

CUTOFF ON ALL RAMANUJAN GRAPHS

EYAL LUBETZKY AND YUVAL PERES

ABSTRACT. We show that on every Ramanujan graph G , the simple random walk exhibits cutoff: when G has n vertices and degree d , the total-variation distance of the walk from the uniform distribution at time $t = \frac{d}{d-2} \log_{d-1} n + s\sqrt{\log n}$ is asymptotically $\mathbb{P}(Z > cs)$ where Z is a standard normal variable and $c = c(d)$ is an explicit constant. Furthermore, for all $1 \leq p \leq \infty$, d -regular Ramanujan graphs minimize the asymptotic L^p -mixing time for SRW among all d -regular graphs. Our proof also shows that, for every vertex x in G as above, its distance from $n - o(n)$ of the vertices is asymptotically $\log_{d-1} n$.

1. INTRODUCTION

A family of d -regular graphs G_n with $d \geq 3$ fixed is called an *expander*, following the works of Alon and Milman [4, 6] from the 1980's, if all nontrivial eigenvalues of the adjacency matrices are uniformly bounded away from d . Lubotzky, Phillips, and Sarnak [25] defined a connected d -regular graph G with $d \geq 3$ to be *Ramanujan* iff every eigenvalue λ of its adjacency matrix is either $\pm d$ or satisfies $|\lambda| \leq 2\sqrt{d-1}$. Such expanders, which in light of the Alon–Boppana Theorem [29] have an asymptotically optimal spectral gap, were first constructed, using deep number theoretic tools, in [25] and independently by Margulis [28] (see also [13, 24] and Fig. 1). Due to their remarkable expansion properties, Ramanujan graphs have found numerous applications (cf. [19] and the references therein). However, after 25 years of study, the geometry of these objects is still mysterious, and in particular, determining the profile of distances between vertices in such a graph and the precise mixing time of simple random walk (SRW) remained open.

Formally, letting $\|\mu - \nu\|_{\text{TV}} = \sup_A [\mu(A) - \nu(A)]$ denote total-variation distance, the (L^1) mixing time of a finite Markov chain with transition kernel P and stationarity distribution π is defined as

$$t_{\text{MIX}}(\varepsilon) = \min\{t : D_{\text{TV}}(t) \leq \varepsilon\} \quad \text{where} \quad D_{\text{TV}}(t) = \max_x \|P^t(x, \cdot) - \pi\|_{\text{TV}}.$$

A sequence of finite ergodic Markov chains is said to exhibit *cutoff* if its total-variation distance from stationarity drops abruptly, over a period of time referred to as the *cutoff window*, from near 1 to near 0; that is, there is cutoff iff $t_{\text{MIX}}(\varepsilon) = (1 + o(1))t_{\text{MIX}}(\varepsilon')$ for any fixed $0 < \varepsilon, \varepsilon' < 1$.

Our main result shows that the Ramanujan assumption implies cutoff with $t_{\text{MIX}}(\varepsilon) = (\frac{d}{d-2} + o(1)) \log_{d-1} n$ and window $O(\sqrt{\log n})$.

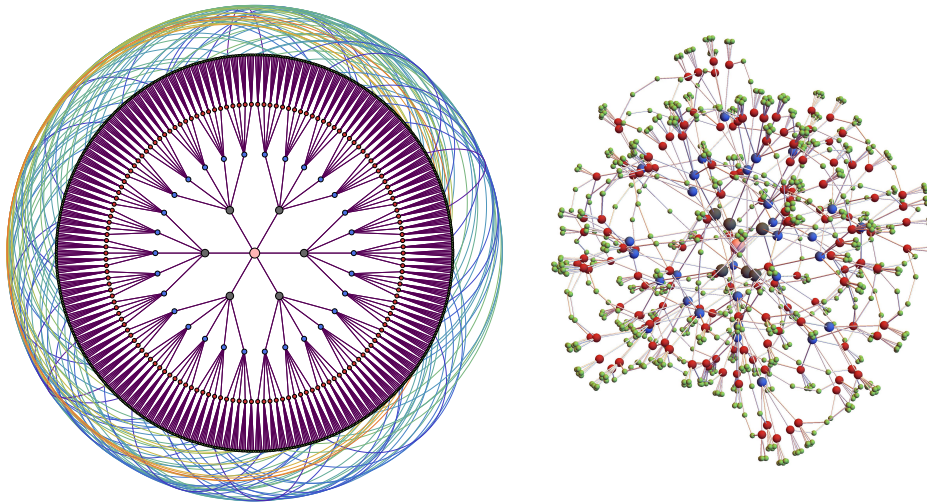


FIGURE 1. A ball of radius 4 in the Lubotzky–Phillips–Sarnak 6-regular Ramanujan graph on $n = 12180$ vertices via $\text{PSL}(2, \mathbb{F}_{29})$.

Theorem 1. *On any sequence of d -regular non-bipartite Ramanujan graphs, SRW exhibits cutoff. More precisely, let G be such a graph on n vertices and*

$$t_*(n) := \frac{d}{d-2} \log_{d-1} n.$$

Then for every fixed $s \in \mathbb{R}$ and every initial vertex x , the SRW satisfies

$$D_{\text{TV}}(t_*(n) + s\sqrt{\log_{d-1} n}) \rightarrow \mathbb{P}(Z > c_d s) \quad \text{as } n \rightarrow \infty, \quad (1.1)$$

where Z is a standard normal random variable and $c_d = \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}}$.

Consequently, we obtain that the profile of graph distances from every vertex x in a d -regular Ramanujan graph G concentrates on $\log_{d-1} n$ (the minimum possible value it can concentrate on in a d -regular graph).

Corollary 2. *Let G be a d -regular Ramanujan graph on n vertices. Then for every vertex x in G ,*

$$\#\{y : |\text{dist}(x, y) - \log_{d-1} n| > 3 \log_{d-1} \log n\} = o(n),$$

and furthermore, for all except $o(n)$ vertices y there is a nonbacktracking cycle¹ through x, y of length at most $2 \log_{d-1} n + 6 \log_{d-1} \log n$.

More can be said about high-girth Ramanujan graphs, *e.g.*, the bipartite LPS expanders whose girth is asymptotically $\frac{4}{3} \log_{d-1} n$ (see, *e.g.*, [24, §7]).

Corollary 3. *Let G be a d -regular Ramanujan graph with n vertices and girth g , and set $R = \lceil \log_{d-1} n + 5 \log_{d-1} \log n \rceil$. For every $k \leq g - R$ and simple path $(x_i)_{i=1}^k$ in G , for all but a $o(1)$ -fraction of simple paths $(y_i)_{i=1}^k$ in G , there are vertex-disjoint paths of length R from x_i to y_i for all i .*

¹A nonbacktracking cycle is a sequence of adjacent vertices v_0, \dots, v_k such that $v_0 = v_k$ and $v_i \neq v_{(i+2) \bmod k}$ for all i .

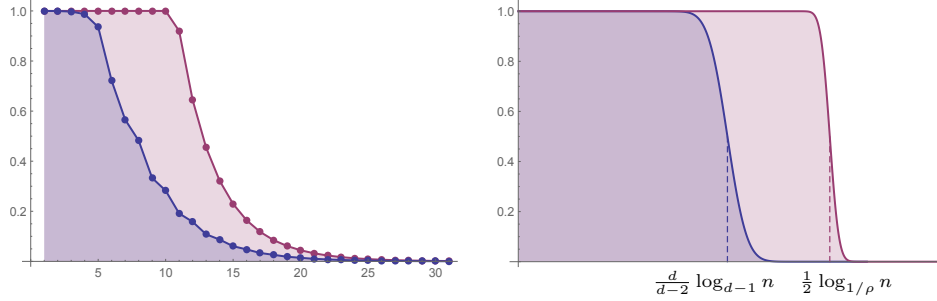


FIGURE 2. Distance of SRW from equilibrium in L^1 (blue) and L^2 (red, capped at 1). On left, the LPS graph on $\text{PSL}(2, \mathbb{F}_{29})$ shown in Fig. 1; on right, asymptotics via Theorem 1 and Proposition 6.

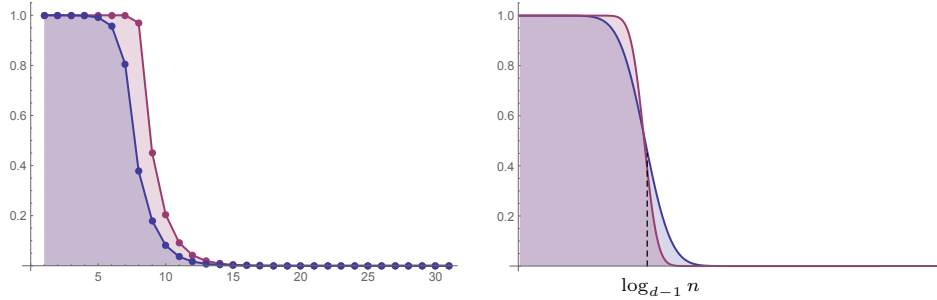


FIGURE 3. The analogue of Fig. 2 for the NBRW (see §1.4).

That Corollaries 2–3 also cover bipartite Ramanujan graphs (recently shown to exist for every degree $d \geq 3$ in [27]) follows from an extension of the proof of Theorem 1 to the bipartite setting (see Corollary 3.9). Moreover, it extends to the case where the graph G is *weakly* Ramanujan (see §1.2).

1.1. Background and related work. The cutoff phenomenon was first identified in pioneering studies of Diaconis, Shahshahani and Aldous [1, 2, 14] in the early 1980’s, and while believed to be widespread, rigorous examples where it was confirmed were scarce. In view of the canonical example where there is no cutoff—SRW on a cycle—and the fact that a necessary condition for any reversible Markov chain to have cutoff is for its inverse spectral-gap to be negligible compared to its mixing time, the second author conjectured [30] in 2004 that on every transitive expander SRW has cutoff.

Durrett [15, §6] conjectured in 2008 that the random walk should have cutoff on a uniformly chosen d -regular graph on n vertices (typically a good expander) with probability tending to 1 as $n \rightarrow \infty$; indeed this is the case, as was verified by the first author and Sly [22] in 2010. Subsequently, expanders without cutoff were constructed in [23], but these were highly asymmetric. The conjectured behavior of cutoff for all transitive expanders was reiterated in the latter works (see [22, Conjecture 6.1] and [23, §3]), yet this was not verified nor refuted on any single example to date.

As a special case, Theorem 1 confirms cutoff on all transitive Ramanujan graphs—in particular for the Lubotzky–Phillips–Sarnak graphs (see Fig. 2).

The concentration of measure phenomenon in expanders, discovered by Alon and Milman [6], implies that the distance from a prescribed vertex is concentrated up to an $O(1)$ -window. Formally, for every sequence of expander graphs G_n on n vertices and vertex $x \in V(G_n)$ there exists a sequence $m_{n,x}$ and constants $a, C > 0$ so that, for every $r > 0$,

$$\#\{y \in V(G_n) : |\text{dist}_{G_n}(x, y) - m_{n,x}| > r\} \leq Ca^{-r}n. \quad (1.2)$$

Corollary 2 shows that $m_{n,x} = \log_{d-1} n + O(\log \log n)$ for Ramanujan graphs.

As for the diameter, Alon and Milman ([6, Theorem 2.7]) showed that $\text{diam}(G) \leq 2\sqrt{2d/(d-\lambda)} \log_2 n$ for every d -regular graph G on n vertices where all nontrivial eigenvalues are at most λ in absolute value. This bound was improved to $\lceil \log_{d/\lambda}(n-1) \rceil$ by Chung [11, Theorem 1], and then to $\lfloor \frac{\cosh^{-1}(n-1)}{\cosh^{-1}(d/\lambda)} \rfloor + 1$ in [12] using properties of $T_k(x)$, the Chebyshev polynomials of the first kind. Since $\cosh(\frac{1}{2} \log(d-1)) = d/(2\sqrt{d-1})$ for any d , this bound translates to $2 \log_{d-1} n + O(1)$ for Ramanujan graphs, and remains the best known upper bound on the diameter of the LPS expanders (for which this was proved directly in [25] via the polynomials $T_k(x)$ as later used in [12]). Corollary 2 implies this asymptotically for every Ramanujan graph: as the distance from any vertex x to most of the vertices is $(1 + o(1)) \log_{d-1} n$, the distance between any two vertices x, y is at most $(2 + o(1)) \log_{d-1} n$. Moreover, one can deduce that for every two vertices x, y and every integer $\ell \geq (2 + o(1)) \log_{d-1} n$, there exists a path of length *exactly* ℓ between x, y .

A new impetus for understanding distances in Ramanujan graphs is due to their role as building blocks in quantum computing; see the influential letter by P. Sarnak [32]. Some of Sarnak’s ideas were developed further by his student N.T. Sardari in an insightful paper [31] posted to the arXiv a few months after the initial posting of the present paper. For a certain infinite family of $(p+1)$ -regular n -vertex Ramanujan graphs, Sardari [31] shows that the diameter is at least $\lfloor \frac{4}{3} \log_p(n) \rfloor$ and also gives an alternative proof of the first part of Corollary 2.

1.2. Extensions. A sequence of connected d -regular graphs ($d \geq 3$ fixed) G_n on n vertices is called *weakly Ramanujan* if, for some $\delta_n = o(1)$ as $n \rightarrow \infty$, every eigenvalue λ of G_n is either $\pm d$ or has $|\lambda| \leq 2\sqrt{d-1} + \delta_n$.

Theorem 4. *On any sequence of d -regular non-bipartite weakly Ramanujan graphs, SRW exhibits cutoff. More precisely, if G_n is such a graph on n vertices then for every initial vertex x , the SRW has*

$$t_{\text{MIX}}(\varepsilon) = \left(\frac{d}{d-2} + o(1)\right) \log_{d-1} n \quad \text{for every fixed } 0 < \varepsilon < 1.$$

Corollary 5. *Let G_n be a d -regular weakly Ramanujan sequence of graphs on n vertices. Then for every vertex x in G_n ,*

$$\#\left\{y : \left| \frac{\text{dist}(x, y)}{\log_{d-1} n} - 1 \right| > \varepsilon \right\} = o(n) \quad \text{for every fixed } \varepsilon > 0.$$

Remark. The weakly Ramanujan hypothesis in Theorem 4 and Corollary 5 may be relaxed to allow some exceptional eigenvalues; for instance, we can allow $n^{o(1)}$ eigenvalues λ to only satisfy $|\lambda| < d - \varepsilon'$ for some $\varepsilon' > 0$ fixed.

By the result of Friedman [17] that a random (uniformly chosen) d -regular graph on n vertices is typically weakly Ramanujan (as conjectured by Alon), Theorem 4 then implies cutoff, re-deriving the above mentioned result of [22].

More generally, for two graphs F and G , a *covering map* $\phi : V(G) \rightarrow V(F)$ is a graph homomorphism that, for every $x \in V(G)$, induces a bijection between the edges incident to x and those incident to $\phi(x)$. If such a map exists, we say G is a *lift* (or a *cover*) of F ; a random n -lift of F is a uniformly chosen lift out of all those with cover number n (i.e., $|\phi^{-1}(x)| = n$ for all x).

Friedman and Kohler [18] recently proved (see also [9, Corollary 20] by Bordenave) that for every fixed d -regular base graph F and $\delta > 0$, if G is a random n -lift of F then typically all of its “new” eigenvalues (those not inherited from F via pullback) are at most $2\sqrt{d-1} + \delta$. By the remark above, Theorem 4 and Corollary 5 apply here (for any fixed regular F).

1.3. Cutoff in L^p -distance. Theorem 1 showed that Ramanujan graphs have an optimal t_{MIX} for SRW: the total-variation distance from (1.1) matches a lower bound valid for every d -regular graph on n vertices (Fact 2.1 in §2). It turns out that Ramanujan graphs are extremal for L^p -mixing for *all* $p \geq 1$.

For $1 \leq p \leq \infty$, the L^p -mixing time of a Markov chain with transition kernel P from its stationary distribution π is defined as

$$t_{\text{MIX}}^{(L^p)}(\varepsilon) = \min\{t : D_p(t) \leq \varepsilon\} \quad \text{where} \quad D_p(t) = \max_x \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^p(\pi)}$$

(note that $p = 1$ measures total-variation mixing since $D_1(t) = 2D_{\text{TV}}(t)$, whereas the L^2 -distance $D_2(t)$ is also known as the chi-square distance). Chen and Saloff-Coste [10, Theorem 1.5] showed that a lazy random walk on a family of expander graphs exhibits L^p -cutoff, at some unknown location, for all $p \in (1, \infty]$. (On the notable exception of $p = 1$, it is said in [10] that there “the question is more subtle and no good general answer is known.”)

The following theorem gives a lower bound on for SRW on a d -regular graph, asymptotically achieved by Ramanujan graphs for all $p \in (1, \infty]$.

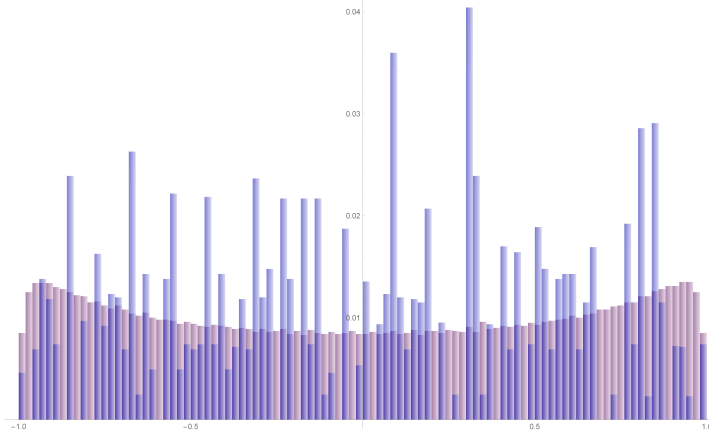


FIGURE 4. Normalized eigenvalues of Ramanujan graphs on $n \approx 10^4$ vertices: the 6-regular LPS expanders on $\text{PSL}(2, \mathbb{F}_q)$ for $q = 29$ (front; every nontrivial eigenvalue has multiplicity at least $\frac{q-1}{2}$) and a 1000-lift of the 3-regular Peterson graph (back).

Proposition 6. Fix $d \geq 3$ and let $\rho = 2\sqrt{d-1}/d$. Then for all connected d -regular graphs G on n vertices and every fixed $\varepsilon > 0$, the SRW satisfies

$$t_{\text{MIX}}^{(L^p)}(\varepsilon) \geq \begin{cases} c_{d,p} \log_{d-1} n - O(\log \log n) & \text{if } p \in (1, 2] \\ \frac{p-1}{p} \log_{1/\rho} n - O(\log \log n) & \text{if } p \in [2, \infty) \end{cases}, \quad (1.3)$$

where $c_{d,p} = [2\beta - 1 + \frac{p}{p-1} H_{d-1}(\beta \parallel \frac{d-1}{d})]^{-1}$ for $\beta = [1 + (d-1)^{(p-2)/p}]^{-1}$ and $H_b(\beta \parallel \alpha) = \beta \log_b(\frac{\beta}{\alpha}) + (1-\beta) \log_b(\frac{1-\beta}{1-\alpha})$ is the relative entropy function. Furthermore, if G is non-bipartite Ramanujan then, with the same notation,

$$t_{\text{MIX}}^{(L^p)}(\varepsilon) = \begin{cases} c_{d,p} \log_{d-1} n + O(\log \log n) & \text{if } p \in (1, 2] \\ \frac{p-1}{p} \log_{1/\rho} n + O(\log \log n) & \text{if } p \in [2, \infty) \end{cases}. \quad (1.4)$$

1.4. Method of proof. The natural route to exploit spectral details on the transition kernel P for an upper bound on the L^1 -distance from the stationary distribution π is via the L^2 -distance (see, e.g., [19, Theorem 3.2]). However, this fails to give the sought bound $\frac{d+o(1)}{d-2} \log_{d-1} n$ for the SRW, as we see from Proposition 6 that the SRW on Ramanujan graphs exhibits an L^2 -cutoff at $\frac{1}{2} \log_{1/\rho} n > (1+\eta) \frac{d}{d-2} \log_{d-1} n$ for some $\eta(d) > 0$ (see (2.7)).

To remedy this, we turn to the nonbacktracking random walk (NBRW), which moves from the directed edge (x, y) to a uniformly chosen edge (y, z) such that $z \neq x$. In recent years, delicate spectral information on random graphs has been extracted by counting nonbacktracking paths; notably, this was essential in the proofs that random d -regular graphs and random lifts are weakly Ramanujan [9, 17, 18]. Here we follow the reverse route, and use spectral information on the graph to control the nonbacktracking paths.

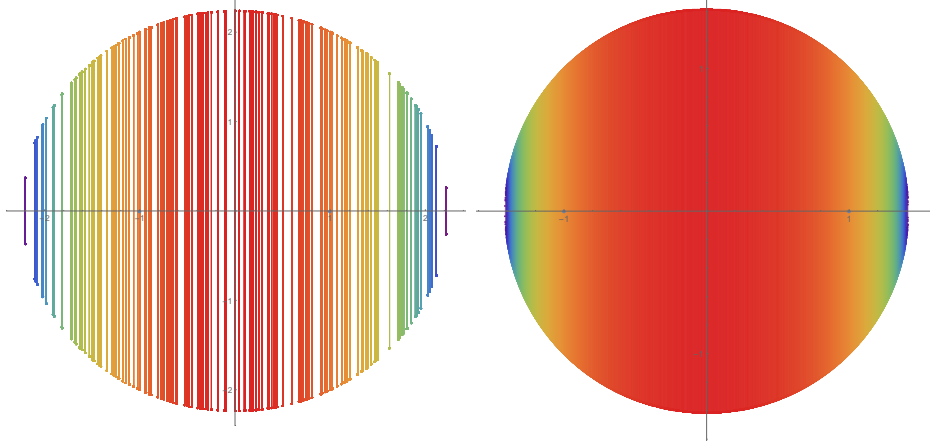


FIGURE 5. Eigenvalues of the nonbacktracking operator B of the graphs from Fig. 4 (with the LPS expander on the left). Colors of chords between pairs of eigenvalues $\theta, \bar{\theta}$ depict the inner product of their corresponding eigenvectors w, w' (blue near 1).

The known relation between the spectrum of G and the spectrum of the nonbacktracking operator B implies that if G is Ramanujan, each of its nontrivial eigenvalues λ is mapped to eigenvalues $\theta, \theta' \in \mathbb{C}$ of B with modulus $\sqrt{d-1}$ (see Fig. 4-5 showing this effect for two Ramanujan graphs with drastically different spectral features). For intuition, note that, had the operator B been self-adjoint and transitive (it is neither), we would have gotten that the L^2 -distance at time t is $O(\sqrt{n}(d-1)^{-t/2})$, implying the correct upper bound of $(1+o(1))\log_{d-1} n$ for the NBRW.

Fortunately, it turns out that, while not a normal operator, B is unitarily similar to a matrix Λ that is block-diagonal with $n-1$ non-singleton blocks ($n-2$ if G is bipartite), each of which has size 2×2 (despite potential high multiplicities in the eigenvalues of G) and corresponds to an eigenvector pair w, w' with matching eigenvalues $\theta, \bar{\theta}$. This description of B appears in Proposition 3.1 and may be of independent interest.

1.5. Organization. The rest of this paper is organized as follows. Section 2 describes the reduction of L^1 -mixing for the SRW to that of the NBRW and establishes the optimality of the L^p -cutoff of SRW on Ramanujan graphs for all $p > 1$ (Proposition 6). Section 3 studies the NBRW, beginning in §3.1 with the aforementioned spectral decomposition and its properties (an exact computation of the off-diagonal entries is deferred to Proposition 4.1 in §4). In §3.2 we give the proof of the non-bipartite case, which implies Theorem 1 and Corollary 2; and §3.3 includes the proofs of the extensions to bipartite and weakly Ramanujan graphs, which imply Theorem 4 and Corollary 5.

2. SIMPLE RANDOM WALK

2.1. Reduction to NBRW. As described in [22] (see §2.3 and §5.2 there), cutoff for SRW can be reduced to cutoff for the NBRW as follows. Let $G = (V, E)$ be a d -regular graph and let \mathbb{T}_d be the infinite regular tree rooted at ξ , the universal cover of G . For a given vertex $x \in V$, consider a cover map $\phi : \mathbb{T}_d \rightarrow V$ with $\phi(\xi) = x$, and observe that if (\mathcal{X}_t) is SRW on \mathbb{T}_d started at ξ , then $X_t = \phi(\mathcal{X}_t)$ is SRW on G started at x . (This was also used in the proof of the Alon–Boppana Theorem given in [25, Proposition 4.2].) Similarly, if \mathcal{Y}_t is NBRW on \mathbb{T}_d started at $(\xi, \sigma) \in \vec{E}(\mathbb{T}_d)$, and we write $\mathcal{Y}_t = (\mathcal{Y}'_t, \mathcal{Y}''_t)$ to denote its endpoint vertices, then $Y_t = (Y'_t, Y''_t)$ given by $Y'_t = \phi(\mathcal{Y}'_t)$ and $Y''_t = \phi(\mathcal{Y}''_t)$ is NBRW on G started at $(x, \phi(\sigma))$. By symmetry, if

$$\mathcal{E}_{t,\ell} := \{\text{dist}(\xi, \mathcal{X}_t) = \ell\},$$

then the conditional distribution of \mathcal{X}_t given $\mathcal{E}_{t,\ell}$ is uniform over the vertices at distance ℓ from ξ in \mathbb{T}_d . Therefore,

$$\mathbb{P}_x(X_t \in \cdot \mid \mathcal{E}_{t,\ell}) = \frac{1}{d} \sum_{\sigma: \xi\sigma \in E(\mathbb{T}_d)} \mathbb{P}_{(x, \phi(\sigma))}(Y''_{\ell-1} \in \cdot).$$

As a projection can only decrease total-variation distance, letting $\ell = t_{\text{MIX}}(\varepsilon)$ for the NBRW on G and π be the uniform distribution over $V(G)$, we get

$$\|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\text{TV}} \leq \varepsilon + \mathbb{P}(\text{dist}(\xi, \mathcal{X}_t) < \ell),$$

and in particular, taking a maximum over x shows that the SRW on G has

$$D_{\text{TV}}(t) \leq \varepsilon + \mathbb{P}(\text{dist}(\xi, \mathcal{X}_t) < \ell). \quad (2.1)$$

Finally, since SRW on \mathbb{T}_d ($d \geq 3$) is transient, \mathcal{X}_t returns to ξ only a finite number of times almost surely. If $\mathcal{X}_t \neq \xi$ then $\text{dist}(\mathcal{X}_{t+1}, \xi) - \text{dist}(\mathcal{X}_t, \xi)$ is equal to -1 with probability $1/d$ and $+1$ otherwise. Therefore, by the CLT,

$$\frac{\text{dist}(\mathcal{X}_t, \xi) - ((d-2)/d)t}{(2\sqrt{d-1}/d)\sqrt{t}} \Rightarrow \mathcal{N}(0, 1). \quad (2.2)$$

Thus, if $\ell \rightarrow \infty$ then by (2.1), for every fixed $s \geq 0$ the SRW on G satisfies

$$\limsup_{n \rightarrow \infty} D_{\text{TV}}\left(\frac{d}{d-2}\ell + s\sqrt{\ell}\right) \leq \varepsilon + \mathbb{P}(Z > c_d s), \quad (2.3)$$

where Z is a standard normal random variable and $c_d = \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}}$.

Conversely, the number of vertices at distance ℓ from a given vertex $x \in V$ is at most $d(d-1)^\ell$. So, on the event $\text{dist}(\mathcal{X}_t, \xi) < \log_{d-1}(\varepsilon n/d)$, the SRW X_t is confined to a set of at most εn vertices of G , thus its total-variation distance from π is at least $1 - \varepsilon$. Altogether, (2.2) implies the following.

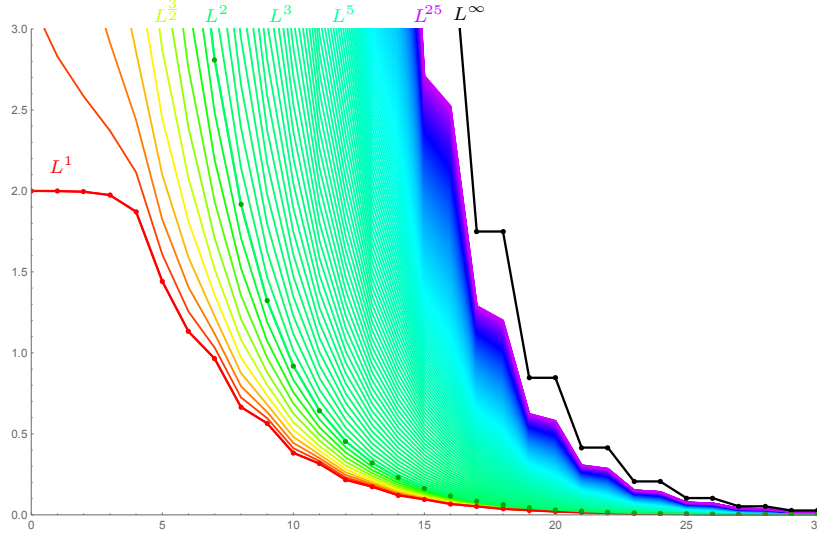


FIGURE 6. The L^p -distance ($p \geq 1$) from equilibrium of SRW on the LPS graph on $\text{PSL}(2, \mathbb{F}_{29})$ shown in Fig. 1, highlighting $p = 1, 2, \infty$.

Fact 2.1. *For every d -regular graph on n vertices with $d \geq 3$ fixed, and every fixed $s, \varepsilon > 0$, the SRW on G satisfies*

$$\liminf_{n \rightarrow \infty} D_{\text{TV}} \left(t - s\sqrt{\log_{d-1} n} \right) \geq 1 - \varepsilon - \mathbb{P}(Z > c_d s) \quad (2.4)$$

at $t = \frac{d}{d-2} \log_{d-1}(\varepsilon n/d)$, where $c_d = \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}}$ and $Z \sim \mathcal{N}(0, 1)$.

Comparing the two bounds (2.3)–(2.4) with the desired estimate (1.1) for the SRW in Theorem 1, we see that the latter will follow if we show that the NBRW has cutoff at time $\log_{d-1} n + o(\sqrt{\log n})$ with window $o(\sqrt{\log n})$. This will be achieved in §3 via a spectral analysis of the nonbacktracking walk.

2.2. Optimal L^p -mixing on Ramanujan graphs. We begin with the special case $p = 2$ of Proposition 6.

Lemma 2.2. *Fix $d \geq 3$ and let $\rho = 2\sqrt{d-1}/d$. For every fixed $\varepsilon > 0$ and every connected d -regular graph G on n vertices, the SRW satisfies*

$$t_{\text{MIX}}^{(L^2)}(\varepsilon) \geq \frac{1}{2} \log_{1/\rho} n - O(\log \log n).$$

Moreover, if G is non-bipartite Ramanujan then this is tight: SRW has an L^2 -cutoff at $\frac{1}{2} \log_{1/\rho} n > (1+\eta) \frac{d}{d-2} \log_{d-1} n$ for some constant $\eta = \eta(d) > 0$.

Proof of Lemma 2.2. Let P^t be the t -step transition kernel of SRW, and let π be the uniform distribution on $V(G)$. For any $x \in V(G)$,

$$\sum_y P^t(x, y)^2 = P^{2t}(x, x) \geq Q^{2t}(\xi, \xi),$$

where Q^t is the t -step transition kernel of SRW on \mathbb{T}_d , the infinite d -regular tree rooted at ξ ; indeed, as argued above, if \mathcal{X}_t is SRW on the cover tree \mathbb{T}_d then $X_t = \phi(\mathcal{X}_t)$ is SRW on G , where ϕ is the cover map, and in particular a return to the root in the former implies a return to the origin in the latter.

The probability $Q^{2t}(\xi, \xi)$ is nothing but the probability of a 1D biased walk, reflected at 0, to be 0 at time $2t$, well-known (cf. [35, §5, p128]) to be

$$Q^{2t}(\xi, \xi) = \frac{2\rho^2}{1-\rho} \frac{\rho^{2t}}{t 2^{2t-2}} \binom{2t-2}{t-1} \sim \frac{2\rho^2}{1-\rho^2} \frac{\rho^{2t}}{\sqrt{\pi t^3}} \quad \text{for } \rho = \frac{2\sqrt{d-1}}{d}. \quad (2.5)$$

In particular, using the standard expansion of the L^2 -distance,

$$\sum_x \left(\frac{\mu(x)}{\pi(x)} - 1 \right)^2 \pi(x) = \sum_x \frac{\mu^2(x)}{\pi(x)} - 1 \quad (2.6)$$

which holds for every probability distribution μ , thus we have

$$\left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \geq c_d n \rho^{2t} t^{-3/2} - 1$$

for $c_d = 2\rho^2[(1-\rho^2)\sqrt{\pi}]^{-1}$. Consequently, from any initial $x \in V$ we have

$$t_{\text{MIX}}^{(L^2)}(x, \varepsilon) \geq \frac{\log(n/\varepsilon)}{2 \log(1/\rho)} - O(\log \log n),$$

where $t_{\text{MIX}}^{(L^2)}(x, \varepsilon)$ is the first t where $\|P^t(x, \cdot)/\pi - 1\|_{L^2(\pi)}$ becomes at most ε .

We next argue that

$$\frac{1}{2} \log_{1/\rho} n > \frac{d}{d-2} \log_{d-1} n \quad \text{for every real } d \in (2, \infty) \text{ and } n \geq 2. \quad (2.7)$$

Indeed, (2.7) is equivalent to having $\frac{d-2}{d} \log(d-1) > 2 \log\left(\frac{d}{2\sqrt{d-1}}\right)$ for all real $d \in (2, \infty)$, which, in turn, immediately follows from the fact that

$$f(d) := \frac{d-2}{d} \log(d-1) - 2 \log\left(\frac{d}{2\sqrt{d-1}}\right)$$

has $f'(d) = 2d^{-2} \log(d-1)$, so $f(2) = 0$ whereas $f'(d) > 0$ for all $d > 1$.

Finally, when G is Ramanujan, the sought upper bound on the L^2 -distance follows from considering the spectral representation (see, *e.g.*, [3])

$$\|P^t(x, \cdot)/\pi - 1\|_{L^2(\pi)}^2 = n \sum_{i=2}^n |f_i(x)|^2 (\lambda_i/d)^{2t} \quad (2.8)$$

for $\{f_i\}_{i=1}^n$ an orthonormal basis of eigenfunctions with eigenvalues $\{\lambda_i\}_{i=1}^n$ of the adjacency matrix and $\lambda_1 = d$, and plugging in $|\lambda_i| \leq 2\sqrt{d-1}$. ■

Remark 2.3. A different perspective on Lemma 2.2 is given by the next proof of a slightly weaker statement. By the generalization by Serre [33] (see [13, Theorem 1.4.9]) of the Alon–Boppana Theorem [29], for every $\varepsilon > 0$ there exists $c_{\varepsilon, d} > 0$ such that G has at least $c_{\varepsilon, d} n$ eigenvalues λ with

$|\lambda| > 2\sqrt{d-1} - \varepsilon$. Applying this fact for some $\varepsilon(d) > 0$ to be specified later, since $\frac{1}{n} \sum_x \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 = \sum_{i=2}^n (\lambda_i/d)^{2t}$ where $\lambda_2, \dots, \lambda_n$ are the nontrivial eigenvalues of G (this follows from (2.8) since an average over x allows one to replace $\sum_x |f_i(x)|^2$ by 1 for each i), we deduce that

$$\max_x \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \geq \frac{1}{n} \sum_x \left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \geq c_{\varepsilon, d} n \left(\frac{2\sqrt{d-1} - \varepsilon}{d} \right)^{2t}.$$

Consequently,

$$t_{\text{MIX}}^{(L^2)}(\delta) \geq \frac{\log(n/\delta)}{2 \log\left(\frac{d}{2\sqrt{d-1} - \varepsilon}\right)} - O(1). \quad (2.9)$$

The proof now follows from (2.7) as we may choose $\varepsilon(d), \eta(d) > 0$ so that the right-hand of (2.9) would be at least $(1 + \eta - o(1)) \frac{d}{d-2} \log_{d-1} n$, as needed.

For the general case of $p \in [1, \infty]$, we need the following simple claims.

Claim 2.4. *Let G be a d -regular graph on n vertices, and let \mathbb{T}_d be the infinite d -regular tree rooted at ξ . For every $1 \leq p < \infty$, SRW on G satisfies*

$$\left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^p(\pi)} \geq n^{(p-1)/p} \|Q^t(\xi, \cdot)\|_p - 1$$

for all x, t , where P and Q are the transition kernels of SRW on G and \mathbb{T}_d .

Proof. By the triangle inequality w.r.t. $\|\cdot\|_{L^p(\pi)}$,

$$\left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^p(\pi)} \geq n^{(p-1)/p} \|P^t(x, \cdot)\|_p - 1.$$

Since $P^t(x, y) = \sum_{\eta \in \phi^{-1}(y)} Q^t(\xi, \eta)$ for every cover map $\phi : V(\mathbb{T}_d) \rightarrow V(G)$ with $\phi(\xi) = x$, using the fact $(\sum_{i=1}^k a_i)^p \geq \sum_{i=1}^k a_i^p$ for every $a_1, \dots, a_k > 0$ and $p \geq 1$ gives

$$(P^t(x, y))^p \geq \sum_{\eta \in \phi^{-1}(y)} (Q^t(\xi, \eta))^p.$$

Summing over all y gives $\|P^t(x, \cdot)\|_p \geq \|Q^t(\xi, \cdot)\|_p$, as required. \blacksquare

Claim 2.5. *Fix $d \geq 3$ and let \mathbb{T}_d be the infinite d -regular tree rooted at ξ . There exist constants $c_1(d), c_2(d) > 0$ such that, for all k and t ,*

$$c_1(d) \leq \frac{\mathbb{P}_\xi(|\mathcal{X}_t| = k)}{\frac{k+1}{t} \mathbb{P}\left(Z_t = \frac{k+t}{2}\right)} \leq c_2(d)$$

where $|\mathcal{X}_t|$ is the distance of \mathcal{X}_t from its origin ξ , and $Z_t \sim \text{Bin}(t, \frac{d-1}{d})$.

Proof. The case $k = 0$ follows from (2.5), since $\mathbb{P}(Z_{2t} = t) = \left(\frac{d-1}{d}\right)^t d^{-t} \binom{2t}{t}$, which is $(\rho/2)^{2t} \binom{2t}{t}$. This extends to all k using the decomposition

$$\mathbb{P}_\xi(|\mathcal{X}_t| = k) = \sum_{\ell=0}^{t-1} \mathbb{P}_\xi(|\mathcal{X}_\ell| = 0) \mathbb{P}_\xi(\{|\mathcal{X}_j| > 0 : 1 \leq j \leq t - \ell\}, |\mathcal{X}_{t-\ell}| = k)$$

and the Ballot Theorem (see, e.g., [16, §III.1]). \blacksquare

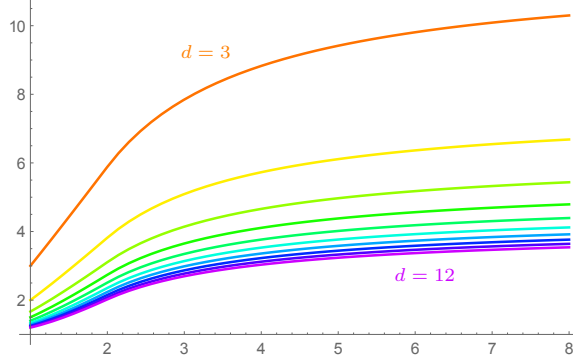


FIGURE 7. L^p -cutoff location (normalized by $\log_{d-1} n$) as a function of $p \geq 1$ for Ramanujan graphs with degree $d = 3, \dots, 12$. The functions are C^1 , but not C^2 at $p = 2$.

For more general local limit theorems on trees, see, *e.g.*, [21].

Proof of Proposition 6. With Claims 2.4 and 2.5 in mind, and using their notation, for every t and $p \in [1, \infty]$ we have

$$\begin{aligned} \|Q^t(\xi, \cdot)\|_p^p &= \sum_{k \geq 0} (d-1)^k \left((d-1)^{-k} \mathbb{P}_\xi(|\mathcal{X}_t| = k) \right)^p \\ &\geq \left(\frac{c_1(d)}{t} \right)^p \sum_{k \geq 0} \left((d-1)^{k(1-p)/p} \mathbb{P}(Z_t = (k+t)/2) \right)^p. \end{aligned} \quad (2.10)$$

Writing $\beta t = (k+t)/2$ (so that $k = (2\beta - 1)t$), the large deviation estimate

$$\mathbb{P}(Z_t = \beta t) \asymp t^{-1/2} \exp[-H_e(\beta \parallel \frac{d-1}{d})t] \quad (2.11)$$

for the binomial variable Z_t thus leads to the following optimization problem:

$$\min \left\{ \frac{p-1}{p}(2\beta-1) + H_{d-1}(\beta \parallel \frac{d-1}{d}) : \frac{1}{2} \leq \beta \leq 1 \right\}. \quad (2.12)$$

(Observe that in fact $\beta \leq \frac{d-1}{d}$ since for $\beta > \frac{d-1}{d}$ both terms are increasing.)

Let $f(\beta)$ denote the objective in (2.12). Then

$$f'(\beta) = \frac{2(p-1)}{p} + \log_{d-1} \left(\frac{\beta}{(1-\beta)(d-1)} \right),$$

and solving $f'(\beta) = 0$ we get $\frac{1-\beta}{\beta} = (d-1)^{(p-2)/p}$. Since $f''(\beta)$ is positive, it follows that the minimizer of (2.12) is at

$$\beta^* = \frac{1}{(d-1)^{(p-2)/p} + 1} \vee \frac{1}{2}. \quad (2.13)$$

(Observe that $\beta^* = 1/2$ iff $p \geq 2$, hence the two regimes for the L^p -cutoff location as a function of p .) By (2.10)–(2.11), for some $c = c_d > 0$,

$$\|Q^t(\xi, \cdot)\|_p \geq c_d t^{-3/2} (d-1)^{-f(\beta^*)t},$$

and therefore, by Claim 2.4, for every starting vertex x ,

$$\left\| \frac{P^t(x, \cdot)}{\pi} - 1 \right\|_{L^p(\pi)} \geq c_d n^{(p-1)/p} t^{-3/2} (d-1)^{-f(\beta_*)t} - 1. \quad (2.14)$$

This implies (1.3) (and is furthermore valid for every starting vertex x).

For matching upper bounds in case G is a Ramanujan graph, first take $p \geq 2$. The lower bound established above is $D_p(t) \geq c_d n^{(p-1)/p} t^{-3/2} \rho^t - 1$. Recalling Lemma 2.2, for Ramanujan graphs,

$$D_2(t) \leq \sqrt{n} \rho^t, \quad D_\infty(t) \leq D_2(\lfloor t/2 \rfloor)^2 \leq n \rho^t,$$

using the well-known fact (a routine application of Cauchy–Schwarz) that

$$D_\infty(s+t) \leq D_2(t) D_2^*(s), \quad (2.15)$$

where $D_2^*(s)$ corresponds to the reversed chain (here $D_2(s) = D_2^*(s)$ as SRW is reversible). So, by the Riesz–Thorin Interpolation Theorem (see, *e.g.*, [34, Theorem 1.3, p. 179]), for $2 \leq p \leq \infty$, we deduce that $D_p(t) \leq n^{(p-1)/p} \rho^t$.

Having established (1.4) for $p \geq 2$, now take $1 < p \leq 2$. Let

$$P^t(x, \cdot) = \sum_k \mathbb{P}_\xi(|\mathcal{X}_t| = k) \tilde{\mu}_k(x, \cdot) \quad (2.16)$$

where $\tilde{\mu}_k$ is the law of the projection of NBRW on the endpoint of its directed edge, started at a uniform edge originating from x . By Jensen’s inequality,

$$(d(d-1)^{k-1})^{-1/p} \|\tilde{\mu}_k(x, \cdot) - \frac{1}{n}\|_p \leq (d(d-1)^{k-1})^{-1/2} \|\tilde{\mu}_k(x, \cdot) - \frac{1}{n}\|_2.$$

Notice that

$$n \left\| \tilde{\mu}_k(x, \cdot) - \frac{1}{n} \right\|_2^2 = \left\| \frac{\tilde{\mu}_k(x, \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq \max_{y: xy \in E(G)} \left\| \frac{\mu_{k-1}((x, y), \cdot)}{\pi_{\vec{E}}} - 1 \right\|_{L^2(\pi_{\vec{E}})}^2$$

where μ_k is the k -step transition kernel of the NBRW and $\pi_{\vec{E}}$ is its stationary distribution. In our analysis of the NBRW in §3, we will show (see (3.13)) that the right-hand side of the last display is $O(nk^2(d-1)^{-k})$, whence

$$\|\tilde{\mu}_k(x, \cdot) - \frac{1}{n}\|_p \leq c_d k (d-1)^{k(1-p)/p}.$$

Recalling (2.16), it now follows that

$$\begin{aligned} \|P^t(x, \cdot) - \frac{1}{n}\|_p &\leq (t+1) \max_{0 \leq k \leq t} (\mathbb{P}_\xi(|\mathcal{X}_t| = k)) \|\tilde{\mu}_k(x, \cdot) - \frac{1}{n}\|_p \\ &\leq c'_d t^2 \max_{0 \leq k \leq t} (d-1)^{k(1-p)/p} (\mathbb{P}_\xi(|\mathcal{X}_t| = k)), \end{aligned}$$

which, in view of (2.10), gives rise to the same optimization problem (2.12). Therefore, the right-hand side of the last display is at most $t^C (d-1)^{-f(\beta_*)t}$ for $C > 0$ fixed. Taking t as in (1.4) with a suitable additive $O(\log \log n)$ term gives $(d-1)^{f(\beta_*)t} \geq n^{(p-1)/p} t^{2C}$. Thus,

$$\|P^t(x, \cdot)/\pi - 1\|_{L^p(\pi)} = n^{(p-1)/p} \|P^t(x, \cdot) - \frac{1}{n}\|_p \leq t^{-C},$$

establishing (1.4) for all $1 < p \leq 2$. ■

3. NONBACKTRACKING WALKS

3.1. Spectral decomposition. The spectrum of the nonbacktracking walk has been thoroughly studied, in part due to the fact that its eigenvalues are precisely the inverse of the poles of the so-called Ihara Zeta function of the graph (cf. [8, 20]). Our analysis here, on the other hand, hinges on the structure of the eigenfunctions, starting with a spectral decomposition of the nonbacktracking operator; this builds on properties of this operator that appear implicitly in [20] (see also [5, 7, 8] as well as [26, Exercise 6.59]). Proposition 3.1 below gives a more complete picture.

Throughout this section, for a graph $G = (V, E)$, we denote its adjacency matrix by $A = A(G)$ and let $\lambda_1 = d \geq \lambda_2 \geq \dots \geq \lambda_n$ be its eigenvalues. Denote by \vec{E} the set of $N = 2|E|$ directed edges of G ; we refer to undirected edges as $xy \in E$ and to directed ones as $(x, y) \in \vec{E}$ for the sake of clarity. The nonbacktracking walk matrix B is the $(\vec{E} \times \vec{E})$ -matrix given by

$$B_{(u,v),(x,y)} = \mathbb{1}_{\{v=x, u \neq y\}} \quad \text{for } (u, v) \text{ and } (x, y) \text{ in } \vec{E}. \quad (3.1)$$

Though B may not be a normal operator, it can be decomposed as follows.

Proposition 3.1. *Let $G = (V, E)$ be a connected d -regular graph ($d \geq 3$) on n vertices. Let $N = dn$ and let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of the adjacency matrix, with $\lambda_1 = d$. Then the operator B from (3.1) is unitarily similar to*

$$\Lambda = \text{diag} \left(d-1, \begin{pmatrix} \theta_2 & \alpha_2 \\ 0 & \theta'_2 \end{pmatrix}, \dots, \begin{pmatrix} \theta_n & \alpha_n \\ 0 & \theta'_n \end{pmatrix}, \overbrace{-1, \dots, -1}^{N/2-n}, \overbrace{1, \dots, 1}^{N/2-n+1} \right) \quad (3.2)$$

where $|\alpha_i| < 2(d-1)$ for all i and $\theta_i, \theta'_i \in \mathbb{C}$ are defined as the solutions to

$$\theta^2 - \lambda_i \theta + d - 1 = 0. \quad (3.3)$$

Remark 3.2. The exact value of $|\alpha_i|$ is shown in Proposition 4.1 to be $d-2$ for every $|\lambda_i| \leq 2\sqrt{d-1}$ and $\sqrt{d^2 - \lambda_i^2}$ for every $2\sqrt{d-1} < |\lambda_i| < d$.

Remark 3.3. We see that every eigenvalue $\theta \neq \pm 1$ of B is of the form $\lambda/2 \pm \sqrt{(\lambda/2)^2 - (d-1)}$ for some eigenvalue λ of A (with $\theta = d-1$ matching the principal eigenvalue $\lambda = d$). Indeed, this well-known fact follows from Bass's Formula [8], which in the d -regular case is equivalent to the statement that $f_B(\theta) = (1 - \theta^2)^{N/2-n} f_A(\theta + (d-1)/\theta)$ for f_A and f_B the characteristic polynomials of A and B , respectively.

- (i) $\lambda = d$ corresponds to $\theta = d-1$, the principal eigenvalue of B matching the eigenvector $w_1 \equiv N^{-1/2}$; the second solution, $\theta' = 1$, was already accounted for in (3.2). An eigenvalue of $\lambda = -d$ (when G is bipartite) yields $\theta = -(d-1)$ and an extra -1 eigenvalue ($N - n + 1$ altogether).

- (ii) $2\sqrt{d-1} < |\lambda| < d$ yields two eigenvalues $\theta \neq \theta' \in \mathbb{R}$ of B .
- (iii) $\lambda < |2\sqrt{d-1}|$ yields $\theta = \bar{\theta}' \in \mathbb{C} \setminus \mathbb{R}$ with $|\theta| = \sqrt{d-1}$ (for instance, $\lambda = 0$ corresponds to $\theta = i\sqrt{d-1}$ and $\theta' = -i\sqrt{d-1}$).
- (iv) $\lambda = \pm 2\sqrt{d-1}$ gives a single solution $\theta = \pm\sqrt{d-1}$ with multiplicity 2.

Remark 3.4. For each $\theta \in \mathbb{C}$, define $T_\theta : \ell^2(V) \rightarrow \ell^2(\vec{E})$ by

$$(T_\theta f)(x, y) := \theta f(y) - f(x). \quad (3.4)$$

Each solution $\theta \neq \pm 1$ of equation (3.3), for some λ such that $Af = \lambda f$, is an eigenvalue of B corresponding to the eigenvector $T_\theta f$; indeed,

$$\begin{aligned} (BT_\theta f)(x, y) &= \sum_{\substack{z:yz \in E \\ z \neq x}} (\theta f(z) - f(y)) = \theta[(Af)(y) - f(x)] - (d-1)f(y) \\ &= [\theta\lambda - (d-1)]f(y) - \theta f(x) = \theta(T_\theta f)(x, y); \end{aligned}$$

where the last equality used (3.3) to replace $\theta\lambda$ by $\theta^2 + d - 1$; thus, $T_\theta f$ is an eigenfunction of B corresponding to θ as long as $T_\theta f \neq 0$, and clearly $T_\theta f = 0$ only if $\theta = \pm 1$ (which, in turn, occurs iff $\lambda = \pm d$).

Proof of Proposition 3.1. Observe that $\ell^2(\vec{E}) = \ell_+^2(\vec{E}) \oplus \ell_-^2(\vec{E})$ where

$$\ell_+^2(\vec{E}) = \{w : w(x, y) = w(y, x)\}, \quad \ell_-^2(\vec{E}) = \{w : w(x, y) = -w(y, x)\},$$

as the term for (x, y) in $\langle w_+, w_- \rangle$ cancels with that of (y, x) if $w_\pm \in \ell_\pm^2(\vec{E})$.

With this in mind, the eigenspaces of 1 and -1 in B are straightforward: the *star spaces* $\mathcal{S}_- \subset \ell_-^2(\vec{E})$ and $\mathcal{S}_+ \subset \ell_+^2(\vec{E})$ are defined by

$$\mathcal{S}_\pm = \text{Span}(\{s_x^\pm : x \in V\}), \quad \text{where} \quad s_x^\pm(u, v) = \begin{cases} 1 & u = x, \\ \pm 1 & v = x, \\ 0 & \text{otherwise.} \end{cases}$$

For every $w \in \ell_-^2(\vec{E})$ and s_x^- as above $\langle w, s_x^- \rangle = 2 \sum_{y:xy \in E} w(x, y)$, and so $(Bw)(x, y) = -w(y, x) = w(x, y)$ when in addition $w \perp s_y^-$. Thus,

$$Bw = w \quad \text{for every} \quad w \in \ell_-^2(\vec{E}) \cap \mathcal{S}_-^\perp, \quad (3.5)$$

and similarly,

$$Bw = -w \quad \text{for every} \quad w \in \ell_+^2(\vec{E}) \cap \mathcal{S}_+^\perp. \quad (3.6)$$

As for the dimension of these spaces, note that if $\{a_x\}_{x \in V}$ is such that $\sum a_x s_x^- = 0$ then $a_x = a_y$ for every $xy \in E$; since G is connected, this implies that $\dim(\mathcal{S}_-) = n - 1$, thus B has an orthonormal system of $N/2 - (n - 1)$ eigenvectors with eigenvalue 1. Similarly, if $\sum a_x s_x^+ = 0$ then $a_x = -a_y$ for every $xy \in E$, so the eigenspace of -1 has dimension $N/2 - (n - 1)$ if G is bipartite and dimension $N/2 - n$ otherwise.

Having specified these eigenvectors of B as well as those corresponding to θ_i, θ'_i in Remark 3.4, we proceed to analyzing their inner products. Observe that after appropriate permutations of its rows and columns, B becomes block diagonal with blocks $J_d - I_d$, where J_d and I_d are the all-one matrix and identity matrix of order d , respectively; thus, B has an inverse, which under the same permutations is block diagonal with blocks $(d-1)^{-1}J_d - I_d$, so the matrix $C := (d-1)B^{-1} + B$ (which, of course, satisfies $Cw = ((d-1)/\theta + \theta)w$ for every eigenfunction w of B with eigenvalue θ) is given by

$$C_{(u,v),(x,y)} = \begin{cases} 1 & \{v = x, u \neq y\} \text{ or } \{y = u, v \neq x\}, \\ -(d-2) & (u,v) = (y,x), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, C is real symmetric, and $\ell_+^2(\vec{E})$ and $\ell_-^2(\vec{E})$ are invariant under it. Furthermore, if $f \in \ell^2(V)$ and $w_f \in \ell^2(\vec{E})$ is given by $w_f(x, y) := f(y)$ then

$$\begin{aligned} (Cw_f)(x, y) &= \sum_{\substack{z:yz \in E \\ z \neq x}} f(z) + \sum_{\substack{v:vx \in E \\ v \neq y}} f(x) - (d-2)f(x) \\ &= \sum_{\substack{z:yz \in E \\ z \neq x}} f(z) + f(x) = (Af)(y), \end{aligned}$$

and similarly, if $w'_f(x, y) := f(x)$ then $(Cw'_f)(x, y) = (Af)(x)$. Moreover, $\langle w_f, w_g \rangle = \langle w'_f, w'_g \rangle = d \langle f, g \rangle$ and $\langle w_f, w'_g \rangle = \langle f, Ag \rangle$ for $f, g \in \ell^2(V)$.

In particular, the eigenfunctions $(f_i)_{i=1}^n$ correspond in this way to pairwise orthogonal eigenspaces of C with eigenvalues $(\lambda_i)_{i=1}^n$; the dimension of each eigenspace is 1 if $\lambda_i = \pm d$ and 2 otherwise (as before, w_f can be a multiple of w'_f only if $w \equiv c$ or when G is bipartite and $w \equiv c$ on one part and $w \equiv -c$ on the other), and they notably include the eigenfunctions $T_{\theta_i} f_i$ of B .

Of course, every such 2-dimensional eigenspace corresponding to $\lambda_i \neq \pm d$ is orthogonal to the eigenvectors of B from (3.5)–(3.6) (corresponding to the eigenvalues ± 1), as those are also eigenvectors of $\pm d$ for the self-adjoint C . Finally, the eigenvector $w \equiv 1$ with the eigenvalue $d-1$ of B (and eigenvalue d of C) is orthogonal to $\ell_-^2(\vec{E})$ (thus to all eigenvectors from (3.5)), whereas if G is bipartite and we take $w \equiv 1$ on outgoing edges from a prescribed part of G and $w \equiv -1$ on the incoming ones (with eigenvalue $-(d-1)$ of B) then $w \perp \ell_+^2(\vec{E})$, thus it is orthogonal to all eigenvectors from (3.6).

Suppose for now that A has no eigenvalue λ_i such that $|\lambda_i| = 2\sqrt{d-1}$. Then there are two distinct solutions to (3.3) for each of the λ_i 's, and so, in particular, the eigenspace of C corresponding to $\lambda_i \neq \pm d$ has two linearly independent eigenvectors of B —corresponding to eigenvalues θ_i and θ'_i . The orthogonality of the eigenspaces from the discussion above now establishes the form of Λ from (3.2).

When there exist eigenvalues of A such that $|\lambda_i| = 2\sqrt{d-1}$, we have the unique solution $\theta_i = \lambda_i/2$ for (3.3), and claim that this gives rise to a Jordan block $\begin{pmatrix} \lambda_i/2 & 1 \\ 0 & \lambda_i/2 \end{pmatrix}$. Indeed, recalling that $BT_\theta f_i = \theta_i f_i$, observe that

$$\begin{aligned} (BT_{1+\theta_i} f_i)(x, y) &= [(1 + \theta_i)\lambda_i - (d-1)]f_i(y) - (1 + \theta_i)f_i(x) \\ &= \theta_i[(1 + \theta_i)f_i(y) - f_i(x)] + [\theta_i^2 + \theta_i - (d-1)]f(y) - f(x) \\ &= \theta_i(T_{1+\theta_i} f_i)(x, y) + (T_{\theta_i} f_i)(x, y), \end{aligned} \quad (3.7)$$

where the second equality used $\theta_i = \lambda_i/2$ and the last one used $\theta_i^2 = d-1$. As these both belong to the corresponding eigenspace of C , we arrive at (3.2).

To conclude the proof, it remains to show that $|\alpha_i| < 2(d-1)$ if $\lambda_i \neq \pm d$. Recall that there exist unit vectors w_i, w'_i such that $Bw'_i = \alpha_i w_i + \theta'_i w'_i$ (these can be taken as columns $2i$ and $2i+1$ of U as above). Hence,

$$(B - \theta'_i I)w'_i = \alpha w_i. \quad (3.8)$$

Let $\|\cdot\|_{2 \rightarrow 2}$ be the $\ell^2(\vec{E}) \rightarrow \ell^2(\vec{E})$ operator norm; we claim $\|B\|_{2 \rightarrow 2} = d-1$. Indeed, it is easy to verify that

$$(BB^*)_{(u,v),(x,y)} = \begin{cases} d-1 & x = u \text{ and } y = v, \\ d-2 & x \neq u \text{ and } y = v, \\ 0 & \text{otherwise.} \end{cases}$$

We see that BB^* has $\|BB^*\|_{\infty \rightarrow \infty} = (d-1)^2$ and an eigenvalue $(d-1)^2$ corresponding to the eigenvector $w \equiv 1$; thus, $\|B\|_{2 \rightarrow 2} = d-1$. By (3.8), using $|\theta'_i| < d-1$ and $\|w_i\| = \|w'_i\| = 1$, and we infer that $|\alpha| < 2(d-1)$, concluding the proof of the proposition. \blacksquare

3.2. Cutoff on non-bipartite Ramanujan graphs. On every d -regular graph on n vertices, the number of directed edges at distance ℓ from a given $(x, y) \in \vec{E}$ is at most $(d-1)^\ell$; this readily implies (as stated in [22, Claim 4.8]) that the nonbacktracking random walk satisfies

$$t_{\text{MIX}}(1 - \varepsilon) \geq \lceil \log_{d-1}(dn) \rceil - \lceil \log_{d-1}(1/\varepsilon) \rceil \quad \text{for any } 0 < \varepsilon < 1. \quad (3.9)$$

Our goal in this section is to show an asymptotically tight upper bound on t_{MIX} using the spectral decomposition of the nonbacktracking operator B .

Theorem 3.5. *Let G be a non-bipartite Ramanujan graph on n vertices with degree $d \geq 3$. Let μ_t be the t -step transition kernel of the NBRW, and let π be the uniform distribution on \vec{E} . Then for some fixed $c(d) > 0$,*

$$\max_{(x,y) \in \vec{E}} \left\| \frac{\mu_t((x,y), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq \frac{c(d)}{\log n} \quad \text{at } t = \lceil \log_{d-1} n + 3 \log_{d-1} \log n \rceil.$$

Consequently, on any sequence of such graphs, the NBRW exhibits L^1 -cutoff and L^2 -cutoff both at time $\log_{d-1} n$.

Remark 3.6. The constant $c(d)$ in the above theorem can be taken to be $8(d-1)\log^{-2}(d-1)+1$ for any sufficiently large enough n (cf. (3.13) below).

Proof of Theorem 3.5. Appealing to Proposition 3.1, let U be the unitary matrix such that $B = U\Lambda U^*$ with Λ from (3.2), and write

$$U = (w_1 \mid w_2 \mid w'_2 \mid w_3 \mid w'_3 \mid \dots \mid w_n \mid w'_n \mid u_1 \mid \dots \mid u_{N-(2n-1)}) ,$$

in which $w_1 \equiv N^{-1/2}$. Recalling Remark 3.3, observe that the assumption that G is non-bipartite Ramanujan implies that for all $i = 2, \dots, n$, the solutions θ_i, θ'_i to (3.3) satisfy $\theta'_i = \bar{\theta}_i$ and $|\theta_i| = \sqrt{d-1}$.

Let $(x_0, y_0) \in \vec{E}$ be some initial edge for the NBRW; by the expansion (2.6) of the L^2 -distance, the t -step transition kernel $\mu_t = (d-1)^{-t}B^t$ satisfies

$$\begin{aligned} \left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 &= N \sum_{(x,y)} |\mu_t((x_0, y_0), (x, y))|^2 - 1 \\ &= N \left\| \mu_t((x_0, y_0), \cdot) \right\|^2 - 1. \end{aligned} \quad (3.10)$$

Using $B = U\Lambda U^*$ with Λ from (3.2) and U as specified above we find that

$$\begin{aligned} B^t((x_0, y_0), \cdot) &= (d-1)^t w_1(x_0, y_0) w_1 + \sum_i (\pm 1)^t u_i(x_0, y_0) u_i \\ &\quad + \sum_{i=2}^n \theta_i^t w_i(x_0, y_0) \bar{w}_i + (\bar{\theta}_i^t w'_i(x_0, y_0) + \gamma_i(t) w_i(x_0, y_0)) \bar{w}'_i, \end{aligned}$$

where

$$\gamma_i(t) := \alpha_i \sum_{j=0}^{t-1} \theta_i^j \bar{\theta}_i^{t-1-j}$$

with α_i from Proposition 3.1. Note that in particular, as $\alpha_i < 2(d-1)$,

$$|\gamma_i(t)| \leq 2(d-1)t|\theta_i|^{t-1}. \quad (3.11)$$

From the above expansion of B^t , since U is unitary and $w_1 \equiv N^{-1/2}$,

$$\begin{aligned} \left\| \mu_t((x_0, y_0), \cdot) \right\|^2 &= \frac{1}{N} + \sum_i (d-1)^{-2t} |u_i(x_0, y_0)|^2 \\ &\quad + (d-1)^{-2t} \sum_{i=2}^n \left(|\theta_i|^{2t} |w_i(x_0, y_0)|^2 + |\bar{\theta}_i^t w'_i(x_0, y_0) + \gamma_i(t) w_i(x_0, y_0)|^2 \right). \end{aligned} \quad (3.12)$$

Now we exploit the fact that G is Ramanujan: since $|\theta_i| = \sqrt{d-1}$ for every $2 \leq i \leq n$, the expression in the second line of (3.12) is at most

$$(d-1)^{-t} \sum_{i=2}^n |w_i(x_0, y_0)|^2 + 2|w'_i(x_0, y_0)|^2 + 2 \frac{|\gamma_i(t)|^2}{(d-1)^t} |w_i(x_0, y_0)|^2 ,$$

using the parallelogram law. Since by Parseval's identity,

$$\sum_i |u_i(x_0, y_0)|^2 + \sum_i |w_i(x_0, y_0)|^2 + |w'_i(x_0, y_0)|^2 = \|\delta_{(x_0, y_0)}\|^2 = 1,$$

and with (3.10) in mind, we infer that

$$\left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq 2N(d-1)^{-t} \left(1 + \max_i \frac{|\gamma_i(t)|^2}{(d-1)^t} \right).$$

Substituting the bound (3.11) on $\gamma_i(t)$, again using that G is Ramanujan,

$$\left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq 2N(d-1)^{-t} (4(d-1)t^2 + 1). \quad (3.13)$$

In particular, for $t = \lceil \log_{d-1} n + 3 \log_{d-1} \log n \rceil$,

$$\left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq O(1/\log n) = o(1),$$

thus concluding the proof of Theorem 3.5. \blacksquare

Using the reduction in §2.1 from SRW to NBRW (see (2.3)–(2.4)), one can deduce Theorem 1 from Theorem 3.5, as the $O(\log \log n)$ window for the NBRW is negligible compared with the term $s\sqrt{\log_{d-1} n}$ in (1.1).

Note that for every integer $\ell \geq (2+o(1)) \log_{d-1} n$ there is a path of length exactly ℓ between every pair of vertices x, y using $D_\infty(2t) \leq D_2(t)D_2^*(t)$ for the NBRW (recall (2.15), and that the chain and its reversal are isomorphic).

Proof of Corollary 2. Since $\max_{(x,y)} \|\mu_t((x,y), \cdot) - \pi\|_{\text{TV}} = o(1)$ at time t as per Theorem 3.5, for every x , all but $o(n)$ directed edges can be reached by a nonbacktracking path of length t from x . The remark above (3.9) on the growth of balls in a d -regular graph thus implies the corollary: the statement on a nonbacktracking cycle follows from applying this argument once on a directed edge originating from x (and reaching almost every y within the proper length bound) and once on an arbitrarily chosen other directed edge ending at x , in the reversed NBRW. \blacksquare

Proof of Corollary 3. Note that at time R , the L^2 -distance of the NBRW from equilibrium is $O(1/\log^{3/2} n)$ by (3.13), and that $k = O(\log n)$ since $k \leq g$. For a uniformly chosen path $(y_i)_{i=1}^k$ in G , each y_i is uniform by the stationarity of the NBRW. Thus, by a union bound over the vertices y_i , for each i there exists a path of length R from the edge (x_i, z_i) to (y_i, z'_i) , except with probability $O(k/\log^{3/2} n) = o(1)$, where z_i and z'_i are not on the paths (x_i) and (y_i) , respectively. The conclusion now follows since, if vertex ℓ of the path from x_i coincides with vertex ℓ' of the path from x_j , then $\ell + \ell' + k > g$ and $(R - \ell) + (R - \ell') + k > g$, so $k > g - R$, a contradiction. \blacksquare

Remark 3.7. In the setting of Theorem 3.5, if G is in addition transitive then, by using the exact value $|\alpha_i| = d - 2$ from Proposition 4.1 below, the L^2 -mixing time of the NBRW can be pinpointed precisely: let

$$\Upsilon_G(k) := (d - 2)^2(d - 1)^{-1} \int U_{k-1}(x)^2 d\mu_G$$

for $\mu_G = \frac{1}{n} \sum_i \delta_{\lambda_i/(2\sqrt{d-1})}$ the empirical spectral distribution (ESD) of G and $U_k(\cos(x)) = \frac{\sin((k-1)x)}{\sin x}$ the Chebyshev polynomial of the second kind. Then for any fixed $\varepsilon > 0$,

$$t_{\text{MIX}}^{(L^2)}(\varepsilon) = \lceil \log_{d-1}(n) + \log_{d-1}(\Upsilon_G(\log_{d-1} n) + 2) + \log_{d-1}(1/\varepsilon) \rceil. \quad (3.14)$$

Indeed, from (3.12) we see that for any non-bipartite Ramanujan graph G (not necessarily transitive), averaging over the initial state (x_0, y_0) gives

$$\frac{1}{N} \sum_{(x_0, y_0)} \|\mu_t((x_0, y_0), \cdot)\|^2 = \frac{1}{N} + \frac{N - 2n + 1}{(d - 1)^{2t}} + \frac{2n - 2}{(d - 1)^t} + \frac{\sum_i |\gamma_i(t)|^2}{(d - 1)^t},$$

using that $w_i \perp w'_i$ and $\|u_i\| = \|w_i\| = \|w'_i\| = 1$ for all i . Thus, by (3.10),

$$\frac{1}{N} \sum_{(x_0, y_0)} \left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 = (1 + o(1)) \left(\sum_i |\gamma_i(t)|^2 + 2 \right) \frac{n}{(d - 1)^t},$$

provided that $t \rightarrow \infty$ with n . Writing $\varphi_i = \lambda_i/(2\sqrt{d-1})$ (so $\theta_i = \cos \varphi_i$ for $i = 2, \dots, n$) and using Proposition 4.1,

$$|\gamma_i(t)| = (d - 2) \left| \frac{\bar{\theta}_i^t - \theta_i^t}{\theta_i - \theta_i} \right| = (d - 2)(d - 1)^{(t-1)/2} \left| \frac{\sin(t\varphi_i)}{\sin \varphi_i} \right|,$$

which implies the analogue of (3.14) for the average of the mixing times over the initial states (x_0, y_0) , thus establishing (3.14) for the transitive case.

3.3. Extensions. We conclude with corollaries of the proof of Theorem 3.5.

3.3.1. Bipartite Ramanujan graphs. Following is the analog for NBRW in the bipartite case; its SRW counterpart follows from the cover-tree reduction.

Corollary 3.8. *Let $G = (V_0, V_1, E)$ be a bipartite Ramanujan graph on n vertices with degree $d \geq 3$. Let μ_t be the t -step transition kernel of the NBRW, and let π_0 and π_1 be the uniform distribution on the $N/2$ directed edges originating from V_0 and V_1 , respectively. Then for some fixed $c(d) > 0$,*

$$\max_{\substack{(x_0, y_0) \in \vec{E} \\ x_0 \in V_0}} \left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi_{(t \bmod 2)}} - 1 \right\|_{L^2(\pi_{(t \bmod 2)})}^2 \leq \frac{c(d)}{\log n}$$

at time

$$t = \lceil \log_{d-1} n + 3 \log_{d-1} \log n \rceil.$$

Consequently, on any sequence of such graphs, the NBRW that is modified to be lazy in its first step exhibits L^1 -cutoff and L^2 -cutoff at time $\log_{d-1} n$.

Proof. Following the arguments used to prove Theorem 3.5, observe that in computing $\mathbb{E} \left[\left| \frac{\mu_t((x_0, y_0), (x, y))}{\pi_{(t \bmod 2)}} - 1 \right|^2 \right]$, the identity (3.10) becomes valid once we replace N by $N/2$. The only other modification needed is to treat $\lambda_n = -d$, which produces the eigenvalue $\theta_n = -(d-1)$. Since all the coordinates of w_n are $\pm N^{-1/2}$, the contribution of this eigenvalue to the right-hand of (3.12) is $1/N$, exactly that of the eigenvalue $d-1$ of B . The combined $2/N$ cancels via the modified identity (3.10), thus (3.13) becomes

$$\left\| \frac{\mu_t((x_0, y_0), \cdot)}{\pi_{(t \bmod 2)}} - 1 \right\|_{L^2(\pi_{(t \bmod 2)})}^2 \leq N(d-1)^{-t} (4(d-1)t^2 + 1),$$

which is $O(1/\log n)$ at the same value of t . \blacksquare

Corollary 3.9. *Let $G = (V_0, V_1, E)$ be a bipartite Ramanujan graph on n vertices with degree $d \geq 3$. Let P^t be the t -step transition kernel of the SRW, and let π_0 and π_1 be the uniform distribution on V_0 and V_1 , respectively. Let T be the first step of the SRW. Then for every fixed $s \in \mathbb{R}$ and every initial vertex x , the SRW at time*

$$t = \frac{d}{d-2} \log_{d-1} n + s \sqrt{\log_{d-1} n}.$$

satisfies

$$\max_{x_0 \in V_0} \left\| P^t(x_0, \cdot) - \pi_{(t \bmod 2)} \right\|_{\text{TV}} \rightarrow \mathbb{P}(Z > c_d s) \quad \text{as } n \rightarrow \infty,$$

where Z is a standard normal random variable and $c_d = \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}}$.

Consequently, on any sequence of such graphs, the SRW that is modified to be lazy in its first step exhibits L^1 -cutoff and L^2 -cutoff at time $\frac{d}{d-2} \log_{d-1} n$.

3.3.2. Weakly Ramanujan graphs. It suffices to establish the result for the NBRW (here we do not specify D_{TV} for the SRW within the cutoff window, thus there is no need to control the NBRW within a window of $o(\sqrt{\log n})$), and Theorem 4 and Corollary 5 will then follow using the above reduction.

Corollary 3.10. *Fix $d \geq 3$ and let G be a d -regular graph on n vertices whose nontrivial eigenvalues $\{\lambda_i\}_{i=2}^n$ all satisfy $|\lambda_i| \leq (1 + \delta_n)2\sqrt{d-1}$ for some δ_n going to 0 as $n \rightarrow \infty$. Let μ_t be the t -step transition kernel of the NBRW, and let π be the uniform distribution on \vec{E} . For some fixed $c(d) > 0$,*

$$\max_{(x,y) \in \vec{E}} \left\| \frac{\mu_t((x,y), \cdot)}{\pi} - 1 \right\|_{L^2(\pi)}^2 \leq \frac{c(d)}{\log n}$$

at time

$$t = \left\lceil (1 + 5\sqrt{\delta_n}) \log_{d-1} n + 3 \log_{d-1} \log n \right\rceil.$$

Consequently, on any sequence of such graphs, the NBRW exhibits L^1 -cutoff and L^2 -cutoff both at time $\log_{d-1} n$.

Proof. The analysis of blocks of Λ corresponding to eigenvalues λ_i ($i \geq 2$) of A such that $|\lambda_i| \leq 2\sqrt{d-1}$ remains valid unchanged, and it remains to consider the effect of

$$|\lambda_i| = (1 + \varepsilon)2\sqrt{d-1} \quad \text{for some } 0 < \varepsilon \leq \delta_n. \quad (3.15)$$

As mentioned in the proof of Theorem 3.5, the fact that G is Ramanujan is exploited when replacing $|\theta_i|^2 t$ by $(d-1)^t$ for all $i \geq 2$ in the spectral decomposition (3.12), and once again (just above (3.13)) in the bound (3.11) on $\gamma_i(t)$. For λ_i as in (3.15), the corresponding real eigenvalues θ_i, θ'_i of B are given, as per (3.3), by

$$\left(1 + \varepsilon \pm \sqrt{\varepsilon(2 + \varepsilon)}\right) \sqrt{d-1};$$

in particular, denoting $|\theta_i| > |\theta'_i|$, we have $|\theta_i| = (1 + \sqrt{2\varepsilon} + O(\varepsilon))\sqrt{d-1}$ (while at the same time $|\theta'_i| < \sqrt{d-1}$). We account for this modified value of $|\theta_i|^{2t}$ in the spectral decomposition of $\|\mu_t((x, y), \cdot)\|^2$ via the pre-factor

$$\left(1 + \sqrt{2\varepsilon} + O(\varepsilon)\right)^{2t} \leq \exp\left[(2\sqrt{2\delta_n} + O(\delta_n))t\right],$$

thus replacing the right-hand of (3.13) by

$$2N(d-1)^{-t} e^{[2\sqrt{2\delta_n} + O(\delta_n)]t} (4(d-1)t^2 + 1).$$

For the designated value of t (in which there is an extra additive term of $5\sqrt{\delta_n} \log_{d-1} n$ compared to t from Theorem 3.5) and using that $\delta_n \rightarrow 0$, we find that there exists some fixed $c(d) > 0$ such that $\|\mu_t((x, y), \cdot)/\pi - 1\|_{L^2(\pi)}^2$ is at most

$$\frac{c(d) + o(1)}{\log n} \exp\left[(2\sqrt{2} - 5 \log(d-1) + o(1))\sqrt{\delta_n} t\right],$$

which is $O(1/\log n)$ since $2\sqrt{2} < 5 \log(d-1)$ for all $d \geq 3$. \blacksquare

Remark 3.11. Suppose that, for some $\delta_n = o(1)$ and fixed $\varepsilon' > 0$, the graph G has $|\lambda| \leq 2\sqrt{d-1} + \delta_n$ for all eigenvalues λ except for $n^{o(1)}$ exceptional ones, which instead satisfy $|\lambda| < d - \varepsilon'$. Each eigenvalue of the latter form corresponds to an additive term of $O(a^{2t})$ in the right-hand of (3.13), where $0 < a < 1$ depends only on d and ε' . For the prescribed t from Corollary 3.10, this amounts to $O(n^{-\varepsilon''})$ for some fixed $\varepsilon'' > 0$, thus the overall contribution of these $n^{o(1)}$ exceptional eigenvalues is negligible and the same result holds.

4. PINPOINTING THE SPECTRAL DECOMPOSITION

The following proposition gives the precise moduli of the off-diagonal terms in Λ from the spectral decomposition (3.2) in Proposition 3.1.

Proposition 4.1. *In the setting of Proposition 3.1, for all $i \geq 2$ we have $\alpha_i = 0$ if $\lambda_i = -d$ (and $i = n$), and otherwise*

$$|\alpha_i| = \begin{cases} d-2 & \text{if } |\lambda_i| \leq 2\sqrt{d-1}, \\ \sqrt{d^2 - \lambda_i^2} & \text{if } |\lambda_i| > 2\sqrt{d-1}. \end{cases}$$

Proof. Let $i \geq 2$, and for simplicity, omit its indices from the corresponding subscripts; namely, let θ, θ' correspond to the eigenvalue $\lambda \neq \pm d$, and let f be so that $Af = \lambda f$ and $\|f\| = 1$, where A is the adjacency matrix of G .

Case (1): $|\lambda| \neq 2\sqrt{d-1}$: Recalling T_θ from (3.4), we claim that

$$\alpha = \frac{\beta(\theta' - \theta)}{\sqrt{1 - |\beta|^2}} \quad \text{where} \quad \beta := \frac{\langle T_{\theta'} f, T_\theta f \rangle}{\|T_{\theta'} f\| \|T_\theta f\|}. \quad (4.1)$$

Indeed, taking

$$w = \frac{T_\theta f}{\|T_\theta f\|}, \quad w' = \frac{T_{\theta'} f}{\|T_{\theta'} f\|}, \quad w'' = \frac{w' - \beta w}{\|w' - \beta w\|}$$

for β as above gives $Bw = \theta w$, $Bw' = \theta' w$, and

$$\|w' - \beta w\|^2 = 1 + |\beta|^2 - \beta \langle w, w' \rangle - \bar{\beta} \langle w', w \rangle = 1 - |\beta|^2, \quad (4.2)$$

so $w'' = (1 - |\beta|^2)^{-1/2}(w' - \beta w)$ satisfies

$$Bw'' = \frac{\theta' w' - \beta \theta w}{\sqrt{1 - |\beta|^2}} = \theta' w'' + \frac{\beta(\theta' - \theta)}{\sqrt{1 - |\beta|^2}} = \theta' w'' + \alpha w,$$

as claimed. To estimate α , observe that for every $f \in \ell^2(V)$,

$$\|T_\theta f\|_{\ell^2(\bar{E})}^2 = d(|\theta|^2 + 1)\|f\|_{\ell^2(V)}^2 - (\theta + \bar{\theta}) \langle Af, f \rangle_{\ell^2(V)}. \quad (4.3)$$

Case (1.a): below the Ramanujan threshold. When $|\lambda| < 2\sqrt{d-1}$ we have $\theta' = \bar{\theta} \in \mathbb{C} \setminus \mathbb{R}$. Since $\theta^2 + d - 1 = \lambda\theta$ and $|\theta| = \sqrt{d-1}$,

$$\begin{aligned} \|T_\theta f\|^2 &= d^2 - (\theta + \bar{\theta})\lambda = d^2 - 2(d-1) - (\theta^2 + \bar{\theta}^2) \\ &= d^2 - 2(d-1)[1 + \cos(2\varphi)] = (d-2)^2 + 2(d-1)(1 - \cos(2\varphi)), \end{aligned}$$

where we let $\theta = \sqrt{d-1} \exp(i\varphi)$. Similarly,

$$\beta = \frac{d(\bar{\theta}^2 + 1) - 2\bar{\theta}\lambda}{(d-2)^2 + 2(d-1)(1 - \cos(2\varphi))} = \frac{(d-2)(\bar{\theta}^2 - 1)}{(d-2)^2 + 2(d-1)(1 - \cos(2\varphi))},$$

and so

$$\begin{aligned} 1 - |\beta|^2 &= 1 - \frac{(d-1)^2 + 1 - 2(d-1)\cos(2\varphi)}{\left[d - 2 + 2\frac{d-1}{d-2}(1 - \cos(2\varphi))\right]^2} \\ &= \frac{4\left(\frac{d-1}{d-2}\right)^2(1 - \cos(2\varphi))^2 + 2(d-1)(1 - \cos(2\varphi))}{\left[d - 2 + 2\frac{d-1}{d-2}(1 - \cos(2\varphi))\right]^2}. \end{aligned}$$

Substituting $\cos(2\varphi) = 1 - 2\sin^2\varphi$ we see that

$$1 - |\beta|^2 = \frac{4(d-1) \left(1 + 4\frac{d-1}{(d-2)^2} \sin^2\varphi\right) \sin^2\varphi}{\left(d-2 + 4\frac{d-1}{d-2} \sin^2\varphi\right)^2} = \frac{4(d-1) \sin^2\varphi}{(d-2)^2 + 4(d-1) \sin^2\varphi},$$

and so

$$\frac{|\beta|^2}{1 - |\beta|^2} = \frac{1}{1 - |\beta|^2} - 1 = \frac{(d-2)^2}{4(d-1) \sin^2\varphi}.$$

Since $\theta - \theta' = 2\sqrt{d-1} \sin\varphi$, we conclude from (4.1) that $|\alpha| = d-2$.

Case (1.b): above the Ramanujan threshold. For $2\sqrt{d-1} < |\lambda| < d$, we have $\theta \neq \theta' \in \mathbb{R}$, and assume w.l.o.g. that $\theta > \theta'$. By (4.3) we get

$$\|T_\theta f\|^2 = d(\theta^2 + 1) - 2\theta\lambda = d(\theta^2 + 1) - 2(\theta^2 + d - 1) = (d-2)(\theta^2 - 1),$$

and for the same reason, $\|T_{\theta'} f\|^2 = (d-2)(\theta'^2 - 1)$. Similarly,

$$\langle T_\theta f, T_{\theta'} f \rangle = d(\theta\theta' + 1) - (\theta + \theta')\lambda = d^2 - \lambda^2,$$

using that $\theta\theta' = d-1$ whereas $\theta + \theta' = \lambda$ through their definition in (3.3). Since we also have $\theta^2 + \theta'^2 = \lambda^2 - 2(d-1)$, we see that

$$(\theta^2 - 1)(\theta'^2 - 1) = (d-1)^2 - (\lambda^2 - 2(d-1)) + 1 = d^2 - \lambda^2,$$

and altogether deduce that

$$\beta = \frac{d^2 - \lambda^2}{[(d-2)(\theta^2 - 1)]^{\frac{1}{2}} [(d-2)(\theta'^2 - 1)]^{\frac{1}{2}}} = \frac{\sqrt{d^2 - \lambda^2}}{d-2}.$$

Therefore,

$$\frac{\beta^2}{1 - \beta^2} = \frac{1}{1 - \beta^2} - 1 = \frac{d^2 - \lambda^2}{(d-2)^2 - (d^2 - \lambda^2)} = \frac{d^2 - \lambda^2}{\lambda^2 - 4(d-1)},$$

Recalling the definition (4.1) of α , and using that $\theta - \theta' = \sqrt{\lambda^2 - 4(d-1)}$, we infer that $\alpha^2 = d^2 - \lambda^2$.

Case (2): at the Ramanujan threshold: For $|\lambda| = 2\sqrt{d-1}$ we claim

$$\alpha = \frac{\|T_\theta f\|}{\|T_{1+\theta} f\| \sqrt{1 - |\beta|^2}} \quad \text{where} \quad \beta := \frac{\langle T_{1+\theta} f, T_\theta f \rangle}{\|T_{1+\theta} f\| \|T_\theta f\|}. \quad (4.4)$$

To see this, take

$$w = \frac{T_\theta f}{\|T_\theta f\|}, \quad w' = \frac{T_{1+\theta} f}{\|T_{1+\theta} f\|}, \quad w'' = \frac{w' - \beta w}{\|w' - \beta w\|};$$

since $Bw' = \theta w' + (\|T_\theta f\|/\|T_{1+\theta} f\|)w$ by (3.7), while $\|w' - \beta w\|^2 = 1 - |\beta|^2$ (by the same calculation as in (4.2)),

$$Bw'' = \frac{\theta w' + \|T_{1+\theta} f\|^{-1} w - \beta \theta w}{\sqrt{1 - |\beta|^2}} = \theta w'' + \alpha w,$$

as claimed. To compute α , we recall that $\theta = \lambda/2$, and infer from (4.3) that

$$\|T_\theta f\|^2 = d(\theta^2 + 1) - 2\theta\lambda = d^2 - 2\theta\lambda = (d-2)^2, \quad (4.5)$$

$$\|T_{1+\theta} f\|^2 = d((1+\theta)^2 + 1) - 2(1+\theta)\lambda = (d-2)(d+2\theta-2) + d,$$

as well as that

$$\langle T_{1+\theta} f, T_\theta f \rangle = d(\theta(1+\theta) + 1) - (2\theta+1)\lambda = (d-2)(d+\theta-2).$$

We therefore have

$$\begin{aligned} 1 - \beta^2 &= 1 - \frac{(d+\theta-2)^2}{(d-2)(d+2\theta-2) + d} = \frac{(d+\theta-2)(-\theta) + \theta(d-2) + d}{(d-2)(d+2\theta-2) + d} \\ &= \frac{d - \theta^2}{(d-2)(d+2\theta-2) + d} = \frac{1}{\|T_{1+\theta} f\|^2} \end{aligned}$$

and so, by (4.4)–(4.5), $\alpha = \|T_\theta f\| = d-2$. \blacksquare

Acknowledgements. We thank Shayan Oveis Gharan for suggesting that we study cutoff on Ramanujan graphs, and Perla Sousi for comments on an earlier version of this manuscript. The research of E.L. was supported in part by NSF grant DMS-1513403.

REFERENCES

- [1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.
- [2] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.
- [3] D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs, 2002. Available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [4] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [5] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin. Non-backtracking random walks mix faster. *Commun. Contemp. Math.*, 9(4):585–603, 2007.
- [6] N. Alon and V. D. Milman. λ_1 , isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.
- [7] O. Angel, J. Friedman, and S. Hoory. The non-backtracking spectrum of the universal cover of a graph. *Trans. Amer. Math. Soc.*, 367(6):4287–4318, 2015.
- [8] H. Bass. The Ihara-Selberg zeta function of a tree lattice. *Internat. J. Math.*, 3(6):717–797, 1992.
- [9] C. Bordenave. A new proof of Friedman’s second eigenvalue Theorem and its extension to random lifts. 2015. Preprint, available at [arXiv:1502.04482](https://arxiv.org/abs/1502.04482).
- [10] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, 13:no. 3, 26–78, 2008.
- [11] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–196, 1989.
- [12] F. R. K. Chung, V. Faber, and T. A. Manteuffel. An upper bound on the diameter of a graph from eigenvalues associated with its Laplacian. *SIAM J. Discrete Math.*, 7(3):443–457, 1994.

- [13] G. Davidoff, P. Sarnak, and A. Valette. *Elementary number theory, group theory, and Ramanujan graphs*, volume 55 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2003.
- [14] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
- [15] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.
- [16] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [17] J. Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.
- [18] J. Friedman and D. Kohler. The relativized second eigenvalue conjecture of Alon. 2014. Preprint, available at [arXiv:1403.3462](https://arxiv.org/abs/1403.3462).
- [19] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.
- [20] M. Kotani and T. Sunada. Zeta functions of finite graphs. *J. Math. Sci. Univ. Tokyo*, 7(1):7–25, 2000.
- [21] S. P. Lalley. Finite range random walk on free groups and homogeneous trees. *Ann. Probab.*, 21(4):2087–2130, 1993.
- [22] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.*, 153(3):475–510, 2010.
- [23] E. Lubetzky and A. Sly. Explicit expanders with cutoff phenomena. *Electron. J. Probab.*, 16:no. 15, 419–435, 2011.
- [24] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures*. Modern Birkhäuser Classics. Birkhäuser Verlag, Basel, 2010.
- [25] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [26] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press. In preparation. Current version available at <http://pages.iu.edu/~rdlyons/>.
- [27] A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families I: bipartite Ramanujan graphs of all degrees. *Ann. of Math.*, 182(1):307–325, 2015.
- [28] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.
- [29] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.
- [30] Y. Peres. American Institute of Mathematics (AIM) research workshop “Sharp Thresholds for Mixing Times”, Palo Alto, December 2004. Summary available at <http://www.aimath.org/WWN/mixingtimes>.
- [31] N. T. Sardari. Diameter of Ramanujan graphs and random Cayley graphs with numerics. 2015. Preprint, available at [arXiv:1511.09340](https://arxiv.org/abs/1511.09340).
- [32] P. Sarnak. Letter to Scott Aaronson and Andrew Pollington on the Solovay–Kitaev Theorem and Golden Gates (with an appendix on optimal lifting of integral points). February 2015. Available at <http://publications.ias.edu/sarnak/paper/2637>.
- [33] J.-P. Serre. Répartition asymptotique des valeurs propres de l’opérateur de Hecke T_p . *J. Amer. Math. Soc.*, 10(1):75–102, 1997.
- [34] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces*. Princeton University Press, Princeton, N.J., 1971.

- [35] W. Woess. *Denumerable Markov chains*. European Mathematical Society (EMS), Zürich, 2009. Generating functions, boundary theory, random walks on trees.

E. LUBETZKY

COURANT INSTITUTE, NEW YORK UNIVERSITY, 251 MERCER ST., NEW YORK, NY 10012.

E-mail address: `eyal@courant.nyu.edu`

Y. PERES

MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052.

E-mail address: `peres@microsoft.com`