d\phi = \int_{\mathbb{C}^n} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha}\Big(\delta_{ij}\partial_j g(\phi\bar{\phi}) + \phi_i\bar{\phi}_j \partial_{ij} g(\phi\bar{\phi})\Big) d\mu_\alpha(\phi).$$

Specifically, if for fixed $l = \phi\bar{\phi}$, we let $g(l) := \int_{\mathbb{C}^n} f(\phi\bar{\phi}+\phi'\bar{\phi}')\, d\mu_\alpha(\phi')$ and apply the above equation to $g(\phi\bar{\phi})$, we get

$$\begin{aligned}
&\int_0^1 \int_{\mathbb{C}^n} \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^{n}} f(\phi\bar{\phi}+\phi'\bar{\phi}') \frac{\partial}{\partial\alpha}\left(\frac{1}{|G_\alpha|} e^{\langle \phi, Q_\alpha \bar{\phi}\rangle}\right) d\mu_\alpha(\phi')\, d\bar{\phi}\, d\phi\, d\alpha \\
&= \int_0^1 \int_{\mathbb{C}^{2n}} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha}\left(\delta_{ij}\partial_j f(\phi\bar{\phi}+\phi'\bar{\phi}') + \phi_i\bar{\phi}_j \partial_{ij} f(\phi\bar{\phi}+\phi'\bar{\phi}')\right) d\mu_\alpha(\phi)\, d\mu_\alpha(\phi')\, d\alpha.
\end{aligned}$$

Similarly, we can fix $\phi, \bar{\phi}$ and let $g(\phi'\bar{\phi}') := \int_{\mathbb{C}^n} f(\phi\bar{\phi} + \phi'\bar{\phi}')d\mu_\alpha(\phi)$. This gives us

$$\begin{aligned}
&\int_0^1 \int_{\mathbb{C}^n} \int_{\mathbb{C}^n} \frac{1}{(2\pi \mathrm{i})^n} f(\phi\bar{\phi} + \phi'\bar{\phi}') \frac{\partial}{\partial \alpha}\left(\frac{1}{|G_\alpha|} e^{\langle \phi', Q_\alpha \bar{\phi}' \rangle}\right) d\mu_\alpha(\phi) d\bar{\phi}' d\phi' d\alpha \\
&= \int_0^1 \int_{\mathbb{C}^{2n}} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha}\left(\delta_{ij}\partial_j f(\phi\bar{\phi} + \phi'\bar{\phi}') + \phi'_i\bar{\phi}'_j \partial_{ij} f(\phi\bar{\phi} + \phi'\bar{\phi}')\right) d\mu_\alpha(\phi) d\mu_\alpha(\phi') d\alpha.
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}
&\mathbb{E}\big[f(\ell_1 + \ell'_1)\big] - \mathbb{E}\big[f(\ell_0 + \ell'_0)\big] \\
&\quad = \int_0^1 \int_{\mathbb{C}^{2n}} \sum_{i,j} \frac{dG_{\alpha,ij}}{d\alpha}\left(2\delta_{ij}\partial_j f(\phi\bar{\phi} + \phi'\bar{\phi}') + (\phi_i\bar{\phi}_j + \phi'_i\bar{\phi}'_j)\partial_{ij} f(\phi\bar{\phi} + \phi'\bar{\phi}')\right) \\
&\qquad\qquad d\mu_\alpha(\phi) d\mu_\alpha(\phi') d\alpha.
\end{aligned}$$

Note that $\forall i \in [n]$,

$$\Delta G_{ii}(2\partial_i f(l) + l_i \partial_{ii} f(l)) \geq 0,$$

by assumption. Also, for any $i \neq j$,

$$\int_{\mathbb{C}^{2n}} \phi_i\bar{\phi}_j \frac{dG_{\alpha,ij}}{d\alpha}\partial_{ij} f(\phi\bar{\phi} + \phi'\bar{\phi}') d\mu_\alpha(\phi) d\mu_\alpha(\phi') = \int_{\mathbb{C}^{2n}} \phi_i\bar{\phi}_j \frac{dG_{\alpha,ij}}{d\alpha}\mathbb{E}\big[\partial_{ij} f(\phi\bar{\phi} + \ell'_\alpha)\big] d\mu_\alpha(\phi).$$

Here ℓ'_α is a 1-permanental process with kernel G_α. Using Lemma A.3.1, we conclude that the integral above has a fixed sign. This guarantees

$$\int_{\mathbb{C}^{2n}} \frac{dG_{\alpha,ij}}{d\alpha}(\phi_i\bar{\phi}_j + \phi'_i\bar{\phi}'_j)\partial_{ij} f(\phi\bar{\phi} + \phi'\bar{\phi}') d\mu_\alpha(\phi) d\mu_\alpha(\phi') \geq 0.$$

Hence $\mathbb{E}[f(\ell_\alpha + \ell'_\alpha)]$ is increasing with $\alpha \in (0,1)$, and $\mathbb{E}[f(\ell_1 + \ell'_1)] \geq \mathbb{E}[f(\ell_0 + \ell'_0)]$. This verifies the case $k = 2$. The general case where k is any positive integer follows from almost the same arguments as above.

Slepian's lemma for 1-permanental processes

We now turn to prove an analog of Slepian's lemma for 1-permanental processes. To do so, we first prove a weaker version as below.

Lemma 2.4.2 (Weak comparision of 1-permanental vectors)**.** *In the same setting as Lemma 2.4.1, if, further, we have $G_{0,ii} = G_{1,ii}$ for all $i \in [n]$ and $G_{1,ij} \geq G_{0,ij}$ for all $i \neq j$, then*

$$\sup_i \ell_{0,i} \succeq \sup_i \ell_{1,i}.$$

Proof. Given $\epsilon > 0$, consider the function $g_\epsilon : \mathbb{R} \to [0, 1]$ as follows,

$$g_\epsilon(x) := \begin{cases} 1 & \text{if } x < 0, \\ 1 - 6\epsilon^{-5}x^5 + 15\epsilon^{-4}x^4 - 10\epsilon^{-3}x^3 & \text{if } x \in [0, \epsilon), \\ 0 & \text{if } x \geq \epsilon. \end{cases}$$

We know that g_ϵ is decreasing and twice differentiable. Given $s_1, \cdots, s_n \geq 0$, define $f_\epsilon(l_1, ..., l_n) := \prod_{i=1}^n g_\epsilon(l_i - s_i)$. Note (even though we do not use this property) that for any $i \in [n]$, we have $\partial_i f(l) \leq 0$. And for any $i \neq j$, $\partial_{ij} f(l) \geq 0$. Hence, by Lemma 2.4.1, we have

$$\mathbb{E}[f_\epsilon(\ell_1)] \geq \mathbb{E}[f_\epsilon(\ell_0)].$$

Taking the limit as $\epsilon \downarrow 0$ on both sides, we have

$$\lim_{\epsilon \downarrow 0} \mathbb{E}[f_\epsilon(\ell_1)] \geq \lim_{\epsilon \downarrow 0} \mathbb{E}[f_\epsilon(\ell_0)].$$

By the dominated convergence theorem, since $\mathbb{E}[|f_\epsilon(\ell_1)|] \leq \mathbb{E}[1] = 1$, we can interchange the integral with the limit and get

$$\mathbb{E}[\lim_{\epsilon \downarrow 0} f_\epsilon(\ell_1)] \geq \mathbb{E}[\lim_{\epsilon \downarrow 0} f_\epsilon(\ell_0)].$$

But $\lim_{\epsilon \downarrow 0} f_\epsilon(l) = \prod_{i=1}^n \mathbf{1}_{(-\infty, s_i]}(l_i)$. We conclude that

$$\mathbb{P}(\ell_{1,i} \leq s_i, \forall i \in [n]) \geq \mathbb{P}(\ell_{0,i} \leq s_i, \forall i \in [n]).$$

Since this is true for any $\{s_i \geq 0 : i \in [n]\}$, we conclude that $\sup_i \ell_{0,i} \succeq \sup_i \ell_{1,i}$. $\square$

Boosting the above lemma using a scaling trick, we have the following Slepian lemma for 1-permanental vectors.

Theorem 2.4.3 (Slepian lemma for 1-permanental vectors)**.** *In the same setting as Lemma 2.4.1, if, further, we have $G_{1,ii} \leq G_{0,ii}$ for all $i \in [n]$ and $G_{1,ij} \geq G_{0,ij}$ for all $i \neq j$, then*

$$\sup_i \ell_{0,i} \succeq \sup_i \ell_{1,i}.$$

Proof. To see this, we scale the 1-permanental vector ℓ_0 by $c = (c_1, \cdots, c_n)$ where $c_i := \frac{G_{1,ii}}{G_{0,ii}} \leq 1$(we use the convention that $\frac{0}{0} = 1$). We claim that the new vector $c \odot \ell_0$, where $\odot$ is the standard Hadamard product between vectors, is also a 1-permanental vector, since for any $\lambda \in \mathbb{R}_+^n$, we have

$$\mathbb{E}[e^{-\langle \lambda, c \odot \ell_0 \rangle}] = \mathbb{E}[e^{-\langle \lambda \odot c, \ell_0 \rangle}] = |I + \Lambda C G_0|^{-1}.$$

Here, we defined $C := \text{diag}(c)$ and used the moment generating function of ℓ_0. Then by definition, $c \odot \ell_0$ is a 1-permanental vector with kernel $\tilde{G} := CG_0$. Note that $\forall i \in [n]$,

$\tilde{G}_{ii} = c_i G_{0,ii} = G_{1,ii}$; and $\forall i \neq j$, $\tilde{G}_{ij} = c_i G_{0,ij} \leq G_{0,ij} \leq G_{1,ij}$. Therefore, we can use Lemma 2.4.2 to conclude that $\sup_i c_i \ell_{0,i} \succeq \sup_i \ell_{1,i}$. Here for two scalar random variables X, Y, X (first-order) stochastically dominates Y, i.e., $X \succeq Y$ iff $\forall x \in \mathbb{R}$, $\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x)$. Let $c_* := \sup_i c_i = \sup_i \frac{G_{1,ii}}{G_{0,ii}}$, then we conclude that

$$\sup_i \ell_{0,i} \succeq c_* \sup_i \ell_{0,i} \succeq \sup_i c_i \ell_{0,i} \succeq \sup_i \ell_{1,i}.$$

□

2.5 Symmetrization bound on cover time

Our main result in this section will be a symmetrization bound on the cover time of a non-reversible Markov chain on a finite set. An important property of this upper bound is that it coincides with that for a random walk on a graph, i.e. a reversible Markov chain on a finite set, given in [36], which is tight up to constant factors in the reversible case. To arrive at our result we apply the density formula, with the generalized second Ray-Knight isomorphism theorem for non-reversible Markov chains, to study the local time process at inverse local time.

Let $P = (P_{ij} : i, j \in [n])$ be an irreducible transition probability matrix on the state space $[n]$, with stationary distribution $\pi = (\pi_i : i \in [n])$. Let the normalized Laplacian be $Q = \Pi(P - I)$, where Π denotes the diagonal matrix with diagonal entries given by the π_i. Let $\{X_t, t \geq 0\}$ denote the continuous time Markov chain with rate matrix Q. Note that the cover time being discussed is defined in terms of the underlying discrete-time Markov chain with transition probability matrix P, as is conventional in the literature on cover times.

We will decompose Q as in Eq. (2.4). The Green's kernel associated with killing upon hitting node 1 is

$$\tilde{G} = (-Q_{**})^{-1}. \tag{2.10}$$

Recall that the local time process is defined as in Eq. (2.6). For a fixed state $a \in [n]$, the inverse local time for any $r \geq 0$ is $\tau_{\text{inv}}(r) := \inf\{t : \mathcal{L}_t(a) \geq r\}$. We may as well renumber the states so that $a = 1$. We will then consider the permanental process with kernel G defined by

$$G := \begin{pmatrix} 1 & \mathbf{1}^T \\ \mathbf{1} & \tilde{G} + \mathbf{1}\mathbf{1}^T \end{pmatrix}.$$

One can think of it as the Green's kernel associated with killing once the process $\{X_t : t \geq 0\}$ has spent an exponentially distributed random time with mean 1 at node 1 [46]. Equivalently, $G = (e_1 e_1^T - Q)^{-1}$ in light of Theorem 2.3.4. Let ℓ be a 1-permanental process with kernel G. Then from Theorem 2.3.4 we know

$$\{\mathcal{L}_{\tau_{\text{inv}}(r)} + \ell \,|\, \ell_1 = 0\} \overset{d}{=} \{\ell \,|\, \ell_1 = r\}.$$

Note that $\{\ell \,|\, \ell_1 = 0\}$ is distributed as a permanental vector with kernel $\tilde{G}$, so we have a density formula for it. For $\{\ell \,|\, \ell_1 = r\}$, we will use Lemma 2.2.2 to bound its density.

We first derive a tail bound for $\tau_{\text{inv}}(t)$, which will be useful when bounding the cover time.

Lemma 2.5.1 (Tail bound lemma)**.** *For an irreducible Markov chain on state space $[n]$, let $S_{**} = \frac{1}{2}(Q_{**} + Q_{**}^T)$ and denote by σ^2 the largest diagonal term of $\frac{1}{2}(-S_{**})^{-1}$. Then for any $\lambda > 0$, we have*

$$\mathbb{P}\bigg(\Big(\sum_{i=1}^n \pi_i \mathcal{L}_{\tau_{\text{inv}}(r)}(i)\Big)^{\frac{1}{2}} - \sqrt{r} \geq \sqrt{\lambda}\sigma\bigg) \leq 5\gamma e^{-\lambda/8}.$$

Here

$$\gamma := \frac{|2\tilde{G}|}{|\tilde{G} + \tilde{G}^T|}. \tag{2.11}$$

Proof. First, in light of the proof of Theorem 2.3.4, we have

$$\rho_r(l_2, \cdots, l_n) = \int_{(0,2\pi)^{n-1}} \frac{1}{(2\pi)^{n-1}|\tilde{G}|} \exp\Big(\langle \phi_* - \sqrt{r}\mathbf{1}, Q_{**}(\bar{\phi}_* - \sqrt{r}\mathbf{1})\rangle\Big) d\theta_*.$$

Also, we have

$$\Big\{\sum_{i=1}^n \pi_i \mathcal{L}_{\tau_{\text{inv}}(r)}(i) + \sum_{i=1}^n \pi_i \ell_i \,\Big|\, \ell_1 = 0\Big\} \overset{d.}{=} \Big\{\sum_{i=1}^n \pi_i \ell_i \,\Big|\, \ell_1 = r\Big\}.$$

Hence we conclude that $\{\sum_{i=1}^n \pi_i \ell_i \,|\, \ell_1 = r\} \succeq \sum_{i=1}^n \pi_i \mathcal{L}_{\tau_{\text{inv}}(r)}(i)$. Let

$$\mathcal{R} := \{l : \sum_{i=1}^n \pi_i l_i \geq c\}.$$

Then we have the following bound:

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^n \pi_i \mathcal{L}_{\tau_{\text{inv}}(r)}(i) \geq c\Big) &\leq \int_{\mathcal{R}} \rho_r(l_2, \cdots, l_n) dl_* \\
&= \int_{\mathcal{R}} \int_{(0,2\pi)^{n-1}} \frac{1}{(2\pi)^{n-1}|\tilde{G}|} \exp\Big(\langle \phi_* - \sqrt{r}\mathbf{1}, Q_{**}(\bar{\phi}_* - \sqrt{r}\mathbf{1})\rangle\Big) d\theta_* dl_* \\
&\leq \int_{\mathcal{R}} \int_{(0,2\pi)^{n-1}} \frac{1}{(2\pi)^{n-1}|\tilde{G}|} \exp\Big(\langle \phi_* - \sqrt{r}\mathbf{1}, S_{**}(\bar{\phi}_* - \sqrt{r}\mathbf{1})\rangle\Big) d\theta_* dl_* \\
&= \gamma \int_{\mathcal{R}} \int_{(0,2\pi)^{n-1}} \frac{|-S_{**}|}{(2\pi)^{n-1}} \exp\Big(\langle \phi_* - \sqrt{r}\mathbf{1}, S_{**}(\bar{\phi}_* - \sqrt{r}\mathbf{1})\rangle\Big) d\theta_* dl_*.
\end{aligned}$$

Note that the integrand is now the density of $Z + \binom{\sqrt{r}\,\mathbf{1}}{\mathbf{0}}$, where $Z = \eta + \mathrm{i}\xi$ is a complex Gaussian random vector with η, ξ being i.i.d. $\mathcal{N}(0, \frac{1}{2}(-S_{**})^{-1})$ random variables. Therefore,

$$\mathbb{P}\left(\sum_{i=1}^{n} \pi_i \mathcal{L}_{\tau_{\mathrm{inv}}(r)}(i) \geq c\right) \leq \gamma \mathbb{P}\big((\eta + \sqrt{r}\mathbf{1})^2 + \xi^2 \in \mathcal{R}\big)$$
$$= \gamma \mathbb{P}\big((\eta^2 + \xi^2 + 2\sqrt{r}\eta + r\mathbf{1}) \in \mathcal{R}\big)$$
$$= \gamma \mathbb{P}\left(\sum_{i=1}^{n} \pi_i(\eta_i^2 + \xi_i^2 + 2\sqrt{r}\eta_i + r) \geq c\right).$$

From Claim 2.2 of [34], we know for any $\lambda > 0$, $\mathbb{P}(\sum_{i=1}^{n} \pi_i \eta_i^2 \geq \lambda\sigma^2) \leq 2e^{-\lambda/4}$, where $\sigma^2 = \max_i \mathbb{E}[\eta_i^2]$. The same is true for $\sum_{i=1}^{n} \pi_i \xi_i^2$. We also know that $\sum_{i=1}^{n} \pi_i \eta_i \prec \mathcal{N}(0, \sigma^2)$, hence $\mathbb{P}(\sum_i \pi_i \eta_i \geq \sqrt{\lambda}\sigma) \leq e^{-\lambda/4}$. Therefore,

$$\mathbb{P}\left(\sum_i \pi_i \eta_i^2 + \sum_i \pi_i \xi_i^2 + 2\sqrt{r}\sum_i \pi_i \eta_i + r \geq \lambda\sigma^2 + 2\sqrt{\lambda r}\sigma + r\right)$$
$$\leq \mathbb{P}\left(\sum_i \pi_i \eta_i^2 \geq \frac{1}{2}\lambda\sigma^2\right) + \mathbb{P}\left(\sum_i \pi_i \xi_i^2 \geq \frac{1}{2}\lambda\sigma^2\right) + \mathbb{P}\left(\sum_i \pi_i \eta_i \geq \sqrt{\lambda}\sigma\right)$$
$$\leq 4e^{-\lambda/8} + e^{-\lambda/4} \leq 5e^{-\lambda/8}.$$

Hence $\mathbb{P}\big(\sum_{i=1}^{n} \pi_i \mathcal{L}_{\tau_{\mathrm{inv}}(r)}(i) \geq (\sqrt{r} + \sqrt{\lambda}\sigma)^2\big) \leq 5\gamma e^{-\lambda/8}$. □

Let us denote the random cover time as $\tau_{\mathrm{cov}} := \inf\{\sum_i \pi_i \mathcal{L}_s(i) :\ s \geq 0,\ \mathcal{L}_s(i) > 0, \forall\, i \in [n]\}$[11]. The cover time t_{cov} is just the maximum expectation of τ_{cov} maximized over all starting states $X_0 \in [n]$. We denote the random cover-and-return time as

$$\tau_{\mathrm{cov}}^{+} := \inf\left\{\sum_i \pi_i \mathcal{L}_s(i) :\ s \geq 0,\ X_s = X_0,\ \mathcal{L}_s(i) > 0, \forall\, i \in [n]\right\}.$$

Similarly, we can define the cover-and-return time t_{cov}^{+}. We remark that using the Markov property, one can easily see $t_{\mathrm{cov}} \leq \mathbb{E}(\tau_{cov}^{+}) \leq 2t_{\mathrm{cov}}$ for any irreducible finite-state chains, with any starting state. See [36] for a short proof.[12] Given $j \in [n]$, the random hitting time, $\tau_{\mathrm{hit}}(j)$, is defined as

$$\tau_{\mathrm{hit}}(j) := \inf\left\{\sum_i \pi_i \mathcal{L}_s(i) :\ s \geq 0,\ X_s = j\right\}.$$

Similarly, define the hitting time as $t_{\mathrm{hit}} = \sup_{i,j\in[n]} \mathbb{E}[\tau_{\mathrm{hit}}(j)|X_0 = i]$. Let us also denote by t'_{cov} and t'_{hit} the cover time and the hitting time of the symmetrized Markov chain with kernel $\frac{Q+Q^T}{2}$, respectively.

We now prove the following symmetrization upper bound on the cover time.

[11] Since we are working with a variable jump rate Markov chain in this section, we need π_i-weighted summation to recover the cover time for a discrete Markov chain.

[12] They only considered symmetric chains, but their proof holds for any non-symmetric chain as well.

Theorem 2.5.2 (Symmetrization upper bound on cover time)**.** *For an irreducible Markov chain on state space $[n]$ with generator Q, we decompose Q as in Eq.* (2.4)*. Let S be defined in terms of Q as in equation* (2.7)*, and decompose S similarly to Q. We write $M := \mathbb{E}\sup_i \eta_i$ for $\eta \sim \mathcal{N}(0, \frac{1}{2}(-S_{**})^{-1})$, and let σ^2 be the maximal diagonal term in $\frac{1}{2}(-S_{**})^{-1}$. Then we have the following upper bound on the cover time*

$$t_{\mathrm{cov}} = O(M^2 + \sigma^2 \log\gamma) = O(t'_{\mathrm{cov}} \log\gamma),$$

where γ is defined in equation (2.11)*.*

Proof. To give an upper bound on the cover time, we need to

- bound $\{\sup_i \ell_i \,|\, \ell_1 = 0\}$ from above asymptotically;
- bound $\{\inf_i \ell_i \,|\, \ell_1 = t\}$ from below asymptotically.

Let η, ξ denote independent $\mathcal{N}(0, \frac{1}{2}(-S_{**})^{-1})$ random vectors. Then by standard results about Gaussian suprema (see, for example, Theorem 5.8 of [17]), we know that for all $\lambda > 0$ we have

$$\mathbb{P}\Big(\sup_i \eta_i > \mathbb{E}\sup_i \eta_i + \sqrt{\lambda}\sigma\Big) \le e^{-\lambda/2}.$$

Since X is symmetric, $-\inf_i \eta_i \overset{d.}{=} \sup_i \eta_i$. Hence we also have

$$\mathbb{P}\Big(-\inf_i \eta_i > \mathbb{E}\sup_i \eta_i + \sqrt{\lambda}\sigma\Big) \le e^{-\lambda/2}.$$

By an application of the union bound, we deduce

$$\mathbb{P}\Big(\sup_i |\eta_i| > M + \sqrt{\lambda}\sigma\Big) = \mathbb{P}\Big(\sup_i \eta_i^2 > \big(M + \sqrt{\lambda}\sigma\big)^2\Big) \le 2e^{-\lambda/2}.$$

For any $c > 0$, let $\mathcal{R}' := \{l : \sup_i l_i \ge c\}$. Then we have the following bound, where $\tilde{G}$ is defined as in equation (2.10),

$$\begin{aligned}
\mathbb{P}\Big(\sup_i \ell_i \ge c \,\Big|\, \ell_1 = 0\Big) &= \int_{\mathcal{R}'} \rho_0(l_2, \cdots, l_n) dl_* \\
&= \int_{\mathcal{R}'} \int_{(0,2\pi)^{n-1}} \frac{1}{(2\pi)^{n-1}|\tilde{G}|} \exp\big(\langle \phi_*, Q_{**}\bar{\phi}_*\rangle\big) d\theta_* dl_* \\
&\le \gamma \int_{\mathcal{R}'} \int_{(0,2\pi)^{n-1}} \frac{|-S_{**}|}{(2\pi)^{n-1}} \exp\big(\langle \phi_*, S_{**}\bar{\phi}_*\rangle\big) d\theta_* dl_* \\
&= \gamma \mathbb{P}\big((\eta^2 + \xi^2) \in \mathcal{R}'\big).
\end{aligned}$$

Recall that here η, ξ are independent $\mathcal{N}(0, \frac{1}{2}(-S_{**})^{-1})$ random vectors. This allows us to conclude that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_i \ell_i \geq 2(M+\sqrt{\lambda}\sigma)^2 \,\Big|\, \ell_1 = 0\Big) \\
&\qquad\qquad \leq \gamma\mathbb{P}\Big(\sup_i(\eta_i^2 + \xi_i^2) \geq 2(M + \sqrt{\lambda}\sigma)^2\Big) \\
&\qquad\qquad \leq \gamma\mathbb{P}\Big(\sup_i \eta_i^2 + \sup_i \xi_i^2 \geq 2(M + \sqrt{\lambda}\sigma)^2\Big) \\
&\qquad\qquad \leq \gamma\mathbb{P}\Big(\sup_i \eta_i^2 \geq (M + \sqrt{\lambda}\sigma)^2\Big) + \gamma\mathbb{P}\Big(\sup_i \xi_i^2 \geq (M + \sqrt{\lambda}\sigma)^2\Big) \\
&\qquad\qquad \leq 4\gamma e^{-\lambda/2}.
\end{aligned}$$

Similarly, note that for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\inf_i \ell_i \leq -2\sqrt{t}M - 2\sqrt{\lambda t}\sigma + t \,\Big|\, \ell_1 = t\Big) \\
&\qquad\qquad \leq \gamma\mathbb{P}\Big(\inf_i \big((\eta_i + \sqrt{t})^2 + \xi_i^2\big) \leq -2\sqrt{t}M - 2\sqrt{\lambda t}\sigma + t\Big) \\
&\qquad\qquad \leq \gamma\mathbb{P}\Big(\inf_i (2\sqrt{t}\eta_i + t) \leq -2\sqrt{t}M - 2\sqrt{\lambda t}\sigma + t\Big) \\
&\qquad\qquad = \gamma\mathbb{P}\Big(\inf_i \eta_i \leq -M - \sqrt{\lambda}\sigma\Big) \\
&\qquad\qquad \leq \gamma e^{-\lambda/2}.
\end{aligned}$$

Now we take $t = 9(M + \sqrt{\lambda}\sigma)^2$, so that $-2\sqrt{t}M - 2\sqrt{\lambda t}\sigma + t = 3(M + \sqrt{\lambda}\sigma)^2 > 2(M + \sqrt{\lambda}\sigma)^2$. Therefore, with probability $\geq 1 - 5\gamma e^{-\lambda/2}$, we have $\tau_{\text{cov}}^+ \leq \sum_i \pi_i \mathcal{L}_{\tau_{\text{inv}}(t)}(i)$ for such t. Also, by Lemma 2.5.1, with probability $\geq 1 - 5\gamma e^{-\lambda/8}$, we know that $\sum_i \pi_i \mathcal{L}_{\tau_{\text{inv}}(t)}(i) \leq (\sqrt{t} + \sqrt{\lambda}\sigma)^2 = (3M + 4\sqrt{\lambda}\sigma)^2$.

So in conclusion,

$$\mathbb{P}(\tau_{\text{cov}}^+ \geq (3M + 4\sqrt{\lambda}\sigma)^2) \leq 10\gamma e^{-\lambda/8}, \quad \forall \lambda > 0.$$

Specifically, we have $\mathbb{P}(\tau_{\text{cov}}^+ \geq 2((3M)^2 + \lambda(4\sigma)^2)) \leq \mathbb{P}(\tau_{\text{cov}}^+ \geq (3M + 4\sqrt{\lambda}\sigma)^2) \leq 10\gamma e^{-\lambda/8}$. Since $\gamma \geq 1$, we know that when $\lambda \geq 16 \log \gamma$, we have $\gamma e^{-\lambda/8} \leq (\gamma e^{-\lambda/16}) e^{-\lambda/16} \leq e^{-\lambda/16}$.

Therefore, we conclude that

$$\begin{aligned}
\mathbb{E}\tau^+_{\text{cov}} &= \int_0^\infty \mathbb{P}(\tau^+_{\text{cov}} \geq t)dt \\
&= \int_0^{18M^2} \mathbb{P}(\tau^+_{\text{cov}} \geq t)dt + \int_0^\infty \mathbb{P}(\tau^+_{\text{cov}} \geq 18M^2 + 32\lambda\sigma^2) \cdot 32\sigma^2 d\lambda \\
&\leq \int_0^{18M^2} 1 \cdot dt + \int_0^{16\log\gamma} 1 \cdot 32\sigma^2 d\lambda + \int_{16\log\gamma}^\infty \mathbb{P}(\tau^+_{\text{cov}} \geq 18M^2 + 32\lambda\sigma^2) \cdot 32\sigma^2 d\lambda \\
&\leq 18M^2 + 16\log\gamma \cdot 32\sigma^2 + \int_{16\log\gamma}^\infty 10e^{-\lambda/16} \cdot 32\sigma^2 d\lambda \\
&\leq 18M^2 + 512\sigma^2\log\gamma + \int_0^\infty 10e^{-\lambda/16} \cdot 32\sigma^2 d\lambda \\
&= 18M^2 + 512\sigma^2\log\gamma + 5120\sigma^2.
\end{aligned}$$

Therefore, we deduce that $\mathbb{E}\tau^+_{\text{cov}} = O(M^2 + \sigma^2\log\gamma)$. Since $t'_{\text{cov}} = \Theta(M^2 + \sigma^2)$, we obtain $t_{\text{cov}} = \Theta(\mathbb{E}\tau^+_{\text{cov}}) = O(t'_{\text{cov}}\log\gamma)$. $\square$

We remark that when $\gamma = 2^{O(M^2/\sigma^2)}$, i.e., when the chain does not deviate too much from symmetry, using the remarkable theorem of Ding-Lee-Peres[36], we have $t_{\text{cov}} = O(t'_{\text{cov}})$. We also note that since non-reversibility reduces hitting time, i.e., $t_{\text{hit}} \leq t'_{\text{hit}}$(Corollary 3.1 of [60]), then by Mathews' bound (see, e.g., Theorem 11.2 of [67]), we know

$$t_{\text{cov}} = O(t_{\text{hit}}\log n) = O(t'_{\text{hit}}\log n) = O(t'_{\text{cov}}\log n).$$

However, this upper bound does not provide a smooth transition between the tight reversible bound and the non-reversible scenarios as Theorem 2.5.2.

2.6 Summary

In this chapter, we used a density formula approach to study several problems on non-reversible finite-state Markov chains and the associated 1-permanental processes, as well as a larger class of related 1-permanental processes. After deriving a density formula for this broad family of 1-permanental processes, we used it, together with non-reversible Ward identities, to give unified proofs of several isomorphism theorems for non-reversible Markov chains. We then applied the density formulas to derive Kahane-type and Slepian-type comparison inequalities for 1-permanental processes. Finally, we applied the technique to bounding the cover time of non-reversible Markov chains. The derived bound becomes tight in the reversible case [36].

We believe that the twisted Gaussian density is a good substitute for the Gaussian free field (GFF) in the non-reversible case (it indeed coincides with GFF when the Markov chain is reversible). In light of this, for many previous results about reversible Markov chains proved via the GFF, one should be able to give a natural generalization (perhaps suboptimal) in non-reversible scenarios using the twisted Gaussian density instead.

Another intriguing research question is as follows. In the Markovian case, the coordinates of the 1-permanental vector corresponding to the kernel with psd symmetric part are the total loop soup local times, where loops are sampled at rate 1. Is there a similar picture for the general kernel with a symmetric psd part?

Chapter 3

The Double Dixie Cup Problem on General Graphs

Building on the coupon-collector and double-Dixie-cup discussion in Chapter 1, this chapter studies the corresponding double Dixie cup problem for random walks on graphs. Specifically, we study the *marginal time to the second cover* of a simple random walk on a finite connected graph (V, E): the additional time after the first cover time needed for every vertex to be visited at least twice.

Let $M := \mathbb{E}[\max_{v\in V} \eta(v)]$ denote the expected maximum of the Gaussian free field pinned at v_0, and let $R := \max_{u,v\in V} r(u, v)$ denote the maximal effective resistance; see Section 3.1 for precise definitions and references. Note that M does not depend on the choice of v_0. We prove that there exist absolute constants $C, c > 0$ such that

$$\mathbb{P}_{v_0}\Big(\tau_{\text{cov}}^{(2)} - \tau_{\text{cov}} > C\big(|E|M\sqrt{R}\big)\beta + C\big(|E|R\big)\beta^2\Big) \;\leq\; Ce^{-c\beta^2}$$

for all $\beta \geq 1$, for every finite connected graph $G = (V, E)$ and every starting vertex v_0.

On a full b-ary tree of height h (fixed $b \geq 2$), starting from the root v_0, there exists $C_b < \infty$ such that

$$\mathbb{P}_{v_0}\big(\tau_{\text{cov}}^{(2)} - \tau_{\text{cov}} \;\leq\; C_b\,|E|\,h\log h\big) \;\geq\; 1 - o(1).$$

This is proved via a variance analysis for inverse local time in the spirit of [38].

3.1 Introduction

We use the CT-SRW and cover-time notation from Chapter 1. In particular, $\mathcal{L}_t(v)$ denotes the degree-normalized local time, $\tau_{\text{inv}}(t)$ the inverse local time at v_0, $N_t(v)$ the jump-time visit count, τ_{cov} and $\tau_{\text{cov}}^{(2)}$ the first and second cover times, and $\Delta_{\text{cov}} := \tau_{\text{cov}}^{(2)} - \tau_{\text{cov}}$. We write $f(n) \asymp g(n)$ or $f(n) = \Theta(g(n))$ to mean $c_1 g(n) \leq f(n) \leq c_2 g(n)$ for absolute constants $c_1, c_2 > 0$.

Since holding times have unit mean, the discrete-time and continuous-time versions of the random walk differ only by constant factors in asymptotic expressions. We therefore work throughout in continuous time, for compatibility with local-time methods.

Our main theorem gives a general upper bound on the marginal time to the second cover in terms of classical graph parameters. Let $r(u,v)$ denote the effective resistance in the unit-resistance network associated with $G=(V,E)$, and write $R := \max_{u,v\in V} r(u,v)$; see [41, 68]. Let $(\eta(v), v\in V)$ denote the GFF pinned at v_0, and set $M := \mathbb{E}[\max_{v\in V}\eta(v)]$; the distribution of M does not depend on the choice of v_0. See [37, 34] for the role of the GFF in cover-time analysis and [77] for background on Gaussian processes and local times.

Using the generalized second Ray–Knight isomorphism theorem (Theorem 1.1.1) together with the concentration lemma for inverse local time (Lemma 1.1.2) and an excursion decomposition of the random walk, we prove in Theorem 3.2.1 that for any $v_0\in V$,

$$\mathbb{E}_{v_0}\big[\Delta_{\mathrm{cov}}\big] \;=\; O\big(|E|\,M\sqrt{R}\big).$$

Since $t_{\mathrm{hit}} = \Theta(|E|R)$ [23, 68] and $t_{\mathrm{cov}} = \Theta(|E|M^2)$ [37, 34], this is equivalently

$$\mathbb{E}_{v_0}\big[\tau^{(2)}_{\mathrm{cov}} - \tau_{\mathrm{cov}}\big] \;=\; O\big(\sqrt{t_{\mathrm{hit}}\,t_{\mathrm{cov}}}\big).$$

For full b-ary trees of height h, we obtain a sharper bound $O(|E|\,h\log h)$ with high probability.

The rest of this chapter is organized as follows. Section 3.2 states the main results. Section 3.3 contains the proofs, including the excursion decomposition and the inverse-local-time arguments.

3.2 Main results on marginal cover time

We adopt the notation and conventions of Chapter 1. In particular, $\mathcal{L}^v_t$ denotes the degree-normalized continuous-time local time, $\tau_{\mathrm{inv}}(t)$ denotes the inverse local time, and $N_t(v)$ denotes the jump-time visit count. We also use G^{v_0}, η, R, and M as in Section 1.1, together with the standard estimates

$$t_{\mathrm{hit}} = \Theta(|E|R), \qquad t_{\mathrm{cov}} = \Theta(|E|M^2), \qquad R \le 2\pi M^2.$$

The generalized second Ray–Knight isomorphism theorem (Theorem 1.1.1) and the inverse-local-time concentration lemma (Lemma 1.1.2) are the main tools from the Overview used below.

Throughout this section, $(X_t)_{t\ge 0}$ denotes the CT-SRW on a finite connected graph $G=(V,E)$ with i.i.d. Exp(1) holding times, started from $v_0\in V$.

A general bound for random walk on graphs

Our main theorem gives a sub-Gaussian tail bound for Δ_{cov} at the scale $|E|M\sqrt{R}$, and hence an expectation bound of the same order. The proof is given in Section 3.3.

Theorem 3.2.1 (General bound on the marginal time to the second cover)**.** *There exist absolute constants $C, c > 0$ such that for every finite connected graph $G = (V, E)$, every starting vertex $v_0 \in V$, and every $\beta \geq 1$,*

$$\mathbb{P}_{v_0}\Big(\Delta_{\mathrm{cov}} > C\,|E|\,M\sqrt{R}\,\beta \,+\, C\,|E|\,R\,\beta^2\Big) \,\leq\, Ce^{-c\beta^2}. \tag{3.1}$$

In particular, we have $\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] = O(|E|\,M\sqrt{R})$, which further implies that

$$\mathbb{E}_{v_0}\big[\Delta_{\mathrm{cov}}\big] \,=\, O\big(\sqrt{t_{\mathrm{hit}}\,t_{\mathrm{cov}}}\big).$$

Corollary 3.2.2. *If $t_{\mathrm{hit}} = o(t_{\mathrm{cov}})$ along a graph sequence, then we have*

$$\mathbb{E}_{v_0}\big[\Delta_{\mathrm{cov}}\big] \,=\, o(t_{\mathrm{cov}}).$$

Remark 3.2.2.1. *Applying Theorem 3.2.1 for some typical choices of the order of $|E|$, R, and M in Table 1.2 yields the following bounds (up to universal constants):*

- *2D grid/torus and full b-ary tree ($b \geq 2$ fixed): $|E| \asymp n$, $R \asymp \log n$, $M \asymp \log n$, hence*

$$\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] = O\big(n(\log n)^{3/2}\big) \ll t_{\mathrm{cov}} = O(n(\log n)^2).$$

- *d-regular expanders and star graphs: $|E| \asymp n$, $R \asymp 1$, $M \asymp \sqrt{\log n}$, hence*

$$\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] = O\big(n\sqrt{\log n}\big) \ll t_{\mathrm{cov}} = O(n\log n).$$

- *Hypercube ($n = 2^d$): $|E| \asymp n\log n$, $R \asymp 1/\log n$, $M \asymp 1$, hence*

$$\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] = O\big(n\sqrt{\log n}\big) \ll t_{\mathrm{cov}} = O(n\log n).$$

The same substitutions can be made in the tail bound (3.1) *to obtain explicit high-probability estimates.*

We remark that the correct order of $\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}]$ for d-regular expanders and star graphs is known to be $\Theta(n\log\log n)$ [86]. So our general bound is not tight for every instance. However, it applies to every finite-state reversible random walk, and rigorously shows that $\mathbb{E}_{v_0}[\Delta_{\mathrm{cov}}] \ll t_{\mathrm{cov}}$ for many graph families of interest, so long as $t_{\mathrm{hit}} \ll t_{\mathrm{cov}}$ (see Table 1.1).

We also have the following improved result for full b-ary trees.

Theorem 3.2.3 (Improved bound for full b-ary trees)**.** *Fix $b \geq 2$. Let $T_{b,h}$ be the full b-ary tree of height h rooted at v_0 (so $|E| \asymp b^h$ and $|V| = (b^{h+1}-1)/(b-1)$). Then there exists a constant $C_b < \infty$ such that, as $h \to \infty$,*

$$\mathbb{P}_{v_0}\Big(\Delta_{\mathrm{cov}} \,\leq\, C_b\,|E|\,h\log h\Big) \,\geq\, 1 - o(1).$$

The proof of this result for trees is given in Section 3.3.

3.3 Proofs of the main results

In this section, we demonstrate proofs of the claimed results.

Proof of Theorem 3.2.1

We begin by proving some useful lemmas.

Lemma 3.3.1 (Bounds on the hitting time). *In the same setting as Theorem 1.1.1, the hitting time of the graph random walk satisfies* $|E|R \le t_{\mathrm{hit}} \le 2|E|R$.

We include a proof in Section B.1 of the appendix.

Lemma 3.3.2 (Lower bound on the maximum effective resistance). *In the same setting as Theorem 1.1.1, we have* $d_{\min}R \ge 1$, *where* $d_{\min} := \min_{v\in V} d(v)$.

We include a proof in Section B.2 of the Appendix.

Fix $v_0 \in V$. Recall that $\tau_{\mathrm{inv}}(t)$ is a stopping time defined in terms of the local time at v_0. Since $\tau_{\mathrm{inv}}(t)$ depends on the continuous holding times, it is not a priori obvious how to describe the law of the vector of local times $\mathcal{L}_{\tau_{\mathrm{inv}}(t)}$ conditional on the vector of discrete visit counts $N_{\tau_{\mathrm{inv}}(t)}$. The following lemma shows that, for each $v \neq v_0$, the conditional distribution of $\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)$ given $N^v_{\tau_{\mathrm{inv}}(t)}$ is given by a Gamma distribution.

Lemma 3.3.3 (Excursion decomposition at v_0). *Consider the continuous-time walk on a finite connected graph* $G = (V, E)$ *with i.i.d.* $\mathrm{Exp}(1)$ *holding times. Fix* $v_0 \in V$ *and* $t > 0$. *Then for any* $v \neq v_0$ *and any* $k \in \mathbb{N}$,

$$\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) \,\big|\, \{N_{\tau_{\mathrm{inv}}(t)}(v) = k\} \stackrel{d}{=} \mathrm{Gamma}(k, d(v)),$$

where $\mathrm{Gamma}(k, d)$ *denotes the Gamma distribution with shape* k *and rate* d, *i.e. the distribution on* $(0,\infty)$ *with probability density function* $\frac{d^k}{\Gamma(k)}x^{k-1}e^{-dx}$, $x > 0$. *Equivalently, it has scale parameter* $1/d$.

Proof. One can construct the continuous walk stopped at $\tau_{\mathrm{inv}}(t)$ via the following procedure:

(a) (Number of excursions) Determine the number of excursions from $v_0 \in V$ up to $\tau_{\mathrm{inv}}(t)$ as $Q \sim \mathrm{Pois}(d(v_0)t)$. This also pins down the number of visits to v_0 as $N_{\tau_{\mathrm{inv}}(t)}(v_0) = Q+1$.

(b) (Generate excursions) Generate Q discrete excursions (via the excursion measure associated with v_0), which will determine the total number of visits to any other node $v \neq v_0$, and hence determine $N_{\tau_{\mathrm{inv}}(t)}(v)$.

(c) (Holding times) Given the number of discrete visits $N_{\tau_{\mathrm{inv}}(t)}(v) = n_v$ for each $v \in V$, generate $\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) \sim \mathrm{Gamma}(n_v, d(v))$ for each $v \neq v_0$ independently.

□

Lemma 3.3.3 shows that, for $v \neq v_0$, the local time $\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)$ conditional on the discrete visit count $N_{\tau_{\mathrm{inv}}(t)}(v)$ behaves as if we first sampled the discrete path (hence $N_{\tau_{\mathrm{inv}}(t)}(v)$) and then attached i.i.d. exponential holding times at each visit.

Theorem 3.3.4 (Second cover level t)**.** *Let $\eta := (\eta(v), v \in V)$ denote the pinned GFF, pinned at v_0, i.e. $\eta(v_0) = 0$. Write $R := \max_{u,v} r(u,v)$ and let $M := \mathbb{E}[\max_{v\in V} \eta(v)]$, whose distribution does not depend on the choice of v_0. Let $d_{\min} := \min_{v\in V} d(v)$. Fix $\beta \geq 1$ and set $t := \frac{1}{2}(M + \beta\sqrt{R})^2$. Then*

$$\mathbb{P}_{v_0}\big(\min_{v\in V} N_{\tau_{\mathrm{inv}}(t)}(v) \leq 1\big) \;\leq\; 7e^{-\beta^2/16}.$$

Proof. Let η' be an independent copy of η. Define

$$\mathcal{E}_{\mathrm{RK}} := \left\{ \min_{v\in V} \big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) + \tfrac{1}{2}\,\eta(v)^2\big) \;\geq\; \tfrac{\beta^2}{8}\,R \right\}.$$

By the generalized second Ray–Knight theorem (Theorem 1.1.1), the random vector $\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) + \frac{1}{2}\eta(v)^2\big)_{v\in V}$ has the same distribution as $\big(\frac{1}{2}(\eta'(v) + \sqrt{2t})^2\big)_{v\in V}$. Since $\sqrt{2t} = M + \beta\sqrt{R}$,

$$\begin{aligned}
\mathbb{P}_{v_0}(\mathcal{E}_{\mathrm{RK}}^c) &= \mathbb{P}_{v_0}\Big(\min_v \tfrac{1}{2}(\eta'(v) + \sqrt{2t})^2 < \tfrac{\beta^2}{8} R\Big) \\
&\leq \mathbb{P}_{v_0}\Big(\max_v(-\eta'(v)) > M + \tfrac{\beta}{2}\sqrt{R}\Big) \\
&\leq e^{-\beta^2/8},
\end{aligned} \tag{3.2}$$

where the last inequality is Borell–TIS for $S^- := \max_v(-\eta'(v))$ with variance proxy $\sup_v \mathrm{var}(\eta'(v)) = \sup_v r(v_0, v) \leq R$; see [16, 103, 3].

Let v^* be a minimizer of $v \mapsto N_{\tau_{\mathrm{inv}}(t)}(v)$ (ties broken arbitrarily). Then

$$\begin{aligned}
\mathbb{P}_{v_0}\big(\min_v N_{\tau_{\mathrm{inv}}(t)}(v) \leq 1\big) \leq \mathbb{P}_{v_0}(\mathcal{E}_{\mathrm{RK}}^c) &+ \mathbb{P}_{v_0}\big(N_{\tau_{\mathrm{inv}}(t)}(v^*) \leq 1,\, v^* \neq v_0,\, \mathcal{E}_{\mathrm{RK}}\big) \\
&+ \mathbb{P}_{v_0}\big(v^* = v_0,\, N_{\tau_{\mathrm{inv}}(t)}(v_0) \leq 1\big).
\end{aligned}$$

Step 1: the branch $v^* \neq v_0$. We split according to $N_{\tau_{\mathrm{inv}}(t)}(v^*) = 0$ or $N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1$.

Case $N_{\tau_{\mathrm{inv}}(t)}(v^*) = 0$. On this event $\mathcal{L}^{v^*}_{\tau_{\mathrm{inv}}(t)} = 0$, hence $\mathcal{E}_{\mathrm{RK}}$ forces

$$\tfrac{1}{2}\,\eta(v^*)^2 \;\geq\; \tfrac{\beta^2}{8}\,R \quad\Rightarrow\quad |\eta(v^*)| \;\geq\; \tfrac{\beta}{2}\,\sqrt{R}.$$

Since v^* is measurable with respect to the walk (hence independent of η),

$$\begin{aligned}
&\mathbb{P}_{v_0}\big(N_{\tau_{\mathrm{inv}}(t)}(v^*) = 0, v^* \neq v_0, \mathcal{E}_{\mathrm{RK}}\big) && (3.3)\\
&\qquad \le \mathbb{E}_{v_0}\Big[\mathbf{1}\{N_{\tau_{\mathrm{inv}}(t)}(v^*) = 0,\, v^* \neq v_0\}\,\mathbb{P}_{v_0}\big(|\eta(v^*)| \ge \tfrac{\beta}{2}\sqrt{R}\,\big|\, v^*\big)\Big]\\
&\qquad \le \sup_{v\neq v_0} \mathbb{P}_{v_0}\Big(|\eta(v)| \ge \tfrac{\beta}{2}\sqrt{R}\Big) \;\le\; 2e^{-\beta^2/8}. && (3.4)
\end{aligned}$$

Case $N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1$. On the event $\{N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1,\, v^* \neq v_0,\, \mathcal{E}_{\mathrm{RK}}\}$ we have $\frac{1}{2}\eta(v^*)^2 + \mathcal{L}^{v^*}_{\tau_{\mathrm{inv}}(t)} \ge \frac{\beta^2}{8}R$. Condition on the walk and on v^* (equivalently, on the discrete visit counts); then by Lemma 3.3.3,

$$\mathcal{L}^{v^*}_{\tau_{\mathrm{inv}}(t)} \,\big|\, \{N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1,\, v^*\} \;\sim\; \mathrm{Exp}(d(v^*)),$$

and this holding-time randomness is independent of $\eta(v^*)$. Therefore,

$$\begin{aligned}
&\mathbb{P}_{v_0}\big(N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1,\, v^* \neq v_0,\, \mathcal{E}_{\mathrm{RK}}\big)\\
&\qquad \le \mathbb{E}_{v_0}\Big[\mathbf{1}\{N_{\tau_{\mathrm{inv}}(t)}(v^*) = 1,\, v^* \neq v_0\}\,\mathbb{P}_{v_0}\Big(\tfrac{1}{2}\eta(v^*)^2 + \mathcal{L}^{v^*}_{\tau_{\mathrm{inv}}(t)} \ge \tfrac{\beta^2}{8}R \,\Big|\, v^*\Big)\Big]\\
&\qquad \le \sup_{v\neq v_0} \mathbb{P}_{v_0}\Big(\tfrac{1}{2}\eta(v)^2 + Z_v \ge \tfrac{\beta^2}{8}R\Big), && (3.5)
\end{aligned}$$

where $Z_v \sim \mathrm{Exp}(d(v))$ is independent of $\eta(v)$.

For a fixed $v \neq v_0$, apply a union bound splitting the threshold:

$$\begin{aligned}
\mathbb{P}_{v_0}\Big(\tfrac{1}{2}\eta(v)^2 + Z_v \ge \tfrac{\beta^2}{8}R\Big) &\le \mathbb{P}_{v_0}\Big(\tfrac{1}{2}\eta(v)^2 \ge \tfrac{\beta^2}{16}R\Big) + \mathbb{P}_{v_0}\Big(Z_v \ge \tfrac{\beta^2}{16}R\Big)\\
&\le 2e^{-\beta^2/16} + e^{-(\beta^2/16)\,d(v)\,R} \;\le\; 2e^{-\beta^2/16} + e^{-(\beta^2/16)\,d_{\min}\,R}. && (3.6)
\end{aligned}$$

Combining (3.5) and (3.6) and then adding (3.4) yields

$$\mathbb{P}_{v_0}\big(N_{\tau_{\mathrm{inv}}(t)}(v^*) \le 1,\, v^* \neq v_0,\, \mathcal{E}_{\mathrm{RK}}\big) \;\le\; 4e^{-\beta^2/16} + e^{-(\beta^2/16)d_{\min}R}. \tag{3.7}$$

Step 2: the brzanch $v^* = v_0$. By definition of $\tau_{\mathrm{inv}}(t)$ and the degree-normalization, the *unnormalized* amount of time spent at v_0 up to $\tau_{\mathrm{inv}}(t)$ equals $d(v_0)\,t$. While at v_0, the walk departs at rate 1, hence the number of departures from v_0 during this time is $\mathrm{Poisson}(d(v_0)t)$. Therefore

$$N_{\tau_{\mathrm{inv}}(t)}(v_0) \;\stackrel{d}{=}\; 1 + \mathrm{Poisson}\big(d(v_0)t\big),$$

and in particular

$$\begin{aligned}
\mathbb{P}_{v_0}\big(v^* = v_0,\, N_{\tau_{\mathrm{inv}}(t)}(v_0) \le 1\big) &= \mathbb{P}_{v_0}\big(\mathrm{Poisson}(d(v_0)t) = 0\big) = e^{-d(v_0)t}\\
&= e^{-\frac{1}{2}d(v_0)(M+\beta\sqrt{R})^2} \le e^{-(\beta^2/2)\,d_{\min}R}.
\end{aligned} \tag{3.8}$$

Finally, combining (3.2), (3.7), and (3.8) gives

$$\mathbb{P}_{v_0}\big(\min_{v} N_{\tau_{\mathrm{inv}}(t)}(v) \le 1\big) \le e^{-\beta^2/8} + 4e^{-\beta^2/16} + e^{-(\beta^2/16)d_{\min}R} + e^{-(\beta^2/2)d_{\min}R}.$$

Using $d_{\min}R \ge 1$, as stated in Lemma 3.3.2, each term is at most $e^{-\beta^2/16}$, hence the right-hand side is bounded by $7e^{-\beta^2/16}$. □

Proof of Theorem 3.2.1. Fix $\beta \ge 1$ and define

$$t_+ := \tfrac{1}{2}\,(M + \beta\sqrt{R})^2, \qquad t_- := \tfrac{1}{2}\,(M - \beta\sqrt{R})_+^2.$$

By the standard exponential concentration for the cover time (e.g. [110]), there exist absolute constants $c_1, c_2 > 0$ such that

$$\mathbb{P}_{v_0}\big(\tau_{\mathrm{inv}}(t_-) \le \tau_{\mathrm{cov}} \le \tau_{\mathrm{inv}}(t_+)\big) \;\ge\; 1 - c_1 e^{-c_2\beta^2}. \tag{3.9}$$

On the other hand, applying Theorem 3.3.4 at level $t = t_+$ gives

$$\mathbb{P}_{v_0}\big(\min_{v\in V} N_{\tau_{\mathrm{inv}}(t_+)}(v) \ge 2\big) \;\ge\; 1 - 7e^{-\beta^2/16}, \tag{3.10}$$

and on the event in (3.10) we have $\tau^{(2)}_{\mathrm{cov}} \le \tau_{\mathrm{inv}}(t_+)$. Therefore, intersecting (3.9) and (3.10), we obtain

$$\mathbb{P}_{v_0}\big(\Delta_{\mathrm{cov}} \le \tau_{\mathrm{inv}}(t_+) - \tau_{\mathrm{inv}}(t_-)\big) \;\ge\; 1 - c_1 e^{-c_2\beta^2} - 7e^{-\beta^2/16}. \tag{3.11}$$

Apply Lemma 1.1.2 with $\lambda = \beta^2$ at $s = t_+$, and also at $s = t_-$ if $t_- > 0$ (if $t_- = 0$ then $\tau_{\mathrm{inv}}(t_-) = 0$ deterministically). By a union bound, with probability at least $1 - 12e^{-\beta^2/16}$ we have simultaneously

$$\tau_{\mathrm{inv}}(t_+) \le 2t_+|E| + \big(\beta\sqrt{t_+R} + \beta^2 R\big)|E|, \qquad \tau_{\mathrm{inv}}(t_-) \ge 2t_-|E| - \big(\beta\sqrt{t_-R} + \beta^2 R\big)|E|.$$

Subtracting yields, on this event,

$$\tau_{\mathrm{inv}}(t_+) - \tau_{\mathrm{inv}}(t_-) \le 2(t_+ - t_-)|E| + \beta\sqrt{R}\,(\sqrt{t_+} + \sqrt{t_-})|E| + 2\beta^2 R|E|. \tag{3.12}$$

We now bound the two deterministic terms. First, for all $\beta \ge 1$,

$$t_+ - t_- \;\le\; 2M\beta\sqrt{R} + \beta^2 R. \tag{3.13}$$

(Indeed, if $t_- > 0$ then $t_+ - t_- = \frac{1}{2}\big[(M+\beta\sqrt{R})^2 - (M-\beta\sqrt{R})^2\big] = 2M\beta\sqrt{R}$; if $t_- = 0$ then $t_+ - t_- = \frac{1}{2}(M+\beta\sqrt{R})^2 \le 2M\beta\sqrt{R} + \beta^2 R$.) Second,

$$\sqrt{t_+} + \sqrt{t_-} = \frac{(M+\beta\sqrt{R}) + (M-\beta\sqrt{R})_+}{\sqrt{2}} \;\le\; \sqrt{2}\,(M+\beta\sqrt{R}). \tag{3.14}$$

Plugging (3.13) and (3.14) into (3.12), we obtain

$$\tau_{\mathrm{inv}}(t_+) - \tau_{\mathrm{inv}}(t_-) \;\le\; C\,|E|\big(M\beta\sqrt{R} + \beta^2 R\big) \tag{3.15}$$

for an absolute constant $C > 0$.

Combining (3.11) with (3.15) and the inverse-local-time concentration failure probability, we conclude that there exist absolute constants $C, c > 0$ such that for all $\beta \geq 1$,

$$\mathbb{P}_{v_0}\Big(\Delta_{\text{cov}} > C|E|\big(M\beta\sqrt{R} + \beta^2 R\big)\Big) \leq Ce^{-c\beta^2}. \tag{3.16}$$

Let $g(\beta) := C|E|\big(M\beta\sqrt{R} + \beta^2 R\big)$ for $\beta \geq 1$. Then g is increasing and $g'(\beta) = C|E|\big(M\sqrt{R} + 2\beta R\big)$. Using $\mathbb{E}_{v_0}[\Delta_{\text{cov}}] = \int_0^\infty \mathbb{P}_{v_0}(\Delta_{\text{cov}} > u)\, du$ and the change of variables $u = g(\beta)$ on $[g(1), \infty)$,

$$\begin{aligned}
\mathbb{E}_{v_0}[\Delta_{\text{cov}}] &\leq g(1) + \int_1^\infty \mathbb{P}_{v_0}\big(\Delta_{\text{cov}} > g(\beta)\big)\, g'(\beta)\, d\beta \\
&\leq C|E|(M\sqrt{R} + R) + \int_1^\infty \big(Ce^{-c\beta^2}\big)\, C|E|\big(M\sqrt{R} + 2\beta R\big)\, d\beta \\
&= O\big(|E|(M\sqrt{R} + R)\big),
\end{aligned}$$

since $\int_1^\infty e^{-c\beta^2}\, d\beta < \infty$ and $\int_1^\infty \beta e^{-c\beta^2}\, d\beta < \infty$.

Finally, using $R = O(M^2)$ (equivalently, $\sqrt{R} \leq CM$), we have $R \leq CM\sqrt{R}$, hence

$$\mathbb{E}_{v_0}\big[\Delta_{\text{cov}}\big] = O\big(|E|M\sqrt{R}\big),$$

as claimed. □

Proof of Theorem 3.2.3

Let T be a full b-ary tree of height h ($b \geq 2$), rooted at v_0 (depth 0), with leaves at depth h. Each of the b^k vertices at depth $k \in \{0, \ldots, h-1\}$ has exactly b children, so $|E| = \sum_{k=1}^h b^k = \frac{b(b^h-1)}{b-1}$, and in particular $|E| = \Theta(b^h)$ for fixed $b \geq 2$.

Fix a leaf v and let $\pi(v) = (v_0, v_1, \ldots, v_h = v)$ be the unique root–leaf path, which we call the *spine* from v_0 to v. Consider the walk observed only when it is on $\pi(v)$: whenever it steps from some v_i into an off-spine child subtree, it must return to v_i before it can reach any other spine vertex (since v_i is the unique articulation point). Thus, if we record the successive spine locations at the return times to $\pi(v)$, the induced chain on $\{0, 1, \ldots, h\}$ is a *lazy* nearest-neighbor walk whose *non-lazy moves* from i go to $i-1$ or $i+1$ with equal probability 1/2. In particular, all hitting probabilities on the spine coincide with those of the simple random walk on the path $\{0, 1, \ldots, h\}$ (self-loops do not change hitting probabilities).

Lemma 3.3.5 (Two-visit level on a full tree)**.** *Given a full b-ary tree of size n and height h. Let v_0 be the root of the tree. For any $\gamma \geq 1$, if*

$$t \geq (\log b)\, h^2 + \gamma h, \tag{3.17}$$

then

$$\mathbb{P}_{v_0}\Big(\min_{u \in V} N_{\tau_{\text{inv}}(t)}(u) \leq 1\Big) \leq e^{-\gamma}\Big(1 + \log b + \frac{\gamma}{h}\Big).$$

Proof. Fix any leaf $v \in V$. We will prove the following fact first: at time $\tau_{\text{inv}}(t)$, the number of root excursions is distributed as

$$Q \sim \text{Poisson}(b\,t),$$

and the number of those excursions that visit v is

$$H_v \sim \text{Poisson}\Big(\frac{t}{h}\Big).$$

To see this, by definition of $\tau_{\text{inv}}(t)$, the (unnormalized) time spent at the root up to $\tau_{\text{inv}}(t)$ equals $d(v_0)\,t = b\,t$. Since the holding times are i.i.d. Exp(1), departures from the root during its total occupation time forms a Poisson process of rate 1, hence $Q \sim \text{Poisson}(b\,t)$.

Fix a leaf v and use the spine viewpoint above. A root excursion begins by choosing one of the b children of v_0 uniformly; it enters the unique top-level branch containing v with probability $1/b$. Conditioned on entering that branch, the trace chain on the spine $\{0, 1, \dots, h\}$ starts at level 1 and (by the discussion above) has the same hitting probability as simple random walk on the path, hence

$$\mathbb{P}_{v_0}(\text{hit level } h \text{ before } 0 \mid \text{entered branch of } v) = \frac{1}{h}.$$

Therefore a given root excursion hits v with probability $(1/b) \cdot (1/h) = 1/(bh)$. Poisson thinning then yields

$$H_v \;\sim\; \text{Poisson}\Big(\mathbb{E}[Q] \cdot \frac{1}{bh}\Big) \;=\; \text{Poisson}\Big(\frac{bt}{bh}\Big) \;=\; \text{Poisson}\Big(\frac{t}{h}\Big).$$

Condition on the event that the excursion hits v. In the trace chain on $\{0, 1, \dots, h\}$, after the first hit of h the next move is to $h-1$, and the probability to return to h before hitting 0 when started from $h-1$ equals $(h-1)/h$. Thus the number of additional returns to h before absorption at 0 is Geometric$(1/h)$ on $\{0, 1, 2, \dots\}$, giving the following compound–Poisson formula for the total number of visits to the leaf v:

$$S_v \overset{d}{=} \sum_{i=1}^{H_v} (1 + G_i), \qquad G_i \overset{\text{i.i.d.}}{\sim} \text{Geometric}(1/h).$$

From the representation of S_v,

$$\begin{aligned}\mathbb{P}_{v_0}(S_v \le 1) &= \mathbb{P}_{v_0}(S_v = 0) + \mathbb{P}_{v_0}(S_v = 1)\\ &= \mathbb{P}_{v_0}(H_v = 0) + \mathbb{P}_{v_0}(H_v = 1)\mathbb{P}_{v_0}(G_1 = 0) = e^{-t/h} + \Big(\frac{t}{h}e^{-t/h}\Big)\frac{1}{h}.\end{aligned}$$

There are b^h leaves, so by the union bound,

$$\mathbb{P}_{v_0}\Big(\min_{\text{leaf } v} S_v \le 1\Big) \le b^h e^{-t/h}\Big(1 + \frac{t}{h^2}\Big).$$

Under (3.17), we have $b^h e^{-t/h} \le b^h e^{-(\log b)h-\gamma} = e^{-\gamma}$ and $t/h^2 \le \log b + \gamma/h$, yielding the claim.

Finally, if u is any internal vertex and v is any descendant leaf of u, then every visit to v forces a visit to u, hence $N_{\tau_{\mathrm{inv}}(t)}(u) \ge S_v$. Therefore, the same level ensures two visits for all vertices as well. □

On a tree, conditioning on the local time at a vertex w separates the walk into independent excursion processes into the disjoint child subtrees below w. This gives a natural top-down (root-to-leaves) generation of $\{\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)\}$: given $\mathcal{L}^w_{\tau_{\mathrm{inv}}(t)}$, the occupation fields in the subtrees rooted at the children of w are conditionally independent, and their marginal laws are driven by Poisson counts of excursions. This is exactly the mechanism used in Ding–Zeitouni (Lemma 3.6 and its proof) [38].

Lemma 3.3.6 (Variance of inverse local time on a full b-ary tree)**.** *Let T be a full b-ary tree of height h. Then for all $t > 0$,*

$$\mathrm{var}\big(\tau_{\mathrm{inv}}(t)\big) \;\le\; C\, b^{2h}\, t,$$

for a constant $C = C(b)$. Equivalently, since $|E| \asymp b^h$,

$$\tau_{\mathrm{inv}}(t) \;=\; 2t\,|E| \;\pm\; O\big(|E|\sqrt{t}\big) \qquad \textit{in } L^2.$$

Proof. Recall that $\tau_{\mathrm{inv}}(t) = \sum_{v\in V} d(v)\,\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)$. Fix $u, v \in V$ and let $w = u \wedge v$ be their least common ancestor (LCA), with depth $k = \mathrm{depth}(w) \in \{0, 1, \ldots, h\}$.

By the tree Markov property for the excursion decomposition (see the proof of Lemma 3.6 in [38]), conditional on $\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)$, the collections of local times in the different child subtrees of w are independent. Also, for every descendant x of w,

$$\mathbb{E}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(x) \mid \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big) = \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w).$$

Thus, whenever u and v lie in different child subtrees of w,

$$\mathbb{E}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(u) \mid \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big) = \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w) \quad \text{and} \quad \mathbb{E}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) \mid \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big) = \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w).$$

Consequently,

$$\begin{aligned}\mathbb{E}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(u)\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)\big) &= \mathbb{E}\Big(\mathbb{E}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(u)\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v) \mid \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big)\Big) \\ &= \mathbb{E}\big((\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w))^2\big) = t^2 + \mathrm{var}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big),\end{aligned}$$

and hence

$$\mathrm{cov}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(u), \mathcal{L}_{\tau_{\mathrm{inv}}(t)}(v)\big) \;\le\; \mathrm{var}\big(\mathcal{L}_{\tau_{\mathrm{inv}}(t)}(w)\big). \tag{3.18}$$

(When $w = u$ or $w = v$ this inequality is also immediate.)

For a vertex w at depth $k \geq 1$, the one-dimensional spine reduction implies that

$$\mathcal{L}_{\tau_{\text{inv}}(t)}(w) \sim \text{PoiGamma}\Big(\frac{t}{k},\, k\Big),$$

in the notation of [38]: a compound Poisson sum $\sum_{i=1}^{N} Y_i$ with $N \sim \text{Poisson}(t/k)$ and $Y_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(k,1)$. (For $b \geq 2$ this is the same spine computation as in the binary case; only constants depend on b.)

Using $\mathbb{E}[Y_i] = k$ and $\text{var}(Y_i) = k$, the compound-Poisson variance formula gives

$$\text{var}\big(\mathcal{L}_{\tau_{\text{inv}}(t)}(w)\big) = \frac{t}{k}\big(k + k^2\big) \leq 2tk. \tag{3.19}$$

Combining (3.18)–(3.19) and bounding degrees by $d(\cdot) \leq b+1$,

$$\text{var}(\tau_{\text{inv}}(t)) = \sum_{u,v\in V} d(u)d(v)\text{cov}(\mathcal{L}_{\tau_{\text{inv}}(t)}(u), \mathcal{L}_{\tau_{\text{inv}}(t)}(v)) \leq 2t(b+1)^2 \sum_{u,v\in V} k(u\wedge v).$$

Group ordered pairs (u,v) by $k = \text{depth}(u \wedge v)$. There are b^k choices of the LCA at depth k, and for each such LCA the number of ordered descendant pairs is $\asymp (b^{h-k})^2$. Hence

$$\sum_{u,v\in V} k(u\wedge v) \lesssim \sum_{k=0}^{h} k \cdot b^k \cdot (b^{h-k})^2 = b^{2h} \sum_{k=0}^{h} k\, b^{-k} \lesssim b^{2h},$$

since $\sum_{k\geq 0} kb^{-k} < \infty$ for $b \geq 2$. This yields $\text{var}(\tau_{\text{inv}}(t)) \leq Cb^{2h}t$ for some $C = C(b)$. The equivalent statement follows from $\mathbb{E}[\tau_{\text{inv}}(t)] = 2t|E|$. $\square$

Proof of Theorem 3.2.3. Set

$$t_\star := (\log b)\, h^2 + \gamma h, \qquad \gamma := 3\log h.$$

By Lemma 3.3.5,

$$\mathbb{P}_{v_0}\Big(\min_{u\in V} N_{\tau_{\text{inv}}(t_\star)}(u) \geq 2\Big) \geq 1 - e^{-\gamma}\Big(1 + \log b + \frac{\gamma}{h}\Big) \geq 1 - h^{-2},$$

for all h large enough.

On this event, every vertex has been visited at least twice by time $\tau_{\text{inv}}(t_\star)$, so

$$\tau_{\text{cov}}^{(2)} \leq \tau_{\text{inv}}(t_\star). \tag{3.20}$$

By the Chebyshev inequality and Lemma 3.3.6, for any $a \geq 1$,

$$\mathbb{P}_{v_0}\Big(\big|\tau_{\text{inv}}(t_\star) - 2t_\star|E|\big| \geq a\,|E|\sqrt{t_\star}\Big) \leq \frac{C}{a^2}.$$

Taking $a = \log h$ and using $\sqrt{t_\star} = O(h)$ gives, with probability $\geq 1 - O((\log h)^{-2}) = 1 - o(1)$,

$$\tau_{\text{inv}}(t_\star) \leq 2t_\star|E| + O\big(|E|\, h \log h\big) = 2|E|(\log b)\, h^2 + O\big(|E|\, h \log h\big).$$

Together with (3.20), we obtain

$$\tau_{\text{cov}}^{(2)} \leq 2|E|(\log b)\, h^2 + O\big(|E|\, h \log h\big) \qquad \text{with probability } \geq 1 - o(1). \tag{3.21}$$

Now recall that Ding–Zeitouni [38] (proved for $b = 2$ and stated there to extend to any $b \geq 3$, see Lemma B.4.2 in the appendix.) gives that τ_{cov} has the same leading term $2|E|(\log b)h^2$ with a second-order correction of order $-|E|h \log h$. In particular, there exists $C_0 = C_0(b)$ such that, with probability $1 - o(1)$,

$$\tau_{\text{cov}} \;\geq\; 2|E|(\log b)\, h^2 - C_0\, |E|\, h \log h.$$

Combining with (3.21) yields

$$\tau_{\text{cov}}^{(2)} - \tau_{\text{cov}} \leq O\big(|E|\, h \log h\big) \qquad \text{with probability } \geq 1 - o(1).$$

Since $|E| \asymp n$ and $h \asymp \log n$ for a full b-ary tree, this is $O(n \log n \log \log n)$, as claimed. $\square$

3.4 Summary

The general bound obtained in this chapter is not always sharp, as shown by stars and expanders. Nevertheless, it identifies a broad regime, $t_{\text{hit}} = o(t_{\text{cov}})$, in which the marginal time to the second cover is asymptotically smaller than the cover time itself. The sharper estimate for full b-ary trees further suggests that graph-specific arguments can improve substantially over the general inverse-local-time bracketing method.

Chapter 4

On Statistical Estimation of Edge-Reinforced Random Walks

Reinforced random walks (RRWs), including vertex-reinforced random walks (VRRWs) and edge-reinforced random walks (ERRWs), model random walks where the transition probabilities evolve based on prior visitation history [79, 50, 99, 106]. These models have found applications in various areas, such as network representation learning [109], reinforced PageRank [56], and modeling animal behaviors [96], among others. However, statistical estimation of the parameters governing RRWs remains underexplored. This work focuses on estimating the initial edge weights of ERRWs using observed trajectory data. Leveraging the connections between an ERRW and a random walk in a random environment (RWRE) [82, 81], as given by the so-called "magic formula", we propose an estimator based on the generalized method of moments. To analyze the sample complexity of our estimator, we exploit the hyperbolic Gaussian structure embedded in the random environment to bound the fluctuations of the underlying random edge conductances.

4.1 Introduction

Statistical estimation of Markov chain parameters from trajectory data is a well-studied problem. However, many real-world processes, from animal foraging to web browsing, exhibit path-dependent behavior that cannot be captured by a fixed transition matrix. Reinforced random walks provide a natural framework for such phenomena, but the statistical problem of recovering their governing parameters from observations is far less understood than its Markovian counterpart. This chapter takes a step in this direction, focusing on the estimation of initial edge weights in edge-reinforced random walks.

The definitions and structural properties of VRRW and ERRW are reviewed in Section 1.2. One key distinction is that ERRW on a finite connected graph visits every vertex infinitely often almost surely, whereas VRRW can localize. This follows from the random-environment representation: for any $v_0 \in V$ and $A \in \mathbb{R}^E_{++}$, the ERRW with initial weights

A is a mixture of reversible Markov chains (Theorem 1.2.1), each of which is irreducible on G since all conductances are almost surely positive.

ERRW as a random walk in a random environment

Via the random-environment representation (Theorem 1.2.1 and Section 1.2), the normalized local time of an ERRW converges to the stationary distribution of the random conductance environment $W \sim \mu_{v_0,A}$. The explicit density of $\mu_{v_0,A}$, i.e., the magic formula, is recalled here since it is needed for the moment and KL computations in this chapter.

Fix a reference edge $e_0 \in E$ incident to v_0, work in the gauge $w_{e_0} = 1$, and write $a_v := \sum_{e \ni v} a_e$ and $w_v := \sum_{u \sim v} w_{uv}$. Then

$$d\mu_{v_0,A}(w_{-e_0}) = Z_{v_0,A}^{-1} \frac{w_{v_0}^{1/2} \prod_{e \in E} w_e^{a_e - 1}}{\prod_{v \in V} w_v^{(a_v+1)/2}} \sqrt{\sum_{T \in \mathcal{T}} \prod_{e \in T} w_e} \, dw_{-e_0}, \tag{4.1}$$

where $w_{-e_0} := (w_e)_{e \neq e_0}$ and the normalizing constant is

$$Z_{v_0,A} = \frac{\pi^{(|V|-1)/2}}{2^{1-|V|+\sum_{e \in E} a_e}} \cdot \frac{\prod_{e \in E} \Gamma(a_e)}{\prod_{v \in V} \Gamma\big(\frac{1}{2}(a_v + 1 - \mathbf{1}_{v=v_0})\big)}. \tag{4.2}$$

Here π is the mathematical constant, and $\mathcal{T}$ denotes the set of all spanning trees of G.

Since G is finite, connected, and all conductances are almost surely strictly positive, the quenched chain has a finite cover time, and therefore the ERRW cover time $t_{\text{cov}} := \max_{v_0} \mathbb{E}_{\mathbb{P}_{\text{errw}}^{v_0,A}}[\tau_{\text{cov}}]$ is also finite. The finiteness of t_{cov} determines how long a single trajectory must run before the quenched environment can be reliably estimated, and will be a key input in the sample-complexity analysis.

The estimation problem

We are interested in the problem of parameter estimation of ERRW using sample trajectories. Given the graph structure and the sampled trajectories, we want to recover the list of initial edge weights $L_0 = A$.

Formally, given a connected simple graph $G = (V = [n], E)$, $v_0 \in V$ and $A := (a_e)_{e \in E} \in \mathbb{R}_{++}^E$[1], we obtain $K \geq 1$ i.i.d. ERRW sample trajectories of length $T \geq 1$ on graph G with starting point $X_0 = v_0$ and initial local time $L_0 = A$. We denote these sample trajectories as $\big\{X_t^{(k)} \,\big|\, 0 \leq t \leq T, 1 \leq k \leq K\big\}$, and we want to find an estimator $\hat{A}$ such that the error criterion $d(A, \hat{A})$ is small enough with high probability. We assume that all the initial edge weights are bounded above and below by positive constants. We assume that $\underline{a} \leq \min_{e \in E} a_e \leq \max_{e \in E} a_e \leq \overline{a}$ for some positive constants $\underline{a}, \overline{a}$ that do not scale with n.

[1] To be precise, G and v_0 are observed while A is not.

To be precise, we want an estimator $\hat{A}$ that is an entry-wise ϵ-multiplicative approximation to A, with high probability. This motivates us to consider the following ratio-based divergence between two entry-wise positive sequences $A, B \in \mathbb{R}^E_{++}$:

$$d(A,B) := \max_{\{i,j\}\in E} \left(\max\left\{ \frac{A_{ij}}{B_{ij}}, \frac{B_{ij}}{A_{ij}} \right\} - 1 \right).$$

It is straightforward to check that we have: (i) $d(A,B) \geq 0$; (ii) $d(A,B) = 0$ if and only if $A = B$; (iii) if $d(A,B) \leq \epsilon$ for some $\epsilon \in (0,1)$ then A is an entry-wise ϵ-multiplicative approximation of B, i.e., $(1-\epsilon)B_{ij} \leq A_{ij} \leq (1+\epsilon)B_{ij}$ for every $\{i,j\} \in E$.

Informally speaking, when K and T are large enough, we should be able to recover A with arbitrarily small error. But fixing the error level $\epsilon > 0$ and a confidence level $\delta \in (0,1)$, how many samples are sufficient for the recovery of A by an estimator $\hat{A}$ with error $d(A,\hat{A}) \leq \epsilon$ with probability $\geq 1-\delta$ (upper bound)? How many samples are needed to do so (lower bound)?

4.2 Insufficiency of a single trajectory

As stated in section 4.1, fixing $G = (V,E)$, $v_0 \in V$ and $A \in \mathbb{R}^E_{++}$, we observe $K \geq 1$ i.i.d. sample trajectories of length $T \geq 1$ from $\mathbb{P}^{v_0,A}_{\mathrm{errw}}$, and want to estimate A. This statistical question arises in modeling user website browsing behavior via an ERRW. The vertices of G are the websites, and the network links between websites form the edges in G, while the vector of initial local times A encodes the prior preferences of the population: larger weights correspond to more "preferred" links. Given a segment of users, one assumes each user's browsing log is generated by $\mathbb{P}^{v_0,A}_{\mathrm{errw}}$, with v_0 representing a common entry point (e.g., a search engine). Here ERRW serves as a way to capture the users' self-reinforcement behaviors. Given users' weblogs, which are treated as i.i.d. sample trajectories, we then aim to recover A. Alternatively, in a personalized context, where we have a single user, multiple continuous browsing sessions can each be viewed as an independent ERRW trajectory, capturing short-term memory effects within that session, and assuming independence between different sessions.

Several works in the literature illustrate some similar modeling approaches:

- [79, 109] modeled user browsing with reinforced random walks, then ranked the websites using certain averaged visitation statistics over multiple observed trajectories.

- [14] employed a variant of ERRW to model taxis' trajectories, assuming each follows the same underlying distribution.

- [89] examined potential applications of reinforced random walk variants in modeling tumor angiogenesis. To infer the model parameters, one can likely obtain near-i.i.d. sample trajectories under controlled lab conditions.

Thus, for a wide range of potential applications, one can frame the task as estimating the model parameter A given multiple sampled i.i.d. ERRW trajectories.

We now introduce some standard statistical notation. The total variation distance between two probability distributions on a measurable space $(\mathcal{X}, \mathcal{B})$ is defined, as usual, as $d_{\mathrm{TV}}(P_X, Q_X) := \sup_{B\in\mathcal{B}} (P_X(B) - Q_X(B))$. Recall that if P_X and Q_X are probability distribution on the finite set $\mathcal{X}$ we have

$$d_{\mathrm{TV}}(P_X, Q_X) := \frac{1}{2}\sum_{x\in\mathcal{X}} |p_X(x) - q_X(x)|.$$

The Kullback-Leibler (KL) divergence of P_X with respect to Q_X is defined to be ∞ if P_X is not absolutely continuous with respect to Q_X, and is defined to be $D_{\mathrm{KL}}(P_X \| Q_X) := \mathbb{E}_{P_X}\left[\log \frac{dP_X}{dQ_X}(X)\right]$ otherwise. Pinsker's inequality, see e.g. Theorem 4.19 of [17], relates KL divergence and total variation distance by

$$d_{\mathrm{TV}}(P_X, Q_X) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_X \| Q_X)}.$$

We use the notation $X_i^j := \{X_i, \cdots, X_j\}$ for $0 \leq i \leq j \leq T$.

When estimating the transition matrix of an irreducible finite Markov chain, given a single trajectory that is long enough, one can expect to recover the model parameters in the limit of large trajectory length. However, we will argue in this section that a single trajectory, even if infinitely long, is not good enough to recover the parameters (i.e. the initial edge weights) for ERRW.

Consider a scenario where we have a sample trajectory coming from either $\mathbb{P}^{v_0,A}_{\mathrm{errw}}$ or $\mathbb{P}^{v_0,\tilde{A}}_{\mathrm{errw}}$, with the same graph structure $G = (V, E)$ and the same starting point $v_0 \in V$. We want to test the null hypothesis H_0:"the initial weight vector is $L_0 = A$" against the alternative hypothesis H_1:"the initial weight vector is $L_0 = \tilde{A}$". To do so, one designs a testing algorithm $\mathcal{A}$ that takes some trajectory $\underline{X}_0^T$ of the ERRW (assumed to be generated under hypothesis $\Theta \in \{H_0, H_1\}$), and outputs a decision whether $\underline{X}_0^T$ is generated from $\mathbb{P}^{v_0,A}_{\mathrm{errw}}$ or it is generated from $\mathbb{P}^{v_0,\tilde{A}}_{\mathrm{errw}}$. The testing error of this algorithm is defined as

$$\mathcal{E} := \frac{1}{2}\big(\mathbb{P}(\mathcal{A}(\underline{X}_0^T) = H_1 \,|\, \Theta = H_0) + \mathbb{P}(\mathcal{A}(\underline{X}_0^T) = H_0 \,|\, \Theta = H_1)\big).$$

In hypothesis testing language, this is a combination of Type-I error which is defined as $\mathcal{E}_{\mathrm{I}} := \mathbb{P}(\mathcal{A}(\underline{X}_0^T) = H_1 \,|\, \Theta = H_0)$, and Type-II error which is defined as $\mathcal{E}_{\mathrm{II}} := \mathbb{P}(\mathcal{A}(\underline{X}_0^T) = H_0 \,|\, \Theta = H_1)$. This testing error $\mathcal{E}$ is also referred to as the Bayesian error of the testing procedure under the uniform prior. $\mathcal{E}$ being small is equivalent to both Type-I error and Type-II error being small. A good testing procedure should make the testing error $\mathcal{E}$ as small as possible.

Let us denote a trajectory generated from H_0 as X_0^T and another independent trajectory generated from H_1 as $\tilde{X}_0^T$. Recall that for any ERRW, one can naturally couple it with an RWRE. Let us assume that the random environment for each trajectory is W and $\tilde{W}$, respectively. Under such a coupling, we have Markov chains $A \to W \to X_0^T$ and $\tilde{A} \to \tilde{W} \to \tilde{X}_0^T$. We then have

$$D_{\mathrm{KL}}(X_0^T, W \| \tilde{X}_0^T, \tilde{W}) = \int \sum_{x_0^T} p_{X_0^T|W}(x_0^T|w) p_W(w) \log \left(\frac{p_{X_0^T|W}(x_0^T|w)}{p_{\tilde{X}_0^T|\tilde{W}}(x_0^T|w)} \cdot \frac{p_W(w)}{p_{\tilde{W}}(w)} \right) dw.$$

Since $p_{X_0^T|W}(x_0^T|w) = p_{\tilde{X}_0^T|\tilde{W}}(x_0^T|w)$, as they are both the probability of generating trajectory x_0^T from a simple random walk $\mathbb{P}_{\mathrm{srw}}^{v_0,w}$, we conclude that

$$\begin{aligned} D_{\mathrm{KL}}(X_0^T, W \| \tilde{X}_0^T, \tilde{W}) &= \int \sum_{x_0^T} p_{X_0^T|W}(x_0^T|w) p_W(w) \log \left(\frac{p_W(w)}{p_{\tilde{W}}(w)} \right) dw \\ &= \int p_W(w) \log \left(\frac{p_W(w)}{p_{\tilde{W}}(w)} \right) dw \\ &= D_{\mathrm{KL}}(W \| \tilde{W}). \end{aligned} \tag{4.3}$$

In the first equation above we used the total probability rule. On the other hand, we also have $p_{X_0^T|W}(x_0^T|w) p_W(w) = p_{X_0^T}(x_0^T) p_{W|X_0^T}(w|x_0^T)$. Hence,

$$\begin{aligned} D_{\mathrm{KL}}(X_0^T, W \| \tilde{X}_0^T, \tilde{W}) &= \int \sum_{x_0^T} p_{X_0^T}(x_0^T) p_{W|X_0^T}(w|x_0^T) \log \left(\frac{p_{X_0^T}(x_0^T)}{p_{\tilde{X}_0^T}(x_0^T)} \cdot \frac{p_{W|X_0^T}(w|x_0^T)}{p_{\tilde{W}|\tilde{X}_0^T}(w|x_0^T)} \right) dw \\ &= \int \sum_{x_0^T} p_{X_0^T}(x_0^T) p_{W|X_0^T}(w|x_0^T) \log \left(\frac{p_{X_0^T}(x_0^T)}{p_{\tilde{X}_0^T}(x_0^T)} \right) dw \\ &\quad + \int \sum_{x_0^T} p_{X_0^T}(x_0^T) p_{W|X_0^T}(w|x_0^T) \log \left(\frac{p_{W|X_0^T}(w|x_0^T)}{p_{\tilde{W}|\tilde{X}_0^T}(w|x_0^T)} \right) dw \\ &= D_{\mathrm{KL}}(X_0^T \| \tilde{X}_0^T) + \sum_{x_0^T} p_{X_0^T}(x_0^T) D_{\mathrm{KL}}(\{W|X_0^T = x_0^T\} \,\|\, \{\tilde{W}|\tilde{X}_0^T = x_0^T\}). \end{aligned} \tag{4.4}$$

Notice that we used $\int p_{W|X_0^T}(w|x_0^T) dw = 1$ in the last step, and

$$D_{\mathrm{KL}}(\{W|X_0^T = x_0^T\} \,\|\, \{\tilde{W}|\tilde{X}_0^T = x_0^T\})$$

is the conditional KL divergence between the law of two conditioned random variables $\{W|X_0^T = x_0^T\}$ and $\{\tilde{W}|\tilde{X}_0^T = x_0^T\}$. Therefore, by non-negativity of KL divergence, combining Eq. (4.3) and Eq. (4.4), we conclude that

$$D_{\mathrm{KL}}(X_0^T \| \tilde{X}_0^T) \le D_{\mathrm{KL}}(X_0^T, W \| \tilde{X}_0^T, \tilde{W}) = D_{\mathrm{KL}}(W \| \tilde{W}).$$

By standard results in hypothesis testing, we also know that the best testing error $\mathcal{E}^*$ is

$$\mathcal{E}^* = \frac{1}{2}(1 - d_{\mathrm{TV}}(X_0^T, \tilde{X}_0^T)).$$

Using Pinsker's inequality together with Section 4.2, we have

$$\mathcal{E}^* \geq \frac{1}{2}\left(1 - \sqrt{\frac{1}{2}D_{\mathrm{KL}}(X_0^T \| \tilde{X}_0^T)}\right) \geq \frac{1}{2}\left(1 - \sqrt{\frac{1}{2}D_{\mathrm{KL}}(W \| \tilde{W})}\right).$$

For some sufficiently small constant $\epsilon \in (0,1)$, if we can construct A and $\tilde{A}$ such that $d(A, \tilde{A}) \geq \epsilon$, and $D_{\mathrm{KL}}(W\|\tilde{W}) = O(\epsilon^2)$, then any algorithm that tests H_0 against H_1 will incur error that is at least $\frac{1}{2} - O(\epsilon)$. The performance is dismal if ϵ is small enough (close to randomly guessing which hypothesis is correct).

We now plug in the mixing measure, Eq. (4.1), into the KL divergence formula and obtain

$$\begin{aligned}
D_{\mathrm{KL}}(W\|\tilde{W}) &= \int \log\left(\frac{d\mu_{v_0,A}(w)}{dw} \Big/ \frac{d\mu_{v_0,\tilde{A}}(w)}{dw}\right) d\mu_{v_0,A}(w) \\
&= \int \log\left(\frac{Z_{v_0,\tilde{A}}}{Z_{v_0,A}} \cdot \frac{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{a_e - 1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a_v+1)}} \cdot \frac{\prod_{v\in V} w_v^{\frac{1}{2}(\tilde{a}_v+1)}}{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{\tilde{a}_e - 1}}\right) d\mu_{v_0,A}(w) \\
&= \log\frac{Z_{v_0,\tilde{A}}}{Z_{v_0,A}} + \int \log \frac{\prod_{e\in E} w_e^{a_e - \tilde{a}_e}}{\prod_{v\in V} w_v^{\frac{1}{2}(a_v - \tilde{a}_v)}} \, d\mu_{v_0,A}(w). \qquad (4.5)
\end{aligned}$$

To calculate the second term in the equation above, we need to use some tricks. Fixing arbitrary $e \in E$ and taking the derivative of both sides of Eq. (4.2) with respect to a_e, we obtain

$$\begin{aligned}
\frac{\partial}{\partial a_e} Z_{v_0,A} &= \int_{\mathbb{R}_{++}^{|E|-1}} \frac{\partial}{\partial a_e}\left(\frac{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a_v+1)}}\right) \sqrt{\sum_{T\in\mathcal{T}} \prod_{e\in T} w_e} \, dw_{-e_0} \\
&= \int_{\mathbb{R}_{++}^{|E|-1}} \left(\log w_e - \frac{1}{2}\sum_{v\in e} \log w_v\right) \frac{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a_v+1)}} \sqrt{\sum_{T\in\mathcal{T}} \prod_{e\in T} w_e} \, dw_{-e_0} \\
&= Z_{v_0,A} \int \left(\log w_e - \frac{1}{2}\sum_{v\in e} \log w_v\right) d\mu_{v_0,A}(w).
\end{aligned}$$

Here we used the fact that $\frac{\partial}{\partial a_e} w_e^{a_e - 1} = (\log w_e) w_e^{a_e-1}$. Also, for any $v \in V\backslash\{v_0\}$ that is incident with e, we have $\frac{\partial}{\partial a_e} w_v^{-\frac{1}{2}(a_v+1)} = (-\frac{1}{2}\log w_v) w_v^{-\frac{1}{2}(a_v+1)}$; and for v_0, we have $\frac{\partial}{\partial a_e} w_{v_0}^{-\frac{1}{2}a_{v_0}} = (-\frac{1}{2}\log w_{v_0}) w_{v_0}^{-\frac{1}{2}a_{v_0}}$.

Scale the above equation by $(a_e - \tilde{a}_e)$ and then sum it up over all $e \in E$. We then have

$$\begin{aligned}
\sum_{e\in E}(a_e-\tilde{a}_e)\cdot\frac{\partial}{\partial a_e}Z_{v_0,A} &= Z_{v_0,A}\cdot\sum_{e\in E}\left((a_e-\tilde{a}_e)\int\left(\log w_e-\frac{1}{2}\sum_{v\in e}\log w_v\right)d\mu_{v_0,A}(w)\right)\\
&= Z_{v_0,A}\cdot\int\left(\sum_{e\in E}(a_e-\tilde{a}_e)\log\frac{w_e}{\prod_{v\in e}w_v^{\frac{1}{2}}}\right)d\mu_{v_0,A}(w)\\
&= Z_{v_0,A}\cdot\int\log\frac{\prod_{e\in E}w_e^{(a_e-\tilde{a}_e)}}{\prod_{e\in E}\prod_{v\in e}w_v^{\frac{1}{2}(a_e-\tilde{a}_e)}}d\mu_{v_0,A}(w)\\
&= Z_{v_0,A}\cdot\int\log\frac{\prod_{e\in E}w_e^{(a_e-\tilde{a}_e)}}{\prod_{v\in V}w_v^{\frac{1}{2}(a_v-\tilde{a}_v)}}\,d\mu_{v_0,A}(w).
\end{aligned}\tag{4.6}$$

Here, in the last step, we used the fact that

$$\prod_{e\in E}\prod_{v\in e}w_v^{\frac{1}{2}(a_e-\tilde{a}_e)}=\prod_{v\in V}\prod_{e:v\in e}w_v^{\frac{1}{2}(a_e-\tilde{a}_e)}=\prod_{v\in V}w_v^{\sum_{e:v\in e}\frac{1}{2}(a_e-\tilde{a}_e)}=\prod_{v\in V}w_v^{\frac{1}{2}(a_v-\tilde{a}_v)}.$$

Therefore, combining Eq. (4.5) and Eq. (4.6), we conclude that

$$D_{\mathrm{KL}}(W\|\tilde{W})=\log\frac{Z_{v_0,\tilde{A}}}{Z_{v_0,A}}+Z_{v_0,A}^{-1}\sum_{e\in E}(a_e-\tilde{a}_e)\cdot\frac{\partial}{\partial a_e}Z_{v_0,A}.\tag{4.7}$$

Recalling the explicit formula for $Z_{v_0,A}$, Eq. (4.2), we have

$$\begin{aligned}
\frac{\partial}{\partial a_e}Z_{v_0,A} &= \frac{\partial}{\partial a_e}\left(\frac{\pi^{\frac{1}{2}(|V|-1)}}{2^{1-|V|+\sum_{e\in E}a_e}}\cdot\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}\right)\\
=&\frac{\partial}{\partial a_e}\left(\frac{\pi^{\frac{1}{2}(|V|-1)}}{2^{1-|V|+\sum_{e\in E}a_e}}\right)\cdot\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}\\
&\qquad+\frac{\pi^{\frac{1}{2}(|V|-1)}}{2^{1-|V|+\sum_{e\in E}a_e}}\cdot\frac{\partial}{\partial a_e}\left(\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}\right)\\
=&-\log 2\left(\frac{\pi^{\frac{1}{2}(|V|-1)}}{2^{1-|V|+\sum_{e\in E}a_e}}\cdot\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}\right)\\
&\quad+\left(\Psi(a_e)-\sum_{v\in e}\frac{1}{2}\Psi\left(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0})\right)\right)\frac{\pi^{\frac{1}{2}(|V|-1)}}{2^{1-|V|+\sum_{e\in E}a_e}}\cdot\frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}\\
=&\left(-\log 2+\Psi(a_e)-\sum_{v\in e}\frac{1}{2}\Psi\left(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0})\right)\right)Z_{v_0,A}.
\end{aligned}\tag{4.8}$$

Here we used the notation for the digamma function $\Psi(x) := \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$. In the third equation, we used $\frac{d}{dx}\Gamma(x) = \Gamma'(x) = \Psi(x)\Gamma(x)$, and $\frac{d}{dx}\frac{1}{\Gamma(x)} = -\frac{1}{(\Gamma(x))^2}\Gamma'(x) = -\frac{\Psi(x)}{\Gamma(x)}$.

Plugging Eq. (4.8) into Eq. (4.7), and using Eq. (4.2) again, we derive that

$$
\begin{aligned}
D_{\mathrm{KL}}(W\|\tilde{W}) =& \log \frac{Z_{v_0,\tilde{A}}}{Z_{v_0,A}} + \sum_{e\in E}(a_e-\tilde{a}_e)\Big(-\log 2+\Psi(a_e)-\sum_{v\in e}\frac{1}{2}\Psi\Big(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0})\Big)\Big) \\
=& \log\left(\frac{2^{\sum_{e\in E}a_e}}{2^{\sum_{e\in E}\tilde{a}_e}}\cdot\frac{\prod_{e\in E}\Gamma(\tilde{a}_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(\tilde{a}_v+1-\mathbf{1}_{v=v_0}))}\cdot\frac{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}{\prod_{e\in E}\Gamma(a_e)}\right) \\
&+\sum_{e\in E}(a_e-\tilde{a}_e)\Big(-\log 2+\Psi(a_e)-\sum_{v\in e}\frac{1}{2}\Psi\Big(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0})\Big)\Big) \\
=& \sum_{e\in E}(a_e-\tilde{a}_e)\log 2+\sum_{v\in V}\log\frac{\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}{\Gamma(\frac{1}{2}(\tilde{a}_v+1-\mathbf{1}_{v=v_0}))}+\sum_{e\in E}\log\frac{\Gamma(\tilde{a}_e)}{\Gamma(a_e)} \\
&+\sum_{e\in E}(a_e-\tilde{a}_e)\Big(-\log 2+\Psi(a_e)-\sum_{v\in e}\frac{1}{2}\Psi\Big(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0})\Big)\Big). \qquad (4.9)
\end{aligned}
$$

We now construct the following instance. Fix a specific edge $\tilde{e}\in E$ that is not incident with v_0 and consider a connected d-regular graph with vertex set $V=[n]$ together with weights $a_e=1,\forall e\in E$ and $\tilde{a}_e=1+\epsilon\mathbf{1}(e=\tilde{e}),\forall e\in E$. We want to test the hypothesis $H_0: L_0=A$ against $H_1: L_0=\tilde{A}$. We clearly have $d(A,\tilde{A})=\epsilon$. However, by plugging the values into Eq. (4.9), it is not hard to compute that

$$
\begin{aligned}
D_{\mathrm{KL}}(W\|\tilde{W}) =& -\epsilon\log 2+2\log\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d+1}{2}+\frac{\epsilon}{2})}+\log\Gamma(1+\epsilon)+\epsilon\Big(\log 2-\Psi(1)+\Psi\Big(\frac{d+1}{2}\Big)\Big) \\
=& -2\log\Gamma\Big(\frac{d+1}{2}+\frac{\epsilon}{2}\Big)+2\log\Gamma\Big(\frac{d+1}{2}\Big)+\log\Gamma(1+\epsilon)+\gamma\epsilon+\Psi\Big(\frac{d+1}{2}\Big)\epsilon.
\end{aligned}
$$

Here we used the fact that $\Gamma(1)=1$ and $\Psi(1)=-\gamma$, where $\gamma\approx 0.577$ is the Euler constant. Using the Taylor expansion, it's not hard to see that $\log\Gamma(1+\epsilon)=-\gamma\epsilon+\frac{1}{2}\Psi'(1)\epsilon^2+O(\epsilon^3)$, and

$$
\log\Gamma\Big(\frac{d+1}{2}+\frac{\epsilon}{2}\Big)=\log\Gamma\Big(\frac{d+1}{2}\Big)+\Psi\Big(\frac{d+1}{2}\Big)\frac{\epsilon}{2}+\frac{1}{2}\Psi'\Big(\frac{d+1}{2}\Big)\frac{\epsilon^2}{4}+O(\epsilon^3).
$$

Using these Taylor approximations, we can compute that

$$
D_{\mathrm{KL}}(W\|\tilde{W})=\left(2\Psi'(1)-\Psi'\Big(\frac{d+1}{2}\Big)\right)\frac{\epsilon^2}{4}+O(\epsilon^3).
$$

Since Ψ' is positive on $(0,\infty)$ and strictly decreasing, we have $2\Psi'(1)>2\Psi'(1)-\Psi'(\frac{d+1}{2})>\Psi'(1)>0$ whenever $d\geq 2$. This means that $D_{\mathrm{KL}}(W\|\tilde{W})=O(\epsilon^2)$. By the previous discussion and Pinsker's inequality, we know that any testing algorithm will have test error $\geq\mathcal{E}^*=\frac{1}{2}-O(\epsilon)$, which would be close to $\frac{1}{2}$ when ϵ is a small constant. Therefore, for any small enough constant $\epsilon>0$, there exists an instance where $d(A,\tilde{A})\geq\epsilon$, but any testing algorithm will have performance close to random guessing in terms of distinguishing whether H_0 or H_1 is true if only given a single trajectory, even if it is infinitely long.

Remark 4.2.0.1 (Positive-integer-valued edge weights). *A special case would be the scenario when all the initial edge weights are positive integers within some fixed range $\{\underline{a}, \underline{a}+1, \cdots, \overline{a}-1, \overline{a}\}$ for some $\underline{a} < \overline{a}$ both in $\mathbb{N}_+$. In this case, since there are only $(\overline{a} - \underline{a} + 1)^{|E|}$ possible choices of initial edge weights A, one may expect to pin it down given enough samples.*

However, essentially the same analysis as above can be used to conclude that a single trajectory would not be enough in this case either. To see this, note that for any two different initial weights $A, \tilde{A}$ in this case, the optimal testing error is $\mathcal{E}^* = \frac{1}{2}(1 - d_{\mathrm{TV}}(X_0^T, \tilde{X}_0^T)) \geq \frac{1}{2}(1 - d_{\mathrm{TV}}(W, \tilde{W}))$, where we used data-processing inequality for total variation distance in the second step. According to the magic formula Eq. (4.1)*, $d_{\mathrm{TV}}(W, \tilde{W})$ is going to be some positive quantity bounded away from* 1*, because $d\mu_{v_0,A}(w_{-e_0})$ and $d\mu_{v_0,\tilde{A}}(w_{-e_0})$ share the same support and each has a positive continuous density. Therefore, $\mathcal{E}^*$ will be bounded away from* 0 *and a single infinitely long trajectory will not be able to drive the testing error to* 0. ☐

This suggests that we have to use multiple trajectories if we want to recover the parameters A. Based on the coupling with RWRE, we will explore the following natural two-level learning algorithm:

- From the independent trajectories $\{X^{(0)}\}_0^T, \cdots, \{X^{(K)}\}_0^T$, we construct estimators, denoted as $\hat{W}^{(1)}, \cdots, \hat{W}^{(K)}$, for the corresponding random environments $W^{(1)}, \cdots, W^{(K)}$;
- Interpret the estimates $\hat{W}^{(1)}, \cdots, \hat{W}^{(K)}$ as approximately i.i.d. samples from the mixing measure $d\mu_{v_0,A}$, and recover A from these approximate samples.

To make the algorithm work, we need T large enough so we can estimate W's well enough. At the same time, we also need K large enough so we can then estimate A well enough. Informally speaking, the above algorithm follows from the philosophy that $A \to W \to X_0^T$ is a Markov chain, so any information we can infer about A from X_0^T is essentially contained in the information that we know about W from X_0^T. Hence, inferring W given X_0^T first and then inferring A from W gives a natural algorithm framework for our purpose. The algorithm we are going to present has a more intricate design than the outlined scheme, but its core concept still centers on leveraging the random environment inherent in ERRW.

The maximum likelihood estimator

Before discussing our algorithm, we deviate a bit to consider the following maximum likelihood estimator and see why it may be difficult to make it work.[2] Given K trajectories $\{X(k)_0^T : k \in [K]\}$, let's denote the number of undirected edge crossings in the trajectory $k \in [K]$ as $\{M_e(k), e \in E\}$. In other words, we have $M_e(k) := L_T^e(k) - L_0^e$, for any $e \in E$ and

[2]We do not exclude the possibility of using this method in practice. It could still perform well on empirical data despite dismal performance in the worst case, and many local optimization algorithms like gradient descent are very efficient.

$k \in [K]$. For every $k \in [K]$ let $\{N_v(k), v \in V\}$ denote the number of outgoing transitions from vertex v (note that we have $\sum_{v\in V} N_v(k) = T$ for all $k \in [K]$). Then we have

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(X(k)_0^T = x(k)_0^T, \forall k \le K) = \prod_{k=1}^{K} \frac{\prod_{e\in E}\prod_{i=0}^{m_e(k)-1}(a_e+i)}{\prod_{i=0}^{n_{v_0}(k)-1}(a_{v_0}+2i)\prod_{v\in V\setminus\{v_0\}}\prod_{i=0}^{n_v(k)-1}(a_v+2i+1)}.$$

This is rigorously proved as Lemma 3.5 in [61]. A natural estimator can be obtained via finding the optimal $(a_e^*, e \in E)$ that maximizes the log-likelihood. However, finding this optimizer can be computationally hard, as this log-likelihood function is non-concave in the parameters $\{a_e : e \in E\}$ in general.

4.3 Estimation via a generalized moment method

To recover A, let's construct some statistics to help the estimation procedure. Note that, by the magic formula, fixing a special vertex $v_0 \in V$ and a special edge $e_0 \in E$ that is incident with v_0, we know that

$$d\mu_{v_0,A}(w_{-e_0}) = Z^{-1}_{v_0,A} \cdot \frac{w_{v_0}^{\frac{1}{2}}\prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a_v+1)}}\sqrt{\sum_{T\in\mathcal{T}}\prod_{e\in T} w_e}\, dw_{-e_0},$$

with the normalizing factor $Z_{v_0,A}$ given as in Eq. (4.2). Recall that in this formula we have $w_{e_0} = 1$.

Without loss of generality, we can assume that the random walk starts at a vertex v_0 that's of degree $\deg(v_0) \ge 2$.[3] Denote the transition matrix induced by the random weights W by $(P_{ij})_{i,j\in V}$. Thus $P_{ij} = \frac{W_{ij}}{W_i}$ for any $i, j \in V$. Note that P_{ij} is thought of as a random variable.

Define the matrix U by

$$U_{ij} := \begin{cases} P_{ij}P_{ji}, & \text{if } e = \{i,j\} \in E, \\ 0, & \text{if } i \nsim j. \end{cases} \tag{4.10}$$

Note that $U_{ij} = U_{ji}$ when $i \sim j$, so we can write this common value as U_e, where $e = \{i, j\} \in E$. We remark that similar quantities are also studied and used in [81] (e.g., Equation (6)) to prove the magic formula. We claim that we then have the following moment formulas involving $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\sqrt{U_e})$, $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)$, etc.

[3] The case where we start from a degree-one node v_0 with initial weights A is equivalent to the case where we start from the unique neighbor of v_0 with initial weights A' that differ from A only by 1 at the unique edge that v_0 is incident with.

Lemma 4.3.1 (Moment formulas)**.** *Consider an ERRW on a connected graph $G = (V, E)$. Given a vertex $v_0 \in V$ with degree $\deg(v_0) \geq 2$ and initial edge weights $A \in \mathbb{R}_{++}^E$, we sample a random environment W according to the mixing measure in Eq. (4.1). Denote by $P = (P_{ij})_{i,j\in V}$ the random transition matrix induced by the random conductances W. Let $U_e := P_{ij}P_{ji}$ for any $e = \{i, j\} \in E$. Then we have*

$$
\begin{aligned}
\mathbb{E}_{\text{errw}}^{v_0,A}(\sqrt{U_e}) =& \frac{a_e}{2} \cdot \frac{\Gamma(\frac{1}{2}(a_i + 1 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j + 1 - \mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i + 2 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j + 2 - \mathbf{1}_{j=v_0}))}, \forall e = \{i, j\} \in E;\\
\mathbb{E}_{\text{errw}}^{v_0,A}(U_e) =& \frac{a_e(a_e + 1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})}, \qquad \forall e = \{i, j\} \in E;\\
\mathbb{E}_{\text{errw}}^{v_0,A}(U_e^2) =& \frac{a_e(a_e + 1)(a_e + 2)(a_e + 3)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_i + 3 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})(a_j + 3 - \mathbf{1}_{j=v_0})},\\
& \forall e = \{i, j\} \in E;\\
\mathbb{E}_{\text{errw}}^{v_0,A}(U_e U_{e'}) =& \frac{a_e(a_e + 1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})} \cdot \frac{a_{e'}(a_{e'} + 1)}{(a_k + 1 - \mathbf{1}_{k=v_0})(a_l + 1 - \mathbf{1}_{l=v_0})},\\
& \forall e = \{i, j\}, e' = \{k, l\} \in E;\\
\mathbb{E}_{\text{errw}}^{v_0,A}(U_e U_{e'}) =& \frac{a_e(a_e + 1)a_{e'}(a_{e'} + 1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_k + 1 - \mathbf{1}_{k=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})(a_j + 3 - \mathbf{1}_{j=v_0})},\\
& \forall e = \{i, j\}, e' = \{j, k\} \in E.
\end{aligned}
$$

Here $\Gamma(x), \forall x > 0$ is the Gamma function, and in the conditions above we assume that i, j, k, l are mutually distinct vertices.

Proof. Consider some edge $e' = \{i, j\}$, we know that $\sqrt{P_{ij}P_{ji}} = \frac{W_{e'}}{\sqrt{W_i W_j}}$, and

$$
\mathbb{E}_{\text{errw}}^{v_0,A}(\sqrt{U_{e'}}) = \int \frac{w_{e'}}{\sqrt{w_i w_j}} d\mu_{v_0,A}(w) = Z_{v_0,A}^{-1} \int \frac{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{a'_e - 1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a'_v+1)}} \sqrt{\sum_{T\in\mathcal{T}} \prod_{e\in T} w_e}\, dw_{-e_0}. \quad (4.11)
$$

In the above equation, we let $a'_e := a_e + \mathbf{1}(e = e'), \forall e \in E$. From Eq. (4.2), we know that

$$
\int \frac{w_{v_0}^{\frac{1}{2}} \prod_{e\in E} w_e^{a'_e - 1}}{\prod_{v\in V} w_v^{\frac{1}{2}(a'_v+1)}} \sqrt{\sum_{T\in\mathcal{T}} \prod_{e\in T} w_e}\, dw_{-e_0} = Z_{v_0,A'},
$$

hence we conclude that

$$\begin{aligned}
\mathbb{E}^{v_0,A}_{\text{errw}}(\sqrt{U_{e'}}) = \frac{Z_{v_0,A'}}{Z_{v_0,A}} &= \frac{2^{\sum_{e\in E} a_e}}{2^{\sum_{e\in E} a'_e}} \cdot \frac{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))}{\prod_{e\in E}\Gamma(a_e)} \cdot \frac{\prod_{e\in E}\Gamma(a'_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a'_v+1-\mathbf{1}_{v=v_0}))} \\
&= 2^{\sum_{e\in E}(a_e-a'_e)} \cdot \frac{\prod_{v\in V}\Gamma(\frac{1}{2}(a_v+1-\mathbf{1}_{v=v_0}))\prod_{e\in E}\Gamma(a'_e)}{\prod_{v\in V}\Gamma(\frac{1}{2}(a'_v+1-\mathbf{1}_{v=v_0}))\prod_{e\in E}\Gamma(a_e)} \\
&= \frac{1}{2} \cdot \frac{\Gamma(\frac{1}{2}(a_i+1-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+1-\mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i+2-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+2-\mathbf{1}_{j=v_0}))} \cdot \frac{\Gamma(a_{e'}+1)}{\Gamma(a_{e'})} \\
&= \frac{a_{e'}}{2} \cdot \frac{\Gamma(\frac{1}{2}(a_i+1-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+1-\mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i+2-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+2-\mathbf{1}_{j=v_0}))}. \qquad (4.12)
\end{aligned}$$

In the second equality above, we used Eq. (4.2) again. This proves the first equation in Lemma 4.3.1. Similarly, we also have

$$\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) = \int \frac{w_{e'}^2}{w_i w_j} d\mu_{v_0,A}(w) = \frac{Z_{v_0,A''}}{Z_{v_0,A}}.$$

Here $a''_e = a_e + 2\cdot\mathbf{1}(e=e'), \forall e \in E$. We can then use Eq. (4.2) to conclude that

$$\begin{aligned}
\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) &= \frac{1}{4} \cdot \frac{\Gamma(\frac{1}{2}(a_i+1-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+1-\mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i+3-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+3-\mathbf{1}_{j=v_0}))} \cdot \frac{\Gamma(a_{e'}+2)}{\Gamma(a_{e'})} \\
&= \frac{1}{4} \cdot \frac{1}{\frac{1}{2}(a_i+1-\mathbf{1}_{i=v_0}) \cdot \frac{1}{2}(a_j+1-\mathbf{1}_{j=v_0})} \cdot a_{e'}(a_{e'}+1) \\
&= \frac{a_{e'}(a_{e'}+1)}{(a_i+1-\mathbf{1}_{i=v_0})(a_j+1-\mathbf{1}_{j=v_0})}. \qquad (4.13)
\end{aligned}$$

Here we used the fact that $\frac{\Gamma(x+1)}{\Gamma(x)} = x$ for all $x > 0$. This proves the second equation in the lemma.

Now let us fix any two edges $e' = \{i,j\}, e'' = \{k,l\} \in E$ s.t. $|e' \cap e''| = 0$, and consider the random variable $U_{e'}U_{e''}$. Since $|e' \cap e''| = 0$, it's straightforward to compute that

$$\begin{aligned}
\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}U_{e''}) &= \int \frac{w_{e'}^2}{w_i w_j} \cdot \frac{w_{e''}^2}{w_k w_l} d\mu_{v_0,A}(w) = \frac{Z_{v_0,A^{(3)}}}{Z_{v_0,A}} \\
&= \frac{1}{4} \frac{\Gamma(\frac{1}{2}(a_i+1-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+1-\mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i+3-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j+3-\mathbf{1}_{j=v_0}))} \cdot \frac{\Gamma(a_{e'}+2)}{\Gamma(a_{e'})} \\
&\quad \cdot \frac{1}{4} \frac{\Gamma(\frac{1}{2}(a_k+1-\mathbf{1}_{k=v_0}))\Gamma(\frac{1}{2}(a_l+1-\mathbf{1}_{i=v_0}))}{\Gamma(\frac{1}{2}(a_k+3-\mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_l+3-\mathbf{1}_{j=v_0}))} \cdot \frac{\Gamma(a_{e''}+2)}{\Gamma(a_{e''})} \\
&= \frac{a_{e'}(a_{e'}+1)}{(a_i+1-\mathbf{1}_{i=v_0})(a_j+1-\mathbf{1}_{j=v_0})} \cdot \frac{a_{e''}(a_{e''}+1)}{(a_k+1-\mathbf{1}_{k=v_0})(a_l+1-\mathbf{1}_{l=v_0})}. \qquad (4.14)
\end{aligned}$$

The intermediate steps are fairly similar to those in Eq. (4.11) and Eq. (4.12), while we have $a_e^{(3)} = a_e + 2\cdot\mathbf{1}(e = e') + 2\cdot\mathbf{1}(e = e'')$ here. On the other hand, if $e' = \{i, j\}$, $e'' = \{j, k\}$ for $i \neq k$, i.e., $|e \cap e'| = 1$, then

$$\begin{aligned}
\mathbb{E}_{\text{errw}}^{v_0,A}(U_{e'}U_{e''}) &= \int \frac{w_{e'}^2}{w_i w_j}\cdot\frac{w_{e''}^2}{w_j w_k} d\mu_{v_0,A}(w) \\
&= \int \frac{w_{e'}^2 w_{e''}^2}{w_i w_j^2 w_k} d\mu_{v_0,A}(w) = \frac{Z_{v_0,A^{(3)}}}{Z_{v_0,A}} \\
&= \frac{1}{16}\,\frac{\Gamma(\frac{1}{2}(a_i + 1 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_k + 1 - \mathbf{1}_{k=v_0}))}{\Gamma(\frac{1}{2}(a_i + 3 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_k + 3 - \mathbf{1}_{k=v_0}))} \\
&\qquad \cdot \frac{\Gamma(\frac{1}{2}(a_j + 1 - \mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_j + 5 - \mathbf{1}_{j=v_0}))}\cdot\frac{\Gamma(a_{e'} + 2)}{\Gamma(a_{e'})}\cdot\frac{\Gamma(a_{e''} + 2)}{\Gamma(a_{e''})} \\
&= \frac{a_{e'}(a_{e'} + 1)a_{e''}(a_{e''} + 1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_k + 1 - \mathbf{1}_{k=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})(a_j + 3 - \mathbf{1}_{j=v_0})}.
\end{aligned} \tag{4.15}$$

Notice that since e' and e'' have an intersection, so we have $a_j^{(3)} = a_j + 4$ in this case, compared to the previous scenario.

Similarly, we can compute $\mathbb{E}_{\text{errw}}^{v_0,A}(U_{e'}^2)$ as

$$\begin{aligned}
\mathbb{E}_{\text{errw}}^{v_0,A}(U_{e'}^2) &= \int \frac{w_{e'}^4}{w_i^2 w_j^2} d\mu_{v_0,A}(w) \\
&= \frac{1}{16}\,\frac{\Gamma(\frac{1}{2}(a_i + 1 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j + 1 - \mathbf{1}_{j=v_0}))}{\Gamma(\frac{1}{2}(a_i + 5 - \mathbf{1}_{i=v_0}))\Gamma(\frac{1}{2}(a_j + 5 - \mathbf{1}_{j=v_0}))}\cdot\frac{\Gamma(a_{e'} + 4)}{\Gamma(a_{e'})} \\
&= \frac{a_{e'}(a_{e'} + 1)(a_{e'} + 2)(a_{e'} + 3)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_i + 3 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})(a_j + 3 - \mathbf{1}_{j=v_0})}.
\end{aligned} \tag{4.16}$$

The intermediate steps are again similar to those in Eq. (4.11) and Eq. (4.12). □

If we could obtain estimates for each $\mathbb{E}_{\text{errw}}^{v_0,A}(U_e)$, $e \in E$, from multiple trajectories of the reinforced random walk, then we can try to solve the set of equations

$$\begin{aligned}
\mathbb{E}_{\text{errw}}^{v_0,A}(U_e) &= \frac{a_e(a_e + 1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})}, \qquad && \forall e = \{i, j\} \in E, \\
a_v &= \sum_{e: v\in e} a_e, && \forall v \in V.
\end{aligned}$$

with $\mathbb{E}_{\text{errw}}^{v_0,A}(U_e)$ substituted for by our estimate for it. However, it is non-trivial to solve for the initial weights A directly from this set of equations. But from these equations, we know that if we can approximate the value $(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})\mathbb{E}_{\text{errw}}^{v_0,A}(U_e)$, for $e = \{i, j\}$, within multiplicative factor $(1\pm\epsilon)$, then we can approximate $a_e(a_e+1)$ within the same multiplicative factor, and moreover, we should be able to recover a_e up to multiplicative error $(1 \pm \epsilon)$ since

$(1-\epsilon)x(x+1) \leq \hat{x}(\hat{x}+1) \leq (1+\epsilon)x(x+1), x, \hat{x} > 0$ implies $\hat{x} \in (1 \pm \epsilon)x$.[4] To estimate terms like $(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)$, for $e = \{i,j\}$, up to some multiplicative error, it suffices to approximate $\{\mathbb{E}^{v_0,A}_{\text{errw}}(U_e), \forall e \in E\}$ and $\{o_i := a_i + 1 - \mathbf{1}_{i=v_0}, \forall i \in V\}$ up to some multiplicative error ϵ respectively.

Consider two edges $e = \{i,j\}, e' = \{j,k\}$ such that $i \neq k$. Recall that by Eq. (4.13) we have

$$\mathbb{E}^{v_0,A}_{\text{errw}}(U_e) = \frac{a_e(a_e+1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})}, \ \mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) = \frac{a_{e'}(a_{e'}+1)}{(a_j + 1 - \mathbf{1}_{j=v_0})(a_k + 1 - \mathbf{1}_{k=v_0})}.$$

Using these in Eq. (4.15), we obtain

$$\mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'}) = \mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) \cdot \frac{a_j + 1 - \mathbf{1}_{j=v_0}}{a_j + 3 - \mathbf{1}_{j=v_0}} = \mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) \cdot \frac{o_j}{o_j+2},$$

where we recall that we defined $o_j := a_j + 1 - \mathbf{1}_{j=v_0}$. This equation indicates that we can estimate o_j's via estimating the second moment $V_{e,e'} := \mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})$ and the first moments $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)$, $\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'})$. Specifically, $\forall e, e' \in E$, if e and e' intersect at one vertex j, then

$$o_j = \frac{2\mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})}{\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) - \mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})}. \tag{4.17}$$

From the moment formulas in Lemma 4.3.1, one can see that $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) \geq \mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})$, i.e., U_e and $U_{e'}$ are strictly negatively correlated. Therefore, we would like to approximate $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})$ and $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) - \mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'})$ up to multiplicative error ϵ in order to approximate o_j up to multiplicative error 2ϵ. This procedure can be used to estimate o_j for any vertex j with $\deg(j) \geq 2$, since in this case, we can find two distinct edges that are incident at j. [5]

On the other hand, for any vertex j of degree one and the only edge incident with it, $e = \{i,j\}$, we have

$$\mathbb{E}^{v_0,A}_{\text{errw}}(U_e) = \frac{a_e(a_e+1)}{(a_i + 1 - \mathbf{1}_{i=v_0})(a_j + 1 - \mathbf{1}_{j=v_0})} = \frac{a_j(a_j+1)}{o_i(a_j+1)} = \frac{a_j}{o_i}.$$

Here we used $a_j = a_e$ since e is the only edge that incident with j. We also used the fact that $j \neq v_0$ since $\deg(j) = 1$ while $\deg(v_0) \geq 2$ by assumption. Therefore, if we can estimate $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)$ and o_i, we can then estimate $o_j = a_j + 1$ via

$$o_j = a_j + 1 - \mathbf{1}_{j=v_0} = a_j + 1 = o_i \mathbb{E}^{v_0,A}_{\text{errw}}(U_e) + 1. \tag{4.18}$$

[4] Note that if $0 < \hat{x} < (1-\epsilon)x$, then $\hat{x}^2 + \hat{x} < (1-\epsilon)^2 x^2 + (1-\epsilon)x < (1-\epsilon)x^2 + (1-\epsilon)x$, a contradiction. Similarly, if $\hat{x} > (1+\epsilon)x$, then $\hat{x}^2 + \hat{x} > (1+\epsilon)^2 x + (1+\epsilon)x > (1+\epsilon)x^2 + (1+\epsilon)x$, a contradiction. Therefore, we must have $\hat{x} \in (1 \pm \epsilon)x$.

[5] In our algorithm, we used an arbitrary pair of such e, e' to construct an estimator for o_j. We remark that one can incorporate more pairs of $e, e' \in E$ that intersect at j to possibly design a better estimator for o_j.

Since vertex i must have degree $\deg(i) \geq 2$ given that $\deg(j) = 1$ (else the graph has only a single edge and is not of interest), we can indeed estimate o_i via the procedure discussed above.

Fix some small constant $\epsilon \in (0, 1)$. Assume that for any $e \in E$, we can obtain an ϵ-multiplicative estimator for $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)$, denoted as $\hat{U}_e$. Also assume that for any $e, e' \in E$ such that $|e \cap e'| = 1$ we can obtain an ϵ-multiplicative estimator for $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})$ as $\hat{V}_{e,e'}$, and an ϵ-multiplicative estimator for $\Delta_{e,e'} := \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_{e'}) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})$ denoted as $\hat{\Delta}_{e,e'}$. We can then estimate the initial weights $(a_e, e \in E)$ via the following procedure.

Algorithm 1 Estimation via generalized moments

Require: Estimators $\{\hat{U}_e,\, e \in E\}$ and $\{\hat{V}_{e,e'},\, e, e' \in E\}$, $\{\hat{\Delta}_{e,e'},\, e, e' \in E\}$;

1: For every node $i \in V$ with $\deg(i) \geq 2$, choose two edges e, e' that are incident at i. Plug in estimators of $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})$ and $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_{e'}) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})$ into Eq. (4.17) to obtain an estimator for o_i as:

$$\hat{o}_i = \frac{2\hat{V}_{e,e'}}{\hat{\Delta}_{e,e'}}. \tag{4.19}$$

2: For every node $j \in V$ with $\deg(j) = 1$, let the only edge that is incident with j be $e = \{i, j\}$. Plug in estimators of $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)$, and o_i (obtained in step 1, since i should have degree at least 2) into Eq. (4.18) to obtain an estimator for o_j as:

$$\hat{o}_j = \hat{o}_i \hat{U}_e + 1. \tag{4.20}$$

3: For every edge $e = \{i, j\} \in E$, obtain an estimator for a_e by solving the equation

$$x(x + 1) = \hat{o}_i \hat{o}_j \hat{U}_e. \tag{4.21}$$

Take the unique positive solution as the estimator for a_e:

$$\hat{a}_e = -\frac{1}{2} + \sqrt{\frac{1}{4} + \hat{o}_i \hat{o}_j \hat{U}_e}. \tag{4.22}$$

We note that $\hat{\Delta}_{e,e'}$ is essentially an estimator for the covariance $\mathrm{cov}(U_e, U_{e'})$, albeit with a sign flip. Assuming that the estimators $\hat{U}_e$ and $\hat{o}_i$, $\hat{o}_j$ are all positive (as will be evident from Algorithm 2), Eq. (4.21) is expected to yield one positive solution and one negative solution. Here, we will consider only the positive solution. In the next section, we will see how one can estimate the moments that we used to construct the above procedure. Specifically, we will give estimators $\{\hat{U}_e,\, e \in E\}$, $\{\hat{V}_{e,e'},\, e, e' \in E\}$, and $\{\hat{\Delta}_{e,e'},\, e, e' \in E\}$ that entry-wise approximate the corresponding quantities up to some small multiplicative error, and which will be seen to satisfy the positivity requirement.

Estimating the Moments

Let's now consider the problem of estimating the following moments:

i. $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e), \forall e \in E$;

ii. $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'}), \forall e, e' \in E$ s.t. $|e \cap e'| = 1$;

iii. $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)\mathbb{E}^{v_0,A}_{\text{errw}}(U_{e'}) - \mathbb{E}^{v_0,A}_{\text{errw}}(U_e U_{e'}), \forall e, e' \in E$ s.t. $|e \cap e'| = 1$.

Recall that we are given K independent trajectories $\{X^{(k)}_t | 0 \leq t \leq T, 1 \leq k \leq K\}$ from $\mathbb{P}^{v_0,A}_{\text{errw}}$. Via coupling with RWRE, we denote the random conductance corresponding to the k-th trajectory as $W^{(k)}$, and the corresponding transition matrix (also a random variable) as $P^{(k)}$. Let $N_{ij}(k)$ denote the number of directed edge crossings over the edge $e = \{i, j\}$, from i to j, in the k-th trajectory. For any $m \geq 1$, let $H_{ij}(k; m)$ represent the number of such transitions among the first m outgoing transitions from i in the k-th trajectory. Additionally, let $N_i(k)$ denote the total number of outgoing transitions from i in the k-th trajectory. A natural estimator for $P^{(k)}$ is the simple counting-based estimator:

$$\hat{P}^{(k)}_{ij} := \begin{cases} \frac{H_{ij}(k;m)}{m}, & \text{if } N_i(k) \geq m \text{ and } i \sim j; \\ \frac{1}{\deg(i)}, & \text{if } N_i(k) < m \text{ and } i \sim j; \quad \forall i, j \in V. \\ 0, & \text{otherwise.} \end{cases} \tag{4.23}$$

Note that $\hat{P}^{(k)}_{ij}$ depends on the choice of $m \geq 1$, but this dependency has been suppressed from the notation for readability. Similarly, a natural estimator for $U^{(k)}$ is then

$$\hat{U}^{(k)}_{ij} := \begin{cases} \hat{P}^{(k)}_{ij}\hat{P}^{(k)}_{ji}, & \forall i, j \in V \text{ s.t. } i \sim j, \\ 0, & \text{if } i \nsim j. \end{cases} \tag{4.24}$$

Note that $\hat{U}^{(k)}_{ij}$ also depends on the choice of $m \geq 1$. Also, since $\hat{U}^{(k)}_{ij} = \hat{U}^{(k)}_{ji}$ for all $i \sim j$, we can write this common value as $\hat{U}^{(k)}_e$, where $e = \{i, j\} = \{j, i\}$.

For any $e \in E$, we can average over all $\{\hat{U}^{(k)}_e, 1 \leq k \leq K\}$ to obtain an estimator for $\mathbb{E}^{v_0,A}_{\text{errw}}(U_e)$. We summarize the estimation procedure as follows.

Algorithm 2 Estimating the moments

Require: The trajectories $\{X(k)^T_0, 1 \leq k \leq K\}$ and the parameter $m \geq 1$;

1: For each trajectory, compute the local times and the corresponding estimator $\hat{P}^{(k)}$ according to Eq. (4.23), as well as the empirical U-matrix $\hat{U}^{(k)}$ according to Eq. (4.24);

2: Compute the average empirical U-matrix as

$$\hat{U}_{ij} := K^{-1} \sum_{k=1}^{K} \hat{U}^{(k)}_{ij} \quad \forall \{i, j\} \in E; \tag{4.25}$$

Note that since $\hat{U}_{ij} = \hat{U}_{ji}$ for all $i \sim j$, we can write this common value as $\hat{U}_e$, where $e = \{i,j\} = \{j,i\}$.

3: For any $e, e' \in E$ s.t. $|e \cap e'| = 1$, compute

$$\hat{V}_{e,e'} := K^{-1} \sum_{k=1}^{K} \hat{U}_e^{(k)} \hat{U}_{e'}^{(k)}, \tag{4.26}$$

as well as

$$\hat{\Delta}_{e,e'} := \begin{cases} \hat{U}_e \hat{U}_{e'} - \hat{V}_{e,e'}, & \text{if } \hat{U}_e \hat{U}_{e'} > \hat{V}_{e,e'}; \\ 1, & \text{otherwise.} \end{cases} \tag{4.27}$$

Note that each $\hat{U}_e^{(k)}$, $\hat{U}_e$, $\hat{V}_{e,e'}$, and $\hat{\Delta}_{e,e'}$ depends on the choice of $m \geq 1$.

Given the above two procedures, the overall estimation scheme is just chaining them together by feeding the outputs of Algorithm 2 to Algorithm 1.[6] We denote the average U-matrix as $\overline{U} = K^{-1} \sum_{k=1}^{K} U^{(k)}$. (Here the matrix $U^{(k)}$ is defined as in Eq. (4.10).) Informally, for any $e \in E$, we expect $\overline{U}_e$ to converge in probability to $\mathbb{E}_{\mathrm{ERRW}}^{v_0,A}(U_e)$ as $K \to \infty$. Similarly, we anticipate $\hat{U}_e \to \overline{U}_e$ in probability as $T \to \infty$ and $m \to \infty$ with m being sufficiently smaller than T that the probability of having made at least m transitions out of every vertex approaches 1 for any fixed trajectory. Consequently, as $K, T, m \to \infty$, it is natural to expect that $\hat{U}_e$ converges to $\mathbb{E}_{\mathrm{ERRW}}^{v_0,A}(U_e)$ in probability. A non-asymptotic sample complexity analysis is provided in the next section.

4.4 Non-asymptotic analysis

To obtain non-asymptotic bounds on sample complexity, it is essential to deepen our understanding of the random conductance associated with ERRWs. In particular, leveraging a "lifted" magic formula that links the random conductance to the so-called hyperbolic Gaussian densities allows us to derive tighter bounds on the cover time of the graph and hence yields a sample complexity bound for our estimation problem. The analysis of the cover time is needed to understand how big we can choose the parameter m as T grows.

The observation we exploit is that, given a connected simple graph $G = (V, E)$, and initial edge local times $A = (a_e)_{e \in E}$, ERRW starting from node v_0 is equivalent in distribution to the simple random walk with random conductance Q determined by the following process:

1. Generate independent $\beta_e \sim \mathrm{Gamma}(a_e, 1)$ for each edge $e \in E$.

[6] In the analysis later, we use the fact that the probability of $\{\exists\, e, e' \in E,\, |e \cap e'| = 1,\, \hat{U}_e \hat{U}_{e'} \leq \hat{V}_{e,e'}\}$ is eventually smaller than the confidence level δ that we chose (when the parameters K, T, and m are large enough). So letting $\hat{\Delta}_{e,e'} = 1$ here is simply a placeholder for an error event.

2. Given $(\beta_e)_{e\in E}$, fixing $\phi_{v_0} = 0$, generate $\{\phi_i : i \neq v_0\}$ according to the following hyperbolic Gaussian density:

$$p_\beta(\phi) = \frac{1}{(2\pi)^{(n-1)/2}} e^{-\sum_{e=\{i,j\}\in E} \beta_e(\cosh(\phi_i-\phi_j)-1)} e^{-\sum_{i\in V\setminus\{v_0\}} \phi_i} \sqrt{\sum_{T\in\mathcal{T}} \prod_{e=\{i,j\}\in T} \beta_e e^{\phi_i+\phi_j}}. \tag{4.28}$$

3. Perform a simple random walk with initial weights

$$Q_e := \beta_e e^{\phi_i+\phi_j}, \quad \forall e = \{i,j\} \in E.$$

We remark that Q is related to the random environment W (Eq. (4.1)) in the sense that $\{Q_e/Q_{e_0} : e \in E\}$ will be equal in distribution to $\{W_e : e \in E\}$, as is proven in Section 5 of [92]. Since multiplying all conductances by a common positive scalar does not change the transition matrix, Q and W induce the same law for the random transition matrix P.

To distinguish the measure on Q from that on W, we use $\nu_{v_0,A}$ to denote the measure on Q. Note that there is no need to choose any edge e_0 incident on v_0 in the definition of $\nu_{v_0,A}$.

The process above directly reflects the probabilistic essence of Theorem 1 and Theorem 2 in [92]. [7] It offers an alternative perspective on the mixing measure, uncovering the independent Gamma field (which we dub the β-field) and the hyperbolic Gaussian field (which we dub the ϕ-field) embedded within the mixing measure. This perspective, highlighting both the independence and the (hyperbolic) Gaussian nature, provides a more effective approach to bounding the fluctuations of the random conductance.

In this subsection, let Q represent the random environment coupled with the ERRW. We assume that β and ϕ are also naturally coupled with Q and the ERRW, consistent with the process described above.

Before proceeding with the analysis, we introduce the standard notations O, Ω, and Θ to describe growth rates of functions:

1. $f(x) = O(g(x))$: $f(x) \leq c \cdot g(x)$ for constants $c > 0$, $x \geq x_0$ (asymptotic upper bound);
2. $f(x) = \Omega(g(x))$: $f(x) \geq c \cdot g(x)$ for constants $c > 0$, $x \geq x_0$ (asymptotic lower bound);
3. $f(x) = \Theta(g(x))$: $c_1 \cdot g(x) \leq f(x) \leq c_2 \cdot g(x)$ for $c_1, c_2 > 0$, $x \geq x_0$ (tight asymptotic bound up to constant factors).

Hyperbolic Gaussian measure on a tree

We begin by studying the supremum of the absolute value of the ϕ-field on a tree. Studying a tree serves as a simple case that illustrates the core idea needed in the case of general

[7] Theorem 2 in [92] uses a normalization condition that restricts the ϕ_i's to the set $\{\phi : \sum_i \phi_i = 0\}$, while we restrict them to the set $\{\phi : \phi_{v_0} = 0\}$. One can deduce the formula we presented here from the one in [92] by a simple change-of-variable.

graphs. Consider an arbitrary orientation of each edge, where for any edge $e = \{i, j\}$, we select either the directed edge $\vec{e} = (i \to j)$ or $\vec{e} = (j \to i)$. Let $\vec{E}$ denote the set of directed edges under this orientation.

Next, we define the gradient field of the ϕ-field as $\nabla\phi : \vec{E} \to \mathbb{R}$, where for each directed edge $\vec{e} = (i \to j) \in \vec{E}$, the gradient is given by $\nabla\phi(\vec{e}) := \phi_i - \phi_j$. This forms an edge field $\{\nabla\phi(\vec{e}) : \vec{e} \in \vec{E}\}$, which can be shown to be comprised of mutually independent random variables when the underlying graph is a tree.[8] In the following lemma and its proof, we identify $a_{\vec{e}}$ with a_e, where e is the unoriented edge corresponding to $\vec{e} \in E$. Similarly, we define $\beta_{\vec{e}}$ to be the same as its unoriented counterpart β_e. We use $\{y_{\vec{e}} : \vec{e} \in \vec{E}\}$ to denote a realization of $\{\nabla\phi(\vec{e}) : \vec{e} \in \vec{E}\}$.

Lemma 4.4.1 (Independence of the gradient field)**.** *Consider a tree $G = (V = [n], E)$ with a root $v_0 \in V$, and positive edge weights given by $A := (a_e)_{e \in E}$. Orient all the edges towards the root to obtain $\vec{E}$. Let the ϕ-field on this graph be $\{\phi_i : i \in V\}$ and let the gradient field be $\{\nabla\phi(\vec{e}) : \vec{e} \in \vec{E}\}$. Then the distribution of the gradient field is given by the following product density:*

$$p(y) = \prod_{\vec{e} \in \vec{E}} \frac{\Gamma\big(a_{\vec{e}} + \frac{1}{2}\big)}{\sqrt{2\pi}\,\Gamma(a_{\vec{e}})}\, e^{-\frac{1}{2} y_{\vec{e}}} (\cosh(y_{\vec{e}}))^{-(a_{\vec{e}} + \frac{1}{2})}, \quad \forall y \in \mathbb{R}^{\vec{E}}.$$

Proof. The hyperbolic Gaussian density is

$$\begin{aligned}
p_\beta(\phi) &= \frac{1}{(2\pi)^{(n-1)/2}}\, e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}}(\cosh(\phi_i - \phi_j) - 1)} e^{-\sum_{i \in V\setminus\{v_0\}} \phi_i} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}}^{\frac{1}{2}} e^{\frac{1}{2}(\phi_i + \phi_j)} \\
&= \frac{1}{(2\pi)^{(n-1)/2}} \prod_{\vec{e}=(i\to j)\in\vec{E}} e^{-\beta_{\vec{e}}(\cosh(\phi_i - \phi_j) - 1)} \cdot e^{-\phi_i} \cdot \beta_{\vec{e}}^{\frac{1}{2}} e^{\frac{1}{2}(\phi_i + \phi_j)} \\
&= \prod_{\vec{e}=(i\to j)\in\vec{E}} \sqrt{\frac{\beta_{\vec{e}}}{2\pi}}\, e^{-\beta_{\vec{e}}(\cosh(\phi_i - \phi_j) - 1)} e^{\frac{1}{2}(\phi_j - \phi_i)}.
\end{aligned}$$

In the second step above, we used the fact that $e^{-\sum_{i \in V\setminus\{v_0\}} \phi_i} = \prod_{\vec{e}=(i\to j)\in\vec{E}} e^{-\phi_i}$.

Using the change of variables $y_{\vec{e}} = \nabla\phi(\vec{e}) = \phi_i - \phi_j$, we obtain

$$\prod_{\vec{e} \in \vec{E}} dy_{\vec{e}} = \prod_{i \neq v_0} d\phi_i.$$

This leads to the following expression for the probability density in the new variables:

$$p_\beta(y) = \prod_{\vec{e} \in \vec{E}} \sqrt{\frac{\beta_{\vec{e}}}{2\pi}}\, e^{-\beta_{\vec{e}}(\cosh(y_{\vec{e}}) - 1)} e^{-\frac{1}{2} y_{\vec{e}}}.$$

[8] Readers familiar with the Gaussian Free Field (GFF) may recognize similarities to the gradient-GFF on a tree; the gradient components of GFF on a tree are also independent.

Integrating out the independent Gamma field $\{\beta_{\vec{e}} : \vec{e} \in \vec{E}\}$, with $b_{\vec{e}}$ denoting the realization of $\beta_{\vec{e}}$, we have

$$\begin{aligned}
p(y) &= \int_{\mathbb{R}_+^{n-1}} \left(\prod_{\vec{e}\in\vec{E}} \sqrt{\frac{b_{\vec{e}}}{2\pi}}\, e^{-b_{\vec{e}}(\cosh(y_{\vec{e}})-1)} e^{-\frac{1}{2}y_{\vec{e}}} \right) \prod_{\vec{e}\in\vec{E}} \frac{1}{\Gamma(a_{\vec{e}})}\, b_{\vec{e}}^{a_{\vec{e}}-1} e^{-b_{\vec{e}}} db_{\vec{e}} \\
&= \prod_{\vec{e}\in\vec{E}} \left(\int_{\mathbb{R}_+} \frac{1}{\sqrt{2\pi}\Gamma(a_{\vec{e}})}\, e^{-\frac{1}{2}y_{\vec{e}}} e^{-b_{\vec{e}}\cosh(y_{\vec{e}})}\, b_{\vec{e}}^{a_{\vec{e}}-\frac{1}{2}} db_{\vec{e}} \right) \\
&= \prod_{\vec{e}\in\vec{E}} \frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{2\pi}\Gamma(a_{\vec{e}})}\, e^{-\frac{1}{2}y_{\vec{e}}} (\cosh(y_{\vec{e}}))^{-(a_{\vec{e}}+\frac{1}{2})}.
\end{aligned}$$

In the last step above, we used the so-called Gamma integral identity:

$$\int_{\mathbb{R}_+} \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1-1} e^{-\theta_2 x} dx = 1, \quad \forall \theta_1, \theta_2 > 0, \tag{4.29}$$

setting $x = b_{\vec{e}}, \quad \theta_1 = a_{\vec{e}} + \frac{1}{2}, \quad \theta_2 = \cosh(y_{\vec{e}})$. This concludes the proof. □

Using this characterization, it is not hard to get upper bounds on $\sup_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})|$ and $\sup_{i\in V} |\phi_i|$.

Lemma 4.4.2 (Upper bounds on $\sup_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})|$ and $\sup_{i\in V} |\phi_i|$)**.** *In the same setting as Lemma 4.4.1, in particular that the graph is a tree, assume that* $\underline{a} \le \min_{e\in E} a_e \le \max_{e\in E} a_e \le \overline{a}$ *for some positive constants* $\underline{a}, \overline{a}$*. For any* $c > 0$*, we have*

$$\mathbb{P}_{\text{errw}}^{v_0,A}\left(\sup_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})| \ge \frac{2(c+\overline{a}+3)}{\underline{a}} n \right) \le \mathbb{P}_{\text{errw}}^{v_0,A}\left(\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})| \ge \frac{2(c+\overline{a}+3)}{\underline{a}} n \right) \le e^{-cn}.$$

In particular, we also have $\mathbb{P}_{\text{errw}}^{v_0,A}\left(\sup_{i\in V} |\phi_i| \ge \frac{2(c+\overline{a}+3)}{\underline{a}} n \right) \le e^{-cn}$.

Proof. Using Lemma 4.4.1, we know that for any λ such that $\lambda_{\vec{e}} < a_{\vec{e}}$ for all $\vec{e} \in \vec{E}$,

$$\begin{aligned}
\int_{\mathbb{R}_+^{\vec{E}}} e^{\sum_{\vec{e}\in\vec{E}} \lambda_{\vec{e}}|y_{\vec{e}}|} p(y) dy &= \int_{\mathbb{R}^{\vec{E}}} \prod_{\vec{e}\in\vec{E}} \left(\frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{2\pi}\Gamma(a_{\vec{e}})}\, e^{-\frac{1}{2}y_{\vec{e}}+\lambda_{\vec{e}}|y_{\vec{e}}|} (\cosh(y_{\vec{e}}))^{-(a_{\vec{e}}+\frac{1}{2})} \right) dy_{\vec{e}} \\
&= \prod_{\vec{e}\in\vec{E}} \int_{-\infty}^{\infty} \left(\frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{2\pi}\Gamma(a_{\vec{e}})}\, e^{-\frac{1}{2}y_{\vec{e}}+\lambda_{\vec{e}}|y_{\vec{e}}|} (\cosh(y_{\vec{e}}))^{-(a_{\vec{e}}+\frac{1}{2})} \right) dy_{\vec{e}} \\
&= \prod_{\vec{e}\in\vec{E}} \int_{-\infty}^{\infty} \left(\frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{\pi}\Gamma(a_{\vec{e}})} \left(\frac{e^{-y_{\vec{e}}}}{e^{y_{\vec{e}}}+e^{-y_{\vec{e}}}} \right)^{\frac{1}{2}} \left(\frac{2}{e^{y_{\vec{e}}}+e^{-y_{\vec{e}}}} \right)^{a_{\vec{e}}} e^{\lambda_{\vec{e}}|y_{\vec{e}}|} \right) dy_{\vec{e}}.
\end{aligned}$$

Using the inequalities $\frac{e^{-y_{\vec{e}}}}{e^{y_{\vec{e}}}+e^{-y_{\vec{e}}}} < 1$ and $e^{|y_{\vec{e}}|} < e^{y_{\vec{e}}} + e^{-y_{\vec{e}}}$, we obtain

$$\begin{aligned}\int_0^\infty e^{\sum_{\vec{e}\in\vec{E}} \lambda_{\vec{e}}|y_{\vec{e}}|} p(y)dy &\le \prod_{\vec{e}\in\vec{E}} \int_{-\infty}^{\infty} \left(\frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{\pi}\Gamma(a_{\vec{e}})} \left(\frac{2}{e^{|y_{\vec{e}}|}}\right)^{a_{\vec{e}}} e^{\lambda_{\vec{e}}|y_{\vec{e}}|}\right) dy_{\vec{e}} \\ &= \prod_{\vec{e}\in\vec{E}} \int_0^\infty \left(\frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{\pi}\Gamma(a_{\vec{e}})} 2^{a_{\vec{e}}+1} e^{-(a_{\vec{e}}-\lambda_{\vec{e}})y_{\vec{e}}}\right) dy_{\vec{e}} \\ &= \prod_{\vec{e}\in\vec{E}} \frac{\Gamma\big(a_{\vec{e}}+\frac{1}{2}\big)}{\sqrt{\pi}\Gamma(a_{\vec{e}})} 2^{a_{\vec{e}}+1} \frac{1}{a_{\vec{e}}-\lambda_{\vec{e}}}.\end{aligned}$$

Here, we used the symmetry of the integrand (which is an even function) in the second step. Using the inequality $\Gamma(a_{\vec{e}}+\frac{1}{2}) \le 2\sqrt{\pi}\Gamma(a_{\vec{e}}+1)$ for all $a_{\vec{e}} \ge 0$, [9] we conclude that

$$\mathbb{E}^{v_0,A}_{\text{errw}}\big(e^{\sum_{\vec{e}} \lambda_{\vec{e}}|\nabla\phi(\vec{e})|}\big) \le \prod_{\vec{e}\in\vec{E}} 2^{a_{\vec{e}}+2} \frac{a_{\vec{e}}}{a_{\vec{e}}-\lambda_{\vec{e}}}.$$

Thus, for any $0 < \theta < \underline{a}$, we conclude that

$$\mathbb{E}^{v_0,A}_{\text{errw}}(e^{\theta\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})|}) \le 2^{(\overline{a}+2)|E|} \left(\frac{\underline{a}}{\underline{a}-\theta}\right)^{|E|}.$$

Thus we have the following tail bound for any $\alpha > 0$ and any $0 < \theta < \underline{a}$:

$$\mathbb{P}^{v_0,A}_{\text{errw}}\left(\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})| \ge \alpha\right) \le \mathbb{E}^{v_0,A}_{\text{errw}}(e^{\theta\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})|}) \cdot e^{-\theta\alpha} \le 2^{(\overline{a}+2)|E|} \left(\frac{\underline{a}}{\underline{a}-\theta}\right)^{|E|} e^{-\theta\alpha}.$$

Take $\theta = \frac{\underline{a}}{2}$. We get $\mathbb{P}^{v_0,A}_{\text{errw}}(\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})| \ge \alpha) \le 2^{(\overline{a}+3)|E|} e^{-\underline{a}\alpha/2} \le e^{(\overline{a}+3)n-\underline{a}\alpha/2}$, where we have noted that for a tree we have $|E| = n-1$. For any $c > 0$, let $\alpha = \frac{2(c+\overline{a}+3)}{\underline{a}} n$. Then we have $\mathbb{P}^{v_0,A}_{\text{errw}}(\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})| \ge \alpha) \le e^{-cn}$, as claimed.

Since the graph is assumed to be a tree with the edges directed towards the root v_0 and with $\phi_{v_0} = 0$, the bound on $\sum_{\vec{e}\in\vec{E}} |\nabla\phi(\vec{e})|$ immediately implies the same bound on $\sup_{i\in V} |\phi_i|$. □

Remark 4.4.2.1 (Tightness of the bound on $\sup_{i\in V} |\phi_i|$.)**.** *Consider a path that consists of $(n+1)$ vertices, let $A = \mathbf{1}$, i.e., all edges have initial local time* 1*. Label the left-most vertex by v_0 and the right-most vertex by v_n. Denote their corresponding entries in ϕ as ϕ_0 and ϕ_n, respectively. Then we have $\phi_n = \phi_n - \phi_0 = \sum_{\vec{e}\in\vec{E}} y_{\vec{e}}$.*

[9] To see this, note that $\Gamma(a)$ is decreasing for $a \in (0,\mu)$, where $\mu \approx 1.46$ with $\Gamma(\mu) \approx 0.8856$, and is increasing for $a \in (\mu,\infty)$. If $a_e \ge \mu - \frac{1}{2}$, then $\Gamma(a_e+\frac{1}{2}) \le \Gamma(a_e+1)$; if $0 < a_e < \mu - \frac{1}{2}$, then $\Gamma(a_e+\frac{1}{2}) \le \Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(a_e+1) \ge \Gamma(\mu) > \frac{1}{2}$, therefore $\Gamma(a_e+\frac{1}{2}) \le 2\sqrt{\pi}\Gamma(a_e+1)$.

Note that $\{y_{\vec{e}} : \vec{e} \in \vec{E}\}$ are i.i.d.. For each $\vec{e} \in \vec{E}$, we have

$$\mathbb{E}_{\text{errw}}^{v_0,A}(y_{\vec{e}}) = \int_{\mathbb{R}} \frac{1}{2\sqrt{2}}\, y_{\vec{e}} e^{-\frac{1}{2}y_{\vec{e}}} (\cosh(y_{\vec{e}}))^{-\frac{3}{2}} dy_{\vec{e}} \approx -0.98.$$

Also, the second moment is bounded as

$$\mathbb{E}_{\text{errw}}^{v_0,A}(y_{\vec{e}}^2) = \int_{\mathbb{R}} \frac{1}{2\sqrt{2}}\, y_{\vec{e}}^2 e^{-\frac{1}{2}y_{\vec{e}}} (\cosh(y_{\vec{e}}))^{-\frac{3}{2}} dy_{\vec{e}} \approx 3.00.$$

Then it is clear from the central limit theorem perspective that ϕ_n, as the summation of n i.i.d. random variables, each with mean -0.98 and bounded variance, should be of order $-0.98n+O(\sqrt{n})$ with high probability when n is large enough. Therefore $\sup_{i\in V}|\phi_i|$ is indeed $\Theta(n)$ with high probability in this example.

Hyperbolic Gaussian measure for a general graph

When the graph is not a tree, the independent structure of the gradient field that was identified for trees no longer holds, requiring a different approach to establish bounds on $\sup_{i\in V}|\phi_i|$. The following observation plays a crucial role.

Lemma 4.4.3 (Generating function bound)**.** *Consider a connected graph $G = (V, E)$ with positive edge weights $A := (a_e)_{e\in E}$. Fix a vertex $v_0 \in V$ and generate the β-field and ϕ-field accordingly. Choose an arbitrary orientation of the edges, and let $\vec{E}$ represent the set of directed edges, inducing the gradient field $\nabla\phi$. For any $\lambda \in \mathbb{R}^{\vec{E}}$ such that $\lambda_{\vec{e}} < \beta_{\vec{e}}$ for all $\vec{e} \in \vec{E}$, we have*

$$\mathbb{E}_{\text{errw}}^{v_0,A}\left(e^{\sum_{\vec{e}=(i\to j)\in\vec{E}} \lambda_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)} \,\middle|\, \beta \right) \le \max_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}\in\vec{T}} \sqrt{\frac{\beta_{\vec{e}}}{\beta_{\vec{e}}-\lambda_{\vec{e}}}}. \tag{4.30}$$

Here $\vec{T}$ is the notation for a spanning tree with oriented edges, and $\vec{\mathcal{T}}$ is the family of all such oriented spanning trees. Note that $\vec{\mathcal{T}}$ is just $\mathcal{T}$ where the edges are oriented according to the chosen orientation that is fixed once and for all.

The inequality in Eq. (4.30) becomes an equality if the graph is a tree.

Proof. Note that

$$\begin{aligned}
\text{LHS of Eq. (4.30)} &= \int_{\mathbb{R}^{n-1}} (2\pi)^{-\frac{n-1}{2}} e^{\sum_{\vec{e}=(i\to j)\in\vec{E}} \lambda_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)} \\
&\quad \cdot e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)-\sum_{k\in V\setminus\{v_0\}}\phi_k} \\
&\quad \cdot \sqrt{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}} e^{\phi_i+\phi_j}}\, d\phi_{-v_0} \\
&= \int_{\mathbb{R}^{n-1}} (2\pi)^{-\frac{n-1}{2}} e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} (\beta_{\vec{e}}-\lambda_{\vec{e}})(\cosh(\phi_i-\phi_j)-1)-\sum_{k\in V\setminus\{v_0\}}\phi_k} \\
&\quad \cdot \sqrt{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}} e^{\phi_i+\phi_j}}\, d\phi_{-v_0}.
\end{aligned}$$

Now let $\beta'_{\vec{e}} = \beta_{\vec{e}} - \lambda_{\vec{e}}$ for $\vec{e} \in \vec{E}$. Then,

$$\begin{aligned}
\text{LHS of Eq. (4.30)} &= \int_{\mathbb{R}^{n-1}} (2\pi)^{-\frac{n-1}{2}} e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)-\sum_{k\in V\setminus\{v_0\}}\phi_k} \\
&\quad \cdot \sqrt{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}} e^{\phi_i+\phi_j}} \cdot \sqrt{\frac{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}} e^{\phi_i+\phi_j}}{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}} e^{\phi_i+\phi_j}}}\, d\phi_{-v_0} \\
&\le \int_{\mathbb{R}^{n-1}} (2\pi)^{-\frac{n-1}{2}} e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)-\sum_{k\in V\setminus\{v_0\}}\phi_k} \\
&\quad \cdot \sqrt{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}} e^{\phi_i+\phi_j}} \cdot \sqrt{\max_{\vec{T}\in\vec{\mathcal{T}}} \frac{\prod_{\vec{e}=(i\to j)\in\vec{E}} \beta_{\vec{e}} e^{\phi_i+\phi_j}}{\prod_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}} e^{\phi_i+\phi_j}}}\, d\phi_{-v_0} \\
&= \max_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}\in\vec{T}} \sqrt{\frac{\beta_{\vec{e}}}{\beta'_{\vec{e}}}} \Bigg(\int_{\mathbb{R}^{n-1}} (2\pi)^{-\frac{n-1}{2}} e^{-\sum_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)-\sum_{k\in V\setminus\{v_0\}}\phi_k} \\
&\quad \cdot \sqrt{\sum_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}=(i\to j)\in\vec{E}} \beta'_{\vec{e}} e^{\phi_i+\phi_j}}\, d\phi_{-v_0} \Bigg) \\
&= \max_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}\in\vec{T}} \sqrt{\frac{\beta_{\vec{e}}}{\beta'_{\vec{e}}}}.
\end{aligned}$$

The inequality follows from the fact that $\frac{\sum_i a_i}{\sum_i b_i} \le \max_i \frac{a_i}{b_i}$ for positive sequences $\{a_i\}, \{b_i\}$. $\square$

A natural consequence is that for any $i \sim j$, $|\phi_i - \phi_j|$ has a sub-exponential tail, as we now show.

Corollary 4.4.4 (Sub-exponential tail of gradients of the ϕ-field)**.** *In the same setting as Lemma 4.4.3, we have* $\forall \vec{e} = (i \to j) \in \vec{E}$, *and* $\forall s > 0$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(|\phi_i - \phi_j| \geq s) \leq \sqrt{2}\Big(\frac{1}{2} + \frac{1}{4}e^s\Big)^{-a_{\vec{e}}} \leq 2^{2a_{\vec{e}}+1}e^{-a_{\vec{e}}s}.$$

Proof. By Lemma 4.4.3, for any $\lambda \in \mathbb{R}^{\vec{E}}$ such that $\lambda_{\vec{e}} < \beta_{\vec{e}}$ for all $\vec{e} \in \vec{E}$, we have

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left(e^{\sum_{\vec{e}\in\vec{E}} \lambda_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)} \,\middle|\, \beta\right) \leq \max_{\vec{T}\in\vec{\mathcal{T}}} \prod_{\vec{e}\in\vec{T}} \sqrt{\frac{\beta_{\vec{e}}}{\beta_{\vec{e}} - \lambda_{\vec{e}}}}.$$

Fix any $\vec{e} = (i \to j) \in \vec{E}$, and let $\lambda_{\vec{e}'} = 0$ for any $\vec{e}' \neq \vec{e}$. Then,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left(e^{\lambda_{\vec{e}}(\cosh(\phi_i-\phi_j)-1)} \,\middle|\, \beta\right) \leq \sqrt{\frac{\beta_{\vec{e}}}{\beta_{\vec{e}} - \lambda_{\vec{e}}}}, \quad \forall \lambda_{\vec{e}} < \beta_{\vec{e}}.$$

Thus, for any $\alpha > 0$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\cosh(\phi_i - \phi_j) - 1 \geq \alpha \mid \beta) \leq \sqrt{\frac{\beta_{\vec{e}}}{\beta_{\vec{e}} - \lambda_{\vec{e}}}}e^{-\lambda_{\vec{e}}\alpha}, \quad \forall 0 \leq \lambda_{\vec{e}} < \beta_{\vec{e}}.$$

Taking $\lambda_{\vec{e}} = \frac{1}{2}\beta_{\vec{e}}$ gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\cosh(\phi_i - \phi_j) - 1 \geq \alpha \mid \beta) \leq \sqrt{2}e^{-\frac{1}{2}\beta_{\vec{e}}\alpha}.$$

Integrating over $\beta_{\vec{e}} \sim \mathrm{Gamma}(a_{\vec{e}}, 1)$, we have

$$\begin{aligned}
\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\cosh(\phi_i - \phi_j) - 1 \geq \alpha) &\leq \sqrt{2}\int_0^\infty \frac{1}{\Gamma(a_{\vec{e}})} e^{-\frac{1}{2}\beta_{\vec{e}}\alpha}\beta_{\vec{e}}^{a_{\vec{e}}-1}e^{-\beta_{\vec{e}}}d\beta_{\vec{e}} \\
&= \sqrt{2}\frac{1}{\Gamma(a_{\vec{e}})}\int_0^\infty e^{-(\frac{1}{2}\alpha+1)\beta_{\vec{e}}}\beta_{\vec{e}}^{a_{\vec{e}}-1}d\beta_{\vec{e}} \\
&= \sqrt{2}\Big(1 + \frac{1}{2}\alpha\Big)^{-a_{\vec{e}}}.
\end{aligned}$$

In the final step, we used the Gamma integral identity of Eq. (4.29) again, with $\theta_1 = a_{\vec{e}}, \theta_2 = 1 + \frac{1}{2}\alpha$. Since

$$\frac{1}{2}e^{|\phi_i-\phi_j|} \leq \frac{1}{2}(e^{\phi_i-\phi_j} + e^{-(\phi_i-\phi_j)}) = \cosh(\phi_i - \phi_j),$$

it follows that

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\frac{1}{2}e^{|\phi_i-\phi_j|} \geq \alpha + 1\right) \leq \mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\cosh(\phi_i - \phi_j) \geq \alpha + 1\right) \leq \sqrt{2}\Big(1 + \frac{1}{2}\alpha\Big)^{-a_{\vec{e}}}, \quad \forall \alpha > 0.$$

By the change-of-variable $s = \log(2\alpha + 2) \in (\log 2, \infty)$, this is equivalent to

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(|\phi_i - \phi_j| \geq s) \leq \sqrt{2}\Big(\frac{1}{2} + \frac{1}{4}e^s\Big)^{-a_{\vec{e}}}, \quad \forall s > \log 2.$$

Note that the inequality is also true when $0 < s \leq \log 2$ since the right-hand side will be greater than 1. We can further bound the right-hand side as

$$\sqrt{2}\Big(\frac{1}{2} + \frac{1}{4}e^s\Big)^{-a_{\vec{e}}} \leq \sqrt{2}\Big(\frac{1}{4}e^s\Big)^{-a_{\vec{e}}} = 2^{2a_{\vec{e}}+\frac{1}{2}}e^{-a_{\vec{e}}s} \leq 2^{2a_{\vec{e}}+1}e^{-a_{\vec{e}}s}.$$

□

We are now able to estimate the tail probabilities of $\sup_{i\in V} |\phi_i|$ on a general graph.

Theorem 4.4.5 (Tail probabilities of $\sup_{i\in V} |\phi_i|$ on a general graph)**.** *In the same setting as Lemma 4.4.3, let* diam *be the diameter of the graph G. Assume that $\underline{a} \leq \min_{e\in E} a_e \leq \max_{e\in E} a_e \leq \overline{a}$ for some positive constants $\underline{a}, \overline{a}$. For any $\delta \in (0,1)$, we have*

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Bigg(\sup_{i\in V} |\phi_i| \leq \mathrm{diam} \cdot \Big(\frac{2\overline{a}+1}{\underline{a}} + \frac{1}{\underline{a}} \cdot \log\,\frac{n}{\delta}\Big)\Bigg) \geq 1 - \delta.$$

Proof. Consider the graph with unit distance on each edge $e \in E$. Given any root $v_0 \in V$, there exists a shortest-path tree T_{v_0} such that the distance between any vertex $i \in V$ and v_0 in the graph G is identical to their distance in T_{v_0}. Hence, the depth of the tree should be no more than diam. Let $\vec{T}_{v_0}$ be the oriented version of the tree such that every edge is oriented towards the root. For any $s > 0$, the probability that there exists some $\vec{e} = (i \to j) \in \vec{T}_{v_0}$, such that $|\phi_i - \phi_j| \geq s$ is bounded as

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\exists \vec{e} = (i \to j) \in \vec{T}_{v_0}, |\phi_i - \phi_j| \geq s) \leq \sum_{\vec{e}\in\vec{T}} 2^{2a_{\vec{e}}+1}e^{-a_{\vec{e}}s} \leq n2^{2\overline{a}+1}e^{-\underline{a}s} \leq ne^{2\overline{a}+1-\underline{a}s}. \tag{4.31}$$

In the first step above, we used Corollary 4.4.4 and a union bound over the edges of the tree. Now, for any $\delta \in (0,1)$, to ensure that $ne^{2\overline{a}+1-\underline{a}s}$ is bounded by δ, we require

$$s = \frac{2\overline{a}+1}{\underline{a}} + \frac{1}{\underline{a}} \cdot \log\Big(\frac{n}{\delta}\Big).$$

Since with probability $\geq 1-\delta$, we have $\forall \vec{e} = (i \to j) \in \vec{T}_{v_0}$ that $|\phi_i - \phi_j| \leq \frac{2\overline{a}+1}{\underline{a}} + \frac{1}{\underline{a}} \cdot \log \frac{n}{\delta}$, it follows that

$$\sup_{i\in V} |\phi_i| \leq \mathrm{diam} \cdot \Big(\frac{2\overline{a}+1}{\underline{a}} + \frac{1}{\underline{a}} \cdot \log\,\frac{n}{\delta}\Big),$$

with probability at least $1 - \delta$. □

An improvement based on $(\psi, \tilde{\beta})$-field

Given the beta field and the hyperbolic field

$$(\beta, \phi) \in (0, \infty)^E \times \mathbb{R}^V,$$

together with an extra independent $\gamma \sim \text{Gamma}(\frac{1}{2}, 1)$ placed at the root v_0, define the vertex field $\psi = (\psi_i)_{i \in V}$ by

$$\psi_i := \frac{1}{2} \sum_{j:j \sim i} \beta_{ij} e^{\phi_j - \phi_i} \quad (i \neq v_0), \qquad \psi_{v_0} := \frac{1}{2} \sum_{j:j \sim v_0} \beta_{v_0 j} e^{\phi_j - \phi_{v_0}} + \gamma. \tag{4.32}$$

We write B for the $n \times n$ symmetric matrix with (i, j) entry $B_{ij} = \beta_{ij}$ if $\{i, j\} \in E$ and $B_{ij} = 0$ otherwise, and $B_{ii} = 0$. Sabot-Tarres-Zeng studied the conditioned ψ-field and proved that([93, Lemma 1])

$$p_\beta(\psi) := p(\psi \mid \beta) = \left(\frac{2}{\pi}\right)^{n/2} \frac{e^{-\sum_i \psi_i + \sum_e \beta_e}}{\sqrt{|2\psi - B|}} \cdot \mathbf{1}(2\psi - B \succ 0), \tag{4.33}$$

with conditional moment-generating function given by

$$\mathbb{E}\big(e^{\sum_i \lambda_i \psi_i} \mid \beta\big) = \exp\Bigg(\sum_{e=\{i,j\} \in E} \beta_e (1 - \sqrt{1 - \lambda_i}\sqrt{1 - \lambda_j})\Bigg) \prod_{i \in V} \frac{1}{\sqrt{1 - \lambda_i}}, \tag{4.34}$$

for any $\{\lambda_i \mid i \in V\}$ such that $\lambda_i < 1$ for all $i \in V$.

We also define the normalized edge weights $\tilde{\beta} = (\tilde{\beta}_e)_{e \in E}$ by

$$\tilde{\beta}_{ij} := \frac{\beta_{ij}}{\sqrt{\psi_i \psi_j}} \qquad (\{i, j\} \in E). \tag{4.35}$$

Equivalently, $\beta_{ij} = \tilde{\beta}_{ij} \sqrt{\psi_i \psi_j}$ for all edges.

When $\beta_e \sim \text{Gamma}(a_e, 1)$ are distributed as independent Gamma random variables, the induced measure on ψ, after averaging on β, becomes independent Gamma random variables.

Lemma 4.4.6 (Unconditioned ψ-field)**.** *The unconditioned ψ-field is distributed as independent Gamma random variables where $\psi_i \sim \text{Gamma}(\frac{1}{2}(a_i + 1), 1)$.*

Proof. We verify it by calculating the mgf. Note that for any $\{\lambda_i \mid i \in V\}$ such that $\lambda_i < 1$

for all $i \in V$,

$$\begin{aligned}
\mathbb{E}\big(e^{\sum_i \lambda_i \psi_i}\big) &= \int_{\mathbb{R}^E_{++}} \mathbb{E}\big(e^{\sum_i \lambda_i \psi_i} \mid \beta\big) p(\beta)\, d\beta \\
&= \int_{\mathbb{R}^E_{++}} e^{\sum_{e=\{i,j\}\in E} \beta_e (1-\sqrt{1-\lambda_i}\sqrt{1-\lambda_j})} \prod_{i\in V} \frac{1}{\sqrt{1-\lambda_i}} \prod_{e\in E} \frac{1}{\Gamma(a_e)} \beta_e^{a_e-1} e^{-\beta_e}\, d\beta \\
&= \prod_{i\in V} \frac{1}{\sqrt{1-\lambda_i}} \prod_{e\in E} \left(\int_{\mathbb{R}_{++}} e^{\beta_e (1-\sqrt{1-\lambda_i}\sqrt{1-\lambda_j})} \frac{1}{\Gamma(a_e)} \beta_e^{a_e-1} e^{-\beta_e}\, d\beta_e \right) \\
&= \prod_{i\in V} \frac{1}{\sqrt{1-\lambda_i}} \prod_{e\in E} \left(\int_{\mathbb{R}_{++}} \frac{1}{\Gamma(a_e)} \beta_e^{a_e-1} e^{-\beta_e \sqrt{1-\lambda_i}\sqrt{1-\lambda_j}}\, d\beta_e \right) \\
&= \prod_{i\in V} \frac{1}{\sqrt{1-\lambda_i}} \prod_{e\in E} \left(\frac{1}{\sqrt{1-\lambda_i}\sqrt{1-\lambda_j}} \right)^{a_e} \\
&= \prod_{i\in V} (1-\lambda_i)^{-\frac{1}{2}(a_i+1)}.
\end{aligned}$$

By uniqueness of the Laplace transform, we know that ψ should be distributed as independent Gamma$(\frac{1}{2}(a_i+1), 1)$ random variables. $\square$

One can then compute the conditional density of β given ψ.

Lemma 4.4.7 (Conditioned β on ψ)**.** *The conditional density of β on ψ is:*

$$p_\psi(\beta) := p(\beta \mid \psi) = \frac{1}{Z_A} \cdot \frac{\prod_{e\in E} \beta_e^{a_e-1}}{\prod_{i\in V} \psi_i^{\frac{1}{2}(a_i-1)}} \cdot \frac{1}{\sqrt{|2\psi - B|}} \cdot \mathbf{1}(2\psi - B \succ 0), \tag{4.36}$$

where the normalizing constant is $Z_A := \left(\frac{\pi}{2}\right)^{n/2} \frac{\prod_{e\in E}\Gamma(a_e)}{\prod_{i\in V}\Gamma(\frac{1}{2}(a_i+1))}$ and $\beta \in \mathbb{R}^E_{++}$.

Proof. Note that by Bayes rule,

$$p_\psi(\beta) = \frac{p(\beta,\psi)}{p(\psi)} = \frac{p(\beta)p_\beta(\psi)}{p(\psi)}.$$

Since β and ψ are marginally independent Gamma fields, we know that

$$p(\beta) = \prod_{e\in E} \frac{1}{\Gamma(a_e)} \beta_e^{a_e-1} e^{-\beta_e}, \quad p(\psi) = \prod_{i\in V} \frac{1}{\Gamma(\frac{1}{2}(a_i+1))} \psi_i^{\frac{1}{2}(a_i-1)} e^{-\psi_i}.$$

We then continue to compute that

$$\begin{aligned}
p_\psi(\beta) &= \frac{\prod_{e\in E} \frac{1}{\Gamma(a_e)} \beta_e^{a_e-1} e^{-\beta_e}}{\prod_{i\in V} \frac{1}{\Gamma(\frac{1}{2}(a_i+1))} \psi_i^{\frac{1}{2}(a_i-1)} e^{-\psi_i}} \cdot \left(\frac{2}{\pi}\right)^{n/2} \frac{e^{-\sum_i \psi_i + \sum_e \beta_e}}{\sqrt{|2\psi - B|}} \cdot \mathbf{1}(2\psi - B \succ 0) \\
&= \left(\frac{2}{\pi}\right)^{n/2} \frac{\prod_{i\in V}\Gamma(\frac{1}{2}(a_i+1))}{\prod_{e\in E}\Gamma(a_e)} \frac{\prod_{e\in E} \beta_e^{a_e-1}}{\prod_{i\in V} \psi_i^{\frac{1}{2}(a_i-1)}} \cdot \frac{1}{\sqrt{|2\psi - B|}} \cdot \mathbf{1}(2\psi - B \succ 0).
\end{aligned}$$

$\square$

A key structural input is that, the induced pair $(\psi, \tilde{\beta})$ factorizes:

$$(\psi_i)_{i\in V} \perp\!\!\!\perp (\tilde{\beta}_e)_{e\in E}. \tag{4.37}$$

Lemma 4.4.8 (Independence of $\tilde{\beta}$ and ψ)**.** *Condition on a realization $\psi \in \mathbb{R}^V_{++}$, and suppose that β has conditional density $p_\psi(\beta)$ as in Eq. (4.36). Then the "normalized" version of β, defined as $\tilde{\beta}_e := \frac{\beta_e}{\sqrt{\psi_i\psi_j}}$ for all $e = \{i,j\} \in E$, is distributed as $\forall\, \psi \in \mathbb{R}^V_{++}$,*

$$p_\psi(\tilde{\beta}) = p(\tilde{\beta} \mid \psi) = \frac{1}{Z_A} \cdot \frac{\prod_{e\in E} \tilde{\beta}_e^{a_e-1}}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0), \tag{4.38}$$

with Z_A defined in Lemma 4.4.7 and $\tilde{\beta} \in \mathbb{R}^E_{++}$. In other words, $\tilde{\beta} \sim p_{\mathbf{1}}(\beta)$. Note that $\tilde{\beta}$ is independent of ψ.

Proof. Under the transformation, we know that $\beta_e = \tilde{\beta}_e\sqrt{\psi_i\psi_j}$ for all $e \in E$ and $d\beta = (\prod_{e\in E}\sqrt{\psi_i\psi_j})d\tilde{\beta}$. Moreover, whenever $2\psi - B \succ 0$, we also have $2I - \psi^{-1/2}B\psi^{-1/2} = 2I - \tilde{B} \succ 0$ and $|2\psi_i - B| = (\prod_i \psi_i)|2I - \tilde{B}|$. Therefore, we conclude that

$$\begin{aligned}
p(\tilde{\beta} \mid \psi) &= \frac{1}{Z_A}\frac{\prod_{e\in E}(\tilde{\beta}_e\sqrt{\psi_i\psi_j})^{a_e-1}}{\prod_{i\in V}\psi_i^{\frac{1}{2}(a_i-1)}} \cdot \frac{1}{\sqrt{(\prod_i \psi_i)|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0) \cdot \big(\prod_{e\in E}\sqrt{\psi_i\psi_j}\big) \\
&= \frac{1}{Z_A}\frac{\prod_{e\in E}\tilde{\beta}_e^{a_e-1}\prod_{e\in E}(\sqrt{\psi_i\psi_j})^{a_e-1}}{\prod_{i\in V}\psi_i^{\frac{1}{2}(a_i-1)}} \cdot \frac{\prod_{e\in E}\sqrt{\psi_i\psi_j}}{\prod_i \psi_i^{\frac{1}{2}}}\frac{1}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0) \\
&= \frac{1}{Z_A}\frac{\prod_{e\in E}(\sqrt{\psi_i\psi_j})^{a_e}}{\prod_{i\in V}\psi_i^{\frac{1}{2}a_i}} \cdot \frac{\prod_{e\in E}\tilde{\beta}_e^{a_e-1}}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0) \\
&= \frac{1}{Z_A}\frac{\prod_{e\in E}\tilde{\beta}_e^{a_e-1}}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0).
\end{aligned}$$

Note that we used the fact that $\frac{\prod_{e\in E}(\sqrt{\psi_i\psi_j})^{a_e}}{\prod_{i\in V}\psi_i^{\frac{1}{2}a_i}} = 1$ in the last step. □

This induces the following interesting way of generating the correlated (β, ψ) field:

- Generate ψ as independent Gamma random variables, i.e., $\psi_i \sim \text{Gamma}(\frac{1}{2}(a_i + 1), 1)$;
- Generate $\tilde{\beta}$ as independent random variable that has the law Eq. (4.38);
- Let $\beta_e = \tilde{\beta}_e\sqrt{\psi_i\psi_j}$ for all edge $e \in E$.

One can compute the moments of $\{\beta_e \mid e \in E\}$ as follows. Given real values $\Theta := \{\theta_e \mid e \in E\}$ such that $\theta_e > -a_e, \forall e \in E$, we compute

$$\mathbb{E}\Big(\prod_{e\in E} \tilde{\beta}_e^{\theta_e} \mid \psi\Big) = \int_{\mathbb{R}^E_{++}} \frac{1}{Z_A} \prod_{e\in E} \tilde{\beta}_e^{\theta_e + a_e - 1} \cdot \frac{1}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0)\, d\tilde{\beta}$$
$$= \frac{Z_{A+\theta}}{Z_A} \int_{\mathbb{R}^E_{++}} \frac{1}{Z_{A+\theta}} \prod_{e\in E} \tilde{\beta}_e^{\theta_e + a_e - 1} \cdot \frac{1}{\sqrt{|2I - \tilde{B}|}} \cdot \mathbf{1}(2I - \tilde{B} \succ 0)\, d\tilde{\beta}.$$

This is exactly $\frac{Z_{A+\Theta}}{Z_A}$ since the last integral evaluates to 1. Using these nice properties of the $(\psi, \tilde{\beta})$-field, one can eventually improve the previous bounds on $\sup_i |\phi_i|$ as follows.

Lemma 4.4.9 (Pathwise telescoping bound)**.** *Fix a connected simple graph $G = (V, E)$, a root $v_0 \in V$, and let (β, ϕ, ψ) be the fields defined as in the begining of Section 4.4 and* (4.32) *(with the independent $\gamma \sim \mathrm{Gamma}(\frac{1}{2}, 1)$ at v_0). For each $v \in V$, fix a shortest path*

$$P(v): \quad v_0 = v_0,\, v_1,\, \ldots,\, v_k = v,$$

and write $|P(v)| := k$. Then

$$e^{-\phi_v} \le 2^{|P(v)|} \sqrt{\frac{\psi_v}{\psi_{v_0}}} \prod_{e\in P(v)} \tilde{\beta}_e^{-1}. \tag{4.39}$$

Proof. By the definition of ψ in (4.32), for any vertex $i' \neq v_0$ we have

$$\psi_{i'} = \frac{1}{2} \sum_{j: j\sim i'} \beta_{i'j} e^{\phi_j - \phi_{i'}} \ge \frac{1}{2}\, \beta_{i\,i'}\, e^{\phi_i - \phi_{i'}},$$

where i denotes the neighbor of i' along the fixed path $P(v)$ in the direction towards v_0. Rearranging gives

$$e^{-\phi_{i'}} \le \frac{2\psi_{i'}}{\beta_{i\,i'}}\, e^{-\phi_i}. \tag{4.40}$$

Iterating (4.40) along the path $v_k \to v_{k-1} \to \cdots \to v_1$ yields

$$e^{-\phi_v} \le \prod_{r=1}^{k} \frac{2\psi_{v_r}}{\beta_{v_{r-1}v_r}} = \prod_{r=1}^{k} \frac{2}{\tilde{\beta}_{v_{r-1}v_r}} \sqrt{\frac{\psi_{v_r}}{\psi_{v_{r-1}}}} = 2^k \sqrt{\frac{\psi_{v_k}}{\psi_{v_0}}} \prod_{r=1}^{k} \tilde{\beta}^{-1}_{v_{r-1}v_r},$$

where in the middle equality we used $\beta_{ij} = \tilde{\beta}_{ij}\sqrt{\psi_i\psi_j}$, and in the last equality the $\sqrt{\psi}$-factors telescope. Since $k = |P(v)|$ and $v_k = v$, this is exactly (4.39). $\square$

Lemma 4.4.10 (Endpoint Gamma–ratio control)**.** *Assume the setup preceding Lemma 4.4.9. In particular, the unconditioned ψ-field is independent with*

$$\psi_i \sim \mathrm{Gamma}\Big(s_i, 1\Big), \qquad s_i := \frac{a_i + 1}{2}, \quad a_i = \sum_{e \ni i} a_e,$$

and $0 < \underline{a} \le a_e \le \overline{a} < \infty$ for all $e \in E$. Then for any $\alpha \ge 0$,

$$\Pr\left(\max_{v \in V} \sqrt{\frac{\psi_v}{\psi_{v_0}}} > n^{\alpha} \sqrt{2(1+\alpha)\log n + (\overline{a}\, d_{\max} + 1)\log 2}\right) \le 3n^{-\alpha}. \tag{4.41}$$

Proof. Fix $\theta = \frac{1}{2} \in (0,1)$. For each v,

$$\Pr(\psi_v \ge x) \le (1-\theta)^{-s_v} e^{-\theta x} \le 2^{s_{\max}} e^{-x/2}, \quad s_{\max} := \max_{i \in V} s_i.$$

By a union bound over $v \in V$,

$$\Pr\left(\max_{v \in V} \psi_v \ge x\right) \le n\, 2^{s_{\max}} e^{-x/2} \le n\, 2^{\frac{\overline{a}\, d_{\max}+1}{2}} e^{-x/2}. \tag{4.42}$$

Choose

$$x := 2\log n + (\overline{a}\, d_{\max} + 1)\log 2 + 2\alpha \log n,$$

to get for any $\alpha \ge 0$,

$$\Pr\Big(\max_{v} \psi_v \ \ge 2\log n + (\overline{a}\, d_{\max} + 1)\log 2 + 2\alpha \log n\Big) \le \ e^{-\alpha \log n} = n^{-\alpha}. \tag{4.43}$$

By the definition (4.32), at the root v_0 we have the extra independent $\gamma \sim \mathrm{Gamma}(\frac{1}{2}, 1)$ term and hence $\psi_{v_0} \ge \gamma$. Therefore, for all $y > 0$,

$$\Pr(\psi_{v_0} \le y) \le \Pr(\gamma \le y).$$

Since $\gamma \sim \mathrm{Gamma}(\frac{1}{2}, 1)$ has cdf $\Pr(\gamma \le y) = \mathrm{erf}(\sqrt{y})$, we use the standard bound $\mathrm{erf}(t) \le \frac{2}{\sqrt{\pi}}\, t$ to obtain, for all $\alpha \ge 0$,

$$\Pr\left(\psi_{v_0} \le \frac{1}{n^{2\alpha}}\right) \le \Pr\left(\gamma \le \frac{1}{n^{2\alpha}}\right) \le \frac{2}{\sqrt{\pi}}\, n^{-\alpha}. \tag{4.44}$$

On the intersection of the events in (4.43) and (4.44) (which has probability at least $1 - 3n^{-\alpha}$), we have

$$\max_{v} \sqrt{\frac{\psi_v}{\psi_{v_0}}} \ \le \ \sqrt{\frac{2(1+\alpha)\log n + (\overline{a}\, d_{\max} + 1)\log 2}{n^{-2\alpha}}}.$$

This proves the claim. $\square$

Lemma 4.4.11 (Chernoff upper bound for Y_P)**.** *Let $G=(V,E)$ be connected, fix a root v_0, and assume $0<\underline{a}\le a_e\le \overline{a}<\infty$ for all $e\in E$. For a fixed (shortest) path P from v_0 to some $v\in V$, note that $|P|\le \operatorname{diam}$ and define*

$$Y_P := \sum_{e\in P}\big(-\log\tilde\beta_e\big), \qquad \tilde\beta_{ij} := \frac{\beta_{ij}}{\sqrt{\psi_i\psi_j}}\ (\{i,j\}\in E).$$

Then:

(a) *The mean is*

$$\mathbb{E}[Y_P] = -\sum_{e\in P}\Psi(a_e) + \frac{1}{2}\sum_{i\in V}\deg_P(i)\,\Psi\Big(\frac{a_i+1}{2}\Big), \tag{4.45}$$

where Ψ is the digamma function and $\deg_P(i)\in\{0,1,2\}$ counts edges of P incident to i.

(b) *The variance admits the upper bound*

$$\operatorname{var}(Y_P) \le \sum_{e\in P}\Psi_1(a_e) \le |P|\,\Psi_1(\underline{a}) \le \operatorname{diam}\cdot\Psi_1(\underline{a}), \tag{4.46}$$

where Ψ_1 is the trigamma function.

(c) *Let $v_* := \Psi_1(\underline{a}/2)$. For all $t\ge 0$,*

$$\Pr\big(Y_P > \mathbb{E}[Y_P]+t\big) \le \exp\bigg(-\min\Big\{\frac{t^2}{2\,v_*\,|P|},\ \frac{\underline{a}}{4}\,t\Big\}\bigg). \tag{4.47}$$

In particular, for any $\alpha>0$ (with $x:=(1+\alpha)\log n$),

$$\Pr\Big(Y_P > \mathbb{E}[Y_P] + \sqrt{2\,v_*\,|P|\,x} + \frac{4}{\underline{a}}\,x\Big) \le e^{-x} = n^{-(1+\alpha)}. \tag{4.48}$$

Proof. For θ in a neighborhood of 0 (small enough so that all arguments of $\Gamma(\cdot)$ below remain positive),

$$\mathbb{E}\big[e^{\theta Y_P}\big] = \mathbb{E}\Big[\prod_{e\in P}\tilde\beta_e^{-\theta}\Big] = \frac{Z_{A-\theta\mathbf{1}_P}}{Z_A},$$

where $\mathbf{1}_P(e)=\mathbf{1}\{e\in P\}$ and

$$Z_{A-\theta\mathbf{1}_P} = \Big(\tfrac{\pi}{2}\Big)^{n/2}\frac{\prod_{e\in E}\Gamma(a_e-\theta\,\mathbf{1}_P(e))}{\prod_{i\in V}\Gamma\big(\frac{a_i-\theta\,\deg_P(i)+1}{2}\big)}, \qquad a_i := \sum_{e\ni i} a_e.$$

Define the cumulant generating function (cgf)

$$\kappa_P(\theta) := \log\mathbb{E}[e^{\theta Y_P}] = \sum_{e\in P}\big(\log\Gamma(a_e-\theta)-\log\Gamma(a_e)\big) \\ -\sum_{i\in V}\Big[\log\Gamma\Big(\tfrac{a_i-\theta\,\deg_P(i)+1}{2}\Big) - \log\Gamma\Big(\tfrac{a_i+1}{2}\Big)\Big].$$

Differentiate using the digamma function $\Psi = \Gamma'/\Gamma$:

$$\kappa_P'(\theta) = -\sum_{e\in P} \Psi(a_e - \theta) + \frac{1}{2}\sum_{i\in V} \deg_P(i)\,\Psi\Big(\tfrac{a_i - \theta\,\deg_P(i)+1}{2}\Big),$$

hence $\kappa_P'(0) = \mathbb{E}[Y_P]$, which yields (4.45).

Differentiate again using $\Psi_1 = \Psi'$:

$$\kappa_P''(\theta) = \sum_{e\in P} \Psi_1(a_e - \theta) - \frac{1}{4}\sum_{i\in V} \deg_P(i)^2\,\Psi_1\Big(\tfrac{a_i - \theta\,\deg_P(i)+1}{2}\Big).$$

Evaluating at $\theta = 0$ gives

$$\kappa_P''(0) = \mathrm{var}(Y_P) = \sum_{e\in P} \Psi_1(a_e) - \frac{1}{4}\sum_{i\in V} \deg_P(i)^2\,\Psi_1\Big(\tfrac{a_i+1}{2}\Big) \le \sum_{e\in P} \Psi_1(a_e),$$

which implies (4.46).

For the tail, fix $\theta \in [0, \underline{a}/2]$. Since Ψ_1 is decreasing and $a_e \ge \underline{a}$,

$$\kappa_P''(s) \le \sum_{e\in P} \Psi_1(a_e - s) \le |P|\,\Psi_1(\underline{a} - s) \le |P|\,\Psi_1(\underline{a}/2) =: v_*\,|P| \qquad \forall s \in [0,\theta].$$

By Taylor's theorem with integral remainder,

$$\kappa_P(\theta) - \theta\kappa_P'(0) = \int_0^\theta (\theta - s)\kappa_P''(s)\,ds \le \frac{\theta^2}{2}\,v_*\,|P|.$$

Therefore, by Chernoff/Markov,

$$\Pr(Y_P - \mathbb{E}[Y_P] > t) \le \exp\Big(-\theta t + \kappa_P(\theta) - \theta\mathbb{E}[Y_P]\Big) \le \exp\Big(-\theta t + \tfrac{\theta^2}{2}\,v_*\,|P|\Big), \qquad \forall \theta \in [0, \underline{a}/2].$$

Optimizing over θ yields (4.47). For (4.48), set $x := (1+\alpha)\log n$ and $t = \sqrt{2v_*|P|x} + \frac{4}{\underline{a}}x$, then

$$\min\left\{\frac{t^2}{2v_*|P|}, \frac{\underline{a}}{4}t\right\} \ge x,$$

so $\Pr(Y_P > \mathbb{E}[Y_P] + t) \le e^{-x} = n^{-(1+\alpha)}$. □

Theorem 4.4.12 (Uniform upper bound for $e^{-\phi}$)**.** *Assume the setup preceding Lemma 4.4.9 and that $0 < \underline{a} \le a_e \le \overline{a} < \infty$ for all $e \in E$. Let $d_{\max} := \max_{v\in V} \deg(v)$. Then for any $\alpha > 0$, with probability at least $1 - 4n^{-\alpha}$,*

$$\max_{v\in V} e^{-\phi_v} \le 2^{\mathrm{diam}}\, n^\alpha \sqrt{2(1+\alpha)\log n + (\overline{a}d_{\max}+1)\log 2}$$
$$\exp\Big(\mathrm{diam}\cdot\Lambda + \sqrt{2(1+\alpha)\,\Psi_1(\underline{a}/2)\,\mathrm{diam}\cdot\log n} + \frac{4}{\underline{a}}(1+\alpha)\log n\Big), \tag{4.49}$$

where

$$\Lambda := -\Psi(\underline{a}) + \log\Big(\frac{\overline{a}\, d_{\max} + 1}{2}\Big).$$

Note that the upper bound can be simplified as $\max_{v\in V} e^{-\phi_v} \le \mathrm{poly}(n)\, d_{\max}^{O(\mathrm{diam})}$*, assuming* α, a, b *to be constant.*

Proof. Fix, for each $v \in V$, a shortest path $P(v)$ from v_0 to v so that $|P(v)| \le \mathrm{diam}$. By Lemma 4.4.9,

$$e^{-\phi_v} \le 2^{|P(v)|} \sqrt{\frac{\psi_v}{\psi_{v_0}}}\, \exp\big(Y_{P(v)}\big), \qquad Y_{P(v)} := \sum_{e\in P(v)} (-\log \tilde{\beta}_e).$$

Taking the maximum over v and using $|P(v)| \le \mathrm{diam}$ gives

$$\max_v e^{-\phi_v} \le 2^{\mathrm{diam}} \Big(\max_u \sqrt{\tfrac{\psi_u}{\psi_{v_0}}}\Big)\, \exp\Big(\max_v Y_{P(v)}\Big).$$

For the endpoint factor, Lemma 4.4.10 yields

$$\Pr\left(\max_{u\in V} \sqrt{\frac{\psi_u}{\psi_{v_0}}} > n^{\alpha}\sqrt{2(1+\alpha)\log n + \log 2\,(\overline{a}\, d_{\max} + 1)}\right) \le 3n^{-\alpha}.$$

Let $x := (1+\alpha)\log n$. By Lemma 4.4.11*(c)* and a union bound over the at most n root–shortest paths,

$$\Pr\Big(\max_{v\in V} Y_{P(v)} > \max_{v\in V} \mathbb{E}[Y_{P(v)}] + \sqrt{2\Psi_1(\underline{a}/2)\,\mathrm{diam}\cdot x} + \frac{4}{\underline{a}}x\Big) \le n\, e^{-x} = n^{-\alpha}.$$

It remains to bound $\max_v \mathbb{E}[Y_{P(v)}]$ deterministically in terms of diam and $(\underline{a}, \overline{a}, d_{\max})$. From Lemma 4.4.11,

$$\mathbb{E}[Y_P] = -\sum_{e\in P} \Psi(a_e) + \frac{1}{2}\sum_{i\in V} \deg_P(i)\, \Psi\Big(\frac{a_i+1}{2}\Big), \qquad a_i = \sum_{e\ni i} a_e,\ \deg_P(i) \in \{0,1,2\}.$$

Using that Ψ is increasing, $a_e \ge \underline{a}$, $a_i \le \overline{a}\ \deg(i) \le \overline{a}\, d_{\max}$, and $\sum_i \deg_P(i) = 2|P|$, we have $-\sum_{e\in P}\Psi(a_e) \le -|P|\Psi(\underline{a})$, and

$$\frac{1}{2}\sum_i \deg_P(i)\, \Psi\Big(\frac{a_i+1}{2}\Big) \le |P|\Psi\Big(\frac{\overline{a}\, d_{\max}+1}{2}\Big) \le |P| \log\Big(\frac{\overline{a}\, d_{\max}+1}{2}\Big).$$

Hence, for any shortest path P,

$$\mathbb{E}[Y_P] \le |P| \cdot \Big(- \Psi(\underline{a}) + \log\Big(\frac{\overline{a}\, d_{\max}+1}{2}\Big)\Big) \le \mathrm{diam}\cdot\Lambda,$$

which matches the definition of Λ.

Combining the three displays and intersecting the two high-probability events (probability at least $1 - 4n^{-\alpha}$) yields (4.49). $\square$

To upper bound e^{ϕ_i}, we use the following re-rooting identity. In this lemma, $p_\beta^{(r)}$ denotes the hyperbolic Gaussian density with the same edge field β, but instead pinned at the vertex r, i.e., with $\phi_r = 0$, and

$$p_\beta^{(r)}(\phi) = \frac{1}{(2\pi)^{(n-1)/2}}\, e^{-\sum_{e=\{i,j\}\in E}\beta_e(\cosh(\phi_i-\phi_j)-1)} e^{-\sum_{i\in V\setminus\{r\}}\phi_i} \sqrt{\sum_{T\in\mathcal{T}}\prod_{e=\{i,j\}\in T}\beta_e e^{\phi_i+\phi_j}}.$$

Using this notation, $p_\beta^{(v_0)}$ is the same as what was called p_β earlier, as defined in Equation (4.28).

Lemma 4.4.13 (Exponential moment)**.** *Fix $G = (V, E)$, edge parameters $A = (a_e)_{e\in E}$, and a fixed edge field β. For any $v_0, j \in V$ and any real t for which the integrals below are finite, one has*

$$\int_{\mathbb{R}^{V\setminus\{v_0\}}} e^{\,t\phi_j}\, p_\beta^{(v_0)}(\phi)\, d\phi_{V\setminus\{v_0\}} \;=\; \int_{\mathbb{R}^{V\setminus\{j\}}} e^{\,(1-t)\phi_{v_0}}\, p_\beta^{(j)}(\phi)\, d\phi_{V\setminus\{j\}}.$$

Equivalently,

$$\mathbb{E}_{p_\beta^{(v_0)}}\big[e^{\,t\phi_j}\big] = \mathbb{E}_{p_\beta^{(j)}}\big[e^{\,(1-t)\phi_{v_0}}\big].$$

Proof. Write

$$p_\beta^{(v_0)}(\phi) = \mathcal{Z}_\beta\, e^{-\Phi(\phi)}\, e^{-L_{v_0}(\phi)}\, \sqrt{\Xi(\phi)},$$

where

$$\Phi(\phi) := \sum_{\{i,k\}} \beta_{ik}\big(\cosh(\phi_i - \phi_k) - 1\big), \qquad L_{v_0}(\phi) := \sum_{i\neq v_0} \phi_i,$$

and

$$\mathcal{Z}_\beta := (2\pi)^{-(n-1)/2}, \qquad \Xi(\phi) := \sum_{T\in\mathcal{T}}\prod_{\{i,k\}\in T} \beta_{ik}\, e^{\phi_i+\phi_k}.$$

Fix $j \in V$ and set $\tilde{\phi}_i := \phi_i - \phi_j$ for all i. Then $\tilde{\phi}_j = 0$, $\phi_j = -\tilde{\phi}_{v_0}$, and the change of coordinates $\phi_{\neq v_0} \mapsto \tilde{\phi}_{\neq j}$ has unit Jacobian.

We track the factors in $e^{\,t\phi_j} p_\beta^{(v_0)}(\phi)$. First, $\Phi(\phi) = \Phi(\tilde{\phi})$, since Φ depends only on differences. Next,

$$L_{v_0}(\phi) = \sum_{i\neq v_0}(\tilde{\phi}_i + \phi_j) = \Big(\sum_{i\neq j}\tilde{\phi}_i - \tilde{\phi}_{v_0}\Big) + (n-1)\phi_j,$$

and therefore, using $\phi_j = -\tilde{\phi}_{v_0}$,

$$e^{-L_{v_0}(\phi)} = e^{-\sum_{i\neq j}\tilde{\phi}_i}\, e^{n\tilde{\phi}_{v_0}}.$$

Moreover, since every spanning tree has $n-1$ edges,

$$\Xi(\phi) = e^{\,2(n-1)\phi_j}\Xi(\tilde{\phi}) = e^{-2(n-1)\tilde{\phi}_{v_0}}\Xi(\tilde{\phi}),$$

and hence

$$\sqrt{\Xi(\phi)} = e^{-(n-1)\tilde{\phi}_{v_0}} \sqrt{\Xi(\tilde{\phi})}.$$

Finally, $e^{t\phi_j} = e^{-t\tilde{\phi}_{v_0}}$. Multiplying these identities gives

$$e^{t\phi_j}\, p_\beta^{(v_0)}(\phi)\, d\phi_{\neq v_0} = p_\beta^{(j)}(\tilde{\phi})\, e^{(1-t)\tilde{\phi}_{v_0}}\, d\tilde{\phi}_{\neq j}.$$

Integrating over the corresponding pinned spaces gives the desired identity. □

Remark 4.4.13.1 (Special case $t = 1$). *In particular, at $t = 1$, the above lemma states that for any $v_0, j \in V$,*

$$\mathbb{E}_{p_\beta^{(v_0)}}\left[e^{\phi_j}\right] = 1.$$

This equation will be used in the proof of Theorem 4.4.16. We remark that this special case has previously appeared in [39, Appendix B, Eq. (B.3)].

Bounds on P_{ij}, π_*, and the cover time

Given the bound on the tail probabilities of $\sup_{i\in V} |\phi_i|$ proved in Theorem 4.4.5, we can establish bounds on the cover time of the ERRW. We accomplish this from the perspective of RWRE, by deriving high probability bounds on the cover time of the random walk with respect to the random conductance. To this end, we recall several bounds on the cover time of a random walk in terms of the underlying conductances.

Consider a connected graph $G = (V, E)$ with edge conductances $(q_e)_{e\in E}$. Let $\mathcal{C} := \sum_{e\in E} q_e$ denote the total conductance. Denote the effective resistance between $i, j \in V$ as $R(i \leftrightarrow j)$, and let

$$\mathcal{R} := \max_{i,j\in V} R(i \leftrightarrow j)$$

represent the maximum effective resistance of the network. For the simple random walk on this graph, starting at any $i \in V$, define the random hitting time of vertex $j \in V$ as $\tau_{\text{hit}}(j) := \inf\{t \geq 0 : X_t = j\}$, and let the hitting time be defined as $t_{\text{hit}} := \max_{i,j\in V} \mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j))$, where $\mathbb{E}^{i,q}_{\text{srw}}$ denotes expectations conditioned on starting at $i \in V$ with the conductances given by q. Similarly, the commute time between vertices $i, j \in V$ is defined as $t_{\text{comm}}(i, j) := \mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j)) + \mathbb{E}^{j,q}_{\text{srw}}(\tau_{\text{hit}}(i))$. Define the random cover time τ_{cov} as the first time when all states are visited at least once, and define the cover time as $t_{\text{cov}} := \max_{i\in V} \mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{cov}})$.

The following bounds on the cover time hold.

Lemma 4.4.14 (Conductance bound on the cover time). *Given a connected graph $G = (V = [n], E)$ with positive edge conductances $(q_e)_{e\in E}$, we have*

$$\frac{1}{2}\mathcal{C}\mathcal{R} \leq t_{\text{cov}} \leq \mathcal{C}\mathcal{R}\log n.$$

Proof. Recall that by Matthew's bound (Theorem 11.2 in [67]) we have $t_{\text{hit}} \leq t_{\text{cov}} \leq t_{\text{hit}} \log n$. Also, by the commute time identity (Proposition 10.6 in [67]), we know that $t_{\text{comm}}(i,j) = \mathcal{C}R(i \leftrightarrow j)$. Since we have

$$\begin{aligned} t_{\text{hit}} &= \max_{i,j} \mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j)) \\ &\leq \max_{i,j} \left(\mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j)) + \mathbb{E}^{j,q}_{\text{srw}}(\tau_{\text{hit}}(i))\right) \\ &\leq \max_{i,j} \mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j)) + \max_{i,j} \mathbb{E}^{j,q}_{\text{srw}}(\tau_{\text{hit}}(i)) \\ &= 2t_{\text{hit}}, \end{aligned}$$

and

$$\max_{i,j} \left(\mathbb{E}^{i,q}_{\text{srw}}(\tau_{\text{hit}}(j)) + \mathbb{E}^{j,q}_{\text{srw}}(\tau_{\text{hit}}(i))\right) = \max_{i,j} t_{\text{comm}}(i,j) = \max_{i,j} \mathcal{C}R(i \leftrightarrow j) = \mathcal{C}\mathcal{R},$$

we conclude that $t_{\text{hit}} \leq \mathcal{C}\mathcal{R} \leq 2t_{\text{hit}}$, or equivalently, $\frac{1}{2}\mathcal{C}\mathcal{R} \leq t_{\text{hit}} \leq \mathcal{C}\mathcal{R}$. Now using Matthew's bound, we finally deduce that $t_{\text{cov}} \geq t_{\text{hit}} \geq \frac{1}{2}\mathcal{C}\mathcal{R}$ and $t_{\text{cov}} \leq t_{\text{hit}} \log n \leq \mathcal{C}\mathcal{R} \log n$. □

This lemma states that, up to a factor of $\log n$, the cover time is of the same order as $\mathcal{C}\mathcal{R}$. Thus, determining the order of the cover time reduces to analyzing the order of the effective resistance and total conductance. To proceed, we require a comparison lemma.

Lemma 4.4.15 (Comparison of resistive networks)**.** *For two resistive networks on the same graph with respective conductances $(q_e)_{e \in E}$ and $(q'_e)_{e \in E}$, if*

$$\frac{q_e}{q'_e} \in [\underline{q}, \overline{q}] \quad \forall e \in E,$$

then

$$\frac{\mathcal{R}'}{\mathcal{R}} \in [\underline{q}, \overline{q}],$$

where $\mathcal{R}$ and $\mathcal{R}'$ are the respective maximum effective resistances.

Proof. By Dirichlet's principle, see Exercise 1, Chapter 3 of [42], the effective resistance between i and j is

$$R(i \leftrightarrow j) = \left(\min_{v \in \mathbb{R}^V : v_i = 1, v_j = 0} \sum_{e = \{k,l\} \in E} q_e (v_k - v_l)^2 \right)^{-1},$$

and similarly for $R'(i \leftrightarrow j)$ with q'_e. Since for every edge $e \in E$ we have

$$\underline{q} \leq \frac{q_e}{q'_e} \leq \overline{q},$$

we know $\underline{q}\, q'_e(v_k - v_l)^2 \le q_e(v_k - v_l)^2 \le \overline{q}\, q'_e(v_k - v_l)^2$ for all $\{k, l\} \in E$, and so

$$\underline{q} \sum_{e=\{k,l\}\in E} q'_e(v_k - v_l)^2 \le \sum_{e=\{k,l\}\in E} q_e(v_k - v_l)^2 \le \overline{q} \sum_{e=\{k,l\}\in E} q'_e(v_k - v_l)^2.$$

Taking the minimum over $\{v \in \mathbb{R}^V : v_i - v_j = 1\}$, it follows that

$$\underline{q} R'(i \leftrightarrow j)^{-1} \le R(i \leftrightarrow j)^{-1} \le \overline{q} R'(i \leftrightarrow j)^{-1},$$

or equivalently,

$$\underline{q} R(i \leftrightarrow j) \le R'(i \leftrightarrow j) \le \overline{q} R(i \leftrightarrow j).$$

Now taking the maximum over $i, j \in V$ yields the desired result. $\square$

We can now demonstrate how to derive bounds on the cover time. Bounds on the order of the β-field, combined with those on the order of the ϕ-field, provide bounds on the order of the random conductance W. These bounds on W lead to bounds on the effective resistance $\mathcal{R}$, which then yield upper bounds on the cover time.

Theorem 4.4.16 (Upper bound on the cover time)**.** *Let $G = (V = [n], E)$ be a connected, weighted graph with positive edge weights $A = (a_e)_{e\in E}$ and diameter* diam*. Let*

$$d_{\max} := \max_{v\in V} \deg(v).$$

Assume that each a_e satisfies $\underline{a} \le a_e \le \overline{a}$, where $\underline{a}, \overline{a}$ are positive constants. There exists a constant $g_1(\underline{a}, \overline{a}) > 0$, depending only on $\underline{a}, \overline{a}$, such that for any $\delta \in (0, 1)$, the random environment Q coupled with the ERRW satisfies

$$t_{\mathrm{cov}} \le n^3 \log n \left(\frac{d_{\max}}{\delta}\right)^{2g_1(\underline{a},\overline{a})\mathrm{diam}}$$

with probability at least $1 - \delta$.

Proof. If $d_{\max} = 1$, then, since G is connected, G consists of a single edge, and the claim is immediate. We therefore assume $d_{\max} \ge 2$. We repeatedly use

$$n \le 1 + d_{\max} + \cdots + d_{\max}^{\mathrm{diam}} \le 2d_{\max}^{\mathrm{diam}}, \qquad |E| \le \frac{nd_{\max}}{2} \le d_{\max}^{2\mathrm{diam}}.$$

The first bound follows by exploring the graph from any vertex up to distance diam, and the second follows from the handshaking lemma with some simple algebra. Write $R := d_{\max}/\delta$. Since $d_{\max} \ge 2$ and $\delta \in (0, 1)$, we have $R \ge 2$.

We first control the random edge field β. Since $\beta_e \sim \mathrm{Gamma}(a_e, 1)$, for $0 < \epsilon < 1$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\beta_e \le \epsilon) = \int_0^\epsilon \frac{1}{\Gamma(a_e)} x^{a_e-1} e^{-x}\, dx \le \frac{\epsilon^{a_e}}{\Gamma(a_e + 1)} \le 2\epsilon^{\underline{a}},$$

where we used $e^{-x} \leq 1$ and $\Gamma(x) > 1/2$ for $x > 0$. For the upper tail, exponential Markov's inequality with $\lambda = 1/2$ gives, for every $\alpha > 0$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\beta_e \geq \alpha) \leq \mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{\beta_e/2}]e^{-\alpha/2} = 2^{a_e} e^{-\alpha/2} \leq e^{-\alpha/2+\overline{a}}.$$

Choose

$$\epsilon_\beta := \left(\frac{\delta}{8|E|}\right)^{1/\underline{a}}, \qquad \alpha_\beta := 2\left(\overline{a} + \log\frac{4|E|}{\delta}\right).$$

Then the union bound gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\exists e \in E : \beta_e < \epsilon_\beta) \leq 2|E|\epsilon_\beta^{\underline{a}} = \frac{\delta}{4},$$

and

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(\exists e \in E : \beta_e > \alpha_\beta) \leq |E|e^{-\alpha_\beta/2+\overline{a}} = \frac{\delta}{4}.$$

Thus, with probability at least $1 - \delta/2$, all edges satisfy $\epsilon_\beta \leq \beta_e \leq \alpha_\beta$.

Next, by Lemma 4.4.13 with $t = 1$, $\mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{\phi_i}] = 1$ for every $i \in V$. Hence Markov's inequality and the union bound give, for any $x > 0$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\max_{i\in V} e^{\phi_i} > x\right) \leq \frac{n}{x}.$$

Choosing $x_\phi := 4n/\delta$, we obtain

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\max_{i\in V} e^{\phi_i} > x_\phi\right) \leq \frac{\delta}{4}.$$

For the negative side of the ϕ-field, choose $\alpha_0 := \log(16/\delta)/\log n$, so that $4n^{-\alpha_0} = \delta/4$. Applying Theorem 4.4.12 with this choice gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\max_{i\in V} e^{-\phi_i} > y_\phi\right) \leq \frac{\delta}{4},$$

where

$$\begin{aligned} y_\phi := {} & 2^{\mathrm{diam}} n^{\alpha_0}\sqrt{2(1+\alpha_0)\log n + (\overline{a}d_{\max}+1)\log 2} \\ & \times \exp\left(\mathrm{diam}\Lambda + \sqrt{2(1+\alpha_0)\Psi_1(\underline{a}/2)\mathrm{diam}\log n} + \frac{4}{\underline{a}}(1+\alpha_0)\log n\right), \end{aligned}$$

and $\Lambda := -\Psi(\underline{a}) + \log((\overline{a}d_{\max}+1)/2)$.

By the union bound, with probability at least $1-\delta$, all four estimates hold simultaneously:

$$\epsilon_\beta \leq \beta_e \leq \alpha_\beta \text{ for all } e \in E, \qquad \max_i e^{\phi_i} \leq x_\phi, \qquad \max_i e^{-\phi_i} \leq y_\phi.$$

On this event, for every edge $e = \{i, j\}$,

$$Q_e = \beta_e e^{\phi_i+\phi_j} \le \alpha_\beta x_\phi^2, \qquad Q_e = \beta_e e^{\phi_i+\phi_j} \ge \epsilon_\beta y_\phi^{-2}.$$

Thus, on this event,

$$\forall e \in E, \qquad Q_e \in [\epsilon_\beta y_\phi^{-2},\, \alpha_\beta x_\phi^2]. \tag{4.50}$$

It remains to compare the four thresholds with powers of $R = d_{\max}/\delta$. Recall that $R \ge 2$, $\mathrm{diam} \ge 1$, $n \le 2d_{\max}^{\mathrm{diam}} \le 2R^{\mathrm{diam}}$, and $|E| \le d_{\max}^{2\mathrm{diam}} \le R^{2\mathrm{diam}}$. We also repeatedly use that $\log R \ge \log 2$. Define

$$C_\epsilon := \frac{6}{\underline{a}}, \qquad C_\alpha := \frac{2\overline{a}}{\log 2} + 10, \qquad C_x := 5,$$

and

$$C_y := 8 + \frac{\log(14 + (\log 2)(\overline{a}+1))}{\log 2} + \frac{|\Psi(\underline{a})| + \log(\overline{a}+1)}{\log 2} + \sqrt{\frac{14\Psi_1(\underline{a}/2)}{\log 2}} + \frac{28}{\underline{a}}.$$

Set

$$g_1(\underline{a}, \overline{a}) := \max\{C_\alpha + 2C_x,\ C_\epsilon + 2C_y\}.$$

First,

$$\log \epsilon_\beta^{-1} = \frac{1}{\underline{a}} \log \frac{8|E|}{\delta}.$$

Since $|E| \le R^{2\mathrm{diam}}$ and $\delta^{-1} \le R$,

$$\log \frac{8|E|}{\delta} \le \log 8 + (2\mathrm{diam} + 1)\log R.$$

Moreover, $\log 8 = 3\log 2 \le 3\mathrm{diam} \log R$, and $\log R \le \mathrm{diam} \log R$. Therefore

$$\log \epsilon_\beta^{-1} \le \frac{1}{\underline{a}}\big(3\mathrm{diam}\log R + 2\mathrm{diam}\log R + \mathrm{diam}\log R\big) = \frac{6}{\underline{a}}\mathrm{diam}\log R.$$

Hence $\epsilon_\beta^{-1} \le R^{C_\epsilon \mathrm{diam}}$.

Next,

$$\alpha_\beta = 2\left(\overline{a} + \log\frac{4|E|}{\delta}\right) \le 2\left(\overline{a} + \log 4 + (2\mathrm{diam}+1)\log R\right).$$

Using $\mathrm{diam} \ge 1$ and $\log R \ge \log 2$, this gives

$$\alpha_\beta \le \left(\frac{2\overline{a}}{\log 2} + 10\right)\mathrm{diam}\log R = C_\alpha \mathrm{diam}\log R = \log(R^{C_\alpha \mathrm{diam}}) \le R^{C_\alpha \mathrm{diam}}.$$

Also,

$$x_\phi = \frac{4n}{\delta} \le 8R^{\mathrm{diam}+1},$$

because $n \leq 2R^{\text{diam}}$ and $\delta^{-1} \leq R$. Taking logarithms gives

$$\log x_\phi \leq \log 8 + (\text{diam} + 1)\log R \leq 3\text{diam}\log R + \text{diam}\log R + \text{diam}\log R = 5\text{diam}\log R.$$

Thus $x_\phi \leq R^{C_x \text{diam}}$.

It remains to bound y_ϕ. Recall that

$$\alpha_0 := \frac{\log(16/\delta)}{\log n}, \qquad n^{\alpha_0} = \frac{16}{\delta}.$$

Hence $(1+\alpha_0)\log n = \log n + \log(16/\delta)$. Since $n \leq 2R^{\text{diam}}$, $\delta^{-1} \leq R$, $R \geq 2$, and $\text{diam} \geq 1$, we have

$$\log n \leq 2\text{diam}\log R, \qquad \log(16/\delta) \leq 5\text{diam}\log R,$$

and therefore

$$(1+\alpha_0)\log n \leq 7\text{diam}\log R.$$

Taking logarithms in the definition of y_ϕ, we get

$$\begin{aligned} \log y_\phi \leq{}& \text{diam}\log 2 + \alpha_0 \log n + \frac{1}{2}\log\left(2(1+\alpha_0)\log n + (\overline{a}d_{\max}+1)\log 2\right) \\ &+ \text{diam}\Lambda + \sqrt{2(1+\alpha_0)\Psi_1(\underline{a}/2)\text{diam}\log n} + \frac{4}{\underline{a}}(1+\alpha_0)\log n, \end{aligned} \tag{4.51}$$

where $\Lambda = -\Psi(\underline{a}) + \log((\overline{a}d_{\max}+1)/2)$.

We now bound the terms on the right crudely. First,

$$\text{diam}\log 2 + \alpha_0\log n = \text{diam}\log 2 + \log(16/\delta) \leq 6\text{diam}\log R.$$

For the logarithmic square-root prefactor, since $(1+\alpha_0)\log n \leq 7\text{diam}\log R$ and $d_{\max} \leq R$,

$$2(1+\alpha_0)\log n + (\overline{a}d_{\max}+1)\log 2 \leq 14\text{diam}\log R + (\log 2)(\overline{a}+1)R. \tag{4.52}$$

Using $\text{diam}\log R \leq R^{\text{diam}}$ and $R \leq R^{\text{diam}}$, the quantity on the right-hand side of Equation (4.52) is at most $(14 + (\log 2)(\overline{a}+1))R^{\text{diam}}$. Thus, after absorbing the constant into $\text{diam}\log R$, half of the logarithm of the left-hand side of Equation (4.51) is at most

$$\left(1 + \frac{\log(14 + (\log 2)(\overline{a}+1))}{\log 2}\right)\text{diam}\log R.$$

Next, since $d_{\max} \leq R$,

$$\Lambda \leq |\Psi(\underline{a})| + \log(\overline{a}+1) + \log R,$$

and hence

$$\text{diam}\Lambda \leq \left(1 + \frac{|\Psi(\underline{a})| + \log(\overline{a}+1)}{\log 2}\right)\text{diam}\log R.$$

For the Gaussian fluctuation term, using $(1+\alpha_0)\log n \le 7\mathrm{diam}\log R$,

$$\sqrt{2(1+\alpha_0)\Psi_1(\underline{a}/2)\mathrm{diam}\log n} \le \sqrt{14\Psi_1(\underline{a}/2)}\,\mathrm{diam}\sqrt{\log R} \le \frac{\sqrt{14\Psi_1(\underline{a}/2)}}{\sqrt{\log 2}}\,\mathrm{diam}\log R.$$

Finally,

$$\frac{4}{\underline{a}}(1+\alpha_0)\log n \le \frac{28}{\underline{a}}\mathrm{diam}\log R.$$

Combining these estimates gives

$$\log y_\phi \le C_y \mathrm{diam}\log R,$$

where one may take

$$C_y := 8 + \frac{\log(14+(\log 2)(\overline{a}+1))}{\log 2} + \frac{|\Psi(\underline{a})| + \log(\overline{a}+1)}{\log 2} + \sqrt{\frac{14\Psi_1(\underline{a}/2)}{\log 2}} + \frac{28}{\underline{a}}.$$

Equivalently, $y_\phi \le R^{C_y \mathrm{diam}}$.

We have shown

$$\epsilon_\beta^{-1} \le R^{C_\epsilon \mathrm{diam}}, \qquad \alpha_\beta \le R^{C_\alpha \mathrm{diam}}, \qquad x_\phi \le R^{C_x \mathrm{diam}}, \qquad y_\phi \le R^{C_y \mathrm{diam}}.$$

Combining these with (4.50), we obtain

$$Q_e \le R^{(C_\alpha + 2C_x)\mathrm{diam}}, \qquad Q_e \ge R^{-(C_\epsilon + 2C_y)\mathrm{diam}}.$$

Thus, with

$$g_1(\underline{a}, \overline{a}) := \max\{C_\alpha + 2C_x,\ C_\epsilon + 2C_y\},$$

we have, with probability at least $1-\delta$,

$$\forall e \in E, \qquad Q_e \in \left[R^{-g_1(\underline{a},\overline{a})\mathrm{diam}}, R^{g_1(\underline{a},\overline{a})\mathrm{diam}}\right].$$

On this event, Lemma 4.4.15 implies that, if $\mathcal{R}$ is the maximum effective resistance of the unit-conductance network and $\mathcal{R}'$ is the corresponding maximum effective resistance for the conductances Q, then $\mathcal{R}' \le \mathcal{R}R^{g_1 \mathrm{diam}}$. Also, $\mathcal{C}' := \sum_{e\in E} Q_e \le |E| R^{g_1 \mathrm{diam}}$. Using Lemma 4.4.14, we conclude that

$$t_{\mathrm{cov}} \le \mathcal{C}'\mathcal{R}' \log n \le |E|\mathcal{R}\log n\, R^{2g_1 \mathrm{diam}}.$$

Finally, since $|E| \le n^2$ and $\mathcal{R} \le n^{10}$, we obtain

$$t_{\mathrm{cov}} \le n^3 \log n \left(\frac{d_{\max}}{\delta}\right)^{2g_1 \mathrm{diam}}.$$

This proves the theorem. □

[10] Note that the maximum resistance one could get is the resistance between two endpoints of an n-path.

We also have the following bound on the entries of the random transition matrix P as well as the minimum value $\pi_* := \min_{i\in V} \pi_i$ of the stationary distribution associated with it.

Corollary 4.4.17 (Lower bounds on P_{ij} and π_*)**.** *In the same setting as Theorem 4.4.16, for any $\delta \in (0,1)$, the following hold with probability at least $1-\delta$:*

$$\pi_* \geq (nd_{\max})^{-1}\left(\frac{d_{\max}}{\delta}\right)^{-2g_1(\underline{a},\overline{a})\mathrm{diam}}, \qquad \min_{i,j\in V:\, i\sim j} P_{ij} \geq d_{\max}^{-1}\left(\frac{d_{\max}}{\delta}\right)^{-2g_1(\underline{a},\overline{a})\mathrm{diam}}.$$

Proof. By the proof of Theorem 4.4.16, with probability at least $1-\delta$, every edge conductance satisfies

$$Q_e \in [B^{-1}, B], \qquad B := \left(\frac{d_{\max}}{\delta}\right)^{g_1\mathrm{diam}}.$$

On this event, since G is connected, each vertex has at least one incident edge, and hence $Q_i := \sum_{j\sim i} Q_{ij} \geq B^{-1}$. Also, $\sum_{v\in V} Q_v = 2\sum_{e\in E} Q_e \leq 2|E|B \leq nd_{\max}B$. Therefore, for every $i \in V$,

$$\pi_i = \frac{Q_i}{\sum_{v\in V} Q_v} \geq \frac{B^{-1}}{nd_{\max}B} = (nd_{\max})^{-1}\left(\frac{d_{\max}}{\delta}\right)^{-2g_1\mathrm{diam}}.$$

Taking the minimum over i gives the lower bound on π_*. Similarly, for any edge $e = \{i,j\}$, we have $Q_{ij} \geq B^{-1}$ and $Q_i = \sum_{k\sim i} Q_{ik} \leq d_{\max}B$, so

$$P_{ij} = \frac{Q_{ij}}{Q_i} \geq \frac{B^{-1}}{d_{\max}B} = d_{\max}^{-1}\left(\frac{d_{\max}}{\delta}\right)^{-2g_1\mathrm{diam}}.$$

Taking the minimum over adjacent pairs completes the proof. □

A non-asymptotic sample complexity bound

Recall the setup and notation in Section 4.3, where U_e is defined as (Equation (4.10)):

$$U_{ij} := \begin{cases} P_{ij}P_{ji}, & \text{if } e = \{i,j\} \in E, \\ 0, & \text{if } i \nsim j. \end{cases}$$

Given K independent trajectories $\{X_t^{(k)} \,|\, t \leq T, k \leq K\}$ from $\mathbb{P}_{\mathrm{errw}}^{v_0,A}$, we want to estimate the following moment-based statistics connected with $\{U_e : e \in E\}$:

i. $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e), \forall e \in E$;

ii. $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}), \forall e, e' \in E$ s.t. $|e \cap e'| = 1$;

iii. $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_{e'}) - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}), \forall e, e' \in E$ s.t. $|e \cap e'| = 1$.

Also, recall that we can couple a random conductance $Q^{(k)}$ with the k-th trajectory. We denote the corresponding transition matrix as $P^{(k)}$. We can assume that each of the K trajectories is a truncation of an infinitely long trajectory. For any $m \in \mathbb{N}_+$, let $H_{ij}^{\infty}(k;m)$ represent the number of transitions from i to j among the first m outgoing transitions from i in the original, non-truncated k-th trajectory. The following random variable is always well-defined due to the ergodic property of irreducible Markov chains:

$$\Theta_{ij}^{(k)} := \begin{cases} \frac{H_{ij}^{\infty}(k;m)}{m}, & \forall\, i,j \in V \text{ s.t. } i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{4.53}$$

Recall that $N_i(k)$ denoted the number of outgoing transitions from i in the k-th (truncated) trajectory, and $N_{ij}(k)$ denoted the number of transitions from i to j in this trajectory, see Section 4.3. Then, on the event $\{N_i(k) \geq m\}$ we have $\Theta_{ij}^{(k)} = \hat{P}_{ij}^{(k)}$ for all $i \sim j$, where $\hat{P}_{ij}^{(k)}$ is defined in Eq. (4.23).

Given the results on the cover time t_{cov} associated with a random environment that we have already proved in Theorem 4.4.16 and Corollary 4.4.17, we want to bound the complexity of approximating the three families of moment-based statistics above, up to a small multiplicative error, non-asymptotically.

Step 1. Estimating $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e), \forall e \in E$.

The proposed estimator for $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)$ is $\hat{U}_e$, which is defined in Eq. (4.25). To bound the estimation error of $\hat{U}_e$ for $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)$, the first step is to bound the estimation error of $\hat{P}^{(k)}$ for $P^{(k)}$. For $\hat{P}^{(k)}$ to achieve a small estimation error for $P^{(k)}$, one must gather sufficiently many transitions from each state $i \in V$, and then reconstruct the outgoing transition probabilities of state i based on those transitions.

Let's first consider the subproblem of statistical estimation for an n-state discrete distribution

$$\mathbf{p} := (p_1, p_2, \ldots, p_n)$$

such as to ensure that each entry is accurate up to a small multiplicative factor $\eta \in (0,1)$. Suppose we draw m i.i.d. samples from $\mathbf{p}$, and let $\hat{p}_i$ be the empirical estimator for p_i, i.e., the proportion of times that i appears in the m samples. By Hoeffding's inequality (see Theorem 2.8 of [17]), for each $i \in [n]$ we have:

$$\Pr\big(\, |\hat{p}_i - p_i| \geq \eta\, p_i\big) \leq 2\, \exp\!\big(-2\, \eta^2\, p_i^2\, m\big). \tag{4.54}$$

Taking a union bound over $i \in [n]$, it follows that

$$\Pr\Big(\exists\, i \in [n] \colon\; |\hat{p}_i - p_i| \geq \eta\, p_i\Big) \leq \sum_{i=1}^{n} \Pr\big(\, |\hat{p}_i - p_i| \geq \eta\, p_i\big) \leq 2\, n\, \exp\!\big(-2\, \eta^2\, p_*^2\, m\big), \tag{4.55}$$

where $p_* := \min_{i \in [n]} p_i$. Consequently, with probability at least $1 - 2\, n\, \exp(-2\, \eta^2 p_*^2 m)$ the empirical estimator $\{\hat{p}_i \,:\, i \in V\}$ is an entry-wise η-multiplicative approximation of $\mathbf{p}$.

Fix any $\delta' \in (0,1)$. In the ERRW setting, with probability at least $1-\delta'$, for each $k \in [K]$ the random environment satisfies $Q_e^{(k)} \in [(\frac{d_{\max}}{\delta'})^{-g_1(\underline{a},\overline{a})\mathrm{diam}}, (\frac{d_{\max}}{\delta'})^{g_1(\underline{a},\overline{a})\mathrm{diam}}]$ for all $e \in E$. Let

$$\mathcal{Q} := \left\{ q \in \mathbb{R}_{++}^E, q_e \in \left[\left(\frac{d_{\max}}{\delta'}\right)^{-g_1(\underline{a},\overline{a})\mathrm{diam}}, \left(\frac{d_{\max}}{\delta'}\right)^{g_1(\underline{a},\overline{a})\mathrm{diam}} \right], \forall e \in E \right\}. \tag{4.56}$$

Conditioned on $Q^{(k)} \in \mathcal{Q}$, we have for all $e = \{i,j\} \in E$ that

$$P_{ij}^{(k)} \geq d_{\max}^{-1} \left(\frac{d_{\max}}{\delta'}\right)^{-2g_1(\underline{a},\overline{a})\,\mathrm{diam}}, \tag{4.57}$$

as established in Corollary 4.4.17.

From Eq. (4.55), for all $Q \in \mathcal{Q}$ we also have

$$\mathbb{P}_{\mathrm{srw}}^{v_0,Q}\big(\exists\, j \sim i,\, |\Theta_{ij}^{(k)} - P_{ij}^{(k)}| \geq \eta\, P_{ij}^{(k)}\big) \leq \sum_{j\sim i} 2\exp\!\big(-2\,\eta^2\,(P_{ij}^{(k)})^2\, m\big). \tag{4.58}$$

Then we have

$$\begin{aligned}
&\mathbb{P}_{\mathrm{errw}}^{v_0,A}\big(\exists i \in V,\, \exists\, j \sim i, |\Theta_{ij}^{(k)} - P_{ij}^{(k)}| \geq \eta\, P_{ij}^{(k)}\big) \\
&\qquad = \int_{\mathcal{Q}} \mathbb{P}_{\mathrm{srw}}^{v_0,q^{(k)}}\big(\exists i \in V,\, \exists\, j \sim i,\, |\Theta_{ij}^{(k)} - P_{ij}^{(k)}| \geq \eta\, P_{ij}^{(k)}\big)\, d\nu_{v_0,A}(q^{(k)}) \\
&\qquad\quad + \int_{\mathcal{Q}^c} \mathbb{P}_{\mathrm{srw}}^{v_0,q^{(k)}}\big(\exists i \in V,\, \exists\, j \sim i,\, |\Theta_{ij}^{(k)} - P_{ij}^{(k)}| \geq \eta\, P_{ij}^{(k)}\big)\, d\nu_{v_0,A}(q^{(k)}) \\
&\qquad \leq \int_{\mathcal{Q}} \sum_{i\in V}\sum_{j\sim i} 2\exp\!\big(-2\,\eta^2\,(P_{ij}^{(k)})^2\, m\big)\, d\nu_{v_0,A}(q^{(k)}) + \int_{\mathcal{Q}^c} 1\, d\nu_{v_0,A}(q^{(k)}) \\
&\qquad \leq \int_{\mathcal{Q}} 4\,|E|\, \exp\!\left(-2\,\eta^2 \left(d_{\max}^{-1}\left(\frac{d_{\max}}{\delta'}\right)^{-2g_1(\underline{a},\overline{a})\,\mathrm{diam}}\right)^2 m\right) d\nu_{v_0,A}(q^{(k)}) \\
&\qquad\qquad + \mathbb{P}_{\mathrm{errw}}^{v_0,A}(Q^{(k)} \in \mathcal{Q}^c) \\
&\qquad \leq 2\,n d_{\max} \left(\int_{\mathcal{Q}} 1\, d\nu_{v_0,A}(q^{(k)})\right)\, \exp\!\left(-2\,\eta^2 m d_{\max}^{-2}\left(\frac{d_{\max}}{\delta'}\right)^{-4g_1(\underline{a},\overline{a})\,\mathrm{diam}}\right) \\
&\qquad\qquad + \delta' \\
&\qquad \leq 2\,n d_{\max}\, \exp\!\left(-2\,\eta^2 m d_{\max}^{-2}\left(\frac{d_{\max}}{\delta'}\right)^{-4g_1(\underline{a},\overline{a})\,\mathrm{diam}}\right) + \delta'.
\end{aligned} \tag{4.59}$$

Here we used the union bound and Eq. (4.58) in the second step, and Eq. (4.57) in the third step. Also, we used Theorem 4.4.16 in the fourth step. In the last step, we used $\int_{\mathcal{Q}} 1\, d\nu_{v_0,A}(q^{(k)}) = \mathbb{P}_{\mathrm{errw}}^{v_0,A}(Q^{(k)} \in \mathcal{Q}) \leq 1$.

To obtain m samples per state, we use the notion of the m-cover time $t_{\mathrm{cov}}(m)$. To define $t_{\mathrm{cov}}(m)$, we first define the random m-cover time $\tau_{\mathrm{cov}}(m)$ as the first time when each state has been visited at least m times. Then the m-cover time can be defined as $t_{\mathrm{cov}}(m) := \max_{v_0\in V} \mathbb{E}_{\mathrm{srw}}^{v_0,q}(\tau_{\mathrm{cov}}(m))$.

For reversible random walks, it is known (Theorem 16, [22]) that there exists a universal constant $g_2 > 0$ (which does not depend on the size of the connected graph or its structure) such that

$$t_{\mathrm{cov}}(m) \leq g_2 \cdot \Big(t_{\mathrm{cov}} + \tfrac{m}{\pi_*}\Big).$$

From the discussion that led to Theorem 4.4.16 and Corollary 4.4.17, we know that on the event $\mathcal{Q}$, which has probability $\geq 1-\delta'$, we have

$$t_{\mathrm{cov}} \leq n^3 \log n \left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\,\mathrm{diam}},$$

and $\pi_* \geq (nd_{\max})^{-1}\,(\frac{d_{\max}}{\delta'})^{-2g_1(\underline{a},\overline{a})\,\mathrm{diam}}$. Hence, on the event $\mathcal{Q}$ we have

$$t_{\mathrm{cov}}(m+1) \leq g_2 \cdot \Big(n^3 \log n + (m+1)nd_{\max}\Big) \cdot \left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\,\mathrm{diam}}.$$

Now realize that by Markov inequality, $\mathbb{P}^{v_0,q}_{\mathrm{srw}}(\tau_{\mathrm{cov}}(m+1) \geq e t_{\mathrm{cov}}(m+1)) \leq e^{-1}$. Actually according to Lemma 5 in [22], we have $\forall l \in \mathbb{Z}_+$ that $\mathbb{P}^{v_0,q}_{\mathrm{srw}}(\tau_{\mathrm{cov}}(m+1) \geq el \cdot t_{\mathrm{cov}}(m+1)) \leq e^{-l}$.

We now use the notation $\tau^{(k)}_{\mathrm{cov}}(m)$ for the m-cover time of the trajectory associated with the k-th ERRW trajectory. Conditioned on $Q^{(k)} = q$, the k-th ERRW is simply a random walk from $\mathbb{P}^{v_0,q}_{\mathrm{srw}}$. The conditional probability of not having at least $m+1$ visits to each state is bounded by

$$\mathbb{P}^{v_0,q}_{\mathrm{srw}}(\tau^{(k)}_{\mathrm{cov}}(m+1) \geq T) \leq \exp\Big(-\Big\lfloor \tfrac{T}{e\,t^{(k)}_{\mathrm{cov}}(m+1)}\Big\rfloor\Big) \leq \exp\Big(-\tfrac{T}{e\,t^{(k)}_{\mathrm{cov}}(m+1)} + 1\Big).$$

Similar to the procedure in Eq. (4.59), we have

$$\begin{aligned}
\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\tau^{(k)}_{\mathrm{cov}}(m+1) \geq T\big) &= \int_{\mathcal{Q}} \mathbb{P}^{v_0,q^{(k)}}_{\mathrm{srw}}\big(\tau^{(k)}_{\mathrm{cov}}(m+1) \geq T\big) d\nu_{v_0,A}(q^{(k)}) \\
&\quad + \int_{\mathcal{Q}^c} \mathbb{P}^{v_0,q^{(k)}}_{\mathrm{srw}}\big(\tau^{(k)}_{\mathrm{cov}}(m+1) \geq T\big) d\nu_{v_0,A}(q^{(k)}) \\
&\leq \int_{\mathcal{Q}} \exp\Big(-\tfrac{T}{e\,t^{(k)}_{\mathrm{cov}}(m+1)} + 1\Big) d\nu_{v_0,A}(q^{(k)}) + \int_{\mathcal{Q}^c} 1\, d\nu_{v_0,A}(q^{(k)}) \\
&\leq \exp\Big(-\tfrac{T}{e\,g_2 \cdot (n^3 \log n + (m+1)nd_{\max}) \cdot \left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\,\mathrm{diam}}} + 1\Big) + \delta'.
\end{aligned} \tag{4.60}$$

On the event $\{\tau^{(k)}_{\mathrm{cov}}(m+1) \leq T\}$, each state has at least m outgoing transitions before time T, and hence $\hat{P}^{(k)}_{ij} = \Theta^{(k)}_{ij}$ for all $i \sim j$. Therefore, combining Eq. (4.59) and Eq. (4.60), we conclude that

$$\begin{aligned}
&\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\{\forall i \in V,\, \forall j \sim i,\, |\hat{P}^{(k)}_{ij} - P^{(k)}_{ij}| \le \eta\, P^{(k)}_{ij}\} \cap \{\tau^{(k)}_{\mathrm{cov}}(m+1) \le T\}\big)\\
&\quad = \mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\{\forall i \in V,\, \forall j \sim i,\, |\Theta^{(k)}_{ij} - P^{(k)}_{ij}| \le \eta\, P^{(k)}_{ij}\} \cap \{\tau^{(k)}_{\mathrm{cov}}(m+1) \le T\}\big)\\
&\quad \ge 1 - 2\,nd_{\max}\,\exp\Big(-2\,\eta^2 m d_{\max}^{-2}\left(\frac{d_{\max}}{\delta'}\right)^{-4g_1(\underline{a},\overline{a})\,\mathrm{diam}}\Big)\\
&\qquad\qquad - \exp\Big(-\frac{T}{e\,g_2\cdot(n^3\log n + (m+1)nd_{\max})\cdot\left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\,\mathrm{diam}}} + 1\Big) - 2\delta'
\end{aligned} \tag{4.61}$$

Here we used the fact that on the event $\{N_i(k) \ge m, \forall i \in V\}$ we have $\Theta^{(k)}_{ij} = \hat{P}^{(k)}_{ij}$ for all $i, j \in V$ s.t. $i \sim j$. Also the condition $\{N_i(k) \ge m, \forall i \in V\}$ is implied by the condition $\{\tau^{(k)}_{\mathrm{cov}}(m+1) \le T\}$,

We now want to choose T and m appropriately so that with high probability, we simultaneously (i) gather at least m samples per state and (ii) achieve the desired η-approximation of all transition probabilities. We then take

$$m = \frac{d_{\max}^2}{2\eta^2}\left(\frac{d_{\max}}{\delta'}\right)^{4g_1(\underline{a},\overline{a})\mathrm{diam}} \cdot \log\left(\frac{2nd_{\max}}{\delta'}\right)$$

to make the term

$$2\,nd_{\max}\,\exp\Big(-2\,\eta^2 m d_{\max}^{-2}\left(\frac{d_{\max}}{\delta'}\right)^{-4g_1(\underline{a},\overline{a})\,\mathrm{diam}}\Big) \le \delta'.$$

We further take

$$T = e\,g_2(n^3\log n + (m+1)nd_{\max})\left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\mathrm{diam}} \log\left(\frac{e}{\delta'}\right)$$

to make the term

$$\exp\Big(-\frac{T}{e\,g_2\cdot(n^3\log n + (m+1)nd_{\max})\cdot\left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\,\mathrm{diam}}} + 1\Big) \le \delta'.$$

Given this choice of m and T, we have

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\{\forall i \in V,\, \forall j \sim i,\, |\hat{P}^{(k)}_{ij} - P^{(k)}_{ij}| \le \eta\, P^{(k)}_{ij}\} \cap \{\tau^{(k)}_{\mathrm{cov}}(m+1) \le T\}\big) \ge 1 - 4\delta'.$$

Chernoff bound will be used for bounding concentration over the K trajectories.

Lemma 4.4.18 (Multiplicative Chernoff for bounded variables)**.** *Let $X_1, \dots, X_K$ be i.i.d. random variables in $[0,1]$, with mean μ. Then for $\epsilon \in (0,1)$,*

$$\Pr\left(\left|\frac{1}{K}\sum_{k=1}^{K} X_k - \mu\right| \ge \epsilon\mu\right) \le 2\exp\left(-\frac{\epsilon^2}{3}\mu K\right).$$

For later use, define

$$\mu_* := \frac{\underline{a}(\underline{a}+1)}{(d_{\max}\overline{a}+1)^2}, \qquad c_0 := \frac{\underline{a}}{\underline{a}+1}, \qquad \mathcal{P}_1 := \{\{e,e'\} : e,e' \in E,\ e \neq e',\ |e\cap e'| = 1\}.$$

We now fix arbitrary $\epsilon \in (0,1/2)$. We first control the bias introduced by using the finite trajectory estimator $\hat{U}_e$ in place of the ideal quantity U_e. As before,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) = \left(\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\overline{U}_e)\right) + \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e),$$

where we used $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\overline{U}_e) = \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)$. Therefore

$$\left|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\right| \le \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}^{v_0,A}_{\mathrm{errw}}\left|\hat{U}^{(k)}_e - U^{(k)}_e\right|.$$

We next bound the summands. For every k,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left|\hat{U}^{(k)}_e - U^{(k)}_e\right| \le 2\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(|\hat{U}^{(k)}_e - U^{(k)}_e| > \eta U^{(k)}_e\right) + \eta. \tag{4.62}$$

Indeed, on the event $|\hat{U}^{(k)}_e - U^{(k)}_e| \le \eta U^{(k)}_e$ the contribution is at most $\eta\mathbb{E}(U^{(k)}_e) \le \eta$, while on the complement we use the crude bound $|\hat{U}^{(k)}_e - U^{(k)}_e| \le 2$.

Moreover, if

$$\hat{P}^{(k)}_{ij} \in \left(1 \pm \frac{\eta}{4}\right) P^{(k)}_{ij} \quad \text{and} \quad \hat{P}^{(k)}_{ji} \in \left(1 \pm \frac{\eta}{4}\right) P^{(k)}_{ji},$$

then $\hat{U}^{(k)}_e \in (1\pm\eta)U^{(k)}_e$ for every $\eta \in (0,1)$. Therefore,

$$\begin{aligned}
&\mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(|\hat{U}^{(k)}_e - U^{(k)}_e| > \eta U^{(k)}_e\right) \\
&\qquad \le \mathbb{P}^{v_0,A}_{\mathrm{errw}}\left(\left\{\exists i \in V,\ j \sim i : |\hat{P}^{(k)}_{ij} - P^{(k)}_{ij}| \ge \frac{\eta}{4}P^{(k)}_{ij}\right\} \cup \{\tau^{(k)}_{\mathrm{cov}}(m+1) > T\}\right).
\end{aligned}$$

By the transition-matrix estimate above, the last probability is at most $4\delta'$ provided

$$m \ge \frac{8d^2_{\max}}{\eta^2}\left(\frac{d_{\max}}{\delta'}\right)^{4g_1\mathrm{diam}} \log\left(\frac{2nd_{\max}}{\delta'}\right), \tag{4.63}$$

and

$$T \ge e\, g_2\big(n^3\log n + (m+1)nd_{\max}\big)\left(\frac{d_{\max}}{\delta'}\right)^{2g_1\mathrm{diam}} \log\left(\frac{e}{\delta'}\right), \tag{4.64}$$

where we write g_1 for $g_1(\underline{a},\overline{a})$.

Using this in (4.62), we get

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left|\hat{U}^{(k)}_e - U^{(k)}_e\right| \le 8\delta' + \eta.$$

Taking $\eta = \delta'$ gives

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left|\hat{U}^{(k)}_e - U^{(k)}_e\right| \le 9\delta',$$

and hence

$$\left|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\right| \le 9\delta'. \tag{4.65}$$

From the explicit moment formula,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e) = \frac{a_e(a_e+1)}{(a_i+1-\mathbf{1}_{i=v_0})(a_j+1-\mathbf{1}_{j=v_0})} \ge \mu_*. \tag{4.66}$$

If

$$\delta' \le \frac{\epsilon}{9}\mu_*, \tag{4.67}$$

then (4.65) implies

$$\left|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\right| \le \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e). \tag{4.68}$$

In particular, since $\epsilon < 1/2$,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) \ge (1-\epsilon)\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e) \ge \frac{1}{2}\mu_*.$$

We now use multiplicative Chernoff instead of Chebyshev. Since $\hat{U}_e^{(1)}, \dots, \hat{U}_e^{(K)}$ are i.i.d. random variables in $[0,1]$ and

$$\hat{U}_e = \frac{1}{K}\sum_{k=1}^{K} \hat{U}_e^{(k)},$$

the bounded-variable multiplicative Chernoff bound gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(|\hat{U}_e - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e)| \ge \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e)\big) \le 2\exp\left(-\frac{\epsilon^2}{3}\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e)K\right).$$

Using $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e) \ge \mu_*/2$, we obtain

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(|\hat{U}_e - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e)| \ge \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{U}_e)\big) \le 2\exp\left(-\frac{\epsilon^2\mu_* K}{6}\right). \tag{4.69}$$

Combining (4.68) and (4.69), we get

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(|\hat{U}_e - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)| \ge 3\epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\big) \le 2\exp\left(-\frac{\epsilon^2\mu_* K}{6}\right).$$

Finally, taking a union bound over $e \in E$ yields

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\exists e \in E:\ |\hat{U}_e - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)| \ge 3\epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\big) \le 2|E|\exp\left(-\frac{\epsilon^2\mu_* K}{6}\right). \tag{4.70}$$

In particular, since $|E| \le nd_{\max}/2$, it suffices to take

$$K \ge \frac{6}{\epsilon^2\mu_*}\log\left(\frac{nd_{\max}}{\delta}\right) \ge \frac{6}{\epsilon^2\mu_*}\log\left(\frac{2|E|}{\delta}\right)$$

to make the probability in (4.70) at most δ.

Step 2. Estimating $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'}), \forall e,e' \in E$ s.t. $|e \cap e'| = 1$.

Recall that the proposed estimator for $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'})$ is

$$\hat{V}_{e,e'} := \frac{1}{K}\sum_{k=1}^{K} \hat{U}^{(k)}_e \hat{U}^{(k)}_{e'},$$

as defined in Eq. (4.26). We first control the bias of this estimator. As in the preceding step,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}^{v_0,A}_{\mathrm{errw}}\big(\hat{U}^{(k)}_e \hat{U}^{(k)}_{e'}\big).$$

Adding and subtracting $U^{(k)}_e U^{(k)}_{e'}$, we get

$$\begin{aligned}\big|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'})\big| &\le \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}^{v_0,A}_{\mathrm{errw}}\big|\hat{U}^{(k)}_e \hat{U}^{(k)}_{e'} - U^{(k)}_e U^{(k)}_{e'}\big| \\ &\le \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}^{v_0,A}_{\mathrm{errw}}\big|\hat{U}^{(k)}_e - U^{(k)}_e\big| + \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}^{v_0,A}_{\mathrm{errw}}\big|\hat{U}^{(k)}_{e'} - U^{(k)}_{e'}\big|,\end{aligned}$$

where we used $0 \le \hat{U}^{(k)}_e, U^{(k)}_e, \hat{U}^{(k)}_{e'}, U^{(k)}_{e'} \le 1$. By Eq. (4.65), under the choices of m, T in Eqs. (4.63) and (4.64) and with $\eta = \delta'$,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\big|\hat{U}^{(k)}_e - U^{(k)}_e\big| \le 9\delta' \qquad \text{for every } e \in E.$$

Therefore

$$\big|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'})\big| \le 18\delta'. \tag{4.71}$$

Next, we lower bound the target moment. By Eq. (4.66), $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e) \ge \mu_*$ for every $e \in E$. If e and e' share the vertex j, then the moment identity in Lemma 4.3.1 gives

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'}) = \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e)\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_{e'}) \cdot \frac{a_j + 1 - \mathbf{1}_{j=v_0}}{a_j + 3 - \mathbf{1}_{j=v_0}}.$$

Moreover,

$$\frac{a_j + 1 - \mathbf{1}_{j=v_0}}{a_j + 3 - \mathbf{1}_{j=v_0}} \ge \frac{a_j}{a_j+2} \ge \frac{2\underline{a}}{2\underline{a}+2} = \frac{\underline{a}}{\underline{a}+1}.$$

Then, for all adjacent edge pairs (e, e'),

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_eU_{e'}) \ge c_0\mu_*^2. \tag{4.72}$$

Assume

$$\delta' \le \frac{\epsilon}{18}c_0\mu_*^2. \tag{4.73}$$

Then Eq. (4.71) implies

$$\left|\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})\right| \le \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'}).$$

In particular, if $\epsilon < 1/2$, then

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) \ge (1-\epsilon)\mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'}) \ge \frac{1}{2} c_0 \mu_*^2.$$

We now apply multiplicative Chernoff. For fixed e, e', the variables

$$\hat{U}^{(1)}_e \hat{U}^{(1)}_{e'}, \ldots, \hat{U}^{(K)}_e \hat{U}^{(K)}_{e'}$$

are i.i.d. and take values in $[0,1]$, with average $\hat{V}_{e,e'}$. Therefore,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(|\hat{V}_{e,e'} - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'})| \ge \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'})\Big) \le 2\exp\left(-\frac{\epsilon^2}{3}\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'})K\right).$$

Using $\mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'}) \ge \frac{1}{2}c_0\mu_*^2$, we get

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(|\hat{V}_{e,e'} - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'})| \ge \epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(\hat{V}_{e,e'})\Big) \le 2\exp\left(-\frac{\epsilon^2 c_0 \mu_*^2 K}{6}\right). \tag{4.74}$$

Combining the concentration event with the bias bound gives, as in Step 1,

$$|\hat{V}_{e,e'} - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})| \le 3\epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})$$

outside the exceptional event in Eq. (4.74). Indeed,

$$|\hat{V}_{e,e'} - \mathbb{E}(U_e U_{e'})| \le \epsilon\mathbb{E}(\hat{V}_{e,e'}) + \epsilon\mathbb{E}(U_e U_{e'}) \le 3\epsilon\mathbb{E}(U_e U_{e'}),$$

where we used $\mathbb{E}(\hat{V}_{e,e'}) \le (1+\epsilon)\mathbb{E}(U_e U_{e'})$ and $\epsilon < 1$.

Recall that $\mathcal{P}_1$ is the set of unordered adjacent edge pairs. Since

$$|\mathcal{P}_1| = \sum_{v\in V} \binom{\deg(v)}{2} \le (d_{\max}-1)|E| \le n d_{\max}^2,$$

a union bound over $\mathcal{P}_1$ gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(\exists (e,e') \in \mathcal{P}_1 :\, |\hat{V}_{e,e'} - \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})| \ge 3\epsilon\, \mathbb{E}^{v_0,A}_{\mathrm{errw}}(U_e U_{e'})\Big) \le 2|\mathcal{P}_1|\exp\left(-\frac{\epsilon^2 c_0 \mu_*^2 K}{6}\right). \tag{4.75}$$

In particular, it suffices to take

$$K \ge \frac{6}{\epsilon^2 c_0 \mu_*^2}\log\left(\frac{2|\mathcal{P}_1|}{\delta}\right)$$

to make the probability in Eq. (4.75) at most δ. Since $|\mathcal{P}_1| \leq nd_{\max}^2$, a simpler sufficient condition is

$$K \geq \frac{6}{\epsilon^2 c_0 \mu_*^2} \log\left(\frac{2nd_{\max}^2}{\delta}\right).$$

Step 3. Estimating $\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_{e'}) - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'})$ for adjacent e, e'.

Recall that $\hat{\Delta}_{e,e'}$, defined in Eq. (4.27), is the proposed estimator for

$$\Delta_{e,e'} := \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_{e'}) - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}).$$

We study $\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}$, which equals $\hat{\Delta}_{e,e'}$ whenever it is nonnegative.

The key point is that no additional variance estimate is needed. Once $\hat{U}_e, \hat{U}_{e'}$, and $\hat{V}_{e,e'}$ are multiplicatively accurate, the corresponding estimate of $\Delta_{e,e'}$ is also accurate.

Fix an internal tolerance $\epsilon_0 \in (0,1)$, to be chosen below. Suppose that for all $e \in E$ and all $(e,e') \in \mathcal{P}_1$,

$$|\hat{U}_e - \mathbb{E}(U_e)| \leq \epsilon_0\mathbb{E}(U_e), \qquad |\hat{V}_{e,e'} - \mathbb{E}(U_eU_{e'})| \leq \epsilon_0\mathbb{E}(U_eU_{e'}).$$

Then, for every $(e,e') \in \mathcal{P}_1$,

$$\begin{aligned}
&\left|(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - (\mathbb{E}(U_e)\mathbb{E}(U_{e'}) - \mathbb{E}(U_eU_{e'}))\right| \\
&\qquad \leq |\hat{U}_e\hat{U}_{e'} - \mathbb{E}(U_e)\mathbb{E}(U_{e'})| + |\hat{V}_{e,e'} - \mathbb{E}(U_eU_{e'})|.
\end{aligned}$$

The second term is at most $\epsilon_0\mathbb{E}(U_eU_{e'})$. For the first term, since $\hat{U}_e \leq (1+\epsilon_0)\mathbb{E}(U_e)$ and $\hat{U}_{e'} \leq (1+\epsilon_0)\mathbb{E}(U_{e'})$, while also $\hat{U}_e \geq (1-\epsilon_0)\mathbb{E}(U_e)$ and $\hat{U}_{e'} \geq (1-\epsilon_0)\mathbb{E}(U_{e'})$, we have

$$|\hat{U}_e\hat{U}_{e'} - \mathbb{E}(U_e)\mathbb{E}(U_{e'})| \leq 3\epsilon_0\,\mathbb{E}(U_e)\mathbb{E}(U_{e'})$$

for $\epsilon_0 \in (0,1)$. Therefore,

$$\left|(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - \Delta_{e,e'}\right| \leq 3\epsilon_0\,\mathbb{E}(U_e)\mathbb{E}(U_{e'}) + \epsilon_0\,\mathbb{E}(U_eU_{e'}). \tag{4.76}$$

Since $\mathbb{E}(U_eU_{e'}) \leq \mathbb{E}(U_e)\mathbb{E}(U_{e'})$, this gives

$$\left|(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - \Delta_{e,e'}\right| \leq 4\epsilon_0\,\mathbb{E}(U_e)\mathbb{E}(U_{e'}).$$

It remains to compare $\mathbb{E}(U_e)\mathbb{E}(U_{e'})$ with $\Delta_{e,e'}$. From the moment formulas in Lemma 4.3.1, if j is the shared vertex of e and e', then

$$\Delta_{e,e'} = \frac{2\mathbb{E}(U_e)\mathbb{E}(U_{e'})}{a_j + 3 - \mathbf{1}_{j=v_0}}.$$

Equivalently,

$$\mathbb{E}(U_e)\mathbb{E}(U_{e'}) = \frac{a_j + 3 - \mathbf{1}_{j=v_0}}{2}\Delta_{e,e'} \leq \frac{\overline{a}d_{\max} + 3}{2}\Delta_{e,e'}.$$

Thus

$$\left|(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - \Delta_{e,e'}\right| \le 2\epsilon_0(\overline{a}d_{\max}+3)\Delta_{e,e'}.$$

Choose

$$\epsilon_0 := \frac{\epsilon}{2(\overline{a}d_{\max}+3)}. \tag{4.77}$$

Then

$$\left|(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - \Delta_{e,e'}\right| \le \epsilon\,\Delta_{e,e'} \qquad \forall (e,e') \in \mathcal{P}_1. \tag{4.78}$$

In particular, since $\epsilon < 1$, the quantity $\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}$ is positive on this event, so it agrees with the estimator $\hat{\Delta}_{e,e'}$ defined in Eq. (4.27).

We now invoke the concentration bounds from Steps 1 and 2 with the final multiplicative tolerance there taken to be ϵ_0, equivalently with the auxiliary tolerance $\epsilon_0/3$ in the last Chernoff step. Thus, provided the corresponding bias conditions hold, namely

$$\delta' \le \frac{\epsilon_0}{27}\mu_*, \qquad \delta' \le \frac{\epsilon_0}{54}c_0\mu_*^2,$$

Steps 1 and 2 give

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(\exists e \in E:\, |\hat{U}_e - \mathbb{E}(U_e)| > \epsilon_0\mathbb{E}(U_e)\Big) \le 2|E|\exp\left(-\frac{\epsilon_0^2\mu_*K}{54}\right),$$

and

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(\exists (e,e') \in \mathcal{P}_1:\, |\hat{V}_{e,e'} - \mathbb{E}(U_eU_{e'})| > \epsilon_0\mathbb{E}(U_eU_{e'})\Big) \le 2|\mathcal{P}_1|\exp\left(-\frac{\epsilon_0^2c_0\mu_*^2K}{54}\right).$$

Therefore, by the deterministic implication above,

$$\begin{aligned}\mathbb{P}^{v_0,A}_{\mathrm{errw}}\Big(\exists (e,e') \in \mathcal{P}_1:\, |(\hat{U}_e\hat{U}_{e'} - \hat{V}_{e,e'}) - \Delta_{e,e'}| > \epsilon\Delta_{e,e'}\Big) &\le 2|E|\exp\left(-\frac{\epsilon_0^2\mu_*K}{54}\right)\\ &\quad + 2|\mathcal{P}_1|\exp\left(-\frac{\epsilon_0^2c_0\mu_*^2K}{54}\right).\end{aligned} \tag{4.79}$$

Since $\epsilon_0 = \epsilon/[2(\overline{a}d_{\max}+3)]$, it suffices to take

$$K \ge \max\left\{\frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2\mu_*}\log\left(\frac{4|E|}{\delta}\right),\ \frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2c_0\mu_*^2}\log\left(\frac{4|\mathcal{P}_1|}{\delta}\right)\right\}$$

to make the failure probability in (4.79) at most δ.

Step 4. Piecing it all together.

We now combine the preceding estimates. Let

$$\epsilon_0 := \frac{\epsilon}{2(\overline{a}d_{\max}+3)}.$$

To make the bias terms negligible at the internal tolerance level needed for the Δ-estimate, it is enough to choose δ' so that

$$\delta' \le \min\left\{\frac{\epsilon_0}{27}\mu_*, \frac{\epsilon_0}{54}c_0\mu_*^2\right\}.$$

Equivalently, it is enough to take, for example,

$$\delta' \le \frac{\epsilon}{108(\overline{a}d_{\max}+3)}\min\{\mu_*, c_0\mu_*^2\}.$$

With this choice of δ', and with m, T satisfying Eqs. (4.63) and (4.64) where $\eta = \delta'$, Steps 1–3 imply the following. For all $e \in E$,

$$|\hat{U}_e - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)| \le \epsilon_0\, \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)$$

fails with probability at most

$$2|E|\exp\left(-\frac{\epsilon_0^2\mu_* K}{54}\right).$$

For all $(e,e') \in \mathcal{P}_1$,

$$|\hat{V}_{e,e'} - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'})| \le \epsilon_0\, \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'})$$

fails with probability at most

$$2|\mathcal{P}_1|\exp\left(-\frac{\epsilon_0^2 c_0\mu_*^2 K}{54}\right).$$

On the intersection of these two good events, Step 3 gives

$$|\hat{\Delta}_{e,e'} - \Delta_{e,e'}| \le \epsilon\,\Delta_{e,e'} \qquad \forall (e,e') \in \mathcal{P}_1,$$

where

$$\Delta_{e,e'} := \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_{e'}) - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}).$$

Moreover, since $\epsilon_0 \le \epsilon$, the same good event also gives

$$|\hat{U}_e - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)| \le \epsilon\, \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e) \qquad \forall e \in E,$$

and

$$|\hat{V}_{e,e'} - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'})| \le \epsilon\, \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}) \qquad \forall (e,e') \in \mathcal{P}_1.$$

Therefore, by the union bound, all three families of quantities are simultaneously ϵ-multiplicatively approximated with probability at least

$$1 - 2|E|\exp\left(-\frac{\epsilon_0^2\mu_* K}{54}\right) - 2|\mathcal{P}_1|\exp\left(-\frac{\epsilon_0^2 c_0\mu_*^2 K}{54}\right).$$

Consequently, for any $\delta \in (0,1)$, it suffices to take

$$K \geq \max\left\{\frac{54}{\epsilon_0^2\mu_*}\log\left(\frac{4|E|}{\delta}\right), \frac{54}{\epsilon_0^2 c_0\mu_*^2}\log\left(\frac{4|\mathcal{P}_1|}{\delta}\right)\right\}.$$

Using $\epsilon_0 = \epsilon/[2(\overline{a}d_{\max}+3)]$, this is equivalent to

$$K \geq \max\left\{\frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2\mu_*}\log\left(\frac{4|E|}{\delta}\right), \frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2 c_0\mu_*^2}\log\left(\frac{4|\mathcal{P}_1|}{\delta}\right)\right\}.$$

Since $|E| \leq nd_{\max}/2$ and $|\mathcal{P}_1| \leq nd_{\max}^2$, a simpler sufficient condition is

$$K \geq \frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2 c_0\mu_*^2}\log\left(\frac{4nd_{\max}^2}{\delta}\right).$$

In particular,

$$K = O_{\underline{a},\overline{a}}\left(\frac{d_{\max}^6}{\epsilon^2}\log\frac{nd_{\max}}{\delta}\right).$$

Theorem 4.4.19 (Non-asymptotic bounds)**.** *Let $G = (V = [n], E)$ be a connected weighted graph with initial edge weights $A = (a_e)_{e\in E}$ and diameter* diam*. Assume that $\underline{a} \leq a_e \leq \overline{a}$ for all $e \in E$, where $\underline{a}, \overline{a}$ are positive constants. Fix $\epsilon, \delta \in (0,1)$. Let*

$$\mathcal{P}_1 := \{\{e, e'\} : e, e' \in E,\ e \neq e',\ |e \cap e'| = 1\}.$$

Define

$$\mu_* := \frac{\underline{a}(\underline{a}+1)}{(d_{\max}\overline{a}+1)^2}, \qquad c_0 := \frac{\underline{a}}{\underline{a}+1}, \qquad \epsilon_0 := \frac{\epsilon}{2(\overline{a}d_{\max}+3)}.$$

Let $\delta' \in (0,1)$ satisfy

$$\delta' \leq \frac{\epsilon}{108(\overline{a}d_{\max}+3)}\min\{\mu_*, c_0\mu_*^2\}.$$

Suppose

$$m \geq \frac{8d_{\max}^2}{(\delta')^2}\left(\frac{d_{\max}}{\delta'}\right)^{4g_1(\underline{a},\overline{a})\mathrm{diam}}\log\left(\frac{2nd_{\max}}{\delta'}\right),$$

and

$$T \geq e\, g_2\big(n^3\log n + (m+1)nd_{\max}\big)\left(\frac{d_{\max}}{\delta'}\right)^{2g_1(\underline{a},\overline{a})\mathrm{diam}}\log\left(\frac{e}{\delta'}\right).$$

If

$$K \geq \max\left\{\frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2\mu_*}\log\left(\frac{4|E|}{\delta}\right), \frac{216(\overline{a}d_{\max}+3)^2}{\epsilon^2 c_0\mu_*^2}\log\left(\frac{4|\mathcal{P}_1|}{\delta}\right)\right\},$$

then, with probability at least $1-\delta$, the estimators simultaneously give ϵ-multiplicative approximations of

$$\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e), \qquad \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'}), \qquad \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_e)\mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_{e'}) - \mathbb{E}_{\mathrm{errw}}^{v_0,A}(U_eU_{e'})$$

for all $e \in E$ and all $(e, e') \in \mathcal{P}_1$. Consequently, the induced estimates of the o_j's, and hence of the edge weights a_e, are $O(\epsilon)$-accurate.

Remark 4.4.19.1 (Constant-degree graphs)**.** *For constant-degree graphs, i.e., when* $d_{\max} = O(1)$*, the preceding theorem gives a particularly simple form. For fixed* $\epsilon, \delta \in (0,1)$ *and fixed constants* $\underline{a}, \overline{a}$*, it suffices to take*

$$K = O(\log n), \qquad m = \exp\{O(\text{diam})\} \log n,$$

and

$$T = O\big(n^3 \log n \, \exp\{O(\text{diam})\}\big) .$$

Thus, for bounded-degree graphs, the number of independent trajectories needed is only logarithmic in the graph size. The main cost is instead the required trajectory length, which is governed by the cover-time factor $\exp\{O(\text{diam})\}$*. In particular, if the graph has both bounded degree and logarithmic diameter, then* T *is polynomial in* n*.*

Remark 4.4.19.2 (Positive-integer-valued edge weights)**.** *In the special case where all edge weights are integers, one might wonder whether the sample complexity can be significantly reduced. However, the requirement to cover the graph imposes a fundamental bottleneck on the trajectory length* T*.*

To see this, consider a path consisting of $(n+1)$ *vertices and let* $A = \mathbf{1}$ *be the initial edge local time. Label the left-most vertex as* v_0 *and the right-most vertex as* v_n*. Define* $\tilde{A}$ *as an alternative hypothesis that matches* A *except for the right-most edge weight, which is set to 2 in* $\tilde{A}$*. To obtain any information about the right-most edge, the trajectory must reach it first, necessitating a traversal of the entire path. This imposes a constraint on* T*, making it at least* $2^{\Omega(n)}$*. The time to reach* v_n *from* v_0 *is asymptotically bounded below like this because it stochastically dominates the time to reach* n *from* 0 *in a biased random walk with negative drift on the nonnegative integers. So, while the positive-integer valued case is likely to be easier than the general case as it may require a lower sample complexity for the number of trajectories* K*, the dependence on the cover time* T *is still unimprovable using our current techniques.*

4.5 Summary

In this chapter, we explored how to estimate the initial weights in an edge-reinforced random walk (ERRW) given a collection of sampled trajectories. We provided lower and upper bounds on the complexity of these estimation tasks, both in terms of the length of the trajectories and in terms of the number of trajectories. We expect that the upper bounds we derived can be significantly improved with more carefully designed algorithms.

We list some other questions that seem to warrant further investigation.

Question 1. Sample-Complexity Upper Bounds for Testing

Consider identity testing: the null hypothesis H_0 is $A = A_0$, while the alternative H_1 states A is at least ϵ-far from A_0 under some metric. An interesting question is whether the sample complexity of the classical likelihood-ratio tests is competitive with tests that use

ERRW-specific structure, perhaps along the lines of the ERRW-specific properties we have used in designing the estimators derived in this chapter.

Question 2. Non-Linear Activation Functions

Beyond the linear activation setting, one can study ERRW with non-linear activation. The law of ERRW with activation function $g : \mathbb{R}^n_{++} \to \mathbb{R}^n_{++}$ starting at v_0 with initial edge local time $L_0 = A$ can be stated as

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}(g)}(X_{t+1} = i \mid \{X_0, X_1, \cdots, X_t\}) = \frac{g\big(L_t^{\{X_t,i\}}\big)}{\sum_{i':X_t\sim i'} g\big(L_t^{\{X_t,i'\}}\big)}, \quad \forall i \in V, \quad \forall t \in \mathbb{Z}_+.$$

Some examples for non-linear activation functions include $g(x) = x^\alpha$ ($\alpha \neq 1$) and $g(x) = \log(x+1)$. While such processes have been analyzed in the literature(e.g., see [70, 26]), their behavior cannot be reduced to a mixture of Markov chains, and no "magic formula" integration identity is known. Nevertheless, one might still attempt to recover the initial weights from trajectory data or, alternatively, prove that exact recovery is impossible.

Chapter 5

Information Quantities Related to Edge-reinforced Random Walks

The ERRW model, its random-environment representation, and the magic formula are reviewed in Section 1.2; we refer to [84] for further background. This chapter studies information-theoretic quantities associated with ERRWs on finite graphs, motivated in part by the problem of statistically distinguishing between different ERRW models from observed trajectories. Leveraging structural properties of the random environment [92, 93], we derive an annealed representation of the entropy rate, a closed-form formula for the environment-level KL divergence, and quantitative bounds on the convergence of trajectory-level KL divergence toward environment-level KL divergence.

These information-theoretic quantities are motivated by the two-point hypothesis testing problem for ERRW trajectories, and in particular by the associated Stein exponent. We also expect them to play a role in other testing problems for ERRWs, including identity testing and closeness testing.

5.1 Introduction

Our main results are as follows.

- *Entropy rate.* The entropy rate of ERRW admits an annealed representation in terms of the mixing measure. We derive an explicit formula for the entropy rate and an upper bound expressed via the normalizing constant in the magic formula.

- *KL divergence between environment laws.* The family of ERRW mixing measures forms a canonical exponential family, which yields a closed-form formula for the environment-level KL divergence. Through the STZ field representation [93], this formula has a probabilistic interpretation as the difference of edge-Gamma and vertex-Gamma KL divergences; see Section 5.3.

- *Convergence of trajectory-level KL divergence.* The trajectory-level KL divergence converges to the environment-level KL divergence as $T \to \infty$ at a sublinear rate on general graphs, with a tight characterization in the n-star (Pólya urn) case.

The results support the following general point of view. For ERRW on finite graphs, the random environment, especially at the level of the STZ field [93], is the natural object in terms of which entropy and KL divergence are most naturally understood. In this sense, the random-environment representation is not only a structural description of the model but also the appropriate framework for its information-theoretic analysis.

Throughout this chapter, let $G = (V, E)$ be a finite, connected, simple undirected graph, with $|V| =: n$ and $|E| =: m$. For $u, v \in V$, we write $u \sim v$ if $\{u, v\} \in E$, and for $v \in V$, we write $\deg(v) := |\{u \in V : u \sim v\}|$. We fix a distinguished initial vertex $v_0 \in V$, and we equip the edges with positive initial weights $A = (a_e)_{e\in E} \in (0, \infty)^E$. For $v \in V$, let $a_v := \sum_{e\ni v} a_e$. When $e = \{u, v\}$, we occasionally write $a_e = a_{uv} = a_{vu}$.

We recall the mixing measure from Definition 1.2.0.1, using notation adapted to this chapter. Fix a reference edge $e_0 \in E$, work in the gauge $w_{e_0} = 1$, and write $w_v := \sum_{e\ni v} w_e$ and $\tau(w) := \sum_{T\in\mathcal{T}} \prod_{e\in T} w_e$. Then

$$d\mu_{v_0,A}(w_{-e_0}) = Z_{v_0,A}^{-1} \frac{w_{v_0}^{1/2} \prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{(a_v+1)/2}} \sqrt{\tau(w)}\, dw_{-e_0}, \tag{5.1}$$

where

$$Z_{v_0,A} := \frac{\prod_{e\in E} \Gamma(a_e)}{\prod_{v\in V} \Gamma\left(\frac{a_v+1-\mathbb{1}(v=v_0)}{2}\right)} \cdot \frac{\pi^{(|V|-1)/2}}{2^{1-|V|+\sum_{e\in E} a_e}}. \tag{5.2}$$

Here Γ denotes the Gamma function. Also note that, since $w_{e_0} = 1$ we sometimes write $d\mu_{v_0,A}(w)$ for $d\mu_{v_0,A}(w_{-e_0})$, with the understanding that an additional factor of $\delta(w_{e_0} - 1)$ has been tagged on.

Conditional on $W = w$, the transition kernel is $p_{ij}(w) := \frac{w_{\{i,j\}}}{w_i}$, $i \sim j$. For a fixed environment w, the chain $P^w = (p_{ij}(w))$ is reversible with stationary distribution $\pi_i(w) := \frac{w_i}{\sum_{v\in V} w_v}$, $i \in V$. Indeed, for $i \sim j$, one has $\pi_i(w)p_{ij}(w) = w_{ij}/\sum_{v\in V} w_v = \pi_j(w)p_{ji}(w)$. The random environment representation will be the basis for our analysis of the entropy rate and KL divergence in the rest of the chapter.

Let w be a fixed environment, and let x_0^T be a trajectory with $x_0 = v_0$. Under the quenched law $\mathbb{P}_{\mathrm{srw}}^{v_0,w}$, the probability of x_0^T for $T \geq 1$ is

$$\mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_0^T = x_0^T) = \prod_{t=1}^{T} p_{x_{t-1},x_t}(w) = \prod_{u\in V} \prod_{v\sim u} \left(\frac{w_{\{u,v\}}}{w_u}\right)^{N_{uv}(x_0^T)}. \tag{5.3}$$

Equivalently, the quenched log-likelihood for $T \geq 1$ is

$$\log \mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_0^T = x_0^T) = \sum_{u\in V} \sum_{v\sim u} N_{uv}(x_0^T) \log w_{\{u,v\}} - \sum_{u\in V} N_u(x_0^T) \log w_u. \tag{5.4}$$

In particular, for fixed w, the quenched law of the trajectory depends on x_0^T only through its directed transition count vector $N(x_0^T) := (N_{uv}(x_0^T))_{u\sim v}$. We shall use this observation repeatedly.

For later use, note also that the stationary mass carried by an undirected edge $e = \{i, j\}$ in the fixed environment w is $2\pi_i(w)p_{ij}(w) = w_e/\sum_{g\in E} w_g$. Accordingly, we consider the stationary occupation measure on edges under the fixed environment w and denote it by $X_e(w) := w_e/\sum_{g\in E} w_g$, $e \in E$. This quantity will play an important role in the rest of the chapter.

The rest of this chapter is organized as follows. In Section 5.2, we study the entropy rate of ERRW. In Section 5.3, we analyze the KL divergence between ERRW environment laws. In Section 5.4, we study the KL divergence between finite trajectory laws and its relation to posterior updating. We then conclude with a discussion of several open problems.

5.2 Entropy Rate of ERRW

In this section, we study the entropy rate of ERRW through its random-environment representation. Throughout, let $(X_t)_{t\geq 0}$ denote the ERRW on (G, v_0, A), and let $W \sim \mu_{v_0,A}$ denote the random environment from the recalled random-environment representation above, where $\mu_{v_0,A}$ is defined in (5.1). Conditional on $W = w$, let $P^w = (p_{ij}(w))$ be the corresponding quenched transition kernel and let $\pi(w)$ be its stationary distribution.

Entropy rate and the finite-state Markov-chain formula

We write $\mathrm{H}(\cdot)$ for Shannon entropy with logarithms to the natural base. Since $X_0 = v_0$ is deterministic, $\mathrm{H}(X_0^T)$ for $T \geq 1$ is the same as the entropy of $(X_1, \ldots, X_T)$; nevertheless, we continue to write $\mathrm{H}(X_0^T)$ for convenience.

For each $T \geq 1$ and each environment w, define the quenched path entropy by $h_T(w) := \mathrm{H}_{\mathbb{P}_{\mathrm{srw}}^{v_0,w}}(X_0^T)$. Since V is finite, this is well-defined and satisfies $0 \leq h_T(w) \leq T\log|V|$. We define the conditional entropy of X_0^T given W by $\mathrm{H}(X_0^T \mid W) := \mathbb{E}[h_T(W)]$, and the mutual information between W and X_0^T by $I(W; X_0^T) := \mathrm{H}(X_0^T) - \mathrm{H}(X_0^T \mid W)$. Equivalently, $I(W; X_0^T) = D_{\mathrm{KL}}(P_{W,X_0^T} \,\|\, P^W \otimes P_{X_0^T})$, where D_{KL} denotes relative entropy. Since X_0^T takes values in the finite set V^{T+1}, all these quantities are finite.

Definition 5.2.0.1 (Entropy rate). *Let $(Y_t)_{t\geq 0}$ be a discrete-time process on a finite state space. Its* entropy rate *is defined as*

$$r := \lim_{T\to\infty} \frac{1}{T}\mathrm{H}(Y_0^T),$$

provided the limit exists.

The entropy rate of a finite irreducible Markov chain is well known.

Proposition 5.2.0.1 (Entropy rate of a finite Markov chain)**.** *Let $P = (p_{ij})_{i,j\in V}$ be an irreducible Markov chain on a finite state space V, and let π be its stationary distribution. Then the entropy rate exists and is given by*

$$r(P) := -\sum_{i\in V}\pi_i\sum_{j\in V}p_{ij}\log p_{ij}.$$

In the present setting, for each fixed environment w, the quenched chain P^w is irreducible on the finite connected graph G, so Proposition 5.2.0.1 applies.

Annealed reduction through the random environment

We first bound the entropy $\mathrm{H}(X_0^T)$ under the law of the ERRW on (G, v_0, A) in terms of an annealed average of the corresponding quantities for the quenched chains corresponding to each environment w.

Lemma 5.2.1 (Annealed reduction)**.** *For every $T \geq 1$,*

$$\mathrm{H}(X_0^T) = \mathrm{H}(X_0^T \mid W) + I(W; X_0^T). \tag{5.5}$$

Moreover,

$$I(W; X_0^T) \leq (2|E| - 1)\log(T + 2|E| - 1). \tag{5.6}$$

In particular, $T^{-1}I(W; X_0^T) \to 0$ as $T \to \infty$.

Proof. The identity (5.5) is immediate from the definition of mutual information.

To prove (5.6), let $N(X_0^T) := (N_{uv}(X_0^T))_{u\sim v}$ be the directed transition-count vector. By (5.3), for every fixed environment w, the quenched probability of a trajectory x_0^T depends on x_0^T only through $N(x_0^T)$. Thus, if $\mathcal{S}(n) := \{x_0^T \in V^{T+1} : N(x_0^T) = n,\ x_0 = v_0\}$, then $\mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_0^T = x_0^T)$ is constant on $\mathcal{S}(n)$. Hence, conditional on $\{N(X_0^T) = n\}$, the quenched law of X_0^T is uniform on $\mathcal{S}(n)$, and in particular does not depend on w. Equivalently, $W \perp X_0^T \mid N(X_0^T)$.

By the chain rule for mutual information, $I(W; X_0^T) = I(W; N(X_0^T)) + I(W; X_0^T \mid N(X_0^T)) = I(W; N(X_0^T))$. Since $N(X_0^T)$ is discrete, $I(W; N(X_0^T)) \leq \mathrm{H}(N(X_0^T))$.

Now $N(X_0^T)$ has $2|E|$ nonnegative integer coordinates and total mass T, so its support is contained in the set of weak compositions of T into $2|E|$ parts. Therefore

$$|\operatorname{supp}(N(X_0^T))| \leq \binom{T + 2|E| - 1}{2|E| - 1},$$

and hence

$$\mathrm{H}(N(X_0^T)) \leq \log\binom{T + 2|E| - 1}{2|E| - 1} \leq (2|E| - 1)\log(T + 2|E| - 1).$$

This proves (5.6). $\square$

In the special case of an n-star the bound of (5.6) is tight up to a constant, see Appendix C.1.

We now deduce the existence of the entropy rate of the ERRW on (G, v_0, A) and provide an annealed representation for this entropy rate.

Theorem 5.2.2 (Entropy-rate formula)**.** *The entropy rate of the ERRW on (G, v_0, A) exists. More precisely, for $\mu_{v_0,A}$-a.e. environment w, the quenched chain P^w has entropy rate*

$$r(P^w) = -\sum_{i\in V} \pi_i(w) \sum_{j\in V} p_{ij}(w) \log p_{ij}(w),$$

and

$$r(v_0, A) := \lim_{T\to\infty} \frac{1}{T} \mathrm{H}(X_0^T) = \mathbb{E}\left[r(P^W)\right]. \tag{5.7}$$

Proof. Fix an environment w. For $i \in V$, define the one-step entropy by

$$h_i(w) := -\sum_{j\in V} p_{ij}(w) \log p_{ij}(w) = -\sum_{j\sim i} \frac{w_{ij}}{w_i} \log \frac{w_{ij}}{w_i}.$$

Since under $\mathbb{P}_{\mathrm{srw}}^{v_0,w}$ the process is Markov, the chain rule for entropy gives

$$h_T(w) = \sum_{t=1}^{T} \mathrm{H}_{\mathbb{P}_{\mathrm{srw}}^{v_0,w}}(X_t \mid X_0^{t-1}),$$

and the Markov property yields $\mathrm{H}_{\mathbb{P}_{\mathrm{srw}}^{v_0,w}}(X_t \mid X_0^{t-1}) = \mathrm{H}_{\mathbb{P}_{\mathrm{srw}}^{v_0,w}}(X_t \mid X_{t-1}) = \mathbb{E}^{v_0,w}[h_{X_{t-1}}(w)]$. Therefore

$$h_T(w) = \sum_{t=0}^{T-1} \mathbb{E}^{v_0,w}[h_{X_t}(w)].$$

Since P^w is irreducible on a finite state space, its Cesàro averages converge to the stationary distribution $\pi(w)$. Thus, for each $i \in V$,

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_t = i) \longrightarrow \pi_i(w).$$

Multiplying by $h_i(w)$ and summing over i, we obtain

$$\frac{1}{T} h_T(w) = \sum_{i\in V} \left(\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_t = i) \right) h_i(w) \longrightarrow \sum_{i\in V} \pi_i(w) h_i(w) = r(P^w).$$

Moreover, $0 \le h_i(w) \le \log\deg(i) \le \log|V|$, so $0 \le T^{-1} h_T(w) \le \log|V|$ and $0 \le r(P^w) \le \log|V|$.

Now divide (5.5) by T. By Lemma 5.2.1, the mutual-information term is $O((\log T)/T) = o(1)$. On the other hand,

$$\frac{1}{T}\mathrm{H}(X_0^T \mid W) = \mathbb{E}\left[\frac{1}{T}h_T(W)\right].$$

Since $T^{-1}h_T(W) \to r(P^W)$ almost surely and $0 \le T^{-1}h_T(W) \le \log |V|$, dominated convergence gives

$$\frac{1}{T}\mathrm{H}(X_0^T \mid W) \longrightarrow \mathbb{E}[r(P^W)].$$

Combining these two limits with (5.5) proves (5.7). □

Exact formulas and bounds for the ERRW entropy rate

We now rewrite the quenched entropy rate in a form better adapted to the random environment.

Theorem 5.2.3 (Exact formula and upper bound)**.** *Let $r(v_0, A)$ be the ERRW entropy rate from Theorem 5.2.2. Then*

$$r(v_0, A) = -\int \left(\sum_{\{i,j\}\in E} \frac{w_{ij}}{\sum_{\{u,v\}\in E} w_{uv}} \log \frac{w_{ij}}{\sqrt{w_i w_j}} \right) \mu_{v_0,A}(dw), \tag{5.8}$$

where $\mu_{v_0,A}$ is defined in (5.1). For $v \in V$, write $b_v := \frac{a_v+1-\mathbb{1}(v=v_0)}{2}$. Then

$$r(v_0, A) \le \sum_{e=\{i,j\}\in E} \left(\frac{a_e}{2} \cdot \frac{\Gamma(b_i)}{\Gamma(b_i+\frac{1}{2})} \cdot \frac{\Gamma(b_j)}{\Gamma(b_j+\frac{1}{2})} \right) \left(\log 2 + \frac{1}{2}\Psi\big(b_i+\tfrac{1}{2}\big) + \frac{1}{2}\Psi\big(b_j+\tfrac{1}{2}\big) - \Psi(a_e+1) \right). \tag{5.9}$$

Here, for a real number $x > 0$, we use the standard notation $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt$ for the Gamma function and $\Psi(x) := \frac{d}{dx}\log\Gamma(x)$ for the digamma function.

Proof. By Theorem 5.2.2, $r(v_0, A) = \mathbb{E}[r(P^W)]$. For a fixed environment w,

$$r(P^w) = -\sum_{i\in V} \pi_i(w) \sum_{j\in V} p_{ij}(w) \log p_{ij}(w) = -\frac{1}{\sum_{v\in V} w_v} \sum_{i\in V} \sum_{j\sim i} w_{ij} \log \frac{w_{ij}}{w_i}.$$

For an undirected edge $\{i, j\}$, the two oriented contributions are $w_{ij} \log \frac{w_{ij}}{w_i}$ and $w_{ij} \log \frac{w_{ij}}{w_j}$, whose sum is $2w_{ij} \log \frac{w_{ij}}{\sqrt{w_i w_j}}$. Summing over all edges and using $\sum_{v\in V} w_v = 2\sum_{e\in E} w_e$, we obtain

$$r(P^w) = -\sum_{\{i,j\}\in E} \frac{w_{ij}}{\sum_{e\in E} w_e} \log \frac{w_{ij}}{\sqrt{w_i w_j}}. \tag{5.10}$$

Averaging over $W \sim \mu_{v_0,A}$ yields (5.8).

For the upper bound, define $Y_e(w) := \frac{w_e}{\sqrt{w_i w_j}}$ for $e = \{i, j\} \in E$. Since $w_e \le w_i$ and $w_e \le w_j$, we have $0 < Y_e(w) \le 1$. Also, $\sum_{g \in E} w_g \ge \frac{1}{2}(w_i + w_j) \ge \sqrt{w_i w_j}$, so

$$\frac{w_e}{\sum_{g \in E} w_g} \le Y_e(w).$$

Because $-\log Y_e(w) \ge 0$, (5.8) and (5.10) imply

$$r(P^w) \le -\sum_{e \in E} Y_e(w) \log Y_e(w), \qquad r(v_0, A) \le -\sum_{e \in E} \mathbb{E}_{\mu_{v_0,A}}[Y_e(W) \log Y_e(W)].$$

Fix $e = \{i, j\} \in E$, and let $A^{(e)} := A + \mathbb{1}_e$. Comparing the densities (5.1) for A and $A^{(e)}$, we see that increasing a_e by 1 multiplies the unnormalized density by

$$\frac{w_e}{\sqrt{w_i w_j}} = Y_e(w),$$

while the remaining change is only the ratio of normalizing constants. This remains true when $e = e_0$: in the pinned gauge $w_{e_0} = 1$, the edge factor w_{e_0} is equal to 1, but the two incident vertex factors $w_i^{-1/2}$ and $w_j^{-1/2}$ are still introduced. Hence the multiplier is still

$$Y_{e_0}(w) = \frac{w_{e_0}}{\sqrt{w_i w_j}} = \frac{1}{\sqrt{w_i w_j}}.$$

Thus

$$\frac{d\mu_{v_0,A^{(e)}}}{d\mu_{v_0,A}}(w) = \frac{Z_{v_0,A}}{Z_{v_0,A^{(e)}}}\, Y_e(w).$$

Integrating gives $\mathbb{E}_{\mu_{v_0,A}}[Y_e(W)] = \frac{Z_{v_0,A^{(e)}}}{Z_{v_0,A}}$, while multiplying by $\log Y_e(w)$ before integrating gives

$$\mathbb{E}_{\mu_{v_0,A}}[Y_e(W) \log Y_e(W)] = \mathbb{E}_{\mu_{v_0,A}}[Y_e(W)]\, \mathbb{E}_{\mu_{v_0,A^{(e)}}}[\log Y_e(W)].$$

The first factor on the RHS is computed directly from (5.2):

$$\mathbb{E}_{\mu_{v_0,A}}[Y_e(W)] = \frac{a_e}{2} \cdot \frac{\Gamma(b_i)}{\Gamma(b_i + \frac{1}{2})} \cdot \frac{\Gamma(b_j)}{\Gamma(b_j + \frac{1}{2})}.$$

For the second factor, the dependence of the density (5.1) on a_e is exactly through the factor $Y_e(w)^{a_e}$, so

$$\frac{\partial}{\partial a_e} \log Z_{v_0,A} = \mathbb{E}_{\mu_{v_0,A}}[\log Y_e(W)].$$

Differentiating the explicit formula (5.2) then gives

$$\mathbb{E}_{\mu_{v_0,A}}[\log Y_e(W)] = \Psi(a_e) - \frac{1}{2}\Psi(b_i) - \frac{1}{2}\Psi(b_j) - \log 2.$$

Applying this identity at the shifted parameter $A^{(e)}$ gives

$$\mathbb{E}_{\mu_{v_0,A^{(e)}}}[\log Y_e(W)] = \Psi(a_e+1) - \frac{1}{2}\Psi\big(b_i + \tfrac{1}{2}\big) - \frac{1}{2}\Psi\big(b_j + \tfrac{1}{2}\big) - \log 2.$$

Substituting these identities into the previous bound yields (5.9). □

Remark 5.2.3.1. *The exact formula* (5.8) *shows that the entropy rate is determined by the random environment through the ratios* $w_{ij}/\sqrt{w_i w_j}$*, while the upper bound* (5.9) *provides an estimate for this dependence explicitly in terms of the initial weights.*

5.3 KL Divergence Between Environment Laws

Throughout this section, fix a finite connected graph $G = (V, E)$, a distinguished initial vertex $v_0 \in V$, and a reference edge $e_0 \in E$. We work throughout in the pinned gauge $w_{e_0} = 1$. For $A = (a_e)_{e\in E} \in (0,\infty)^E$, let $\mu_{v_0,A}$ be the ERRW mixing measure from Definition 1.2.0.1, let $Z_{v_0,A}$ be the corresponding normalizing constant, and write $\Phi(A) := \log Z_{v_0,A}$. As before, for $v \in V$ we write $a_v := \sum_{e\ni v} a_e$ and $b_v = b_v(A) := \frac{a_v+1-\mathbb{1}(v=v_0)}{2}$.

For each edge $e = \{u, v\} \in E$, define $Y_e(w) := \frac{w_e}{\sqrt{w_u w_v}}$ and $T_e(w) := \log Y_e(w)$. Thus T_e is exactly the logarithm of the quantity that appeared already in Section 5.2.

The starting point is that, once the graph and the initial vertex are fixed, the family $\{\mu_{v_0,A}\}_{A\in(0,\infty)^E}$ forms a canonical exponential family in the standard sense of exponential-family theory; see, for example, [18, 107]. As a consequence, the KL divergence between two environment laws is determined by the log-normalizing constant $\Phi(A)$, where A denotes the vector of initial edge weights.

Explicit formula for environment-level KL divergence

We have the following formula for the KL divergence between two random environments.

Theorem 5.3.1 (Exponential-family structure and KL formula)**.** *The family* $\{\mu_{v_0,A}\}_{A\in(0,\infty)^E}$ *is a regular canonical exponential family on the pinned space* $\mathcal{W}_{e_0} := \{w \in (0,\infty)^E : \ w_{e_0} = 1\}$*, with natural parameter* A*, sufficient statistic* $T = (T_e)_{e\in E}$*, and log-partition function* Φ*. More precisely,*

$$d\mu_{v_0,A}(w) = \exp\Big(\sum_{e\in E} a_e T_e(w) - \Phi(A)\Big)\,\nu(dw), \tag{5.11}$$

where the base measure ν *does not depend on* A*. Consequently, for every* $e, e' \in E$*,*

$$\frac{\partial}{\partial a_e}\Phi(A) = \mathbb{E}_{\mu_{v_0,A}}[T_e(W)], \qquad \frac{\partial^2}{\partial a_e \partial a_{e'}}\Phi(A) = \mathrm{cov}_{\mu_{v_0,A}}\big(T_e(W), T_{e'}(W)\big). \tag{5.12}$$

If $A^{(0)}, A^{(1)} \in (0,\infty)^E$, and $\mu_s := \mu_{v_0,A^{(s)}}$ for $s \in \{0,1\}$, then the KL divergence is the Bregman divergence of Φ:

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = \Phi(A^{(1)}) - \Phi(A^{(0)}) - \sum_{e\in E}(a_e^{(1)} - a_e^{(0)})\,\frac{\partial}{\partial a_e}\Phi(A)\Big|_{A=A^{(0)}}. \tag{5.13}$$

Define, for $p, q > 0$, $\Lambda(p,q) := \log\Gamma(q) - \log\Gamma(p) - (q-p)\Psi(p)$. Then, with $b_v^{(s)} := b_v(A^{(s)})$,

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = \sum_{e\in E}\Lambda(a_e^{(0)}, a_e^{(1)}) - \sum_{v\in V}\Lambda(b_v^{(0)}, b_v^{(1)}). \tag{5.14}$$

Proof. Starting from (5.1), the density of $\mu_{v_0,A}$ is

$$d\mu_{v_0,A}(w) = Z_{v_0,A}^{-1}\,\frac{w_{v_0}^{1/2}\prod_{e\in E} w_e^{a_e-1}}{\prod_{v\in V} w_v^{(a_v+1)/2}}\sqrt{\tau(w)}\,dw_{-e_0}.$$

For each edge $e = \{u,v\}$, the dependence on a_e is exactly through the factor

$$w_e^{a_e} w_u^{-a_e/2} w_v^{-a_e/2} = Y_e(w)^{a_e} = e^{a_e T_e(w)}.$$

Thus

$$d\mu_{v_0,A}(w) = \exp\Big(\sum_{e\in E} a_e T_e(w) - \Phi(A)\Big)\,\nu(dw),$$

where

$$\nu(dw) := w_{v_0}^{1/2}\Big(\prod_{e\in E} w_e^{-1}\Big)\Big(\prod_{v\in V} w_v^{-1/2}\Big)\sqrt{\tau(w)}\,dw_{-e_0}$$

does not depend on A. This proves (5.11).

The identities (5.12) and (5.13) are the standard gradient, Hessian, and KL–Bregman identities for regular canonical exponential families; see, for example, [18, 107]. Indeed, differentiating the log-partition function in (5.11) gives the mean and covariance identities in (5.12), while direct substitution of (5.11) into $D_{\mathrm{KL}}(\mu_0\|\mu_1)$ gives the Bregman formula (5.13).

To verify (5.14), we compute the following explicit formulas. By (5.2), we have

$$\Phi(A) = \sum_{e\in E}\log\Gamma(a_e) - \sum_{v\in V}\log\Gamma(b_v) + \frac{|V|-1}{2}\log\pi - \Big(1 - |V| + \sum_{e\in E} a_e\Big)\log 2.$$

Differentiating with respect to a_e, where $e = \{u,v\}$, gives

$$\frac{\partial}{\partial a_e}\Phi(A) = \Psi(a_e) - \frac{1}{2}\Psi(b_u) - \frac{1}{2}\Psi(b_v) - \log 2. \tag{5.15}$$

Substituting these expressions in (5.13) yields (5.14).

$\square$

Remark 5.3.1.1. *The exponential-family structure in* (5.11) *is already implicit in the magic formula and in the Bayesian interpretation of ERRW as a prior on reversible Markov chains; see [25, 62, 91, 33, 80]. The point of Theorem 5.3.1 is to make explicit the associated Bregman-divergence formula for the KL divergence between two ERRW environment laws, and to record the resulting closed form in terms of Gamma and digamma functions.*

A probabilistic interpretation via the STZ field

The formula (5.14) also admits a natural probabilistic interpretation through the random-field representation developed by Sabot, Tarrès, and Zeng [92, 93]. In the original STZ field representation, the pinned field includes an auxiliary "ghost" component: the field at v_0 has an extra independent $\mathrm{Gamma}(\frac{1}{2}, 1)$ factor. We consider a version of the field with this component removed, and call the resulting reduced representation the *deghosted* parametrization. In this parametrization, the underlying field separates into an ERRW environment component together with independent Gamma variables on edges and independent Gamma variables on vertices. The KL identity then appears as the net contribution of these two Gamma structures.

Let $\beta = (\beta_e)_{e\in E}$ have independent coordinates with $\beta_e \sim \mathrm{Gamma}(a_e, 1)$. Conditional on β, and in the gauge $\phi_{v_0} = 0$, let $\phi = (\phi_i)_{i\in V}$ be sampled according to the hyperbolic Gaussian density

$$p_\beta(\phi) = \frac{1}{(2\pi)^{(n-1)/2}} e^{-\sum_{e=\{i,j\}\in E} \beta_e(\cosh(\phi_i-\phi_j)-1)} e^{-\sum_{i\in V\setminus\{v_0\}} \phi_i} \sqrt{\sum_{T\in\mathcal{T}} \prod_{e=\{i,j\}\in T} \beta_e e^{\phi_i+\phi_j}}.$$

We then define, for $i \in V$ and $\{i,j\} \in E$, $\tilde{\psi}_i := \frac{1}{2}\sum_{j\sim i} \beta_{ij} e^{\phi_j-\phi_i}$ and $W_{ij} := \beta_{ij} e^{\phi_i+\phi_j}$. Since $W_i := \sum_{j\sim i} W_{ij} = 2\tilde{\psi}_i e^{2\phi_i}$, it is natural to introduce the normalized field

$$\tilde{\beta}_{ij} = \frac{\beta_{ij}}{\sqrt{\tilde{\psi}_i\tilde{\psi}_j}} = \frac{2W_{ij}}{\sqrt{W_iW_j}}.$$

The first point is that the vertex variables separate from the environment.

Lemma 5.3.2 (The $\tilde{\psi}$ vertex field (Section 4.4))**.** *The coordinates of $\tilde{\psi}$ are independent, with $\tilde{\psi}_v \sim \mathrm{Gamma}(b_v, 1)$ for $v \in V$.*

We remark that since the deghosted ψ field differs from the original one only by an independent $\mathrm{Gamma}(\frac{1}{2}, 1)$ random variable at v_0, the proof in Section 4.4, which is carried out for the original ψ field, applies verbatim to the $\tilde{\psi}$ field.

We next record the basic structural properties of the deghosted parametrization.

Lemma 5.3.3 (Deghosted parametrization, normalized field, and factorization)**.** *Fix the gauge $\phi_{v_0} = 0$.*

1. Bijection between (β, ϕ) and $(W, \tilde{\psi})$. *The map $(\beta, \phi) \mapsto (W, \tilde{\psi})$ is measurable and one-to-one onto its image. Its inverse is given by*

$$\phi_i = \frac{1}{2} \log \frac{W_i}{2\tilde{\psi}_i}, \qquad i \in V,$$

and $\beta_{ij} = W_{ij} e^{-(\phi_i + \phi_j)}$.

2. Pinned environments and normalized fields. *Let $\mathcal{W}_{e_0} := \{w \in (0, \infty)^E : \ w_{e_0} = 1\}$, and define*

$$H : \mathcal{W}_{e_0} \to (0, \infty)^E, \qquad H(w)_{ij} := \frac{2w_{ij}}{\sqrt{w_i w_j}}, \qquad \{i, j\} \in E.$$

Then H is injective. Moreover, a vector $\tilde{\beta} \in (0, \infty)^E$ lies in the image of H if and only if the matrix $\tilde{B}$ with rows and column indexed by the vertices and with entries $\tilde{\beta}_{ij}$ has spectral radius 2. In that case, if $u > 0$ satisfies $\tilde{B}u = 2u$ and is normalized by $\frac{1}{2}\tilde{\beta}_{i_0 j_0} u_{i_0} u_{j_0} = 1$, then the unique $w \in \mathcal{W}_{e_0}$ with $H(w) = \tilde{\beta}$ is given by $w_{ij} = \frac{1}{2}\tilde{\beta}_{ij} u_i u_j$.

3. Factorization. *Let P_A denote the law of (β, ϕ) under parameter A. Then the pushforward of P_A under $(\beta, \phi) \mapsto (W, \tilde{\psi})$ factorizes as*

$$(W, \tilde{\psi}) \sim \mu_{v_0, A} \otimes \bigotimes_{v \in V} \mathrm{Gamma}(b_v, 1). \tag{5.16}$$

Equivalently, if $\tilde{\nu}_{v_0, A} := H_{\#}\mu_{v_0, A}$ denotes the law of the normalized field, then

$$(\tilde{\beta}, \tilde{\psi}) \sim \tilde{\nu}_{v_0, A} \otimes \bigotimes_{v \in V} \mathrm{Gamma}(b_v, 1). \tag{5.17}$$

Proof. For part (1), the identities $W_i = \sum_{j \sim i} W_{ij} = 2\tilde{\psi}_i e^{2\phi_i}$, $i \in V$, determine ϕ uniquely from $(W, \tilde{\psi})$, since $\phi_{v_0} = 0$ is fixed. Once ϕ is known, the variables β_{ij} are recovered from $W_{ij} = \beta_{ij} e^{\phi_i + \phi_j}$. This proves the claimed bijection.

For part (2), let $w \in \mathcal{W}_{e_0}$, let W be the matrix with rows and columns indexed by the vertices and with the entries W_{ij}, and let $D := \mathrm{diag}(w_i : i \in V)$. Then the matrix associated with $H(w)$ is $\tilde{B} = 2D^{-1/2} W D^{-1/2}$. If $P := D^{-1}W$, then P is the transition matrix of the reversible Markov chain with conductances w, and $\tilde{B} = 2D^{1/2} P D^{-1/2}$. Thus $\tilde{B}$ is similar to $2P$. Since G is connected and all edge weights are positive, P is irreducible and stochastic, so $\rho(P) = 1$. Hence $\rho(\tilde{B}) = 2$.

Now let $u_i := \sqrt{w_i}$. Then

$$(\tilde{B}u)_i = \sum_{j \sim i} \frac{2w_{ij}}{\sqrt{w_i w_j}} \sqrt{w_j} = \frac{2}{\sqrt{w_i}} \sum_{j \sim i} w_{ij} = 2\sqrt{w_i} = 2u_i.$$

Thus u is a positive eigenvector of $\tilde{B}$ for eigenvalue 2. By the Perron–Frobenius theorem, such an eigenvector is unique up to scale. Since the pinning condition $w_{e_0} = 1$ fixes the scale, H is injective.

Conversely, let $\tilde{\beta} \in (0,\infty)^E$ be such that the matrix $\tilde{B}$ with rows and columns indexed by the vertices and with entries $\tilde{\beta}_{ij}$ satisfies $\rho(\tilde{B}) = 2$. By the Perron–Frobenius theorem, there exists $u > 0$ with $\tilde{B}u = 2u$. Define $w_{ij} := \frac{1}{2}\tilde{\beta}_{ij} u_i u_j$. Then

$$w_i = \sum_{j\sim i} w_{ij} = \frac{1}{2}\, u_i \sum_{j\sim i} \tilde{\beta}_{ij} u_j = \frac{1}{2}\, u_i (\tilde{B}u)_i = u_i^2.$$

Hence

$$H(w)_{ij} = \frac{2w_{ij}}{\sqrt{w_i w_j}} = \frac{2\cdot \frac{1}{2}\tilde{\beta}_{ij} u_i u_j}{u_i u_j} = \tilde{\beta}_{ij}.$$

If u is normalized by $\frac{1}{2}\tilde{\beta}_{i_0 j_0} u_{i_0} u_{j_0} = 1$, then $w_{e_0} = 1$, so $w \in \mathcal{W}_{e_0}$. This proves the image characterization and reconstruction formula.

For part (3), part (1) shows that the change of variables from (β, ϕ) to $(W, \tilde{\psi})$ is bijective onto its image. The STZ representation implies that, after this change of variables, the W-component has law $\mu_{v_0,A}$, while the coordinates of $\tilde{\psi}$ are independent with laws Gamma$(b_v, 1)$; see [92, 93]. This proves (5.16). Applying the map H to the W-coordinate yields (5.17). $\square$

We may now recover the KL identity from the deghosted factorization.

Theorem 5.3.4 (Probabilistic interpretation of the environment KL formula)**.** *Let $A^{(0)}, A^{(1)} \in (0,\infty)^E$, and let $\mu_s := \mu_{v_0,A^{(s)}}$, $\tilde{\nu}_s := H_{\#}\mu_{v_0,A^{(s)}}$, and $b_v^{(s)} := b_v(A^{(s)})$, for $s \in \{0,1\}$. Then*

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = D_{\mathrm{KL}}(\tilde{\nu}_0\|\tilde{\nu}_1) = \sum_{e\in E} \Lambda(a_e^{(0)}, a_e^{(1)}) - \sum_{v\in V} \Lambda(b_v^{(0)}, b_v^{(1)}). \tag{5.18}$$

In particular, the environment-level KL divergence is the net contribution of the edge-Gamma and vertex-Gamma parts of the deghosted field.

Proof. Let P_s be the law of (β, ϕ) under parameter $A^{(s)}$, for $s \in \{0,1\}$. Since the conditional law of ϕ given β does not depend on A, the chain rule for relative entropy gives

$$D_{\mathrm{KL}}(P_0\|P_1) = D_{\mathrm{KL}}(\beta^{(0)}\|\beta^{(1)}) = \sum_{e\in E} \Lambda(a_e^{(0)}, a_e^{(1)}),$$

because the coordinates β_e are independent Gamma variables.

On the other hand, by Lemma 5.3.3(1), relative entropy is preserved under the bijection $(\beta, \phi) \mapsto (W, \tilde{\psi})$. By Lemma 5.3.3(3), the pushforward law factorizes as

$$(W, \tilde{\psi}) \sim \mu_{v_0,A} \otimes \bigotimes_{v\in V} \mathrm{Gamma}(b_v, 1).$$

Hence

$$D_{\mathrm{KL}}(P_0\|P_1) = D_{\mathrm{KL}}(\mu_0\|\mu_1) + \sum_{v\in V} \Lambda(b_v^{(0)}, b_v^{(1)}).$$

Comparing the two expressions yields

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = \sum_{e\in E}\Lambda(a_e^{(0)}, a_e^{(1)}) - \sum_{v\in V}\Lambda(b_v^{(0)}, b_v^{(1)}).$$

Finally, by Lemma 5.3.3(2), the map H is a bijection between the pinned environment and the normalized field. Since relative entropy is invariant under measurable bijections,

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = D_{\mathrm{KL}}(\tilde{\nu}_0\|\tilde{\nu}_1).$$

This proves (5.18). □

In light of this, the environment-level KL divergence should be understood as

$$D_{\mathrm{KL}}(\mu_0\|\mu_1) = D_{\mathrm{KL}}(\beta^{(0)}\|\beta^{(1)}) - D_{\mathrm{KL}}(\tilde{\psi}^{(0)}\|\tilde{\psi}^{(1)}) = \sum_{e\in E} D_{\mathrm{KL}}(\beta_e^{(0)}\|\beta_e^{(1)}) - \sum_{v\in V} D_{\mathrm{KL}}(\tilde{\psi}_v^{(0)}\|\tilde{\psi}_v^{(1)}).$$

Indeed, under the deghosted factorization, the raw field splits into an environment component and an independent vertex Gamma field, while the edge variables $(\beta_e)_{e\in E}$ are themselves independent Gamma variables. The ghost variable, when included, has the same law under both parameters and therefore contributes no relative entropy. Thus the KL divergence of the environment is obtained by subtracting the vertex-Gamma contribution from the total KL divergence of the independent edge-Gamma field.

5.4 KL Divergence Between Trajectory Laws

We now turn from the hidden environment to the observable trajectory. For $T \geq 1$ and $s \in \{0,1\}$, let

$$P_s^{(T)} := \mathcal{L}_{\mathbb{P}_{\mathrm{errw}}^{v_0, A^{(s)}}}(X_0^T)$$

denote the law of the first T steps of ERRW with initial weights $A^{(s)}$. These trajectory laws are the natural objects in statistical testing problems based on finite observed paths. In particular, if one observes i.i.d. T-step trajectories drawn either from $P_0^{(T)}$ or from $P_1^{(T)}$, then Stein's lemma identifies $D_{\mathrm{KL}}(P_0^{(T)}\|P_1^{(T)})$ as the optimal type-II error exponent at fixed type-I error level. The main point of this section is that the difference between the environment-level and trajectory-level KL divergences is exactly an expected posterior KL divergence.

Trajectory laws, testing, and posterior conjugacy

The relevant posterior update is especially simple for ERRW: after observing a path, one adds the undirected edge counts to the parameters and moves the distinguished root to the terminal vertex. This is the conjugacy property underlying the Bayesian approach of Diaconis and Rolles [33].

Proposition 5.4.0.1 (Posterior conjugacy and path probabilities)**.** *Fix $A \in (0,\infty)^E$, let $W \sim \mu_{v_0,A}$, and, conditional on $W = w$, let X_0^T be sampled from the quenched chain $\mathbb{P}^{v_0,w}_{\mathrm{srw}}$. For a realized path x_0^T with $x_0 = v_0$, let $N = N(x_0^T) = (N_e)_{e\in E}$ be its undirected edge-count vector and let x_T be its endpoint. Then*

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)\,\mu_{v_0,A}(dw) = \frac{Z_{x_T,A+N}}{Z_{v_0,A}}\,\mu_{x_T,A+N}(dw). \tag{5.19}$$

Consequently,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(X_0^T = x_0^T) = \frac{Z_{x_T,A+N}}{Z_{v_0,A}}, \tag{5.20}$$

and

$$\mu_{v_0,A}(dw \mid X_0^T = x_0^T) = \mu_{x_T,A+N}(dw). \tag{5.21}$$

Moreover, since the endpoint is determined by the parity rule (C.1)*, the posterior depends on the path only through N:*

$$\mu_{v_0,A}(dw \mid N) = \mu_{x_T(N),A+N}(dw). \tag{5.22}$$

Proof. See Appendix C.2. □

The posterior-gap identity and general bounds

We now compare the environment-level and trajectory-level KL divergences. Let

$$\mu_s := \mu_{v_0,A^{(s)}}, \qquad P_s^{(T)} := \mathcal{L}_{\mathbb{P}^{v_0,A^{(s)}}_{\mathrm{errw}}}(X_0^T), \qquad s \in \{0,1\},$$

and define the gap

$$\mathrm{Gap}_T := D_{\mathrm{KL}}(\mu_0\|\mu_1) - D_{\mathrm{KL}}(P_0^{(T)}\|P_1^{(T)}).$$

Theorem 5.4.1 (Posterior-gap identity and explicit formula)**.** *For every $T \geq 0$,*

$$\mathrm{Gap}_T = \mathbb{E}_{P_0^{(T)}}\left[D_{\mathrm{KL}}\big(P_0(W \mid N) \,\|\, P_1(W \mid N)\big)\right]. \tag{5.23}$$

Equivalently, if $N = N(X_0^T)$ and $x = x_T(N)$, then

$$\mathrm{Gap}_T = \mathbb{E}_{P_0^{(T)}}\left[D_{\mathrm{KL}}\big(\mu_{x,A^{(0)}+N} \,\|\, \mu_{x,A^{(1)}+N}\big)\right]. \tag{5.24}$$

For $s \in \{0,1\}$, write $a_v^{(s)} := \sum_{e\ni v} a_e^{(s)}$, and let $d_v(N) := \sum_{e\ni v} N_e$. Define

$$b_v^{(s,N)} := \frac{a_v^{(s)} + d_v(N) + 1 - \mathbb{1}(v = x_T(N))}{2}, \qquad v \in V.$$

Then

$$\mathrm{Gap}_T = \mathbb{E}_{P_0^{(T)}}\left[\sum_{e\in E}\Lambda(a_e^{(0)} + N_e, a_e^{(1)} + N_e) - \sum_{v\in V}\Lambda(b_v^{(0,N)}, b_v^{(1,N)})\right]. \tag{5.25}$$

Proof. Let Q_s be the joint law of (W, X_0^T) under parameter $A^{(s)}$. Since the conditional law of X_0^T given $W = w$ is the same quenched law $\mathbb{P}_{\mathrm{srw}}^{v_0,w}$ under both parameters, the chain rule for relative entropy gives

$$D_{\mathrm{KL}}(Q_0 \| Q_1) = D_{\mathrm{KL}}(\mu_0 \| \mu_1).$$

Decomposing in the opposite order yields

$$D_{\mathrm{KL}}(Q_0 \| Q_1) = D_{\mathrm{KL}}(P_0^{(T)} \| P_1^{(T)}) + \mathbb{E}_{P_0^{(T)}}\left[D_{\mathrm{KL}}\big(P_0(W \mid X_0^T) \,\|\, P_1(W \mid X_0^T)\big)\right].$$

Subtracting the two identities proves

$$\mathrm{Gap}_T = \mathbb{E}_{P_0^{(T)}}\left[D_{\mathrm{KL}}\big(P_0(W \mid X_0^T) \,\|\, P_1(W \mid X_0^T)\big)\right].$$

By Proposition 5.4.0.1, $P_s(W \mid X_0^T) = P_s(W \mid N) = \mu_{x, A^{(s)}+N}$, where $x = x_T(N)$. This gives (5.23) and (5.24).

It remains to apply the environment-level KL formula from Theorem 5.3.1. For the posterior measure $\mu_{x,A^{(s)}+N}$, the edge parameters are $a_e^{(s)} + N_e$, while the corresponding vertex parameters are

$$\frac{\sum_{e \ni v}(a_e^{(s)} + N_e) + 1 - \mathbb{1}(v = x)}{2} = \frac{a_v^{(s)} + d_v(N) + 1 - \mathbb{1}(v = x)}{2} = b_v^{(s,N)}.$$

Substituting these into (5.14) yields (5.25). $\square$

The exact expression (5.25) is convenient because the vertex term is nonpositive, so one can obtain upper bounds by estimating only the edge contribution. The next proposition records both an integral representation and a simple upper bound.

Proposition 5.4.1.1 (General bounds on Gap_T)**.** *Let $\delta_e := a_e^{(1)} - a_e^{(0)}$, let $\underline{a} := \min_{e\in E} \min\{a_e^{(0)}, a_e^{(1)}\}$, and let $\Psi_1 = \Psi'$ be the trigamma function. Then*

$$\mathrm{Gap}_T = \int_0^1 (1-t)\Bigg(\sum_{e\in E} \delta_e^2\, \mathbb{E}_{P_0^{(T)}}\Big[\Psi_1\big(a_e^{(0)} + N_e + t\delta_e\big)\Big] - \frac{1}{4}\sum_{v\in V} \delta_v^2\, \mathbb{E}_{P_0^{(T)}}\Big[\Psi_1\big(b_v^{(0,N)} + \tfrac{t\delta_v}{2}\big)\Big]\Bigg)\, dt, \tag{5.26}$$

where $\delta_v := a_v^{(1)} - a_v^{(0)} = \sum_{e\ni v} \delta_e$.

In particular,

$$\mathrm{Gap}_T \le \frac{1}{2}\sum_{e\in E} \delta_e^2\, \mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e + \underline{a}} + \frac{1}{(N_e + \underline{a})^2}\right]. \tag{5.27}$$

Hence also

$$\mathrm{Gap}_T \le \frac{1}{2}\Big(1 + \frac{1}{\underline{a}}\Big)\sum_{e\in E} \delta_e^2\, \mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e + \underline{a}}\right]. \tag{5.28}$$

Proof. We use the elementary identity

$$\Lambda(p,q) = (q-p)^2 \int_0^1 (1-t)\, \Psi_1\big(p+t(q-p)\big)\, dt, \qquad p,q>0,$$

which follows from Taylor's theorem with integral remainder applied to $\log\Gamma$. Applying this to each term in (5.25) yields (5.26).

For the upper bound, discard the nonpositive vertex term in (5.25). Since $\Psi_1(x) \le x^{-1} + x^{-2}$ for $x > 0$, we obtain

$$\Lambda(a_e^{(0)} + N_e, a_e^{(1)} + N_e) \le \frac{\delta_e^2}{2} \left(\frac{1}{m_e(N)} + \frac{1}{m_e(N)^2} \right),$$

where $m_e(N) := \min\{a_e^{(0)} + N_e, a_e^{(1)} + N_e\}$. Since $m_e(N) \ge N_e + \underline{a}$, this gives (5.27) after taking expectation. Finally,

$$\frac{1}{(N_e+\underline{a})^2} \le \frac{1}{\underline{a}} \cdot \frac{1}{N_e+\underline{a}},$$

which implies (5.28). □

The n-star case

We next consider the n-star, in which ERRW reduces to a Pólya urn. This gives a tractable model in which the inverse local-time estimates underlying the KL gap can be analyzed sharply.

Let G be the n-star with center v_0 and n leaves, and suppose that all initial edge weights are equal to $a > 0$. Fix one leaf edge e. Since each excursion from the center consists of traversing exactly one leaf edge out and then back, it is natural to index the process by excursions from v_0. For $T \ge 1$, let $N_e(T)$, with a harmless abuse of notation, denote the number of the first T excursions that use the edge e.

Theorem 5.4.2 (Inverse local-time bound on the n-star). *Fix $\eta > 0$. Then starting at the center node v_0, as $T \to \infty$,*

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] = \begin{cases} O(T^{-1}), & a>1,\\ O((\log T)\, T^{-1}), & a=1,\\ O(T^{-a}), & 0<a<1, \end{cases}$$

with constants depending only on a, n, and η.

The proof is based on the beta-binomial representation of $N_e(T)$ together with a sharp estimate on the binomial inverse moment $g_T(x) := \mathbb{E}[(\mathrm{Bin}(T,x)+\eta)^{-1}]$; see Appendix C.3.

Combining Theorem 5.4.2 with Proposition 5.4.1.1 yields the corresponding decay rates for the trajectory-level KL gap.

Corollary 5.4.3 (Gap decay on the n-star). *Let G be the n-star, and let $A^{(s)}$, $s \in \{0,1\}$, be constant edge-weight vectors with values $a^{(s)} > 0$. Then starting at the center node v_0, we have*

$$\mathrm{Gap}_T = \begin{cases} O(T^{-1}), & a^{(0)}, a^{(1)} > 1, \\ O((\log T)\, T^{-1}), & \min\{a^{(0)}, a^{(1)}\} = 1, \\ O\big(T^{-\min\{a^{(0)}, a^{(1)}\}}\big), & \min\{a^{(0)}, a^{(1)}\} < 1. \end{cases}$$

Proof. Apply Proposition 5.4.1.1. In the n-star with constant initial weights, all edge counts are identically distributed under $P_0^{(T)}$, so the right-hand side of (5.28) is bounded by a constant multiple of $\mathbb{E}[(N_e(T)+\eta)^{-1}]$, for any fixed $\eta > 0$. The stated rates then follow from Theorem 5.4.2. □

The general graph case

We carry out the proof in three steps, followed by a step to try to optimize the bounds.

Step 1: tails of the random environment. We begin by controlling the lower tail of the normalized environment

$$X_e := \frac{W_e}{\sum_{g \in E} W_g}, \qquad e \in E,$$

where $W \sim \mu_{v_0,A}$. By [83], the law of $X = (X_g)_{g \in E}$ has a density on the open simplex

$$\Delta_E := \Big\{ x = (x_g)_{g \in E} \in (0,1)^E : \sum_{g \in E} x_g = 1 \Big\}$$

of the form

$$\rho_{v_0,A}(x) = \tilde{Z}_{v_0,A}^{-1} \frac{\prod_{g \in E} x_g^{a_g - 1}}{\prod_{v \in V} x_v^{b_v}} \sqrt{\tau(x)}, \qquad x \in \Delta_E, \tag{5.29}$$

where $x_v := \sum_{g \ni v} x_g$, $b_v := \frac{a_v + 1 - \mathbb{1}(v = v_0)}{2}$, $\tau(x) := \sum_{T \in \mathcal{T}} \prod_{g \in T} x_g$, and $\tilde{Z}_{v_0,A} = \Gamma(|E|) \cdot Z_{v_0,A}$ is the normalizing constant.

For a nonempty set $F \subseteq E$, let $a(F) := \sum_{g \in F} a_g$, and let

$$S(F) := \{ v \in V : \text{ every edge incident to } v \text{ belongs to } F \}.$$

We also write $\kappa(G \setminus F)$ for the number of connected components of the graph obtained by deleting the edges in F, and define

$$A(F) := a(F) - \sum_{v \in S(F)} b_v + \frac{\kappa(G \setminus F) - 1}{2}. \tag{5.30}$$

The key combinatorial input is the following fact.

Lemma 5.4.4. *Assume that $a_g \geq \underline{a} > 0$ for all $g \in E$. Then, for every nonempty proper subset $F \subsetneq E$, one has*

$$A(F) \geq \frac{\underline{a}}{2}.$$

Proof. Deferred to Appendix C.4. Note that the claim does not hold if $F = \emptyset$ or if $F = E$, because $A(F) = 0$ in both these cases. $\square$

We will also make use of the following lemma.

Lemma 5.4.5 (Ordered small-edge coordinates)**.** *Let $F \subsetneq E$ be nonempty, write $m := |F|$, and fix an ordering π, such that the edges in F is ordered as $\{g_1, \ldots, g_m\}$ under π. Fix $\delta_0 := (2|E|)^{-1}$, define*

$$\Omega(F, \pi) := \{x \in \Delta_E : \ 0 < x_{g_1} \leq \cdots \leq x_{g_m} < \delta_0, \quad x_h \geq \delta_0 \ \forall h \in E \setminus F\}. \tag{5.31}$$

For $x \in \Omega(F, \pi)$, define $r := x_{g_m}$ and $t_j := x_{g_j}/x_{g_{j+1}} \in (0, 1]$ for $1 \leq j \leq m-1$. Then $x_{g_j} = r \prod_{\ell=j}^{m-1} t_\ell$ for every $1 \leq j \leq m$, and

$$dx_F := \prod_{j=1}^{m} dx_j = r^{m-1} \prod_{j=1}^{m-1} t_j^{j-1} \, dr \, dt_1 \cdots dt_{m-1}.$$

More generally, for any real numbers $(\alpha_g)_{g \in F}$, if $\alpha(F_j^\pi) := \sum_{k=1}^{j} \alpha_{g_k}$, then

$$\Big(\prod_{g \in F} x_g^{\alpha_g - 1}\Big) \, dx_F = r^{\alpha(F)-1} \prod_{j=1}^{m-1} t_j^{\alpha(F_j^\pi)-1} \, dr \, dt_1 \cdots dt_{m-1}.$$

In particular, if $e = g_q \in F$ and $s \geq 0$, then

$$x_e^{-s} = r^{-s} \prod_{j=q}^{m-1} t_j^{-s} = r^{-s} \prod_{j=1}^{m-1} t_j^{-s \, \mathbb{1}(e \in F_j^\pi)}.$$

Proof. Since $x_{g_j} = t_j x_{g_{j+1}}$ for $1 \leq j \leq m-1$ and $x_{g_m} = r$, iterating gives $x_{g_j} = r \prod_{\ell=j}^{m-1} t_\ell$. Writing $y_j := x_{g_j}$, we have $y_m = r$ and $y_j = t_j y_{j+1}$, so

$$dy_1 \cdots dy_m = (y_2 \cdots y_m) \, dt_1 \cdots dt_{m-1} \, dr.$$

Since $y_k = r \prod_{\ell=k}^{m-1} t_\ell$, the product $y_2 \cdots y_m$ equals $r^{m-1} \prod_{j=1}^{m-1} t_j^{j-1}$, because t_j appears in exactly $j-1$ of the factors $y_2, \ldots, y_m$. This proves the Jacobian formula.

For the monomial identity, substitute $x_{g_k} = r \prod_{\ell=k}^{m-1} t_\ell$ into $\prod_{g \in F} x_g^{\alpha_g - 1}$. The exponent of r from this product is $\sum_{k=1}^{m} (\alpha_{g_k} - 1) = \alpha(F) - m$, while the exponent of t_j is $\sum_{k=1}^{j} (\alpha_{g_k} - 1) = \alpha(F_j^\pi) - j$, since t_j appears exactly in $x_{g_1}, \ldots, x_{g_j}$. After multiplying by the Jacobian, these become $\alpha(F) - 1$ and $\alpha(F_j^\pi) - 1$, respectively, proving the second display.

Finally, if $e = g_q$, then $x_e = x_{g_q} = r \prod_{j=q}^{m-1} t_j$, and raising to the power $-s$ gives the last identity. The alternative form follows because $e = g_q \in F_j^\pi$ holds exactly when $q \leq j$. $\square$

We may now prove the required negative-moment and tail bounds.

Theorem 5.4.6 (Small-edge tail bound)**.** *Assume that $a_g \geq \underline{a} > 0$ for all $g \in E$. Fix $e \in E$ and $0 < s < \underline{a}/2$. Let $\delta_0 := (2|E|)^{-1}$. For each nonempty proper subset $F \subsetneq E$ with $e \in F$, define*

$$B_{\mathrm{out}}(F) := \sum_{v \notin S(F)} b_v \qquad \text{and} \qquad M_{\mathrm{edge}}(F) := \prod_{h \in E \setminus F} \max\{1, \delta_0^{a_h - 1}\}.$$

Then

$$\mathbb{E}[X_e^{-s}] \leq (2|E|)^s + \frac{1}{\tilde{Z}_{v_0,A}} \sum_{\substack{\varnothing \neq F \subsetneq E \\ e \in F}} |F|! \sqrt{|\mathcal{T}|}\, M_{\mathrm{edge}}(F)\, \delta_0^{-B_{\mathrm{out}}(F)} \frac{\delta_0^{\underline{a}/2 - s}}{(\underline{a}/2 - s)^{|F|}}. \tag{5.32}$$

In particular, $\mathbb{E}[X_e^{-s}] < \infty$. Consequently, if

$$C_{e,s} := (2|E|)^s + \frac{1}{\tilde{Z}_{v_0,A}} \sum_{\substack{\varnothing \neq F \subsetneq E \\ e \in F}} |F|! \sqrt{|\mathcal{T}|}\, M_{\mathrm{edge}}(F)\, \delta_0^{-B_{\mathrm{out}}(F)} \frac{\delta_0^{\underline{a}/2 - s}}{(\underline{a}/2 - s)^{|F|}}, \tag{5.33}$$

then

$$\mathbb{P}(X_e < \varepsilon) \leq C_{e,s}\, \varepsilon^s, \qquad 0 < \varepsilon < 1. \tag{5.34}$$

Proof. Write

$$\mathbb{E}[X_e^{-s}] = \int_{\Delta_E} x_e^{-s} \rho_{v_0,A}(x)\, dx = \int_{\{x_e \geq \delta_0\}} x_e^{-s} \rho_{v_0,A}(x)\, dx + \int_{\{x_e < \delta_0\}} x_e^{-s} \rho_{v_0,A}(x)\, dx.$$

On $\{x_e \geq \delta_0\}$, one has $x_e^{-s} \leq \delta_0^{-s} = (2|E|)^s$, so

$$\int_{\{x_e \geq \delta_0\}} x_e^{-s} \rho_{v_0,A}(x)\, dx \leq (2|E|)^s.$$

It remains to estimate the integral over $\{x_e < \delta_0\}$. For each x in this set, let $F(x) := \{g \in E : \ x_g < \delta_0\}$. Since $\delta_0 = (2|E|)^{-1} < |E|^{-1}$, not all coordinates of x can be $< \delta_0$, so $F(x)$ is always a nonempty proper subset of E, and it contains e. Thus

$$\{x_e < \delta_0\} = \bigcup_{\substack{\varnothing \neq F \subsetneq E \\ e \in F}} \bigcup_{\pi \in \mathfrak{S}(F)} \Omega(F, \pi),$$

where, for $F = \{g_1, \ldots, g_m\}$ listed in the order π, $\Omega(F, \pi)$ is defined in Eq. (5.31). The union is disjoint up to null sets.

Fix such a pair (F, π), write $m := |F|$, and let $F_j^\pi := \{g_1, \ldots, g_j\}$. On $\Omega(F, \pi)$, introduce the coordinates $r := x_{g_m}$ and $t_j := x_{g_j}/x_{g_{j+1}} \in (0, 1]$ for $j = 1, \ldots, m-1$. By Lemma 5.4.5,

$$\Big(\prod_{g \in F} x_g^{a_g - 1}\Big)\, dx_F = r^{a(F)-1} \prod_{j=1}^{m-1} t_j^{a(F_j^\pi)-1}\, dr\, dt_1 \cdots dt_{m-1}.$$

For the remaining factors, we use the following bounds on $\Omega(F,\pi)$. First, for $h \in E \setminus F$, since $x_h \in [\delta_0, 1]$, we have $x_h^{a_h-1} \le \max\{1, \delta_0^{a_h-1}\}$, and hence $\prod_{h\in E\setminus F} x_h^{a_h-1} \le M_{\text{edge}}(F)$.

Second, if $v \notin S(F)$, then some incident edge lies outside F, so $x_v \ge \delta_0$, and therefore $x_v^{-b_v} \le \delta_0^{-b_v}$. For $v \in S(F)$, all incident edges of v belong to F. Let

$$m_v := \max\{k \in \{1,\dots,m\} : g_k \text{ is incident to } v\}.$$

Then $x_v = \sum_{e\ni v} x_e \ge x_{g_{m_v}}$, and since $x_{g_k} = r\prod_{\ell=k}^{m-1} t_\ell$, we get

$$x_v \ge x_{g_{m_v}} = r\prod_{\ell=m_v}^{m-1} t_\ell, \qquad \text{so} \qquad x_v^{-b_v} \le r^{-b_v}\prod_{\ell=m_v}^{m-1} t_\ell^{-b_v}.$$

Multiplying these bounds over $v \notin S(F)$ gives the factor $\delta_0^{-B_{\text{out}}(F)}$, where

$$B_{\text{out}}(F) := \sum_{v\notin S(F)} b_v.$$

Multiplying over $v \in S(F)$ gives the factor

$$r^{-\sum_{v\in S(F)} b_v} \prod_{v\in S(F)} \prod_{\ell=m_v}^{m-1} t_\ell^{-b_v}.$$

Now fix $j \in \{1,\dots,m-1\}$. The variable t_j appears in the contribution of v exactly when $m_v \le j$. Since m_v is the largest index of an edge in F incident to v, the condition $m_v \le j$ is equivalent to saying that every edge incident to v lies in $F_j^\pi = \{g_1,\dots,g_j\}$, i.e. $v \in S(F_j^\pi)$. Indeed, we show both implications.

($\Rightarrow$) Suppose $m_v \le j$. By definition, m_v is the largest index k such that g_k is incident to v. Hence no edge g_k with $k > j$ is incident to v. Since $v \in S(F)$, every edge incident to v belongs to F, so every edge incident to v must in fact belong to $F_j^\pi = \{g_1,\dots,g_j\}$. Therefore $v \in S(F_j^\pi)$.

($\Leftarrow$) Conversely, suppose $v \in S(F_j^\pi)$. Then every edge incident to v lies in F_j^π. In particular, if g_k is incident to v, then $k \le j$. Since m_v is the largest index of an edge in F incident to v, it follows that $m_v \le j$.

Thus

$$m_v \le j \quad \Longleftrightarrow \quad v \in S(F_j^\pi).$$

Therefore the total exponent of t_j is $-\sum_{v\in S(F):\, m_v\le j} b_v = -\sum_{v\in S(F_j^\pi)} b_v$. Combining everything, we obtain

$$\prod_{v\in V} x_v^{-b_v} \le \delta_0^{-B_{\text{out}}(F)}\, r^{-\sum_{v\in S(F)} b_v} \prod_{j=1}^{m-1} t_j^{-\sum_{v\in S(F_j^\pi)} b_v}.$$

Third, for the spanning-tree factor, fix $T \in \mathcal{T}$. Since $x_h \le 1$ for $h \in E \setminus F$, we have

$$\prod_{g \in T} x_g \le r^{|T \cap F|} \prod_{j=1}^{m-1} t_j^{|T \cap F_j^\pi|}.$$

Indeed,

$$\prod_{g \in T} x_g = \Big(\prod_{g \in T \cap F} x_g \Big) \Big(\prod_{h \in T \setminus F} x_h \Big).$$

Since $x_h \le 1$ for every $h \in E \setminus F$, we have $\prod_{g \in T} x_g \le \prod_{g \in T \cap F} x_g$. Now write $T \cap F = \{g_{k_1}, \dots, g_{k_q}\}$, where $q = |T \cap F|$. For each such edge g_k, we have $x_{g_k} = r \prod_{\ell=k}^{m-1} t_\ell$. Therefore

$$\prod_{g \in T \cap F} x_g = \prod_{g_k \in T \cap F} \left(r \prod_{\ell=k}^{m-1} t_\ell \right) = r^{|T \cap F|} \prod_{j=1}^{m-1} t_j^{\#\{g_k \in T \cap F:\ k \le j\}}.$$

But $\#\{g_k \in T \cap F:\ k \le j\} = |T \cap F_j^\pi|$, since $F_j^\pi = \{g_1, \dots, g_j\}$. Hence

$$\prod_{g \in T} x_g \le r^{|T \cap F|} \prod_{j=1}^{m-1} t_j^{|T \cap F_j^\pi|}.$$

Because a spanning tree must reconnect the $\kappa(G \setminus F_j^\pi)$ components of $G \setminus F_j^\pi$, one has $|T \cap F_j^\pi| \ge \kappa(G \setminus F_j^\pi) - 1$. This can be seen more directly by comparing $G \setminus F_j^\pi$ with the spanning subgraph $T \setminus F_j^\pi$: since $T \setminus F_j^\pi \subseteq G \setminus F_j^\pi$ and both graphs have the same vertex set, adding edges can only decrease the number of connected components, so $\kappa(G \setminus F_j^\pi) \le \kappa(T \setminus F_j^\pi)$. Now T is a tree, so removing k edges from T produces exactly $k+1$ connected components; taking $k = |T \cap F_j^\pi|$ gives $\kappa(T \setminus F_j^\pi) = |T \cap F_j^\pi| + 1$. Combining these yields

$$\kappa(G \setminus F_j^\pi) \le |T \cap F_j^\pi| + 1.$$

Hence

$$\sqrt{\tau(x)} \le \sqrt{|\mathcal{T}|}\, r^{(\kappa(G \setminus F)-1)/2} \prod_{j=1}^{m-1} t_j^{(\kappa(G \setminus F_j^\pi)-1)/2}.$$

Also, by Lemma 5.4.5,

$$x_e^{-s} = r^{-s} \prod_{j=1}^{m-1} t_j^{-s\, \mathbb{1}(e \in F_j^\pi)}.$$

Recall that on $\Omega(F, \pi)$,

$$x_e^{-s} \rho_{v_0,A}(x)\, dx = \tilde{Z}_{v_0,A}^{-1}\, x_e^{-s} \Big(\prod_{g \in F} x_g^{a_g - 1} \Big) \Big(\prod_{h \in E \setminus F} x_h^{a_h - 1} \Big) \Big(\prod_{v \in V} x_v^{-b_v} \Big) \sqrt{\tau(x)}\, dx_F\, dz.$$

Using the preceding bounds together with Lemma 5.4.5,

$$\Big(\prod_{g\in F} x_g^{a_g-1}\Big)\,dx_F = r^{a(F)-1}\prod_{j=1}^{m-1} t_j^{a(F_j^\pi)-1}\,dr\,dt_1\cdots dt_{m-1}, \qquad x_e^{-s} = r^{-s}\prod_{j=1}^{m-1} t_j^{-s\,\mathbb{1}(e\in F_j^\pi)}.$$

Hence

$$x_e^{-s}\rho_{v_0,A}(x)\,dx \le \frac{\sqrt{|\mathcal{T}|}\,M_{\mathrm{edge}}(F)\,\delta_0^{-B_{\mathrm{out}}(F)}}{\tilde Z_{v_0,A}}\,r^{A(F)-1-s}\prod_{j=1}^{m-1} t_j^{A(F_j^\pi)-1-s\,\mathbb{1}(e\in F_j^\pi)}\,dr\,dt_1\cdots dt_{m-1}\,dz,$$

because the exponent of r is $a(F)-1-\sum_{v\in S(F)} b_v + (\kappa(G\setminus F)-1)/2 - s = A(F)-1-s$, and similarly the exponent of t_j is $a(F_j^\pi)-1-\sum_{v\in S(F_j^\pi)} b_v + (\kappa(G\setminus F_j^\pi)-1)/2 - s\,\mathbb{1}(e\in F_j^\pi) = A(F_j^\pi)-1-s\,\mathbb{1}(e\in F_j^\pi)$. Therefore

$$x_e^{-s}\rho_{v_0,A}(x)\,dx \le \frac{\sqrt{|\mathcal{T}|}\,M_{\mathrm{edge}}(F)\,\delta_0^{-B_{\mathrm{out}}(F)}}{\tilde Z_{v_0,A}}\,r^{A(F)-1-s}\prod_{j=1}^{m-1} t_j^{A(F_j^\pi)-1-s\,\mathbb{1}(e\in F_j^\pi)}\,dr\,dt_1\cdots dt_{m-1}\,dz.$$

where dz denotes the remaining simplex coordinates outside F. To make the z-coordinates precise, fix once and for all an edge $h_*\in E\setminus F$ (this is possible since $F\subsetneq E$). Write

$$E\setminus F = \{h_*\}\cup H, \qquad |H| = |E|-|F|-1.$$

For $h\in H$, set $z_h := x_h$, and let $dz := \prod_{h\in H} dz_h$. Then, on Δ_E, the remaining outside coordinate is determined by the simplex constraint:

$$x_{h_*} = 1-\sum_{g\in F} x_g - \sum_{h\in H} z_h.$$

Thus $(r,t_1,\dots,t_{m-1},z)$ gives a coordinate system on $\Omega(F,\pi)$, where $z=(z_h)_{h\in H}\in\mathbb{R}^{|E|-|F|-1}$. More explicitly, the admissible z-set is

$$D(r,t) := \Big\{(z_h)_{h\in H}:\ z_h\ge\delta_0\ \forall h\in H,\ \ x_{h_*} = 1-\sum_{g\in F} x_g - \sum_{h\in H} z_h \ge \delta_0\Big\}.$$

In particular, since each coordinate satisfies $\delta_0\le z_h\le 1$, we have

$$D(r,t)\subseteq[\delta_0,1]^{|E|-|F|-1}\subseteq[0,1]^{|E|-|F|-1}.$$

Hence its Lebesgue measure is at most 1:

$$|D(r,t)|\le 1.$$

Therefore, after obtaining an upper bound that is independent of z, integrating over the z-variables contributes at most a factor 1. If $|E\setminus F|=1$, then $H=\varnothing$, there are no z-variables, and this factor is interpreted as 1.

By Lemma 5.4.4, every exponent $A(F)$ and $A(F_j^\pi)$ is at least $\underline{a}/2$. Since $0 < s < \underline{a}/2$, all the resulting integrals are finite, and

$$\int_0^{\delta_0} r^{A(F)-1-s}\,dr \le \frac{\delta_0^{\underline{a}/2-s}}{\underline{a}/2-s}, \qquad \int_0^1 t_j^{A(F_j^\pi)-1-s\,\mathbb{1}(e\in F_j^\pi)}\,dt_j \le \frac{1}{\underline{a}/2-s}.$$

Hence

$$\int_{\Omega(F,\pi)} x_e^{-s}\rho_{v_0,A}(x)\,dx \le \frac{\sqrt{|\mathcal{T}|}\,M_{\mathrm{edge}}(F)\,\delta_0^{-B_{\mathrm{out}}(F)}}{\tilde{Z}_{v_0,A}}\frac{\delta_0^{\underline{a}/2-s}}{(\underline{a}/2-s)^{|F|}}.$$

Summing over the $|F|!$ possible orderings π and then over all nonempty proper $F \subsetneq E$ containing e gives (5.32).

Finally, since $\mathbb{E}[X_e^{-s}] < \infty$, Markov's inequality gives

$$\mathbb{P}(X_e < \varepsilon) = \mathbb{P}(X_e^{-s} > \varepsilon^{-s}) \le \varepsilon^s\,\mathbb{E}[X_e^{-s}],$$

and substituting (5.32) yields (5.34). □

The preceding theorem has two immediate consequences that will be used later.

Lemma 5.4.7 (Root-mass tail and Laplace consequence)**.** *Assume that $a_g \ge \underline{a} > 0$ for all $g \in E$, and fix $0 < s < \underline{a}/2$.*

1. *Let π^W be the stationary distribution associated with the random environment W. Then, for every $\varepsilon > 0$,*
$$\mathbb{P}\big(\pi^W(v_0) < \varepsilon\big) \le 2^s\Big(\min_{g:v_0\in g} C_{g,s}\Big)\varepsilon^s. \tag{5.35}$$

2. *For every $u > 0$ and every edge $e \in E$,*
$$\mathbb{E}[e^{-uX_e^2}] \le C_{s/2}\,u^{-s/2}\,\mathbb{E}[X_e^{-s}], \qquad C_{s/2} := \sup_{y>0} y^{s/2}e^{-y} = \Big(\frac{s}{2e}\Big)^{s/2}. \tag{5.36}$$
In particular, $\mathbb{E}[e^{-uX_e^2}] \lesssim u^{-s/2}$, with an explicit constant depending only on e and s.

Proof. Since

$$\pi^W(v_0) = \frac{W_{v_0}}{\sum_{u\in V} W_u} = \frac{1}{2}\sum_{g\ni v_0} X_g,$$

the event $\{\pi^W(v_0) < \varepsilon\}$ implies $\sum_{g\ni v_0} X_g < 2\varepsilon$. In particular, for every edge g incident to v_0, one has $X_g < 2\varepsilon$. Fix one such edge $g_0 \ni v_0$ that minimizes $\{C_{g,s} : g \ni v_0\}$. Then

$$\{\pi^W(v_0) < \varepsilon\} \subseteq \{X_{g_0} < 2\varepsilon\}.$$

Therefore, by Theorem 5.4.6,

$$\mathbb{P}(\pi^W(v_0) < \varepsilon) \le \mathbb{P}(X_{g_0} < 2\varepsilon) \le 2^s C_{g_0,s}\varepsilon^s.$$

For the Laplace estimate, let $Z > 0$ and $r > 0$. Since $e^{-y} \le C_r y^{-r}$ for all $y > 0$, where $C_r := \sup_{y>0} y^r e^{-y} = (r/e)^r$, one has $e^{-uZ} \le C_r u^{-r} Z^{-r}$. Taking $Z = X_e^2$ and $r = s/2$ gives (5.36). □

Step 2: quenched concentration. We now pass from tail bounds on the random environment to concentration estimates for the quenched walk. For a fixed environment w, let P^w be the quenched transition matrix, let π^w be its stationary distribution, and recall that $X_e(w) := w_e / \sum_{g\in E} w_g$ for $e \in E$. We also write

$$m(w) := \min_{e\in E} X_e(w).$$

For the spectral gap, let $\lambda_2(P^w)$ denote the second largest eigenvalue of P^w, let

$$\gamma(P^w) := 1 - \lambda_2(P^w). \tag{5.37}$$

Proposition 5.4.7.1 (Spectral gap and edge-count concentration)**.** *For every fixed environment w, the quenched chain P^w satisfies*

$$\gamma(P^w) \ge \frac{m(w)^2}{2}. \tag{5.38}$$

Consequently, for every $0 < s < \underline{a}/2$,

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}\big(\gamma(P^W) < \varepsilon\big) \le \Big(\sum_{e\in E} C_{e,s}\Big) (2\varepsilon)^{s/2}, \qquad 0 < \varepsilon < \frac{1}{2}, \tag{5.39}$$

where $W \sim \mu_{v_0,A}$ and the constants $C_{e,s}$ are those from Theorem 5.4.6.

Now fix an edge $e \in E$, and let

$$N_e(T) := \sum_{t=0}^{T-1} \mathbb{1}\big(\{X_t, X_{t+1}\} = e\big)$$

be its traversal count up to time T. If the chain starts from stationarity π^w, then for every $\eta > 0$ and every $T \ge 1$,

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\big(\big|N_e(T) - T X_e(w)\big| \ge \eta T\big) \le 2\exp\left(-\frac{\gamma(P^w)}{4}\eta^2 T\right) + 4\exp\left(-\frac{1}{8}\eta^2 T\right). \tag{5.40}$$

In particular, on the event $\{\gamma(P^w) \ge \varepsilon_1,\ X_e(w) \ge \varepsilon_2\}$, one has for every $0 < \delta \le 1$,

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\big(\big|N_e(T) - T X_e(w)\big| \ge \delta\, T X_e(w)\big) \le 2\exp\left(-\frac{\varepsilon_1\varepsilon_2^2}{4}\delta^2 T\right) + 4\exp\left(-\frac{\varepsilon_2^2}{8}\delta^2 T\right). \tag{5.41}$$

Proof. For the spectral gap bound, let $Q^w(i,j) := \pi^w(i) P^w_{ij}$. Since $\pi^w(i) = w_i / \sum_{u\in V} w_u$ and $P^w_{ij} = w_{ij}/w_i$, one has

$$Q^w(i,j) = \frac{w_{ij}}{\sum_{u\in V} w_u} = \frac{X_{\{i,j\}}(w)}{2}.$$

Hence, for every nonempty proper subset $S \subsetneq V$,

$$Q^w(S,S^c) := \sum_{i\in S,\ j\in S^c} Q^w(i,j) = \frac{1}{2}\sum_{e\in\partial S} X_e(w).$$

Here $Q^w(S,S^c)$ is the total stationary flow from S to its complement S^c, where $Q^w(i,j) := \pi_i^w P_{ij}^w$, and $\partial S := \{\{i,j\} \in E :\ i \in S,\ j \in S^c\}$ is the edge boundary of S. Because G is connected, $\partial S \neq \varnothing$, so $Q^w(S,S^c) \geq m(w)/2$. Therefore the conductance of the chain satisfies

$$\Phi(P^w) := \min_{\substack{S\subset V\\ 0<\pi^w(S)\leq 1/2}} \frac{Q^w(S,S^c)}{\pi^w(S)} \geq \min_{\emptyset\neq S\subsetneq V} 2Q^w(S,S^c) = \min_{\emptyset\neq S\subsetneq V} \sum_{e\in\partial S} X_e(w) \geq m(w).$$

Cheeger's inequality for finite reversible Markov chains states that

$$\gamma(P^w) \geq \frac{\Phi(P^w)^2}{2};$$

see, for example, [69, Theorem 13.14]. Together with the inequality $\Phi(P^w) \geq m(w)$, we obtain (5.38).

To obtain (5.39), note that

$$\{\gamma(P^W) < \varepsilon\} \subseteq \{m(W) < \sqrt{2\varepsilon}\} = \bigcup_{e\in E} \{X_e(W) < \sqrt{2\varepsilon}\}.$$

Hence, by the union bound and Theorem 5.4.6,

$$\mathbb{P}\big(\gamma(P^W) < \varepsilon\big) \leq \sum_{e\in E} \mathbb{P}\big(X_e(W) < \sqrt{2\varepsilon}\big) \leq \Big(\sum_{e\in E} C_{e,s}\Big)(2\varepsilon)^{s/2}.$$

We now turn to the concentration of $N_e(T)$ for fixed w. Define

$$Y_t := \mathbb{1}\big(\{X_t, X_{t+1}\} = e\big), \qquad f_e(i) := \mathbb{P}_{\mathrm{srw}}^{v_0,w}\big(\{X_t, X_{t+1}\} = e \mid X_t = i\big) = \sum_{j:\{i,j\}=e} P_{ij}^w.$$

Then $0 \leq f_e(i) \leq 1$, and, under stationarity,

$$\pi^w(f_e) := \sum_{i\in V} \pi^w(i) f_e(i) = \mathbb{P}_{\mathrm{srw}}^{\pi^w,w}\big(\{X_t, X_{t+1}\} = e\big) = X_e(w).$$

Therefore $\mathbb{E}_{\mathrm{srw}}^{\pi^w,w}[N_e(T)] = T X_e(w)$.

Write

$$A_T := \sum_{t=0}^{T-1}\big(f_e(X_t) - X_e(w)\big) \qquad \text{and} \qquad M_T := \sum_{t=0}^{T-1}\big(Y_t - f_e(X_t)\big).$$

Then $N_e(T) - TX_e(w) = A_T + M_T$.

To bound A_T, we apply Theorem 1 of Léon–Perron [66] to the stationary reversible chain (X_t) under $\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}$ and the function $f_e \colon V \to [0,1]$. Since

$$\mathbb{E}^{\pi^w,w}_{\mathrm{srw}}[f_e(X_0)] = X_e(w),$$

and, by stationarity, $\sum_{t=0}^{T-1} f_e(X_t)$ has the same law as $\sum_{t=1}^{T} f_e(X_t)$, Theorem 1 of [66], with $n = T$, $\mu = X_e(w)$, and $\varepsilon = \eta/2$ yields,

$$\begin{aligned}\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\Big(A_T \geq \frac{\eta T}{2}\Big) &= \mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\Big(\sum_{t=0}^{T-1} f_e(X_t) \geq T\Big(X_e(w) + \frac{\eta}{2}\Big)\Big) \\ &\leq \exp\Big(-\min\{1, \gamma(P^w)\}\, T\Big(\frac{\eta}{2}\Big)^2\Big).\end{aligned}$$

To control the lower tail, set $g_e := 1 - f_e \in [0,1]$. Then $\mathbb{E}_{\pi^w,w}[g_e(X_0)] = 1 - X_e(w)$, and

$$\Big\{A_T \leq -\frac{\eta T}{2}\Big\} = \Big\{\sum_{t=0}^{T-1} f_e(X_t) \leq T\Big(X_e(w) - \frac{\eta}{2}\Big)\Big\} = \Big\{\sum_{t=0}^{T-1} g_e(X_t) \geq T\Big(1 - X_e(w) + \frac{\eta}{2}\Big)\Big\}.$$

Therefore, we can apply the previous upper tail inequality to g_e, getting

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\Big(A_T \leq -\frac{\eta T}{2}\Big) \leq \exp\Big(-\min\{1, \gamma(P^w)\}\, T\Big(\frac{\eta}{2}\Big)^2\Big).$$

Therefore, by the union bound,

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\Big(|A_T| \geq \frac{\eta T}{2}\Big) \leq 2\exp\Big(-\frac{1}{4}\eta^2 T\Big) + 2\exp\Big(-\frac{1}{4}\eta^2\gamma(P^w)T\Big).$$

For M_T, note that it is a martingale with respect to $\mathcal{F}_t := \sigma(X_0, \dots, X_t)$, because $\mathbb{E}_w[Y_t \mid \mathcal{F}_t] = f_e(X_t)$. Its increments satisfy $|M_{t+1} - M_t| = |Y_t - f_e(X_t)| \leq 1$, so the Azuma–Hoeffding inequality [11, 58] gives

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}\Big(|M_T| \geq \frac{\eta T}{2}\Big) \leq 2\exp\Big(-\frac{1}{8}\eta^2 T\Big).$$

Since

$$\{|N_e(T) - TX_e(w)| \geq \eta T\} \subseteq \{|A_T| \geq \eta T/2\} \cup \{|M_T| \geq \eta T/2\},$$

we obtain (5.40).

Finally, substituting $\eta = \delta X_e(w)$ into (5.40) and using $\gamma(P^w) \geq \varepsilon_1, X_e(w) \geq \varepsilon_2$ yields (5.41). $\square$

To pass from a stationary start to a start from the deterministic initial state v_0, we use the following standard comparison estimate. The following lemma is inspired by [88, Proposition 3.15].

Lemma 5.4.8 (Transfer from a stationary start to a start from a deterministic root)**.** *Fix an environment w, and let π^w be the a stationary distribution of the quenched chain P^w. Then, for every event A measurable with respect to $(X_0, \dots, X_T)$,*

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A) \le (\pi^w(v_0))^{-1/2}\, \mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A)^{1/2}. \tag{5.42}$$

In particular, on the event $\{\pi^w(v_0) \ge \varepsilon_0\}$, one has

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A) \le \varepsilon_0^{-1/2}\, \mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A)^{1/2}. \tag{5.43}$$

Proof. We argue directly on path space. Let $\mathcal{X}_T := V^{T+1}$, and for $x_0, \dots, x_T \in V$, note

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T) = \mathbb{1}(x_0 = v_0) \prod_{t=0}^{T-1} p^w(x_t, x_{t+1})$$

for the quenched path law started from v_0, and

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(X_0^T = x_0^T) := \pi^w(x_0) \prod_{t=0}^{T-1} p^w(x_t, x_{t+1})$$

for the quenched path law started from stationarity π^w. Since $\pi^w(x) > 0$ for every $x \in V$, the first measure is absolutely continuous with respect to the second, and

$$\frac{\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)}{\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(X_0^T = x_0^T)} = \frac{\mathbb{1}(x_0 = v_0)}{\pi^w(v_0)}.$$

Therefore, for every event A measurable with respect to $(X_0, \dots, X_T)$,

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A) = \mathbb{E}^{\pi^w,w}_{\mathrm{srw}}\left[\frac{\mathbb{1}(X_0 = v_0)}{\pi^w(v_0)}\, \mathbb{1}_A\right].$$

Applying the Cauchy–Schwarz inequality, we have

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A) \le \left(\mathbb{E}^{\pi^w,w}_{\mathrm{srw}}\left[\left(\frac{\mathbb{1}(X_0 = v_0)}{\pi^w(v_0)}\right)^2\right]\right)^{1/2} \left(\mathbb{E}^{\pi^w,w}_{\mathrm{srw}}[\mathbb{1}_A]\right)^{1/2}.$$

Now under $\mathbb{P}_{\pi^w,w}$, one has $X_0 \sim \pi^w$, so

$$\mathbb{E}^{\pi^w,w}_{\mathrm{srw}}\left[\left(\frac{\mathbb{1}(X_0 = v_0)}{\pi^w(v_0)}\right)^2\right] = \sum_{x \in V} \pi^w(x) \left(\frac{\mathbb{1}(x = v_0)}{\pi^w(v_0)}\right)^2 = \frac{1}{\pi^w(v_0)},$$

while $\mathbb{E}^{\pi^w,w}_{\mathrm{srw}}[\mathbb{1}_A] = \mathbb{E}^{\pi^w,w}_{\mathrm{srw}}(A)$. Hence

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A) \le \pi^w(v_0)^{-1/2}\, \mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A)^{1/2},$$

which is (5.42). The bound (5.43) follows immediately on the event $\{\pi^w(v_0) \ge \varepsilon_0\}$. $\square$

Step 3: thresholded and refined inverse-local-time bounds. We now combine the environment tail bounds from Step 1 with the quenched concentration estimates from Step 2. The first estimate is obtained by imposing an explicit lower threshold on X_e; the second treats X_e directly through its negative moments and Laplace transform.

Theorem 5.4.9 (Thresholded and refined inverse-local-time bounds)**.** *Assume that $G = (V, E)$ is finite and connected, that the ERRW starts from v_0, and that the initial edge weights satisfy $0 < \underline{a} \le a_g \le \overline{a}$ for all $g \in E$. Fix an edge $e \in E$ and $\eta > 0$. Let $W \sim \mu_{v_0,A}$, let P^W be the quenched chain, let π^W be its stationary distribution, and write $X_e := W_e / \sum_{g\in E} W_g$. Also let*

$$N_e(T) := \sum_{t=0}^{T-1} \mathbb{1}\big(\{X_t, X_{t+1}\} = e\big)$$

be the traversal count of e up to time T.

(i) Let $0 < s < \underline{a}/2$, and define

$$C_{\mathrm{gap},s} := \sum_{h\in E} C_{h,s}, \qquad C_{v_0,s} := 2^s \min_{g\ni v_0} C_{g,s},$$

where the constants $C_{h,s}$ are those from Theorem 5.4.6. Then, for every $0 < \varepsilon_0 < 1/2$, $0 < \varepsilon_1 \le 1/2$, $0 < \varepsilon_2 < 1$, and every $T \ge 1$,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] \le \frac{2}{T\varepsilon_2} + \frac{1}{\eta}\Big[C_{e,s}\,\varepsilon_2^s + C_{\mathrm{gap},s}\,(2\varepsilon_1)^{s/2} + C_{v_0,s}\,\varepsilon_0^s\Big] + \frac{2}{\eta}\,\varepsilon_0^{-1/2}\left(e^{-\varepsilon_1\varepsilon_2^2 T/32} + e^{-\varepsilon_2^2 T/64}\right). \tag{5.44}$$

(ii) If in addition $1 < s < \underline{a}/2$, then for every $0 < \varepsilon_0 < 1/2$, $0 < \varepsilon_1 \le 1/2$, and every $T \ge 1$,

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] \le 2\,C_{e,s}\,T^{-1} + \frac{2^{s/2}}{\eta}\,C_{\mathrm{gap},s}\,\varepsilon_1^{s/2} + \frac{1}{\eta}\,C_{v_0,s}\,\varepsilon_0^s + \frac{2\cdot 32^{s/2}\,C_{s/2}\,C_{e,s}}{\eta}\,\varepsilon_0^{-1/2}(\varepsilon_1 T)^{-s/2} + \frac{2\cdot 64^{s/2}\,C_{s/2}\,C_{e,s}}{\eta}\,\varepsilon_0^{-1/2} T^{-s/2}. \tag{5.45}$$

Proof. We prove (i) and (ii) separately.
Proof of (i). Let

$$B := \{X_e < \varepsilon_2\} \cup \{\gamma(P^W) < \varepsilon_1\} \cup \{\pi^W(v_0) < \varepsilon_0\},$$

where we recall that $\gamma(P^W)$ is defined in (5.37). Splitting according to B, we write

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] = \mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_B\right] + \mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{B^c}\right].$$

On B, the trivial bound $(N_e(T)+\eta)^{-1} \le \eta^{-1}$ gives

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_B\right] \le \frac{1}{\eta}\,\mathbb{P}(B).$$

By the union bound together with Theorem 5.4.6, (5.39), and (5.35), we have

$$\mathbb{P}(B) \le C_{e,s}\,\varepsilon_2^s + C_{\mathrm{gap},s}\,(2\varepsilon_1)^{s/2} + C_{v_0,s}\,\varepsilon_0^s.$$

Hence

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_B\right] \le \frac{1}{\eta}\Big[C_{e,s}\,\varepsilon_2^s + C_{\mathrm{gap},s}\,(2\varepsilon_1)^{s/2} + C_{v_0,s}\,\varepsilon_0^s\Big].$$

We next estimate the contribution on B^c. Fix $w \in B^c$. Then $X_e(w) \ge \varepsilon_2$, $\gamma(P^w) \ge \varepsilon_1$, and $\pi^w(v_0) \ge \varepsilon_0$. Let

$$A_w := \left\{\big|N_e(T) - TX_e(w)\big| \ge \frac{1}{2}\,TX_e(w)\right\}.$$

Applying (5.41) with $\delta = 1/2$, we obtain

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A_w) \le 2\exp\!\big(-\varepsilon_1\varepsilon_2^2T/16\big) + 4\exp\!\big(-\varepsilon_2^2T/32\big)\,.$$

By Lemma 5.4.8,

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A_w) \le \pi^w(v_0)^{-1/2}\,\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A_w)^{1/2} \le 2\,\varepsilon_0^{-1/2}\left(e^{-\varepsilon_1\varepsilon_2^2T/32} + e^{-\varepsilon_2^2T/64}\right).$$

On A_w^c, one has $N_e(T) \ge \frac{1}{2}\,TX_e(w) \ge \frac{1}{2}\,T\varepsilon_2$, so $(N_e(T)+\eta)^{-1} \le 2/(T\varepsilon_2)$. On A_w, we again use $(N_e(T)+\eta)^{-1} \le \eta^{-1}$. Thus

$$\mathbb{E}^{v_0,w}_{\mathrm{srw}}\left[\frac{1}{N_e(T)+\eta}\right] \le \frac{2}{T\varepsilon_2} + \frac{1}{\eta}\,\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A_w).$$

Averaging this bound over all $w \in B^c$ gives

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{B^c}\right] \le \frac{2}{T\varepsilon_2} + \frac{2}{\eta}\,\varepsilon_0^{-1/2}\left(e^{-\varepsilon_1\varepsilon_2^2T/32} + e^{-\varepsilon_2^2T/64}\right).$$

Combining the estimates on B and B^c yields (5.44).

Proof of (ii). Let

$$B' := \{\gamma(P^W) < \varepsilon_1\} \cup \{\pi^W(v_0) < \varepsilon_0\}.$$

Again we split

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] = \mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{B'}\right] + \mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{(B')^c}\right].$$

On B', we use $(N_e(T)+\eta)^{-1} \le \eta^{-1}$, so

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{B'}\right] \le \frac{1}{\eta}\,\mathbb{P}^{v_0,A}_{\mathrm{errw}}(B').$$

By (5.39) and (5.35),

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(B') \le C_{\mathrm{gap},s}\,(2\varepsilon_1)^{s/2} + C_{v_0,s}\,\varepsilon_0^s = 2^{s/2}C_{\mathrm{gap},s}\,\varepsilon_1^{s/2} + C_{v_0,s}\,\varepsilon_0^s.$$

Hence

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{B'}\right] \le \frac{2^{s/2}}{\eta}C_{\mathrm{gap},s}\,\varepsilon_1^{s/2} + \frac{1}{\eta}C_{v_0,s}\,\varepsilon_0^s.$$

Fix now $w \in (B')^c$. Then $\gamma(P^w) \ge \varepsilon_1$ and $\pi^w(v_0) \ge \varepsilon_0$. Define

$$A_w := \left\{\big|N_e(T) - TX_e(w)\big| \ge \frac{1}{2}\,TX_e(w)\right\}.$$

Applying (5.40) with $\eta = X_e(w)/2$, and using $\gamma(P^w) \ge \varepsilon_1$, gives

$$\mathbb{P}^{\pi^w,w}_{\mathrm{srw}}(A_w) \le 2\exp\!\big(-\varepsilon_1 X_e(w)^2 T/16\big) + 4\exp\!\big(-X_e(w)^2 T/32\big)\,.$$

By Lemma 5.4.8,

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A_w) \le 2\,\varepsilon_0^{-1/2}\left(e^{-\varepsilon_1 X_e(w)^2 T/32} + e^{-X_e(w)^2 T/64}\right).$$

On A_w^c, one has $N_e(T) \ge \frac{1}{2}\,TX_e(w)$, so $(N_e(T)+\eta)^{-1} \le 2/(TX_e(w))$. Hence

$$\mathbb{E}^{v_0,w}_{\mathrm{srw}}\left[\frac{1}{N_e(T)+\eta}\right] \le \frac{2}{TX_e(w)} + \frac{1}{\eta}\,\mathbb{P}^{v_0,w}_{\mathrm{srw}}(A_w).$$

Averaging over $w \in (B')^c$, we obtain

$$\begin{aligned}\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{(B')^c}\right] &\le \frac{2}{T}\,\mathbb{E}^{v_0,A}_{\mathrm{errw}}[X_e^{-1}] \\ &\quad + \frac{2}{\eta}\,\varepsilon_0^{-1/2}\left(\mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{-\varepsilon_1 X_e^2 T/32}] + \mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{-X_e^2 T/64}]\right).\end{aligned}$$

Since $1 < s < \underline{a}/2$ and $0 < X_e \le 1$, one has $X_e^{-1} \le X_e^{-s}$, so

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}[X_e^{-1}] \le \mathbb{E}^{v_0,A}_{\mathrm{errw}}[X_e^{-s}] \le C_{e,s}$$

by Theorem 5.4.6. Also, by (5.36),

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{-\varepsilon_1 X_e^2 T/32}] \le C_{s/2}\,(\varepsilon_1 T/32)^{-s/2}\,\mathbb{E}^{v_0,A}_{\mathrm{errw}}[X_e^{-s}] \le 32^{s/2}C_{s/2}C_{e,s}\,(\varepsilon_1 T)^{-s/2},$$

and similarly

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}[e^{-X_e^2 T/64}] \le C_{s/2}\,(T/64)^{-s/2}\,\mathbb{E}^{v_0,A}_{\mathrm{errw}}[X_e^{-s}] \le 64^{s/2} C_{s/2} C_{e,s}\, T^{-s/2}.$$

Substituting these estimates gives

$$\begin{aligned}\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\,\mathbb{1}_{(B')^c}\right] &\le 2\,C_{e,s}\,T^{-1}\\ &\qquad + \frac{2\cdot 32^{s/2} C_{s/2} C_{e,s}}{\eta}\,\varepsilon_0^{-1/2}(\varepsilon_1 T)^{-s/2}\\ &\qquad + \frac{2\cdot 64^{s/2} C_{s/2} C_{e,s}}{\eta}\,\varepsilon_0^{-1/2} T^{-s/2}.\end{aligned}$$

Combining this with the estimate on B' yields (5.45). $\square$

Step 4: optimization. We now optimize the two estimates from Theorem 5.4.9 to obtain the final polynomial decay rate.

Theorem 5.4.10 (Inverse-local-time decay). *Assume that $G = (V, E)$ is finite and connected, that the ERRW starts from v_0, and that the initial edge weights satisfy $0 < \underline{a} \le a_g \le \overline{a}$ for all $g \in E$. Fix an edge $e \in E$ and $\eta > 0$.*

(i) If $0 < \underline{a} \le 4 + 2\sqrt{5}$, then for every

$$0 < r < \min\left\{\frac{\underline{a}}{8}, \frac{3}{4}\right\},$$

there exists a constant $C_r < \infty$, depending only on G, v_0, A, e, η, r, such that

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] \le C_r\,T^{-r}, \qquad T \ge 1.$$

(ii) If $\underline{a} > 4 + 2\sqrt{5}$, then there exists a constant $C < \infty$, depending only on G, v_0, A, e, η, such that

$$\mathbb{E}^{v_0,A}_{\mathrm{errw}}\left[\frac{1}{N_e(T)+\eta}\right] \le \frac{C}{T}, \qquad T \ge 1.$$

Proof. Set $E_T := \mathbb{E}^{v_0,A}_{\mathrm{errw}}[(N_e(T)+\eta)^{-1}]$.
Proof of (i). Assume $0 < \underline{a} \le 4 + 2\sqrt{5}$, and fix any constant r such that

$$0 < r < \min\left\{\frac{\underline{a}}{8}, \frac{3}{4}\right\}.$$

Choose any constant s so that $4r < s < \underline{a}/2$. This is possible because $r < \underline{a}/8$ implies $4r < \underline{a}/2$.

Next choose any constant $\kappa > 0$ so that

$$\kappa < \frac{1}{2}, \qquad \kappa < \frac{3}{4} - r, \qquad \kappa < \frac{2}{s}\Big(\frac{s}{4} - r\Big).$$

This is possible because $r < 3/4$ and $r < s/4$. With this choice, $r < 3/4 - \kappa$ and $r < s/4 - s\kappa/2$.

Apply Theorem 5.4.9(i) with

$$\varepsilon_0 = T^{-1}, \qquad \varepsilon_1 = T^{-1/2+\kappa}, \qquad \varepsilon_2 = T^{-1/4}.$$

Then $\frac{2}{T\varepsilon_2} = 2T^{-3/4}$, $\varepsilon_2^s = T^{-s/4}$, $\varepsilon_1^{s/2} = T^{-s/4+s\kappa/2}$, and $\varepsilon_0^s = T^{-s}$. Moreover,

$$\varepsilon_1\varepsilon_2^2 T = T^{-1/2+\kappa} \cdot T^{-1/2} \cdot T = T^{\kappa}, \qquad \varepsilon_2^2 T = T^{-1/2} \cdot T = T^{1/2},$$

and $\varepsilon_0^{-1/2} = T^{1/2}$. Hence (5.44) gives

$$E_T \le C\Big(T^{-3/4} + T^{-s/4} + T^{-s/4+s\kappa/2} + T^{-s} + T^{1/2}e^{-cT^{\kappa}} + T^{1/2}e^{-cT^{1/2}}\Big)$$

for constants $C, c > 0$ depending only on G, v_0, A, e, η, r.

We now compare the exponents with r. By construction, $3/4 > r$, $s/4 > r$, $s/4 - s\kappa/2 > r$, and $s > r$, so each polynomial term on the right-hand side is $O(T^{-r})$.

It remains to control the exponential terms. We use the elementary inequality that for every $c > 0$ and every $\beta > 0$,

$$\sup_{u\ge 0} u^{\beta} e^{-cu} < \infty.$$

Indeed, the function $\phi(u) := u^{\beta}e^{-cu}$ is continuous on $[0, \infty)$, tends to 0 as $u \to \infty$, and for $u > 0$ satisfies $\phi'(u) = u^{\beta-1}e^{-cu}(\beta - cu)$. Hence its maximum is attained at $u = \beta/c$, and therefore

$$u^{\beta}e^{-cu} \le \left(\frac{\beta}{ce}\right)^{\beta}, \qquad u \ge 0.$$

Applying this with $u = T^{\kappa}$ and $\beta = (r + \frac{1}{2})/\kappa$, we obtain

$$T^{r+1/2}e^{-cT^{\kappa}} = (T^{\kappa})^{(r+1/2)/\kappa}e^{-cT^{\kappa}} \le \left(\frac{(r+1/2)/\kappa}{ce}\right)^{(r+1/2)/\kappa}.$$

Thus $T^{1/2}e^{-cT^{\kappa}} \le C_{r,\kappa,c}T^{-r}$ for all $T \ge 1$.

Similarly, applying the same inequality with $u = T^{1/2}$ and $\beta = 2r + 1$, we get

$$T^{r+1/2}e^{-cT^{1/2}} = (T^{1/2})^{2r+1}e^{-cT^{1/2}} \le \left(\frac{2r+1}{ce}\right)^{2r+1},$$

and hence $T^{1/2}e^{-cT^{1/2}} \le C'_{r,c}T^{-r}$ for all $T \ge 1$.

Therefore every term on the right-hand side is $O(T^{-r})$, and hence

$$E_T \le C_r T^{-r}, \qquad T \ge 1.$$

Proof of (ii). Assume $\underline{a} > 4 + 2\sqrt{5}$. Choose s so that

$$2 + \sqrt{5} < s < \frac{\underline{a}}{2}.$$

This is possible because $\underline{a} > 4 + 2\sqrt{5}$ implies $\underline{a}/2 > 2 + \sqrt{5}$.

Apply Theorem 5.4.9(ii) with

$$\varepsilon_0 = T^{-1/s}, \qquad \varepsilon_1 = T^{-2/s}.$$

Then $\varepsilon_1^{s/2} = (T^{-2/s})^{s/2} = T^{-1}$ and $\varepsilon_0^s = (T^{-1/s})^s = T^{-1}$. Also,

$$\varepsilon_0^{-1/2}(\varepsilon_1 T)^{-s/2} = T^{1/(2s)}\big(T^{-2/s}T\big)^{-s/2} = T^{1/(2s)}\,T^{-s/2+1} = T^{-(\,s/2-1-1/(2s)\,)},$$

and similarly

$$\varepsilon_0^{-1/2}T^{-s/2} = T^{1/(2s)}T^{-s/2} = T^{-(\,s/2-1/(2s)\,)}.$$

We now check that both exponents are strictly larger than 1 in absolute value. First,

$$\frac{s}{2} - 1 - \frac{1}{2s} > 1 \iff s - \frac{1}{s} > 4 \iff s^2 - 4s - 1 > 0.$$

The positive root of $s^2 - 4s - 1 = 0$ is $2 + \sqrt{5}$, so this holds because $s > 2 + \sqrt{5}$. Second,

$$\frac{s}{2} - \frac{1}{2s} > 1 \iff s - \frac{1}{s} > 2 \iff s^2 - 2s - 1 > 0,$$

whose positive root is $1 + \sqrt{2}$; this also holds because $s > 2 + \sqrt{5}$.

Therefore both terms $\varepsilon_0^{-1/2}(\varepsilon_1 T)^{-s/2}$ and $\varepsilon_0^{-1/2}T^{-s/2}$ are $O(T^{-1-\delta})$ for some $\delta > 0$, and in particular are $O(T^{-1})$. Since the first three terms in (5.45) are already $O(T^{-1})$, it follows that every term on the right-hand side of (5.45) is $O(T^{-1})$. Hence

$$E_T \le \frac{C}{T}, \qquad T \ge 1,$$

for a finite constant C depending only on G, v_0, A, e, η. □

Corollary 5.4.11 (Decay of the trajectory-level KL gap)**.** *Let $A^{(0)}, A^{(1)} \in (0,\infty)^E$, and let*

$$\mu_s := \mu_{v_0,A^{(s)}}, \qquad P_s^{(T)} := \mathcal{L}_{\mathbb{P}_{\mathrm{errw}}^{v_0,A^{(s)}}}(X_0^T), \qquad s \in \{0,1\}.$$

Define

$$\mathrm{Gap}_T := D_{\mathrm{KL}}(\mu_0\|\mu_1) - D_{\mathrm{KL}}(P_0^{(T)}\|P_1^{(T)}), \quad \delta_e := a_e^{(1)} - a_e^{(0)}, \quad \underline{a} := \min_{e\in E}\min\{a_e^{(0)}, a_e^{(1)}\} > 0.$$

Then the following hold.

(i) If $0 < \underline{a} \le 4 + 2\sqrt{5}$, then for every

$$0 < r < \min\left\{\frac{\underline{a}}{8}, \frac{3}{4}\right\},$$

there exists a constant $C_r < \infty$, depending only on $G, v_0, A^{(0)}, A^{(1)}, r$, such that

$$\mathrm{Gap}_T \le C_r\, T^{-r}, \qquad T \ge 1.$$

(ii) If $\underline{a} > 4+2\sqrt{5}$, then there exists a constant $C < \infty$, depending only on $G, v_0, A^{(0)}, A^{(1)}$, such that

$$\mathrm{Gap}_T \le \frac{C}{T}, \qquad T \ge 1.$$

Proof. By Proposition 5.4.1.1,

$$\mathrm{Gap}_T \le \frac{1}{2}\Big(1 + \frac{1}{\underline{a}}\Big) \sum_{e\in E} \delta_e^2\, \mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e(T) + \underline{a}}\right].$$

Here the expectation is taken under the ERRW law with initial weights $A^{(0)}$. Since E is finite and $A^{(0)} \in (0,\infty)^E$, we have

$$0 < \underline{a} \le a_e^{(0)} \le \overline{a}_0 := \max_{e\in E} a_e^{(0)} < \infty, \qquad e \in E.$$

Thus Theorem 5.4.10 applies to ERRW with initial weight vector $A^{(0)}$ and $\eta = \underline{a}$.

Assume first that $0 < \underline{a} \le 4 + 2\sqrt{5}$, and fix

$$0 < r < \min\left\{\frac{\underline{a}}{8}, \frac{3}{4}\right\}.$$

For each edge $e \in E$, Theorem 5.4.10(i) yields a finite constant $C'_{r,e}$, depending only on $G, v_0, A^{(0)}, e, \underline{a}, r$, such that

$$\mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e(T) + \underline{a}}\right] \le C'_{r,e}\, T^{-r}, \qquad T \ge 1.$$

Since E is finite, we may define

$$C'_r := \max_{e\in E} C'_{r,e} < \infty.$$

Then

$$\mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e(T) + \underline{a}}\right] \le C'_r T^{-r}, \qquad T \ge 1,$$

uniformly over $e \in E$. Substituting this into the preceding bound gives

$$\mathrm{Gap}_T \le \frac{1}{2}\Big(1 + \frac{1}{\underline{a}}\Big)\Big(\sum_{e\in E} \delta_e^2\Big)\, C_r' T^{-r},$$

which proves part (i).

Now assume that $\underline{a} > 4 + 2\sqrt{5}$. For each edge $e \in E$, Theorem 5.4.10(ii) yields a finite constant C_e'', depending only on $G, v_0, A^{(0)}, e, \underline{a}$, such that

$$\mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e(T) + \underline{a}}\right] \le C_e''\, T^{-1}, \qquad T \ge 1.$$

Again, since E is finite, we may define

$$C'' := \max_{e\in E} C_e'' < \infty,$$

so that

$$\mathbb{E}_{P_0^{(T)}}\left[\frac{1}{N_e(T) + \underline{a}}\right] \le C''\, T^{-1}, \qquad T \ge 1,$$

uniformly over $e \in E$. Inserting this into the same bound from Proposition 5.4.1.1 yields

$$\mathrm{Gap}_T \le \frac{1}{2}\Big(1 + \frac{1}{\underline{a}}\Big)\Big(\sum_{e\in E} \delta_e^2\Big)\, C''\, T^{-1},$$

which proves part (ii). □

5.5 Summary

This chapter studies several information-theoretic quantities for ERRW on finite graphs, including the entropy rate, the KL divergence between environment laws, and the KL divergence between finite-trajectory laws. The random-environment representation makes these quantities accessible in a rather explicit form, and in particular reduces the convergence of trajectory-level KL divergence toward environment-level KL divergence to the control of inverse local times.

A first natural open problem is to obtain a sharp characterization of this convergence in the general graph case. In the n-star, the Pólya-urn structure yields the precise rates T^{-1}, $(\log T)T^{-1}$, and T^{-a}. By contrast, for general finite graphs, our argument gives polynomial upper bounds, but these are unlikely to be optimal for $\underline{a} < 4 + 2\sqrt{5}$. It would be very interesting to identify the correct rate, and more broadly to understand how it depends on the graph geometry and the initial weights.

A second, and perhaps more important, direction is to use these information-theoretic results to derive quantitative upper and lower bounds for testing problems involving ERRWs.

The most immediate example is two-point testing, where the trajectory-level KL divergence governs the Stein exponent. More generally, one may ask about identity testing, closeness testing, and related minimax testing problems. From this perspective, the quantities studied here should be viewed not only as intrinsic objects of interest, but also as tools for proving information-theoretic lower bounds and for designing statistically efficient tests.

Chapter 6

True Self-Avoiding Walk for Accelerating MCMC Estimation

This chapter studies a TSAW-based mechanism for improving empirical integral estimation in Markov chain Monte Carlo. Given a finite irreducible Markov kernel P with stationary distribution π, we construct an adaptive walk whose transition probabilities penalize empirical overuse of directed transitions relative to their target frequencies.

The main result is an almost-sure pathwise discrepancy bound. If $L_t(i)$ denotes the occupation count of state i, and $N_t(i,j)$ denotes the directed transition count from i to j, then

$$L_t(i) - t\pi_i = O(\sqrt{\log t}), \qquad N_t(i,j) - t\pi_i P_{ij} = O(\sqrt{\log t})$$

almost surely for every state i and every support edge (i,j). Consequently, for every bounded $f : V \to \mathbb{R}$,

$$\left| \frac{1}{t} \sum_{s=0}^{t-1} f(X_s) - \sum_{i \in V} \pi_i f(i) \right| = O\left(\frac{\sqrt{\log t}}{t} \right) \qquad \text{almost surely.}$$

6.1 Introduction

Let V be a finite state space, let P be an irreducible Markov kernel on V with stationary distribution π, and let $f : V \to \mathbb{R}$ be bounded. Writing the vertex local time as $L_t(i) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}$, the empirical integral error decomposes as

$$\frac{1}{t} \sum_{s=0}^{t-1} f(X_s) - \pi(f) = \frac{1}{t} \sum_{i \in V} f(i) \big(L_t(i) - t\pi_i \big),$$

so empirical estimation is directly controlled by the discrepancy between the occupation counts $L_t(i)$ and their targets $t\pi_i$.

A substantial literature addresses occupation imbalance by modifying the base chain to avoid backtracking or break reversibility, and through nonlinear self-interacting dynamics;

see Section 1.3 for a review. The closest prior work is Doshi, Hu, and Eun [40], who develop self-repellent random walks on finite graphs, prove almost sure convergence of the empirical distribution, and establish a CLT whose asymptotic variance decreases with the repulsion strength. Their estimator achieves the usual $t^{-1/2}$ scale. The improvement here is not only a variance-constant effect but a pathwise discrepancy bound at the sharper $O(\sqrt{\log t}/t)$ scale.

We design a TSAW-based adaptive dynamics in which the empirical transition counts are balanced row-wise against the target probabilities P_{ij}, giving

$$L_t(i) - t\pi_i = O(\sqrt{\log t}), \qquad N_t(i,j) - t\pi_i P_{ij} = O(\sqrt{\log t})$$

almost surely for every state i and every support edge (i, j), and hence an $O(\sqrt{\log t}/t)$ almost-sure bound for empirical averages.

The argument proceeds in three steps. We begin with a warm-up on the star graph, where the TSAW dynamics admits an exponential-race embedding that allows detailed analysis of the cover time. We then study a finite-alphabet self-balancing model and construct a Lyapunov function for the occupation-deficit process, obtaining uniform exponential moment bounds by a drift argument. Finally, we apply the same balancing mechanism row-wise to the transition counts of a general finite irreducible kernel, and combine the resulting row-wise estimates with a Poisson equation for the occupation vector to derive the global almost-sure bound.

Related Work

The closest prior work is Doshi, Hu, and Eun [40], who study self-repellent random walks on finite graphs as nonlinear Markov chains targeting a prescribed stationary distribution. They prove almost sure convergence of the empirical distribution and a central limit theorem for empirical averages, with asymptotic variance decreasing as the repulsion strength increases. In contrast, our result gives an explicit almost-sure $O(\sqrt{\log t}/t)$ estimator error bound through pathwise control of occupation and transition discrepancies.

True self-avoiding walk and weakly self-avoiding walk have a much longer history in probability theory and statistical physics. Classical works such as [8, 101, 102, 57] study self-repelling walks mainly through path properties, including scaling limits, recurrence, and self-repulsion on lattice-type graphs. These works provide the probabilistic background for TSAW, but they are not formulated in terms of empirical integral estimation for a prescribed target distribution on a finite state space. Our use of TSAW is different in purpose: the self-avoidance mechanism is introduced here in order to control occupation imbalance relative to a target law.

There is also a substantial literature showing that non-backtracking or nonreversible modifications can improve the efficiency of random-walk-based Monte Carlo methods; see, for example, [32, 85, 7, 65, 104, 24, 74, 100]. These methods typically improve performance by preventing immediate reversals or by introducing irreversible flow into the dynamics. Our setting is different: the transition rule depends on empirical usage accumulated along the trajectory, and the main quantity of interest is the resulting occupation and transition discrepancy. In this sense, the present paper is closer to trajectory-dependent self-avoidance than to nonreversible acceleration.

From a broader perspective, the model studied here belongs naturally to the theory of nonlinear and self-interacting Markov chains. In this direction, Del Moral and Miclo [28, 29], Andrieu, Jasra, Doucet, and Del Moral [9, 10], Del Moral and Doucet [27], and Fort and coauthors [55, 54] develop general frameworks for Markov dynamics whose transition mechanism depends on occupation statistics or interacting empirical measures. Our contribution fits into this history-dependent framework, but focuses on a concrete TSAW-based construction for which one can prove explicit almost-sure bounds on vertex and edge discrepancies.

Taken together, these works place the present paper at the intersection of self-avoiding random walks, nonlinear Markov chains, and MCMC variance reduction. Relative to the existing literature, the main novelty here is the explicit finite-time pathwise control of empirical occupation and transition errors, and the resulting almost-sure bound of order $O(\sqrt{\log t}/t)$ for empirical integral estimation.

The remainder of the chapter is organized as follows. Section 6.2 studies the star-graph warm-up. Section 6.3 develops the self-balancing framework, first in the finite-alphabet setting and then for general finite irreducible Markov kernels.

6.2 A warm-up: cover time of TSAW on the star graph

We begin with a warm-up model that isolates the effect of self-avoidance on exploration. Let G_n be the star graph with hub 0 and leaves $1, \ldots, n$. Under TSAW, every jump from the hub to a leaf is followed by an immediate return to the hub, so the dynamics is completely determined by the sequence of hub departures. It is therefore natural to study the reduced process obtained by recording only the successive choices of leaves at the hub.

Fix $\rho \in (0, 1)$. For $m \geq 0$ and $i \in [n] := \{1, \ldots, n\}$, let

$$C_m(i) := \sum_{r=1}^{m} \mathbf{1}\{I_r = i\}$$

be the number of times leaf i has been chosen in the first m hub departures. Here $C_0(i) = 0$ for all i. We define the reduced TSAW dynamics by

$$\mathbb{P}(I_{m+1} = i \mid I_1, \ldots, I_m) = \frac{\rho^{C_m(i)}}{\sum_{j=1}^{n} \rho^{C_m(j)}}, \qquad i \in [n], \tag{6.1}$$

with I_1 being chosen uniformly at random. This is exactly the hub-departure chain induced by TSAW on the star graph, after absorbing the deterministic factor coming from the return step into the parameter ρ. Let

$$\tau_{\text{cov}}^{(n)} := \inf\bigl\{m \geq 0 : \; C_m(i) \geq 1 \text{ for all } i \in [n]\bigr\}$$

be the cover time, measured in hub departures. The usual discrete-time cover time differs only by a factor of 2.

For random variables Y_n and a deterministic constant y, we write $Y_n \xrightarrow{\mathbb{P}} y$ if for every $\varepsilon > 0$,

$$\mathbb{P}(|Y_n - y| > \varepsilon) \to 0 \qquad \text{as } n \to \infty.$$

We also write $R_n = o_{\mathbb{P}}(1)$ if $R_n \xrightarrow{\mathbb{P}} 0$.

Our main result in this section is the following.

Theorem 6.2.1 (cover time of TSAW on the star graph)**.** *Fix $\rho \in (0, 1)$ and set $c_\rho := \frac{1}{\log(1/\rho)}$. Let $\tau_{\mathrm{cov}}^{(n)}$ be the cover time, measured in hub departures, of the TSAW dynamics on the star graph with n leaves and effective discount parameter ρ. Then for every $\varepsilon > 0$ and every $\delta > 0$, there exists $n_0 = n_0(\varepsilon, \delta, \rho)$ such that for all $n \geq n_0$,*

$$\mathbb{P}\left(\left|\frac{\tau_{\mathrm{cov}}^{(n)}}{n \log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1 - \delta.$$

The proof is based on an exponential-race embedding. For each leaf $i \in [n]$, let $\{E_{i,k} : k \geq 0\}$ be independent random variables with

$$E_{i,k} \sim \mathrm{Exp}(\rho^k), \qquad i \in [n],\ k \geq 0,$$

and assume that these families are mutually independent across i. Define the arrival times

$$S_{i,r} := \sum_{k=0}^{r-1} E_{i,k}, \qquad r \geq 1, \qquad S_{i,0} := 0,$$

and the associated birth processes

$$N_i(s) := \max\{r \geq 0 :\ S_{i,r} \leq s\}, \qquad s \geq 0.$$

Let

$$K(s) := \sum_{i=1}^{n} N_i(s)$$

be the total number of births by time s, and let

$$S_{\max} := \max_{1 \leq i \leq n} S_{i,1} = \max_{1 \leq i \leq n} E_{i,0}$$

be the time at which the last leaf receives its first birth.

The first lemma shows that this continuous-time system reproduces the discrete TSAW dynamics exactly.

Lemma 6.2.2 (exponential-race embedding)**.** *The reduced TSAW process* (6.1) *can be coupled with the birth processes* $\{N_i(s)\}_{i=1}^n$ *so that the ordered sequence of births in the merged timeline* $\{(i,r) : S_{i,r}\}_{i\in[n],\,r\geq 1}$ *has the same law as the sequence of hub departures* $(I_m)_{m\geq 1}$*. Under this coupling,* $\tau_{\mathrm{cov}}^{(n)} = K(S_{\max})$ *almost surely, where* $S_{\max} = \max_{1\leq i\leq n} S_{i,1}$*. For every* $\varepsilon \in (0,1)$*, if* $C_\varepsilon := \log(2/\varepsilon)$*, then for all sufficiently large* n*,*

$$\mathbb{P}\big(|S_{\max} - \log n| \leq C_\varepsilon\big) \geq 1 - \varepsilon.$$

On the event $\{|S_{\max} - \log n| \leq C_\varepsilon\}$*, we have*

$$K(\log n - C_\varepsilon) \leq \tau_{\mathrm{cov}}^{(n)} \leq K(\log n + C_\varepsilon).$$

Proof. Suppose that after m births the current birth counts are $(c_1, \ldots, c_n)$. By the memoryless property of the exponential distribution, the residual waiting time to the next birth in process i is $\mathrm{Exp}(\rho^{c_i})$, independently across i. Therefore the next birth occurs in process i with probability $\rho^{c_i}/\sum_{j=1}^n \rho^{c_j}$, which is exactly the transition rule (6.1). Iterating proves that the merged birth order reproduces the reduced TSAW dynamics.

For each $i \in [n]$, the birth process N_i is non-explosive. Indeed, since $\rho^k \leq 1$ for all $k \geq 0$, each waiting time $E_{i,k} \sim \mathrm{Exp}(\rho^k)$ stochastically dominates an $\mathrm{Exp}(1)$ random variable. Therefore the partial sums

$$S_{i,r} = \sum_{k=0}^{r-1} E_{i,k}$$

tend to $+\infty$ almost surely as $r \to \infty$, and hence $N_i(s) < \infty$ almost surely for every finite $s \geq 0$. Since $S_{\max} = \max_{1\leq i\leq n} E_{i,0}$ is the maximum of finitely many exponential random variables, we also have $S_{\max} < \infty$ almost surely. It follows that

$$K(S_{\max}) = \sum_{i=1}^n N_i(S_{\max}) < \infty \qquad \text{almost surely.}$$

Since $S_{\max} = \max_i S_{i,1}$ is the time by which every process has had its first birth, the number of births by time $S_{\max}$ is exactly the number of hub departures required to visit every leaf at least once. Hence $\tau_{\mathrm{cov}}^{(n)} = K(S_{\max})$ almost surely. Because $K(\cdot)$ is nondecreasing, the sandwich bound follows immediately.

To control $S_{\max}$, note that $E_{1,0}, \ldots, E_{n,0}$ are i.i.d. $\mathrm{Exp}(1)$, so $\mathbb{P}(S_{\max} \leq y) = (1 - e^{-y})^n$ for $y \geq 0$. If $x \in [0, \log n]$, then

$$\mathbb{P}(S_{\max} \leq \log n - x) = \left(1 - \frac{e^x}{n}\right)^n \leq e^{-e^x},$$

while

$$\mathbb{P}(S_{\max} \geq \log n + x) = 1 - \left(1 - \frac{e^{-x}}{n}\right)^n \leq e^{-x}.$$

Here we used the basic inequalities $1 - u \leq e^{-u}$ and $1 - (1-u)^n \leq nu$ for $u \in [0,1]$. Thus $\mathbb{P}(|S_{\max} - \log n| > x) \leq e^{-x} + e^{-e^x}$. Taking $x = C_\varepsilon = \log(2/\varepsilon)$ gives $e^{-x} + e^{-e^x} \leq \varepsilon/2 + e^{-2/\varepsilon} \leq \varepsilon$ (note that $ue^{-u} \leq 1$ for $u > 0$), and the claim follows. $\square$

Therefore, to bound the cover time it suffices to study $K(s_n)$ for some deterministic $s_n = \log n + O(1)$. And furthermore, to study $K(s_n) = \sum_{i=1}^n N_i(s_n)$, which is a sum of i.i.d. random variables $N_i(s_n)$'s, it suffices to study the individual distributions. This is reflected in the following lemmas.

Lemma 6.2.3 (birth process with exponentially decaying rates)**.** *Fix $\rho \in (0,1)$ and, for $s \geq 1$, let $N(s)$ denote a generic copy of $N_i(s)$. Define*

$$i_*(s) := \lfloor \log_{1/\rho} s \rfloor + 1, \qquad \text{so that} \qquad \rho^{i_*(s)-1} s \in [1, 1/\rho].$$

Set

$$C_1(\rho) := e^{1/\rho}, \qquad C_2(\rho) := \prod_{m=0}^{\infty} \frac{1}{1 - \frac{1}{2}\rho^m},$$

$$A_\rho := 1 + \frac{C_1(\rho)}{1-\sqrt{\rho}} + \frac{2C_2(\rho)}{1-\rho},$$

and

$$B_\rho := 2\left(C_1(\rho)\frac{\sqrt{\rho}(1+\sqrt{\rho})}{(1-\sqrt{\rho})^2} + 2C_2(\rho)\frac{1+\rho}{(1-\rho)^2}\right) + 2\left(\frac{C_1(\rho)}{1-\sqrt{\rho}} + \frac{2C_2(\rho)}{1-\rho}\right)^2.$$

Then, for every $s \geq 1$, the following hold:

(i) *For every integer $k \geq 0$,*

$$\mathbb{P}\big(N(s) \geq i_*(s) + k\big) \leq C_1(\rho)\, \rho^{k(k+1)/2} \leq C_1(\rho)\, \rho^{k/2}.$$

(ii) *For every integer $k \geq 1$,*

$$\mathbb{P}\big(N(s) \leq i_*(s) - k\big) \leq C_2(\rho) \exp\big(-\tfrac{1}{2}\rho^{-k+1}\big),$$

where the event is understood to be empty when $i_*(s) - k < 0$.

(iii) *One has*

$$\left|\mathbb{E}[N(s)] - \frac{\log s}{\log(1/\rho)}\right| \leq A_\rho, \qquad \mathrm{Var}(N(s)) \leq B_\rho.$$

This lemma can be proved using moment generating functions. We defer this proof to Appendix D.1.

Lemma 6.2.4 (concentration of the total number of departures)**.** *Fix $\rho \in (0,1)$, set $c_\rho := 1/\log(1/\rho)$, and let $c \in \mathbb{R}$ be fixed. For each n sufficiently large, define $s_n := \log n + c$. Then for every $\varepsilon > 0$ and every $\delta > 0$, there exists $n_0 = n_0(\varepsilon, \delta, \rho, c)$ such that for all $n \geq n_0$,*

$$\mathbb{P}\left(\left|\frac{K(s_n)}{n \log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1 - \delta.$$

The proof of this lemma uses the fact that for any deterministic sequence $(s_n)_{n\geq 1}$, $K(s_n)$ is the sum of n i.i.d. random variables. The proof of this lemma can be found in Appendix D.2.

Proof of Theorem 6.2.1. Fix $\varepsilon > 0$ and $\delta > 0$. We must prove that, for all sufficiently large n,

$$\mathbb{P}\left(\left|\frac{\tau_{\text{cov}}^{(n)}}{n\log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1-\delta.$$

Set $\delta_1 = \delta_2 = \delta_3 = \delta/3$ and let $C := \log(2/\delta_1)$. By Lemma 6.2.2, there exists n_1 such that for all $n \geq n_1$, the event $E_n := \{|S_{\max} - \log n| \leq C\}$ satisfies $\mathbb{P}(E_n) \geq 1-\delta_1$. On E_n, we have the deterministic sandwich

$$K(\log n - C) \leq \tau_{\text{cov}}^{(n)} \leq K(\log n + C).$$

Now define $s_n^- := \log n - C$ and $s_n^+ := \log n + C$. Since C is fixed, both sequences satisfy $s_n^\pm = \log n + O(1)$ and $s_n^\pm \geq 1$ for all sufficiently large n. Applying Lemma 6.2.4 to (s_n^-) with parameters (ε, δ_2), we obtain some n_2 such that for all $n \geq n_2$,

$$\mathbb{P}\left(\left|\frac{K(s_n^-)}{n\log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1-\delta_2.$$

Likewise, applying the same proposition to (s_n^+) with parameters (ε, δ_3), there exists n_3 such that for all $n \geq n_3$,

$$\mathbb{P}\left(\left|\frac{K(s_n^+)}{n\log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1-\delta_3.$$

For $n \geq \max\{n_1, n_2, n_3\}$, let

$$A_n^- := \left\{\left|\frac{K(s_n^-)}{n\log\log n} - c_\rho\right| \leq \varepsilon\right\} \quad \text{and} \quad A_n^+ := \left\{\left|\frac{K(s_n^+)}{n\log\log n} - c_\rho\right| \leq \varepsilon\right\}.$$

Then $\mathbb{P}(E_n) \geq 1-\delta_1$, $\mathbb{P}(A_n^-) \geq 1-\delta_2$, and $\mathbb{P}(A_n^+) \geq 1-\delta_3$.

Consider the good event $G_n := E_n \cap A_n^- \cap A_n^+$. On G_n, the inequalities defining $A_n^\pm$ give

$$(c_\rho - \varepsilon)n\log\log n \leq K(s_n^-) \leq (c_\rho + \varepsilon)n\log\log n$$

and

$$(c_\rho - \varepsilon)n\log\log n \leq K(s_n^+) \leq (c_\rho + \varepsilon)n\log\log n.$$

Since E_n also implies $K(s_n^-) \leq \tau_{\text{cov}}^{(n)} \leq K(s_n^+)$, we deduce that on G_n,

$$(c_\rho - \varepsilon)n\log\log n \leq \tau_{\text{cov}}^{(n)} \leq (c_\rho + \varepsilon)n\log\log n.$$

Equivalently,

$$\left|\frac{\tau_{\text{cov}}^{(n)}}{n\log\log n} - c_\rho\right| \leq \varepsilon.$$

Thus

$$G_n \subseteq \left\{ \left| \frac{\tau_{\mathrm{cov}}^{(n)}}{n \log \log n} - c_\rho \right| \leq \varepsilon \right\}.$$

Finally, by the union bound,

$$\mathbb{P}(G_n) \geq 1 - \mathbb{P}(E_n^c) - \mathbb{P}((A_n^-)^c) - \mathbb{P}((A_n^+)^c) \geq 1 - \delta_1 - \delta_2 - \delta_3 = 1 - \delta.$$

Therefore, for all $n \geq \max\{n_1, n_2, n_3\}$,

$$\mathbb{P}\left(\left| \frac{\tau_{\mathrm{cov}}^{(n)}}{n \log \log n} - c_\rho \right| \leq \varepsilon \right) \geq 1 - \delta.$$

This proves the theorem. □

Theorem 6.2.1 shows that even in this simple model, exponential self-avoidance changes the exploration scale from the classical coupon-collector order $n \log n$ to the much smaller order $n \log \log n$. The remainder of the paper turns from cover time to the main theme of empirical integral estimation, where a related self-avoidance mechanism yields explicit control of occupation and transition discrepancies.

6.3 TSAW-Based Balancing for MCMC Estimation

Before discussing how TSAW transition rules can be used in a random walk setting, we first analyze the discrete distribution scenario. This will be the building block of a non-Markovian random walk we use to accelerate the MCMC estimation.

Fix $d \geq 2$, a target distribution $p = (p_1, \ldots, p_d)$ satisfying

$$p_i > 0, \qquad \sum_{i=1}^{d} p_i = 1, \qquad p_{\min} := \min_{i \in [d]} p_i > 0,$$

and a parameter $\lambda > 0$.

Let $(X_t)_{t \geq 1}$ be an $[d] := \{1, \ldots, d\}$-valued process adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t := \sigma(X_1, \ldots, X_t)$. Define the empirical counts

$$L_t(i) := \sum_{s=1}^{t} \mathbf{1}\{X_s = i\}, \qquad L_0(i) := 0,$$

and the excess/deficit process

$$\Delta_i(t) := L_t(i) - t p_i, \qquad \Delta(t) := (\Delta_1(t), \ldots, \Delta_d(t)).$$

Then

$$\sum_{i=1}^{d} \Delta_i(t) = 0 \qquad \text{for all } t \geq 0.$$

The local TSAW sampling rule. Fix $\lambda > 0$. For $t \geq 0$, define the normalizing factor

$$Z_t := \sum_{k=1}^{d} e^{-\lambda \Delta_k(t)},$$

and assume that

$$\mathbb{P}(X_{t+1} = i \mid \mathcal{F}_t) = q_t(i) := \frac{e^{-\lambda \Delta_i(t)}}{Z_t}, \qquad i \in [d]. \tag{6.2}$$

The deficit recursion is then

$$\Delta_i(t+1) = \Delta_i(t) + \mathbf{1}\{X_{t+1} = i\} - p_i, \qquad i \in [d]. \tag{6.3}$$

Define the maximum excess and the near-maximum set

$$M(t) := \max_{i \in [d]} \Delta_i(t), \qquad A(t) := \{i \in [d] : \Delta_i(t) \geq M(t) - 1\}.$$

The unit width of $A(t)$ is chosen to guarantee that if $X_{t+1} \notin A(t)$, then after the update in (6.3), the chosen coordinate still cannot exceed the previous maximizers. The following theorem is the basic Lyapunov estimate for the maximal deficit.

Theorem 6.3.1 (Exponential drift for the maximal excess)**.** *Fix $\alpha > 0$ and define*

$$M(t) := \max_{i \in [d]} \Delta_i(t), \qquad A(t) := \{i \in [d] : \Delta_i(t) \geq M(t) - 1\}, \qquad V(t) := e^{\alpha M(t)}.$$

Let

$$c_d := \frac{d}{d-1}, \qquad \rho := \exp\Big(-\frac{\alpha p_{\min}}{2}\Big) \in (0,1),$$

and set

$$R := \frac{1}{\lambda c_d} \log\left(\frac{d e^{\lambda}(e^{\alpha} - 1)}{e^{\alpha p_{\min}/2} - 1}\right), \qquad B := \exp\{\alpha(R + 1 - p_{\min})\}. \tag{6.4}$$

Note that $R > 0$ and $B \geq 1$. Then for all $t \geq 0$,

$$\mathbb{E}[V(t+1) \mid \mathcal{F}_t] \leq \rho V(t) + B\,\mathbf{1}\{M(t) \leq R\}. \tag{6.5}$$

Consequently,

$$\sup_{t \geq 0} \mathbb{E}\big[e^{\alpha M(t)}\big] \leq \max\Big\{e^{\alpha M(0)}, \frac{B}{1-\rho}\Big\} < \infty. \tag{6.6}$$

Proof. Fix $t \geq 0$. Since $\sum_{i=1}^{d} \Delta_i(t) = 0$ and $\max_i \Delta_i(t) = M(t)$, we have

$$\min_{k \in [d]} \Delta_k(t) \leq -\frac{M(t)}{d-1}. \tag{6.7}$$

Indeed, if $\min_k \Delta_k(t) > -M(t)/(d-1)$, then

$$\sum_{i=1}^{d} \Delta_i(t) > M(t) + (d-1)\Big(-\frac{M(t)}{d-1}\Big) = 0,$$

contradicting $\sum_i \Delta_i(t) = 0$. Therefore

$$Z_t = \sum_{k=1}^{d} e^{-\lambda \Delta_k(t)} \geq e^{-\lambda \min_k \Delta_k(t)} \geq \exp\Big(\frac{\lambda}{d-1} M(t)\Big). \tag{6.8}$$

If $i \in A(t)$, then $\Delta_i(t) \geq M(t) - 1$, so $e^{-\lambda \Delta_i(t)} \leq e^{-\lambda(M(t)-1)} = e^{\lambda} e^{-\lambda M(t)}$. Summing over $A(t)$ gives

$$\sum_{i \in A(t)} e^{-\lambda \Delta_i(t)} \leq |A(t)|\, e^{\lambda} e^{-\lambda M(t)} \leq d\, e^{\lambda} e^{-\lambda M(t)}.$$

Dividing by Z_t and using (6.2) and (6.8), we obtain

$$\mathbb{P}(X_{t+1} \in A(t) \mid \mathcal{F}_t) \leq d\, e^{\lambda} \exp\Big(-\lambda \frac{d}{d-1} M(t)\Big) = d\, e^{\lambda} e^{-\lambda c_d M(t)}. \tag{6.9}$$

We next compare $M(t+1)$ with $M(t)$. If $X_{t+1} = j \notin A(t)$, then $\Delta_j(t) \leq M(t) - 1$, so by (6.3),

$$\Delta_j(t+1) = \Delta_j(t) + 1 - p_j \leq (M(t) - 1) + 1 - p_{\min} = M(t) - p_{\min}.$$

For $k \neq j$, we have $\Delta_k(t+1) = \Delta_k(t) - p_k \leq M(t) - p_{\min}$. Hence

$$X_{t+1} \notin A(t) \quad \Rightarrow \quad M(t+1) \leq M(t) - p_{\min}. \tag{6.10}$$

Also, for every $i \in [d]$,

$$\Delta_i(t+1) = \Delta_i(t) + \mathbf{1}\{X_{t+1} = i\} - p_i \leq M(t) + 1 - p_{\min},$$

so we always have

$$M(t+1) \leq M(t) + (1 - p_{\min}). \tag{6.11}$$

Using (6.10) and (6.11), we get

$$e^{\alpha M(t+1)} \leq e^{\alpha(M(t)-p_{\min})} \mathbf{1}\{X_{t+1} \notin A(t)\} + e^{\alpha(M(t)+1-p_{\min})} \mathbf{1}\{X_{t+1} \in A(t)\}.$$

Taking conditional expectation and factoring out $e^{\alpha(M(t)-p_{\min})}$ gives

$$\mathbb{E}[e^{\alpha M(t+1)} \mid \mathcal{F}_t] \leq e^{\alpha(M(t)-p_{\min})}\Big(1 + (e^{\alpha} - 1)\, \mathbb{P}(X_{t+1} \in A(t) \mid \mathcal{F}_t)\Big). \tag{6.12}$$

Applying (6.9) yields

$$\mathbb{E}[V(t+1) \mid \mathcal{F}_t] \leq e^{\alpha M(t)} e^{-\alpha p_{\min}}\Big(1 + d e^{\lambda}(e^{\alpha} - 1) e^{-\lambda c_d M(t)}\Big). \tag{6.13}$$

If $M(t) \geq R$, with R given by (6.4), then

$$de^{\lambda}(e^{\alpha}-1)e^{-\lambda c_d M(t)} \leq de^{\lambda}(e^{\alpha}-1)e^{-\lambda c_d R} = e^{\alpha p_{\min}/2}-1,$$

so the term in the bracket on the RHS of (6.13) is at most $e^{\alpha p_{\min}/2}$. Therefore

$$\mathbb{E}[V(t+1) \mid \mathcal{F}_t] \leq e^{\alpha M(t)} e^{-\alpha p_{\min}} e^{\alpha p_{\min}/2} = \rho V(t).$$

If instead $M(t) \leq R$, then (6.11) gives $M(t+1) \leq R + (1-p_{\min})$, hence

$$V(t+1) \leq e^{\alpha(R+1-p_{\min})} = B.$$

Thus on $\{M(t) \leq R\}$ we have

$$\mathbb{E}[V(t+1) \mid \mathcal{F}_t] \leq B.$$

Combining the two regimes proves (6.5).

Finally, set $a_t := \mathbb{E}[V(t)]$. Taking expectations in (6.5) gives $a_{t+1} \leq \rho a_t + B$ for all $t \geq 0$. Let [1]

$$C := \max\left\{e^{\alpha M(0)}, \frac{B}{1-\rho}\right\}.$$

Then $a_0 \leq C$, and if $a_t \leq C$ then

$$a_{t+1} \leq \rho C + B \leq C,$$

since $C \geq B/(1-\rho)$. By induction, $a_t \leq C$ for all $t \geq 0$, which proves (6.6). □

A direct consequence of (6.6) is the following uniform-in-t exponential tail bound for each coordinate. Writing $C := \max\{e^{\alpha M(0)}, \frac{B}{1-\rho}\}$, Markov's inequality gives

$$\mathbb{P}\big(\Delta_i(t) > x\big) \leq \mathbb{P}\big(M(t) > x\big) \leq C_\alpha e^{-\alpha x}, \qquad x > 0.$$

On the other hand, since $\sum_{j=1}^{d} \Delta_j(t) = 0$, we have $-\Delta_i(t) \leq \sum_{j \neq i} \Delta_j(t) \leq (d-1)M(t)$, so

$$\mathbb{P}\big(\Delta_i(t) < -x\big) \leq \mathbb{P}\Big(M(t) > \frac{x}{d-1}\Big) \leq C_\alpha e^{-\alpha x/(d-1)}.$$

Therefore, for every $i \in [d]$, every $t \geq 0$, and every $x > 0$,

$$\mathbb{P}\big(|L_t(i) - tp_i| > x\big) = \mathbb{P}\big(|\Delta_i(t)| > x\big) \leq C_\alpha\Big(e^{-\alpha x} + e^{-\alpha x/(d-1)}\Big) \leq 2C_\alpha e^{-\alpha x/(d-1)}.$$

And the empirical counts remain within $O(1)$ of their targets tp_i at any finite level of trajectory length $t \geq 0$.

[1] Since $M(0) = 0$ and $B \geq 1$, we have $C = \frac{B}{1-\rho}$, but we prefer to write C this way for compatibility with the version of this calculation that appears when we discuss general graphs rather than just the star-graph.

Lifting the local TSAW to a random walk

Let V be a nonempty finite set, and let $(X_t)_{t\geq 0}$ be a V-valued process adapted to the natural filtration

$$\mathcal{F}_t := \sigma(X_0, X_1, \ldots, X_t), \qquad t \geq 0.$$

Fix a row-stochastic matrix $P = (P_{ij})_{i,j\in V}$:

$$P_{ij} \geq 0, \qquad \sum_{j\in V} P_{ij} = 1 \quad \text{for all } i \in V.$$

Throughout this subsection, irreducibility is assumed, but reversibility is not.

For $t \geq 0$ define the departure counts and directed edge counts

$$L_t(i) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i\}, \qquad N_t(i,j) := \sum_{s=0}^{t-1} \mathbf{1}\{X_s = i,\ X_{s+1} = j\}, \qquad i, j \in V.$$

This convention counts visits at times $0, \ldots, t-1$, so that $L_t(i)$ is exactly the number of departures from i before time t. Write L_t for the column vector $(L_t(i))_{i\in V}$ and N_t for the matrix $(N_t(i,j))_{i,j\in V}$.

Row sums satisfy

$$N_t\mathbf{1} = L_t, \tag{6.14}$$

where $\mathbf{1}$ is the all-ones column vector. Column sums satisfy, for each $j \in V$,

$$\sum_{i\in V} N_t(i,j) = L_t(j) - \mathbf{1}\{X_0 = j\} + \mathbf{1}\{X_t = j\}, \tag{6.15}$$

equivalently,

$$N_t^\top \mathbf{1} = L_t - \mathbf{e}_{X_0} + \mathbf{e}_{X_t}, \tag{6.16}$$

where $\mathbf{e}_x$ is the standard basis vector at x.

Define for $t \geq 0$ and $i, j \in V$ the edge discrepancy

$$\epsilon_t(i,j) := N_t(i,j) - L_t(i)\, P_{ij}. \tag{6.17}$$

In matrix form,

$$N_t = \operatorname{diag}(L_t)\, P + \epsilon_t. \tag{6.18}$$

Each row of ϵ_t sums to zero:

$$\epsilon_t\mathbf{1} = \mathbf{0}, \tag{6.19}$$

since $\sum_j N_t(i,j) = L_t(i)$ and $\sum_j P_{ij} = 1$.

The true self-avoiding walk variant. Fix $\lambda > 0$. The (time-inhomogeneous) transition rule is

$$\mathbb{P}(X_{t+1} = j \mid \mathcal{F}_t,\ X_t = i) = Q_t(i,j) := \frac{P_{ij} \exp\{-\lambda\, \epsilon_t(i,j)\}}{\sum_{k \in V} P_{ik} \exp\{-\lambda\, \epsilon_t(i,k)\}}. \tag{6.20}$$

When $P_{ij} = 0$ the numerator is 0, so (6.20) automatically restricts to the support $\{j : P_{ij} > 0\}$.

For $i \in V$ define the out-neighborhood and out-degree

$$\mathcal{N}(i) := \{j \in V :\ P_{ij} > 0\}, \qquad d_i := |\mathcal{N}(i)|.$$

If $d_i = 1$ then $\epsilon_t(i, \cdot) \equiv 0$ deterministically, so the nontrivial case is $d_i \geq 2$.

For $m \geq 0$ define the m-th departure time from row i by

$$\tau_m^{(i)} := \inf\{t \geq 0 :\ X_t = i,\ L_t(i) = m\},$$

with the convention $\inf \emptyset = \infty$.

The first step is to isolate a single row i and observe the process only at successive departure times from that row. The resulting process is closely related to the finite-alphabet local TSAW studied earlier, allowing us to exploit the results proved there.

Lemma 6.3.2 (Row-wise reduction)**.** *Assume $d_i \geq 2$ and $p_{\min}(i) := \min_{j \in \mathcal{N}(i)} P_{ij} > 0$.*

For $m \geq 0$ define the m-th departure time from row i by

$$\tau_m^{(i)} := \inf\{t \geq 0 :\ X_t = i,\ L_t(i) = m\},$$

with the convention $\inf \emptyset = \infty$. Thus, on the event $\{\tau_m^{(i)} < \infty\}$, the chain is at state i at time $\tau_m^{(i)}$, and exactly m departures from i have occurred before time $\tau_m^{(i)}$.

On $\{\tau_m^{(i)} < \infty\}$ define the destination of the next departure from i by

$$Y_{m+1}^{(i)} := X_{\tau_m^{(i)}+1}.$$

For $m \geq 0$ and $j \in V$, define the embedded row-counts

$$C_m^{(i)}(j) := \sum_{r=0}^{m-1} \mathbf{1}\{\tau_r^{(i)} < \infty,\ Y_{r+1}^{(i)} = j\}, \qquad \epsilon_m^{(i)}(j) := C_m^{(i)}(j) - mP_{ij},$$

with the empty sum convention $C_0^{(i)}(j) = 0$.

For $j \in \mathcal{N}(i)$, define further, for every $m \geq 0$,

$$\bar{c}_i := \frac{1}{d_i} \sum_{k \in \mathcal{N}(i)} \frac{1}{\lambda} \log P_{ik}, \qquad \Delta_m^{(i)}(j) := \epsilon_m^{(i)}(j) - \frac{1}{\lambda} \log P_{ij} + \bar{c}_i.$$

Then the following hold.

(i) *For every* $t \geq 0$ *and* $j \in V$,

$$N_t(i,j) = C^{(i)}_{L_t(i)}(j), \qquad \epsilon_t(i,j) = \epsilon^{(i)}_{L_t(i)}(j). \tag{6.21}$$

(ii) *For every* $m \geq 0$*, on the event* $\{\tau^{(i)}_m < \infty\}$,

$$\mathbb{P}\left(Y^{(i)}_{m+1} = j \,\middle|\, \mathcal{F}_{\tau^{(i)}_m}\right) = \frac{P_{ij}\exp\{-\lambda\,\epsilon^{(i)}_m(j)\}}{\sum_{k\in V} P_{ik}\exp\{-\lambda\,\epsilon^{(i)}_m(k)\}}, \qquad j \in \mathcal{N}(i). \tag{6.22}$$

(iii) *For every* $m \geq 0$ *and every* $j \in \mathcal{N}(i)$,

$$\epsilon^{(i)}_m(j) = \Delta^{(i)}_m(j) + \frac{1}{\lambda}\log P_{ij} - \bar{c}_i. \tag{6.23}$$

Moreover, on $\{\tau^{(i)}_m < \infty\}$,

$$\sum_{j\in\mathcal{N}(i)} \Delta^{(i)}_m(j) = 0,$$

and on $\{\tau^{(i)}_m < \infty\}$ *the transition rule becomes*

$$\mathbb{P}\left(Y^{(i)}_{m+1} = j \,\middle|\, \mathcal{F}_{\tau^{(i)}_m}\right) = \frac{e^{-\lambda\Delta^{(i)}_m(j)}}{\sum_{k\in\mathcal{N}(i)} e^{-\lambda\Delta^{(i)}_m(k)}}, \qquad j \in \mathcal{N}(i), \tag{6.24}$$

and

$$\Delta^{(i)}_{m+1}(j) = \Delta^{(i)}_m(j) + \mathbf{1}\{Y^{(i)}_{m+1} = j\} - P_{ij} \qquad \text{on } \{\tau^{(i)}_m < \infty\}.$$

Proof. For (i), note that $L_t(i)$ is exactly the number of departures from i before time t, and the r-th such departure, when it exists, occurs at time $\tau^{(i)}_{r-1}$ and goes to $Y^{(i)}_r$. Therefore the number of departures from i to j before time t is precisely the number of indices $r \in \{1,\ldots,L_t(i)\}$ for which $Y^{(i)}_r = j$, which gives the first identity in (6.21). The second then follows immediately from the definitions of $\epsilon_t(i,j)$ and $\epsilon^{(i)}_m(j)$.

For (ii), fix $m \geq 0$ and $j \in \mathcal{N}(i)$. First, $\tau^{(i)}_m$ is a stopping time with respect to $(\mathcal{F}_t)$, because for each $t \geq 0$,

$$\{\tau^{(i)}_m \leq t\} = \bigcup_{r=0}^{t}\{X_r = i,\ L_r(i) = m\} \in \mathcal{F}_t.$$

Now let $A \in \mathcal{F}_{\tau^{(i)}_m}$. Then $A \cap \{\tau^{(i)}_m = r\} \in \mathcal{F}_r$ for every $r \geq 0$. Therefore

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_A\,\mathbf{1}\{\tau^{(i)}_m < \infty,\ Y^{(i)}_{m+1} = j\}\Big] \\
&\quad = \sum_{r=0}^{\infty}\mathbb{E}\Big[\mathbf{1}_{A\cap\{\tau^{(i)}_m = r\}}\mathbf{1}\{X_{r+1} = j\}\Big] \\
&\quad = \sum_{r=0}^{\infty}\mathbb{E}\Big[\mathbf{1}_{A\cap\{\tau^{(i)}_m = r\}}\mathbb{E}[\mathbf{1}\{X_{r+1} = j\} \mid \mathcal{F}_r]\Big].
\end{aligned}$$

By (6.20),

$$\mathbb{E}[\mathbf{1}\{X_{r+1} = j\} \mid \mathcal{F}_r] = Q_r(X_r, j).$$

On the event $\{\tau_m^{(i)} = r\}$, we have $X_r = i$, and by (i),

$$\epsilon_r(i, \ell) = \epsilon_m^{(i)}(\ell) \qquad \text{for every } \ell \in V.$$

Hence on $\{\tau_m^{(i)} = r\}$,

$$Q_r(X_r, j) = \frac{P_{ij} \exp\{-\lambda\, \epsilon_m^{(i)}(j)\}}{\sum_{k \in V} P_{ik} \exp\{-\lambda\, \epsilon_m^{(i)}(k)\}}.$$

Substituting this back yields

$$\mathbb{E}\Big[\mathbf{1}_A\, \mathbf{1}\{\tau_m^{(i)} < \infty,\ Y_{m+1}^{(i)} = j\}\Big] = \mathbb{E}\left[\mathbf{1}_A\, \mathbf{1}\{\tau_m^{(i)} < \infty\} \frac{P_{ij} \exp\{-\lambda\, \epsilon_m^{(i)}(j)\}}{\sum_{k \in V} P_{ik} \exp\{-\lambda\, \epsilon_m^{(i)}(k)\}}\right],$$

which proves (6.22) on $\{\tau_m^{(i)} < \infty\}$.

For (iii), identity (6.23) is immediate from the definition of $\Delta_m^{(i)}$. Now fix $m \geq 0$ and work on the event $\{\tau_m^{(i)} < \infty\}$. Then the first m departures from row i have all occurred, so for each $r \in \{0, \dots, m-1\}$, the variable $Y_{r+1}^{(i)}$ is defined and takes exactly one value in $\mathcal{N}(i)$. Therefore

$$\sum_{j \in \mathcal{N}(i)} C_m^{(i)}(j) = m, \qquad \sum_{j \in \mathcal{N}(i)} \epsilon_m^{(i)}(j) = 0.$$

Since $\sum_{j \in \mathcal{N}(i)} P_{ij} = 1$, it follows that on the event $\{\tau_{m-1}^{(i)} < \infty\}$, and hence on $\{\tau_m^{(i)} < \infty\}$, we have

$$\sum_{j \in \mathcal{N}(i)} \Delta_m^{(i)}(j) = \sum_{j \in \mathcal{N}(i)} \left(\epsilon_m^{(i)}(j) - \frac{1}{\lambda} \log P_{ij} + \bar{c}_i\right) = 0.$$

Also,

$$e^{-\lambda \Delta_m^{(i)}(j)} = e^{-\lambda \bar{c}_i}\, P_{ij}\, e^{-\lambda \epsilon_m^{(i)}(j)},$$

so multiplying the factor $e^{-\lambda \bar{c}_i}$ to both the numerator and the denominator in (6.22), we arrive at (6.24). Finally, on $\{\tau_m^{(i)} < \infty\}$,

$$C_{m+1}^{(i)}(j) = C_m^{(i)}(j) + \mathbf{1}\{Y_{m+1}^{(i)} = j\},$$

hence

$$\epsilon_{m+1}^{(i)}(j) = \epsilon_m^{(i)}(j) + \mathbf{1}\{Y_{m+1}^{(i)} = j\} - P_{ij},$$

and therefore

$$\Delta_{m+1}^{(i)}(j) = \Delta_m^{(i)}(j) + \mathbf{1}\{Y_{m+1}^{(i)} = j\} - P_{ij}.$$

Finally, on $\{\tau_m^{(i)} < \infty\}$, the variable $Y_{m+1}^{(i)}$ is well-defined: at time $\tau_m^{(i)}$, the walk is at i and exactly m departures from i have occurred, so the next transition is the $(m+1)$-st departure from i. $\square$

Once the row-wise reduction has been identified, the finite-alphabet drift theorem, Theorem 6.3.1, applies to the embedded row process on the event that the next departure from row i is defined. This yields the same exponential-drift inequality as in the local model, with indicators keeping track of the possibility that the row process may terminate.

Corollary 6.3.3 (Row-wise exponential moment bound)**.** *Under the assumptions of Lemma 6.3.2, for every $\alpha > 0$ there exist constants $\rho_i(\alpha) \in (0,1)$ and $B_i(\alpha) < \infty$, depending only on α, λ, and the row $P_{i\cdot}$, such that, writing*

$$M^{(i)}_{\mathrm{urn}}(m) := \max_{j\in\mathcal{N}(i)} \Delta^{(i)}_m(j), \qquad V^{(i)}_{\mathrm{urn}}(m) := \exp\{\alpha M^{(i)}_{\mathrm{urn}}(m)\},$$

one has, for every $m \geq 0$,

$$\mathbb{E}\left[V^{(i)}_{\mathrm{urn}}(m+1)\,\mathbf{1}\{\tau^{(i)}_m < \infty\}\,\middle|\,\mathcal{F}_{\tau^{(i)}_m}\right] \leq \rho_i(\alpha)\,V^{(i)}_{\mathrm{urn}}(m)\,\mathbf{1}\{\tau^{(i)}_m < \infty\} + B_i(\alpha)\,\mathbf{1}\{\tau^{(i)}_m < \infty\}. \tag{6.25}$$

Moreover, with the convention $\tau^{(i)}_{-1} := 0$,

$$\sup_{m\geq 0}\,\mathbb{E}\left[\exp\left\{\alpha \max_{j\in\mathcal{N}(i)} \Delta^{(i)}_m(j)\right\}\mathbf{1}\{\tau^{(i)}_{m-1} < \infty\}\right] < \infty. \tag{6.26}$$

In fact one may take

$$\rho_i(\alpha) := \exp\Big(-\frac{\alpha p_{\min}(i)}{2}\Big), \qquad c_i := \frac{d_i}{d_i - 1},$$

$$R_i(\alpha) := \frac{1}{\lambda c_i}\log\left(\frac{d_i e^{\lambda}(e^{\alpha}-1)}{e^{\alpha p_{\min}(i)/2}-1}\right), \qquad B_i(\alpha) := \exp\{\alpha(R_i(\alpha) + 1 - p_{\min}(i))\}.$$

Note that $\rho_i(\alpha) \in (0,1)$, $R_i(\alpha) > 0$ and $B_i(\alpha) > 1$. Also, since we are making the assumptions of Lemma 6.3.2, we are in a scenario where $d_i \geq 2$, so c_i is well-defined.

Proof. Fix $m \geq 0$. On the event $\{\tau^{(i)}_m < \infty\}$, Lemma 6.3.2(iii) shows that the vector $\big(\Delta^{(i)}_m(j)\big)_{j\in\mathcal{N}(i)}$ has zero sum and that the next-step transition rule is exactly the finite-alphabet softmax rule (6.2) with target distribution $p_j = P_{ij}$ on the alphabet $\mathcal{N}(i)$. Therefore, conditional on $\mathcal{F}_{\tau^{(i)}_m}$ and on $\{\tau^{(i)}_m < \infty\}$, the one-step estimate from Theorem 6.3.1 applies to the row process. This yields (6.25).

Now define

$$a_m := \mathbb{E}\left[V^{(i)}_{\mathrm{urn}}(m)\,\mathbf{1}\{\tau^{(i)}_{m-1} < \infty\}\right], \qquad m \geq 0,$$

with the convention $\tau^{(i)}_{-1} := 0$. Since

$$\Delta^{(i)}_0(j) = -\frac{1}{\lambda}\log P_{ij} + \bar{c}_i,$$

we have $a_0 = e^{\alpha M^{(i)}_{\mathrm{urn}}(0)} < \infty$. Taking expectations in (6.25) gives

$$a_{m+1} = \mathbb{E}\left[V^{(i)}_{\mathrm{urn}}(m+1)\,\mathbf{1}\{\tau^{(i)}_m < \infty\}\right] \le \rho_i(\alpha)\,\mathbb{E}\left[V^{(i)}_{\mathrm{urn}}(m)\,\mathbf{1}\{\tau^{(i)}_m < \infty\}\right] + B_i(\alpha)\,\mathbb{P}(\tau^{(i)}_m < \infty).$$

Since $\{\tau^{(i)}_m < \infty\} \subseteq \{\tau^{(i)}_{m-1} < \infty\}$, it follows that

$$a_{m+1} \le \rho_i(\alpha) a_m + B_i(\alpha).$$

Hence

$$a_m \le \max\left\{e^{\alpha M^{(i)}_{\mathrm{urn}}(0)},\ \frac{B_i(\alpha)}{1-\rho_i(\alpha)}\right\} \qquad \text{for all } m \ge 0,$$

which is exactly (6.26). □

The next step is to pass from urn time along a single row to a deterministic global time t. The idea is to condition on the realized number of departures from each row up to time t, invoke the row-wise exponential moment bound, and then take a union bound over the finitely many rows. It is also natural to record at this stage an explicit choice of the exponential-moment constant, since that is exactly what enters the later Borel–Cantelli argument.

Lemma 6.3.4 (Fixed-time tail bound for the global edge discrepancy)**.** *Assume that for every $i \in V$, either $d_i = 1$ or $d_i \ge 2$ with $p_{\min}(i) := \min_{j\in\mathcal{N}(i)} P_{ij} > 0$. If $d_{\max} := \max_{i\in V} d_i = 1$, then $M(t) \equiv 0$ for all t. Otherwise define*

$$p_{\min} := \min_{i\in V}\ \min_{j\in\mathcal{N}(i)} P_{ij} \in (0,1).$$

Then there exist finite constants

$$C_0 := \log\left(1 + \frac{2}{p_{\min}}\right), \qquad C_1 := 1 - p_{\min} + \frac{1}{\lambda}\left(\log d_{\max} + \lambda + \log 2 + 2\log\frac{1}{p_{\min}}\right),$$

such that, for every $\alpha \ge 1$, the quantity

$$H_\alpha := \exp\left(\frac{\alpha^2}{\lambda} + C_1\alpha + C_0\right)$$

results in the bound

$$\mathbb{P}(M(t) \ge x) \le |V|\,(t+1)\,H_\alpha\,e^{-\alpha x} \qquad \textit{for all } t \ge 0 \textit{ and } x \ge 0.$$

Proof. If $d_{\max} = 1$, then every row has a single possible transition, so $\epsilon_t(i,\cdot) \equiv 0$ for every i, and hence $M(t) \equiv 0$. We then have $\epsilon_t(i,\cdot) \equiv 0$ for every i, and hence $M(t) \equiv 0$. Thus suppose $d_{\max} \ge 2$.

Fix $i \in V$. If $d_i = 1$, define $M_i(t) := 0$ for all t. If $d_i \geq 2$, define

$$M_i(t) := \max_{j \in \mathcal{N}(i)} \epsilon_t(i, j).$$

Then

$$M(t) = \max_{i \in V} M_i(t).$$

Fix now a row i with $d_i \geq 2$. By Lemma 6.3.2(i),

$$M_i(t) = \max_{j \in \mathcal{N}(i)} \epsilon_t(i, j) = \max_{j \in \mathcal{N}(i)} \epsilon^{(i)}_{L_t(i)}(j).$$

By Lemma 6.3.2(iii), for every $m \geq 0$ and $j \in \mathcal{N}(i)$,

$$\epsilon^{(i)}_m(j) = \Delta^{(i)}_m(j) + \frac{1}{\lambda} \log P_{ij} - \bar{c}_i.$$

Define

$$K_i := \max_{j \in \mathcal{N}(i)} \left(\frac{1}{\lambda} \log P_{ij} - \bar{c}_i \right).$$

Then for every $m \geq 0$,

$$\max_{j \in \mathcal{N}(i)} \epsilon^{(i)}_m(j) \leq \max_{j \in \mathcal{N}(i)} \Delta^{(i)}_m(j) + K_i.$$

Therefore, by Corollary 6.3.3,

$$\begin{aligned} &\sup_{m \geq 0} \mathbb{E}\Bigg[\exp\Big\{ \alpha \max_{j \in \mathcal{N}(i)} \epsilon^{(i)}_m(j) \Big\} \mathbf{1}\{\tau^{(i)}_{m-1} < \infty\} \Bigg] \\ &\qquad\qquad \leq e^{\alpha K_i} \sup_{m \geq 0} \mathbb{E}\Bigg[\exp\Big\{ \alpha \max_{j \in \mathcal{N}(i)} \Delta^{(i)}_m(j) \Big\} \mathbf{1}\{\tau^{(i)}_{m-1} < \infty\} \Bigg]. \end{aligned} \tag{6.27}$$

From Corollary 6.3.3,

$$\sup_{m \geq 0} \mathbb{E}\left[\exp\Big\{ \alpha \max_{j \in \mathcal{N}(i)} \Delta^{(i)}_m(j) \Big\} \mathbf{1}\{\tau^{(i)}_{m-1} < \infty\} \right] \leq \max\left\{ e^{\alpha M^{(i)}_{\mathrm{urn}}(0)},\ \frac{B_i(\alpha)}{1 - \rho_i(\alpha)} \right\},$$

where

$$\rho_i(\alpha) := \exp\Big(-\frac{\alpha p_{\min}(i)}{2} \Big),\ R_i(\alpha) := \frac{1}{\lambda c_i} \log\left(\frac{d_i e^{\lambda}(e^{\alpha} - 1)}{e^{\alpha p_{\min}(i)/2} - 1} \right),$$
$$B_i(\alpha) := \exp\{\alpha(R_i(\alpha) + 1 - p_{\min}(i))\}.$$

Because $P_{ij} \geq p_{\min}$ on every support edge, every number $\lambda^{-1} \log P_{ij}$ lies in the interval $[-\lambda^{-1} \log(1/p_{\min}), 0]$. Since $\bar{c}_i$ is the average of these numbers,

$$K_i \leq \frac{1}{\lambda} \log \frac{1}{p_{\min}}, \qquad M^{(i)}_{\mathrm{urn}}(0) = \max_{j \in \mathcal{N}(i)} \left(-\frac{1}{\lambda} \log P_{ij} + \bar{c}_i \right) \leq \frac{1}{\lambda} \log \frac{1}{p_{\min}}.$$

Also, since $d_i \le d_{\max}$, $p_{\min}(i) \ge p_{\min}$, $c_i \ge 1$, and $\alpha \ge 1$,

$$R_i(\alpha) \le \frac{1}{\lambda} \log\left(\frac{d_{\max} e^{\lambda}(e^{\alpha}-1)}{e^{\alpha p_{\min}/2}-1}\right) \le \frac{\alpha}{\lambda} + \frac{1}{\lambda}\left(\log d_{\max} + \lambda + \log 2 + \log \frac{1}{p_{\min}}\right),$$

where we used $e^{\alpha}-1 \le e^{\alpha}$ and $e^{u}-1 \ge u$ for $u>0$. Moreover,

$$\frac{1}{1-\rho_i(\alpha)} = \frac{1}{1-e^{-\alpha p_{\min}(i)/2}} \le 1 + \frac{2}{\alpha p_{\min}} \le 1 + \frac{2}{p_{\min}},$$

since $\alpha \ge 1$. Combining these estimates yields

$$e^{\alpha K_i} \max\left\{e^{\alpha M_{urn}^{(i)}(0)},\ \frac{B_i(\alpha)}{1-\rho_i(\alpha)}\right\} \le \exp\left(\frac{\alpha^2}{\lambda} + C_1\alpha + C_0\right) = H_\alpha.$$

Together with (6.27), this gives

$$\sup_{m\ge 0} \mathbb{E}\left[\exp\left\{\alpha \max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j)\right\} \mathbf{1}\{\tau_{m-1}^{(i)} < \infty\}\right] \le H_\alpha \tag{6.28}$$

for every row i with $d_i \ge 2$.

Now fix $t \ge 0$, $x \ge 0$, and a row i with $d_i \ge 2$. Since $L_t(i) \in \{0,1,\dots,t\}$,

$$\{M_i(t) \ge x\} = \bigcup_{m=0}^{t} \left(\{L_t(i) = m\} \cap \left\{\max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j) \ge x\right\}\right),$$

and these events are disjoint. Therefore

$$\mathbb{P}(M_i(t) \ge x) = \sum_{m=0}^{t} \mathbb{P}\Big(L_t(i) = m,\ \max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j) \ge x\Big).$$

If $m=0$, then $\{L_t(i)=0\}$ implies that no departure from i has occurred before time t, hence $N_t(i,j)=0$ for all j, so $\epsilon_t(i,j)=0$ for all j, and therefore $M_i(t)=0$. Thus the $m=0$ term is zero for $x>0$, and is trivially bounded by $H_\alpha e^{-\alpha x}$ for $x=0$.

If $m \ge 1$, then

$$\{L_t(i) = m\} \subseteq \{\tau_{m-1}^{(i)} < \infty\},$$

because having m departures from i before time t implies that the $(m-1)$-st departure time from i exists. Hence, by Markov's inequality and (6.28),

$$\begin{aligned}
&\mathbb{P}\Big(L_t(i) = m,\ \max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j) \ge x\Big) \\
&\quad\le e^{-\alpha x}\mathbb{E}\left[\exp\left\{\alpha \max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j)\right\} \mathbf{1}\{L_t(i) = m\}\right] \\
&\quad\le e^{-\alpha x}\mathbb{E}\left[\exp\left\{\alpha \max_{j\in\mathcal{N}(i)} \epsilon_m^{(i)}(j)\right\} \mathbf{1}\{\tau_{m-1}^{(i)} < \infty\}\right] \\
&\quad\le H_\alpha e^{-\alpha x}.
\end{aligned}$$

Summing over $m = 0, \dots, t$, we obtain

$$\mathbb{P}(M_i(t) \geq x) \leq (t+1)H_\alpha e^{-\alpha x} \qquad \text{for every row } i \text{ with } d_i \geq 2.$$

For rows with $d_i = 1$, we have $M_i(t) \equiv 0$, so the same bound is trivial. Therefore, for all $i \in V$,

$$\mathbb{P}(M_i(t) \geq x) \leq (t+1)H_\alpha e^{-\alpha x}.$$

Finally, since $M(t) = \max_{i \in V} M_i(t)$, a union bound over $i \in V$ gives

$$\mathbb{P}(M(t) \geq x) \leq |V|\,(t+1)\,H_\alpha\,e^{-\alpha x},$$

as claimed. □

The almost-sure envelope is now obtained by choosing a time-dependent exponential parameter α_t, applying the fixed-time tail bound with target error probability η_t, and then using Borel–Cantelli.

Corollary 6.3.5. *Assume the hypotheses of Lemma 6.3.4. If $d_{\max} := \max_{i \in V} d_i = 1$, then $M(t) \equiv 0$ for all t. Otherwise define*

$$p_{\min} := \min_{i:\, d_i \geq 2} \min_{j \in \mathcal{N}(i)} P_{ij} \in (0,1], \qquad d_{\max} := \max_{i \in V} d_i.$$

Fix $\delta > 0$ and set

$$\eta_t := (t+2)^{-(1+\delta)},\ \Lambda_t := \log \frac{|V|\,(t+1)}{\eta_t} = \log\big(|V|\,(t+1)(t+2)^{1+\delta}\big),\ \alpha_t := \max\{1, \sqrt{\lambda \Lambda_t}\}.$$

Then, almost surely, for all sufficiently large t,

$$M(t) \leq C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}, \tag{6.29}$$

where C_0 and C_1 are the constants from Lemma 6.3.4.

Proof. If $d_{\max} = 1$, then $M(t) \equiv 0$, so there is nothing to prove. Assume $d_{\max} \geq 2$. Apply Lemma 6.3.4 with $\alpha = \alpha_t \geq 1$. Since $H_{\alpha_t} = \exp(\alpha_t^2/\lambda + C_1\alpha_t + C_0)$, we obtain

$$\mathbb{P}(M(t) \geq x) \leq |V|\,(t+1)\exp\left(\frac{\alpha_t^2}{\lambda} + C_1\alpha_t + C_0 - \alpha_t x\right).$$

Now set

$$x_t := \frac{\alpha_t}{\lambda} + C_1 + \frac{C_0}{\alpha_t} + \frac{\Lambda_t}{\alpha_t}.$$

Because $\Lambda_t = \log\big(|V|\,(t+1)/\eta_t\big)$, the preceding display becomes $\mathbb{P}(M(t) \geq x_t) \leq \eta_t$. Since $\sum_{t \geq 0} \eta_t < \infty$, Borel–Cantelli implies that almost surely the event $\{M(t) \geq x_t\}$ occurs only finitely often. Hence almost surely, for all sufficiently large t, $M(t) \leq x_t$.

Finally, for all sufficiently large t we have $\alpha_t = \sqrt{\lambda \Lambda_t}$, so

$$\frac{\alpha_t}{\lambda} + \frac{\Lambda_t}{\alpha_t} = 2\sqrt{\frac{\Lambda_t}{\lambda}}.$$

Substituting this into the definition of x_t yields (6.29). □

Vertex and edge discrepancy bounds

The previous subsection gives an almost-sure bound on the edge discrepancy

$$M(t) := \max_{i \in V} \max_{j \in \mathcal{N}(i)} \epsilon_t(i,j).$$

We now convert this bound into control of the vertex discrepancy $L_t - t\pi$. The key point is that the column-flow identity for N_t yields a Poisson equation whose forcing term is determined by ϵ_t, and hence by $M(t)$.

Lemma 6.3.6 (Poisson equation and deterministic control of the forcing term)**.** *For each $t \geq 0$, define*

$$b_t := \epsilon_t^\top \mathbf{1} + \mathbf{e}_{X_0} - \mathbf{e}_{X_t}.$$

Then

$$(I - P^\top)L_t = b_t, \tag{6.30}$$

and moreover $\mathbf{1}^\top b_t = 0$. If $d_{\max} := \max_{i \in V} d_i$, then for every $i, j \in V$,

$$|\epsilon_t(i,j)| \leq (d_{\max} - 1)M(t),$$

and consequently

$$\|b_t\|_\infty \leq |V|\,(d_{\max} - 1)\,M(t) + 1.$$

Proof. By (6.16), we have $N_t^\top \mathbf{1} = L_t - \mathbf{e}_{X_0} + \mathbf{e}_{X_t}$. Using (6.18), this becomes

$$L_t - \mathbf{e}_{X_0} + \mathbf{e}_{X_t} = N_t^\top \mathbf{1} = (\mathrm{diag}(L_t)P + \epsilon_t)^\top \mathbf{1} = P^\top L_t + \epsilon_t^\top \mathbf{1}.$$

Rearranging gives (6.30).

Next, $\mathbf{1}^\top b_t = \mathbf{1}^\top \epsilon_t^\top \mathbf{1} + \mathbf{1}^\top(\mathbf{e}_{X_0} - \mathbf{e}_{X_t}) = 0$, because $\epsilon_t \mathbf{1} = \mathbf{0}$ by (6.19) and $\mathbf{1}^\top \mathbf{e}_{X_0} = \mathbf{1}^\top \mathbf{e}_{X_t} = 1$.

To bound ϵ_t, fix $i \in V$. If $d_i = 1$, then $\epsilon_t(i,\cdot) \equiv 0$, so there is nothing to prove. Assume $d_i \geq 2$. Since the i-th row of ϵ_t is supported on $\mathcal{N}(i)$ and sums to zero, if

$$M_i(t) := \max_{j \in \mathcal{N}(i)} \epsilon_t(i,j),$$

then

$$\min_{j \in \mathcal{N}(i)} \epsilon_t(i,j) \geq -(d_i - 1)M_i(t).$$

Indeed, if some entry were smaller than $-(d_i - 1)M_i(t)$, then the sum of the row would be strictly negative. Hence, for every $j \in \mathcal{N}(i)$,

$$|\epsilon_t(i,j)| \leq (d_i - 1)M_i(t) \leq (d_{\max} - 1)M(t).$$

If $j \notin \mathcal{N}(i)$, then $P_{ij} = 0$ and $N_t(i,j) = 0$, so $\epsilon_t(i,j) = 0$. Thus

$$|\epsilon_t(i,j)| \leq (d_{\max} - 1)M(t) \qquad \text{for all } i, j \in V.$$

Finally, for each $j \in V$,

$$|b_t(j)| = \Big| \sum_{i \in V} \epsilon_t(i,j) + \mathbf{1}\{X_0 = j\} - \mathbf{1}\{X_t = j\} \Big| \le \sum_{i \in V} |\epsilon_t(i,j)| + 1 \le |V| \, (d_{\max} - 1) \, M(t) + 1.$$

Taking the maximum over j gives the stated bound on $\|b_t\|_\infty$. □

The linear system $(I - P^\top)x = b$ is singular because π lies in the kernel of $I - P^\top$. The correct inverse object is therefore the *group inverse*, which acts on the codimension-one subspace $\{b : \mathbf{1}^\top b = 0\}$.

Let π denote the unique stationary distribution of the irreducible kernel P, and define

$$\mathcal{L} := I - P^\top, \qquad \Pi := \pi \, \mathbf{1}^\top.$$

Lemma 6.3.7 (Group inverse and normalized solution of the Poisson equation)**.** *The matrix $\mathcal{L} + \Pi$ is invertible. Define*

$$\mathcal{L}^{\#} := (\mathcal{L} + \Pi)^{-1} - \Pi. \tag{6.31}$$

This matrix is the group inverse of $\mathcal{L}$, in the sense that

$$\mathcal{L}\mathcal{L}^{\#} = \mathcal{L}^{\#}\mathcal{L} = I - \Pi, \qquad \mathcal{L}^{\#}\Pi = \Pi\mathcal{L}^{\#} = 0.$$

Moreover, for every $b \in \mathbb{R}^V$ satisfying $\mathbf{1}^\top b = 0$, the equation

$$\mathcal{L}x = b$$

has a unique solution satisfying $\mathbf{1}^\top x = 0$, and this solution is given by

$$x = \mathcal{L}^{\#} b.$$

Proof. We first show that $\mathcal{L} + \Pi$ is invertible. Suppose $(\mathcal{L} + \Pi)x = 0$. Left-multiplying by $\mathbf{1}^\top$ gives

$$0 = \mathbf{1}^\top(\mathcal{L} + \Pi)x = \underbrace{\mathbf{1}^\top\mathcal{L}}_{=0} x + \mathbf{1}^\top \Pi x = \mathbf{1}^\top x,$$

because $\mathbf{1}^\top \Pi x = \mathbf{1}^\top(\pi \, \mathbf{1}^\top x) = \mathbf{1}^\top x$. Hence $\mathbf{1}^\top x = 0$, so $\Pi x = 0$, and therefore $\mathcal{L}x = 0$. Since P is irreducible on the finite set V, $\ker(\mathcal{L}) = \operatorname{span}\{\pi\}$. Thus $x = c\pi$ for some scalar c, and then $0 = \mathbf{1}^\top x = c \, \mathbf{1}^\top \pi = c$, so $x = 0$. Hence $\mathcal{L} + \Pi$ is invertible.

Now define $\mathcal{L}^{\#}$ by (6.31). Because $\mathcal{L}\Pi = 0$ and $\Pi\mathcal{L} = 0$, we have

$$(\mathcal{L} + \Pi)\Pi = \Pi \qquad \text{and} \qquad \Pi(\mathcal{L} + \Pi) = \Pi.$$

Multiplying by $(\mathcal{L} + \Pi)^{-1}$ gives

$$(\mathcal{L} + \Pi)^{-1}\Pi = \Pi \qquad \text{and} \qquad \Pi(\mathcal{L} + \Pi)^{-1} = \Pi.$$

Therefore

$$\mathcal{L}^{\#}\Pi = ((\mathcal{L}+\Pi)^{-1} - \Pi)\Pi = 0, \qquad \Pi\mathcal{L}^{\#} = \Pi((\mathcal{L}+\Pi)^{-1} - \Pi) = 0.$$

Also,

$$(\mathcal{L}+\Pi)\mathcal{L}^{\#} = (\mathcal{L}+\Pi)\big((\mathcal{L}+\Pi)^{-1} - \Pi\big) = I - \Pi,$$

and since $\Pi\mathcal{L}^{\#} = 0$, this implies $\mathcal{L}\mathcal{L}^{\#} = I - \Pi$. Similarly,

$$\mathcal{L}^{\#}(\mathcal{L}+\Pi) = \big((\mathcal{L}+\Pi)^{-1} - \Pi\big)(\mathcal{L}+\Pi) = I - \Pi,$$

and because $\mathcal{L}^{\#}\Pi = 0$, it follows that $\mathcal{L}^{\#}\mathcal{L} = I - \Pi$.

Finally, let $b \in \mathbb{R}^V$ satisfy $\mathbf{1}^\top b = 0$. Then $\Pi b = \pi(\mathbf{1}^\top b) = 0$. Set $x := \mathcal{L}^{\#}b$. Using $\mathcal{L}\mathcal{L}^{\#} = I - \Pi$, we obtain

$$\mathcal{L}x = \mathcal{L}\mathcal{L}^{\#}b = (I - \Pi)b = b.$$

Using $\Pi\mathcal{L}^{\#} = 0$, we also have

$$\Pi x = \Pi\mathcal{L}^{\#}b = 0.$$

Since $\Pi x = \pi(\mathbf{1}^\top x)$ and π has strictly positive entries, this implies $\mathbf{1}^\top x = 0$.

It remains to prove uniqueness. Suppose x and y both satisfy $\mathcal{L}x = \mathcal{L}y = b$ and $\mathbf{1}^\top x = \mathbf{1}^\top y = 0$. Then $z := x - y$ satisfies $\mathcal{L}z = 0$, so $z \in \operatorname{span}\{\pi\}$. Thus $z = c\pi$ for some scalar c, and $0 = \mathbf{1}^\top z = c\,\mathbf{1}^\top \pi = c$. Hence $z = 0$, so $x = y$. □

We now apply the previous two lemmas with $b = b_t$ and $x = D_t := L_t - t\pi$. The almost-sure $\sqrt{\log t}$ envelope for $M(t)$ then immediately yields a corresponding bound on the vertex discrepancy.

Corollary 6.3.8 (Vertex and edge discrepancy bounds; estimator error)**.** *Assume:*

1. *P is row-stochastic on finite V and the adaptive rule* (6.20) *holds;*
2. *for each $i \in V$, either $d_i = 1$ or $d_i \geq 2$ with $p_{\min}(i) := \min_{j \in \mathcal{N}(i)} P_{ij} > 0$;*
3. *P is irreducible.*

Let π be the stationary distribution of P, define

$$D_t := L_t - t\pi,$$

and fix $\delta > 0$. Set

$$\Lambda_t := \log\big(|V|\,(t+1)(t+2)^{1+\delta}\big), \qquad \alpha_t := \max\{1, \sqrt{\lambda\Lambda_t}\}.$$

Then, almost surely, for all sufficiently large t,

(i) *for every $i \in V$,*

$$|L_t(i) - t\pi_i| \le \|\mathcal{L}^{\#}\|_{\infty\to\infty}\left(|V|(d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right) + 1\right); \tag{6.32}$$

(ii) *for every (i,j) with $P_{ij} > 0$,*

$$\begin{aligned} |N_t(i,j) - t\pi_i P_{ij}| \le& P_{ij}\,\|\mathcal{L}^{\#}\|_{\infty\to\infty}\left(|V|(d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right) + 1\right) \\ &+ (d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right); \end{aligned} \tag{6.33}$$

(iii) *for every bounded function $f : V \to \mathbb{R}$,*

$$\begin{aligned} &\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_s) - \sum_{i\in V}\pi_i f(i)\right| \\ &\qquad \le \frac{\|f\|_\infty\,|V|}{t}\,\|\mathcal{L}^{\#}\|_{\infty\to\infty}\left(|V|(d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right) + 1\right). \end{aligned} \tag{6.34}$$

In particular,

$$L_t(i) - t\pi_i = O(\sqrt{\log t}), \qquad N_t(i,j) - t\pi_i P_{ij} = O(\sqrt{\log t}),$$

almost surely for every $i \in V$ and every (i,j) with $P_{ij} > 0$, and for every bounded $f : V \to \mathbb{R}$,

$$\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_s) - \sum_{i\in V}\pi_i f(i)\right| = O\left(\frac{\sqrt{\log t}}{t}\right) \qquad \textit{almost surely.}$$

Proof. By Corollary 6.3.5, almost surely, for all sufficiently large t,

$$M(t) \le C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}.$$

This is exactly the bound recorded in (6.32) after applying the Poisson equation estimates below.

Next, Lemma 6.3.6 gives

$$(I - P^\top)L_t = b_t,$$

while $(I - P^\top)(t\pi) = 0$ because $P^\top\pi = \pi$. Hence, with $D_t := L_t - t\pi$,

$$\mathcal{L}D_t = (I - P^\top)(L_t - t\pi) = b_t.$$

Also,

$$\mathbf{1}^\top D_t = \mathbf{1}^\top L_t - t\,\mathbf{1}^\top \pi = t - t = 0.$$

Therefore Lemma 6.3.7 yields

$$D_t = \mathcal{L}^{\#} b_t.$$

Taking ℓ^∞-norms and using Lemma 6.3.6, we obtain

$$\|D_t\|_\infty \le \|\mathcal{L}^{\#}\|_{\infty\to\infty}\, \|b_t\|_\infty \le \|\mathcal{L}^{\#}\|_{\infty\to\infty}\big(|V|(d_{\max}-1)M(t)+1\big).$$

Substituting the bound on $M(t)$ from Corollary 6.3.5 gives (6.32).

For the edge discrepancy, use

$$N_t(i,j) = L_t(i)P_{ij} + \epsilon_t(i,j),$$

so

$$N_t(i,j) - t\pi_i P_{ij} = P_{ij}\big(L_t(i) - t\pi_i\big) + \epsilon_t(i,j).$$

Hence

$$|N_t(i,j) - t\pi_i P_{ij}| \le P_{ij}\,|L_t(i) - t\pi_i| + |\epsilon_t(i,j)|.$$

By Lemma 6.3.6, we have

$$|\epsilon_t(i,j)| \le (d_{\max}-1)M(t),$$

so (6.33) follows from (6.32) and the bound from Corollary 6.3.5.

Finally,

$$\frac{1}{t}\sum_{s=0}^{t-1} f(X_s) - \sum_{i\in V}\pi_i f(i) = \frac{1}{t}\sum_{i\in V} f(i)\big(L_t(i) - t\pi_i\big),$$

and therefore

$$\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_s) - \sum_{i\in V}\pi_i f(i)\right| \le \frac{\|f\|_\infty}{t}\sum_{i\in V}|L_t(i) - t\pi_i| \le \frac{\|f\|_\infty\,|V|}{t}\,\|D_t\|_\infty.$$

Combining this with (6.32) yields (6.34).

The asymptotic statements are immediate from the fact that

$$\Lambda_t = \log\left(|V|\,(t+1)(t+2)^{1+\delta}\right) = O(\log t)$$

and

$$\alpha_t = \max\{1, \sqrt{\lambda\Lambda_t}\} = O(\sqrt{\log t}).$$

□

It is then a natural consequence of this result that this variant of TSAW on the graph is indeed recurrent.

Corollary 6.3.9 (Recurrence of states and directed edges)**.** *Under the assumptions of Corollary 6.3.8, almost surely every state $i \in V$ is visited infinitely often. Moreover, every directed edge (i, j) with $P_{ij} > 0$ is traversed infinitely often.*

Proof. Let $\delta > 0$ be as in Corollary 6.3.8. Since P is irreducible on the finite state space V, its stationary distribution π is strictly positive: $\pi_i > 0$ for every $i \in V$.

By Corollary 6.3.8, there is an event Ω_0 of probability one on which, for all sufficiently large t, the bounds (6.32) and (6.33) hold simultaneously for every $i \in V$ and every support edge (i, j) with $P_{ij} > 0$. Work on this event Ω_0.

Let

$$A_t := \|\mathcal{L}^{\#}\|_{\infty\to\infty}\left(|V|(d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right) + 1\right).$$

Then (6.32) gives $|L_t(i) - t\pi_i| \le A_t$ for all sufficiently large t and every $i \in V$. Since $\Lambda_t = \log\left(|V|(t+1)(t+2)^{1+\delta}\right) = O(\log t)$ and $\alpha_t = \max\{1, \sqrt{\lambda\Lambda_t}\}$, we have $A_t = O(\sqrt{\log t}) = o(t)$. Hence, for every $i \in V$, $L_t(i) \ge t\pi_i - A_t$. Because $\pi_i > 0$ and $A_t = o(t)$, the right-hand side tends to $+\infty$, and therefore $L_t(i) \to \infty$ for every $i \in V$. Thus every state is visited infinitely often.

Next fix a directed edge (i, j) with $P_{ij} > 0$. Define

$$B_t(i,j) := P_{ij}A_t + (d_{\max}-1)\left(C_1 + 2\sqrt{\frac{\Lambda_t}{\lambda}} + \frac{C_0}{\alpha_t}\right).$$

By (6.33), $|N_t(i,j) - t\pi_i P_{ij}| \le B_t(i,j)$ for all sufficiently large t. Again $B_t(i,j) = O(\sqrt{\log t}) = o(t)$. Since $\pi_i > 0$ and $P_{ij} > 0$, we have $\pi_i P_{ij} > 0$, and therefore $N_t(i,j) \ge t\pi_i P_{ij} - B_t(i,j) \to +\infty$. Thus every directed edge (i, j) with $P_{ij} > 0$ is traversed infinitely often. □

6.4 Summary

This chapter developed a TSAW-based adaptive sampling dynamics for finite irreducible Markov kernels. The key mechanism is local row-wise balancing: each row penalizes directed transitions that have been empirically overused relative to the target proportions P_{ij}. A finite-alphabet Lyapunov argument gives uniform exponential control of the row-wise discrepancy, and a Poisson-equation argument then converts this local transition control into global occupation control.

The resulting almost-sure bounds

$$L_t(i) - t\pi_i = O(\sqrt{\log t}), \qquad N_t(i,j) - t\pi_i P_{ij} = O(\sqrt{\log t})$$

imply an empirical integral error of order $O(\sqrt{\log t}/t)$ for every bounded observable. Thus, the chapter gives a concrete example in which a history-dependent self-avoidance mechanism yields pathwise improvement over the usual empirical averaging scale for Markov-chain trajectories.

Bibliography

[1] Stefan Adams and Quirin Vogel. "Space-Time Random Walk Loop Measures". In: *Stochastic Processes and their Applications* 130.4 (2020), pp. 2086–2126. DOI: 10.1016/j.spa.2019.06.006.

[2] Louigi Addario-Berry and Bruce Reed. "Minima in Branching Random Walks". In: *Annals of Probability* 37.3 (2009), pp. 1044–1079.

[3] Robert J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer, 2007.

[4] Sinan G. Aksoy et al. "The Maximum Relaxation Time of a Random Walk". In: *Advances in Applied Mathematics* 101 (2018), pp. 1–14. DOI: 10.1016/j.aam.2018.07.002.

[5] David Aldous and James A. Fill. *Reversible Markov Chains and Random Walks on Graphs*. Unfinished monograph; recompiled as a single PDF in 2014. 2002. URL: https://www.stat.berkeley.edu/users/aldous/RWG/book.pdf.

[6] David Aldous and James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. Unfinished monograph. Available online at https://www.stat.berkeley.edu/~aldous/RWG/book.html. 2002.

[7] Noga Alon et al. "Non-Backtracking Random Walks Mix Faster". In: *Communications in Contemporary Mathematics* 9.4 (2007), pp. 585–603.

[8] Daniel J. Amit, Giorgio Parisi, and Luca Peliti. "Asymptotic Behavior of the "True" Self-Avoiding Walk". In: *Physical Review B* 27.3 (1983), pp. 1635–1645. DOI: 10.1103/PhysRevB.27.1635.

[9] Christophe Andrieu et al. "Non-Linear Markov Chain Monte Carlo". In: *ESAIM: Proceedings*. Vol. 19. 2007, pp. 79–84.

[10] Christophe Andrieu et al. "On Nonlinear Markov Chain Monte Carlo". In: *Bernoulli* 17.3 (2011), pp. 987–1014. DOI: 10.3150/10-BEJ307.

[11] Kazuoki Azuma. "Weighted sums of certain dependent random variables". In: *Tohoku Mathematical Journal* 19.3 (1967), pp. 357–367. DOI: 10.2748/tmj/1178243286.

[12] Roland Bauerschmidt, Tyler Helmuth, and Andrew Swan. “The Geometry of Random Walk Isomorphism Theorems”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 57.1 (2021), pp. 408–454. DOI: 10.1214/20-AIHP1083.

[13] David Belius. “Gumbel Fluctuations for Cover Times in the Discrete Torus”. In: *Probability Theory and Related Fields* 157.3–4 (2013), pp. 635–689. DOI: 10.1007/s00440-012-0467-7.

[14] Austin R. Benson, David F. Gleich, and Lek-Heng Lim. “The Spacey Random Walk: A Stochastic Process for Higher-Order Data”. In: *SIAM Review* 59.2 (2017), pp. 321–345. DOI: 10.1137/16M1074023.

[15] Marek Biskup. *Extrema of the Two-Dimensional Discrete Gaussian Free Field*. Lecture notes, PIMS-CRM Summer School in Probability, University of British Columbia. 2017. arXiv: 1712.09972 [math.PR].

[16] Christer Borell. “The Brunn–Minkowski Inequality in Gauss Space”. In: *Inventiones mathematicae* 30 (1975), pp. 207–216.

[17] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[18] Lawrence D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Vol. 9. Lecture Notes–Monograph Series. Institute of Mathematical Statistics, 1986.

[19] David Brydges, Jürg Fröhlich, and Thomas Spencer. “The Random Walk Representation of Classical Spin Systems and Correlation Inequalities”. In: *Communications in Mathematical Physics* 83.1 (1982), pp. 123–150. DOI: 10.1007/BF01947075.

[20] David Brydges, Remco van der Hofstad, and Wolfgang König. “Joint Density for the Local Times of Continuous-Time Markov Chains”. In: *The Annals of Probability* 35.4 (2007), pp. 1307–1332.

[21] David C. Brydges and Ismael Muñoz Maya. “An Application of Berezin Integration to Large Deviations”. In: *Journal of Theoretical Probability* 4.2 (1991), pp. 371–389. DOI: 10.1007/BF01258743.

[22] Siu On Chan, Qinghua Ding, and Sing Hei Li. “Learning and Testing Irreducible Markov Chains via the k-Cover Time”. In: *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*. Vol. 132. Proceedings of Machine Learning Research. 2021, pp. 458–480.

[23] Ashok K. Chandra et al. “The Electrical Resistance of a Graph Captures its Commute and Cover Times”. In: *Proceedings of the 21st ACM Symposium on Theory of Computing*. STOC. See also the journal version in Computational Complexity 6 (1997), 312–340. 1989, pp. 574–586. DOI: 10.1145/73007.73062.

[24] Guan-Yu Chen and Chii-Ruey Hwang. “Accelerating Reversible Markov Chains”. In: *Statistics & Probability Letters* 83.9 (2013), pp. 1956–1962.

[25] Don Coppersmith and Persi Diaconis. "Random Walk with Reinforcement". Unpublished manuscript. 1986.

[26] Codina Cotar and Daniel Thacker. "Edge- and Vertex-Reinforced Random Walks with Super-Linear Reinforcement on Infinite Graphs". In: *The Annals of Probability* 45.4 (2017), pp. 2655–2706.

[27] Pierre Del Moral and Arnaud Doucet. "Interacting Markov Chain Monte Carlo Methods for Solving Nonlinear Measure-Valued Equations". In: *The Annals of Applied Probability* 20.2 (2010), pp. 593–639. DOI: 10.1214/09-AAP628.

[28] Pierre Del Moral and Laurent Miclo. "On Convergence of Chains with Occupational Self-Interactions". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 460.2041 (2004), pp. 325–346. DOI: 10.1098/rspa.2003.1245.

[29] Pierre Del Moral and Laurent Miclo. "Self-Interacting Markov Chains". In: *Stochastic Analysis and Applications* 24.3 (2006), pp. 615–660.

[30] Amir Dembo et al. "Cover Times for Brownian Motion and Random Walks in Two Dimensions". In: *Annals of Mathematics* 160.2 (2004), pp. 433–464.

[31] Persi Diaconis and David Freedman. "De Finetti's Theorem for Markov Chains". In: *The Annals of Probability* 8.1 (1980), pp. 115–130. DOI: 10.1214/aop/1176994828.

[32] Persi Diaconis, Susan Holmes, and Radford M. Neal. "Analysis of a Nonreversible Markov Chain Sampler". In: *The Annals of Applied Probability* 10.3 (2000), pp. 726–752. DOI: 10.1214/aoap/1019487508.

[33] Persi Diaconis and Silke W. W. Rolles. "Bayesian analysis for reversible Markov chains". In: *The Annals of Statistics* 34.3 (2006), pp. 1270–1292. DOI: 10.1214/009053606000000290.

[34] Jian Ding. "Asymptotics of Cover Times via Gaussian Free Fields: Bounded-Degree Graphs and General Trees". In: *The Annals of Probability* 42.2 (2014), pp. 464–496.

[35] Jian Ding. "On Cover Times for 2D Lattices". In: *Electronic Journal of Probability* 17.45 (2012), pp. 1–18. DOI: 10.1214/EJP.v17-2206.

[36] Jian Ding, James R. Lee, and Yuval Peres. "Cover Times, Blanket Times, and Majorizing Measures". In: *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*. 2011, pp. 61–70.

[37] Jian Ding, James R. Lee, and Yuval Peres. "Cover Times, Blanket Times, and Majorizing Measures". In: *Annals of Mathematics* 175.3 (2012), pp. 1409–1471.

[38] Jian Ding and Ofer Zeitouni. "A Sharp Estimate for Cover Times on Binary Trees". In: *Stochastic Processes and their Applications* 122.5 (2012), pp. 2117–2143.

[39] Margherita Disertori, Thomas Spencer, and Martin R. Zirnbauer. "Quasi-Diffusion in a 3D Supersymmetric Hyperbolic Sigma Model". In: *Communications in Mathematical Physics* 300.2 (2010), pp. 435–486. DOI: 10.1007/s00220-010-1117-5.

[40] Vishwaraj Doshi, Jie Hu, and Do Young Eun. "Self-Repellent Random Walks on General Graphs: Achieving Minimal Sampling Variance via Nonlinear Markov Chains". In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. PMLR, 2023, pp. 8403–8423.

[41] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*. Carus Mathematical Monographs 22. Mathematical Association of America, 1984.

[42] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electrical Networks*. Vol. 22. Carus Mathematical Monographs. Mathematical Association of America, 1984.

[43] E. B. Dynkin. "Local Times and Quantum Fields". In: *Seminar on Stochastic Processes, 1983*. Ed. by E. Çinlar, K. L. Chung, and R. K. Getoor. Vol. 7. Progress in Probability and Statistics. Boston, MA: Birkhäuser Boston, 1984, pp. 69–83. DOI: 10.1007/978-1-4684-9169-2_5.

[44] Nathalie Eisenbaum. "Inequalities for Permanental Processes". In: *Electronic Journal of Probability* 18.99 (2013), pp. 1–15. DOI: 10.1214/EJP.v18-2919.

[45] Nathalie Eisenbaum. "Permanental Vectors with Nonsymmetric Kernels". In: *The Annals of Probability* 45.1 (2017), pp. 210–224. DOI: 10.1214/15-AOP1008.

[46] Nathalie Eisenbaum and Haya Kaspi. "On Permanental Processes". In: *Stochastic Processes and their Applications* 119.5 (2009), pp. 1401–1415.

[47] Nathalie Eisenbaum and Haya Kaspi. "On the Continuity of Local Times of Borel Right Markov Processes". In: *The Annals of Probability* 35.3 (2007), pp. 915–934. DOI: 10.1214/009117906000000980.

[48] Nathalie Eisenbaum and Franck Maunoury. "Existence Conditions of Permanental and Multivariate Negative Binomial Distributions". In: *The Annals of Probability* 45.6B (2017), pp. 4786–4820. DOI: 10.1214/17-AOP1179.

[49] Nathalie Eisenbaum et al. "A Ray–Knight Theorem for Symmetric Markov Processes". In: *Annals of Probability* 28.4 (2000), pp. 1781–1813.

[50] William F. Fagan, Finn McBride, and Leonid Koralov. "Reinforced Diffusions as Models of Memory-Mediated Animal Movement". In: *Journal of the Royal Society Interface* 20.200 (2023), p. 20220700.

[51] William Feller. *An Introduction to Probability Theory and Its Applications*. 3rd ed. Vol. 1. Wiley, 1968.

[52] Patrick J. Fitzsimmons and Jay Rosen. "Markovian Loop Soups: Permanental Processes and Isomorphism Theorems". In: *Electronic Journal of Probability* 19.60 (2014), pp. 1–30. DOI: 10.1214/EJP.v19-3255.

[53] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. 2nd ed. Wiley, 1999.

[54] Gersende Fort. "Central Limit Theorems for Stochastic Approximation with Controlled Markov Chain Dynamics". In: *ESAIM: Probability and Statistics* 19 (2015), pp. 60–80.

[55] Gersende Fort, Eric Moulines, and Philippe Priouret. "Convergence of Adaptive and Interacting Markov Chain Monte Carlo Algorithms". In: *The Annals of Statistics* 39.6 (2011), pp. 3262–3289. DOI: `10.1214/11-AOS938`.

[56] David F. Gleich, Lek-Heng Lim, and Yinyu Yu. "Multilinear PageRank". In: *SIAM Journal on Matrix Analysis and Applications* 36.4 (2015), pp. 1507–1541.

[57] Peter Grassberger. "Self-Trapping Self-Repelling Random Walks". In: *Physical Review Letters* 119 (2017), p. 140601. DOI: `10.1103/PhysRevLett.119.140601`.

[58] Wassily Hoeffding. "Probability inequalities for sums of bounded random variables". In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30. DOI: `10.1080/01621459.1963.10500830`.

[59] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. 2nd ed. Cambridge University Press, 2012.

[60] Lu-Jing Huang and Yong-Hua Mao. "On Some Mixing Times for Nonreversible Finite Markov Chains". In: *Journal of Applied Probability* 54.2 (2017), pp. 627–637.

[61] Michael S. Keane and Silke W. W. Rolles. "Edge-Reinforced Random Walk on Finite Graphs". In: *Infinite Dimensional Stochastic Analysis*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2000, pp. 217–234.

[62] Michael S. Keane and Silke W. W. Rolles. "Edge-reinforced random walk on finite graphs". In: *Infinite Dimensional Stochastic Analysis (Amsterdam, 1999)*. Ed. by Ph. Clément et al. Vol. 52. Verhandelingen, Afdeling Natuurkunde. Eerste Reeks. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2000, pp. 217–234.

[63] Gregory F. Lawler and José A. Trujillo Ferreras. "Random Walk Loop Soup". In: *Transactions of the American Mathematical Society* 359.2 (2007), pp. 767–787.

[64] Yves Le Jan. "Dynkin's Isomorphism without Symmetry". In: *Stochastic Analysis in Mathematical Physics*. Hackensack, NJ: World Scientific, 2008, pp. 43–53.

[65] Chul-Ho Lee, Xin Xu, and Do Young Eun. "Beyond Random Walk and Metropolis–Hastings Samplers: Why You Should Not Backtrack for Unbiased Graph Sampling". In: *arXiv preprint arXiv:1204.4140* (2012). arXiv: `1204.4140 [cs.SI]`.

[66] Carlos A. Léon and François Perron. "Optimal Hoeffding bounds for discrete reversible Markov chains". In: *The Annals of Applied Probability* 14.2 (2004), pp. 958–970. DOI: `10.1214/105051604000000170`.

[67] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. 2nd ed. Vol. 107. Mathematical Surveys and Monographs. With contributions by Elizabeth L. Wilmer. American Mathematical Society, 2017.

[68] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. 2nd ed. American Mathematical Society, 2017.

[69] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. 2nd ed. Vol. 107. Mathematical Surveys and Monographs. American Mathematical Society, 2017. DOI: 10.1090/mbk/107.

[70] Vlada Limic and Pierre Tarrès. "Attracting Edge and Strongly Edge Reinforced Walks". In: *The Annals of Probability* 35.5 (2007), pp. 1783–1806.

[71] Titus Lupu. "From Loop Clusters and Random Interlacements to the Free Field". In: *The Annals of Probability* 44.3 (2016), pp. 2117–2146.

[72] Russell Lyons and Shayan Oveis Gharan. "Sharp Bounds on Random Walk Eigenvalues via Spectral Embedding". In: *International Mathematics Research Notices* 2018.24 (2018), pp. 7555–7605. DOI: 10.1093/imrn/rnx082.

[73] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, 2016.

[74] Yi-An Ma et al. "Irreversible Samplers from Jump and Continuous Markov Processes". In: *arXiv preprint arXiv:1608.05973* (2016). arXiv: 1608.05973 [stat.ME].

[75] Michael B. Marcus and Jay Rosen. "A Sufficient Condition for the Continuity of Permanental Processes with Applications to Local Times of Markov Processes". In: *The Annals of Probability* 41.2 (2013), pp. 671–698.

[76] Michael B. Marcus and Jay Rosen. "Conditions for Permanental Processes to Be Unbounded". In: *The Annals of Probability* 45.4 (2017), pp. 2059–2086. DOI: 10.1214/16-AOP1091.

[77] Michael B. Marcus and Jay Rosen. *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press, 2006.

[78] Michael B. Marcus and Jay Rosen. *Markov Processes, Gaussian Processes, and Local Times*. Vol. 100. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[79] Qiaozhu Mei, Jian Guo, and Dragomir Radev. "DivRank: The Interplay of Prestige and Diversity in Information Networks". In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2010, pp. 1009–1018.

[80] Franz Merkl, Aniko Öry, and Silke W. W. Rolles. "The 'magic formula' for linearly edge-reinforced random walks". In: *Statistica Neerlandica* 62.3 (2008), pp. 345–363. DOI: 10.1111/j.1467-9574.2008.00402.x.

[81] Franz Merkl, Anna Öry, and Silke W. W. Rolles. "The 'Magic Formula' for Linearly Edge-Reinforced Random Walks". In: *Statistica Neerlandica* 62.3 (2008), pp. 345–363.

[82] Franz Merkl and Silke W. W. Rolles. “A Random Environment for Linearly Edge-Reinforced Random Walks on Infinite Graphs”. In: *Probability Theory and Related Fields* 138.1–2 (2007), pp. 157–176.

[83] Franz Merkl and Silke W. W. Rolles. “Bounding a random environment for two-dimensional edge-reinforced random walk”. In: *Electronic Journal of Probability* 13.18 (2008), pp. 530–565. DOI: `10.1214/EJP.v13-495`.

[84] Franz Merkl and Silke W. W. Rolles. “Linearly edge-reinforced random walks”. In: *Dynamics and Stochastics*. Vol. 48. IMS Lecture Notes–Monograph Series. Institute of Mathematical Statistics, 2006, pp. 66–77. DOI: `10.1214/074921706000000103`.

[85] Radford M. Neal. “Improving Asymptotic Variance of MCMC Estimators: Non-Reversible Chains Are Better”. In: *arXiv preprint math/0407281* (2004). arXiv: `math/0407281`.

[86] D. J. Newman and L. A. Shepp. “The Double Dixie Cup Problem”. In: *The American Mathematical Monthly* 67.1 (1960), pp. 58–61.

[87] Alexander M. Ostrowski and Olga Taussky. “On the Variation of the Determinant of a Positive Definite Matrix”. In: *Indagationes Mathematicae* 13 (1951), pp. 383–385.

[88] Daniel Paulin. “Concentration inequalities for Markov chains by Marton couplings and spectral methods”. In: *Electronic Journal of Probability* 20 (2015). DOI: `10.1214/EJP.v20-4039`.

[89] Michael J. Plank and Brian D. Sleeman. “A Reinforced Random Walk Model of Tumour Angiogenesis and Anti-Angiogenic Strategies”. In: *Mathematical Medicine and Biology* 20.2 (2003), pp. 135–181.

[90] Robert J. Plemmons. “M-Matrix Characterizations. I—Nonsingular M-Matrices”. In: *Linear Algebra and its Applications* 18.2 (1977), pp. 175–188. DOI: `10.1016/0024-3795(77)90073-8`.

[91] Silke W. W. Rolles. “How edge-reinforced random walk arises naturally”. In: *Probability Theory and Related Fields* 126.2 (2003), pp. 243–260. DOI: `10.1007/s00440-003-0260-8`.

[92] Christophe Sabot and Pierre Tarrès. “Edge-reinforced random walk, vertex-reinforced jump process and the supersymmetric hyperbolic sigma model”. In: *Journal of the European Mathematical Society* 17.9 (2015), pp. 2353–2378. DOI: `10.4171/JEMS/559`.

[93] Christophe Sabot, Pierre Tarrès, and Xiaolin Zeng. “The Vertex-Reinforced Jump Process and a Random Schrödinger Operator on Finite Graphs”. In: *The Annals of Probability* 45.6A (2017), pp. 3967–3986.

[94] Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[95] David Slepian. "The One-Sided Barrier Problem for Gaussian Noise". In: *The Bell System Technical Journal* 41.2 (1962), pp. 463–501. DOI: `10.1002/j.1538-7305.1962.tb02419.x`.

[96] Peter E. Smouse et al. "Stochastic Modelling of Animal Movement". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1550 (2010), pp. 2201–2211.

[97] Alain-Sol Sznitman. *Topics in Occupation Times and Gaussian Free Fields*. Vol. 16. Zurich Lectures in Advanced Mathematics. European Mathematical Society, 2012.

[98] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer, 2005.

[99] Pierre Tarrès. "Vertex-Reinforced Random Walk on **Z** Eventually Gets Stuck on Five Points". In: *The Annals of Probability* 32.3B (2004), pp. 2650–2701.

[100] Achille Thin et al. "Nonreversible MCMC from Conditional Invertible Transforms: A Complete Recipe with Convergence Guarantees". In: *arXiv preprint arXiv:2012.15550* (2020). arXiv: `2012.15550 [stat.ME]`.

[101] Bálint Tóth. "The "True" Self-Avoiding Walk with Bond Repulsion on $\mathbb{Z}$: Limit Theorems". In: *The Annals of Probability* 23.4 (1995), pp. 1523–1556. DOI: `10.1214/aop/1176987793`.

[102] Bálint Tóth and Bálint Vető. "Self-Repelling Random Walk with Directed Edges on $\mathbb{Z}$". In: *Electronic Journal of Probability* 13 (2008), pp. 1909–1926.

[103] B. S. Tsirelson, I. A. Ibragimov, and V. N. Sudakov. "Norms of Gaussian Sample Functions". In: *Proceedings of the Third Japan–USSR Symposium on Probability Theory*. Vol. 550. Lecture Notes in Mathematics. Springer, 1976, pp. 20–41.

[104] Konstantin S. Turitsyn, Michael Chertkov, and Marija Vucelja. "Irreversible Monte Carlo Algorithms for Efficient Sampling". In: *Physica D: Nonlinear Phenomena* 240.4–5 (2011), pp. 410–414. DOI: `10.1016/j.physd.2010.10.003`.

[105] David Vere-Jones. "Alpha-Permanents and Their Applications to Multivariate Gamma, Negative Binomial and Ordinary Binomial Distributions". In: *New Zealand Journal of Mathematics* 26 (1997), pp. 125–149.

[106] Stanislav Volkov. "Vertex-Reinforced Random Walk on Arbitrary Graphs". In: *The Annals of Probability* 29.1 (2001), pp. 66–91.

[107] Martin J. Wainwright and Michael I. Jordan. "Graphical Models, Exponential Families, and Variational Inference". In: *Foundations and Trends in Machine Learning* 1.1–2 (2008), pp. 1–305. DOI: `10.1561/2200000001`.

[108] R. A. Willoughby. "The Inverse M-Matrix Problem". In: *Linear Algebra and its Applications* 18.1 (1977), pp. 75–94.

[109] Wenyi Xiao et al. “Vertex-Reinforced Random Walk for Network Embedding”. In: *Proceedings of the 2020 SIAM International Conference on Data Mining*. 2020, pp. 595–603. DOI: `10.1137/1.9781611976236.67`.

[110] Alex Zhai. “Exponential Concentration of Cover Times”. In: *Electronic Journal of Probability* 23.32 (2018), pp. 1–22. DOI: `10.1214/18-EJP149`.

Appendix A

Proofs for Chapter 2

A.1 Absolute integrability for Lemma 2.3.2.

We adopt the setting and notation of Lemma 2.3.2. Write $\phi = u + \mathrm{i}v$ with $u, v \in \mathbb{R}^n$, and set

$$l(u,v) := (u_1^2 + v_1^2, \dots, u_n^2 + v_n^2) \in \mathbb{R}_+^n, \qquad r(u,v) := \sum_{j=1}^{n} (u_j^2 + v_j^2).$$

Then $\sum_{j=1}^n \phi_j \bar{\phi}_j = r(u,v)$ and $(2\pi\mathrm{i})^{-n}\, d\bar{\phi}\, d\phi = \pi^{-n}\, du\, dv$. Also, recalling that $S := \frac{1}{2}(Q + Q^\top)$, see (2.7), and using that S is symmetric, we have

$$\left| e^{\langle \phi, Q\bar{\phi}\rangle} \right| = e^{\langle \phi, S\bar{\phi}\rangle} = e^{\langle u, Su\rangle + \langle v, Sv\rangle}.$$

We first bound $T_i g$. Property $\mathfrak{G}$ implies that, for each $i \in [n]$, the function $T_i g$ has polynomial growth. Since $[n]$ is finite, there exists $m > 0$ such that for each $i \in [n]$ there are constants $a_i, b_i > 0$ with $|T_i g(\phi)| \le a_i \|\phi\|^{2m}$ whenever $\|\phi\| > b_i$. Taking maxima over i, and then enlarging the resulting constant to cover the compact set $\{\phi \in \mathbb{C}^n : \|\phi\| \le \max_i b_i\}$, we obtain a constant $C_g > 0$ such that

$$|T_i g(\phi)| \le C_g (1 + \|\phi\|^{2m}), \qquad \phi \in \mathbb{C}^n,\ i \in [n].$$

Since $\|\phi\|^2 = r(u,v)$, this becomes

$$|T_i g(u + \mathrm{i}v)| \le C_g (1 + r(u,v))^m, \qquad u, v \in \mathbb{R}^n,\ i \in [n].$$

Next we bound f. Since f has property $\mathfrak{F}$, applying the definition with exponent $m+2$ shows that for each $i \in [n]$ there are constants $a_i', b_i' > 0$ such that

$$|f(i,l)| \le a_i' \|l\|^{-(m+2)} \qquad \text{whenever } \|l\| > b_i'.$$

Using again that $[n]$ is finite and that all norms on $\mathbb{R}^n$ are equivalent, we find constants $A, B > 0$ such that

$$|f(i,l)| \le A \Big(\sum_{j=1}^{n} l_j \Big)^{-(m+2)} \qquad \text{whenever } \sum_{j=1}^{n} l_j > B,\ i \in [n].$$

Enlarging A if necessary to cover the compact set $\{l \in \mathbb{R}_+^n : \sum_j l_j \leq B\}$, we obtain $C_f > 0$ such that

$$|f(i,l)| \leq C_f \Big(1 + \sum_{j=1}^n l_j\Big)^{-(m+2)}, \qquad l \in \mathbb{R}_+^n,\ i \in [n].$$

Now let

$$I := \int_0^\infty \sum_i \int_{\mathbb{C}^n} \big|(T_i g)(\phi)\, f_t(i, \phi\bar{\phi})\big|\, (2\pi \mathrm{i})^{-n} \big|e^{\langle \phi, Q\bar{\phi}\rangle}\big|\, d\bar{\phi}\, d\phi\, dt.$$

Passing to (u,v)-coordinates, applying $|\mathbb{E}Z| \leq \mathbb{E}|Z|$, and using the two bounds above, we get

$$\begin{aligned}
I &= \int_0^\infty \sum_i \int_{\mathbb{R}^{2n}} \big|(T_i g)(u + \mathrm{i}v)\big|\, \big|\mathbb{E}_{i,l(u,v)}(f(X_t, \mathcal{L}_t))\big|\, \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv\, dt \\
&\leq \int_0^\infty \sum_i \int_{\mathbb{R}^{2n}} \big|(T_i g)(u + \mathrm{i}v)\big|\, \mathbb{E}_{i,l(u,v)} |f(X_t, \mathcal{L}_t)|\, \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv\, dt \\
&\leq C_g C_f \int_0^\infty \sum_i \int_{\mathbb{R}^{2n}} (1 + r(u,v))^m\, \mathbb{E}_{i,l(u,v)} \Big(1 + \sum_{j=1}^n \mathcal{L}_t^j\Big)^{-(m+2)} \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv\, dt.
\end{aligned}$$

Since the total local time increases at unit rate, we have $\sum_{j=1}^n \mathcal{L}_t^j = \sum_{j=1}^n l_j(u,v) + t = r(u,v) + t$. Hence

$$I \leq C_g C_f \int_0^\infty \sum_i \int_{\mathbb{R}^{2n}} (1 + r(u,v))^m (1 + r(u,v) + t)^{-(m+2)} \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv\, dt.$$

The integrand is nonnegative, so Tonelli's theorem applies. Therefore

$$\begin{aligned}
I &\leq C_g C_f \sum_i \int_{\mathbb{R}^{2n}} (1 + r(u,v))^m \left(\int_0^\infty (1 + r(u,v) + t)^{-(m+2)}\, dt\right) \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv \\
&= \frac{C_g C_f}{m+1} \sum_i \int_{\mathbb{R}^{2n}} (1 + r(u,v))^{-1} \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv \\
&\leq \frac{C_g C_f}{m+1} \sum_i \int_{\mathbb{R}^{2n}} \pi^{-n} e^{\langle u, Su\rangle + \langle v, Sv\rangle}\, du\, dv.
\end{aligned}$$

The last integral is finite by the same Gaussian integrability used in Lemma 2.3.2. Hence $I < \infty$.

Thus the integrand is absolutely integrable, and Fubini's theorem allows us to exchange the relevant integrals and the summation.

A.2 Interchange of Differentiation and Integration in Lemma 2.4.1.

Recall that

$$Q_\alpha := -(G_\alpha)^{-1}, \qquad d\mu_\alpha(\phi) := \frac{1}{(2\pi \mathrm{i})^n |G_\alpha|} e^{\langle \phi, Q_\alpha \bar{\phi}\rangle}\, d\bar{\phi}\, d\phi,$$

see equations (2.8) and (2.9). For every $\alpha \in [0,1]$, let

$$F(\alpha) := \mathbb{E}[f(\ell_\alpha)] = \int_{\mathbb{C}^n} f(\phi\bar{\phi})\, d\mu_\alpha(\phi) = \int_{\mathbb{C}^n} H(\alpha, \phi)\, d\bar{\phi}\, d\phi,$$

where

$$H(\alpha, \phi) := \frac{1}{(2\pi \mathrm{i})^n |G_\alpha|} f(\phi\bar{\phi}) e^{\langle \phi, Q_\alpha \bar{\phi}\rangle}.$$

Since $\alpha \mapsto G_\alpha$ is $C^1([0,1])$ entrywise, it is a C^1 map into $\mathbb{R}^{n\times n}$. The determinant map $A \mapsto \det(A)$ is polynomial in the entries of A, hence smooth, and the inversion map $A \mapsto A^{-1}$ is smooth on the open set of invertible matrices. Since each G_α is invertible, it follows that

$$\alpha \mapsto |G_\alpha|^{-1}, \qquad \alpha \mapsto G_\alpha^{-1}, \qquad \alpha \mapsto Q_\alpha$$

are all $C^1([0,1])$. Hence, for each fixed $\phi \in \mathbb{C}^n$, the map $\alpha \mapsto H(\alpha, \phi)$ is $C^1([0,1])$.

To obtain a uniformly integrable majorant, write

$$A_\alpha := -Q_\alpha = G_\alpha^{-1}, \qquad S_\alpha := \frac{A_\alpha + A_\alpha^T}{2}.$$

By Lemma A.4.1, S_α is positive definite for every $\alpha \in [0,1]$. Since $\alpha \mapsto A_\alpha$ is continuous, so is $\alpha \mapsto S_\alpha$. Since the smallest eigenvalue is continuous on the space of real symmetric matrices, the map

$$\alpha \mapsto \lambda_{\min}(S_\alpha)$$

is continuous on $[0,1]$. As each S_α is positive definite, $\lambda_{\min}(S_\alpha) > 0$ for all α, and compactness of $[0,1]$ yields

$$c := \min_{\alpha\in[0,1]} \lambda_{\min}(S_\alpha) > 0.$$

Now write $\phi = u + \mathrm{i}v$ with $u, v \in \mathbb{R}^n$. Since A_α is real,

$$\Re\langle \phi, A_\alpha \bar{\phi}\rangle = u^T A_\alpha u + v^T A_\alpha v = u^T S_\alpha u + v^T S_\alpha v \geq c(\|u\|^2 + \|v\|^2) = c\|\phi\|^2.$$

Equivalently,

$$\Re\langle \phi, Q_\alpha \bar{\phi}\rangle \leq -c\|\phi\|^2.$$

Therefore,

$$\left| e^{\langle \phi, Q_\alpha \bar{\phi}\rangle} \right| = e^{\Re\langle \phi, Q_\alpha \bar{\phi}\rangle} \leq e^{-c\|\phi\|^2}, \qquad \forall \alpha \in [0,1],\ \phi \in \mathbb{C}^n.$$

Next, because $\alpha \mapsto G_\alpha$, $\alpha \mapsto Q_\alpha$, and $\alpha \mapsto \frac{d}{d\alpha} G_\alpha$ are continuous on the compact interval $[0,1]$, there exist finite constants

$$M_0 := \sup_{\alpha\in[0,1]} |G_\alpha|^{-1}, \qquad M_1 := \sup_{\alpha\in[0,1]} \|Q_\alpha\|, \qquad M_2 := \sup_{\alpha\in[0,1]} \Big\|\frac{d}{d\alpha}G_\alpha\Big\|.$$

Using Jacobi's formula (see, e.g., Section 0.8.2 of [59]) and

$$\frac{d}{d\alpha}Q_\alpha = Q_\alpha\Big(\frac{d}{d\alpha}G_\alpha\Big)Q_\alpha,$$

we get

$$\frac{d}{d\alpha}|G_\alpha|^{-1} = |G_\alpha|^{-1}\mathrm{tr}\Big(Q_\alpha \frac{d}{d\alpha}G_\alpha\Big),$$

and

$$\partial_\alpha e^{\langle \phi, Q_\alpha\bar\phi\rangle} = e^{\langle \phi, Q_\alpha\bar\phi\rangle}\Big\langle \phi, Q_\alpha\Big(\frac{d}{d\alpha}G_\alpha\Big)Q_\alpha\bar\phi\Big\rangle.$$

Hence

$$\partial_\alpha H(\alpha,\phi) = \frac{1}{(2\pi \mathrm{i})^n|G_\alpha|} f(\phi\bar\phi)e^{\langle \phi, Q_\alpha\bar\phi\rangle}\Bigg(\mathrm{tr}\Big(Q_\alpha\frac{d}{d\alpha}G_\alpha\Big) + \Big\langle \phi, Q_\alpha\Big(\frac{d}{d\alpha}G_\alpha\Big)Q_\alpha\bar\phi\Big\rangle\Bigg).$$

Taking absolute values and using

$$\Big|\mathrm{tr}\Big(Q_\alpha\frac{d}{d\alpha}G_\alpha\Big)\Big| \le n\|Q_\alpha\|\,\Big\|\frac{d}{d\alpha}G_\alpha\Big\|,$$

and

$$\Big|\Big\langle \phi, Q_\alpha\Big(\frac{d}{d\alpha}G_\alpha\Big)Q_\alpha\bar\phi\Big\rangle\Big| \le \|Q_\alpha\|^2\Big\|\frac{d}{d\alpha}G_\alpha\Big\|\,\|\phi\|^2,$$

we obtain

$$|\partial_\alpha H(\alpha,\phi)| \le C\,|f(\phi\bar\phi)|(1+\|\phi\|^2)e^{-c\|\phi\|^2}, \qquad \forall \alpha\in[0,1],\ \phi\in\mathbb{C}^n,$$

for some constant $C<\infty$ independent of α and ϕ.

By the definition of subexponential growth,

$$\lim_{\|l\|_2\to\infty} \frac{\log(1+|f(l)|)}{\|l\|_2} = 0.$$

Fix $\varepsilon>0$. Then there exists $R_\varepsilon<\infty$ such that whenever $\|l\|_2 \ge R_\varepsilon$,

$$\frac{\log(1+|f(l)|)}{\|l\|_2} \le \varepsilon,$$

and hence

$$|f(l)| \le e^{\varepsilon\|l\|_2}, \qquad \|l\|_2 \ge R_\varepsilon.$$

Since f is continuous, it is bounded on the compact set $\{l \in \mathbb{R}_+^n : \|l\|_2 \le R_\varepsilon\}$; let

$$M_\varepsilon := \sup_{\|l\|_2 \le R_\varepsilon} |f(l)| < \infty.$$

Therefore, for all $l \in \mathbb{R}_+^n$,

$$|f(l)| \le C_\varepsilon e^{\varepsilon \|l\|_2}, \qquad C_\varepsilon := \max\{1, M_\varepsilon\}.$$

Applying this with $l = \phi\bar{\phi} = (|\phi_1|^2, \dots, |\phi_n|^2)$, and using

$$\|\phi\bar{\phi}\|_2 = \Big(\sum_{j=1}^n |\phi_j|^4\Big)^{1/2} \le \sum_{j=1}^n |\phi_j|^2 = \|\phi\|_2^2,$$

we obtain

$$|f(\phi\bar{\phi})| \le C_\varepsilon e^{\varepsilon \|\phi\|_2^2}.$$

Choose $\varepsilon > 0$ so small that $\varepsilon < c/2$. Then

$$|\partial_\alpha H(\alpha, \phi)| \le C \cdot C_\epsilon (1 + \|\phi\|_2^2) e^{-(c-\varepsilon)\|\phi\|_2^2}, \qquad \forall \alpha \in [0,1],\ \phi \in \mathbb{C}^n.$$

Since $c - \varepsilon > 0$ and $\mathbb{C}^n \simeq \mathbb{R}^{2n}$, the right-hand side belongs to $L^1(\mathbb{C}^n)$, because any polynomial times a non-degenerate Gaussian density is integrable.

By the Leibniz integral rule for differentiation under the integral sign (see, e.g., Theorem 2.27 in [53]), justified here by the dominated convergence theorem, the exchange of differentiation and integration in the proof of Lemma 2.4.1 is valid.

A.3 Nonnegativity of a certain integral

Consider G, which is an inverse M-matrix with a positive definite symmetric part. Let $A := G^{-1}$. Then A also has a positive definite symmetric part, as proved in Lemma A.4.1 in Section A.4. It follows that the complex measure $d\mu(\phi) := (2\pi\mathrm{i})^{-n} e^{-\langle \phi, A\bar{\phi}\rangle} d\bar{\phi} d\phi$ is integrable (see, e.g., Eq. (2.3)).

Lemma A.3.1. *Let G be an inverse M-matrix with a positive definite symmetric part. For any nonnegative function $f : \mathbb{R}_+^n \to \mathbb{R}_+$ with subexponential growth, the integral $\int_{\mathbb{C}^n} \phi_a \bar{\phi}_b f(\phi\bar{\phi}) d\mu(\phi)$ is finite and nonnegative for any $a, b \in [n]$.*

Proof. To see that the integral is nonnegative, note that

$$\begin{aligned}
\int_{\mathbb{C}^n} \phi_a \bar{\phi}_b f(\phi\bar{\phi}) d\mu(\phi) &= \frac{1}{(2\pi\mathrm{i})^n} \int_{\mathbb{C}^n} \phi_a \bar{\phi}_b f(\phi\bar{\phi}) e^{-\sum_{ij} A_{ij}\phi_i\bar{\phi}_j} d\bar{\phi} d\phi \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}_+^n} \int_{[0,2\pi)^n} \sqrt{l_a l_b} e^{\mathrm{i}(\theta_a - \theta_b)} f(l) e^{-\sum_{ij} A_{ij}\sqrt{l_i l_j} e^{\mathrm{i}(\theta_i - \theta_j)}} d\theta dl \\
&= \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} \left(\frac{1}{(2\pi)^n} \int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a - \theta_b)} e^{-\sum_{i\neq j} A_{ij}\sqrt{l_i l_j} e^{\mathrm{i}(\theta_i - \theta_j)}} d\theta \right) dl, \quad \text{(A.1)}
\end{aligned}$$

where we have used equation (2.1) in writing the second equality.

Moreover, we have

$$
\begin{aligned}
e^{-\sum_{i\neq j} A_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)}} &= \prod_{i\neq j} e^{-A_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)}} \\
&= \prod_{i\neq j}\sum_{k_{ij}\in\mathbb{N}} \frac{(-A_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)})^{k_{ij}}}{k_{ij}!} \\
&= \sum_{\{k_{ij}\in\mathbb{N},\, i\neq j\}}\prod_{i\neq j} \frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!} e^{\mathrm{i}k_{ij}(\theta_i-\theta_j)}.
\end{aligned}
$$

Fix some $l \in \mathbb{R}^n_+$. The inner integral in equation (A.1) becomes

$$
\begin{aligned}
&\frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)} e^{-\sum_{i\neq j} A_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)}} d\theta \\
&\quad = \frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)}\left(\sum_{\{k_{ij}\in\mathbb{N},\, i\neq j\}}\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}e^{\mathrm{i}k_{ij}(\theta_i-\theta_j)}\right)d\theta \\
&\quad = \frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)}\left(\sum_{\{k_{ij}\in\mathbb{N},\, i\neq j\}}\left(\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}\right)e^{\sum_{i\neq j}\mathrm{i}k_{ij}(\theta_i-\theta_j)}\right)d\theta \\
&\quad = \sum_{\{k_{ij}\in\mathbb{N},\, i\neq j\}}\left(\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}\right)\cdot\left(\frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)}e^{\sum_{i\neq j}\mathrm{i}k_{ij}(\theta_i-\theta_j)}d\theta\right).
\end{aligned}
$$

Notice that $\frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)}e^{\sum_{i\neq j}\mathrm{i}k_{ij}(\theta_i-\theta_j)}d\theta = \frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\sum_{i\neq j}\mathrm{i}r_i\theta_i}d\theta$ where $r_i := \sum_{j\neq i}(k_{ij}-k_{ji}) + \mathbf{1}(i=a) - \mathbf{1}(i=b)$. Since $\frac{1}{2\pi}\int_{[0,2\pi)} e^{\mathrm{i}r_i\theta_i}d\theta_i = \mathbf{1}(r_i=0)$, we know that $\frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)}e^{\sum_{i\neq j}\mathrm{i}k_{ij}(\theta_i-\theta_j)}d\theta = \mathbf{1}(r_i = 0, \forall i\in[n])$. Therefore, we conclude that

$$
\begin{aligned}
&\frac{1}{(2\pi)^n}\int_{[0,2\pi)^n} e^{\mathrm{i}(\theta_a-\theta_b)} e^{-\sum_{i\neq j} A_{ij}\sqrt{l_il_j}e^{\mathrm{i}(\theta_i-\theta_j)}} d\theta \\
&\qquad = \sum_{\{k_{ij}\in\mathbb{N},\, i\neq j\}}\left(\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}\right)\cdot\mathbf{1}(r_i=0,\forall i\in[n]) \\
&\qquad = \sum_{\{k_{ij}\in\mathbb{N},\, i\neq j:\, r_i(k)=0,\forall i\in[n]\}}\left(\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}\right).
\end{aligned}
$$

The original integral should be

$$
\begin{aligned}
&\int_{\mathbb{C}^n}\phi_a\bar{\phi}_b f(\phi\bar{\phi})d\mu(\phi) \\
&\quad = \int_{\mathbb{R}^n_+}\sqrt{l_al_b}f(l)e^{-\sum_i A_{ii}l_i}\cdot \sum_{\{k_{ij}\in\mathbb{N},\, i\neq j:\, r_i(k)=0,\forall i\in[n]\}}\left(\prod_{i\neq j}\frac{(-A_{ij})^{k_{ij}}(l_il_j)^{\frac{1}{2}k_{ij}}}{k_{ij}!}\right)dl \geq 0.
\end{aligned}
$$

Here, we used the sign property of M-matrices, which guarantees $-A_{ij} \geq 0, \forall i \neq j$. We now justify the series expansions and the integration-summation exchanges, critically using positive definiteness. Note that we have

$$\begin{aligned}
&\int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} \left(\frac{1}{(2\pi)^n} \int_{[0,2\pi)^n} \prod_{i \neq j} \sum_{k_{ij} \in \mathbb{N}} \left| e^{\mathrm{i}(\theta_a - \theta_b)} \frac{(-A_{ij} \sqrt{l_i l_j} e^{\mathrm{i}(\theta_i - \theta_j)})^{k_{ij}}}{k_{ij}!} \right| d\theta \right) dl \\
&\qquad = \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} \left(\frac{1}{(2\pi)^n} \int_{[0,2\pi)^n} \prod_{i \neq j} \sum_{k_{ij} \in \mathbb{N}} \frac{(-A_{ij})^{k_{ij}} (l_i l_j)^{\frac{1}{2} k_{ij}}}{k_{ij}!} d\theta \right) dl \\
&\qquad = \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} \left(\frac{1}{(2\pi)^n} \int_{[0,2\pi)^n} \prod_{i \neq j} e^{-A_{ij} \sqrt{l_i l_j}} d\theta \right) dl \\
&\qquad = \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} \prod_{i \neq j} e^{-A_{ij} \sqrt{l_i l_j}} dl \\
&\qquad = \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\sum_i A_{ii} l_i} e^{-\sum_{i \neq j} A_{ij} \sqrt{l_i l_j}} dl \\
&\qquad = \int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-\langle \sqrt{l}, A \sqrt{l} \rangle} dl.
\end{aligned}$$

Since A has positive definite symmetric part, we know there exists $c > 0$, such that

$$\langle \sqrt{l}, A\sqrt{l} \rangle \geq c \|\sqrt{l}\|^2 = c \sum_i l_i.$$

Therefore, the above integral should be less than $\int_{\mathbb{R}_+^n} \sqrt{l_a l_b} f(l) e^{-c \sum_i l_i} dl$. Now $f(l)$ having subexponential growth guarantees this integral to be finite. Then, the dominated convergence theorem guarantees the expansion and integration-summation exchanges. □

A.4 Miscellaneous supporting results

Lemma A.4.1. *Given a non-singular matrix $G \in M_n(\mathbb{R})$ with a positive definite symmetric part, its inverse $A := G^{-1}$ also has a positive definite symmetric part.*

Proof. Note that

$$\frac{A + A^T}{2} = \frac{G^{-1} + G^{-T}}{2} = G^{-T} \frac{G + G^T}{2} G^{-1}.$$

Now for any $x \in \mathbb{R}^n$ such that $x \neq 0$, we have

$$\left\langle x, \frac{A + A^T}{2} x \right\rangle = \left\langle x, G^{-T} \frac{G + G^T}{2} G^{-1} x \right\rangle = \left\langle G^{-1} x, \frac{G + G^T}{2} G^{-1} x \right\rangle > 0,$$

where the last step used the assumption that G has a positive definite symmetric part. Therefore, we conclude that A also has a positive definite symmetric part. □

Appendix B

Proofs for Chapter 3

B.1 Proof of Lemma 3.3.1

Proof of Lemma 3.3.1. Recall $R := \max_{u,v\in V} r(u,v)$, where $r(\cdot,\cdot)$ denotes effective resistance in the unit-resistance network associated with G. For any $u, v \in V$, the *commute time identity* states that

$$\mathbb{E}_u[\tau_{\text{hit}}(v)] + \mathbb{E}_v[\tau_{\text{hit}}(u)] \;=\; 2|E|\, r(u,v). \tag{B.1}$$

See, e.g., [41, Ch. 9], [6, Ch. 10], or [68, Ch. 10].

By (B.1) and nonnegativity,

$$\mathbb{E}_u[\tau_{\text{hit}}(v)] \;\leq\; 2|E|\, r(u,v) \qquad \text{for all } u, v \in V.$$

Taking the maximum over u, v yields

$$t_{\text{hit}} = \max_{u,v} \mathbb{E}_u[\tau_{\text{hit}}(v)] \;\leq\; 2|E| \max_{u,v} r(u,v) = 2|E|R.$$

Let $(u^\star, v^\star)$ realize the maximum resistance: $r(u^\star, v^\star) = R$. Then (B.1) gives

$$\mathbb{E}_{u^\star}[\tau_{\text{hit}}(v^\star)] + \mathbb{E}_{v^\star}[\tau_{\text{hit}}(u^\star)] = 2|E|R,$$

so at least one of the two expectations is $\geq |E|R$. Hence $t_{\text{hit}} \geq |E|R$. Combining the bounds gives $|E|R \leq t_{\text{hit}} \leq 2|E|R$. □

B.2 Proof of Lemma 3.3.2

Proof of Lemma 3.3.2. Let $v \in V$ satisfy $d(v) = d_{\min}$, and let u be any neighbor of v. We claim that $r(u,v) \geq 1/d(v)$. Since $R \geq r(u,v)$, this implies $R \geq 1/d_{\min}$, i.e. $d_{\min}R \geq 1$.

Indeed, we can see that $r(u,v) \geq 1/d(v)$ for all $v \in V$ and all neighbors u of v. To see this, construct a new network G' from G by shorting (identifying) all neighbors of v into a single vertex $u^\star$, and in particular identifying u with $u^\star$. In G', the only edges incident to v are the $d(v)$ unit resistors connecting v to $u^\star$, now in parallel. Hence

$$r_{G'}(u^\star, v) = \frac{1}{d(v)}.$$

By Rayleigh's monotonicity principle (shorting vertices can only decrease effective resistance; see, e.g., [41, Ch. 2] or [73, Ch. 9]), we have

$$r_G(u,v) \;\geq\; r_{G'}(u^\star, v) \;=\; \frac{1}{d(v)}.$$

□

B.3 Proof of the claim $R \leq 2\pi\, M^2$

Proof. Recall that $\eta := (\eta(v), v \in V)$ is the pinned Gaussian free field, pinned by setting $\eta(v_0) = 0$, and $M := \mathbb{E}[\max_{v\in V} \eta(v)]$, which does not depend on the choice of v_0. Since η is a centered Gaussian vector, it is symmetric in law: $\eta \overset{d}{=} -\eta$. In particular,

$$\mathbb{E}\Big[\min_v \eta(v)\Big] = -\mathbb{E}\Big[\max_v \eta(v)\Big] = -M,$$

and therefore

$$\mathbb{E}\Big[\max_{u,v\in V} |\eta(u) - \eta(v)|\Big] = \mathbb{E}\Big[\max_v \eta(v) - \min_v \eta(v)\Big] = 2M.$$

For any integrable family of random variables $\{Z_i\}$ one has $\mathbb{E}[\max_i Z_i] \geq \max_i \mathbb{E}[Z_i]$. Applying this to $Z_{u,v} := |\eta(u) - \eta(v)|$ gives

$$2M = \mathbb{E}\Big[\max_{u,v} |\eta(u) - \eta(v)|\Big] \geq \max_{u,v} \mathbb{E}\big[|\eta(u) - \eta(v)|\big].$$

For fixed u, v, $\eta(u) - \eta(v) \sim \mathcal{N}(0, r(u,v))$, hence $\mathbb{E}[|\eta(u) - \eta(v)|] = \sqrt{2/\pi}\,\sqrt{r(u,v)}$. Therefore,

$$2M \;\geq\; \max_{u,v} \sqrt{\frac{2}{\pi}}\,\sqrt{r(u,v)} = \sqrt{\frac{2}{\pi}}\,\sqrt{R}.$$

Squaring yields

$$M^2 \;\geq\; \frac{1}{2\pi} R, \qquad \text{i.e.} \qquad R \;\leq\; 2\pi\, M^2.$$

□

B.4 A one-sided first-cover lower bound for full b-ary trees

Lemma B.4.1 (Ray–Knight/Zhai lower-bound)**.** *Let $G = (V, E)$ be a finite connected graph rooted at v_0, and let $(\eta_v)_{v\in V}$ be the Gaussian free field pinned at v_0. Write $M := \max_{v\in V} \eta_v$. Then, for every $t > 0$,*

$$\mathbb{P}_{v_0}\big(\tau_{\text{cov}} < \tau_{\text{inv}}(t)\big) \leq \mathbb{P}\big(M \leq \sqrt{2t}\big).$$

Consequently, if m_h, a_h are deterministic sequences with $a_h \to \infty$ and

$$\mathbb{P}(M \geq m_h - a_h) \to 1,$$

then

$$\mathbb{P}_{v_0}\left(\tau_{\text{cov}} > \tau_{\text{inv}}\left(\frac{1}{2}(m_h - 2a_h)^2\right)\right) \to 1.$$

Proof. We use the stochastic domination consequence of the generalized second Ray–Knight theorem, in the form used by Zhai [110, Theorem 3.1]. It gives

$$\left(\sqrt{L^v_{\tau_{\text{inv}}(t)}}\right)_{v\in V} \preceq \left(\frac{1}{\sqrt{2}}(\eta_v + \sqrt{2t})_+\right)_{v\in V},$$

where $x_+ = \max\{x, 0\}$, and $\preceq$ denotes stochastic domination for increasing events.

Since

$$\{\tau_{\text{cov}} < \tau_{\text{inv}}(t)\} \subseteq \{L^v_{\tau_{\text{inv}}(t)} > 0 \text{ for all } v \in V\},$$

and the event on the right is increasing in the local-time field, the stochastic domination gives

$$\mathbb{P}_{v_0}\big(\tau_{\text{cov}} < \tau_{\text{inv}}(t)\big) \leq \mathbb{P}\left((\eta_v + \sqrt{2t})_+ > 0 \text{ for all } v \in V\right).$$

The event on the right is

$$\{\eta_v > -\sqrt{2t} \text{ for all } v \in V\} = \{\min_{v\in V} \eta_v > -\sqrt{2t}\}.$$

Since the pinned GFF is centered and symmetric,

$$\mathbb{P}\left(\min_{v\in V} \eta_v > -\sqrt{2t}\right) = \mathbb{P}\left(\max_{v\in V} \eta_v < \sqrt{2t}\right) \leq \mathbb{P}(M \leq \sqrt{2t}).$$

This proves the first claim.

For the consequence, take $t = \frac{1}{2}(m_h - 2a_h)^2$. Then $\sqrt{2t} = m_h - 2a_h$, and hence

$$\mathbb{P}_{v_0}\left(\tau_{\text{cov}} < \tau_{\text{inv}}\left(\frac{1}{2}(m_h - 2a_h)^2\right)\right) \leq \mathbb{P}(M \leq m_h - 2a_h) \leq \mathbb{P}(M < m_h - a_h) \to 0.$$

This proves the consequence. $\square$

Lemma B.4.2 (One-sided first-cover lower bound for full b-ary trees)**.** *Fix $b \geq 2$. Let $T_{b,h}$ be the full b-ary tree of height h, rooted at v_0. Then there exists $C_b < \infty$ such that*

$$\mathbb{P}_{v_0}\left(\tau_{\mathrm{cov}} \geq 2|E|(\log b)h^2 - C_b|E|h\log h\right) \to 1$$

as $h \to \infty$.

Proof. Let $(\eta_v)_{v\in V}$ be the Gaussian free field on $T_{b,h}$ pinned at the root v_0, and set $M_h := \max_{v\in V}\eta_v$. Since $M_h \geq \max_{v\in L_h}\eta_v$, it is enough to lower-bound the maximum over the leaves. Restricted to L_h, the pinned GFF is a Gaussian b-ary branching random walk of height h, with i.i.d. $N(0,1)$ increments along edges. We apply Addario-Berry–Reed [2, Theorem 3] to the reflected branching random walk $(-\eta_v)_{v\in L_h}$. In their notation, M_h^- denotes the minimum at generation h, and here

$$M_h^- := \min_{v\in L_h}(-\eta_v) = -\max_{v\in L_h}\eta_v.$$

The step distribution of the reflected walk is still $N(0,1)$, so

$$\Lambda(t) := \log \mathbb{E}e^{tX} = \frac{t^2}{2}.$$

Taking $t = -\sqrt{2\log b} < 0$, we have

$$t\Lambda'(t) - \Lambda(t) = t^2 - \frac{t^2}{2} = \frac{t^2}{2} = \log b, \qquad \Lambda'(t) = t = -\sqrt{2\log b}.$$

Thus Addario-Berry–Reed [2, Theorem 3] gives

$$\mathbb{E}M_h^- = -\sqrt{2\log b}\,h + \frac{3}{2\sqrt{2\log b}}\log h + O_b(1),$$

and exponential tails around the mean. Therefore, choosing $C_b < \infty$ large enough,

$$\mathbb{P}\left(M_h^- \leq -\sqrt{2\log b}\,h + C_b\log h\right) \to 1.$$

Since $M_h^- = -\max_{v\in L_h}\eta_v$, this is equivalent to

$$\mathbb{P}\left(\max_{v\in L_h}\eta_v \geq \sqrt{2\log b}\,h - C_b\log h\right) \to 1.$$

Finally, $M_h \geq \max_{v\in L_h}\eta_v$, so

$$\mathbb{P}\left(M_h \geq \sqrt{2\log b}\,h - C_b\log h\right) \to 1.$$

Applying Lemma B.4.1 with

$$m_h = \sqrt{2\log b}\,h, \qquad a_h = C_b\log h,$$

and increasing C_b if necessary, gives

$$\mathbb{P}_{v_0}\left(\tau_{\mathrm{cov}} > \tau_{\mathrm{inv}}(t_-)\right) \to 1, \qquad t_- := \left(\sqrt{\log b}\, h - C_b \log h\right)^2.$$

In particular,

$$t_- = (\log b)h^2 - O_b(h \log h).$$

It remains to translate inverse local time into ordinary time. By Lemma 3.3.6, for the full b-ary tree,

$$\mathrm{var}_{v_0}(\tau_{\mathrm{inv}}(t)) \leq C_b |E|^2 t.$$

Also $\mathbb{E}_{v_0}\tau_{\mathrm{inv}}(t) = 2|E|t$. Applying Chebyshev's inequality at $t = t_-$, and using $t_- = O_b(h^2)$, gives

$$\mathbb{P}_{v_0}\left(|\tau_{\mathrm{inv}}(t_-) - 2|E|t_-| \geq (\log h)|E|\sqrt{t_-}\right) \leq \frac{C_b}{(\log h)^2} \to 0.$$

Since $\sqrt{t_-} = O_b(h)$, it follows that

$$\tau_{\mathrm{inv}}(t_-) \geq 2|E|t_- - C_b|E|h \log h$$

with probability tending to one.

Finally,

$$2|E|t_- = 2|E|\left(\sqrt{\log b}\, h - C_b \log h\right)^2 = 2|E|(\log b)h^2 - O_b(|E| h \log h).$$

Combining the preceding estimates and increasing C_b if necessary gives

$$\mathbb{P}_{v_0}\left(\tau_{\mathrm{cov}} \geq 2|E|(\log b)h^2 - C_b|E|h \log h\right) \to 1.$$

□

Appendix C

Proofs for Chapter 5

C.1 Tightness of the mutual-information upper bound on $I(W; X_0^T)$

In the special case of the n-star, the logarithmic upper bound from Lemma 5.2.1 is sharp up to the value of the constant. Let G be the n-star with center v_0 and leaves $\ell_1, \dots, \ell_n$, and assume first that the walk starts from the center v_0. Write $e_i := \{v_0, \ell_i\}$, let $a_i := a_{e_i}$, set $\alpha_i := a_i/2$, and let $\alpha_0 := \sum_{i=1}^n \alpha_i$. For the random environment W, define

$$P_i := \frac{W_{e_i}}{\sum_{j=1}^n W_{e_j}}, \qquad i = 1, \dots, n.$$

Then

$$I(W; X_0^T) = \frac{n-1}{2} \log T + O(1) \qquad (T \to \infty),$$

or equivalently,

$$\frac{1}{T} I(W; X_0^T) = \frac{n-1}{2} \frac{\log T}{T} + O(T^{-1}).$$

Proof. Let $M := \lceil T/2 \rceil$. Since the graph is a star and the walk starts from the center, the trajectory is deterministic except when it departs from v_0. Let $Y_r \in \{1, \dots, n\}$ be the leaf chosen on the r-th departure from v_0, for $r = 1, \dots, M$. Then X_0^T and $(Y_1, \dots, Y_M)$ determine each other, so

$$I(W; X_0^T) = I(W; Y_1, \dots, Y_M).$$

Moreover, the law of $(Y_1, \dots, Y_M)$ depends on W only through the vector

$$P = (P_1, \dots, P_n),$$

since P_i is the quenched probability of choosing leaf ℓ_i upon departure from the center. Hence $I(W; Y_1, \dots, Y_M) = I(P; Y_1, \dots, Y_M)$.

Now let $K_i(m) := \sum_{r=1}^{m} \mathbb{1}\{Y_r = i\}$. On the $(m+1)$-st departure from the center, the edge e_i has weight $a_i + 2K_i(m)$, because each previous excursion using e_i contributes two traversals of that edge. Therefore

$$\Pr(Y_{m+1} = i \mid Y_1, \ldots, Y_m) = \frac{a_i + 2K_i(m)}{\sum_{j=1}^{n} a_j + 2m} = \frac{\alpha_i + K_i(m)}{\alpha_0 + m}.$$

Thus $(Y_r)_{r \geq 1}$ is exactly a Pólya urn sequence with de Finetti measure $\mathrm{Dir}(\alpha_1, \ldots, \alpha_n)$. Equivalently, conditional on $P \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_n)$, the variables $Y_1, \ldots, Y_M$ are independent and satisfy

$$\Pr(Y_r = i \mid P) = P_i, \qquad i = 1, \ldots, n.$$

In other words, conditional on P, each Y_r has the categorical distribution on $\{1, \ldots, n\}$ with probability vector P. This is the standard de Finetti representation of Pólya's urn, or equivalently the usual Dirichlet–multinomial mixture representation; see, for example, [51, Chapter VII, Section 4].

Let $K_i := K_i(M)$ and $K := (K_1, \ldots, K_n)$. Since

$$\Pr(Y_1 = y_1, \ldots, Y_M = y_M \mid P) = \prod_{i=1}^{n} P_i^{K_i},$$

the sample enters only through the count vector K. Hence K is sufficient for P, and therefore

$$I(P; Y_1, \ldots, Y_M) = I(P; K).$$

Let $h(\cdot)$ denote differential entropy. By Dirichlet–multinomial conjugacy, $P \mid K \sim \mathrm{Dir}(\beta_1, \ldots, \beta_n)$, where $\beta_i := \alpha_i + K_i$ and $\beta_0 := \sum_i \beta_i = \alpha_0 + M$. Thus

$$I(P; K) = h(P) - \mathbb{E}[h(P \mid K)].$$

Since $P \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_n)$, the first term $h(P)$ is a finite constant independent of M.

For a general Dirichlet law $\mathrm{Dir}(\beta_1, \ldots, \beta_n)$, one has

$$h(\mathrm{Dir}(\beta)) = \log B(\beta) + (\beta_0 - n)\Psi(\beta_0) - \sum_{i=1}^{n} (\beta_i - 1)\Psi(\beta_i),$$

where

$$B(\beta) := B(\beta_1, \ldots, \beta_n) := \frac{\prod_{i=1}^{n} \Gamma(\beta_i)}{\Gamma(\beta_0)}, \quad \beta_0 := \sum_{i=1}^{n} \beta_i,$$

denotes the multivariate beta function. Using Stirling's formula $\log \Gamma(x) = x \log x - x - \frac{1}{2} \log x + O(1)$ and the digamma asymptotic $\Psi(x) = \log x + O(x^{-1})$, one computes

$$\begin{aligned}
\log B(\beta) &= \textstyle\sum_i \beta_i \log \beta_i - \beta_0 \log \beta_0 - \frac{1}{2} \sum_i \log \beta_i + \frac{1}{2} \log \beta_0 + O(1), \\
(\beta_0 - n)\Psi(\beta_0) &= \beta_0 \log \beta_0 - n \log \beta_0 + O(1), \\
-\textstyle\sum_i (\beta_i - 1)\Psi(\beta_i) &= -\textstyle\sum_i \beta_i \log \beta_i + \sum_i \log \beta_i + O(1).
\end{aligned}$$

Summing these three lines and simplifying yields, uniformly over all $\beta_i \geq \alpha_i > 0$,

$$h(\mathrm{Dir}(\beta)) = -\frac{n-1}{2}\log\beta_0 + \frac{1}{2}\sum_{i=1}^{n}\log\frac{\beta_i}{\beta_0} + O(1).$$

Applying this with $\beta_i = \alpha_i + K_i$ yields

$$\mathbb{E}[h(P \mid K)] = -\frac{n-1}{2}\log(\alpha_0 + M) + \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\log\frac{\alpha_i + K_i}{\alpha_0 + M}\right] + O(1).$$

Set $Z_{M,i} := (\alpha_i + K_i)/(\alpha_0 + M)$. Since $P \mid K \sim \mathrm{Dir}(\alpha_1 + K_1, \ldots, \alpha_n + K_n)$, the posterior mean of P_i is $Z_{M,i} = \mathbb{E}[P_i \mid K]$. In particular $0 < Z_{M,i} \leq 1$, so $\log Z_{M,i} \leq 0$. On the other hand, by concavity of log,

$$\log Z_{M,i} = \log\mathbb{E}[P_i \mid K] \geq \mathbb{E}[\log P_i \mid K].$$

Taking expectations gives $\mathbb{E}[\log Z_{M,i}] \geq \mathbb{E}[\log P_i]$.

Since the marginal law of P_i is $\mathrm{Beta}(\alpha_i, \alpha_0 - \alpha_i)$, we have $\mathbb{E}[\log P_i] = \Psi(\alpha_i) - \Psi(\alpha_0) > -\infty$. Therefore

$$\Psi(\alpha_i) - \Psi(\alpha_0) \leq \mathbb{E}[\log Z_{M,i}] \leq 0,$$

so $\mathbb{E}[\log Z_{M,i}] = O(1)$ uniformly in M. It follows that

$$\mathbb{E}[h(P \mid K)] = -\frac{n-1}{2}\log M + O(1).$$

Finally, since $h(P)$ is constant,

$$I(W; X_0^T) = I(P; K) = h(P) - \mathbb{E}[h(P \mid K)] = \frac{n-1}{2}\log M + O(1).$$

Because $M = \lceil T/2 \rceil$, we have $\log M = \log T + O(1)$, and hence

$$I(W; X_0^T) = \frac{n-1}{2}\log T + O(1).$$

If the walk starts from a leaf rather than from the center, then $X_1 = v_0$ deterministically, so $I(W; X_0^T) = I(W; X_1^T)$, and the same argument applies after shifting time by one step. □

C.2 Proof of Proposition 5.4.0.1

Proof. Fix a path x_0^T, and write $N := N(x_0^T)$ for its undirected edge-count vector and $x_T := x_T(x_0^T)$ for its endpoint. Let N_v denote the number of departures from v along the path. By the quenched likelihood formula,

$$\mathbb{P}_{\mathrm{srw}}^{v_0,w}(X_0^T = x_0^T) = \prod_{e\in E} w_e^{N_e} \prod_{v\in V} w_v^{-N_v}.$$

Multiplying this by the magic-formula density of $\mu_{v_0,A}$, we obtain

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)\,\mu_{v_0,A}(dw) \propto \frac{w_{v_0}^{1/2}\prod_{e\in E} w_e^{a_e-1+N_e}}{\prod_{v\in V} w_v^{(a_v+1)/2+N_v}}\sqrt{\tau(w)}\,dw_{-e_0}.$$

Now set $a'_e := a_e + N_e$, so that $a'_v := \sum_{e\ni v} a'_e = a_v + \sum_{e\ni v} N_e$. Writing $d_v := \sum_{e\ni v} N_e$, we have the standard identity

$$d_v = 2N_v + \mathbb{1}(v = x_T) - \mathbb{1}(v = v_0),$$

because each departure from v contributes 1 to d_v, each arrival contributes 1, and along a path started at v_0 and ending at x_T, arrivals and departures differ by $\mathbb{1}(v = x_T) - \mathbb{1}(v = v_0)$. Hence

$$\frac{a_v+1}{2} + N_v = \frac{a_v + d_v + 1 - \mathbb{1}(v = x_T) + \mathbb{1}(v = v_0)}{2} = \frac{a'_v + 1 - \mathbb{1}(v = x_T) + \mathbb{1}(v = v_0)}{2}.$$

Substituting this into the previous display gives

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)\,\mu_{v_0,A}(dw) \propto \frac{w_{v_0}^{1/2}\prod_{e\in E} w_e^{a'_e-1}}{\prod_{v\in V} w_v^{(a'_v+1-\mathbb{1}(v=x_T)+\mathbb{1}(v=v_0))/2}}\sqrt{\tau(w)}\,dw_{-e_0}.$$

The extra vertex factor simplifies because

$$\frac{w_{v_0}^{1/2}}{\prod_{v\in V} w_v^{(-\mathbb{1}(v=x_T)+\mathbb{1}(v=v_0))/2}} = w_{x_T}^{1/2}.$$

Therefore

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)\,\mu_{v_0,A}(dw) \propto \frac{w_{x_T}^{1/2}\prod_{e\in E} w_e^{a'_e-1}}{\prod_{v\in V} w_v^{(a'_v+1)/2}}\sqrt{\tau(w)}\,dw_{-e_0}.$$

This is exactly the magic-formula density of $\mu_{x_T,A+N}$, up to normalization. Restoring the normalizing constants yields

$$\mathbb{P}^{v_0,w}_{\mathrm{srw}}(X_0^T = x_0^T)\,\mu_{v_0,A}(dw) = \frac{Z_{x_T,A+N}}{Z_{v_0,A}}\,\mu_{x_T,A+N}(dw),$$

which is (5.19).

Integrating over w gives

$$\mathbb{P}^{v_0,A}_{\mathrm{errw}}(X_0^T = x_0^T) = \frac{Z_{x_T,A+N}}{Z_{v_0,A}},$$

which is (5.20). Dividing the previous identity by this path probability yields

$$\mu_{v_0,A}(dw \mid X_0^T = x_0^T) = \mu_{x_T,A+N}(dw),$$

which is (5.21).

Finally, for a path started at v_0, the endpoint is determined by the parity rule

$$\sum_{e \ni v} N_e \equiv \mathbb{1}(v = v_0) + \mathbb{1}(v = x_T) \pmod 2. \tag{C.1}$$

Thus x_T is a function of N, and the posterior depends on the realized path only through N. Hence

$$\mu_{v_0,A}(dw \mid N) = \mu_{x_T(N),A+N}(dw),$$

which is (5.22). □

C.3 Proof of Theorem 5.4.2

We retain the notation from Section 5.4: G is the n-star with center v_0, all initial edge weights are equal to $a > 0$, e is a fixed leaf edge, and $N_e(T)$, with an abuse of notation, denotes the number of the first T excursions that use e.

The first ingredient is the standard beta-binomial representation.

Lemma C.3.1 (Beta-binomial mixture)**.** *Let $b := (n-1)a$. There exists a random variable $X \sim \mathrm{Beta}(a, b)$ such that, conditional on $X = x$, one has $N_e(T) \mid X = x \sim \mathrm{Bin}(T, x)$. Consequently, for every $\gamma > 0$,*

$$\mathbb{E}\left[\frac{1}{N_e(T) + \gamma}\right] = \frac{1}{B(a,b)} \int_0^1 x^{a-1}(1-x)^{b-1}\, g_T(x)\, dx, \tag{C.2}$$

where $g_T(x) := \mathbb{E}[(\mathrm{Bin}(T, x) + \gamma)^{-1}]$.

Proof. The sequence of leaf edges chosen from the center is exactly a symmetric Pólya urn with n colors and initial weight a in each color. Hence the empirical frequencies are mixed by a $\mathrm{Dir}(a, \ldots, a)$ law, and the marginal frequency of one fixed edge is $\mathrm{Beta}(a, b)$ with $b = (n-1)a$. Conditional on this frequency $X = x$, the number of uses of e among the first T excursions is $\mathrm{Bin}(T, x)$. Integrating out X gives (C.2). □

For the proof of the theorem, we only need upper bounds on the binomial inverse moment.

Lemma C.3.2 (Upper bounds for the binomial inverse moment)**.** *Let $c_\gamma := \max\{1, \gamma^{-1}\}$. Then, for every $T \geq 2$,*

$$g_T(x) \leq c_\gamma \qquad \text{for } 0 < x \leq \frac{1}{2T},$$

and

$$g_T(x) \leq \frac{6c_\gamma}{Tx} \qquad \text{for } \frac{1}{2T} \leq x \leq 1.$$

Proof. Let $K \sim \mathrm{Bin}(T, x)$. We have $(K+\gamma)^{-1} \le c_\gamma$ pointwise, so $g_T(x) := \mathbb{E}[(K+\gamma)^{-1}] \le c_\gamma$ for all $x \in [0,1]$ and so, in particular, for $0 < x \le (2T)^{-1}$.

Now assume $x \ge (2T)^{-1}$, so $Tx \ge 1/2$. Write

$$g_T(x) = \mathbb{E}\left[\frac{1}{K+\gamma}\mathbf{1}_{\{K \le Tx/2\}}\right] + \mathbb{E}\left[\frac{1}{K+\gamma}\mathbf{1}_{\{K > Tx/2\}}\right].$$

On $\{K > Tx/2\}$, we have $(K+\gamma)^{-1} \le 2/(Tx)$. On the other hand, $(K+\gamma)^{-1} \le \gamma^{-1} \le c_\gamma$, so

$$g_T(x) \le c_\gamma \, \mathbb{P}(K \le Tx/2) + \frac{2}{Tx}.$$

Since $\mathbb{E}K = Tx$, the standard Chernoff–Hoeffding lower-tail bound [58] gives

$$\mathbb{P}(K \le Tx/2) \le e^{-Tx/8}.$$

Also, for all $y > 0$, one has $e^{-y/8} \le 4/y$ because $\sup_{y>0} y e^{-y/8} = 8/e < 4$. Applying this with $y = Tx$ yields $e^{-Tx/8} \le 4/(Tx)$, and therefore

$$g_T(x) \le \frac{4c_\gamma}{Tx} + \frac{2}{Tx} \le \frac{6c_\gamma}{Tx}.$$

□

We now prove the theorem with explicit constants. Write $b := (n-1)a$, set $M_b := 2^{(1-b)_+}$, and define

$$C^{(>)}_{a,n,\gamma} := \frac{c_\gamma M_b}{a\,2^a\,B(a,b)} + \frac{6c_\gamma M_b\,2^{1-a}}{(a-1)B(a,b)} + \frac{6c_\gamma\,2^{(2-a)_+}}{b\,B(a,b)},$$

for $a > 1$,

$$C^{(=)}_{n,\gamma} := \frac{c_\gamma M_b}{2\,B(1,n-1)} + \frac{6c_\gamma M_b}{B(1,n-1)} + \frac{12c_\gamma}{(n-1)B(1,n-1)},$$

for $a = 1$, and

$$C^{(<)}_{a,n,\gamma} := \frac{c_\gamma M_b}{a\,2^a\,B(a,b)} + \frac{6c_\gamma M_b\,2^{1-a}}{(1-a)B(a,b)} + \frac{6c_\gamma\,2^{(2-a)_+}}{b\,B(a,b)},$$

for $0 < a < 1$.

Proof of Theorem 5.4.2. Let

$$M_T := \mathbb{E}\left[\frac{1}{N_e(T)+\gamma}\right].$$

By Lemma C.3.1, we have

$$M_T = \frac{1}{B(a,b)}\left(\int_0^{1/(2T)} x^{a-1}(1-x)^{b-1} g_T(x)\,dx + \int_{1/(2T)}^1 x^{a-1}(1-x)^{b-1} g_T(x)\,dx\right).$$

Using Lemma C.3.2, we obtain

$$M_T \le \frac{c_\gamma}{B(a,b)} \int_0^{1/(2T)} x^{a-1}(1-x)^{b-1}\,dx + \frac{6c_\gamma}{T\,B(a,b)} \int_{1/(2T)}^{1} x^{a-2}(1-x)^{b-1}\,dx.$$

We estimate the two integrals separately. On $[0,1/2]$, we have $(1-x)^{b-1} \le M_b$, where $M_b = 2^{(1-b)_+}$. Hence

$$\int_0^{1/(2T)} x^{a-1}(1-x)^{b-1}\,dx \le M_b \int_0^{1/(2T)} x^{a-1}\,dx = \frac{M_b}{a\,2^a}\,T^{-a}.$$

For the second integral, split at $x = 1/2$. On $[1/(2T), 1/2]$, we still have $(1-x)^{b-1} \le M_b$, so

$$\int_{1/(2T)}^{1/2} x^{a-2}(1-x)^{b-1}\,dx \le M_b \int_{1/(2T)}^{1/2} x^{a-2}\,dx.$$

On $[1/2, 1]$, we have $x^{a-2} \le 2^{(2-a)_+}$, and therefore

$$\int_{1/2}^{1} x^{a-2}(1-x)^{b-1}\,dx \le 2^{(2-a)_+} \int_{1/2}^{1} (1-x)^{b-1}\,dx \le \frac{2^{(2-a)_+}}{b}.$$

We now distinguish the three regimes.

If $a > 1$, then

$$\int_{1/(2T)}^{1/2} x^{a-2}\,dx \le \int_0^{1/2} x^{a-2}\,dx = \frac{2^{1-a}}{a-1}.$$

Thus

$$M_T \le \frac{c_\gamma M_b}{a\,2^a\,B(a,b)}\,T^{-a} + \frac{6c_\gamma}{T\,B(a,b)} \left(\frac{M_b\,2^{1-a}}{a-1} + \frac{2^{(2-a)_+}}{b} \right).$$

Since $a > 1$ and $T \ge 1$, one has $T^{-a} \le T^{-1}$. Therefore

$$M_T \le C^{(>)}_{a,n,\gamma}\,T^{-1}.$$

If $a = 1$, then $b = n-1$ and

$$\int_{1/(2T)}^{1/2} x^{-1}\,dx = \log T.$$

Hence

$$M_T \le \frac{c_\gamma M_b}{2\,B(1,n-1)}\,T^{-1} + \frac{6c_\gamma}{T\,B(1,n-1)} \left(M_b \log T + \frac{2}{n-1} \right).$$

Since $\log T \le 1 + \log T$ and $1 \le 1 + \log T$, this implies

$$M_T \le C^{(=)}_{n,\gamma}\,\frac{1+\log T}{T}.$$

If $0 < a < 1$, then

$$\int_{1/(2T)}^{1/2} x^{a-2}\,dx = \frac{(1/(2T))^{a-1} - (1/2)^{a-1}}{1-a} \le \frac{2^{1-a}}{1-a}\,T^{1-a}.$$

Thus

$$M_T \le \frac{c_\gamma M_b}{a\,2^a\,B(a,b)}\,T^{-a} + \frac{6c_\gamma}{T\,B(a,b)}\left(\frac{M_b\,2^{1-a}}{1-a}\,T^{1-a} + \frac{2^{(2-a)_+}}{b}\right).$$

Since $0 < a < 1$ and $T \ge 1$, one has $T^{-1} \le T^{-a}$. Therefore

$$M_T \le C^{(<)}_{a,n,\gamma}\,T^{-a}.$$

This proves the theorem. □

C.4 Proof of Lemma 5.4.4

Proof. Fix a nonempty proper subset $F \subsetneq E$, and write

$$S := S(F).$$

Partition the edges of F into three classes:

$$\begin{aligned} F_2 &:= \{g \in F : \text{ both endpoints of } g \text{ belong to } S\}, \\ F_1 &:= \{g \in F : \text{ exactly one endpoint of } g \text{ belongs to } S\}, \\ F_0 &:= \{g \in F : \text{ no endpoint of } g \text{ belongs to } S\}. \end{aligned}$$

Then

$$F = F_0 \sqcup F_1 \sqcup F_2.$$

We have

$$\sum_{v \in S} a_v = 2a(F_2) + a(F_1).$$

Therefore

$$a(F) - \frac{1}{2}\sum_{v\in S} a_v = a(F_0) + \frac{1}{2}a(F_1). \tag{1}$$

Next,

$$\sum_{v\in S} b_v = \frac{1}{2}\sum_{v\in S} a_v + \frac{1}{2}|S \setminus \{v_0\}|,$$

so

$$A(F) = \Big(a(F) - \frac{1}{2}\sum_{v\in S} a_v\Big) + \frac{\kappa(G\setminus F) - 1 - |S\setminus\{v_0\}|}{2}. \tag{2}$$

We now estimate the two terms on the right-hand side.

First term. By (1),

$$a(F) - \frac{1}{2}\sum_{v\in S} a_v = a(F_0) + \frac{1}{2}a(F_1).$$

Since $a_g \geq \underline{a}$ for every edge,

$$a(F_0) + \frac{1}{2}a(F_1) \geq \underline{a}\,|F_0| + \frac{\underline{a}}{2}|F_1|. \tag{3}$$

We claim that $F_0 \cup F_1 \neq \varnothing$. Indeed, suppose by contradiction that $F_0 = F_1 = \varnothing$. Then every edge in F belongs to F_2, so both endpoints of every edge in F belong to S. Since every vertex in S has all its incident edges in F, there is no edge in $E \setminus F$ incident to any vertex of S. Hence every connected component of the subgraph (S, F) is also a connected component of G. Because F is nonempty, (S, F) has at least one edge, so G has a connected component contained in (S, F). Since $F \subsetneq E$, there is also at least one edge in $E \setminus F$, hence at least one vertex not in S. This contradicts the connectedness of G. Therefore $F_0 \cup F_1 \neq \varnothing$, and thus (3) implies

$$a(F) - \frac{1}{2}\sum_{v\in S} a_v \geq \frac{\underline{a}}{2}. \tag{4}$$

Second term. Every vertex in S becomes isolated in $G \setminus F$, because all incident edges of such a vertex belong to F and are removed. Hence $G \setminus F$ has at least $|S|$ connected components coming from these isolated vertices.

Since $F \subsetneq E$, there exists an edge in $E \setminus F$. The connected component of $G \setminus F$ containing that edge is different from all the isolated-vertex components coming from S. Therefore

$$\kappa(G \setminus F) \geq |S| + 1.$$

Consequently,

$$\kappa(G \setminus F) - 1 - |S \setminus \{v_0\}| \geq |S| - |S \setminus \{v_0\}| \geq 0. \tag{5}$$

Combining (2), (4), and (5), we obtain

$$A(F) \geq \frac{\underline{a}}{2}.$$

This proves the lemma. $\square$

Appendix D

Proofs for Chapter 6

D.1 Proofs for Lemma 6.2.3.

Proof. Write $S_r := \sum_{m=0}^{r-1} E_m$, with $S_0 := 0$, where the E_m are independent with $E_m \sim \mathrm{Exp}(\rho^m)$, so that $N(s) \geq r$ if and only if $S_r \leq s$.

To prove (i), fix $k \geq 0$ and set $r := i_*(s) + k$. Since $s \geq 1$, we have $i_*(s) \geq 1$ and so $r \geq 1$. For any $\theta > 0$, Chernoff's bound gives

$$\mathbb{P}(N(s) \geq r) = \mathbb{P}(S_r \leq s) \leq e^{\theta s} \prod_{m=0}^{r-1} \frac{\rho^m}{\rho^m + \theta}.$$

Choose $\theta := \rho^{i_*(s)-1}$. Then $\theta s \in [1, 1/\rho)$, hence $e^{\theta s} \leq e^{1/\rho} = C_1(\rho)$. Also, for every m we have $\rho^m/(\rho^m + \theta) \leq 1$, while for $m \geq i_*(s) - 1$,

$$\frac{\rho^m}{\rho^m + \theta} \leq \frac{\rho^m}{\theta} = \rho^{m-(i_*(s)-1)}.$$

Therefore

$$\mathbb{P}(N(s) \geq r) \leq C_1(\rho) \prod_{m=i_*(s)-1}^{r-1} \rho^{m-(i_*(s)-1)} = C_1(\rho)\, \rho^{\sum_{\ell=0}^{k} \ell} = C_1(\rho)\, \rho^{k(k+1)/2},$$

which proves (i).

For (ii), first suppose $1 \leq k \leq i_*(s) - 1$ and set $r := i_*(s) - k \geq 1$. For any $\theta \in (0, \rho^{r-1})$, Chernoff's bound yields

$$\mathbb{P}(N(s) \leq r) = \mathbb{P}(S_r \geq s) \leq e^{-\theta s} \prod_{m=0}^{r-1} \frac{\rho^m}{\rho^m - \theta}.$$

Take $\theta := \frac{1}{2}\rho^{r-1}$. Then

$$\prod_{m=0}^{r-1} \frac{\rho^m}{\rho^m - \theta} = \prod_{m=0}^{r-1} \frac{1}{1 - \frac{1}{2}\rho^{r-1-m}} \leq \prod_{m=0}^{\infty} \frac{1}{1 - \frac{1}{2}\rho^m} = C_2(\rho).^1$$

Moreover, since $\rho^{i_*(s)-1} s \in [1, 1/\rho)$, we have

$$\theta s = \frac{1}{2}\rho^{r-1} s = \frac{1}{2}\rho^{i_*(s)-1-k} s \in \left[\frac{1}{2}\rho^{-k}, \frac{1}{2}\rho^{-k-1}\right),$$

so $e^{-\theta s} \leq \exp(-\frac{1}{2}\rho^{-k}) \leq \exp(-\frac{1}{2}\rho^{-k+1})$. Hence

$$\mathbb{P}\big(N(s) \leq i_*(s) - k\big) \leq C_2(\rho) \exp\big(-\tfrac{1}{2}\rho^{-k+1}\big) \qquad \text{for } 1 \leq k \leq i_*(s) - 1.$$

If $k = i_*(s)$, then $N(s) \leq i_*(s) - k$ means $N(s) = 0$, so $\mathbb{P}(N(s) \leq i_*(s) - k) = \mathbb{P}(E_0 > s) = e^{-s}$. Since $s \geq \rho^{-i_*(s)+1}$, we get $e^{-s} \leq e^{-\rho^{-i_*(s)+1}} \leq C_2(\rho) \exp(-\frac{1}{2}\rho^{-i_*(s)+1})$, because $C_2(\rho) \geq 1$. If $k > i_*(s)$, the event is empty. This proves (ii).

For (iii), write $i_* := i_*(s)$. Since $N(s)$ is a nonnegative integer-valued random variable, we have the pointwise identity

$$N(s) = \sum_{r=1}^{\infty} \mathbf{1}\{N(s) \geq r\}.$$

Applying Tonelli's theorem yields[2]

$$\mathbb{E}[N(s)] = \sum_{r=1}^{\infty} \mathbb{P}(N(s) \geq r).$$

For the upper bound on the mean, split the sum at i_* and use (i):

$$\mathbb{E}[N(s)] = \sum_{r=1}^{i_*} \mathbb{P}(N(s) \geq r) + \sum_{r=i_*+1}^{\infty} \mathbb{P}(N(s) \geq r) \leq i_* + \sum_{k=1}^{\infty} \mathbb{P}(N(s) \geq i_* + k).$$

---

[1] This actually is $(\frac{1}{2}; \rho)_\infty$, where $(a; q)_\infty := \prod_{m=0}^{\infty}(1 - aq^m)$ denotes the q-Pochhammer symbol.

[2] Here we use Tonelli's theorem in the form

$$\mathbb{E}\left[\sum_{r=1}^{\infty} X_r\right] = \sum_{r=1}^{\infty} \mathbb{E}[X_r]$$

for any sequence of nonnegative random variables $(X_r)_{r\geq 1}$, where both sides are allowed to take the value $+\infty$. Applying this with $X_r = \mathbf{1}\{N(s) \geq r\}$ gives

$$\mathbb{E}[N(s)] = \sum_{r=1}^{\infty} \mathbb{P}(N(s) \geq r).$$

Therefore

$$\mathbb{E}[N(s)] \leq i_* + C_1(\rho)\sum_{k=1}^{\infty}\rho^{k/2} = i_* + \frac{C_1(\rho)\sqrt{\rho}}{1-\sqrt{\rho}} \leq i_* + \frac{C_1(\rho)}{1-\sqrt{\rho}}.$$

For the lower bound, we again split at i_*:

$$\mathbb{E}[N(s)] = \sum_{r=1}^{i_*}\mathbb{P}(N(s) \geq r) + \sum_{r=i_*+1}^{\infty}\mathbb{P}(N(s) \geq r).$$

The second term is nonnegative, so

$$\mathbb{E}[N(s)] \geq \sum_{r=1}^{i_*}\mathbb{P}(N(s) \geq r) = \sum_{r=1}^{i_*}\big(1-\mathbb{P}(N(s) < r)\big) = i_* - \sum_{r=1}^{i_*}\mathbb{P}(N(s) \leq r-1).$$

Reindexing with $k := i_* - r + 1$ gives

$$\sum_{r=1}^{i_*}\mathbb{P}(N(s) \leq r-1) = \sum_{k=1}^{i_*}\mathbb{P}(N(s) \leq i_* - k).$$

Hence

$$\mathbb{E}[N(s)] \geq i_* - \sum_{k=1}^{i_*}\mathbb{P}(N(s) \leq i_* - k).$$

Applying (ii) and then extending the finite sum to an infinite one, we obtain

$$\mathbb{E}[N(s)] \geq i_* - \sum_{k=1}^{\infty} C_2(\rho)\exp\big(-\tfrac{1}{2}\rho^{-k+1}\big).$$

Using $e^{-x} \leq 1/x$ for $x > 0$, we get

$$\exp\big(-\tfrac{1}{2}\rho^{-k+1}\big) \leq 2\rho^{k-1},$$

and therefore

$$\mathbb{E}[N(s)] \geq i_* - \frac{2C_2(\rho)}{1-\rho}.$$

Since $\big|i_*(s) - \log s/\log(1/\rho)\big| \leq 1$, combining the two bounds gives

$$\left|\mathbb{E}[N(s)] - \frac{\log s}{\log(1/\rho)}\right| \leq 1 + \frac{C_1(\rho)}{1-\sqrt{\rho}} + \frac{2C_2(\rho)}{1-\rho} = A_\rho.$$

To bound the variance, let $Y_s := |N(s) - i_*(s)|$. By (i) and (ii), for every $m \geq 1$,

$$\mathbb{P}(Y_s \geq m) \leq C_1(\rho)\rho^{m/2} + C_2(\rho)\exp\big(-\tfrac{1}{2}\rho^{-m+1}\big).$$

Since Y_s is nonnegative and integer-valued, we have the pointwise identity

$$Y_s^2 = \sum_{m=1}^{\infty} (2m-1)\mathbf{1}\{Y_s \geq m\}.$$

Applying Tonelli's theorem again gives

$$\mathbb{E}[Y_s^2] = \sum_{m=1}^{\infty} (2m-1)\,\mathbb{P}(Y_s \geq m).$$

Hence

$$\mathbb{E}[Y_s^2] \leq C_1(\rho) \sum_{m=1}^{\infty} (2m-1)\rho^{m/2} + C_2(\rho) \sum_{m=1}^{\infty} (2m-1)\exp\bigl(-\tfrac{1}{2}\rho^{-m+1}\bigr).$$

Using again $e^{-x} \leq 1/x$, the second sum is bounded by

$$2C_2(\rho) \sum_{m=1}^{\infty} (2m-1)\rho^{m-1}.$$

Note that for $|x| < 1$, one has

$$\sum_{m=1}^{\infty} (2m-1)x^m = 2\sum_{m=1}^{\infty} mx^m - \sum_{m=1}^{\infty} x^m = \frac{2x}{(1-x)^2} - \frac{x}{1-x} = \frac{x(1+x)}{(1-x)^2}.$$

Applying this with $x = \sqrt{\rho}$ and $x = \rho$ gives, respectively,

$$\sum_{m=1}^{\infty} (2m-1)\rho^{m/2} = \frac{\sqrt{\rho}(1+\sqrt{\rho})}{(1-\sqrt{\rho})^2}, \qquad \sum_{m=1}^{\infty} (2m-1)\rho^{m} = \frac{\rho(1+\rho)}{(1-\rho)^2}.$$

Therefore

$$\mathbb{E}[Y_s^2] \leq C_1(\rho)\frac{\sqrt{\rho}(1+\sqrt{\rho})}{(1-\sqrt{\rho})^2} + 2C_2(\rho)\frac{1+\rho}{(1-\rho)^2}.$$

Finally, since

$$\mathrm{Var}(N(s)) = \mathbb{E}\bigl[(N(s) - \mathbb{E}[N(s)])^2\bigr] \leq 2\,\mathbb{E}\bigl[(N(s) - i_*(s))^2\bigr] + 2\,(\mathbb{E}[N(s)] - i_*(s))^2,$$

the first term is bounded by twice the previous display, and the second is at most

$$2\left(\frac{C_1(\rho)}{1-\sqrt{\rho}} + \frac{2C_2(\rho)}{1-\rho}\right)^2.$$

Combining these bounds yields $\mathrm{Var}(N(s)) \leq B_\rho$. $\square$

D.2 Proof of Lemma 6.2.4.

Proof. Since $s_n = \log n + c$, we have $s_n \geq 1$ for all sufficiently large n. For each fixed n, the random variables $N_1(s_n), \ldots, N_n(s_n)$ are i.i.d., so $\mathbb{E}[K(s_n)] = n\,\mathbb{E}[N(s_n)]$ and $\mathrm{Var}(K(s_n)) = n\ \mathrm{Var}(N(s_n))$.

By Lemma 6.2.3, for every such n,

$$\left|\mathbb{E}[N(s_n)] - \frac{\log s_n}{\log(1/\rho)}\right| \leq A_\rho \qquad \text{and} \qquad \mathrm{Var}(N(s_n)) \leq B_\rho.$$

Multiplying the first bound by n and dividing by $n \log\log n$ yields

$$\left|\frac{\mathbb{E}[K(s_n)]}{n \log\log n} - c_\rho \frac{\log s_n}{\log\log n}\right| \leq \frac{A_\rho}{\log\log n}.$$

We now compare $\log s_n$ with $\log\log n$ explicitly. Since $s_n = \log n + c$, we can write

$$\log s_n = \log(\log n + c) = \log\log n + \log\left(1 + \frac{c}{\log n}\right).$$

Choose n large enough that $\log n \geq 2|c|$, so that $|c/\log n| \leq 1/2$. Then the elementary bound $|\log(1+u)| \leq 2|u|$ for $|u| \leq 1/2$ gives

$$|\log s_n - \log\log n| = \left|\log\left(1 + \frac{c}{\log n}\right)\right| \leq \frac{2|c|}{\log n}.$$

Dividing by $\log\log n$, we obtain

$$\left|\frac{\log s_n}{\log\log n} - 1\right| \leq \frac{2|c|}{\log n\ \log\log n}.$$

Fix $\varepsilon > 0$ and $\delta > 0$. Since $\log\log n \to \infty$ and $\log n\ \log\log n \to \infty$, there exists n_1 such that for all $n \geq n_1$,

$$\frac{A_\rho}{\log\log n} \leq \frac{\varepsilon}{4} \qquad \text{and} \qquad c_\rho \frac{2|c|}{\log n\ \log\log n} \leq \frac{\varepsilon}{4}.$$

For such n, the two previous displays imply

$$\left|\frac{\mathbb{E}[K(s_n)]}{n \log\log n} - c_\rho\right| \leq \left|\frac{\mathbb{E}[K(s_n)]}{n \log\log n} - c_\rho \frac{\log s_n}{\log\log n}\right| + c_\rho \left|\frac{\log s_n}{\log\log n} - 1\right| \leq \frac{\varepsilon}{2}.$$

Similarly,

$$\mathrm{Var}\left(\frac{K(s_n)}{n \log\log n}\right) = \frac{\mathrm{Var}(K(s_n))}{n^2(\log\log n)^2} = \frac{\mathrm{Var}(N(s_n))}{n(\log\log n)^2} \leq \frac{B_\rho}{n(\log\log n)^2}.$$

Since $n(\log\log n)^2 \to \infty$, there exists n_2 such that for all $n \geq n_2$,

$$\frac{4B_\rho}{\varepsilon^2\, n(\log\log n)^2} \leq \delta.$$

Now let $X_n := K(s_n)/(n\log\log n)$. For $n \geq \max\{n_1, n_2\}$, Chebyshev's inequality gives

$$\mathbb{P}\left(|X_n - \mathbb{E}[X_n]| > \frac{\varepsilon}{2}\right) \leq \frac{4\ \mathrm{Var}(X_n)}{\varepsilon^2} \leq \frac{4B_\rho}{\varepsilon^2\, n(\log\log n)^2} \leq \delta.$$

On the event $\{|X_n - \mathbb{E}[X_n]| \leq \varepsilon/2\}$, the bound $|\mathbb{E}[X_n] - c_\rho| \leq \varepsilon/2$ implies $|X_n - c_\rho| \leq \varepsilon$. Therefore

$$\mathbb{P}\left(\left|\frac{K(s_n)}{n\log\log n} - c_\rho\right| \leq \varepsilon\right) \geq 1 - \delta$$

for all $n \geq \max\{n_1, n_2\}$. This proves the lemma. $\square$