Nice Summer Scool Random walks and complex networks July 8-19, 2019.

Cutoff for random walks on trees and sparse random graphs

Course on random walks: Louigi Addario-Berry Anna Ben-Hamou Perla Sousi

Contents

1	Cut	off on trees	3
2	The configuration model		8
3	Cut	off for NBRW on regular random graphs	11
4	Cut	off for SRW on random regular graphs	16
5	Cut	off for NBRW on random graphs with given degrees	18
6	Con	paring NBRW and SRW: entropies on Galton-Watson trees	24
7	Exercises		31
	7.1	The product condition is not sufficient for cutoff for reversible chains	31
	7.2	Cutoff for the lazy RW on the hypercube	31
	7.3	Cutoff for the top-to-random shuffle	32
	7.4	Carne-Varopoulos upper bound	33
	7.5	Typical distance in random regular graphs	33
	7.6	Concentration with exchangeable pairs (proof of Lemma 5.2)	33
8	Exercises: correction		35
	8.1	The product condition is not sufficient for cutoff for reversible chains	35
	8.2	Cutoff for the lazy RW on the hypercube	35
	8.3	Cutoff for the top-to-random shuffle	37
	8.4	Carne-Varopoulos upper bound	38
	8.5	Typical distance in random regular graphs	39
	8.6	Concentration with exchangeable pairs (proof of Lemma 5.2)	39

1 Cutoff on trees

In this first section, we present a result of Basu, Hermon, and Peres [1] which states that the product condition is a sufficient condition for cutoff for random walks on trees. This follows from the characterization of cutoff in terms of concentration of hitting times of "worst" sets, given in the same paper.

Let us first recall some definitions. Let P be the transition matrix of an irreducible Markov chain $(X_t)_{t\geq 0}$ on a finite state space Ω , with stationary distribution π . The worst-case total variation distance to equilibrium at time t is defined as

$$\mathcal{D}(t) = \max_{x \in \Omega} \mathcal{D}_x(t)$$
, where $\mathcal{D}_x(t) = \|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}}$,

and, for $0 < \varepsilon < 1$, the ε -mixing time is defined as

$$t_{\text{MIX}}(\varepsilon) = \min\{t \ge 0, \ \mathcal{D}(t) \le \varepsilon\}$$

We set $t_{\text{MIX}} = t_{\text{MIX}}(1/4)$. We say that the chain is *reversible* if, for all $x, y \in \Omega$, $\pi(x)P(x, y) = \pi(y)P(y, x)$, and that it is *lazy* if, for all $x \in \Omega$, $P(x, x) \ge 1/2$. For an irreducible, reversible, lazy chain, let

$$1 = \lambda_1 > \lambda_2 \ge \cdots \ge \lambda_{|\Omega|} \ge 0$$

be the eigenvalues of P in decreasing order. The *relaxation time* is defined as

$$t_{\rm REL} = \frac{1}{1 - \lambda_2}$$

Mixing and relaxation times are related by the following inequality (see Levin and Peres [11, Theorems 12.4 and 12.5]):

$$(t_{\text{REL}} - 1) \log\left(\frac{1}{2\varepsilon}\right) \le t_{\text{MIX}}(\varepsilon) \le t_{\text{REL}} \log\left(\frac{1}{2\varepsilon\pi_{\min}}\right),$$
 (1.1)

where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$.

A sequence of chains (Ω_n, P_n, π_n) is said to exhibit *cutoff* if for all $0 < \varepsilon < 1$,

$$\frac{t_{\rm MIX}^{(n)}(\varepsilon)}{t_{\rm MIX}^{(n)}(1-\varepsilon)} \xrightarrow[n \to \infty]{} 1$$

Equivalently,

$$\mathcal{D}^{(n)}\left(c\,t_{\mathrm{MIX}}^{(n)}\right) \underset{n\to\infty}{\longrightarrow} \begin{cases} 1 & \text{if } c<1,\\ 0 & \text{if } c>1. \end{cases}$$

We say that the sequence has *cutoff window* w_n if $w_n = o(t_{\text{MIX}})$ and for any $\varepsilon \in]0,1[$, there exists $c_{\varepsilon} > 0$ such that

$$t_{\text{MIX}}^{(n)}(\varepsilon) - t_{\text{MIX}}^{(n)}(1-\varepsilon) \le c_{\varepsilon} w_n$$

From inequality (1.1), one easily deduces that a necessary condition for cutoff is

$$\frac{t_{\text{MIX}}^{(n)}}{t_{\text{REL}}^{(n)}} \xrightarrow[n \to \infty]{} +\infty.$$
(1.2)

This is known as the *product condition* and it is conjectured to be sufficient for a "large" class of Markov chains. In [1], it was shown to be sufficient for lazy weighted nearest-neighbor random walks on trees.

Theorem 1.1. Let P be the transition matrix of a lazy reversible Markov chain on a tree T = (V, E), with $|V| \ge 3$. Then, for all $\varepsilon \in [0, 1/4]$,

$$t_{\text{MIX}}(\varepsilon) - t_{\text{MIX}}(1-\varepsilon) \le 35\sqrt{\varepsilon^{-1}t_{\text{REL}}t_{\text{MIX}}}$$

In particular, if (P_n) is a sequence of such chains and if the product condition holds, then the sequence has a cutoff with window $w_n = \sqrt{t_{\text{REL}}^{(n)} t_{\text{MIX}}^{(n)}}$.

This result is a consequence of the relations between mixing times and hitting times, presented in previous lectures. For $A \subset \Omega$, let

$$\tau_A = \min\{t \ge 0, \ X_t \in A\}$$

be the hitting time of A. For $\alpha, \varepsilon \in [0, 1[$, let

$$\mathsf{hit}_{\alpha}(\varepsilon) = \min\{t \ge 0, \max_{x \in \Omega} \max_{A \subset \Omega, \, \pi(A) \ge \alpha} \mathbb{P}_{x}(\tau_{A} > t) \le \varepsilon\}.$$

For any reversible irreducible finite lazy chain and any $\varepsilon \in [0, 1/4]$, we have

$$\operatorname{hit}_{1/2}(3\varepsilon/2) - \lceil 2t_{\operatorname{REL}}\log(1/\varepsilon)\rceil \le t_{\operatorname{MIX}}(\varepsilon) \le \operatorname{hit}_{1/2}(\varepsilon/2) + \lceil t_{\operatorname{REL}}\log(4/\varepsilon)\rceil , \qquad (1.3)$$

and

$$\operatorname{hit}_{1/2}(1-\varepsilon/2) - \left\lceil 2t_{\operatorname{REL}}\log(1/\varepsilon) \right\rceil \le t_{\operatorname{MIX}}(1-\varepsilon) \le \operatorname{hit}_{1/2}(1-2\varepsilon) + \mathbb{1}_{\varepsilon > 1/18} \left\lceil \frac{1}{2}t_{\operatorname{REL}}\log(8) \right\rceil .$$
(1.4)

Before proving Theorem 1.1, let us state two results on hitting times that hold for any finite irreducible reversible chains (actually, the second one does not require reversibility). Recall that the conductance of $A \subset \Omega$, $A \neq \emptyset$ is defined as

$$\Phi(A) = \frac{1}{\pi(A)} \sum_{x \in A} \pi(x) \sum_{y \in A^c} P(x, y) = \mathbb{P}_{\pi_A}(X_1 \notin A),$$

where π_A is the distribution π conditioned on A, i.e. $\pi_A(z) = \pi(A)^{-1}\pi(z)\mathbb{1}_{z \in A}$.

Lemma 1.2. Let (Ω, P, π) be a finite irreducible reversible Markov chain and $A \subset \Omega$, $A \neq \emptyset$. Then, for all $t \ge 0$,

$$\mathbb{P}_{\pi_{A^c}}\left(\tau_A > t\right) \le \left(1 - \frac{\pi(A)}{t_{\text{REL}}}\right)^t$$

Lemma 1.3. Let (Ω, P, π) be a finite irreducible Markov chain and $A \subset \Omega$, $A \neq \emptyset$. Let ψ_{A^c} be the probability measure on A^c given by

$$\forall y \in A^c, \quad \psi_{A^c}(y) = \mathbb{P}_{\pi_A} \left(X_1 = y \mid X_1 \in A^c \right) \,.$$

Then, for all $t \geq 1$,

$$\frac{\mathbb{P}_{\pi_{A^c}}(\tau_A = t)}{\Phi(A^c)} = \mathbb{P}_{\psi_{A^c}}(\tau_A \ge t).$$
(1.5)

Consequently,

$$\mathbb{E}_{\psi_{A^c}}[\tau_A] = \frac{1}{\Phi(A^c)},\tag{1.6}$$

and

$$\mathbb{E}_{\psi_{A^c}}[\tau_A^2] = \mathbb{E}_{\psi_{A^c}}[\tau_A] \left(2\mathbb{E}_{\pi_{A^c}}[\tau_A] - 1\right) \,. \tag{1.7}$$

Proof of Lemma 1.3. First note that for $t \ge 1$, $\pi(A^c)\mathbb{P}_{\pi_{A^c}}(\tau_A = t) = \mathbb{P}_{\pi}(\tau_A = t)$. Writing

$$\{\tau_A = t\} = \{X_0 \notin A, \dots, X_{t-1} \notin A, X_t \in A\},\$$

and using stationarity, we have

$$\mathbb{P}_{\pi}(\tau_{A} = t) = \mathbb{P}_{\pi} (X_{1} \notin A, \dots, X_{t} \notin A, X_{t+1} \in A)$$

$$= \mathbb{P}_{\pi} (X_{1} \notin A, \dots, X_{t} \notin A) - \mathbb{P}_{\pi} (X_{1} \notin A, \dots, X_{t} \notin A, X_{t+1} \notin A)$$

$$= \mathbb{P}_{\pi} (X_{1} \notin A, \dots, X_{t} \notin A) - \mathbb{P}_{\pi} (X_{0} \notin A, \dots, X_{t} \notin A)$$

$$= \mathbb{P}_{\pi} (X_{0} \in A, X_{1} \notin A, \dots, X_{t} \notin A)$$

$$= \pi(A) \Phi(A) \mathbb{P}_{\psi_{A^{c}}} (\chi_{0} \notin A, \dots, X_{t-1} \notin A)$$

$$= \pi(A) \Phi(A) \mathbb{P}_{\psi_{A^{c}}} (\tau_{A} \ge t).$$

Equation (1.5) is then established after noticing that

$$\pi(A)\Phi(A) = \pi(A^c)\Phi(A^c)$$

Summing (1.5) over $t \ge 1$ yields (1.6), and (1.7) follows from

$$\sum_{t\geq 1} (2t-1)\mathbb{P}_{\psi_{A^c}}(\tau_A \geq t) = \mathbb{E}_{\psi_{A^c}}\left[\sum_{t\geq 1} (2t-1)\mathbb{1}_{t\leq \tau_A}\right] = \mathbb{E}_{\psi_{A^c}}\left[\tau_A^2\right].$$

Let us now move to the proof of Theorem 1.1. Let T = (V, E) be a finite tree and let P be the transition matrix of a lazy chain on T, i.e. such that P(x, y) > 0 iff $\{x, y\} \in E$ or x = y, in which case $P(x, x) \ge 1/2$. Let π be the stationary distribution of P. By Kolmogorov's cycle condition, P is reversible with respect to π .

The first step of the proof of Theorem 1.1 consists in showing that $hit_{1/2}(\varepsilon)$ is well approximated by

$$\mathbf{t}_{\rho}(\varepsilon) = \min\{t \ge 0, \max_{x \in \Omega} \mathbb{P}_x(\tau_{\rho} > t) \le \varepsilon\},\$$

where ρ is a *central vertex* in tree T, i.e. a vertex $v \in V$ such that each connected component of $T \setminus \{v\}$ has stationary probability at most 1/2. There always either one or two central vertices in a tree. Throughout this section, we fix a central vertex ρ and call it the root of the tree.

For all $v \in V \setminus \{\rho\}$, if $(v_0 = v, v_1, \ldots, v_k = \rho)$ is the shortest path from v to ρ in T, we call $\mathbf{p}(v) = v_1$ the *parent* of v and we write $u \prec v$ if $u = v_\ell$ for some $\ell \in \{0, \ldots, k\}$ (inducing a partial order on V).

Lemma 1.4. For all $0 < \varepsilon < 1$,

$$\mathbf{t}_{\rho}(\varepsilon) \leq \mathsf{hit}_{1/2}(\varepsilon) \leq \mathbf{t}_{\rho}(\varepsilon/2) + \left\lceil 4t_{\mathrm{REL}} \log\left(\frac{9}{2\varepsilon}\right) \right\rceil$$

Proof of Lemma 1.4. By definition of a central vertex, for all $x \in V \setminus \{\rho\}$, there exists A with $\pi(A) \ge 1/2$ such that the chain started at x cannot hit A without first hitting ρ . Hence

$$\mathbf{t}_{\rho}(\varepsilon) \leq \mathsf{hit}_{1/2}(\varepsilon)$$
 .

In the other direction, take $A \subset V$ with $\pi(A) \geq 1/2$, $x \in V$ and $s_{\varepsilon} = \lceil 4t_{\text{REL}} \log(9/2\varepsilon) \rceil$ for $0 < \varepsilon < 1$. By Markov property and definition of $\tau_{\rho}(\varepsilon/2)$,

$$\mathbb{P}_{x}\left(\tau_{A} > \mathbf{t}_{\rho}(\varepsilon/2) + s_{\varepsilon}\right) \leq \mathbb{P}_{x}\left(\tau_{\rho} > \mathbf{t}_{\rho}(\varepsilon/2)\right) + \mathbb{P}_{\rho}\left(\tau_{A} > s_{\varepsilon}\right) \leq \frac{\varepsilon}{2} + \mathbb{P}_{\rho}\left(\tau_{A} > s_{\varepsilon}\right)$$

It remains to show that $\mathbb{P}_{\rho}(\tau_A > s_{\varepsilon}) \leq \varepsilon/2$. If $\rho \in A$, this is trivially verified. Let us assume that $\rho \notin A$. Note that $T \setminus \{\rho\}$ can be partitioned into $T_1 \cup T_2$ such that both T_1 and T_2 are unions of components of $T \setminus \{\rho\}$ and have stationary mass at most 2/3. Moreover, since $\pi(A) \geq 1/2$, we may assume without loss of generality that $\pi(A_1) \geq 1/4$, where $A_1 = A \cap T_1$. Let $B = T_2 \cup \{\rho\}$. Since the chain started from any vertex in B must hit ρ before hitting A_1 , we have

$$\mathbb{P}_{\rho}\left(\tau_{A} > s_{\varepsilon}\right) \leq \mathbb{P}_{\rho}\left(\tau_{A_{1}} > s_{\varepsilon}\right) \leq \mathbb{P}_{\pi_{B}}\left(\tau_{A_{1}} > s_{\varepsilon}\right) \,.$$

By Lemma 1.2,

$$\mathbb{P}_{\pi_B}\left(\tau_{A_1} > s_{\varepsilon}\right) \leq \frac{\pi(A_1^c)}{\pi(B)} \left(1 - \frac{\pi(A_1)}{t_{\text{REL}}}\right)^{s_{\varepsilon}} \leq 3 \cdot \frac{3}{4} \left(1 - \frac{1}{4t_{\text{REL}}}\right)^{4t_{\text{REL}}\log(9/2\varepsilon)} \leq \frac{\varepsilon}{2} \,,$$

where we used that $\pi(B) \ge 1/3$ and that $1/4 \le \pi(A_1) \le 1/2$.

Lemma 1.5. Let $\Delta = \max_{x \in V} \mathbb{E}_x[\tau_{\rho}]$. For all $\varepsilon \in]0, 1/4[$,

$$\mathbf{t}_{\rho}(\varepsilon) \leq \Delta + \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}} \quad and \quad \mathbf{t}_{\rho}(1-\varepsilon) \geq \Delta - \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}$$

Proof of Lemma 1.5. For all $x \in V$,

$$\mathbb{P}_x\left(\tau_{\rho} > \Delta + \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}\right) \le \mathbb{P}_x\left(\tau_{\rho} - \mathbb{E}_x[\tau_{\rho}] > \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}\right) \le \frac{\varepsilon \operatorname{Var}_x(\tau_{\rho})}{4\Delta t_{\text{REL}}},$$

where the last inequality is by Chebychev Inequality. Let $(v_0 = x, v_1, \ldots, v_k = \rho)$ be the path from x to ρ . Define $\tau_i = \tau_{v_i} - \tau_{v_{i-1}}$. Note that, under \mathbb{P}_x , $\tau_{\rho} = \sum_{i=1}^k \tau_i$ and that τ_1, \ldots, τ_k are independent. Hence

$$\operatorname{Var}_{x}(\tau_{\rho}) = \sum_{i=1}^{k} \operatorname{Var}_{v_{i-1}}(\tau_{v_{i}}) \leq \sum_{i=1}^{k} \mathbb{E}_{v_{i-1}}\left[\tau_{v_{i}}^{2}\right] \,.$$

Recall that for $u \in V \setminus \{\rho\}$, $\mathbf{p}(u)$ denotes the parent of u in T. Applying Lemma 1.3 to $A^c = \{v \in T, u \prec v\}$, noticing that, by the tree structure, $\psi_{A^c}(y) = \delta_u(y)$ and $\mathbb{E}_u[\tau_A] = \mathbb{E}_u[\tau_{\mathbf{p}(u)}]$, we obtain

$$\mathbb{E}_{u}\left[\tau_{\mathbf{p}(u)}^{2}\right] = \mathbb{E}_{u}\left[\tau_{\mathbf{p}(u)}\right]\left(2\mathbb{E}_{\pi_{A^{c}}}[\tau_{A}]-1\right) \leq 2\mathbb{E}_{u}\left[\tau_{\mathbf{p}(u)}\right]\frac{t_{\text{REL}}}{\pi(A)} \leq 4\mathbb{E}_{u}\left[\tau_{\mathbf{p}(u)}\right]t_{\text{REL}}$$

where the penultimate inequality uses $\mathbb{E}_{\pi_{A^c}}[\tau_A] = \pi(A^c)^{-1}\mathbb{E}_{\pi}[\tau_A] \leq \pi(A)^{-1}t_{\text{REL}}$ by summing over $t \geq 0$ in Lemma 1.2, and the last inequality uses $\pi(A) \geq 1/2$ by centrality of ρ . Hence

$$\operatorname{Var}_{x}(\tau_{\rho}) \leq 4\Delta t_{\operatorname{REL}},$$
(1.8)

and

$$\mathbb{P}_x\left(\tau_\rho > \Delta + \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}\right) \le \varepsilon\,,$$

establishing the first inequality of the lemma. For the second one, take $x \in V$ such that $\mathbb{E}_x[\tau_{\rho}] = \Delta$. Then, again by Chebychev Inequality and by (1.8),

$$\mathbb{P}_x\left(\tau_{\rho} \leq \Delta - \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}\right) \leq \varepsilon$$

Hence $\mathbb{P}_x(\tau_{\rho} > \Delta - \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}) \ge 1 - \varepsilon$ and $\mathbf{t}_{\rho}(1-\varepsilon) \ge \Delta - \sqrt{4\varepsilon^{-1}\Delta t_{\text{REL}}}$.

We are now ready to prove Theorem 1.1. Fix $\varepsilon \in [0, 1/4]$. By (1.3) and (1.4), we have

$$t_{\rm MIX}(\varepsilon) - t_{\rm MIX}(1-\varepsilon) \le \mathsf{hit}_{1/2}(\varepsilon/2) - \mathsf{hit}_{1/2}(1-\varepsilon/2) + \left\lceil t_{\rm REL} \log\left(\frac{4}{\varepsilon}\right) \right\rceil + \left\lceil 2t_{\rm REL} \log\left(\frac{1}{\varepsilon}\right) \right\rceil.$$

Using Lemma 1.4 with ε replaced by $\varepsilon/2$, we get

$$\mathsf{hit}_{1/2}(\varepsilon/2) - \mathsf{hit}_{1/2}(1 - \varepsilon/2) \le \mathbf{t}_{\rho}(\varepsilon/4) - \mathbf{t}_{\rho}(1 - \varepsilon/2) + \left\lceil 4t_{\text{REL}} \log\left(\frac{9}{\varepsilon}\right) \right\rceil \,.$$

By Lemma 1.5,

$$\mathbf{t}_{\rho}(\varepsilon/4) - \mathbf{t}_{\rho}(1 - \varepsilon/2) \le (4 + 2\sqrt{2})\sqrt{\varepsilon^{-1}\Delta t_{\text{REL}}}$$

The proof is then concluded (modulo some additional work to get the constant 35, which we won't do here) by observing that

 $\Delta \leq 4t_{\text{MIX}}$.

Indeed, let $x \in V$ and let $A = V \setminus C_x$ where C_x is the component of $T \setminus \{\rho\}$ which contains x. Clearly, $\mathbb{E}_x[\tau_\rho] = \mathbb{E}_x[\tau_A]$. Letting $\tilde{\tau}_A = \min\{k \ge 0, X_{kt_{\text{MIX}}} \in A\}$, we have $\tau_A \le t_{\text{MIX}}\tilde{\tau}_A$, and, by definition of t_{MIX} and the fact that $\pi(A) \ge 1/2$, the variable $\tilde{\tau}_A$ is stochastically dominated by a Geometric variable with parameter 1/4, which yields $\mathbb{E}_x[\tau_\rho] \le 4t_{\text{MIX}}$.

2 The configuration model

Given a finite set V of size |V| = n and a function $\mathbf{d} \colon V \to \{2, 3, \ldots\}$ such that

$$N = \sum_{v \in V} \mathbf{d}(v)$$

is even, we can construct a graph G with vertex set V and degrees $(\mathbf{d}(v))_{v \in V}$ as follows. We form a set \mathcal{H} by "attaching" $\mathbf{d}(v)$ half-edges to each vertex $v \in V$:

$$\mathcal{H} = \{ (v, i) \colon v \in V, 1 \le i \le \mathbf{d}(v) \}.$$

We then choose a pairing η on \mathcal{H} (i.e., an involution without fixed points), and interpret every pair of matched half-edges $\{x, \eta(x)\}$ as an edge between the corresponding vertices. Loops and multiple edges are allowed. The *configuration model* is the random graph obtained by choosing η uniformly at random among the (N-1)!! possible pairings of \mathcal{H} . We will say that



Figure 1: A set of half-edges \mathcal{H} , a pairing η and the resulting graph G

two half-edges x = (u, i) and y = (v, j) are neighbors if u = v and $i \neq j$. The degree of a half-edge x = (u, i) is then defined as its number of neighbors, i.e.

$$\deg(x) = \mathbf{d}(u) - 1.$$

The non-backtracking random walk (NBRW) on G is the Markov chain on \mathcal{H} with transition matrix

$$P(x,y) = \begin{cases} \frac{1}{\deg(\eta(x))} & \text{if } y \text{ is a neighbor of } \eta(x), \\ 0 & \text{otherwise.} \end{cases}$$

The matrix P is not symmetric. However, it enjoys the following symmetry property with respect to the pairing η :

$$\forall x, y \in \mathcal{H}, \ P(x, y) = P(\eta(y), \eta(x)).$$
(2.1)

The simple random walk (SRW) on G is the Markov chain on V with transition matrix

$$Q(u, v) = \begin{cases} \frac{1}{\mathbf{d}(u)} & \text{if } v \text{ is a neighbor of } u, \\ 0 & \text{otherwise.} \end{cases}$$



Figure 2: A non-backtracking move

Matrix P has stationary distribution uniform over \mathcal{H} and matrix Q has stationary distribution given by

$$\forall u \in V, \ \pi(u) = \frac{\mathbf{d}(u)}{N} \cdot$$

Some definitions and notation

- Let $\Delta = \max_{v \in V} \mathbf{d}(v)$ be the maximum degree.
- For $x \in \mathcal{H}$ and $k \in \mathbb{N}$, let $\mathcal{B}_k(x)$ be the subgraph induced by the half-edges which are at non-backtracking distance less than or equal to k from x.
- The *excess* of $\mathcal{B}_k(x)$ is the maximum number of edges that can be deleted from $\mathcal{B}_k(x)$ while keeping it connected.
- We call x a k-root if $\mathcal{B}_k(x)$ is a tree, i.e. if the excess of $\mathcal{B}_k(x)$ is zero.

Let us start with a simple but crucial result, which illustrates the locally tree-like structure of G in the sparse regime.

Lemma 2.1. Let $L = \lfloor \frac{1}{5} \log_{\Delta - 1}(N) \rfloor$. For all $x \in \mathcal{H}$,

$$\mathbb{P}(\mathrm{Ex}(\mathcal{B}_L(x) \ge 1) = o(1) \quad and \quad \mathbb{P}(\mathrm{Ex}(\mathcal{B}_L(x) \ge 2) = o\left(\frac{1}{N}\right).$$

In particular, whp, the excess is at most 1 in all L-neighborhoods.

Proof of Lemma 2.1. The ball of radius L around x can be generated sequentially, its halfedges being paired one after the other with uniformly chosen other unpaired half-edges, until the whole ball has been paired. Observe that at most $k = (\Delta - 1)^L$ pairs are formed. Moreover, for each of them, the number of unpaired half-edges having an already paired neighbor is at most $(\Delta - 1)^L$ and hence the conditional chance of hitting such a half-edge (thereby creating a cycle) is at most $p = \frac{(\Delta - 1)^L - 1}{N - 2k - 1}$. Thus, the probability that one cycle is formed is at most

$$kp = O\left(\frac{(\Delta - 1)^{2L}}{N}\right) = O(N^{-3/5}),$$

and the probability that more than one cycle is found is at most

$$(kp)^2 = O\left(\frac{(\Delta - 1)^{4L}}{N^2}\right) = O(N^{-6/5}).$$

Lemma 2.2. Assume that $d = \min_{v \in V} \mathbf{d}(v) \ge 3$. Let $K = \lfloor \frac{1}{6} \log_{\Delta - 1}(N) \rfloor$ and let \mathcal{R} be the set of K-roots. Then, for all $s \le L - K$,

$$\max_{x \in \mathcal{H}} P^s(x, \mathcal{H} \setminus \mathcal{R}) \le 2(d-1)^{-s} + o_{\mathbb{P}}(1)$$

Proof of Lemma 2.2. By Lemma 2.1, whp, for all $x \in \mathcal{H}$, the ball $\mathcal{B}_L(x)$ has at most one cycle, with $L = \lfloor \frac{1}{5} \log_{\Delta - 1}(N) \rfloor$. Fix a graph G with this property. We prove that the NBRW on G starting from any $x \in \mathcal{H}$ satisfies

$$P^{s}(x, \mathcal{H} \setminus \mathcal{R}) \le 2(d-1)^{-s}, \tag{2.2}$$

for all $s \leq L - K$. The claim is trivial if the ball of radius L around x is acyclic. Otherwise, it contains a single cycle C, by assumption. Write d(x,C) for the minimum length of a non-backtracking path from x to some $z \in C$. The non-backtracking property ensures that if $d(X_s, C) < d(X_{s+1}, C)$ for some s < L - K, then $X_{s+1}, X_{s+2}, \ldots, X_{L-K}$ are all K-roots. Indeed, as soon as the NBRW makes a step away from C on one of the disjoint trees rooted to C, it can only go further away from it. The conditional chance that $d(X_{s+1}, C) \leq d(X_s, C)$ given the past is at most $\frac{1}{d-1}$ (unless $d(X_s, C) = 1$, which can only happen once). This shows (2.2).

3 Cutoff for NBRW on regular random graphs

Let $d \ge 3$ be a fixed integer and let $G = G_{n,d}$ be the random graph formed by the configuration model on vertex set V with |V| = n and constant degree sequence: $\mathbf{d}(v) = d$ for all $v \in V$. The number of half-edges is N = dn. We are interested in the asymptotics (in n) of

$$t_{\text{MIX}}(\varepsilon) = \min\{t \ge 0, \mathcal{D}(t) \le \varepsilon\},\$$

where

$$\mathcal{D}(t) = \max_{x \in \mathcal{H}} \sum_{y \in \mathcal{H}} \left(\frac{1}{N} - P^t(x, y) \right)_+$$

In a seminal paper, Lubetzky and Sly [13] showed the following.

Theorem 3.1. For $d \ge 3$, whp, the NBRW on $G_{n,d}$ has cutoff at time $\log_{d-1}(N)$, with a window of constant order¹. More precisely, for all $0 < \varepsilon < 1$, whp,

$$t_{\text{MIX}}(1-\varepsilon) \ge \lfloor \log_{d-1}(N) \rfloor - \lceil \log_{d-1}(1/\varepsilon) \rceil,$$

and

$$t_{\text{MIX}}(\varepsilon) \leq \lceil \log_{d-1}(N) \rceil + 3 \lceil \log_{d-1}(1/\varepsilon) \rceil + 4 \rceil$$

Let us start with the lower bound, which comes from a simple counting argument and is actually valid on *any d*-regular graph. Fix a starting point $x \in \mathcal{H}$. Let $t = \lfloor \log_{d-1}(\varepsilon N) \rfloor$, and let A be the set of half-edges which are reachable by a NBRW of length t started at x. Then, $P^t(x, A) = 1$ and

$$\pi(A) \leq \frac{(d-1)^t}{N} \leq \varepsilon,$$

implying $\mathcal{D}_x(t) \geq 1 - \varepsilon$. Hence $t_{\text{MIX}}(1-\varepsilon) \geq t \geq \lfloor \log_{d-1}(N) \rfloor - \lceil \log_{d-1}(1/\varepsilon) \rceil$.

Let us now move to the upper bound. The first step is to reduce the maximization over "nice" starting points. Recall that $K = \lfloor \frac{1}{6} \log_{d-1}(N) \rfloor$ and that \mathcal{R} is the set of K-roots. By Lemma 2.2 applied to $s = \lceil \log_{d-1}(2/\varepsilon) \rceil$, we have

$$\max_{x \in \mathcal{H}} P^s(x, \mathcal{H} \setminus \mathcal{R}) \le \varepsilon + o_{\mathbb{P}}(1) \,.$$

Take

$$t = 2\left[\frac{1}{2}\log_{d-1}(N) + \frac{3}{2}\log_{d-1}(1/\varepsilon)\right]$$

By the inequality

$$\mathcal{D}(t+s) \leq \max_{x \in \mathcal{H}} P^s(x, \mathcal{H} \setminus \mathcal{R}) + \max_{x \in \mathcal{R}} \mathcal{D}_x(t),$$

we see that we may consider starting points which are K-roots.

¹For x, a > 0, $\log_a(x) = \frac{\log(x)}{\log(a)}$.

To get the constant order for the window, we have to take advantage of the averaging over $y \in \mathcal{H}$ in the definition of the total-variation distance. To do so, choose a partition of \mathcal{H} into $\lfloor N/M \rfloor$ blocks of size $M = \lceil (\log N)^2 \rceil$, and possibly one last block of size strictly less than M (\mathcal{P} are fixed before the graph is formed). Fix $x \in \mathcal{H}$. Letting \mathcal{P}_{\star} be the set of blocks of size exactly M which do not contain x and bounding the summands by 1/N for y not in the support of \mathcal{P}_{\star} , we have

$$\mathcal{D}_x(t) \le \sum_{\mathcal{S} \in \mathcal{P}_\star} \sum_{y \in \mathcal{S}} \left(\frac{1}{N} - P^t(x, \eta(y)) \right)_+ + \frac{2M}{N}.$$

Let S be one of the blocks of size M in this partition. For $y \in S$, observe that, thanks to (2.1),

$$P^{t}(x,\eta(y)) = \sum_{u,v \in \mathcal{H}} P^{t/2}(x,u) P^{t/2}(y,v) \mathbb{1}_{\eta(v)=u} \,.$$

We consider an exploration process which generates the paring η along with M + 1 disjoint trees \mathcal{T}_x and $(\mathcal{T}_y)_{y \in S}$, rooted at x and $y \in S$ respectively. Initially, all half-edges are unpaired, \mathcal{T}_x is reduced to x and for all $y \in S$, \mathcal{T}_y is reduced to y. Then at each time step,

- 1. An unpaired half-edge z of $\bigcup_{r \in S \cup \{x\}} \mathcal{T}_r$ is chosen, provided its distance to the corresponding root is strictly less than t/2.
- 2. The chosen half-edge z is then paired to a uniformly chosen other unpaired half-edge z'.
- 3. If z' was not already in $\bigcup_{r \in S \cup \{x\}} \mathcal{T}_r$ and is not a neighbor of either x or $y \in S$, then the neighbors of z' are added to $\bigcup_{r \in S \cup \{x\}} \mathcal{T}_r$ as children of z. Otherwise, both z and z' are marked with the color RED.
- 4. This process continues until no unpaired half-edge in $\bigcup_{r \in S \cup \{x\}} \mathcal{T}_r$ is at distance strictly less than t/2 from its root. For $r \in S \cup \{x\}$, we denote by $\partial \mathcal{T}_r$ the set of leaves of \mathcal{T}_r (including the RED half-edges), and by \mathcal{F}_r the subset of leaves of $\partial \mathcal{T}_r$ which are at distance t/2 of r (i.e., those that are not RED).
- 5. Then finally, for each $y \in S$ and each $z \in \mathcal{F}_y$, we draw a Bernoulli random variable to decide whether $\eta(z) \in \bigcup_{r \in S} \mathcal{F}_r$ or not. If it is, then we choose $z' = \eta(z)$ uniformly at random in $\bigcup_{r \in S} \mathcal{F}_r \setminus \{z\}$ and we mark both z and z' with the color GREEN. We let $\widetilde{\mathcal{F}}_y$ be set of half-edges in \mathcal{F}_y that are not GREEN.

The pairing η is then completed to form the graph G.

Note that, at the end of the exploration stage, at most $(M+1)(d-1)^{t/2} = O(\log^2(N)\sqrt{N})$ half-edges have been revealed.

Retaining only those paths that remain in $\mathcal{T}_x \cup \mathcal{T}_y$, we have

$$P^t(x,\eta(y)) \ge \sum_{\substack{u \in \mathcal{F}_x \\ v \in \widetilde{\mathcal{F}}_y}} \frac{1}{(d-1)^t} \mathbb{1}_{\eta(v)=u} = \frac{1}{(d-1)^t} \sum_{v \in \widetilde{\mathcal{F}}_y} \mathbb{1}_{\eta(v)\in\mathcal{F}_x}.$$

A key observation is that, conditionally on the exploration stage, the variables $(\mathbb{1}_{\eta(v)\in\mathcal{F}_x})_{v\in\widetilde{\mathcal{F}}_y}$ enjoy a strong negative dependence property known as *negative association*.

Definition 3.2. Real-valued random variables X_1, \ldots, X_K are said to be negatively associated if, for any two disjoint subsets A and B of $\{1, \ldots, K\}$, and any two real-valued functions $f: \mathbb{R}^{|A|} \to \mathbb{R}$ and $g: \mathbb{R}^{|B|} \to \mathbb{R}$ that are both coordinate-wise increasing, we have

$$\mathbb{E}\left[f(X_A) \cdot g(X_B)\right] \le \mathbb{E}\left[f(X_A)\right] \cdot \mathbb{E}\left[g(X_B)\right] \,.$$

Conditionally on the exploration stage, the whole sequence $(\mathbb{1}_{\eta(v)\in\mathcal{F}_x})_{v\in\cup_{y\in\mathcal{S}}\widetilde{\mathcal{F}}_y}$ is negatively associated. To see this, compare that sequence to a sequence recording draws of black balls when sampling without replacement in a urn containing $|\mathcal{F}_x|$ black balls and $|\mathcal{I}| - |\mathcal{F}_x| - \sum_{y\in\mathcal{S}} |\widetilde{\mathcal{F}}_y|$ white balls, where \mathcal{I} is the set of half-edges that have not been paired during the exploration stage. Negative association implies that the Laplace transform of the sum can be bounded by the product of the Laplace transforms of each Bernoulli variable. In this sense, " a sum of negatively associated variables can only concentrate better than a sum of independent variables with the same marginal distributions". More precisely, for all $\lambda > 0$, denoting by **P** and **E** the conditional probability and expectation given the exploration stage, and letting

$$M_y = \sum_{v \in \widetilde{\mathcal{F}}_y} \mathbb{1}_{\eta(v) \in \mathcal{F}_x} \,,$$

we have

$$\mathbf{E}\left[e^{-\lambda(M_y-\mathbf{E}M_y)}\right] \leq \prod_{v\in\widetilde{\mathcal{F}}_y} \mathbf{E}\left[e^{-\lambda(\mathbb{1}_{\eta(v)\in\mathcal{F}_x}-\mathbf{E}\mathbb{1}_{\eta(v)\in\mathcal{F}_x})}\right] \leq \exp\left\{\frac{\lambda^2|\mathcal{F}_x||\widetilde{\mathcal{F}}_y|}{2\left(|\mathcal{I}|-\sum_{z\in\mathcal{S}}|\widetilde{\mathcal{F}}_z|\right)}\right\},\$$

where the last inequality is from Bennett Inequality. This entails that

$$\mathbf{P}\left(M_y < \frac{|\mathcal{F}_x| |\widetilde{\mathcal{F}}_y|}{N} - \frac{\varepsilon (d-1)^t}{N}\right) \le \exp\left\{-\frac{1}{2\varepsilon}\right\}$$
(3.1)

where we used that $|\mathcal{I}| - \sum_{z \in S} |\widetilde{\mathcal{F}}_z| \leq N$, that $|\mathcal{F}_x| |\widetilde{\mathcal{F}}_y| \leq (d-1)^t$ and that $(d-1)^t \geq \varepsilon^{-3}N$. Let us now use averaging on S to get a smaller error probability. Let

$$Z = \sum_{y \in \mathcal{S}} Z_y \quad \text{where } Z_y = \mathbb{1} \left\{ M_y < \frac{|\mathcal{F}_x| |\widetilde{\mathcal{F}}_y|}{N} - \frac{\varepsilon (d-1)^t}{N} \right\}.$$

Negative association also holds for $(Z_y)_{y \in S}$, since the functions Z_y are coordinate-wise decreasing functions of $(\mathbb{1}_{\eta(v)\in\mathcal{F}_x})_{v\in\tilde{\mathcal{F}}_y}$.

This implies that for all $\lambda > 0$,

$$\mathbf{E}\left[e^{\lambda(Z-\mathbf{E}\,Z)}\right] \leq \prod_{y\in\mathcal{S}} \mathbf{E}\left[e^{\lambda(Z_y-\mathbf{E}\,Z_y)}\right] \leq e^{\frac{\lambda^2 M}{8}},$$

where the last bound comes from Hoeffding Inequality. By (3.1), $\mathbf{E} Z \leq M e^{-\frac{1}{2\varepsilon}} < \varepsilon M$, hence

$$\mathbf{P}\left(Z > \varepsilon M\right) \le \mathbf{P}\left(Z - \mathbf{E}Z > (\varepsilon - e^{-\frac{1}{2\varepsilon}})M\right) \le \exp\left(-2(\varepsilon - e^{-\frac{1}{2\varepsilon}})^2 M\right).$$

Taking expectation, we get

$$\mathbb{P}\left(Z > \varepsilon M\right) = o\left(\frac{1}{N^2}\right) \,.$$

Taking a union bound over all x and over blocks of \mathcal{P}_{\star} , we obtain that **whp**, for all $x \in \mathcal{H}$,

$$\mathcal{D}_x(t) \le \frac{1}{N} \sum_{\mathcal{S} \in \mathcal{P}_\star} \sum_{y \in \mathcal{S}} \left(1 - \frac{|\mathcal{F}_x| |\widetilde{\mathcal{F}}_y|}{(d-1)^t} \right)_+ + \varepsilon + \frac{2M}{N} \cdot$$
(3.2)

It remains to control the sizes $|\mathcal{F}_x|$ and $|\mathcal{F}_y|$ for $y \in S$. To do so, let us first bound the number of RED half-edges in \mathcal{T}_r . We will then need to show that those RED half-edges do not arrive too early in the trees.

Lemma 3.3. Let |RED| be the number of RED half-edges at the end of the exploration stage.

$$\mathbb{P}\left(|\text{RED}| > \log^5 N\right) = o\left(\frac{1}{N^2}\right)$$

Proof of Lemma 3.3. When looking at the exploration process sequentially, we see that the total number of RED half-edges is stochastically dominated by

$$X = Bin\left((M+1)(d-1)^{t/2}, \frac{(M+1)(d-1)^{t/2}}{N-2(M+1)(d-1)^{t/2}}\right).$$

By Bennett's Inequality, for all a > 0,

$$\mathbb{P}(X - \mathbb{E}X > a) \le \exp\left(-\frac{a^2}{2(\mathbb{E}X + a)}\right)$$
.

Taking $a = \log^4 N$ and observing that $\mathbb{E}X = O(\log^4 N)$, we obtain

 $\mathbb{P}\left(X - \mathbb{E}X > \log^4 N\right) = \exp\left(-\Omega(\log^4 N)\right) \,,$

and we obtain the desired result, after taking some room to compensate for the constant that depends on ε in $\mathbb{E}X$.

Lemma 3.4. Let |GREEN| be the number of GREEN half-edges at the end of the exploration stage.

$$\mathbb{P}\left(|\text{GREEN}| > \log^5 N\right) = o\left(\frac{1}{N^2}\right)$$

Proof of Lemma 3.4. When looking at the last step of the exploration stage, we see that the total number of GREEN half-edges is stochastically dominated by

$$Y = Bin\left(M(d-1)^{t/2}, \frac{M(d-1)^{t/2}}{N-2(M+1)(d-1)^{t/2}}\right).$$

By the same argument as in the previous proof, we obtain the desired result.

Lemma 3.5. Let $\mathcal{E} = \mathcal{E}(\mathcal{S})$ be the event that all $r \in \mathcal{S} \cup \{x\}$ are K-roots and that all pairwise distances in $\mathcal{S} \cup \{x\}$ are larger than 2K. Then

$$\mathbb{P}\left(\mathcal{E}\cap\left\{\min\{|\mathcal{F}_x|,\min_{y\in\mathcal{S}}|\widetilde{\mathcal{F}}_y|\}\le (d-1)^{t/2}(1-N^{-1/7})\right\}\right)=o\left(\frac{1}{N^2}\right).$$

Proof of Lemma 3.5. By Lemma 3.3 and 3.4, with probability $1 - o(1/N^2)$, both |RED| and |GREEN| are smaller than $\log^5 N$. On the event \mathcal{E} , no RED half-edges can occur before level K, hence

$$\min_{r \in \mathcal{S} \cup \{x\}} |\mathcal{F}_r| \ge (d-1)^{t/2} \left(1 - \frac{\log^5 N}{(d-1)^K} \right) \,,$$

and

$$\min_{y \in \mathcal{S}} |\widetilde{\mathcal{F}}_y| \ge (d-1)^{t/2} \left(1 - \frac{\log^5 N}{(d-1)^K}\right) - \log^5(N),$$

which gives the desired result.

Plugging those bounds into (3.2) and bounding the summands by 1 when S does not satisfies \mathcal{E} , we obtain that whp, for all $x \in \mathcal{H}$,

$$\mathcal{D}_x(t) \le \frac{M}{N} \sum_{\mathcal{S} \in \mathcal{P}_\star} \mathbb{1}_{\mathcal{E}(\mathcal{S})^c} + \varepsilon + \frac{2M}{N} + 2N^{-1/7}$$

To conclude the proof, observe that

$$\max_{x \in \mathcal{R}} \sum_{S \in \mathcal{P}_{\star}} \mathbb{1}_{\mathcal{E}(S)^{c}} \leq |\mathcal{H} \setminus \mathcal{R}| + \sum_{S \in \mathcal{P}} \sum_{z \in S} |\mathcal{B}_{2K}(z) \cap S| = o_{\mathbb{P}}(N/M),$$

and recall that by Lemma 2.2, we can restrict to $x \in \mathcal{R}$. We have shown that, whp

$$\max_{x \in \mathcal{R}} \mathcal{D}_x(t) \le 2\varepsilon$$

which gives

$$\max_{x \in \mathcal{H}} \mathcal{D}_x(t+s) \le 3\varepsilon$$

4 Cutoff for SRW on random regular graphs

Let us now move to the SRW. In the same paper [13], Lubetzky and Sly showed the following result.

Theorem 4.1. For $d \ge 3$, whp, the SRW on $G_{n,d}$ has cutoff at time $\frac{d}{d-2}\log_{d-1}(n)$, with a window of order $\sqrt{\log n}$. Moreover, for all $0 < \varepsilon < 1$,

$$t_{\mathrm{MIX}}(\varepsilon) = \frac{d}{d-2} \log_{d-1}(n) + (\Lambda + o_{\mathbb{P}}(1)) \overline{\Phi}^{-1}(\varepsilon) \sqrt{\log n} \,,$$

where $\Lambda = \frac{2\sqrt{d(d-1)}}{(d-2)^{3/2}}$ and $\overline{\Phi}^{-1}(\varepsilon)$ is the $(1-\varepsilon)$ -quantile of the standard Gaussian distribution.

On regular graphs, cutoff for SRW can actually be deduced from cutoff for NBRW as follows. If G = (V, E) is a *d*-regular graph and if \mathbb{T}_d denotes a *d*-regular tree rooted at ρ , then the cover tree of G at $x \in V$ is defined as a map $\varphi : \mathbb{T}_d \to V$ satisfying

$$\begin{cases} \varphi(\rho) = x, \\ \forall \gamma \in \mathbb{T}_d, \ \{\varphi(\zeta), \ \zeta \sim_{\mathbb{T}_d} \gamma\} = \{z \sim_G \varphi(\gamma)\}. \end{cases}$$

In other words, the root of \mathbb{T}_d is mapped to x and φ preserves 1-neighborhoods.

Observe that if \mathcal{X}_t is a SRW on \mathbb{T}_d started at ρ , then $X_t = \varphi(\mathcal{X}_t)$ is a SRW on G started at x. Similarly, if $\mathcal{Y}_t = (\mathcal{Y}_t^-, \mathcal{Y}_t^+)$ is a NBRW on \mathbb{T}_d started at (ρ, ζ) , then $Y_t = (\varphi(\mathcal{Y}_t^-), \varphi(\mathcal{Y}_t^+))$ is a NBRW on G started at $(x, \varphi(\zeta))^2$. By symmetry, we have

$$\mathbb{P}_x\left(X_t \in \cdot \mid \operatorname{dist}(\rho, \mathcal{X}_t) = \ell\right) = \frac{1}{d} \sum_{\zeta \sim_{\mathbb{T}_d} \rho} \mathbb{P}_{(x, \varphi(\zeta))}\left(Y_{\ell-1}^+ \in \cdot\right) \,.$$

As projections can not increase total-variation distance, taking $\ell = t_{\text{MIX}}^{\text{NBRW}}(\varepsilon)$, we get

$$\|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \le \varepsilon + \mathbb{P}\left(\operatorname{dist}(\rho, \mathcal{X}_t) < \ell\right)$$

Maximizing over $x \in V$,

$$\mathcal{D}^{\text{SRW}}(t) \leq \varepsilon + \mathbb{P}\left(\operatorname{dist}(\rho, \mathcal{X}_t) < \ell\right) \,.$$

Now the SRW on \mathbb{T}_d is transient (for $d \geq 3$) and, for $\mathcal{X}_t \neq \rho$,

$$\operatorname{dist}(\rho, \mathcal{X}_{t+1}) - \operatorname{dist}(\rho, \mathcal{X}_t) = \begin{cases} 1 & \text{with probability } \frac{d-1}{d}, \\ -1 & \text{with probability } \frac{1}{d}. \end{cases}$$

By the Central Limit Theorem,

$$\frac{\operatorname{dist}(\rho, \mathcal{X}_t) - \frac{d-2}{d}t}{\sqrt{\frac{4(d-1)}{d^2}t}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \,.$$

 $^{^{2}}$ Here, it is more convenient to define the NBRW on directed edges rather than on half-edges, which is completely equivalent.

We obtain that for all s > 0,

$$\limsup_{n \to +\infty} \mathcal{D}^{\text{SRW}}\left(\frac{d}{d-2}\ell + s\sqrt{\ell}\right) \le \varepsilon + \mathbb{P}\left(\mathcal{N}(0,1) > \Lambda^{-1}s\right) \,. \tag{4.1}$$

Conversely, the number of vertices at distance ℓ from x is at most $d(d-1)^{\ell}$. So on the event $\operatorname{dist}(\rho, \mathcal{X}_t) < \log_{d-1}(\varepsilon n/d)$, the SRW X_t is confined to a set of at most εn vertices, and the total-variation distance is at leat $1 - \varepsilon$. This implies that for all s > 0,

$$\liminf_{n \to +\infty} \mathcal{D}^{\text{srw}}\left(\frac{d}{d-2}\log_{d-1}(\varepsilon n/d) - s\sqrt{\log_{d-1}(n)}\right) \ge 1 - \varepsilon - \mathbb{P}\left(\mathcal{N}(0,1) > \Lambda^{-1}s\right).$$
(4.2)

Combining (4.1) and (4.2), and the fact that $\ell = \log_{d-1}(n) + o(\sqrt{\log n})$, we obtain the desired result.

5 Cutoff for NBRW on random graphs with given degrees

We now consider the NBRW on the configuration model with a given degree sequence. We will assume that

$$\Delta = \max_{v \in V} \mathbf{d}(v) = O(1) \quad \text{and} \quad \min_{v \in V} \mathbf{d}(v) \ge 3.$$
(5.1)

Remarkably enough, the asymptotics in this regime depends on the degrees through two simple statistics: the mean logarithmic degree of an half-edge and the corresponding variance

$$\mu = \frac{1}{N} \sum_{x \in \mathcal{H}} \log \deg(x) , \qquad \sigma^2 = \frac{1}{N} \sum_{x \in \mathcal{H}} \left(\log \deg(x) - \mu \right)^2 .$$

We further assume that

$$\liminf_{n \to +\infty} \sigma^2 > 0.$$
(5.2)

In [2], the following result was shown (under much weaker degree assumptions).

Theorem 5.1. For every $0 < \varepsilon < 1$,

$$t_{\rm MIX}(\varepsilon) = \frac{\log N}{\mu} + (1 + o_{\mathbb{P}}(1))\overline{\Phi}^{-1}(\varepsilon)\sqrt{\frac{\sigma^2 \log N}{\mu^3}}$$

Let us first establish the lower bound. Let $x \in \mathcal{H}$ be a fixed starting point and let

$$t = \frac{\log N}{\mu} + (\lambda + o(1)) \sqrt{\frac{\sigma^2}{\mu^3} \log N} \,.$$

For $\theta = \frac{\log N}{N}$, let A_{θ} be the set of $y \in \mathcal{H}$ such that there exists a path from x to y which has probability larger than θ to be seen by a NBRW of length t. Since, for all $y \in A_{\theta}$, we have $P^t(x, y) \ge \theta$, and since $P^t(x, \cdot)$ is a probability, the set A_{θ} has size less than $1/\theta$, hence

$$\mathcal{D}_x(t) \ge P^t(x, A_\theta) - \pi(A_\theta) \ge P^t(x, A_\theta) - \frac{1}{\theta N}$$
.

Taking expectation with respect to the pairing, we have

$$\mathbb{E}P^t(x, A_{\theta}) \ge \mathbb{P}_x \left(\prod_{s=1}^t \frac{1}{\deg(X_s)} > \theta\right).$$

A useful property of the uniform pairing is that it can be constructed sequentially, the pairs being revealed along the way, as we need them. We exploit this degree of freedom to generate the walk $\{X_k\}_{k\geq 0}$ and the pairing simultaneously, as follows. Initially, all half-edges are unpaired and $X_0 = x$; then at each time $k \geq 1$,

- 1. if X_{k-1} is unpaired, we pair it with a uniformly chosen other unpaired half-edge; otherwise, $\eta(X_{k-1})$ is already defined and no new pair is formed.
- 2. in both cases, we let X_k be a uniformly chosen neighbour of $\eta(X_{k-1})$.

The sequence $\{X_k\}_{k\geq 0}$ is then exactly distributed according to the annealed law. Now, if we sample uniformly from \mathcal{H} instead of restricting the random choice made at (i) to unpaired half-edges, then the uniform neighbour chosen at step (ii) also has the uniform law on \mathcal{H} . This creates a coupling between the process $\{X_k\}_{k\geq 1}$ and a sequence $\{X_k^{\star}\}_{k\geq 1}$ of IID samples from \mathcal{H} , valid until the first time T where the uniformly chosen half-edge or its uniformly chosen neighbour is already paired. As there are less than 2k paired half-edges by step k, a crude union-bound yields

$$\mathbb{P}\left(T \le t\right) \le \frac{2t^2}{N} \,. \tag{5.3}$$

Consequently,

$$\mathbb{E}P^t(x, A_{\theta}) \ge \mathbb{P}_x\left(\prod_{s=1}^t \frac{1}{\deg(X_s^{\star})} > \theta\right) + o(1) \,.$$

Taking the logarithm and using Berry-Esseen Inequality, we have

$$\mathbb{P}_x\left(\prod_{s=1}^t \frac{1}{\deg(X_s^\star)} > \theta\right) \ge \overline{\Phi}(\lambda) + o(1),$$

entailing

$$\min_{x \in \mathcal{H}} \mathbb{E}\mathcal{D}_x(t) \ge \overline{\Phi}(\lambda) + o(1)$$

Let us now move to the upper bound. The first step is the same as in the regular case: reducing to starting points which are roots. Letting as before $K = \lfloor \frac{1}{6} \log_{\Delta - 1}(N) \rfloor$ and \mathcal{R} the set of K-roots, we have, by Lemma 2.2 applied to $s = \lfloor \log \log N \rfloor$,

$$\max_{x \in \mathcal{H}} P^s(x, \mathcal{H} \setminus \mathcal{R}) = o_{\mathbb{P}}(1) \,.$$

By the triangle inequality,

$$\mathcal{D}(t+s) \le \max_{x \in \mathcal{H}} P^s \left(x, \mathcal{H} \setminus \mathcal{R}\right) + \max_{x \in \mathcal{R}} \mathcal{D}_x(t)$$

The first term is $o_{\mathbb{P}}(1)$. For the second one, we write

$$\mathcal{D}_x(t) = \sum_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \left(\frac{1}{N} - P^t(x, \eta(y)) \right)_+ + \sum_{y \in \mathcal{B}_K(x) \cup (\mathcal{H} \setminus \mathcal{R})} \left(\frac{1}{N} - P^t(x, \eta(y)) \right)_+.$$

The second sum is $o_{\mathbb{P}}(1)$ uniformly in $x \in \mathcal{R}$. Indeed, it suffices to bound its summands by 1/N and observe that $|\mathcal{B}_K(x)| \leq \Delta^K = o(N)$, while $|\mathcal{H} \setminus \mathcal{R}| = o_{\mathbb{P}}(N)$ by Lemma 2.1.

The remainder of this section is devoted to establishing that, for $t = \frac{\log N}{\mu} + (\lambda + o(1))\sqrt{\frac{\sigma^2 \log N}{\mu^3}}$

$$\min_{x \in \mathcal{R}} \min_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} P^t(x, \eta(y)) \ge \frac{1 - \overline{\Phi}(\lambda) - o_{\mathbb{P}}(1)}{N}.$$
(5.4)

We start by writing

$$P^{t}(x,\eta(y)) = \sum_{u,v} P^{t/2}(x,u) P^{t/2}(y,v) \mathbb{1}_{\{\eta(u)=v\}}.$$
(5.5)

One problem that arises here (and that did not exist in the regular case) is that we really can not afford to reveal the whole neighborhoods of radius t/2 around x and y. We have to adapt our exploration process so that not too many half-edges are revealed. We proceed as follows. Initially, all half-edges are unpaired and no type has been revealed. Tree \mathcal{T}_x is reduced to xand tree \mathcal{T}_y is reduced to y. Then at each time step,

1. An unpaired half-edge z of $\mathcal{T}_x \cup \mathcal{T}_y$ is chosen, provided it satisfies

$$\mathbf{w}(z) \ge \mathbf{w}_{\text{MIN}} = N^{-\frac{1}{2} - \frac{\log(2)}{16 \log(\Delta)}}$$
 and $\mathbf{h}(z) < t/2$, (5.6)

where $\mathbf{w}(z)$ and $\mathbf{h}(z)$ correspond to the *weight* and *height* of z,defined as follows: if $z \in \mathcal{T}_r$ for $r \in \{x, y\}$, there is a unique path (z_0, \ldots, z_h) from r to z, with $z_0 = r$ and $z_h = z$. The value h is then called the height of z, denoted $\mathbf{h}(z)$, and its weight is

$$\mathbf{w}(z) = \prod_{i=1}^{h} \frac{1}{\deg(z_i)}$$

- 2. z is paired with a uniformly chosen other unpaired half-edge.
- 3. If z' was not already in $\mathcal{T}_x \cup \mathcal{T}_y$ and is not a neighbor of either x or y, then the neighbors of z' are added to $\mathcal{T}_x \cup \mathcal{T}_y$ as children of z. Otherwise, both z and z' are marked with the color RED.

This exploration process continues until no unpaired half-edge in $\mathcal{T}_x \cup \mathcal{T}_y$ satisfies (5.6). The pairing η is then completed to form the graph G. For $r \in \{x, y\}$, we denote by $\partial \mathcal{T}_r$ the set of leaves of \mathcal{T}_r , and by \mathcal{F}_r the subset of leaves of $\partial \mathcal{T}_r$ which are at distance t/2 of r.

Note that, by (5.6), for $r \in \{x, y\}$,

$$\frac{t}{2} \ge \sum_{k=1}^{t/2} \sum_{z \in \mathcal{T}_r} \mathbb{1}_{\{\mathbf{h}(z)=k\}} \mathbf{w}(z) \ge (|\mathcal{T}_r| - 1) \frac{\mathbf{w}_{\text{MIN}}}{\Delta},$$

which, together with (5.1), implies

$$|\mathcal{T}_x \cup \mathcal{T}_y| = O\left(N^{\frac{1}{2} + \frac{\log(2)}{16\log(\Delta)}}\log N\right) = O\left(N^{\frac{1}{2} + \frac{\log(2)}{15\log(\Delta)}}\right).$$
(5.7)

In particular,

$$|\mathcal{T}_x \cup \mathcal{T}_y| = O\left(N^{5/8}\right) \,. \tag{5.8}$$

Retaining only those paths of length t/2 which remain in $\mathcal{T}_x \cup \mathcal{T}_y$ and which have weight less than

$$\theta = \frac{1}{N(\log N)^3} \,,$$

we have

$$P^{t}(x,\eta(y)) \geq \sum_{\substack{u \in \mathcal{F}_{x} \\ v \in \mathcal{F}_{y}}} \mathbf{w}(u) \mathbf{w}(v) \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \leq \theta\}} \mathbb{1}_{\{\eta(u)=v\}}.$$

Conditionally on the exploration stage, the quantity above writes as a weighted sum of Bernoulli variables which are (weakly) dependent. The following lemma (which will be proved in the exercise session) allows us to obtain a strong concentration bound for such variables. **Lemma 5.2.** Let \mathcal{I} be an even set, $\{\omega_{i,j}\}_{(i,j)\in\mathcal{I}\times\mathcal{I}}$ an array of non-negative weights, and η a uniform random pairing on \mathcal{I} . Then for all a > 0,

$$\mathbb{P}\left(\sum_{i\in\mathcal{I}}\omega_{i,\eta(i)} < m-a\right) \le \exp\left\{-\frac{a^2}{4\theta m}\right\}\,.$$

where $m = \frac{1}{|\mathcal{I}|-1} \sum_{i \in \mathcal{I}} \sum_{j \neq i} \omega_{i,j}$ and $\theta = \max_{i \neq j} (\omega_{i,j} + \omega_{j,i})$.

Applying Lemma 5.2 conditionally on the exploration stage, with \mathcal{I} being the set of half-edges that did not get paired, and

$$\omega_{u,v} = \mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \le \theta\}}\mathbb{1}_{u \in \mathcal{F}_x}\mathbb{1}_{v \in \mathcal{F}_y}, \quad \theta = \frac{1}{N(\log N)^3}, \quad \text{and } a = \frac{\varepsilon}{|\mathcal{I}| - 1},$$

we obtain, using that $|\mathcal{I}| - 1 \leq N$ and that $m \leq \frac{1}{|\mathcal{I}| - 1}$,

$$\mathbf{P}\left(NP^{t}(x,\eta(y)) < \sum_{\substack{u \in \mathcal{F}_{x} \\ v \in \mathcal{F}_{y}}} \mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \le \theta\}} - \varepsilon\right) \le \exp\left\{-\frac{\varepsilon^{2}(\log N)^{3}}{4}\right\} = o\left(\frac{1}{N^{2}}\right).$$
(5.9)

where \mathbf{P} is the conditional expectation given the exploration stage.

Let us now show that, if $x \in \mathcal{R}$ and $y \in \mathcal{R} \setminus \mathcal{B}_K(x)$, then whp the total weight of paths of length t/2 that end in \mathcal{F}_r is at least $1 - \varepsilon$.

Lemma 5.3. For all $\varepsilon > 0$, with probability 1 - o(1), for all $x \in \mathcal{R}$ and $y \in \mathcal{R} \setminus \mathcal{B}_K(x)$, we have

$$\sum_{u \in \partial \mathcal{T}_x \setminus \mathcal{F}_x} \mathbf{w}(u) + \sum_{v \in \partial \mathcal{T}_y \setminus \mathcal{F}_y} \mathbf{w}(v) \le \varepsilon \,.$$

Proof of Lemma 5.3. The trees' exploration can be stopped before height t/2 for two reasons: either the weight of the half-edge is too small, or it has been colored RED, namely, for $r \in \{x, y\}$,

$$\sum_{u \in \partial \mathcal{T}_r \setminus \mathcal{F}_r} \mathbf{w}(u) = \sum_{u \in \partial \mathcal{T}_r} \mathbf{w}(u) \mathbb{1}_{\{\mathbf{w}(u) < \mathbf{w}_{\text{MIN}}\}} + \sum_{u \in \partial \mathcal{T}_r} \mathbf{w}(u) \mathbb{1}_{\{u \text{ is RED}\}}.$$

Let us first control the weight of RED half-edges. For $x \in \mathcal{R}$ and $y \in \mathcal{R} \setminus \mathcal{B}_K(x)$, all RED halfedges are at distance at least K from r, and thus have weight smaller than $2^{-K} \leq N^{-\frac{\log(2)}{6\log(\Delta-1)}}$ by our assumption that vertex degrees are at least 3. Moreover, using the upper bound (5.7), the total number of RED half-edges in \mathcal{T}_r is stochastically dominated by twice a binomial random variable $\operatorname{Bin}(k,q)$ where $k = O(N^{\frac{1}{2} + \frac{\log(2)}{15\log(\Delta)}})$ and $q = O(N^{-\frac{1}{2} + \frac{\log(2)}{15\log(\Delta)}})$. By Bennett's Inequality,

$$\mathbb{P}\left(\sum_{u\in\partial\mathcal{T}_r}\mathbbm{1}_{\{u \text{ is } \operatorname{RED}\}} > N^{\frac{\log(2)}{7\log(\Delta)}}\right) \le \exp\left(-\Omega\left(N^{\frac{\log(2)}{7\log(\Delta)}}\right)\right).$$

Hence, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\exists x \in \mathcal{R}, y \in \mathcal{R} \setminus \mathcal{B}_x, r \in \{x, y\}, \sum_{u \in \partial \mathcal{T}_r} \mathbf{w}(u) \mathbb{1}_{\{u \text{ is } \text{RED}\}} > \varepsilon\right) = o(1).$$

Let us now control the weight of paths with weight smaller than \mathbf{w}_{MIN} . To this end, consider $m = \lfloor \log N \rfloor$ independent NBRWS on G starting at r, each being stopped as soon as its weight falls below \mathbf{w}_{MIN} , and let A be the event that their trajectories form a tree of height less than t/2. Clearly,

$$\mathbb{P}\left(A \mid G\right) \geq \left(\sum_{u \in \partial \mathcal{T}_r} \mathbf{w}(u) \mathbb{1}_{\{\mathbf{w}(u) < \mathbf{w}_{\text{MIN}}\}}\right)^m$$

Taking expectation and using Markov inequality, we deduce that

$$\mathbb{P}\left(\sum_{u\in\partial\mathcal{T}_r}\mathbf{w}(u)\mathbb{1}_{\{\mathbf{w}(u)<\mathbf{w}_{\mathrm{MN}}\}}>\varepsilon\right)\leq\frac{\mathbb{P}\left(A\right)}{\varepsilon^m},$$

where the average is now taken over both the walks and the graph. To prove that the above probability is $o(1/N^2)$, it is enough to show that $\mathbb{P}(A) = o(1)^m$. To do so, we generate the *m* stopped NBRWs one after the other, revealing pairs along the way, as described in the proof of the lower bound. Given that the first $\ell - 1$ walks form a tree of height less than t/2, the conditional probability that the ℓ^{th} walk also fulfills the requirement is o(1), uniformly in $1 \leq \ell \leq m$. Indeed,

- either it attains length $s = \lceil 4 \log \log N \rceil$ before leaving the graph spanned by the first $\ell 1$ trajectories and reaching an unpaired half-edge: thanks to the tree structure, there are at most $\ell 1 < m$ possible trajectories to follow, each having weight at most 2^{-s} , so the conditional probability is at most $m2^{-s} = o(1)$.
- or the remainder of its trajectory after the first unpaired half-edge has weight less than $\Delta^s \mathbf{w}_{\text{MIN}}$: this part consists of at most t/2 half-edges which can be coupled with $(X_k^{\star})_{k=1}^{t/2}$ for a total-variation cost of $O(mt^2/N)$, and for N large enough

$$\mathbb{P}\left(\prod_{k=1}^{t/2} \frac{1}{\deg(X_k^{\star})} \le \Delta^s \mathbf{w}_{\text{MIN}}\right) \le \mathbb{P}\left(S_{t/2} - \frac{\mu t}{2} \ge \frac{\log(2)}{18\log(\Delta)}\log N\right),\,$$

which is o(1) by Chebychev Inequality.

Combining Lemma 5.3 and inequality 5.9, we obtain that

$$\max_{x \in \mathcal{R}} \max_{y \in \mathcal{R} \setminus \mathcal{B}_K(x)} \left\{ 1 - NP^t(x, \eta(y)) \right\} \le \sum_{\substack{u \in \mathcal{F}_x \\ v \in \mathcal{F}_y}} \mathbf{w}(u) \mathbf{w}(v) \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) > \theta\}} + o_{\mathbb{P}}(1) \,.$$

The proof of (5.4) will then be concluded by the following Lemma.

Lemma 5.4. For all $\varepsilon > 0$,

$$\mathbb{P}\left(\sum_{\substack{u\in\mathcal{F}_x\\v\in\mathcal{F}_y}}\mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v)>\theta\}}>\overline{\Phi}(\lambda)+\varepsilon\right)=o\left(\frac{1}{N^2}\right)\,.$$

Proof of Lemma 5.4. Set $m = \lceil (\log N)^2 \rceil$. Let $X^{(1)}, \ldots, X^{(m)}$ and $Y^{(1)}, \ldots, Y^{(m)}$ be 2m independent NBRWS of length t/2 starting at x and y respectively. Let B denote the event that their trajectories form two disjoint trees and that for all $1 \le k \le m$,

$$\prod_{\ell=1}^{t/2} \frac{1}{\deg(X_{\ell}^{(k)})} \prod_{\ell=1}^{t/2} \frac{1}{\deg(Y_{\ell}^{(k)})} > \theta.$$

Then clearly,

$$\mathbb{P}\left(B \mid G\right) \geq \left(\sum_{\substack{u \in \mathcal{F}_x \\ v \in \mathcal{F}_y}} \mathbf{w}(u) \mathbf{w}(v) \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) > \theta\}}\right)^m.$$

Averaging w.r.t. the graph, we see that

$$\mathbb{P}\left(\sum_{\substack{u\in\mathcal{F}_x\\v\in\mathcal{F}_y}}\mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v)>\theta\}} > \overline{\Phi}(\lambda) + \varepsilon\right) \leq \frac{\mathbb{P}\left(B\right)}{(\overline{\Phi}(\lambda)+\varepsilon)^m}.$$

Thus, it is enough to establish that $\mathbb{P}(B) \leq (\overline{\Phi}(\lambda) + o(1))^m$. We do so by generating the 2m walks $X^{(1)}, Y^{(1)}, \ldots, X^{(m)}, Y^{(m)}$ one after the other along with the underlying pairing, as above. Given that $X^{(1)}, Y^{(1)}, \ldots, X^{(\ell-1)}, Y^{(\ell-1)}$ already satisfy the desired property, the conditional chance that $X^{(\ell)}, Y^{(\ell)}$ also does is at most $\Phi(\lambda) + o(1)$, uniformly in $1 \leq \ell \leq m$. Indeed,

- either one of the two walks attains length $s = \lceil 4 \log \log N \rceil$ before leaving the graph spanned by the first $2(\ell - 1)$ trajectories and reaching an unpaired half-edge: thanks to the tree structure, there are at most $\ell - 1 < m$ possible trajectories to follow for each walk, each having weight at most 2^{-s} , so the conditional chance is at most $2m2^{-s} = o(1)$.
- or at least t 2s unpaired half-edges are encountered, and the product of their degrees falls below $\frac{1}{\theta}$ with conditional probability at most

$$\frac{4mt^2}{N} + \mathbb{P}\left(\prod_{k=1}^{t-2s} \deg(X_k^{\star}) < \frac{1}{\theta}\right) = \overline{\Phi}(\lambda) + o(1),$$

by the same coupling as above and Berry-Essen's inequality.

6 Comparing NBRW and SRW: entropies on Galton-Watson trees

In previous sections, we have seen that a crucial property of sparse random graphs generated by the configuration model is their locally tree-like structure (see for instance Lemma 2.1). Roughly put, if one sits at a random vertex chosen with probability proportional to its degree and looks around in a neighborhood of "small" radius, then one would see a tree. In the regular case, this would simply be a *d*-regular tree. For the configuration model with degree sequence $(\mathbf{d}(v))_{v \in V}$, the law of this tree would be close to the law of a Galton–Watson tree whose offspring distribution is directly related to the degree sequence in the following manner. Let $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, ...)$ be the vector of degree frequencies, i.e.

$$\mathbf{p}_k = \frac{1}{n} \sum_{v \in V} \mathbb{1}_{\mathbf{d}(v) = k} \,.$$

Since the root is chosen with probability proportional to its degree, the probability that the root vertex has degree k is $\mathbf{p}_{\star}(k) = \frac{n}{N}k\mathbf{p}_{k}$. Then, at subsequent steps, the pairing procedure entails that, provided "not too many" half-edges have been paired already, the probability that a given half-edge is paired to a half-edge attached to a vertex of degree k can also be approximated by $\mathbf{p}_{\star}(k)$. The law \mathbf{p}_{\star} is called the *size-biased distribution* of \mathbf{p} .

The graph G is then locally approximated by a random tree (T, ρ) , which is called an *augmented Galton–Watson tree* with *degree* distribution \mathbf{p}_{\star} , or equivalently with *offspring* distribution \mathbf{q}_{\star} given by $\mathbf{q}_{\star}(k) = \mathbf{p}_{\star}(k+1)$. It has the same law as a tree formed by joining by an edge the roots of two independent Galton–Watson trees with offspring distribution \mathbf{q}_{\star} , one of those two roots being ρ .

Typically, if $\Delta = O(1)$, then the total variation distance between the law of the neighborhood of radius $L = \frac{1}{5} \log_{\Delta - 1}(N)$ around a random vertex in G chosen with probability proportional to its degree and the law of (T, ρ) tends to 0 as $n \to +\infty$.

Let now (T, ρ) be an augmented Galton–Watson tree with offspring variable Z, such that $Z \ge 2$ and $\mathbb{E}Z < \infty$. For $k \ge 1$, let $T_k = \{z \in T, \operatorname{dist}(\rho, z) = k\}$ and use the notation $y \succ x$ to denote that y is a child of x and Z_x to be the number of children of x in T. Note that $Z_\rho \sim Z + 1$ and for all $x \ne \rho$, $Z_x \sim Z$. Let (X_t) and (Y_t) be respectively a NBRW and a SRW on T started at ρ^3 . Conditionally on the environment (T, ρ) , let $\mathbf{H}_{\rho}^T(X_t)$ the entropy of X_t :

$$\mathbf{H}_{\rho}^{T}(X_{t}) = \sum_{x \in T} \mathbb{P}_{\rho}^{T} \left(X_{t} = x \right) \log \frac{1}{\mathbb{P}_{\rho}^{T} \left(X_{t} = x \right)} \quad \text{and} \quad h_{t}^{X} = \mathbb{E}[\mathbf{H}_{\rho}^{T}(X_{t})].$$

Similarly, let $\mathbf{H}_{\rho}^{T}(Y_{t})$ the entropy of Y_{t} conditionally on (T, ρ) :

$$\mathbf{H}_{\rho}^{T}(Y_{t}) = \sum_{x \in T} \mathbb{P}_{\rho}^{T} \left(Y_{t} = x \right) \log \frac{1}{\mathbb{P}_{\rho}^{T} \left(Y_{t} = x \right)} \quad \text{and} \quad h_{t}^{Y} = \mathbb{E}[\mathbf{H}_{\rho}^{T}(Y_{t})].$$

³On a rooted tree, there is no problem with defining the NBRW on vertices.

We have

$$\mathbf{H}_{\rho}^{T}(X_{1}) = \mathbf{H}_{\rho}^{T}(Y_{1}) = \log Z_{\rho},$$

hence

$$h_1^X = h_1^Y = \mathbb{E}[\log(Z+1)]\,.$$

For the NBRW, observe that for all $t \ge 2$,

$$\begin{aligned} \mathbf{H}_{\rho}^{T}(Y_{t}) &= \sum_{y \in T_{t-1}} \sum_{x \succ y} \mathbb{P}_{\rho}^{T}(Y_{t-1} = y) \frac{1}{Z_{y}} \log \left(\frac{Z_{y}}{\mathbb{P}_{\rho}^{T}(Y_{t-1} = y)} \right) \\ &= \mathbf{H}_{\rho}^{T}(Y_{t-1}) + \sum_{y \in T_{t-1}} \mathbb{P}_{\rho}^{T}(Y_{t-1} = y) \log(Z_{y}) \,, \end{aligned}$$

hence

$$\mathbb{E}\left[\mathbf{H}_{\rho}^{T}(Y_{t}) \mid \mathcal{F}_{t-1}\right] = \mathbf{H}_{\rho}^{T}(Y_{t-1}) + \mathbb{E}[\log Z]$$

where (\mathcal{F}_t) is the filtration corresponding to the successive levels of the tree. The sequence $(\mathbf{H}_{\rho}^T(Y_t) - \mathbb{E}[\log(Z+1)] - (t-1)\mathbb{E}[\log Z])_{t\geq 1}$ is a centered martingale, and by the Martingale Convergence Theorem

$$\frac{\mathbf{H}_{\rho}^{T}(Y_{t})}{t} \xrightarrow[t \to +\infty]{} \mathbf{h}_{Y} = \mathbb{E}[\log Z] \quad \text{a.s.} .$$

As for the SRW, it was shown by Lyons, Pemantle, and Peres [14, Theorem 9.7] that there is $\mathbf{h}_X > 0$ such that

$$\frac{\mathbf{H}_{\rho}^{T}(X_{t})}{t} \xrightarrow[t \to +\infty]{} \mathbf{h}_{X} \quad \text{a.s.}$$

Taking expectations, we thus have

$$\lim_{t \to \infty} \frac{h_t^X}{t} = \mathbf{h}_X, \quad \text{and} \quad \lim_{t \to \infty} \frac{h_t^Y}{t} = \mathbf{h}_Y.$$
(6.1)

If $Z \sim \mathbf{q}_{\star}$ with \mathbf{q}_{\star} defined as above from the degree distribution, then $\mathbf{h}_{Y} = \mu$, with μ as defined in Section 5. Theorem 5.1 then states that **whp**, the NBRW on G generated by the configuration model with degree sequence $(\mathbf{d}(v))_{v \in V}$ has cutoff at time $\frac{\log N}{\mathbf{h}_{Y}}$ with window $\sqrt{\log N}$.

Cutoff for the SRW on G has been established by Berestycki, Lubetzky, Peres, and Sly [6] from a given starting vertex. The result has then been extended to the worst starting point by B., Lubetzky and Peres [3].

Theorem 6.1. Let G = (V, E) be a random graph on n vertices with vertex set V and degree distribution $(\mathbf{p}_k)_{k\geq 1}$; that is, the degree sequence $(\mathbf{d}(v))_{v\in V}$ is I.I.D. with distribution $\mathbb{P}(\mathbf{d}(v) = k) = \mathbf{p}_k$, conditioned on $\sum_{v\in V} \mathbf{d}(v)$ being even, and G is thereafter generated by the configuration model. Let Z be a random variable with distribution

$$\forall k \geq 1$$
, $\mathbb{P}(Z=k) \propto (k+1)\mathbf{p}_{k+1}$.

Assume that

$$\mathbb{P}(Z=0) = \mathbb{P}(Z=1) = 0$$
, $\mathbb{E}Z < \infty$, and $\mathbb{P}(Z > \Delta_n) = o\left(\frac{1}{n}\right)$.

where $\Delta_n = \exp\left\{(\log n)^{1/2-\delta}\right\}$ for some fixed $\delta > 0$. Then **whp**, the SRW on G has cutoff at time $\frac{\log n}{\mathbf{h}_X}$ with window $\sqrt{\log n}$.

Remark 6.2. Both for NBRW and SRW, the mixing time on G can be expressed in terms of the entropy on the Galton–Watson tree which approximates G locally. Roughly put, since the entropy of X_t grows like $\mathbf{h}_X t$ and the entropy of Y_t like $\mathbf{h}_Y t$, the mixing time of both walks can be interpreted as the time when the entropy becomes asymptotics to $\log n$ (or $\log N$ for NBRW), which, in the sparse regime is equivalent to the entropy of the stationary distribution. This cutoff phenomenon at the "entropic time" was also observed for other models of Markov chains in random environment. Let us mention the result of Bordenave, Caputo, and Salez [7]. For $i \in [n]$, let $\mathbf{p}_i = (\mathbf{p}_{i,j})_{j=1}^n$ be a probability distribution over [n]. From those vectors, generate a random transition matrix as follows: choose n independent uniform permutation over [n], denoted $\sigma_1, \ldots, \sigma_n$, and let P be the $n \times n$ matrix with entries

$$P(i,j) = \mathbf{p}_{i,\sigma_i^{-1}(j)}.$$

Let **h** be the average row entropy:

$$\mathbf{h} = \frac{1}{n} \sum_{i,j=1}^{n} \mathbf{p}_{i,j} \log \left(\frac{1}{\mathbf{p}_{i,j}} \right)$$

Assume that $\mathbf{h} = O(1)$, that

$$\max_{i \in [n]} \sum_{j=1}^{n} \mathbf{p}_{i,j} \left(\log \mathbf{p}_{i,j} \right)^2 = o(\log n) \,,$$

and that

$$\limsup_{n \to +\infty} \left\{ \frac{1}{n} \sum_{i,j} \mathbb{1}_{\mathbf{p}_{i,j} > 1-\varepsilon} \right\} \xrightarrow[\varepsilon \to 0^+]{} 0.$$

Then whp, the Markov chain with transition matrix P has cutoff at time $\frac{\log N}{h}$.

Given Theorem 6.1, a natural question to ask is: which walk mixes faster? Can we compare \mathbf{h}_X and \mathbf{h}_Y ? In the regular case, the answer is straightforward: as seen in Section 3 and 4, the mixing time on a random *d*-regular graph coincides with the time *t* at which this distance from the starting point is about $\log_{d-1} n$ (so as to contain almost all vertices in its range). This corresponds to $t = \log_{d-1} n + O(1)$ for the NBRW, and a slowdown of the SRW by a factor of d/(d-2) due to the reduced speed of random walk on a tree (with a coarser $O(\sqrt{\log n})$ -window due to the normal fluctuations of its height). The walks are mixed once they reach distance $\log_{d-1}(n)$, which is the typical distance in *G*. The SRW is slowed down by its reduced speed and the NBRW mixes faster. In the non-regular case, this interpretation in terms of reaching the typical distance falls apart. For the NBRW, $\mathbf{h}_Y = \mathbb{E} \log Z$, which satisfies $\mathbf{h}_Y < \log \mathbb{E}Z$ whenever Z is not a constant by Jensen's inequality. Hence, the NBRW mixes well after the time at which its range covers most vertices, unlike the regular setting. The same phenomenon occurs for the SRW: denoting by ν the limiting speed of the SRW on T, then

 $\mathbf{h}_X/\nu < \log \mathbb{E}Z$ whenever Z is not a constant (the "dimension drop" of harmonic measure, as shown in [14]). In the non-regular case, the distribution of the walk on a given level is no longer uniform, and different paths with equal length can have very different weights. Mixing occurs not only when almost all vertices are in the support of the walk, but when all paths have "reasonable" probability to be seen by the walk. To compare mixing times, comparing speeds is not sufficient anymore. One will have to take another effect into account: how well the walk is mixed *conditioned on a given level*.

To do so, let us introduce the notion of harmonic measure. Since $Z \ge 2$, the SRW (X_t) on T is transient. It escapes with asymptotic speed

$$\nu \stackrel{\text{p.s.}}{=} \lim_{t \to \infty} \frac{\operatorname{dist}(\rho, X_t)}{t} \,.$$

As shown in [14], $\nu = \mathbb{E}\left[\frac{Z-1}{Z+1}\right]$. Knowing that the walk escapes (at linear speed), we now want to understand *where* it escapes. The loop-erasure of X_t defines a unique infinite ray $\xi = (\xi_t)$, whose distribution is called the harmonic measure of the walk. If ∂T denotes the boundary of T, i.e. the set of infinite rays from ρ , one can endow ∂T with the following metric d: for all $\beta, \eta \in \partial T$, $d(\beta, \eta) = e^{-|\beta \wedge \eta|}$, where $|\beta \wedge \eta|$ is the length of the longest common prefix of β and η . With this metric, the Hölder exponent of the harmonic measure at ξ , which is also the Hausdorff dimension, is

$$\mathbf{d} \stackrel{\text{a.s.}}{=} \lim_{t \to \infty} -\frac{1}{t} \log \mathbf{P}_{\rho}^{T}(\xi_{t}) , \qquad (6.2)$$

where $\mathbf{P}_{\rho}^{T}(\cdot) = \mathbb{P}_{\rho}^{T}(\cdot \in \xi)$. As shown in [14], letting η_{t} be the location of the first visit of the walk to T_{t} and $\mathbf{Q}_{\rho}^{T}(\cdot) = \mathbb{P}_{\rho}^{T}(\cdot \in \eta)$, we have

$$\mathbf{d} \stackrel{\text{a.s.}}{=} \lim_{t \to \infty} -\frac{1}{t} \log \mathbf{Q}_{\rho}^{T}(\eta_{t}) \ .$$

With this characterization, we see that **d** captures a notion of asymptotic entropy of the walk when it hits a given level. The asymptotic entropy of (X_t) can be decomposed into

$$\mathbf{h}_X = \nu \mathbf{d}$$

Clearly ν , the speed of X, is less than 1, the speed of Y. On the other hand, as soon as Z is not constant, $\mathbf{d} > \mathbb{E}[\log Z]$: the Hausdorff dimension of the harmonic measure of X is larger than the one of Y. This inequality was conjectured since [14] and was recently proved by Lin [12]. The two effects thus go in opposite direction and comparing entropies is not as direct as in the regular case.

Proposition 6.3. Assume $Z \ge 2$ and $\mathbb{E}Z < \infty$. Then $\mathbf{h}_X < \mathbf{h}_Y$.

Proof of Proposition 6.3. We will prove Proposition 6.3 under the stronger assumption that $Z \ge 3$. We first need the following result (cf., e.g., the proof of Theorem 3.2 in [4] and Corollary 10 in [5]), which was first observed in the case of random walks on groups by [10]. (Entropy of random walks on random stationary environments were thereafter studied in [9]).

Lemma 6.4. The map $t \mapsto (h_t^X - h_{t-1}^X)$ is decreasing.

Proof of Lemma 6.4. Consider the joint entropy of X_1 and X_t given T:

$$\mathbf{H}_{\rho}^{T}(X_{1}, X_{t}) = \sum_{x, y \in T} \mathbb{P}_{\rho}^{T} (X_{1} = x, X_{t} = y) \log \frac{1}{\mathbb{P}_{\rho}^{T} (X_{1} = x, X_{t} = y)},$$

and let $h_{1,t}^X = \mathbb{E}[\mathbf{H}_{\rho}^T(X_1, X_t)]$. We have

$$\mathbf{H}_{\rho}^{T}(X_{1}, X_{t}) = \mathbf{H}_{\rho}^{T}(X_{1}) + \sum_{x \in T} \mathbb{P}_{\rho}^{T}(X_{1} = x) \sum_{y \in T} \mathbb{P}_{x}^{T}(X_{t-1} = y) \log \frac{1}{\mathbb{P}_{x}^{T}(X_{t-1} = y)}.$$

and taking expectation gives

$$h_{1,t}^X = h_1^X + \mathbb{E}[\mathbf{H}_{X_1}^T(X_{t-1})] = h_1^X + h_{t-1}^X,$$

where the last equality is due to the fact that the environment (T, ρ) is stationary for the SRW, i.e. (T, ρ) and (T, X_1) have the same distribution. Therefore,

$$h_t^X - h_{t-1}^X = h_t^X - h_{1,t}^X + h_1^X = \mathbb{E}[\mathbf{H}_{\rho}^T(X_t) - \mathbf{H}_{\rho}^T(X_1, X_t)] + h_1^X$$

Conditioned on T, the term $\mathbf{H}_{\rho}^{T}(X_{1}, X_{t}) - \mathbf{H}_{\rho}^{T}(X_{t})$ is the conditional entropy of X_{1} given X_{t} , denoted by $\mathbf{H}_{\rho}^{T}(X_{1} \mid X_{t})$. Since X_{1}, X_{t+1} are conditionally independent given X_{t} , and extra information cannot increase entropy, we have

$$\mathbf{H}_{\rho}^{T}(X_{1} \mid X_{t}) = \mathbf{H}_{\rho}^{T}(X_{1} \mid X_{t}, X_{t+1}) \leq \mathbf{H}_{\rho}^{T}(X_{1} \mid X_{t+1}).$$

Hence the sequence $\left(\mathbf{H}_{\rho}^{T}(X_{t}) - \mathbf{H}_{\rho}^{T}(X_{1}, X_{t})\right)_{t \geq 1}$ is decreasing, and so is its expectation.

The fact that $(h_t^X - h_{t-1}^X)$ is decreasing in t implies that, for every t,

$$h_t^X - h_1^X \le (t-1)(h_2^X - h_1^X),$$

from which we see that it suffices to show that $h_2^X < h_2^Y = \mathbb{E}[\log(Z+1)] + \mathbb{E}[\log Z]$ in order to conclude that $\mathbf{h}_X < \mathbf{h}_Y^4$. We have

$$\mathbf{H}_{\rho}^{T}(X_{2}) = \mathbb{P}_{\rho}(X_{2} = \rho) \log \frac{1}{\mathbb{P}_{\rho}(X_{2} = \rho)} + \sum_{y \in T_{2}} \mathbb{P}_{\rho}(X_{2} = y) \log \frac{1}{\mathbb{P}_{\rho}(X_{2} = y)}.$$
 (6.3)

Let A and B denote respectively the first and second terms in the right-hand side of (6.3). By concavity of the logarithm,

$$A = -\sum_{x \in T_1} \frac{1}{Z_{\rho}(Z_x + 1)} \log \left(\sum_{y \in T_1} \frac{1}{Z_{\rho}(Z_y + 1)} \right)$$

$$\leq \frac{1}{(Z_{\rho})^2} \sum_{x,y \in T_1} \frac{\log(Z_y + 1)}{Z_x + 1}.$$

⁴To prove the result under the weaker assumption $Z \ge 2$, looking at the entropy at time 2 is not sufficient, one needs to look at the entropy at time 3 and calculations are (way) more involved.

Hence

$$\mathbb{E}[A] \leq \mathbb{E}\left[\frac{1}{Z+1}\right] \mathbb{E}\left[\frac{\log(Z+1)}{Z+1}\right] + \mathbb{E}\left[\frac{Z}{Z+1}\right] \mathbb{E}\left[\frac{1}{Z+1}\right] \mathbb{E}\left[\log(Z+1)\right]$$

$$\leq \mathbb{E}\left[\frac{1}{Z+1}\right] \mathbb{E}\left[\log(Z+1)\right],$$

where the last inequality comes from the fact that $\operatorname{Cov}\left(\frac{Z}{Z+1}, \log(Z+1)\right) \ge 0$. As for B, we have

$$B = \sum_{x \in T_1} \sum_{y \succ x} \frac{1}{Z_{\rho}(Z_x + 1)} \log \left(Z_{\rho}(Z_x + 1) \right) = \sum_{x \in T_1} \frac{Z_x}{Z_{\rho}(Z_x + 1)} \log \left(Z_{\rho}(Z_x + 1) \right) \,,$$

which entails

$$\mathbb{E}[B] = \mathbb{E}\left[\frac{Z}{Z+1}\right] \mathbb{E}\left[\log(Z+1)\right] + \mathbb{E}\left[\frac{Z}{Z+1}\log(Z+1)\right] \,.$$

We obtain

$$h_2^X \le \mathbb{E}[\log(Z+1)] + \mathbb{E}\left[\frac{Z}{Z+1}\log(Z+1)\right],$$

and we conclude by noticing that the function $x \mapsto \log x - \frac{x}{x+1}\log(x+1)$ is positive for $x \ge 3$.

References

- R. Basu, J. Hermon, and Y. Peres. Characterization of cutoff for reversible markov chains. The Annals of Probability, 45(3):1448–1487, 2017.
- [2] A. Ben-Hamou and J. Salez. Cutoff for non-backtracking random walks on sparse random graphs. The Annals of Probability, 45(3):1752–1770, 2017.
- [3] A. Ben-Hamou, E. Lubetzky, and Y. Peres. Comparing mixing times on sparse random graphs. Annales de l'Institut Henri Poincaré, 55(2):1116–1130, 2019.
- [4] I. Benjamini and N. Curien. Ergodic theory on stationary random graphs. *Electron. J.* Probab, 17(93):1–20, 2012.
- [5] I. Benjamini, H. Duminil-Copin, G. Kozma, and A. Yadin. Disorder, entropy and harmonic functions. Ann. Probab., 43(5):2332-2373, 2015. doi: 10.1214/14-AOP934. URL http://dx.doi.org/10.1214/14-AOP934.
- [6] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *The Annals of Probability*, 46(1):456–490, 2018.
- [7] C. Bordenave, P. Caputo, and J. Salez. Cutoff at the ?entropic time? for sparse markov chains. *Probability Theory and Related Fields*, 173(1-2):261–292, 2019.
- [8] S. Chatterjee. Stein's method for concentration inequalities. Probability theory and related fields, 138(1):305–321, 2007.
- [9] V. Kaimanovich. Boundary and entropy of random walks in random environment. Prob. Theory and Math. Stat, 1:573–579, 1990.
- [10] V. A. Kaimanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. Ann. Probab., pages 457–490, 1983.
- [11] D. A. Levin and Y. Peres. Markov chains and mixing times, volume 107. American Mathematical Soc., 2017.
- [12] S. Lin. Harmonic measure for biased random walk in a supercritical galton-watson tree. arXiv preprint arXiv:1707.01811, 2017.
- [13] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. Duke Mathematical Journal, 153(3):475–510, 2010.
- [14] R. Lyons, R. Pemantle, and Y. Peres. Ergodic theory on galton?watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory and Dynamical Systems*, 15(3):593–619, 1995.

7 Exercises

7.1 The product condition is not sufficient for cutoff for reversible chains

- 1. Show that, for a sequence of irreducible, aperiodic, reversible chains, the condition $\frac{t_{\text{MIX}}^{(n)}}{t_{\text{PEL}}^{(n)}} \xrightarrow{\rightarrow \infty} +\infty \text{ is a necessary condition for cutoff.}$
- 2. Let P be an irreducible, aperiodic, reversible transition matrix, with stationary distribution π . For $\theta \in]0,1[$, let $\widehat{P} = (1-\theta)P + \theta\Pi$, where Π whose rows are given by the stationary probability vector π .
 - (a) Show that \widehat{P} is still reversible with respect to π .
 - (b) Show that $\widehat{\mathcal{D}}(t) = (1-\theta)^t \mathcal{D}(t)$.
 - (c) Deduce that

$$\widehat{t_{\rm REL}} = \frac{t_{\rm REL}}{\theta t_{\rm REL} + 1 - \theta} \, \cdot \,$$

3. Let (P_n) be a sequence of irreducible, aperiodic, reversible transition matrices, with stationary distribution π_n , and assume that it has cutoff. For $\theta_n \in]0,1[$, let $\hat{P}_n = (1 - \theta_n)P_n + \theta_n \Pi_n$, where Π_n whose rows are given by the stationary probability vector π_n . Assume that

$$t_{\text{REL}}^{(n)} \ll \frac{1}{\theta_n} \ll t_{\text{MIX}}^{(n)} \,.$$

(a) Show that, for all $\alpha > 0$,

$$\widehat{\mathcal{D}}_n\left(\frac{\alpha}{\theta_n}\right) \underset{n \to \infty}{\longrightarrow} e^{-\alpha}.$$

(b) Show that (\widehat{P}_n) still satisfies the product condition.

7.2 Cutoff for the lazy RW on the hypercube

Let $\Omega = \{0, 1\}^n$ be the *n*-dimensional cube and let (X_t) be the lazy SRW on Ω : at each step, choose a coordinate uniformly at random in $\{1, \ldots, n\}$ and refresh the bit on this coordinate (set it to 0 with probability 1/2, or 1 with probability 1/2). Since the hypercube is transitive, $\mathcal{D}(t) = \|\mathbb{P}_0(X_t \in \cdot) - \pi\|_{\mathrm{TV}}$, where $\mathbf{0} = (0, \ldots, 0)$ and π is uniform over Ω .

1. Lower bound.

(a) Let $W_t = H(X_t)$ be the Hamming weight of X_t (the number of ones in X_t). Show that

$$\mathbb{E}_0[W_t] = \frac{n}{2} \left(1 - \left(1 - \frac{1}{n}\right)^t \right)$$

(b) Writing $W_t = \sum_{i=1}^n \xi_i$, with

$$\xi_i = \begin{cases} 1 & \text{if } i \text{ has been actually updated an odd number of times before } t, \\ 0 & \text{otherwise.} \end{cases}$$

show that

$$\operatorname{Var}_{0}(W_{t}) \leq \sum_{i=1}^{n} \operatorname{Var}(\xi_{i}) \leq \frac{n}{2}.$$

(c) Letting
$$A = \left\{ x \in \Omega, \ H(x) \leq \frac{n}{2} - \frac{n}{4} \left(1 - \frac{1}{n} \right)^t \right\}$$
, show that, if $t = \frac{c}{2} n \log n$ with $c < 1$,
 $\mathbb{P}_0(X_t \in A) \to 1$ and $\pi(A) \to 0$.

2. Upper bound.

(a) Show that

$$4\mathcal{D}(t)^2 \le 2^n P^{2t}(\mathbf{0}, \mathbf{0}) - 1 = \sum_{j=2}^{2^n} \lambda_j^{2t},$$

where $1 > \lambda_2 \ge \ldots \lambda_{2^n} \ge 0$ are the eigenvalues of P.

- (b) Determine the spectrum of P.
- (c) Conclude by showing that if c > 1, $\mathcal{D}\left(\frac{c}{2}n\log n\right) \to 0$ as $n \to +\infty$.

7.3 Cutoff for the top-to-random shuffle

Consider the following method for shuffling a deck of n cards: at each step, take the card which is on the top of the deck, and relocate it uniformly at random in one of the n possible locations. This defines a Markov chain (X_t) on the symmetric group S_n , with uniform stationary distribution. Since it is transitive, we may assume without loss of generality that we start at the identity permutation (card 1 is at the top, card n at the bottom).

1. Upper bound.

- (a) Let τ be the time following the first time when card n is on the top of the desk. Show that τ is a strong stationary time.
- (b) Show that, for all λ > 0, P_{id} (τ > n log n + λn) ≤ e^{-λ}. *Hint:* you may view τ as a sum of independent Geometric random variables, as in the coupon collector problem.

2. Lower bound.

(a) Consider the event A_r that the $r \ge 1$ cards that were originally at the bottom of the deck are still in the same relative order. Notice that $\mathbb{P}_{id}(X_t \in A_r) \ge \mathbb{P}_{id}(\tau_r \ge t)$, where τ_r is the first time when card number n - r + 1 arrives at the top. Show that, for $t = n \log n - \lambda n$, and $r = \lfloor e^{\lambda/2} \rfloor$,

$$\mathbb{P}_{\mathbf{id}}(\tau_r \ge t) \ge 1 - \varepsilon(\lambda) \,,$$

where $\varepsilon(\lambda) \to 0$ as $\lambda \to +\infty$.

(b) Show that, for $r = |e^{\lambda/2}|, \pi(A_r) \leq \varepsilon(\lambda)$, where $\varepsilon(\lambda) \to 0$ as $\lambda \to +\infty$.

7.4 Carne-Varopoulos upper bound

The goal of this exercise is to prove Carne-Varopoulos inequality, which states that for a reversible chain P,

$$P^{t}(x,y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}} \exp\left\{-\frac{\mathrm{d}(x,y)^{2}}{2t}\right\} ,$$

where d(x, y) is the length of the minimal path from x to y.

Recall that Chebychev polynomials are defined by the recurrence relation

$$Q_{k+1}(z) = 2zQ_k(z) - Q_{k-1}(z),$$

and that we have $Q_k(\cos\theta) = \cos(k\theta)$.

- 1. Show that $z^t = \sum_{k=0}^t \mathbb{P}(|S_t| = k)Q_k(z)$, where S_t is a simple random walk on \mathbb{Z} started at the origin.
- 2. Deduce that $P^t(x,y) = \sum_{k=d(x,y)}^t \mathbb{P}(|S_t| = k)Q_k(P)(x,y).$
- 3. Show that the spectrum of $Q_k(P)$ is included in [-1,1]. In particular, for all $f, g \in \ell^2(\pi)$, $| < Q_k(P)f, g >_{\pi} | \le ||f||_{\pi} ||g||_{\pi}.$
- 4. Show that $Q_k(P)(x,y) \le \sqrt{\frac{\pi(y)}{\pi(x)}}$.
- 5. Conclude with Hoeffding Inequality.
- 6. Show that Carne-Varopoulos upper bound implies the following lower bound on the mixing time of a simple random walk on a graph with n vertices and diameter D: for all $\varepsilon \in]0, 1/2[$,

$$t_{\rm MIX}(\varepsilon) \ge \frac{D^2}{13\log n}$$

7.5 Typical distance in random regular graphs

Let G_n be a random *d*-regular graph on *n* vertices, with $d \ge 3$ fixed. Show that $\frac{D_n}{\log_{d-1} n} \xrightarrow{\mathbb{P}} 1$, where D_n is the distance between two uniformly chosen vertices in G_n .

7.6 Concentration with exchangeable pairs (proof of Lemma 5.2)

A pair of real-valued random variables (W, W') is called *exchangeable* of $(W, W') \stackrel{d}{=} (W', W) *$. If in addition, it satisfies $\mathbb{E}[W' - W \mid W] = -\lambda W$, for some $0 < \lambda < 1$, then it is called an λ -Stein pair. The following theorem is due to Chatterjee [8].

Theorem 7.1. Let (W, W') be an λ -Stein pair with $Var(W) = \sigma^2 < \infty$. Assume that there exist $a, b \geq 0$ such that

$$\mathbb{E}\left[(W' - W)^2 \mid W \right] \le 2\lambda \left(bW + c \right) \,.$$

Then, for all a > 0,

$$\mathbb{P}(W > a) \le \exp\left\{-\frac{a^2}{2c+2ba}\right\}$$
 and $\mathbb{P}(W < -a) \le \exp\left\{-\frac{a^2}{2c}\right\}$.

Part A. The goal of this first part is to prove Theorem 7.1.

1. Let $m(\theta) = \mathbb{E}\left[e^{\theta W}\right]$. Show that

$$m'(\theta) = \frac{\mathbb{E}\left[(W' - W)(e^{\theta W'} - e^{\theta W}) \right]}{2\lambda}$$

2. Using that for all x > y, $\frac{e^x - e^y}{x - y} \le \frac{e^x + e^y}{2}$, show that for all $\theta \in \mathbb{R}$, $|m'(\theta)| \le |\theta| (bm'(\theta) + cm(\theta))$. 3. Show that for $0 < \theta < 1/b$,

$$\log m(\theta) \leq \int_0^\theta \frac{cu}{1-bu} \mathrm{d} u \leq \frac{c\theta^2}{2(1-b\theta)}$$

- 4. Using the Markov bound $\mathbb{P}(W > a) \leq e^{-\theta a} m(\theta)$ with $\theta > 0$, conclude for the upper tail.
- 5. For $\theta < 0$, use the fact that $m'(\theta) < 0$ to obtain $\log m(\theta) \le \frac{c\theta^2}{2}$.
- 6. Using the Markov bound $\mathbb{P}(W < -a) \leq e^{\theta a} m(\theta)$ for $\theta < 0$, conclude with the bound on the lower tail.

Part B. The goal of this second part is to prove Lemma 5.2 using Theorem 7.1.

Let \mathcal{I} be an even set, $(\omega_{i,j})_{(i,j)\in\mathcal{I}^2}$ an array of non-negative weights, and η a uniform random pairing on \mathcal{I} . Consider the centered variable

$$W = \sum_{i \in \mathcal{I}} \omega_{i,\eta(i)} - m,$$

where $m = \frac{1}{|\mathcal{I}|-1} \sum_{i \in \mathcal{I}} \sum_{j \neq i} \omega_{i,j}$. We want to show that for all a > 0,

$$\mathbb{P}(W > a) \le \exp\left\{-\frac{a^2}{4\theta m + 2\theta a}\right\} \quad \text{and} \quad \mathbb{P}(W < -a) \le \exp\left\{-\frac{a^2}{4\theta m}\right\}$$

where $\theta = \max_{i \neq j} (\omega_{i,j} + \omega_{j,i}).$

To this end, let W' be the corresponding quantity for the pairing η' obtained from η by performing a random switch: two indices i, j are sampled uniformly at random from \mathcal{I} without replacement, and the pairs $\{i, \eta(i)\}, \{j, \eta(j)\}$ are replaced with the pairs $\{i, j\}, \{\eta(i), \eta(j)\}$. Note that if $\Delta_{i,j}$ is the induced change in the total weight when i, j are chosen, then

$$\Delta_{i,j} = \omega_{i,j} + \omega_{j,i} + \omega_{\eta(i),\eta(j)} + \omega_{\eta(j),\eta(i)} - \omega_{i,\eta(i)} - \omega_{\eta(i),i} - \omega_{j,\eta(j)} - \omega_{\eta(j),j}$$

- 1. Show that (W, W') is a λ -Stein pair, with $\lambda = \frac{4}{|\mathcal{I}|}$.
- 2. Regarding the square $\Delta_{i,j}^2 = |\Delta_{i,j}| |\Delta_{i,j}|$, bounding the first copy of $|\Delta_{i,j}|$ by 2θ and the second by changing all minus signs to plus signs, show that

$$\mathbb{E}\left[(W' - W)^2 \mid W \right] \le \frac{8\theta}{|\mathcal{I}|} \left(2m + W \right).$$

and conclude.

8 Exercises: correction

8.1 The product condition is not sufficient for cutoff for reversible chains

1. Assume that the chain has cutoff, i.e. $\frac{t_{\text{MIX}}(\varepsilon)}{t_{\text{MIX}}} \to 1$ as $n \to +\infty$ Using the inequality $t_{\text{MIX}}(\varepsilon) \ge (t_{\text{REL}} - 1) \log(1/2\varepsilon)$, we have, for all $\varepsilon \in (0, 1)$,

$$\frac{t_{\text{MIX}}}{t_{\text{REL}}} \underset{n \to \infty}{\sim} \frac{t_{\text{MIX}}(\varepsilon)}{t_{\text{REL}}} \ge \left(1 - \frac{1}{t_{\text{REL}}}\right) \log\left(\frac{1}{2\varepsilon}\right) \,.$$

Since the right-hand side can be arbitrarily large by taking ε small enough, this shows that $\frac{t_{\text{MIX}}}{t_{\text{REL}}} \to +\infty$.

2. (a) Since P is reversible w.r.t. π ,

$$\pi(x)\widehat{P}(x,y) = (1-\theta)\pi(x)P(x,y) + \theta\pi(x)\pi(y) = (1-\theta)\pi(y)P(y,x) + \theta\pi(x)\pi(y) = \pi(y)\widehat{P}(y,x) \,.$$

(b) As soon as the chain makes a transition according to π , it remains stationary from then on. Hence $\widehat{P}^t(x,\cdot) = (1-\theta)^t P^t(x,\cdot) + (1-(1-\theta)^t) \pi(\cdot)$ and

$$\widehat{\mathcal{D}}(t) = \sum_{y} \left((1-\theta)^t P^t(x,y) - (1-\theta)^t \pi(y) \right)_+ = (1-\theta)^t \mathcal{D}(t) \,.$$

(c) Let us denote by λ_{\star} and $\widehat{\lambda}_{\star}$ the the second largest eigenvalue in absolute value of the patrix P and \widehat{P} respectively. Using that $\mathcal{D}(t)^{1/t} \to \lambda_{\star}$ and $\widehat{\mathcal{D}}(t)^{1/t} \to \widehat{\lambda}_{\star}$, we have

$$\widehat{\lambda}_{\star} = (1 - \theta) \lambda_{\star}$$

Writing $\lambda_{\star} = 1 - \frac{1}{t_{\text{REL}}}$ and $\hat{\lambda}_{\star} = 1 - \frac{1}{\widehat{t_{\text{REL}}}}$ and rearranging, we obtain

$$\widehat{t_{\text{REL}}} = \frac{t_{\text{REL}}}{\theta t_{\text{REL}} + 1 - \theta} \, \cdot \,$$

- 3. (a) Using question 2.(b), we have $\widehat{\mathcal{D}}_n\left(\frac{\alpha}{\theta_n}\right) = (1-\theta_n)^{\alpha/\theta_n} \mathcal{D}_n\left(\frac{\alpha}{\theta_n}\right)$. Since $1/\theta_n \ll t_{\text{MIX}}^{(n)}$, we have $\mathcal{D}_n\left(\frac{\alpha}{\theta_n}\right) \to 1$, which yields $\widehat{\mathcal{D}}_n\left(\frac{\alpha}{\theta_n}\right) \to e^{-\alpha}$. In particular, the mixing time of (\widehat{P}_n) is of order $\frac{1}{\theta_n}$ and there is no cutoff.
 - (b) By question 2.(c) and the assumptions on θ_n , we have $\widehat{t_{\text{REL}}}^{(n)} \sim t_{\text{REL}}^{(n)}$ and $\widehat{t_{\text{MIX}}}^{(n)} \approx \frac{1}{\theta_n} \gg t_{\text{REL}}^{(n)}$. Hence the sequence of chains (\widehat{P}_n) still satisfies the product condition.

8.2 Cutoff for the lazy RW on the hypercube

1. Lower bound.

(a) Let us define $Z_t = H(X_t) - \frac{n}{2}$. Conditionally on X_t , we have

$$Z_{t+1} - Z_t = \begin{cases} 1 & \text{with probability } \frac{1}{2} \left(\frac{1}{2} - \frac{Z_t}{n} \right), \\ -1 & \text{with probability } \frac{1}{2} \left(\frac{1}{2} + \frac{Z_t}{n} \right), \\ 0 & \text{otherwise.} \end{cases}$$

Taking expectation, we have $\mathbb{E}_0[Z_{t+1}] = (1 - \frac{1}{n}) \mathbb{E}_0[Z_t]$. By induction and using $Z_0 = -\frac{n}{2}$, we obtain $\mathbb{E}_0[Z_t] = -\frac{n}{2} (1 - \frac{1}{n})^t$, i.e. $\mathbb{E}_0[W_t] = \frac{n}{2} (1 - (1 - \frac{1}{n})^t)$.

(b) For $1 \le i \le n$, let N_i be the number of times coordinate *i* is chosen between times 1 and *t*. Since each time coordinate *i* is chosen, it is independently flipped with probability 1/2, we have

$$\mathbb{P}(\xi_i = 1) = \sum_{k=1}^t \mathbb{P}(N_i = k) \mathbb{P}(\operatorname{Bin}(k, 1/2) \text{ is odd}) = \frac{1}{2} \mathbb{P}(N_i > 0) = \frac{1}{2} \left(1 - \left(1 - \frac{1}{n}\right)^t \right).$$

Now, let us show that, for $i \neq j$, the variables ξ_i and ξ_j are negatively correlated. We have

$$\mathbb{P}(\xi_i = 1, \xi_j = 1) = \frac{1}{4} \mathbb{P}(N_i > 0, N_j > 0) = \frac{1}{4} \left(1 - 2\left(1 - \frac{1}{n}\right)^t + \left(1 - \frac{2}{n}\right)^t \right).$$

Since $\left(1-\frac{2}{n}\right)^t \leq \left(1-\frac{1}{n}\right)^{2t}$, we have $\mathbb{P}(\xi_i=1,\xi_j=1) \leq \mathbb{P}(\xi_i=1)\mathbb{P}(\xi_j=1)$. Hence

$$\operatorname{Var}_{0}(W_{t}) \leq \sum_{i=1}^{n} \operatorname{Var}(\xi_{i}) \leq \sum_{i=1}^{n} \mathbb{P}(\xi_{i}=1) \leq \frac{n}{2}$$

(c) Let $t = \frac{c}{2}n \log n$ with c < 1. Since $\mathbb{E}_0[W_t] = \frac{n}{2} \left(1 - \left(1 - \frac{1}{n}\right)^t \right)$, and by Chebyshev Inequality,

$$\mathbb{P}_{0}(X_{t} \notin A) = \mathbb{P}_{0}\left(W_{t} - \mathbb{E}_{0}[W_{t}] > \frac{n}{4}\left(1 - \frac{1}{n}\right)^{t}\right) \leq \frac{16\operatorname{Var}_{0}(W_{t})}{n^{2}\left(1 - \frac{1}{n}\right)^{2t}} \leq \frac{8}{n\left(1 - \frac{1}{n}\right)^{cn\log n}},$$

which tends to 0 as $n \to +\infty$ since c < 1. On the other hand, again by Chebyshev Inequality,

$$\pi(A) = \mathbb{P}\left(\operatorname{Bin}(n, 1/2) \le \frac{n}{2} - \frac{n}{4}\left(1 - \frac{1}{n}\right)^t\right) \le \frac{4}{n\left(1 - \frac{1}{n}\right)^{cn\log n}} \longrightarrow 0.$$

This shows that, for c < 1, $\mathcal{D}_n\left(\frac{cn\log n}{2}\right) \to 1$.

2. Upper bound.

(a) By Cauchy-Schwarz Inequality and symmetry of P^t ,

$$4\mathcal{D}(t)^2 = \left(\sum_{y\in\Omega} \left| P^t(\mathbf{0}, y) - \frac{1}{2^n} \right| \right)^2 \le 2^n \sum_{y\in\Omega} P^t(\mathbf{0}, y)^2 - 1 = 2^n P^{2t}(\mathbf{0}, \mathbf{0}) - 1.$$

Since the hypercube is transitive, the return probability $P^{2t}(x, x)$ does not depend on x and we can write

$$2^n P^{2t}(\mathbf{0}, \mathbf{0}) = \sum_{x \in \Omega} P^{2t}(x, x)$$

Using the spectral representation with (φ_j) an orthonormal basis of eigenvectors, we have

$$\sum_{x \in \Omega} P^{2t}(x, x) - 1 = \sum_{x \in \Omega} 2^{-n} \sum_{j=2}^{2^n} \lambda_j^{2t} \varphi_j(x)^2 = \sum_{j=2}^{2^n} \lambda_j^{2t}.$$

(b) For $n \ge 1$ and $A \subset [n]$, let f_A be defined by

$$\forall x \in \{0,1\}^n, \ f_A(x) = (-1)^{\sum_{i \in A} x_i}.$$

Note that $Pf_A = \left(1 - \frac{|A|}{n}\right) f_A$ and that the family $(f_A)_{A \subset [n]}$ is independent. Hence, for all $0 \le k \le n$, $1 - \frac{k}{n}$ is an eigenvalue of P with multiplicity $\binom{n}{k}$.

(c) Let $t = \frac{c}{2}n \log n$, with c > 1.

$$\sum_{j=2}^{2^n} \lambda_j^{2t} = \sum_{k=1}^n \binom{n}{k} \left(1 - \frac{k}{n}\right)^{2t} \le \sum_{k=1}^n \binom{n}{k} \left(\frac{1}{n^c}\right)^k = \left(1 + \frac{1}{n^c}\right)^n - 1 \longrightarrow 0.$$

Hence, for c > 1, $\mathcal{D}_n\left(\frac{cn\log n}{2}\right) \to 0$.

8.3 Cutoff for the top-to-random shuffle

1. Upper bound.

- (a) Given that at time t there are k cards under card n, each of the k! possible orderings of those k cards are equally likely (this can be seen by induction). Hence at time $\tau 1$, all (n 1)! orderings of cards $1, \ldots, n 1$ under card n are equally likely, and one step after that, the distribution is uniform over S_n .
- (b) Note that τ can be decomposed as

$$\tau = G_{n-1} + \dots + G_1 + 1 \,,$$

where G_i is the time taken by card n to go from position i + 1 to position i. The variables (G_i) are independent and $G_i \sim \text{Geom}\left(\frac{n-i}{n}\right)$. The variable τ has exactly the same distribution as the first time all items have been sampled at least once when sampling with replacement in a set of n items. By a union bound,

$$\mathbb{P}_{\mathbf{id}}\left(\tau > n\log n + \lambda n\right) \le n\left(1 - \frac{1}{n}\right)^{n\log n + \lambda n} \le e^{-\lambda}$$

2. Lower bound.

(a) As above the variable τ_r can be decomposed as

$$\tau_r = G_{n-r} + \dots + G_1 \,,$$

where G_i is the time taken by card n - r + 1 to go from position i + 1 to position i. Since $G_i \sim \text{Geom}\left(\frac{n-i}{n}\right)$, we have

$$\mathbb{E}\tau_r = n \sum_{i=r}^{n-1} \frac{1}{i} \ge n \left(\log(n) - \log(r-1) - 1 \right) \ge n \log n - \frac{\lambda n}{2} - n,$$

for $r = \lfloor e^{\lambda/2} \rfloor$. Hence by Chebyshev Inequality,

$$\mathbb{P}_{\mathbf{id}}(\tau_r < n\log n - \lambda n) \le \mathbb{P}_{\mathbf{id}}\left(\tau_r - \mathbb{E}\tau_r < -\frac{\lambda n}{2} + n\right) \le \frac{\operatorname{Var}(\tau_r)}{(\lambda/2 + 1)^2 n^2}.$$

Since $\operatorname{Var}(\tau_r) \leq n^2 \sum_{i=r}^{n-1} \frac{1}{i^2} = O(n^2)$, this yields the desired result. (b) We have $\pi(A_r) = \frac{1}{r!}$.

8.4 Carne-Varopoulos upper bound

1. For $\theta \in \mathbb{R}$, we have

$$(\cos\theta)^{t} = \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^{t} = \mathbb{E}\left[e^{i\theta S_{t}}\right] = \mathbb{E}\left[\cos\left(\theta|S_{t}|\right)\right]$$
$$= \sum_{k=0}^{t} \mathbb{P}(|S_{t}| = k)\cos(k\theta) = \sum_{k=0}^{t} \mathbb{P}(|S_{t}| = k)Q_{k}(\cos\theta)$$

Taking $z = \cos \theta$, we get the desired identity.

2. Applying the above identity to the matrix P and taking entry (x, y), we have

$$P^{t}(x,y) = \sum_{k=0}^{t} \mathbb{P}(|S_{t}| = k)Q_{k}(P)(x,y).$$

Since Q_k is of degree k and since $P^k(x, y) = 0$ for all k < d(x, y), we may start the sum above at k = d(x, y).

- 3. The spectrum of $Q_k(P)$ is given by $\{Q_k(\lambda), \lambda \in \operatorname{Sp}(P)\}$. Since $\operatorname{Sp}(P) \subset [-1, 1]$, and since $Q_k([-1, 1]) \subset [-1, 1]$ (recall that $Q_k(\cos \theta) = \cos(k\theta)$), the spectrum of $Q_k(P)$ is included in [-1, 1]. In particular, $Q_k(P)$ is a contraction.
- 4. By the contracting property of $Q_k(P)$ applied to the functions δ_y and δ_x , we have

$$|\langle Q_k(P)\delta_y, \delta_x \rangle_{\pi}| \le ||\delta_x||_{\pi} ||\delta_y||_{\pi} = \sqrt{\pi(x)\pi(y)}.$$

Now

$$\langle Q_k(P)\delta_y, \delta_x \rangle_{\pi} = \pi(x)Q_k(P)\delta_y(x) = \pi(x)Q_k(P)(x,y).$$

5. We obtain

$$P^{t}(x,y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}} \mathbb{P}\left(|S_{t}| \geq d(x,y)\right) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}} \exp\left\{-\frac{d(x,y)^{2}}{2t}\right\},$$

where the last inequality is by Hoeffding Inequality.

6. Let $d = \lceil D/2 \rceil$. Since D > 2(d-1), one may find two disjoint balls of radius d-1, so that there exists x such that $\pi (\mathcal{B}_{d-1}(x)) \leq 1/2$. On the other hand, we have

$$P^{t}(x, \mathcal{B}_{d-1}(x)^{c}) \leq 2 \sum_{y \notin \mathcal{B}_{d-1}(x)} \sqrt{\frac{\pi(y)}{\pi(x)}} \exp\left\{-\frac{\mathrm{d}(x, y)^{2}}{2t}\right\} \leq 2n^{3/2} e^{-\frac{d^{2}}{2t}} \leq 2n^{3/2} e^{-\frac{D^{2}}{8t}}$$

where we used that $|\mathcal{B}_{d-1}(x)^c| \leq n$ and $\frac{\pi(y)}{\pi(x)} = \frac{\deg(y)}{\deg(x)} \leq n$. For $t = \frac{D^2}{13\log n}$, we have

$$\mathcal{D}(t) \ge \frac{1}{2} - 2n^{-\frac{1}{8}}$$

8.5 Typical distance in random regular graphs

The idea is to sequentially generate the levels around two given vertices x and y and to show that for all $\varepsilon > 0$, with high probability, if the level is less than $\frac{(1-\varepsilon)}{2} \log_{d-1} n$, then the two balls are disjoint, and if it is greater than $\frac{(1+\varepsilon)}{2} \log_{d-1} n$, then they have intersected.

Lower bound.

Let $\varepsilon > 0$ and $k = \frac{(1-\varepsilon)}{2} \log_{d-1} n$. Then (roughly...) the probability $\mathbb{P}(\mathcal{B}_k(x) \cap \mathcal{B}_k(y) \neq \emptyset)$ can be upper-bounded by the probability than a Binomial random variable with parameters $(d-1)^k$ and $\frac{(d-1)^k}{n}$ is non-zero. By Markov Inequality, this is less than $\frac{(d-1)^{2k}}{n} = n^{-\varepsilon} \to 0$.

Upper bound.

Let as before $k = \frac{(1-\varepsilon)}{2} \log_{d-1} n$, and $K = \frac{(1+\varepsilon)}{2} \log_{d-1} n$. By the same reasoning, we can actually show that, with high probability, $\mathcal{B}_k(x)$ and $\mathcal{B}_k(y)$ are not only disjoint but are both *d*-regular trees of depth *k*. On the other hand, when growing from $\mathcal{B}_k(x)$ to $\mathcal{B}_K(x)$, the total number of "bad events" (matching a half-edge with a previously revealed half-edge) is stochastically dominated by a Binomial r.v. with parameters $(d-1)^K$ and $\frac{(d-1)^K}{n}$. By Markov Inequality, with high probability it is less than $n^{2\varepsilon}$. Hence, with high probability, there are at least $(d-1)^K - n^{2\varepsilon}(d-1)^{K-k} = n^{\frac{1+\varepsilon}{2}} - n^{3\varepsilon}$ half-edges at distance *K* from *x*, and the same holds for *y*. Hence, assuming $\mathcal{B}_K(x)$ and $\mathcal{B}_K(y)$ have not yet intersected, the probability that they do at the next level tends to one since the probability that a Binomial r.v. with parameters $n^{\frac{1+\varepsilon}{2}}$ and $\frac{n^{\frac{1+\varepsilon}{2}}}{n}$ is equal to 0 tends to 0.

8.6 Concentration with exchangeable pairs (proof of Lemma 5.2)

Part A.

See Nathan Ross, *Fundamentals of Stein's method*, Theorem 7.4: http://emis.ams.org/ journals/PS/images/getdoc402e.pdf?id=729.

Part B.

See Ben-Hamou and Salez [2][Lemma 6.1]: https://projecteuclid.org/download/pdfview_1/euclid.aop/1494835230.