Non-backtracking random walks mix faster

Noga Alon* Itai Benjamini † Eyal Lubetzky ‡ Sasha Sodin §

Abstract

We compute the mixing rate of a non-backtracking random walk on a regular expander. Using some properties of Chebyshev polynomials of the second kind, we show that this rate may be up to twice as fast as the mixing rate of the simple random walk. The closer the expander is to a Ramanujan graph, the higher the ratio between the above two mixing rates is.

As an application, we show that if G is a high-girth regular expander on n vertices, then a typical non-backtracking random walk of length n on G does not visit a vertex more than $(1+o(1))\frac{\log n}{\log\log n}$ times, and this result is tight. In this sense, the multi-set of visited vertices is analogous to the result of throwing n balls to n bins uniformly, in contrast to the simple random walk on G, which almost surely visits some vertex $\Omega(\log n)$ times.

1 Introduction

1.1 Background and definitions

Let $G = (V, E)$ be an undirected graph. A random walk of length k on G, from some given vertex $w_0 \in V$, is a uniformly chosen member of:

$$\mathcal{W}^{(k)} = \{(w_0, w_1, \ldots, w_k) \;:\; w_t \in V,\; w_{t-1}w_t \in E \;\text{ for all } t \in [k]\} .$$

Equivalently, such a walk is a finite Markov chain $\mathcal{M} = (X_0, \ldots, X_k)$ on the state space V, where $X_0 = w_0$ and the transition probabilities are $P_{uv} = \Pr[X_i = v \mid X_{i-1} = u] = \mathbf{1}_{\{uv\in E\}}/\deg(u)$. For further information on Markov chains, see, e.g, [10], [18].

The extensive study of random walks on graphs was motivated by the following useful property, which we first state informally. While the random walk is simple to analyze and to implement in

*School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by a USA-Israeli BSF grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†Weizmann Institute, Rehovot, 76100, Israel. Email: itai.benjamini@weizmann.ac.il

‡School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: lubetzky@tau.ac.il. Research partially supported by a Charles Clore Foundation Fellowship.

§School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: sodinale@tau.ac.il.

many frameworks, it "mixes" in G after a relatively small number of steps, provided G satisfies some natural requirements. Thus, the random walk provides an efficient method of sampling the graph vertices, a fact which has many applications in Theoretical and Applied Computer Science. See [13] for a survey on the subject.

The following facts are well known (see, e.g., [13], [14], [19]). If G is a connected and non-bipartite undirected graph, then the Markov chain $\mathcal{M}$, corresponding to the random walk on G, is irreducible and aperiodic. In this case, $\mathcal{M}$ converges to a unique stationary distribution, π, regardless of its starting position, where $\pi(u) = \frac{\deg(u)}{2|E|}$. The *mixing rate* of the random walk on G measures how fast $\mathcal{M}$ converges to the stationary distribution, and is defined as follows:

$$\rho = \rho(G) = \limsup_{k\to\infty} \max_{u,v\in V} \left| P_{uv}^{(k)} - \pi(v) \right|^{1/k} , \tag{1}$$

where $P_{uv}^{(k)} = \Pr[X_{t+k} = v \mid X_t = u]$. The notion of *mixing time*, the number of steps it takes $\mathcal{M}$ to get "sufficiently close" to π, has several commonly used definitions, and for each of these definitions there are lower and upper bounds as a function of ρ and n. For instance, letting $P_u^{(k)}$ denote the distribution of $\mathcal{M}$ at time k given that $X_0 = u$, one may define the mixing time τ_ε as the minimal number of steps it takes $P_u^{(k)}$ and π to be at most ε-far in terms of their total variation distance, maximized over all vertices $u \in V$.

An important special case of the above is the one where the graph G is regular. In this case, the stationary distribution π is the uniform distribution, being an eigenvector of the transition probabilities matrix $P = A/d$ (where A is the adjacency matrix of the graph and d is its regularity degree). Hence, whenever G is connected and non-bipartite, the random walk eventually approximates the uniform distribution. As we next specify, these sufficient and necessary conditions, required for the random walk on G to mix, are determined by the spectrum of G.

Let G be a d-regular graph. The eigenvalues of G, that is, the eigenvalues of its (symmetric) adjacency matrix are $d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and $|\lambda_i| \leq d$ for all i (by the Perron-Frobenius Theorem). The multiplicity of the eigenvalue d is equal to the number of connected components of G, and $\lambda_n = -d$ iff G is bipartite (proofs of these well known facts can be found, for instance, in [7]). Therefore, whenever G is d-regular, the conditions that G should be connected and non-bipartite become equivalent to requiring that $\lambda = \max\{\lambda_2, |\lambda_n|\}$ would satisfy $\lambda < d$. Define the following:

Definition. *An (n,d,λ)-graph, for some integer d and some $\lambda < d$, is a d-regular graph on n vertices whose second largest eigenvalue in absolute value is λ.*

This notion was introduced by the first author in the 80's, motivated by the fact that if λ is much smaller than d, then the graph has strong pseudo-random properties. We mention a few of the properties of these graphs, and refer the readers to [11] for an extensive survey of the subject. Let $G = (V,E)$ denote an (n,d,λ)-graph. First, the behavior of G resembles that of a random graph of edge density d/n in the following sense: if A, B are (not necessarily disjoint) subsets of vertices, then $\left|e(A,B) - \frac{d}{n}|A||B|\right| \leq \lambda\sqrt{|A||B|}$, where $e(A,B)$ denotes the number of ordered pairs

$\{(a,b) : a \in A, b \in B, ab \in E\}$ (see [3], Corollary 9.25). In other words, every two sets of vertices A, B have roughly the "right" number of edges between them. Second, the expansion property of G is closely related to the eigenvalue gap $d - \lambda$, as stated next. Defining the vertex boundary of X, δX, as the set of neighbors of X in $V \setminus X$, it is known that $|\delta X| \geq \frac{2(d-\lambda)}{3d-2\lambda}|X|$ for all sets X of size at most $n/2$ ([2]). Conversely, if

$$|\delta X| \geq c|X| \quad \text{for some } c > 0 \text{ and all } X \subset V\ , |X| \leq \frac{n}{2}\ , \tag{2}$$

then $d - \lambda \geq c^2/(4 + 2c^2)$, implying a discrete version of Cheeger's inequality ([1]). A graph satisfying (2) with c bounded away from 0 is commonly referred to as an *expander*, and according to this definition (n,d,λ)-graphs with λ bounded away from d and regular expanders are very close notions.

In many applications of random walks on expanders, there is not much sense in allowing the walk to backtrack, besides making the model easier to understand and to analyze. A non-backtracking random walk on an undirected graph G, is a walk which does not traverse the same edge twice in a row. In the first part of this paper, we determine the mixing rate of non-backtracking random walks on expanders, using some properties of Chebyshev polynomials of the second kind (the connection between these polynomials and non-backtracking walks follows ideas from [15, 12]). We obtain that for $3 \leq d \leq n^{o(1)}$, the mixing-rate of a non-backtracking random walk on an (n,d,λ)-graph is at most the mixing-rate of a simple random walk on the same graph. In fact, the ratio between the two may reach up to $\frac{2(d-1)}{d}$, as formulated in the next Subsection.

Let G be a d-regular expander. The following definition of the mixing-time of a random walk on G corresponds to an L_∞ distance of $\frac{1}{2n}$, as well as to a relative pointwise distance (r.p.d.) of $\frac{1}{2}$, between π and $P_u^{(k)}$, for all $u \in V$:

$$\tau = \tau(G) = \min_t \left\{ \left| P_{uv}^{(k)} - \frac{1}{n} \right| \leq \frac{1}{2n} \text{ for all } u,v \in V \text{ and } k \geq t \right\}\ . \tag{3}$$

As G is a regular expander, $\tau = \Theta(\log n)$. Notice that sampling the position of the random walk $\mathcal{M}$ at time-points, which are at least τ-apart, gives a more or less independent and uniformly distributed set of vertices. On the other hand, a set of vertices sampled at constant-distance time-points is clearly very much dependent. As we next show, there is a special interest in the distribution of the set of vertices along $\Theta(n)$ consecutive steps of the random walk.

An example of this is the amplification of randomized algorithms (such as the Rabin-Miller primality testing algorithm). Let $\mathcal{A}$ denote such an algorithm which uses $\log n$ random bits; the naive parallel repetition of $\mathcal{A}$ spends $\Theta(n \log n)$ bits in order to reduce the error probability exponentially in n. It is well known that the probability that a random walk of length k avoids a given set of vertices of constant proportion, decreases exponentially with k (see, e.g., [3], Corollary 9.28). Therefore, if G is a regular expander of fixed degree, feeding the positions of a random walk of length $\Theta(n)$ on G as the random seeds for the algorithm, reduces the error probability of the algorithm exponentially, using only $\Theta(n)$ random bits.

In the above application of conserving randomness when amplifying randomized algorithms, our concern was the probability that a random walk of length n misses a large given set of vertices. Instead, in load balancing applications, the concern is the maximal number of times that a random walk of length n visits a vertex. This corresponds to the classical *balls and bins* paradigm (see, e.g, [6], [9]), which discusses the result of throwing n balls to n bins, independently and uniformly at random. In the balls and bins experiment, the bin with the largest number of balls typically contains $(1+o(1))\frac{\log n}{\log\log n}$ balls (see [8]).

As we later show, the random walk is unsuitable for conserving randomness in this case, as a typical random walk of length n has a maximal load of $\Omega(\log n)$. As an application for non-backtracking random walks, we show here that the maximal number of times that such a walk of length n visits a vertex, is $(1+o(1))\log n/\log\log n$ times on high girth expanders with n vertices.

Throughout the paper, we say that an event, which is defined for an infinite series of graphs, occurs *with high probability*, or *almost surely*, or that *almost every* graph of an infinite series of graphs satisfies some property, if the probability for the corresponding event tends to 1 as the number of vertices tends to infinity. Unless stated otherwise, all logarithms are in the natural basis.

1.2 Main results

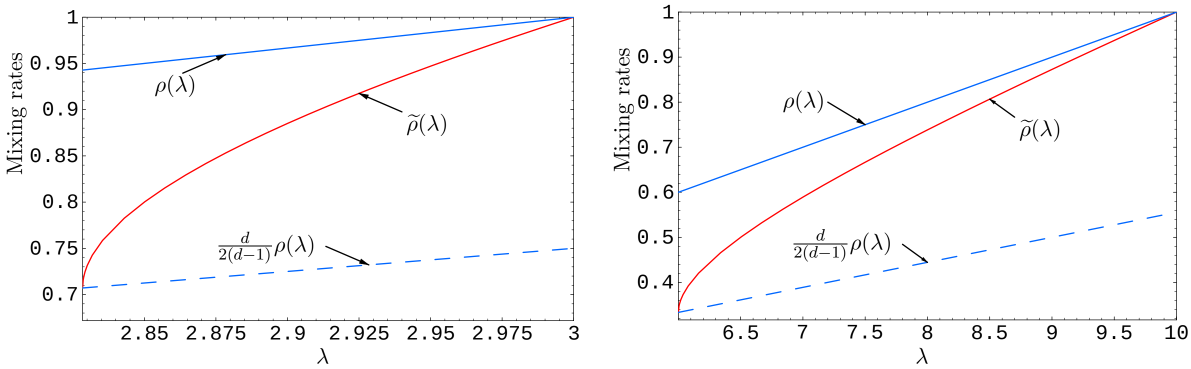
Let $G=(V,E)$ denote an undirected graph. Define a *non-backtracking random walk* of length k on G, from some given vertex $w_0\in V$, as a uniformly chosen member of:

$$\widetilde{\mathcal{W}}^{(k)} = \left\{ (w_0,w_1,\ldots,w_k)\ :\ \begin{array}{c} w_t\in V\ ,\ \ w_{t-1}w_t\in E\ \ \text{for all } t\in[k]\ ,\\ w_{t-1}\neq w_{t+1}\ \ \text{for all } t\in[k-1] \end{array} \right\}\ .$$

Equivalently, a non-backtracking random walk on G from w_0 is a finite Markov chain $\widetilde{\mathcal{M}}$, whose state space is $\overrightarrow{E}$, the set of directed edges of G, taking each edge in both orientations. The distribution of the initial state is given by $\Pr[X_0=(w_0,u)] = \mathbf{1}_{\{w_0u\in E\}}/\deg(w_0)$ (and 0 elsewhere), and the transition probabilities are $P_{(u,v),(v,w)} = \mathbf{1}_{\{u\neq w\}}/(\deg(v)-1)$ (and 0 elsewhere). If G is d-regular, then the transition probabilities matrix is double-stochastic, hence the uniform distribution is a stationary distribution of $\widetilde{\mathcal{M}}$. Notice that if G is 2-regular, then it is a disjoint union of cycles, hence a non-backtracking random walk on G is periodic and does not converge to a stationary distribution. We therefore require that $d\geq 3$, in addition to the requirements that G should be connected and non-bipartite, and these necessary conditions prove to be sufficient for $\widetilde{\mathcal{M}}$ to converge to the uniform distribution.

Let $G=(V,E)$ denote an (n,d,λ)-graph for $d\geq 3$. Recalling (1), define the mixing rate of a non-backtracking random walk on G as:

$$\widetilde{\rho}(G) = \limsup_{k\to\infty}\max_{u,v\in V}\left|\widetilde{P}^{(k)}_{uv}-\frac{1}{n}\right|^{1/k}\ , \tag{4}$$



(a) 3-regular graphs

(b) 10-regular graphs

Figure 1: Mixing rates of simple and non-backtracking random walks on regular graphs.

where $\widetilde{P}_{uv}^{(k)}$ is the probability that a non-backtracking random walk of length k on G, which starts in u, ends in v. The following theorem, proved in Section 2, determines the value of $\widetilde{\rho}$ in this case:

Theorem 1.1. *Let $d \geq 3$ denote some integer, and let G be an (n, d, λ)-graph for some $\lambda < d$. Define $\psi : [0, \infty) \to \mathbb{R}$ by:*

$$\psi(x) = \begin{cases} x + \sqrt{x^2 - 1} & \textit{If } \; x \geq 1 \, , \\ 1 & \textit{If } \; 0 \leq x \leq 1 \, . \end{cases} \tag{5}$$

Then a non-backtracking random walk on G converges to the uniform distribution, and its mixing rate, $\widetilde{\rho}$, satisfies:

$$\widetilde{\rho} = \psi\left(\frac{\lambda}{2\sqrt{d-1}}\right) / \sqrt{d-1} \, . \tag{6}$$

It is well known (see, e.g., [13]), that if G is an (n, d, λ)-graph, then the mixing-rate of the simple random walk on G is $\rho = \lambda/d$. As we state in Section 2, combining this with the properties of the function ψ, defined in (5), gives the inequality $\widetilde{\rho} \leq \rho$, provided $d \leq n^{o(1)}$. The closer λ is to $2\sqrt{d-1}$ (that is, the closer the graph is to being Ramanujan), the closer the ratio $\widetilde{\rho}/\rho$ is to $\frac{d}{2(d-1)}$, as demonstrated in Figure 1. This is formulated in the following corollary:

Corollary 1.2. *Let G be a non-bipartite and connected d-regular graph on n vertices, for some $d \geq 3$, and let ρ and $\widetilde{\rho}$ denote the mixing rates of simple and non-backtracking random walks on G, respectively. The following holds: let λ be the second largest eigenvalue of G in absolute value. If $\lambda \geq 2\sqrt{d-1}$, then*

$$\frac{d}{2(d-1)} \leq \frac{\widetilde{\rho}}{\rho} \leq 1 \, . \tag{7}$$

If $\lambda < 2\sqrt{d-1}$ and $d = n^{o(1)}$, then $\widetilde{\rho}/\rho = \frac{d}{2(d-1)} + o(1)$, where the $o(1)$-term tends to 0 as $n \to \infty$.

In Section 3, we discuss the maximal load of a set of vertices along n consecutive positions of a non-backtracking random walk. The next theorem states that the maximal number of times that such a walk on a regular expander of high girth visits a vertex is equal to $(1+o(1))\frac{\log n}{\log \log n}$, precisely the maximal load in the balls and bins experiment.

Theorem 1.3. *Let G be an (n, d, λ) graph for some fixed $d \geq 3$ and some fixed $\lambda < d$, whose girth is $g \geq 10 \log_{d-1} \log n$. With high probability, the maximal number of times that a non-backtracking random walk of length n on G visits a vertex is equal to $(1 + o(1)) \frac{\log n}{\log \log n}$.*

Furthermore, the above requirement on the girth is essentially tight: in Section 3 we show that, for all $g = g(n)$, there are graphs as described in Theorem 1.3 with girth g, for which the above maximal number of visits is $\Omega(\frac{\log n}{g})$ almost surely.

The final section, Section 4, is devoted to several open problems, further related to random walks on expanders and to similar notions of conserving randomness.

2 The mixing rate of a non-backtracking random walk

Proof of Theorem 1.1. We begin with some preliminaries on Chebyshev polynomials; for further information, see, e.g., [20]. The Chebyshev polynomials of the second kind, of degree $k \geq 0$, are the following polynomials:

$$U_k(\cos \theta) = \frac{\sin((k+1)\theta)}{\sin \theta} \ . \tag{8}$$

Also, it is convenient to define $U_{-1}(x) \equiv 0$. The Chebyshev polynomials satisfy the following three-term recurrence relation:

$$U_{k+1}(x) = 2xU_k(x) - U_{k-1}(x) \ , \text{ for all } k \geq 0 \ , \tag{9}$$

and are orthogonal with respect to the Wigner semicircle measure $d\sigma(x) = \frac{2}{\pi}\sqrt{1-x^2}\mathbf{1}_{[-1,1]}(x)dx$.

Let $A = A(G)$ denote the adjacency matrix of G, and define the $n \times n$ matrix $A^{(k)}$ for $k \geq 1$:

$$A^{(k)}_{u,v} = |\widetilde{\mathcal{W}}^{(k)}_{u,v}| \text{ for all } u, v \in V \ .$$

That is, the entry of $A^{(k)}$ at indices u, v is equal to the number of non-backtracking walks of length k from u to v. By definition, the matrices $A^{(k)}$ satisfy the following recurrence relation:

$$\begin{cases} A^{(1)} = A \ , \ A^{(2)} = A^2 - dI \ , \\ A^{(k+1)} = AA^{(k)} - (d-1)A^{(k-1)} \text{ for } k = 2, 3, \dots \ . \end{cases} \tag{10}$$

where the last term above, $(d-1)A^{(k-1)}$, eliminates the walks which backtrack in the $k+1$ step. We claim that:

$$A^{(k)} = \sqrt{d(d-1)^{k-1}} q_k \left(\frac{A}{2\sqrt{d-1}} \right) \quad \text{for all } k \geq 1 \ , \tag{11}$$

where:

$$q_k(x) = \sqrt{\frac{d-1}{d}} U_k(x) - \frac{1}{\sqrt{d(d-1)}} U_{k-2}(x) \ \text{ for all } k \geq 1 \ . \tag{12}$$

To see this, let $f(A, k) = \sqrt{d(d-1)^{k-1}} q_k \left(A/(2\sqrt{d-1})\right)$ denote the right hand side of (11). Substituting the polynomials $U_{-1}(x) = 0$, $U_0(x) = 1$, $U_1(x) = 2x$ and $U_2(x) = 4x^2 - 1$ in (12) implies that $f(A,1) = A = A^{(1)}$ and that $f(A,2) = A^2 - dI = A^{(2)}$, confirming (11) for $k = 1,2$. In order to verify that (11) holds for all $k \geq 3$, recall that $q_k(x)$ is a linear combination of the polynomials U_{k-2} and U_k, hence it satisfies the recurrence (9):

$$q_{k+1}(x) = 2xq_k(x) - q_{k-1}(x) \quad \text{for all } k \geq 2.$$

Therefore, by induction, the following holds for all $k \geq 2$:

$$\begin{aligned} f(A,k+1) &= \sqrt{d(d-1)^k} q_{k+1}\left(\frac{A}{2\sqrt{d-1}}\right) \\ &= \sqrt{d(d-1)^k}\left[\frac{A}{\sqrt{d-1}} q_k\left(\frac{A}{2\sqrt{d-1}}\right) - q_{k-1}\left(\frac{A}{2\sqrt{d-1}}\right)\right] \\ &= AA^{(k)} - (d-1)A^{(k-1)} = A^{(k+1)} , \end{aligned}$$

where the last inequality is by (10).

Remark 2.1: One can verify that the polynomials $q_k(x)$ are orthogonal polynomials with respect to the Kesten-McKay measure $d\sigma(x) = \frac{2d(d-1)}{\pi} \frac{\sqrt{1-x^2}}{d^2 - 4(d-1)x^2} \mathbf{1}_{[-1,1]}(x)dx$.

Take $k \geq 1$, and recall that $A^{(k)}_{u,v}$ is the number of non-backtracking walks of length k from u to v. Normalizing the matrix $A^{(k)}$ as follows:

$$\widetilde{P}^{(k)} = \frac{A^{(k)}}{d(d-1)^{k-1}} , \tag{13}$$

we obtain that $\widetilde{P}^{(k)}$ is precisely the transition probability matrix of a non-backtracking random walk of length k. Let $\mu_1 = 1, \mu_2, \ldots, \mu_n$ denote the eigenvalues of $\widetilde{P}^{(k)}$, and let

$$\mu = \mu(k) = \max\{|\mu_2|, \ldots, |\mu_n|\} . \tag{14}$$

Claim 2.2. *Let $\widetilde{P}^{(k)}_{ij}$ and $\mu(k)$ be as above. The following holds:*

$$\frac{\mu(k)}{n} \leq \max_{i,j} \left|\widetilde{P}^{(k)}_{ij} - \frac{1}{n}\right| \leq \mu(k) . \tag{15}$$

Proof. The vector $v_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ is an eigenvector of $\widetilde{P}^{(k)}$ corresponding to its largest eigenvalue $\mu_1 = 1$, and therefore:

$$\max_{i,j}\left|\widetilde{P}^{(k)}_{ij} - \frac{1}{n}\right| = \max_{i,j}\left|\left\langle\left(\widetilde{P}^{(k)} - v_1 \otimes v_1\right)e_i, e_j\right\rangle\right| \leq \max_{|u|=|v|=1}\left|\left\langle\left(\widetilde{P}^{(k)} - v_1 \otimes v_1\right)u, v\right\rangle\right| = \mu(k) .$$

On the other hand:

$$\max_{i,j}\left|\widetilde{P}^{(k)}_{ij} - \frac{1}{n}\right| \geq \frac{1}{n}\sqrt{\sum_{i,j}|\widetilde{P}^{(k)}_{ij} - \frac{1}{n}|^2} = \frac{1}{n}\sqrt{\sum_{2\leq s\leq n}\mu_s^2} \geq \mu(k)/n\,.$$

■

We deduce that:

$$\widetilde{\rho} = \limsup_{k\to\infty} \mu(k)^{1/k} = \max_{2\le i\le n} \limsup_{k\to\infty} |\mu_i(k)|^{1/k} \ , \tag{16}$$

and it remains to compute the right hand side above. By (11) and (13), the following holds for all $i \in [n]$:

$$\mu_i = \frac{1}{\sqrt{d(d-1)^{k-1}}} q_k\Big(\frac{\lambda_i}{2\sqrt{d-1}}\Big) \ ,$$

where λ_i are the eigenvalues of A. Therefore, the proof of the theorem will follow from the next lemma:

Lemma 2.3. *The polynomials q_k, defined in* (12)*, satisfy:*

$$\limsup_{k\to\infty} |q_k(x)|^{1/k} = \psi(|x|) = \begin{cases} 1, & -1 \le x \le 1 \ , \\ |x| + \sqrt{x^2-1}, & x \in \mathbb{R} \setminus [-1,\,1] \ . \end{cases}$$

Proof. If $x \in [-1,1]$, then $x = \cos\theta$ for some $\theta \in [0,\pi]$, and hence:

$$q_k(x) = \sqrt{\frac{d-1}{d}} \frac{\sin((k+1)\,\theta)}{\sin\theta} - \frac{1}{\sqrt{d(d-1)}} \frac{\sin((k-1)\,\theta)}{\sin\theta} \ . \tag{17}$$

Therefore:

$$|q_k(x)| \le \sqrt{\frac{d-1}{d}}(k+1) + \frac{1}{\sqrt{d(d-1)}}(k-1)$$

and $\limsup_{k\to\infty} |q_k(x)|^{1/k} \le 1$. The reverse inequality follows from an appropriate subsequence k_j for which the right hand side of (17) is bounded from below by some $c = c(\theta) > 0$.

It remains to treat $x \notin [-1,1]$. In this case, $x = (z+z^{-1})/2$ for $z = x + \mathrm{sign}(x)\sqrt{x^2-1} \notin [-1,1]$. Setting $z = \mathrm{sign}(x)\mathrm{e}^{\theta}$ for some real θ, we get $x = \mathrm{sign}(x)\cos(i\theta)$, and therefore:

$$\begin{aligned} q_k(x) &= \mathrm{sign}(x)^k \left(\sqrt{\frac{d-1}{d}} \frac{\sin((k+1)i\theta)}{\sin(i\theta)} - \frac{1}{\sqrt{d(d-1)}} \frac{\sin((k-1)i\theta)}{\sin(i\theta)} \right) \\ &= \sqrt{\frac{d-1}{d}} \frac{z^{k+1} - z^{-(k+1)}}{z - z^{-1}} - \frac{1}{\sqrt{d(d-1)}} \frac{z^{k-1} - z^{-(k-1)}}{z - z^{-1}} \ , \end{aligned}$$

and $\limsup |q_k(x)|^{1/k} = \lim |q_k(x)|^{1/k} = |z|$.

This completes the proof of the lemma and of Theorem 1.1. ■

Proof of Corollary 1.2. Let λ denote the largest absolute value of a nontrivial eigenvalue of G. Note that $\psi(x)$, as defined in Theorem 1.1, satisfies the following properties:

$$\begin{cases} \psi \text{ is strictly monotone increasing on } [1,\infty) \ , \ \psi(1) = 1 \ , \ \dfrac{\psi(x)}{x} \underset{x\to\infty}{\longrightarrow} 2 \ , \\ \dfrac{\psi\left(\frac{x}{2\sqrt{d-1}}\right)}{\sqrt{d-1}} = \dfrac{x + \sqrt{x^2-4d+4}}{2(d-1)} \le \dfrac{x}{d} \ \text{ for all } d \text{ and all } 2\sqrt{d-1} \le x \le d \ , \\ \psi\left(\frac{d}{2\sqrt{d-1}}\right) = \sqrt{d-1} \ . \end{cases} \tag{18}$$

Therefore, if $\lambda > 2\sqrt{d-1}$, Theorem 1.1 implies that $\widetilde{\rho} = \psi\left(\frac{\lambda}{2\sqrt{d-1}}\right)/\sqrt{d-1}$, and that:

$$\frac{\lambda}{2(d-1)} < \widetilde{\rho} \leq \frac{\lambda}{d} \ .$$

As $\rho = \lambda/d$, we obtain (7). Furthermore, as λ decreases to $2\sqrt{d-1}$, $\psi\left(\frac{\lambda}{2\sqrt{d-1}}\right)$ tends to 1, implying that $\widetilde{\rho} \to \frac{1}{\sqrt{d-1}}$, and $\widetilde{\rho}/\rho \to \frac{d}{2(d-1)}$.

It remains to handle the case $\lambda \leq 2\sqrt{d-1}$. To this end, recall the following result of Nilli [17], which implies the Alon-Boppana Theorem:

Theorem 2.4 ([17])**.** *If G is a simple undirected d-regular graph with diameter at least $2(k+1)$, then the second largest eigenvalue of G, λ_2, satisfies $\lambda_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{k+1}$.*

As the diameter of a d-regular graph on n vertices is at least $(1-o(1))\log_{d-1} n$, we deduce that in the above case, if $d = n^{o(1)}$ then $\lambda = (1-o(1))2\sqrt{d-1}$. In this case, by Theorem 1.1 we have $\widetilde{\rho} = 1/\sqrt{d-1}$, and $\widetilde{\rho}/\rho = \frac{d}{2(d-1)} + o(1)$. ■

Remark 2.5: Examining the trace of the square of the adjacency matrix of a graph, it is easy to see that for every d-regular graph on n vertices, the second largest eigenvalue in absolute value is at least $\sqrt{\frac{d(n-d)}{(n-1)}}$. It thus follows that if $d = o(n)$ then $\widetilde{\rho} \leq (1+o(1))\rho$.

Remark 2.6: For d-regular graphs with $d = \Theta(n)$ the mixing rate of the simple random walk may indeed be faster than that of the non-backtracking random walk. For instance, if G is the complete graph on n vertices, K_n, then by Theorem 1.1, $\widetilde{\rho} = \frac{1}{\sqrt{n-2}}$, and $\rho = \frac{1}{n-1}$.

3 Random walks and the balls and bins paradigm

Proof of Theorem 1.3: Let G be as described in Theorem 1.3. The following definition of the mixing-time of a non-backtracking random walk on G corresponds to an L_∞ distance of $1/n^2$ between π and $\widetilde{P}_u^{(k)}$, for all $u \in V$:

$$\tau = \min_t \left\{ \left|\widetilde{P}_{uv}^{(k)} - \frac{1}{n}\right| \leq \frac{1}{n^2} \text{ for all } u, v \in V \text{ and } k \geq t \right\} \ . \tag{19}$$

Theorem 1.1 implies that a non-backtracking random walk on G converges to the uniform distribution at a mixing-rate of $\widetilde{\rho} = \psi\left(\frac{\lambda}{2\sqrt{d-1}}\right)/\sqrt{d-1}$, and we deduce that $\tau = O(\log n)$ (by usual arguments linking the mixing-rate to the mixing-time).

The proof of Theorem 1.3 will follow from the next two lemmas, which we prove using first and second moment arguments (see, e.g., [3]), combined with some additional ideas.

Lemma 3.1. *Let G be as in Theorem 1.3. With high probability, a non-backtracking random walk of length n on G does not visit a vertex more than $(1+o(1))\frac{\log n}{\log\log n}$ times.*

Lemma 3.2. *Let G be as in Theorem 1.3. With high probability, a non-backtracking random walk of length n on G visits some vertex at least $(1+o(1))\frac{\log n}{\log\log n}$ times.*

The key element in the proofs of both lemmas is showing that the number of times that a non-backtracking random walk visits some vertex, or some pair of vertices, is governed by visits at locations which are at least τ apart. This implies a behavior which is essentially the same as the one in the balls and bins experiment.

Proof of Lemma 3.1. Let $u, v \in V$ denote two vertices, so that either $u = v$ or the distance between u and v in G is at least

$$L = 10 \log_{d-1} \log n \ , \tag{20}$$

and let $\widetilde{P}^{(\ell)}_{uv}$ denote the probability that a non-backtracking random walk of length ℓ on G, which starts at u, ends in v. We claim that:

$$\widetilde{P}^{(\ell)}_{uv} \leq \begin{cases} (d-1)/(\log n)^5 & \text{If } \ell < \tau \ , \\ (1+n^{-1})/n & \text{If } \ell \geq \tau \ . \end{cases} \tag{21}$$

The case $\ell \geq \tau$ follows directly from the definition (19) of the mixing time τ. For the case $\ell < \tau$, let $W = (u = w_0, w_1, \ldots, w_\ell)$ denote a non-backtracking random walk of length ℓ on G, starting at u. The choice of u, v and the fact that g, the girth of G, is at least L (this applies to the case $u = v$), imply that there is no non-empty path between u, v of length shorter than L. Therefore, if $\ell < L$ then $\Pr[w_\ell = v] = 0$. Otherwise, let $h = \lfloor \frac{L-1}{2} \rfloor$, and notice that the neighborhood of v up to distance h is precisely a d-regular tree (as $L \leq g$). Let U denote the $d(d-1)^{h-1}$ leaves of this tree. Since the random walk W cannot backtrack, the event $w_\ell = v$ implies that $w_{\ell-h} \in U$, hence:

$$\begin{aligned} \widetilde{P}^{(\ell)}_{uv} &= \Pr[w_\ell = v] = \Pr[w_\ell = v \mid w_{\ell-h} \in U] \Pr[w_{\ell-h} \in U] \\ &\leq \Pr[w_\ell = v \mid w_{\ell-h} \in U] = (d-1)^{-h} \leq \frac{d-1}{(\log n)^5} \ . \end{aligned}$$

Let $\varepsilon > 0$, and set $k = (1+\varepsilon)\frac{\log n}{\log\log n}$. Consider a non-backtracking random walk of length n on G, $W = (w_0, w_1, \ldots, w_n)$, where w_0 is a fixed vertex of V. For each vertex $v \in V$, and for each $t \in \{0, \ldots, k\}$, define the following event:

$$A_{v,t} = \left(\begin{array}{l} W \text{ visits } v \text{ at least } k \text{ times at some indices } 1 \leq i_1 < \ldots < i_k < \ldots \ , \\ \text{and } |\{j \in [k-1] : i_{j+1} - i_j < \tau\}| = t \ . \end{array} \right) \ .$$

That is, $A_{v,t}$ describes the event in which precisely t of the first $k-1$ segments of W, which are bounded by consecutive visits to v, are of length smaller than τ. Considering all the possible ways to choose indices $i_0, \ldots, i_k$ according to the definition of $A_{v,t}$, we derive the following from (21):

$$\Pr[A_{v,t}] \leq \binom{n}{k-t} \binom{k-1}{t} \tau^t \left(\frac{1+n^{-1}}{n}\right)^{k-t} \left(\frac{d-1}{(\log n)^5}\right)^t \ . \tag{22}$$

For $0 \le t < k-1$, replacing t by $t+1$ in the right hand side of (22) results in a multiplicative factor of:

$$\frac{(k-t)(k-t-1)}{(n-k+t+1)(t+1)} \cdot \frac{\tau n}{1+n^{-1}} \cdot \frac{d-1}{(\log n)^5} = O\left(\frac{k^2 \tau}{(\log n)^5}\right) = o(1) \ .$$

Therefore, the largest term is obtained for $t = 0$. Letting $A_v = \cup_{t=0}^{k-1} A_{v,t}$ denote the event that W visits the vertex v at least k times, we get:

$$\Pr[A_v] \le k \binom{n}{k} \left(\frac{2}{n}\right)^k \le \frac{2^k}{(k-1)!} = o(1/n) \ ,$$

where the last inequality is by the assumption on k. Therefore, $\Pr[\cup_{v \in V} A_v] = o(1)$, and with high probability, W does not visit any vertex of V more than k times. ∎

Proof of Lemma 3.2. Let $\varepsilon > 0$, and set $k = \lceil (1-\varepsilon) \frac{\log n}{\log \log n} \rceil$. Let W denote a non-backtracking random walk of length n on G, $W = (w_0, w_1, \ldots, w_n)$, where w_0 is a fixed vertex of V. We wish to show that, with high probability, W visits some vertex $v \in V$ at least k times. We will show that, in fact, this statement holds even if we restrict ourselves to a predefined subset of the vertices $U \subset V$, and in addition, restrict the pattern of the visiting locations.

Let $U \subset V$ denote a set of vertices of G of size

$$|U| = \lceil n / \left(d(\log n)^{10}\right) \rceil \ , \tag{23}$$

so that the distance between any pair of vertices $u, v \in U$ is at least $L = 10 \log_{d-1} \log n$ (as defined in (20)). To see that such a set U indeed exists, notice that the number of vertices, whose distance from some $u \in U$ is at most $L-1$, does not exceed $\sum_{i=0}^{L-1} d(d-1)^i \le d(d-1)^L$. Therefore, a greedy algorithm which begins with an empty set, and repeatedly adds a new legal vertex to U, always succeeds in producing a set of size at least $n / \left(d(\log n)^{10}\right)$.

The restriction we impose on the pattern of visits is defined next:

Definition. *Let $T \subset \binom{[n]}{k}$ denote a set of k indices in $[n]$. We say that T is a k-**pattern** iff $T \cap [2\tau] = \emptyset$ and $|i-j| > 2\tau$ for all $i, j \in T$. In other words, the value of the elements of T, and the pairwise distances between these elements, all exceed 2τ.*

The above definition implies that, if T is a k-pattern, then for all $i \in [n]$, there is at most one element $j \in T$ so that $|i-j| \le \tau$. This makes it useful to define the correlation between k-patterns as follows:

Definition. *Let T_1 and T_2 denote two k-patterns. The correlation between T_1 and T_2, $\delta(T_1, T_2)$, is defined as the number of pairs in $T_1 \times T_2$ with distance at most τ:*

$$\delta(T_1, T_2) = \left|\left\{(a, b) \in T_1 \times T_2 : |a-b| \le \tau\right\}\right| \ .$$

Let $\mathcal{K}$ denote the collection of all k-patterns, and notice that:

$$|\mathcal{K}| = \binom{n-2\tau k}{k} . \tag{24}$$

Define the following set of indicator variables for all $u \in U$ and $T \in \mathcal{K}$:

$$X_{u,T} = \begin{cases} 1 & \text{If } w_i = u \text{ for all } i \in T \ , \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

In other words, $X_{u,T}$ is the indicator for the event according to which the non-backtracking walk W visits u in all the time-points specified by T. By definition, the first of these time-points exceeds τ, and the same holds for the distance between each consecutive pair of these time-points, and by the definition of τ we deduce that:

$$\left(\frac{1-n^{-1}}{n}\right)^k \leq \Pr[X_{u,T}=1] \leq \left(\frac{1+n^{-1}}{n}\right)^k . \tag{26}$$

Setting $X = \sum_{u,T} X_{u,T}$, we get:

$$\mathbb{E}X \geq |U|\binom{n-2\tau k}{k}\left(\frac{1-n^{-1}}{n}\right)^k = n^{\varepsilon-o(1)} , \tag{27}$$

where the last equality is by the definition of k and (23).

In order to show that X is concentrated around its expected value, we consider its second moment. Let $u, v \in U$ so that $u \neq v$, and let $t \in \{0, 1, \ldots, k\}$. Take $T_1, T_2 \in \mathcal{K}$ so that $\delta(T_1, T_2) = t$. By the definition of U, the distance between u and v is at least L. Hence, if $|a-b| < L$ for some $(a,b) \in T_1 \times T_2$, then the events $(X_{u,T_1} = 1)$ and $(X_{v,T_2} = 1)$ are disjoint. Otherwise, consider the probability of the event $(X_{u,T_1} = 1) \wedge (X_{v,T_2} = 1)$. By (21), the largest of each of the t pairs of indices $(a_i, b_i) \in T_1 \times T_2$, which satisfy $|a_i - b_i| \leq \tau$, contributes a probability of at most $(d-1)/(\log n)^5$ to this event. The definition of τ implies that each of the remaining indices contributes a probability of at most $(1+n^{-1})/n$ for visiting the required vertex (either u or v), and altogether:

$$\Pr[X_{u,T_1} = 1 \wedge X_{v,T_2} = 1] \leq \left(\frac{1+n^{-1}}{n}\right)^{2k-t}\left(\frac{d-1}{(\log n)^5}\right)^t . \tag{28}$$

Combining (26) and (28) gives:

$$\begin{aligned}
&\sum_{T_1\in\mathcal{K}} \sum_{\substack{T_2\in\mathcal{K} \\ \delta(T_1,T_2)=t}} \mathrm{Cov}(X_{u,T_1}, X_{v,T_2}) \\
&\leq \binom{n-2\tau k}{k}\binom{k}{t}(2\tau)^t\binom{n-2\tau k}{k-t}\left(\left(\frac{1+n^{-1}}{n}\right)^{2k-t}\left(\frac{d-1}{(\log n)^5}\right)^t - \left(\frac{1-n^{-1}}{n}\right)^{2k}\right) \\
&= \binom{n-2\tau k}{k}\binom{k}{t}(2\tau)^t\binom{n-2\tau k}{k-t} n^{-2k}\left(\left(1+n^{-1}\right)^{2k}\left(\frac{(1+o(1))(d-1)n}{(\log n)^5}\right)^t - \left(1-n^{-1}\right)^{2k}\right) .
\end{aligned}$$

Let $C_{uv}(t)$ denote the right hand side in the above inequality. Since $(1+n^{-1})^{2k}$ and $(1-n^{-1})^{2k}$ both tend to 1 as n, and hence k, tend to ∞, the following holds for all $t \geq 1$:

$$\frac{C_{uv}(t+1)}{C_{uv}(t)} = \frac{(k-t)^2(2\tau)}{(t+1)(n-(2\tau+1)k+t+1)} \cdot (1+o(1))\frac{(d-1)n}{(\log n)^5} = O\left(\frac{k^2\tau}{(\log n)^5}\right) = o(1) \ .$$

In particular, for a sufficiently large n we deduce that

$$\sum_{t=1}^{k} C_{uv}(t) \leq 2C_{uv}(1) \ , \tag{29}$$

and it remains to examine $C_{uv}(t)$ for $t \in \{0,1\}$:

$$\begin{aligned} C_{uv}(0) &= \binom{n-2\tau k}{k}^2 \left(\left(\frac{1+n^{-1}}{n}\right)^{2k} - \left(\frac{1-n^{-1}}{n}\right)^{2k} \right) \\ &\leq \left(\frac{\mathbb{E}X}{|U|}\right)^2 \left(\frac{n}{1-n^{-1}}\right)^{2k} \cdot \frac{2}{n^2} \cdot 2k \left(\frac{1+n^{-1}}{n}\right)^{2k-1} = O\left(\frac{k}{n}\left(\frac{\mathbb{E}X}{|U|}\right)^2\right) \ , \end{aligned} \tag{30}$$

$$\begin{aligned} C_{uv}(1) &= \binom{n-2\tau k}{k} \cdot 2\tau k \binom{n-2\tau k}{k-1} n^{-2k} \cdot (1+o(1))\frac{(d-1)n}{(\log n)^5} \\ &\leq \left(\frac{\mathbb{E}X}{|U|}\right)^2 \left(\frac{n}{1-n^{-1}}\right)^{2k} \cdot \frac{2\tau k}{n^{2k}} \cdot \frac{(1+o(1))k(d-1)}{(\log n)^5} = O\left(\frac{\tau k^2}{(\log n)^5}\left(\frac{\mathbb{E}X}{|U|}\right)^2\right) \ . \end{aligned} \tag{31}$$

By (29), (30) and (31) we get:

$$\sum_{u \in U} \sum_{\substack{v \in U \\ u \neq v}} \sum_{(T_1,T_2) \in \mathcal{K}^2} \mathrm{Cov}(X_{u,T_1}, X_{v,T_2}) \leq \sum_{u \in U} \sum_{\substack{v \in U \\ u \neq v}} \sum_{t=0}^{k} C_{uv}(t) = o((\mathbb{E}X)^2) \ . \tag{32}$$

Next, take $u \in U$, and consider all k-patterns $T_1 \neq T_2$ which contain $l \in \{0,\ldots,k-1\}$ common indices, and whose correlation, $\delta(T_1,T_2)$, is some $t \in \{l,\ldots,k\}$. The following holds:

$$\begin{aligned} &\sum_{T_1 \in \mathcal{K}} \sum_{\substack{T_2 \in \mathcal{K} \\ |T_1 \cap T_2| = l \\ \delta(T_1,T_2)=t}} \mathrm{Cov}(X_{u,T_1}, X_{u,T_2}) \\ &\leq \binom{n-2\tau k}{k}\binom{k}{t}\binom{t}{l}(2\tau)^{t-l}\binom{n-2\tau k}{k-t} \left(\left(\frac{1+n^{-1}}{n}\right)^{2k-t} \left(\frac{d-1}{(\log n)^5}\right)^{t-l} - \left(\frac{1-n^{-1}}{n}\right)^{2k} \right) \\ &\leq (1+o(1))\binom{n-2\tau k}{k}\binom{k}{t}\binom{t}{l}\binom{n-2\tau k}{k-t} n^{-2k+t} \left(\frac{2\tau(d-1)}{(\log n)^5}\right)^{t-l} \ . \end{aligned} \tag{33}$$

Let $C_u(l,t)$ denote the final expression of (33). For all l and t so that $l \leq t < k$ we have:

$$\frac{C_u(l,t+1)}{C_u(l,t)} = \frac{(1+o(1))(k-t)^2}{t-l+1} \cdot \frac{2\tau(d-1)}{(\log n)^5} = O\left(\frac{k^2\tau}{(\log n)^5}\right) = o(1) \ . \tag{34}$$

This implies that the leading order term in the sum $\sum_{t=l}^{k} C_u(l,t)$ is $C_u(l,l)$. Next,

$$\frac{C_u(l+1,l+1)}{C_u(l,l)} = \frac{(1+o(1))(k-l)^2}{l+1} \ ,$$

hence, if we define:

$$l_0 = k - 2\sqrt{k} \ , \ l_1 = k - \frac{1}{2}\sqrt{k} \ ,$$

then the following holds:

$$\begin{cases} \frac{C_u(l+1,l+1)}{C_u(l,l)} \geq 4 + o(1) & \text{if } l \leq l_0 \ , \\ \frac{C_u(l+1,l+1)}{C_u(l,l)} \leq \frac{1}{4} + o(1) & \text{if } l \geq l_1 \ . \end{cases} \tag{35}$$

On the other hand, for every $l \in [l_0, l_1]$ we have:

$$\begin{aligned} C_u(l,l) &= (1+o(1))\binom{n-2\tau k}{k}\binom{k}{l}\binom{n-2\tau k}{k-l} n^{-2k+l} \\ &\leq (1+o(1))\frac{\mathbb{E}X}{|U|}\left(\frac{\mathrm{e}^2 k(n-2\tau k)}{(k-l)^2 n}\right)^{k-l} \leq (1+o(1))\frac{\mathbb{E}X}{|U|}\left(4\mathrm{e}^2 + o(1)\right)^{2\sqrt{k}} \\ &= (1+o(1))\frac{\mathbb{E}X}{|U|} n^{o(1)} = o\left(\frac{(\mathbb{E}X)^2}{|U| n^{\varepsilon/2}}\right) \ , \end{aligned} \tag{36}$$

where the last equality is by (27). We deduce from (34), (35) and (36) that for all sufficiently large values of n:

$$\begin{aligned} \sum_{l=0}^{k-1}\sum_{t=l}^{k} C_u(l,t) &\leq 2\sum_{l=0}^{k-1} C_u(l,l) \leq 4C_u(l_0,l_0) + 4C_u(l_1,l_1) + 2\sum_{l=l_0}^{l_1} C_u(l,l) \\ &= o\left(\sqrt{k}\frac{(\mathbb{E}X)^2}{|U| n^{\varepsilon/2}}\right) = o\left(\frac{(\mathbb{E}X)^2}{|U|}\right) \ , \end{aligned}$$

and thus:

$$\sum_{u\in U}\sum_{T_1\in\mathcal{K}}\sum_{\substack{T_2\in\mathcal{K} \\ T_1\neq T_2}} \mathrm{Cov}(X_{u,T_1}, X_{u,T_2}) \leq \sum_{u\in U}\sum_{l=0}^{k-1}\sum_{t=l}^{k} C_u(l,t) = o\left((\mathbb{E}X)^2\right) \ . \tag{37}$$

Combining (32) and (37) (and recalling that $\mathbb{E}X = \omega(1)$) gives:

$$\mathrm{Var}(X) \leq \mathbb{E}X + \sum_{u,T_1}\sum_{\substack{v,T_2 \\ (u,T_1)\neq(v,T_2)}} \mathrm{Cov}(X_{u,T_1}, X_{v,T_2}) = o((\mathbb{E}X)^2) \ ,$$

and Chebyshev's inequality implies that:

$$\Pr[X=0] \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2} = o(1) \ .$$

This completes the proof of Lemma 3.2 and of Theorem 1.3. ■

We note that the $\Omega(\log\log n)$ requirement on the girth of G in Theorem 1.3 is tight, as there are (n,d,λ)-graphs with girth g, where a non-backtracking random walk visits some vertex at least $\Omega(\frac{\log n}{g})$ times almost surely. This is stated in the next claim.

Claim 3.3. *Let G be a d-regular graph on n vertices, in which each vertex is contained in a cycle of length $g = g(n)$. If $k = k(n)$ satisfies:*

$$k = \frac{\log_{d-1}(n/\log n) - \omega(1)}{g} ,$$

then, with high probability, a non-backtracking random walk of length n on G visits some vertex at least k times. In particular, such a walk almost surely visits some vertex $\Omega(\frac{\log n}{g})$ times.

Proof. For each $v \in V$, let C_v denote a cycle of length g which contains v in G. Let $W = (w_0, w_1, \ldots, w_n)$ denote a non-backtracking random walk of length n on G, and divide W into $T = \lfloor n/kg \rfloor$ disjoint segments, $I_1, \ldots, I_T$, each of length kg:

$$I_j = \left(w_{(j-1)kg}, \ldots, w_{jkg-1}\right) \text{ for all } j \in T .$$

Define the following event for each $j \in [T]$:

$$A_j = \left(\begin{array}{l} \text{The segment } I_j \text{ of } W \text{ is precisely } k \text{ consecutive walks} \\ \text{along the same cycle } C_v, \text{ where } v = w_{(j-1)kg}. \end{array} \right) .$$

To prove the claim, it suffices to show that, with high probability, at least one of the events A_j ($j \in [T]$) occurs. Since these events are independent, and $\Pr[A_j] = (d-1)^{-kg}$ for all j, we get:

$$\Pr[\cap_{j=1}^{T} A_j^c] = \left(1 - (d-1)^{-kg}\right)^T \leq \exp(-T(d-1)^{-kg}) .$$

The choice of k ensures that $T(d-1)^{-kg} = \omega(1)$, and the result follows. ∎

Remark 3.4: Theorem 1.3 stated that the maximal load in a non-backtracking walk of length n on a d-regular expander of high girth is $(1+o(1))\frac{\log n}{\log \log n}$ with high probability, similar to the maximal load in the classical balls and bins experiment. In contrast to this, a simple calculation shows that a typical simple random walk of length n, on any d-regular graph for a fixed d, has a maximal load of $\Omega(\log n)$. This can be seen as a special case of Claim 3.3, taking $g = 2$: the probability that the simple random walk traverses the same edge repeatedly for, say, $M = \frac{1}{2}\log_d n$ consecutive steps, is $1/\sqrt{n}$. Dividing the walk to disjoint segments of length M implies that, with probability $1 - o(1)$, at least one segment exhibits this behavior, thus the maximal load is at least M.

Remark 3.5: The classical Birthday Paradox states that, when throwing balls to n bins, independently and uniformly at random, we expect a collision after $\Theta(\sqrt{n})$ balls (see, e.g., [6]). Relating this to random walks on expanders, one may ask when do simple and non-backtracking random walks on expanders self-intersect. Clearly, most *simple* random walks on an expander encounter a collision after $O(1)$ steps (the first time at which an edge is traversed twice in a row). An argument similar to the one used in the proof of Claim 3.3 shows that, for every small $\varepsilon > 0$, there are (n, d, λ)-graphs with girth $g = \varepsilon \log_{d-1} n$, on which a non-backtracking random walk will self intersect after at most $n^{\varepsilon + o(1)}$ steps almost surely. Similarly, for $g = o(\log n)$, there are such graphs where the self-intersection time of the non-backtracking random walk is at most $(d-1)^{g+o(1)}$.

4 Concluding remarks and open problems

- We have shown that a non-backtracking random walk on every connected and non-bipartite d-regular graph G, where $d \geq 3$, converges to the uniform distribution, and computed its precise mixing-rate. We obtained that this mixing-rate is always asymptotically at least as fast as that of the simple random walk on the same graph provided $d = o(n)$ (and is faster provided $d \leq n^{o(1)}$), and their ratio may reach up to $2(d-1)/d$.

- As an application, we showed that if G is a high-girth d-regular expander on n vertices, for some fixed $d \geq 3$, then the maximal load while sampling n consecutive positions of a non-backtracking random walk on G is almost surely $(1+o(1))\frac{\log n}{\log\log n}$, similar to the maximal load in the classical balls and bins experiment. Performing a simple random walk, instead of a non-backtracking one, results in a maximal load of $\Omega(\log n)$ with high probability.

- Following the Poisson approximations in the balls and bins model, it would be interesting to establish the precise distribution of a sample of n consecutive positions of a non-backtracking random walk on an expander of high girth.

- The well known *power-of-two* result ([4], see also [16], Chapter 14) states that if n balls are thrown into n bins, where each ball is placed in the least loaded bin, out of two independently chosen random ones, then the maximal load decreases from $\Theta(\frac{\log n}{\log\log n})$ to $\Theta(\log\log n)$. Let W_1 and W_2 denote two non-backtracking random walks on an expander of high girth, and suppose that in each step we are given a choice between the two current locations of W_1 and W_2, and pick the least loaded one. Does the maximal load decrease from $\Theta(\frac{\log n}{\log\log n})$ to $\Theta(\log\log n)$ in this setting as-well?

- One way of proving the above power-of-two result in the balls and bins model is to consider the Erdős-Rényi random graph process $\mathcal{G}^t$, $t \in \{0,1,\ldots,\binom{n}{2}\}$ (where $\mathcal{G}^0$ is the empty graph on n vertices, and in each step a new edge is added, uniformly chosen over all missing edges; see, e.g., [5], Chapter 2). Each pair of bins corresponds to a uniformly chosen edge in the graph (we may ignore self-loops or repeating edges, as we are dealing with a linear number of balls). Selecting a bin corresponds to choosing an orientation for this edge. One can show that the greedy online algorithm, which orients an edge towards the vertex with the lower in-degree, gives an overall maximal in-degree of $O(\log\log n)$ with high probability. This is based on the following properties of $\mathcal{G}^t$, which hold with high probability for all $t \leq \alpha n$, where $0 < \alpha < \frac{1}{2}$ is a constant:

  (1) Each connected component of $\mathcal{G}^t$ is of logarithmic size.

  (2) For some fixed δ, the average degree of every induced subgraph of $\mathcal{G}^t$ is at most δ.

  The above discussion suggests the following approach: let G be a d-regular expander of high girth, for some fixed $d \geq 3$, and let W_1 and W_2 denote two non-backtracking random walks on

G. Define a random (multi) graph process by adding the edge $(W_1(t), W_2(t))$ at step t, where $W_i(t)$ is the position of W_i at time t. This can be viewed as a certain de-randomization of the random graph process, where the graph at time $\Theta(n)$ is produced using only $\Theta(n)$ random bits (instead of $\Theta(n \log n)$ bits). This model, on its own account, seems interesting, with respect to the commonly studied questions on graph processes, e.g., whether there exists a sharp threshold for the appearance of a giant component. In particular, proving that properties (1) and (2) hold for this graph process for all $t \leq \alpha n$ and some $0 < \alpha < \frac{1}{2}$ will imply a positive answer to the previous question, regarding the power-of-two with non-backtracking random walks.

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica, 6 (1986), 83-96.

[2] N. Alon and V.D. Milman, λ_1, isoperimetric inequalities for graphs and superconcentrators, J. Combinatorial Theory, Ser. B 38 (1985), 73-88.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, Wiley, New York, 2000.

[4] Y. Azar, A. Broder, A. Karlin and E. Upfal, Balanced allocation, Proc. of 26th STOC (1994), 593-602. Also in SIAM J. Computing, 29 (1999), 180-200.

[5] B. Bollobás, *Random Graphs*, volume 73 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, second edition, 2001.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol I, Wiley, 1968.

[7] C. Godsil and G. Royle, *Algebraic Graph Theory*, volume 207 of Graduate Text in Mathematics, Springer, New York, 2001.

[8] G.H. Gonnet, Expected length of the longest probe sequence in hash code searching, J. ACM, 28 (1981), 289-304.

[9] N.L. Johnson and S. Kotz, *Urn Models and their Applications*, Wiley, 1977.

[10] J. Keilson, *Markov Chain Models - Rarity and Exponentiality*, Springer-Verlag, New York, 1979.

[11] M. Krivelevich and B. Sudakov, Pseudo-random graphs, in: More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer, 2006, 199-262.

[12] W.W. Li and P. Solé, Spectra of regular graphs and hypergraphs and orthogonal polynomials, European J. Combin. 17 (1996), no. 5, 461–477.

[13] L. Lovász, Random walks on graphs: a survey, in: Combinatorics, Paul Erdős is Eighty, Vol. 2 (ed. D. Miklós, V. T. Sós, T. Szőnyi), János Bolyai Mathematical Society, Budapest, 1996, 353-398.

[14] L. Lovász and P. Winkler, Mixing of random walks and other diffusions on a graph, in: Surveys in Combinatorics (ed. P. Rowlinson), London Math. Soc. Lecture Notes Series 218, Cambridge Univ. Press (1995), 119-154.

[15] B.D. McKay, The expected eigenvalue distribution of a random labelled regular graph, Linear Algebra and its Applications, 40 (1981), 203-216.

[16] M.D. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge, 2005.

[17] A. Nilli, On the second eigenvalue of a graph, Discrete Math., 91 (1991), 207-210.

[18] J.R. Norris, *Markov Chains*, Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.

[19] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, Combinatorics, Probability and Computing 1 (1992), pp. 351-370.

[20] G. Szegő, *Orthogonal Polynomials*, 4th Edition, AMS Colloquium Publications, vol. 23, Providence, RI, 1975.