Electron. Commun. Probab. **25** (2020), article no. 77, 1–8.
https://doi.org/10.1214/20-ECP358
ISSN: 1083-589X

**ELECTRONIC COMMUNICATIONS in PROBABILITY**

# An entropic proof of cutoff on Ramanujan graphs

Narutaka Ozawa*

**Abstract**

It is recently proved by Lubetzky and Peres that the simple random walk on a Ramanujan graph exhibits a cutoff phenomenon, that is to say, the total variation distance of the random walk distribution from the uniform distribution drops abruptly from near $1$ to near $0$. There are already a few alternative proofs of this fact. In this note, we give yet another proof based on functional analysis and entropic consideration.

**Keywords:** random walk; cutoff; entropy; Ramanujan graph.
**MSC2020 subject classifications:** 05C81; 60J10; 94A17.
Submitted to ECP on September 7, 2020, final version accepted on November 2, 2020.

## 1 Introduction

Throughout this note, we fix $d \geq 3$ and consider finite simple connected $d$-regular non-bipartite graphs $G$. We identify $G$ with its vertex set and denote by $P \in \mathbb{B}(\ell_2 G)$ the normalized adjacency matrix on $G$, which is given by $P(x,y) = \frac{1}{d}$ if $x$ and $y$ are adjacent and else $0$. The hermitian matrix $P$ leaves the constant vectors invariant and the corresponding eigenvalue $1$ has multiplicity $1$. Thus the eigenvalues of $P$ are $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|G|} > -1$. We are interested in the *reduced spectral radius*

$$\rho := \|P|_{\ell_2 G \cap (\mathbb{C}1)^\perp}\|_{\mathbb{B}(\ell_2 G \cap (\mathbb{C}1)^\perp)} = \max\{\lambda_2, -\lambda_{|G|}\} < 1. \tag{1}$$

We consider the simple random walk $(X^t)_{t\in\mathbb{N}}$ on $G$ starting at some $X^0 \in G$. The probability distribution $\mu^t := P^t(\,\cdot\,, X^0)$ of $X^t$ converges to the uniform distribution $\pi := \frac{1}{|G|}1_G$. For $\alpha \in (0,1)$, the *total variation mixing time* is defined to be

$$T^{\mathrm{mix}}(\alpha) := \min\{t \in \mathbb{N} : \|\mu^t - \pi\|_{\mathrm{TV}} < \alpha\}, \tag{2}$$

where

$$\|\nu - \eta\|_{\mathrm{TV}} = \max_{A\subset G}\{|\nu(A) - \eta(A)|\} = \frac{1}{2}\sum_{x\in G}|\nu(x) - \eta(x)| = \frac{1}{2}\|\nu - \eta\|_1 \tag{3}$$

is the total variation norm. The total variation mixing time $T^{\mathrm{mix}}(\alpha)$ is clearly monotone in $\alpha$. Since the random walk $(X^t)_t$ backtracks with probability $1/d$, the expected distance from the origin $X^0$ to $X^t$ is at most $\frac{d-2}{d}t$. Thus $\mu^t$ is concentrated on the ball of radius $\frac{d-2}{d}t + O(\sqrt{t})$, which has cardinality at most $d(d-1)^{\frac{d-2}{d}t+O(\sqrt{t})}$. This observation leads to the *entropic lower bound* of the total variation mixing time,

$$T^{\mathrm{mix}}(\alpha) \geq \frac{\log|G|}{h_d} - o_\alpha(\log|G|), \tag{4}$$

*RIMS, Kyoto University, 606-8502, Japan. E-mail: narutaka@kurims.kyoto-u.ac.jp

where

$$h_d := \frac{(d-2)\log(d-1)}{d} \tag{5}$$

is the asymptotic entropy of the simple random walk on the $d$-regular tree (see [LP, BL] for a more precise estimate). On the other hand, since $\|\mu^t - \pi\|_1 \le |G|^{1/2}\|\mu^t - \pi\|_2 \le |G|^{1/2}\rho^t$, one has the *spectral upper bound*

$$T^{\mathrm{mix}}(\alpha) \le \frac{\log|G| + O_\alpha(1)}{-2\log\rho} \le \frac{\log|G| + O_\alpha(1)}{2(1-\rho)}. \tag{6}$$

This upper bound also allows an entropic interpretation, see Lemma 2 below.

Now, let $\mathcal{G} = \{G\}$ be a family of graphs. Recall that $\mathcal{G}$ is said to be a *family of expanders* (see e.g., [Lu]) if

$$\sup\{\rho_G : G \in \mathcal{G}\} < 1. \tag{7}$$

When needed, we add the subscript $G$ to the symbol $\rho$ etc to specify the graph $G$ under consideration. Note that the *Alon–Boppana bound* gives the lower estimate

$$\liminf_{G\in\mathcal{G},\,|G|\to\infty} \rho_G \ge \frac{2\sqrt{d-1}}{d} =: \rho_d \tag{8}$$

for any family $\mathcal{G}$. Here the value $\rho_d$ is the spectral radius of the simple random walk on the $d$-regular tree. The family $\mathcal{G}$ is said to be *asymptotically Ramanujan* if it marks the Alon–Boppana bound,

$$\lim_{G\in\mathcal{G},\,|G|\to\infty} \rho_G = \rho_d. \tag{9}$$

An explicit family of Ramanujan graphs was first discovered by Lubotzky–Phillips–Sarnak ([LPS]) and it was shown by Friedman ([Fr]) that the random graphs are asymptotically Ramanujan. We have seen in the above that every expander family satisfies

$$T_G^{\mathrm{mix}}(\alpha) \asymp_\alpha \log|G| \tag{10}$$

as $|G| \to \infty$. The family $\mathcal{G}$ is said to *exhibit cutoff* if the implied constants in the above asymptotic equality are independent of $\alpha$, or more precisely, if

$$\lim_{G\in\mathcal{G},\,|G|\to\infty} \frac{T_G^{\mathrm{mix}}(\alpha)}{T_G^{\mathrm{mix}}(\alpha')} = 1 \tag{11}$$

for every $\alpha, \alpha' \in (0,1)$ and every choice of the starting points $X_G^0 \in G$. It is proved by Lubetzky and Peres ([LP]) that asymptotically Ramanujan graphs exhibit cutoff.

**Theorem 1.** A family $\mathcal{G}$ of asymptotically Ramanujan graphs exhibits cutoff. In fact, the total variation mixing time marks the entropic lower bound: For any $\alpha \in (0,1)$,

$$\lim_{G\in\mathcal{G},\,|G|\to\infty} \frac{T_G^{\mathrm{mix}}(\alpha)}{\log|G|} = \frac{d}{(d-2)\log(d-1)} = \frac{1}{h_d}. \tag{12}$$

See [He, BL] for alternative proofs of this theorem. In this note, we give yet another proof of it based on functional analysis and entropic consideration. Entropic considerations have been also instrumental in some of the recent works on cutoff phenomena, see [BHLP, BHS, BCS, CK], to just name a few.

## 2 Entropy

Let $\mathrm{Prob}(G)$ denote the set of probability measures on a (finite or infinite) graph $G$. Recall that the *Shannon entropy* (or just *entropy*) of $\nu \in \mathrm{Prob}(G)$ is the quantity

$$H(\nu) = -\sum_x \nu(x) \log \nu(x). \tag{13}$$

When a random walk $(X^t)_{t\in\mathbb{N}}$ on $G$ is under consideration, we also write

$$H(t) := H(\mu^t) = \mathbb{E}\big[-\log \mu^t(X^t)\big]. \tag{14}$$

Let's assume for the moment that the graph $G$ is transitive. Then the quantity $H(t)$ is increasing and concave (Proposition 1.3 in [KV]). The limit

$$h := \lim_t \frac{1}{t} H(t) = \lim_t H(t) - H(t-1) \tag{15}$$

is called the *asymptotic entropy* of the random walk $(X^t)_t$. Moreover, the Shannon–McMillan–Breiman type Theorem holds that $\frac{1}{t}|-\log \mu^t(X^t) - H(t)|$ converges to zero almost surely (Section IV in [De], Theorem 2.1 in [KV]). This result is vacuous when $G$ is finite.

In this note, we are interested in the case of a finite simple connected $d$-regular non-bipartite graph $G$ and the quantitative growth property of $H(t)$. Since $G$ is connected and non-bipartite, one has $\mu^t \to \pi$ and $H(t) \nearrow H(\pi) = \log|G|$.

We review a concavity property of the entropy functional. For $\nu \in \mathrm{Prob}(G)$, we put $P\nu := \sum_x P(\,\cdot\,, x)\nu(x) \in \mathrm{Prob}(G)$. Note that $\mu^{t+1} = P\mu^t$. A caveat is in order here: the notation $\nu P$ instead of $P\nu$ may be more common among probability theory, but we stick to $P\nu$, because we view $\nu$ as a vector in $\ell_1$ and deal with $\nu^{1/2} \in \ell_2$. This notation should not cause any confusion, particularly because our $P$ is always symmetric.

**Lemma 2.** For any $\nu \in \mathrm{Prob}(G)$, one has

$$H(P\nu) - H(\nu) \geq \frac{1-\rho}{16}\|\nu - \pi\|_1^2. \tag{16}$$

*Proof.* By Hall's Matching Theorem, there are permutation matrices $P_i$ such that $P = \sum_{i=1}^d \lambda_i P_i$, where $\lambda_i = 1/d$ (just to see that the proof works in more general setting). Thus $\bar\nu := P\nu$ is the average of $\nu_i := P_i\nu$. Note that $H(\nu_i) = H(\nu)$ since $\nu_i$ has the same distribution up to a permutation as $\nu$. Put $f(t) = t\log t$ and observe that $f''(t) = 1/t \geq 1/(a+b)$ for $t \in [a,b]$. For any $a_i \geq 0$ and $\bar a = \sum_i \lambda_i a_i$, Taylor's theorem yields

$$\sum_i \lambda_i (f(a_i) - f(\bar a)) \geq \frac{1}{2}\sum_i \lambda_i \frac{(a_i - \bar a)^2}{a_i + \bar a} \geq \frac{1}{2}\sum_i \lambda_i (a_i^{1/2} - \bar a^{1/2})^2. \tag{17}$$

It follows that

$$H(\bar\nu) - H(\nu) \geq \frac{1}{2}\sum_i \lambda_i \|\nu_i^{1/2} - \bar\nu^{1/2}\|_2^2. \tag{18}$$

Since $\nu_i^{1/2} = (P_i\nu)^{1/2} = P_i\nu^{1/2}$ and $\bar\nu^{1/2}$ are unit vectors in $\ell_2 G$,

$$\frac{1}{2}\sum_i \lambda_i \|\nu_i^{1/2} - \bar\nu^{1/2}\|_2^2 = 1 - \langle P\nu^{1/2}, \bar\nu^{1/2}\rangle \geq 1 - \|P\nu^{1/2}\|_2. \tag{19}$$

By Pythagorean theorem, for $\xi_0 := \nu^{1/2} - \langle \nu^{1/2}, \pi^{1/2}\rangle \pi^{1/2} \in (\mathbb{C}\pi^{1/2})^\perp \cap \ell_2 G$,

$$1 - \|P\nu^{1/2}\|_2 \geq \frac{1}{2}(\|\nu^{1/2}\|_2^2 - \|P\nu^{1/2}\|_2^2) \geq \frac{1}{2}(1-\rho^2)\|\xi_0\|_2^2 \tag{20}$$

and

$$\|\xi_0\|_2^2 = 1 - \langle \nu^{1/2}, \pi^{1/2} \rangle^2 \geq 1 - \langle \nu^{1/2}, \pi^{1/2} \rangle = \frac{1}{2} \|\nu^{1/2} - \pi^{1/2}\|_2^2. \tag{21}$$

Finally, one has

$$\|\nu^{1/2} - \pi^{1/2}\|_2 \geq \frac{1}{2} \|\nu^{1/2} + \pi^{1/2}\|_2 \|\nu^{1/2} - \pi^{1/2}\|_2 \geq \frac{1}{2} \|\nu - \pi\|_1. \tag{22}$$

By putting the above inequalities (18)-(22) together, we obtain the claim. □

Here is an entropic characterization of cutoff phenomena for expander graphs. It is widely believed that every expander family of transitive graphs should exhibit cutoff.

**Theorem 3.** Let $\mathcal{G} = \{G\}$ be an expander family of finite simple connected $d$-regular non-bipartite graphs. Then $\mathcal{G}$ exhibits cutoff if and only if the following holds: For any $\varepsilon > 0$ and any $\delta > 0$, there is $N \in \mathbb{N}$ satisfying that

$$H(T_G^{\mathrm{mix}}(1-\varepsilon)) > (1-\delta) \log |G| \tag{23}$$

for every $G \in \mathcal{G}$ with $|G| > N$.

*Proof.* We prove the 'if' part. Since $H(t+1) - H(t) \geq (1-\rho)\varepsilon^2/16$ for all $t < T_G^{\mathrm{mix}}(\varepsilon)$ by Lemma 2, $H(T_G^{\mathrm{mix}}(1-\varepsilon)) > (1-\delta)\log|G|$ implies

$$T_G^{\mathrm{mix}}(\varepsilon) - T_G^{\mathrm{mix}}(1-\varepsilon) < \delta \frac{16}{\varepsilon^2} \frac{\log |G|}{1-\rho}. \tag{24}$$

This proves the 'if' part. For the 'only if' part, we observe that if $\|\nu - \pi\|_{\mathrm{TV}} < \delta/2$, then $A := \{x \in G : \nu(x) > \frac{2}{\delta |G|}\}$ satisfies $\nu(A) < \pi(A) + \delta/2 < \delta$, which implies

$$H(\nu) \geq - \sum_{x \in G \setminus A} \nu(x) \log \nu(x) \geq (1-\delta)(\log |G| + \log(\delta/2)). \tag{25}$$

This means that $H(T_G^{\mathrm{mix}}(\delta/2)) \geq (1-\delta)\log|G| - O_\delta(1)$. In case the graphs $G \in \mathcal{G}$ are transitive (but not necessarily expanders), one has by concavity that $H(t_2) - H(t_1) \leq H(t_1)\frac{t_2 - t_1}{t_1}$ for every $t_1 < t_2$ and hence cutoff requires $H(T_G^{\mathrm{mix}}(1-\varepsilon)) > (1-\delta)\log|G|$ for $G$ large enough. In general, since $H(t+1) - H(t) \in [0, \log d]$ always holds,

$$1 - \delta - \frac{H(T_G^{\mathrm{mix}}(1-\varepsilon))}{\log |G|} - \frac{O_\delta(1)}{\log |G|} \leq \frac{T_G^{\mathrm{mix}}(\delta/2) - T_G^{\mathrm{mix}}(1-\varepsilon)}{\log |G|} \log d \to 0, \tag{26}$$

by (10), provided that $\mathcal{G}$ exhibits cutoff. This proves the 'only if' part. □

*Remark* 4. The referee has pointed out that a similar result to Theorem 3 has been used implicitly in the literature, e.g., in [BLPS]. Here is a sketch of a more direct proof along this line, which is communicated by the referee. Let $\delta > 0$ be given and $\nu \in \mathrm{Prob}(G)$ be such that $H(\nu) > (1-\delta)\log|G|$. Then for $A := \{x \in G : -\log \nu(x) \geq (1-\delta^{1/2})\log|G|\}$ and $m := \nu(A)$, one has

$$H(\nu) < (1-m)(1-\delta^{1/2}) \log |G| + m \log \frac{|G|}{m}, \tag{27}$$

which implies $1 - m < \delta^{1/2} + O_\delta((\log|G|)^{-1})$. Thus one has $\|\nu - \pi - (\nu|_A - m\pi)\|_1 \leq 2(1-m)$, $\|(\nu|_A - m\pi)\|_2 \leq 2|G|^{-(1-\delta^{1/2})/2}$, and

$$\|P^k \nu - \pi\|_1 \leq 2\rho_G^k |G|^{\delta^{1/2}/2} + 2\delta^{1/2} + O_\delta((\log|G|)^{-1}) \leq 4\delta^{1/2} \tag{28}$$

at $k = \frac{\delta^{1/2}}{-2\log \rho_G} \log|G| + O_\delta(1)$ when $\log|G|$ is large enough.

It is well-known to experts that a cutoff phenomenon is generally related to a measure concentration phenomenon for $-\log \mu^t(X^t)$.

**Corollary 5.** Let $\mathcal{G} = \{G\}$ be an expander family of finite simple connected $d$-regular non-bipartite graphs. Assume that $\mathcal{G}$ has the following quantitative Shannon–McMillan–Breiman type property: For any $\delta, \kappa > 0$ there is $T_0 \in \mathbb{N}$ such that every $G \in \mathcal{G}$ and $t \geq T_0$ satisfy

$$\mu_G^t(\mathcal{N}_{\delta \log |G|}(\{x \in G : -\log \mu_G^t(x) < H(t) + \delta t\})) > 1 - \kappa, \tag{29}$$

where $\mathcal{N}_r(A)$ denotes the $r$-neighborhood of $A \subset G$. Then $\mathcal{G}$ exhibits cutoff.

*Proof.* Put $t := T^{\mathrm{mix}}(1-\varepsilon)$, $\kappa := \varepsilon/2$, and $A := \{x : -\log \mu^t(x) < H(t) + \delta t\}$. Then $|A| \leq e^{H(t)+\delta t}$ and so $B := \mathcal{N}_{\delta \log |G|}(A)$ has cardinality $|B| \leq d^{\delta \log |G|} e^{H(t)+\delta t}$. For $t \geq T_0$, one has

$$1 - \varepsilon > \|\mu^t - \pi\|_{\mathrm{TV}} \geq \mu^t(B) - \pi(B) > 1 - \varepsilon/2 - |G|^{-1+\delta \log d} e^{H(t)+\delta t}. \tag{30}$$

This implies $e^{H(t)+\delta t} > |G|^{1-\delta \log d}\varepsilon/2$. Hence for $\beta := t/\log |G|$ one has

$$H(T_{\mathrm{mix}}(1-\varepsilon)) > (1 - (\beta + \log d)\delta) \log |G| - O_\varepsilon(1). \tag{31}$$

Since $\beta$ is bounded by (6), this yields the cutoff phenomenon by Theorem 3. □

## 3 Proof of Theorem 1

In this section, we prove Theorem 1. In light of Theorem 3 and the entropic lower bound (4), it suffices to show $H(t) - H(t-1) \geq (1-\delta)h_d$ holds for $t \leq T^{\mathrm{mix}}(1-\varepsilon)$. Upon change of $\alpha \in (0,1)$, we may replace $T^{\mathrm{mix}}(\alpha)$ by the *Hellinger distance mixing time*

$$T^{\mathrm{mix},2}(\alpha) := \min\{t \in \mathbb{N} : \|(\mu^t)^{1/2} - \pi^{1/2}\|_2^2 < 2\alpha\}. \tag{32}$$

We state this well-known fact as a lemma for the readers' convenience.

**Lemma 6.** For every $\varepsilon > 0$, there is $\delta > 0$ such that every $\nu \in \mathrm{Prob}(G)$ satisfies the following.

- If $\|\nu - \pi\|_1 < \delta$, then $\|\nu^{1/2} - \pi^{1/2}\|_2^2 < \varepsilon$.
- If $\|\nu^{1/2} - \pi^{1/2}\|_2^2 < \delta$, then $\|\nu - \pi\|_1 < \varepsilon$.
- If $\|\nu - \pi\|_1 < 2 - \varepsilon$, then $\|\nu^{1/2} - \pi^{1/2}\|_2^2 < 2 - \delta$.
- If $\|\nu^{1/2} - \pi^{1/2}\|_2^2 < 2 - \varepsilon$, then $\|\nu - \pi\|_1 < 2 - \delta$.

*Proof.* The first three are easy consequences of

$$\|\nu^{1/2} - \pi^{1/2}\|_2^2 \leq \|\nu - \pi\|_1 \leq \|\nu^{1/2} - \pi^{1/2}\|_2 \|\nu^{1/2} + \pi^{1/2}\|_2. \tag{33}$$

For the last, put $A := \{x : \nu(x) > \frac{\varepsilon^2}{16|G|}\}$. Then $\|\nu^{1/2} - \pi^{1/2}\|_2^2 < 2 - \varepsilon$ implies

$$\varepsilon/2 < \langle \nu^{1/2}, \pi^{1/2} \rangle \leq (|A|/|G|)^{1/2} + \varepsilon/4 \tag{34}$$

and hence $|A| \geq (\varepsilon^2/16)|G|$. It follows that $\|\nu - \pi\|_1 < 2 - \varepsilon^4/128$. □

Here is another auxiliary lemma about concavity of the square root.

**Lemma 7.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a standard probability space and $\mathcal{B} \subset \mathcal{A}$ be a $\sigma$-subalgebra. For any bounded random variable $f \geq 0$, the conditional expectation $g = \mathbb{E}[f \,|\, \mathcal{B}]$ satisfies

$$\|g^{1/2} - f^{1/2}\|_{L^2(\Omega,\mathbb{P})}^2 \leq 2\|f\|_{L^\infty(\Omega,\mathbb{P})}^{1/2} (\mathbb{E}[g^{1/2}] - \mathbb{E}[f^{1/2}]). \tag{35}$$

*Proof.* First, assume that $\mathcal{B}$ is trivial and hence $g = \mathbb{E}\big[f\big]$ is a constant. Then,

$$\|g^{1/2} - f^{1/2}\|_2^2 = 2g^{1/2} \int g^{1/2} - f^{1/2}\, d\mathbb{P} \leq 2\|f\|_\infty^{1/2} \int g^{1/2} - f^{1/2}\, d\mathbb{P} \tag{36}$$

and the assertion holds true. The case where $\mathcal{B}$ is purely atomic follows by integrating this estimate over atoms. The general case then follows by approximation (or the disintegration formula). □

*Proof of Theorem 1.* Recall from the beginning of the section that we want to estimate

$$H(t) - H(t-1) = \mathbb{E}\big[-\log f_t\big], \tag{37}$$

where

$$f_t := \frac{\mu^t(X^t)}{\mu^{t-1}(X^{t-1})}. \tag{38}$$

The random variable $f_t$ satisfies $f_t \geq \frac{1}{d}$ and

$$\mathbb{E}\big[f_t\big] = \sum_{x,y} \frac{\mu^t(y)}{\mu^{t-1}(x)} \cdot P(y,x)\mu^{t-1}(x) = \sum_{x,y\in G,\, x\in \operatorname{supp}\mu^{t-1}} \mu^t(y)P(y,x) \leq 1. \tag{39}$$

Note that since $\sum_{x\in\operatorname{supp}\mu^{t-1}} P(y,x) = 1$ for $y \in \operatorname{supp}\mu^{t-2}$, one has

$$1 - \mathbb{E}\big[f_t\big] \leq \sum_{y\notin\operatorname{supp}\mu^{t-2}} \mu^t(y) \leq (\frac{d-1}{d})^{t-1}. \tag{40}$$

To relate the random variable $f_t$ to the functional analysis on $\ell_2 G$, we consider the unit vectors $\xi_t := (\mu^t)^{1/2}$ in $\ell_2 G$ and observe that

$$\langle P\xi_{t-1}, \xi_t\rangle = \sum_{x,y\in G} P(y,x)\mu^{t-1}(x)\Big(\frac{\mu_t(y)}{\mu_{t-1}(x)}\Big)^{1/2} = \mathbb{E}\big[f_t^{1/2}\big]. \tag{41}$$

Hence, for any $t \leq T^{\mathrm{mix},2}(1-\varepsilon)$, one has

$$\mathbb{E}\big[f_t^{1/2}\big] \leq \|P\xi_{t-1}\| \leq \langle \xi_{t-1}, \pi^{1/2}\rangle + \rho < \rho + \varepsilon. \tag{42}$$

We will compare the value $\mathbb{E}\big[f_t^{1/2}\big]$ with that for the covering tree. Let's consider the covering of $G$ by the $d$-regular tree $T_d$ and lift the random walk $(X^t)_t$ to the simple random walk $(\tilde{X}^t)_t$ on $T_d$. We write $\tilde{f}_t := \tilde{\mu}^t(\tilde{X}^t)/\tilde{\mu}^{t-1}(\tilde{X}^{t-1})$ and observe that

$$\mathbb{E}\big[\tilde{f}_t \,|\, X^{t-1}, X^t\big] \approx f_t. \tag{43}$$

Indeed, suppose $X^{t-1} = x$ and $X^t = y$. Then $x$ and $y$ are adjacent and $\mathbb{P}(X^{t-1} = x, X^t = y) = \mu^{t-1}(x)/d$. There is a bijection $q_{x,y}$ from the lifts $[x]$ of $x$ onto the lifts $[y]$ of $y$ such that $p$ and $q_{x,y}(p)$ are adjacent for all $p \in [x]$. Moreover, $\mathbb{P}(\tilde{X}^{t-1} = p, \tilde{X}^t = q_{x,y}(p)) = \tilde{\mu}^{t-1}(p)/d$. Thus at the event $X^{t-1} = x$ and $X^t = y$, one has

$$\mathbb{E}\big[\tilde{f}_t \,|\, X^{t-1}, X^t\big] = \frac{d}{\mu^{t-1}(x)} \sum_{p\in[x]\cap\operatorname{supp}\tilde{\mu}^{t-1}} \frac{\tilde{\mu}^t(q_{x,y}(p))}{\tilde{\mu}^{t-1}(p)} \frac{\tilde{\mu}^{t-1}(p)}{d} \leq \frac{\mu^t(y)}{\mu^{t-1}(x)} = f_t. \tag{44}$$

Note that $\|f_t - \mathbb{E}\big[\tilde{f}_t \,|\, X^{t-1}, X^t\big]\|_1 \leq (\frac{d-1}{d})^{t-1}$ is virtually negligible. By the concavity of the square root, one has

$$\mathbb{E}\big[f_t^{1/2}\big] \geq \mathbb{E}\big[(\mathbb{E}\big[\tilde{f}_t \,|\, X^{t-1}, X^t\big])^{1/2}\big] \geq \mathbb{E}\big[\tilde{f}_t^{1/2}\big]. \tag{45}$$

It is not hard to show that the value of $\tilde{f}_t$ is asymptotically $d-1$ with probability $\frac{1}{d}$ (when the step $\tilde{X}^t$ is in the direction to the origin) and $\frac{1}{d-1}$ with probability $\frac{d-1}{d}$. Since $\tilde{f}_t \geq \frac{1}{d}$ and $\mathbb{E}\big[\tilde{f}_t\big] \leq 1$ are unconditionally true, one has

$$\lim_t \mathbb{E}\big[-\log \tilde{f}_t\big] = -\frac{1}{d}\log(d-1) + \frac{d-1}{d}\log(d-1) = \frac{(d-2)\log(d-1)}{d} = h_d \tag{46}$$

and

$$\lim_t \mathbb{E}\big[\tilde{f}_t^{1/2}\big] = \frac{1}{d}\sqrt{d-1} + \frac{d-1}{d}\sqrt{\frac{1}{d-1}} = \frac{2\sqrt{d-1}}{d} = \rho_d. \tag{47}$$

The first equality merely confirms that the asymptotic entropy of the tree random walk $(\tilde{X}^t)_t$ is $h_d$. The second equality is the lucky(?) coincidence that we are going to exploit.

Let $T_\varepsilon \in \mathbb{N}$ be such that $\mathbb{E}\big[\tilde{f}_t^{1/2}\big] \geq \rho_d - \varepsilon$ for all $t \geq T_\varepsilon$. By summarizing the above discussion, we obtain the following. For any $0 < \varepsilon < 1$, if an asymptotically Ramanujan graph $G$ is large enough, then one has

$$\rho_d + \varepsilon \geq \mathbb{E}\big[f_t^{1/2}\big] \geq \mathbb{E}\big[\tilde{f}_t^{1/2}\big] \geq \rho_d - \varepsilon \tag{48}$$

for all $t \in [T_\varepsilon, T_G^{\mathrm{mix},2}(1-\varepsilon)]$. It is intuitively clear (and will be explained later) that this implies

$$\mathbb{E}\big[-\log f_t\big] \geq (1-\delta(\varepsilon))h_d \tag{49}$$

for all $t \in [T_\varepsilon, T_G^{\mathrm{mix},2}(1-\varepsilon)]$, where $\delta \geq 0$ is a continuous function with $\delta(0) = 0$.

To conclude the proof of Theorem 1 via Theorem 3, let $\varepsilon > 0$ and $\delta > 0$ be given. Take $\varepsilon_0 \in (0, \varepsilon)$ small enough so that $\delta(\varepsilon_0) < \delta$. Then by (37), (49), and the entropic lower bound (4), one obtains

$$H(T_G^{\mathrm{mix},2}(1-\varepsilon)) \geq H(T_G^{\mathrm{mix},2}(1-\varepsilon_0)) \geq (1-\delta)\log|G| - o_{\varepsilon_0}(\log|G|). \tag{50}$$

This finishes the proof of Theorem 1.

For completeness, we explain why (48) implies (49). For this, we may assume that $\|f_t - \mathbb{E}\big[\tilde{f}_t \,|\, X^{t-1}, X^t\big]\|_1 < \varepsilon/2$ and $\|\tilde{f}_t - f\|_1 < \varepsilon/2$ for some random variable $f$ that takes value $d-1$ with probability $\frac{1}{d}$ and $\frac{1}{d-1}$ with probability $\frac{d-1}{d}$. Then $g := \mathbb{E}\big[f \,|\, X^{t-1}, X^t\big]$ satisfies $\|g - f_t\|_1 < \varepsilon$ and

$$\|g^{1/2} - f_t^{1/2}\|_1 \leq \||g - f_t|^{1/2}\|_1 \leq \|g - f_t\|_1^{1/2} < \varepsilon^{1/2}. \tag{51}$$

By (48), this implies that $\mathbb{E}\big[g^{1/2}\big] - \mathbb{E}\big[f^{1/2}\big] \leq 2\varepsilon + \varepsilon^{1/2} < 3\varepsilon^{1/2}$ and, by Lemma 7, that

$$\|g^{1/2} - f^{1/2}\|_2 \leq (2\|f\|_\infty^{1/2}(\mathbb{E}\big[g^{1/2}\big] - \mathbb{E}\big[f^{1/2}\big]))^{1/2} \leq 3(d-1)^{1/4}\varepsilon^{1/4}. \tag{52}$$

Thus, $\|g - f\|_1 \leq \|g^{1/2} - f^{1/2}\|_2 \|g^{1/2} + f^{1/2}\|_2 = O_d(\varepsilon^{1/4})$ and $\|f_t - f\|_1 = O_d(\varepsilon^{1/4})$. Since $f \geq \frac{1}{d-1}$, this implies $\mathbb{E}\big[-\log f_t\big] \geq (1-\delta(\varepsilon))\,\mathbb{E}\big[-\log f\big]$ for some continuous function $\delta \geq 0$ with $\delta(0) = 0$. □

*Remark* 8. In the above proof, we saw $\mathbb{E}\big[f_t^{1/2}\big] \approx \mathbb{E}\big[\tilde{f}_t^{1/2}\big]$. This implies that $(f_t)_t$ are asymptotically independent family and, by the central limit theorem, that Ramanujan graphs verify the quantitative Shannon–McMillan–Breiman type theorem (the statement without $\mathcal{N}_{\delta\log|G|}$ in Corollary 5). In case of random walks on transitive graphs, one has $\mathbb{E}\big[f_t^{1/2}\big] = \mathbb{E}\big[g_t^{1/2}\big]$, where $g_t := \frac{P^t(X^t, X^0)}{P^{t-1}(X^t, X^1)}$, and $g_{t+1} \approx \mathbb{E}\big[g_t \,|\, X^1, X^{t+1}\big]$. In view of Kaimanovich and Vershik's proof of the Shannon–McMillan–Breiman type theorem (Theorem 2.1 in [KV]), it would be interesting to know whether or not $g_t \approx \mathbb{E}\big[g_t \,|\, X_1, \ldots, X_m\big]$ for $m = o(\log|G|)$.

In general, for a finitely generated (infinite) group $\Gamma$ and a finitely-supported non-degenerate symmetric probability measure $\mu$ on $\Gamma$, the value $\langle \lambda(\mu)(\mu^{t-1})^{1/2}, (\mu^t)^{1/2}\rangle$ converges to $\langle \sigma(\mu)1_\Pi, 1_\Pi\rangle_{L^2(\Pi,\nu)}$, where $(\Pi, \nu)$ is the Poisson boundary and $\sigma$ is the unitary Koopman representation. This value can be strictly smaller than the spectral radius $\rho$. For example, if $G$ is amenable, then $\rho = 1$; while $\langle \lambda(\mu)(\mu^{t-1})^{1/2}, (\mu^t)^{1/2}\rangle \to 1$ if and only if the Poisson boundary is trivial. See [KV].

## References


[BHLP] A. Ben-Hamou, E. Lubetzky, and Y. Peres; Comparing mixing times on sparse random graphs. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** (2019), 1116–1130. MR-3949966

[BHS] A. Ben-Hamou and J. Salez; Cutoff for nonbacktracking random walks on sparse random graphs. *Ann. Probab.* **45** (2017), 1752–1770. MR-3650414

[BLPS] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly; Random walks on the random graph. *Ann. Probab.* **46** (2018), 456–490. MR-3758735

[BCS] C. Bordenave, P. Caputo, and J. Salez; Cutoff at the "entropic time" for sparse Markov chains. *Probab. Theory Related Fields* **173** (2019), 261–292. MR-3916108

[BL] C. Bordenave and H. Lacoin; Cutoff at the entropic time for random walks on covered expander graphs. *Preprint.* arXiv:1812.06769.

[CK] Guillaume Conchon–Kerjan; Cutoff for random lifts of weighted graphs. *Preprint.* arXiv:1908.02898.

[De] Y. Derriennic; Quelques applications du théorème ergodique sous-additif. *Conference on Random Walks (Kleebach, 1979). Astérisque* **74** (1980), 183–201. MR-0588163

[Fr] J. Friedman; A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.* **195** (2008), no. 910, viii+100 pp. MR-2437174

[He] J. Hermon; Cutoff for Ramanujan graphs via degree inflation. *Electron. Commun. Probab.* **22** (2017), Paper No. 45, 10 pp. MR-3693771

[KV] V. A. Kaimanovich and A. M. Vershik; Random walks on discrete groups: boundary and entropy. *Ann. Probab.* **11** (1983), 457–490. MR-0704539

[LP] E. Lubetzky and Y. Peres; Cutoff on all Ramanujan graphs. *Geom. Funct. Anal.* **26** (2016), 1190–1216. MR-3558308

[Lu] A. Lubotzky; *Discrete groups, expanding graphs and invariant measures.* Progress in Mathematics, 125. Birkhäuser Verlag, Basel, 1994. xii+195 pp. MR-1308046

[LPS] A. Lubotzky, R. Phillips, and P. Sarnak; Ramanujan graphs. *Combinatorica* **8** (1988), 261–277. MR-0963118


**Acknowledgments.** The author would like to thank the anonymous referee for valuable comments and Remark 4. A part of this work was done during the research camp in Tsurui village in August 2020. The author acknowledges the kind hospitality of the people of Tsurui village. The camp was supported by Operator Algebra Supporters' Fund and JSPS KAKENHI 17K05277.