

# Poisson approximation for non-backtracking random walks

Noga Alon\* Eyal Lubetzky †

## Abstract

Random walks on expander graphs were thoroughly studied, with the important motivation that, under some natural conditions, these walks mix quickly and provide an efficient method of sampling the vertices of a graph. The authors of [2] studied non-backtracking random walks on regular graphs, and showed that their mixing rate may be up to twice as fast as that of the simple random walk. As an application, they showed that the maximal number of visits to a vertex, made by a non-backtracking random walk of length  $n$  on a high-girth  $n$ -vertex regular expander, is typically  $(1 + o(1))\frac{\log n}{\log \log n}$ , as in the case of the balls and bins experiment. They further asked whether one can establish the precise distribution of the visits such a walk makes.

In this work, we answer the above question by combining a generalized form of Brun's sieve with some extensions of the ideas in [2]. Let  $N_t$  denote the number of vertices visited precisely  $t$  times by a non-backtracking random walk of length  $n$  on a regular  $n$ -vertex expander of fixed degree and girth  $g$ . We prove that if  $g = \omega(1)$ , then for any fixed  $t$ ,  $N_t/n$  is typically  $\frac{1}{e!} + o(1)$ . Furthermore, if  $g = \Omega(\log \log n)$ , then  $N_t/n$  is typically  $\frac{1+o(1)}{e!}$  uniformly on all  $t \leq (1 - o(1))\frac{\log n}{\log \log n}$  and 0 for all  $t \geq (1 + o(1))\frac{\log n}{\log \log n}$ . In particular, we obtain the above result on the typical maximal number of visits to a single vertex, with an improved threshold window. The essence of the proof lies in showing that variables counting the number of visits to a set of sufficiently distant vertices are asymptotically independent Poisson variables.

## 1 Introduction

### 1.1 Background and definitions

Random walks on graphs have played a major role in Theoretical and Applied Computer Science, as under some natural requirements (related to the notion of expander graphs), these walks converge quickly to a unique stationary distribution, and enable efficient sampling of this distribution. This fact was exploited for example in [1], [5] and [13], in the study of space efficient algorithms for  $S-T$

---

\*Schools of Mathematics and Computer Science, 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 a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†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.

connectivity in undirected graphs. Another well known example is the conservation of random bits in the amplification of randomized algorithms (we will elaborate on this point later on). In many applications for random walks, it seems that using non-backtracking random walks may yield better results, and that these walks possess better random-looking properties than those of simple random walks. This motivated the authors of [2] to study the mixing-rate of a non-backtracking random walk, and show that it may be up to twice as fast as that of a simple random walk. They further show that the number of times that such a walk visits the vertices of a high-girth expander is random-looking, in the sense that its maximum is typically asymptotically the same as the maximal load of the classical balls and bins experiment. In this paper, we further examine this setting, and answer a question raised in [2] by giving a precise description of the limiting distribution of these visits.

We briefly mention some well known properties of random walks on regular graphs; for further information, see, e.g., [11], [14]. Let  $G = (V, E)$  denote a  $d$ -regular undirected graph on  $n$  vertices. A random walk of length  $k$  on  $G$  from a given vertex  $w_0$  is a uniformly chosen member  $W \in \mathcal{W}$ , where  $\mathcal{W} = \{(w_0, w_1, \dots, w_k) : w_{i-1}w_i \in E\}$  is the set of all paths of length  $k$  starting from  $w_0$ . Alternatively, a random walk on  $G$ ,  $\mathcal{M}$ , is a Markov chain whose state space is  $V$ , where the transition probability from  $u$  to  $v$  is  $P_{uv} = \mathbf{1}_{uv \in E}/d$ . The transition probability matrix of  $\mathcal{M}$  is doubly stochastic, and the uniform distribution  $\pi(u) = 1/n$  is a stationary distribution of  $\mathcal{M}$ . If  $G$  is connected and non-bipartite then  $\mathcal{M}$  is irreducible and aperiodic, in which case it converges to the unique stationary distribution  $\pi$ , regardless of the starting point  $w_0$ . These two sufficient and necessary conditions have a clear formulation in terms of the spectrum of  $G$ , which also dictates the rate at which  $\mathcal{M}$  converges to  $\pi$ .

The adjacency matrix of  $G$  is symmetric and thus has  $n$  real eigenvalues, all at most  $d$  in absolute value (by the Perron-Frobenius Theorem). Let  $d = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  denote these eigenvalues. It is simple and well known (see, e.g., [8]) that the multiplicity of the eigenvalue  $d$  is equal to the number of connected components of  $G$ , and the minimal eigenvalue  $\lambda_n$  is equal to  $-d$  iff  $G$  has a bipartite connected component. Therefore, letting  $\lambda = \max_{i>1} |\lambda_i|$  denote the maximal absolute value of all non-trivial eigenvalues, we obtain that  $G$  is connected and non-bipartite iff  $\lambda < d$ . The quantity  $d - \lambda$  is often referred to as the *spectral gap* of  $G$ , and is strongly related to the expansion properties of the graph. In particular, when  $d - \lambda$  is bounded from below by a constant, we call the graph an *expander* (a closely related notion of expander graphs is defined by the expansion ratio of each set of at most  $n/2$  vertices to its neighbor vertices in the graph). See [10] for a survey on the many fascinating pseudo-random properties exhibited by such graphs.

An  $(n, d, \lambda)$  graph is a  $d$ -regular graph on  $n$  vertices, whose largest non-trivial eigenvalue in absolute value is  $\lambda$ . As mentioned above, the condition  $\lambda < d$  is sufficient and necessary for the random walk on  $G$  to converge to  $\pi$ . In this case, the quantity  $\lambda/d$  governs the rate of this convergence: the *mixing rate* of  $\mathcal{M}$  is defined to be  $\limsup_{k \rightarrow \infty} |P_{uv}^{(k)} - \pi(v)|^{1/k}$ , where  $P_{uv}^{(k)}$  is the probability that  $\mathcal{M}$  reaches  $v$  in the  $k$ -th step given that it started from  $u$ . It is well known (see, for instance, [11]) that the mixing rate of the simple random walk on an  $(n, d, \lambda)$  graph is  $\lambda/d$ , and

that in fact, the  $L_2$  distance between  $\pi$  and the distribution of  $\mathcal{M}$  after  $k$  steps is at most  $(\lambda/d)^k$ . Therefore, after  $\Omega(\log_{d/\lambda} n)$  steps, the  $L_2$  distance between the distribution of the simple random walk and the uniform distribution is at most  $1/n^{\Omega(1)}$ .

A useful and well known application of random walks on expander graphs is the following result, related to the conservation of random bits (cf., e.g., [4], Corollary 9.28). Suppose  $G$  is an  $(n, d, \lambda)$  graph, and  $U$  is a predefined set of  $\alpha n$  vertices, for some  $\alpha > 0$ . Then a random walk of length  $k$ , starting from a random vertex, avoids  $U$  with probability at most  $(1 - \alpha + \alpha \frac{\lambda}{d})^k$ . Indeed, the term  $(1 - \alpha)^k$  is the probability to miss  $U$  when selecting  $k$  vertices, uniformly and independently, in which case we would require  $k \log n$  random bits for the selection process. Using a random walk, we require only  $\log n + k \log d$  random bits, at the cost of increasing the base of the exponent by an additive term of  $\alpha \frac{\lambda}{d}$ . This enables amplifying the error-probability in randomized algorithms (such as the Rabin-Miller primality testing algorithm) using fewer random bits: an algorithm utilizing  $s$ -bit seeds can be amplified  $k$  times using  $s + \Theta(k)$  random bits via a walk on a constant-degree expander, instead of  $sk$  random bits in the naive approach.

In many applications of random walks on graphs, forbidding the random walk to backtrack appears to produce better results; an example of this is the construction of sparse universal graphs in [3], where a crucial element is a non-backtracking random walk on a high-girth expander. A non-backtracking random walk of length  $k$  on  $G$ , starting from some vertex  $w_0$ , is a sequence  $\widetilde{W} = (w_0, \dots, w_k)$ , where  $w_i$  is chosen uniformly over all neighbors of  $w_{i-1}$  excluding  $w_{i-2}$ . The mixing-rate of a non-backtracking random walk on a regular graph, in terms of its eigenvalues, was computed in [2], using some properties of Chebyshev polynomials of the second kind. It is shown in [2] that this rate is always better than that of the simple random walk, provided that  $d = n^{o(1)}$ . In fact, the mixing rate of the non-backtracking random walk may be up to twice faster, and the closer the graph is to being a Ramanujan graph (that is, a graph satisfying  $\lambda \leq 2\sqrt{d-1}$ ), the closer the ratio between the two mixing-rates is to  $2(d-1)/d$ .

As an application, the authors of [2] analyzed the maximal number of visits that a non-backtracking random walk of length  $n$  makes to a vertex of  $G$ , an  $(n, d, \lambda)$  graph of fixed degree and girth  $\Omega(\log \log n)$ . Using a careful second moment argument, they proved that this quantity is typically  $(1 + o(1)) \frac{\log n}{\log \log n}$ , as is the typical maximal number of balls in a single bin when throwing  $n$  balls to  $n$  bins uniformly at random (more information on the classical balls and bins experiment may be found in [7]). In contrast to this, it is easy to see that a typical **simple** random walk of length  $n$  on a graph as above visits some vertex  $\Omega(\log n)$  times. The authors of [2] further asked whether it is possible to establish the precise distribution of the number of visits that a non-backtracking random walk makes on a graph  $G$  as above.

In this paper, we answer the above question, by combining a generalized form of Brun's Sieve with extensions of some of the ideas in [2]. This approach shows that even if the girth of  $G$  grows to infinity arbitrarily slowly with  $n$ , then for any fixed  $t$ , the fraction of vertices visited precisely  $t$  times is typically  $(1 + o(1))n/(et!)$ , where the  $o(1)$ -term tends to 0 as  $n \rightarrow \infty$ . The extension of

Brun's Sieve, which includes an estimate on the convergence rate of the variables, is used to treat the case where  $t$  depends on  $n$ , and allows the characterization of the number of vertices visited  $t$  times for all  $t$ . In particular, this provides an alternative proof for the result of [2] of the typical maximum number of visits to a vertex, with an improved error term. These results are summarized below.

Throughout the paper, all logarithms are in the natural basis, and an event, defined for every  $n$ , is said to occur *with high probability* or *almost surely* if its probability tends to 1 as  $n \rightarrow \infty$ .

## 1.2 Results for expanders with a non-fixed girth

A simple argument shows that the requirement for a non-fixed girth is essentially necessary if one wishes that the number of visits at vertices will exhibit a Poisson distribution. Indeed, if  $G$  is a graph where every vertex is contained in a cycle of a fixed length, the probability that the walk  $\widetilde{W}$  will traverse a cycle  $t$  consecutive times, becomes much larger than the Poisson probability of  $t$  visits to a vertex for a sufficiently large  $t$ . Simple random walks correspond to the case of cycles of length 2.

On the other hand, this requirement on the girth turns out to be sufficient as-well: as long as the girth of an  $(n, d, \lambda)$  graph  $G$  tends to infinity arbitrarily slowly with  $n$ , the number of visits that a non-backtracking random walk makes to vertices exhibits a Poisson distribution, as the following theorem states:

**Theorem 1.1.** *For fixed  $d \geq 3$  and fixed  $\lambda < d$ , let  $G$  be an  $(n, d, \lambda)$  graph whose girth is at least  $g = \omega(1)$ . Let  $\widetilde{W} = (w_0, \dots, w_n)$  denote a non-backtracking random walk of length  $n$  on  $G$  from  $w_0$ , and  $N_t$  the number of vertices which  $\widetilde{W}$  visits precisely  $t$  times:*

$$N_t = \left| \left\{ v \in V(G) : \left| \{ 1 \leq i \leq n : w_i = v \} \right| = t \right\} \right|. \quad (1)$$

*Then for every fixed  $t$ ,  $N_t/n = 1/(et!) + o(1)$  almost surely, the  $o(1)$ -term tending to 0 as  $n \rightarrow \infty$ .*

As we later mention, the above theorem in fact holds even for non-fixed  $\lambda, d$  (as long as the spectral gap is large compared to  $d/g$ ). The essence of the proof of Theorem 1.1 lies in proving that the variables counting the visits at vertices, whose pairwise distances are large, are asymptotically independent Poisson variables:

**Proposition 1.2.** *Let  $G$  be a graph as in Theorem 1.1. For some fixed  $r$  and  $\mu > 0$ , let  $v_1, \dots, v_r$  denote vertices of  $G$  whose pairwise distances are at least  $g$ . Let  $\widetilde{W}$  be a non-backtracking random walk of length  $m = \mu n$  on  $G$  starting from  $v_1$ , and  $X_i$  be the number of visits that  $\widetilde{W}$  makes to  $v_i$ . Then  $(X_1, \dots, X_r) \xrightarrow{d} (Z_1, \dots, Z_r)$  as  $n \rightarrow \infty$ , where the  $Z_i$ -s are i.i.d. Poisson random variables with means  $\mu$ ,  $Z_i \sim \text{Po}(\mu)$ .*

**Remark 1.3:** The statement of Proposition 1.2 holds (with the same proof) even if we replace the requirement on the girth of  $G$  with the weaker assumption, that  $v_1, \dots, v_r$  are not contained in

a closed nontrivial walk of length smaller than  $g$ . In this case, the presence of other possibly short cycles in  $G$  has no effect on the limiting distribution of  $(X_1, \dots, X_r)$ .

**Remark 1.4:** The parameters  $\lambda$  and  $d$  in Theorem 1.1 and in Proposition 1.2 need not be fixed, as long as the spectral gap,  $d - \lambda$ , is  $\omega(d/g)$  (where  $g = \omega(1)$  was the lower bound on the girth of  $G$ ). For the sake of simplicity, we prove the case of fixed  $d, \lambda$ , and later describe the required adjustments for the general case.

### 1.3 Stronger results for high-girth expanders

Letting  $N_t$  continue to denote the number of vertices visited  $t$  times by a non-backtracking random walk of length  $n$  on  $G$ , as in (1), consider the case where  $t$  is no longer fixed. In case we want to extend the result of Theorem 1.1 for values of  $t$  which depend on  $n$ , and approximate  $N_t$  by  $n/(et!)$  uniformly over all  $t$ , a behavior analogous to the balls and bins model, we need to assume a larger girth. As noted in [2], there are  $d$ -regular expander graphs with girth  $g$ , where a typical non-backtracking random walk visits some vertex  $\Omega(\log n/g)$  times. Therefore, the girth should be at least  $\Omega(\log \log n)$  to allow the number of visits to exhibit a Poisson distribution for all  $t$ .

Indeed, an  $\Omega(\log \log n)$  girth suffices in order to approximate the above number of visits uniformly over all values of  $t$  up to the asymptotically maximal number of visits:

**Theorem 1.5.** *Let  $G$  be as in Theorem 1.1. If  $g > 10 \log_{d-1} \log n$  then for every fixed  $\delta > 0$ , the following holds with high probability:*

$$\begin{aligned} \left| \frac{N_t}{n/(et!)} - 1 \right| &\leq \frac{1}{\log \log \log n} && \text{for all } t < F(1 - \delta) , \\ N_t = 0 &&& \text{for all } t > F(1 + \delta) , \end{aligned} \tag{2}$$

where  $\{N_t\}$  are the variables defined in (1) and  $F(x) = \left(1 + x \frac{\log \log \log n}{\log \log n}\right) \frac{\log n}{\log \log n}$ .

The above theorem is also related to the notion of conserving random bits, which was mentioned earlier (where the goal was to simulate a distribution which had an exponentially small probability of avoiding a set, using a linear number of random bits). Indeed, using  $\Theta(n)$  random bits, it is possible to simulate a distribution which resembles the resulting distribution of throwing  $n$  balls to  $n$  bins uniformly and independently, as opposed to the naive approach, which requires  $n \log n$  random bits.

Theorem 1.5 also immediately gives the result of [2] regarding the maximal number of visits that a non-backtracking random walk makes to a single vertex, with an improved threshold window, replacing the  $o\left(\frac{\log n}{\log \log n}\right)$  error term by  $o\left(\frac{(\log n)(\log \log \log n)}{(\log \log n)^2}\right)$ :

**Corollary 1.6.** *For any fixed  $d \geq 3$  and fixed  $\lambda < d$  the following holds: if  $G$  is an  $(n, d, \lambda)$  graph whose girth is larger than  $10 \log_{d-1} \log n$ , then the maximal number of visits to a single vertex made*

by a non-backtracking random walk of length  $n$  on  $G$  is with high probability

$$\left(1 + (1 + o(1)) \frac{\log \log \log n}{\log \log n}\right) \frac{\log n}{\log \log n},$$

where the  $o(1)$ -term tends to 0 as  $n \rightarrow \infty$ .

The proof of Theorem 1.5 follows from a result analogous to Proposition 1.2, however, since we require an estimate on the rate of convergence to Poisson variables, we require an extended form of a multivariate Brun's Sieve (Proposition 3.1). The proof may be applied to a more general setting, where the set of visited vertices has size depending on  $n$ , and the walk is of length  $\omega(n)$ . However, for the sake of simplicity, we work in the setting of Theorem 1.5, that is, a fixed set of vertices and a walk of length  $\Theta(n)$ , as stated in the following proposition:

**Proposition 1.7.** *Let  $G$  be a graph as in Theorem 1.1. For some fixed  $r$  and  $\mu > 0$ , let  $v_1, \dots, v_r$  denote vertices of  $G$  whose pairwise distances are at least  $g$ . Let  $\widetilde{W}$  be a non-backtracking random walk of length  $m = \mu n$  on  $G$  starting from  $v_1$ , and  $X_i$  be the number of visits that  $\widetilde{W}$  makes to  $v_i$ . If  $g \geq c \log_{d-1} \log n$  for some fixed  $c > 6$  then*

$$\left| \frac{\Pr[\bigcap_{i=1}^r X_i = t_i]}{\prod_{i=1}^r \Pr[Z = t_i]} - 1 \right| \leq O\left((\log n)^{\frac{6-c}{4}}\right) \text{ for all } t_1, \dots, t_r \in \{0, 1, \dots, \lfloor \log n \rfloor\},$$

where  $Z \sim \text{Po}(\mu)$  and the  $o(1)$ -term tends to 0 as  $n \rightarrow \infty$ .

**Remark 1.8:** As in the case of Proposition 1.2 (see Remark 1.3), the requirement on the girth in Proposition 1.7 may be replaced with the assumption that  $\{v_1, \dots, v_r\}$  are not contained in a closed nontrivial walk of length smaller than  $g$ .

## 1.4 Organization

The rest of the paper is organized as follows: in Section 2 we prove Theorem 1.1 and Proposition 1.2 concerning expanders with a non-fixed girth. In Section 3 we formulate and prove the multivariate version of Brun's Sieve which specifies the rate of convergence to the limiting distribution. This version is subsequently used in Section 4 to prove the above mentioned Theorem 1.5 and Proposition 1.7. The final section, Section 5, is devoted to concluding remarks and some open problems.

# 2 A Poisson approximation for expanders with a non-fixed girth

## 2.1 Proof of Proposition 1.2

For the simpler goal of proving Poisson convergence without estimating its rate, we will need the following known results. The well known univariate version of Brun's Sieve states the following:

**Theorem 2.1** (Brun's Sieve). *Let  $X = X(n)$  be a sum of indicator variables, and let  $\mu > 0$ . If for every  $r$ ,  $\lim_{n \rightarrow \infty} \mathbb{E} \binom{X}{r} = \mu^r / r!$ , then  $X \xrightarrow{d} Z$  as  $n \rightarrow \infty$ , where  $Z \sim \text{Po}(\mu)$ .*

See, e.g., [4] (pp. 119-122) for the derivation of this result from the Inclusion-Exclusion Principle and the Bonferroni inequalities [6], as well as for several applications. A multivariate version of Brun's Sieve is stated in [9], and a proof of the multivariate version by induction (using Brun's Sieve once for the base of the induction, and once more for the induction step) appears in [15]:

**Theorem 2.2** (Multivariate Brun's Sieve). *Let  $X_1 = X(n), \dots, X_r = X_r(n)$  denote sums of indicator variables, and let  $\mu_1, \dots, \mu_r > 0$ . If for every  $t_1, \dots, t_r$ ,  $\lim_{n \rightarrow \infty} \mathbb{E} [\prod_{i=1}^r \binom{X_i}{t_i}] = \prod_{i=1}^r \mu_i^{t_i} / t_i!$ , then  $(X_1, \dots, X_r) \xrightarrow{d} (Z_1, \dots, Z_r)$ , where the  $Z_i$ -s are independent Poisson variables,  $Z_i \sim \text{Po}(\mu_i)$ .*

While a stronger version of Brun's Sieve is proved in Section 3 (Proposition 3.1), Theorem 2.2 suffices for the proof of Proposition 1.2, where the rate of convergence to the Poisson distribution is not specified. Indeed, letting  $X_1, \dots, X_r$  and  $\mu$  be as in Proposition 1.2, we need to prove that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \prod_{i=1}^r \binom{X_i}{t_i} \right] = \prod_{i=1}^r \frac{\mu^{t_i}}{t_i!} \quad \text{for all } t_1, \dots, t_r. \quad (3)$$

Fix integers  $t_1, \dots, t_r$ , and set  $t = \sum_{i=1}^r t_i$ . Let  $(v_1, w_1, \dots, w_m)$  denote the path of the non-backtracking random walk  $\widetilde{W}$ , and for all  $i \in [r]$  and  $j \in [m]$  let  $X_{ij}$  denote the indicator for the event that  $\widetilde{W}$  visits  $v_i$  in position  $j$ ; that is,  $X_{ij} = \mathbf{1}_{\{w_j = v_i\}}$ , and by definition,  $X_i = \sum_{j=1}^m X_{ij}$ . It follows that:

$$\mathbb{E} \left[ \prod_{i=1}^r \binom{X_i}{t_i} \right] = \sum'_{I_1, \dots, I_r} \Pr \left[ \bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1 \right], \quad (4)$$

where  $\sum'$  ranges over  $I_1, \dots, I_r \subset [m]$  with  $|I_i| = t_i$ . We will rewrite the right-hand-side of the above equation. To this end, set

$$L = (\log n)^2, \quad (5)$$

and let  $g = \omega(1)$  be a lower bound for the girth of  $G$ , which satisfies  $g = o(L)$  (such a  $g$  exists by the assumption on  $G$ ). For all  $s \in \{0, \dots, t = \sum t_i\}$ , let  $\mathcal{I}_s$  denote the collection of  $r$ -tuples  $(I_1, \dots, I_r)$  where:

- $I_1, \dots, I_r$  are *disjoint* subsets of  $[m]$  and  $|I_j| = t_j$  for all  $j$ .
- There are precisely  $s$  consecutive elements of  $\cup_j I_j \cup \{0\}$  whose distance is less than  $L$ .

In other words:

$$\mathcal{I}_s = \left\{ (I_1, \dots, I_r) : \begin{array}{l} \cup_j I_j = \{x_1, x_2, \dots, x_t\} \subset [m], \quad x_0 = 0, \quad |I_j| = t_j \text{ for all } j, \\ x_{i-1} < x_i \text{ for all } i, \text{ and } |\{1 \leq i \leq t : x_i - x_{i-1} < L\}| = s \end{array} \right\}. \quad (6)$$

Notice that the events  $X_{ij} = 1$  and  $X_{i'j} = 1$  are disjoint for  $i \neq i'$ . Therefore, (4) takes the following form:

$$\mathbb{E} \left[ \prod_{i=1}^r \binom{X_i}{t_i} \right] = \sum_{s=0}^t \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr \left[ \bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1 \right]. \quad (7)$$

The following claim estimates the probability that a non-backtracking random walk, starting from some given  $v_i$ , would end up in some given  $v_j$  after less than  $L$  steps, as well as after some given  $k \geq L$  number of steps. Here and in what follows, the notation  $\tilde{P}_{uv}^{(k)}$  denotes the probability that a non-backtracking random walk of length  $k$ , which starts in  $u$ , ends in  $v$ .

**Claim 2.3.** *Let  $G$  be as above, and define:  $M = \max_{i,j \in [r]} \sum_{k < L} \tilde{P}_{v_i v_j}^{(k)}$ . Then  $M = o(1)$  and  $\tilde{P}_{v_i v_j}^{(k)} = \frac{1+o(1)}{n}$  for all  $k \geq L$  and  $i, j \in [r]$ , where in both cases the  $o(1)$ -term tends to 0 as  $n \rightarrow \infty$ .*

*Proof.* We need a few results on the mixing of non-backtracking random walks, proved in [2]. The *mixing-rate* of a non-backtracking random walk on  $G$  is defined as:

$$\rho(G) = \limsup_{k \rightarrow \infty} \max_{u,v \in V} \left| \tilde{P}_{uv}^{(k)} - \frac{1}{n} \right|^{1/k}. \quad (8)$$

Theorem 1.1 of [2] determines the value of  $\rho$  as a function of  $\lambda$  and  $d$ :

$$\rho = \frac{\psi\left(\frac{\lambda}{2\sqrt{d-1}}\right)}{\sqrt{d-1}}, \text{ where } \psi(x) = \begin{cases} x + \sqrt{x^2 - 1} & \text{If } x \geq 1, \\ 1 & \text{If } 0 \leq x \leq 1. \end{cases} \quad (9)$$

As shown in [2], one can verify that  $\rho \leq \max\{\frac{\lambda}{d}, \frac{1}{\sqrt{d-1}}\}$ , and in our case, as  $\lambda$  and  $d$  are both fixed, so is  $0 < \rho < 1$ . Furthermore, by the proof of the above theorem,

$$\max_{uv} \left| \tilde{P}_{uv}^{(k)} - \frac{1}{n} \right| \leq (1 + o(1))\rho^k, \quad (10)$$

where the  $o(1)$ -term tends to 0 as  $k \rightarrow \infty$ , and is independent of  $n$ . In particular, by the choice of  $L$  to be  $(\log n)^2$ , for every sufficiently large  $n$  we have

$$\tilde{P}_{uv}^{(k)} = \frac{1 + o(1)}{n} \text{ for all } k \geq L \text{ and all } u, v. \quad (11)$$

Take some  $i, j \in [r]$  (not necessarily distinct). By the assumption on the pairwise distances of  $v_1, \dots, v_r$  and the girth of  $G$ , we have  $\tilde{P}_{v_i v_j}^{(k)} = 0$  for all  $k < g$ . On the other hand, (10) and the fact that  $g = \omega(1)$  imply that  $\tilde{P}_{uv}^{(k)} \leq \frac{1}{n} + (1 + o(1))\rho^k$  for all  $k \geq g$ , giving the upper bound:

$$\sum_{k=1}^{L-1} \tilde{P}_{v_i v_j}^{(k)} = \sum_{k=g}^{L-1} \tilde{P}_{v_i v_j}^{(k)} \leq \frac{L-g}{n} + \frac{(1 + o(1))\rho^g}{1 - \rho}.$$

The required result now follows from the fact that  $L = o(n)$ ,  $\rho$  is fixed and  $g = \omega(1)$ . ■

For convenience, when examining some element  $(I_1, \dots, I_r) \in \mathcal{I}_s$ , we use the following notation: denote by  $i_1, \dots, i_s \in [m]$  the  $s$  indices of the  $x_i$ -s which satisfy  $|x_i - x_{i-1}| < L$ , as in (6), the definition of  $\mathcal{I}_s$ . In addition, for every  $i \in [m]$ , let  $v(x_i)$  denote the vertex  $v_j$ , where  $j \in [r]$  is the

single index satisfying  $x_i \in I_j$ . The following holds:

$$\begin{aligned}
& \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr\left[\bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1\right] \\
& \leq \binom{t}{t_1, \dots, t_r} \binom{m}{t-s} \binom{t}{s} \left(\frac{1+o(1)}{n}\right)^{t-s} \sum_{k_1=1}^{L-1} \dots \sum_{k_s=1}^{L-1} \prod_{j=1}^s \tilde{P}_{v(x_{i_{j-1}})v(x_{i_j})}^{(k_j)} \\
& \leq \binom{t}{t_r, \dots, t_1} \binom{m}{t-s} \binom{t}{s} \left(\frac{1+o(1)}{n}\right)^{t-s} M^s. \tag{12}
\end{aligned}$$

Letting  $\xi(s)$  denote the right hand side of (12), it follows that for all  $s < t$ :

$$\frac{\xi(s+1)}{\xi(s)} = \frac{(t-s)^2}{(m-t+s+1)(s+1)} \cdot \frac{n}{1+o(1)} \cdot M = \Theta(M) = o(1).$$

We deduce that

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^r \binom{X_i}{t_i}\right] &= \sum_{s=0}^t \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr\left[\bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1\right] \\
&\leq \sum_{s=0}^t \xi(s) \leq (1+o(1))\xi(0) = (1+o(1)) \binom{t}{t_1, \dots, t_r} \binom{m}{t} \left(\frac{1+o(1)}{n}\right)^t \\
&= (1+o(1)) \frac{\mu^t}{\prod_{i=1}^r t_i!}. \tag{13}
\end{aligned}$$

For the other direction, consider  $\mathcal{J}$ , the collection of all  $r$ -tuples of disjoint subsets of  $[m] \setminus [L]$ ,  $(I_1, \dots, I_r)$ , where  $|I_j| = t_j$  and the pairwise distances of the indices all exceed  $L$ :

$$\mathcal{J} = \left\{ (I_1, \dots, I_r) : \begin{array}{l} \bigcup_j I_j = \{x_1, x_2, \dots, x_t\} \subset \{L+1, \dots, m\}, \\ |I_j| = t_j \text{ for all } j \text{ and } x_i > x_{i-1} + L \text{ for all } i \end{array} \right\}. \tag{14}$$

Since  $\mathcal{J} \subset \bigcup_s \mathcal{I}_s$  and  $\tilde{P}_{v_i v_j}^{(k)} = \frac{1+o(1)}{n}$  for all  $k \geq L$  and  $i, j \in [r]$  (Claim 2.3), we get:

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^r \binom{X_i}{t_i}\right] &\geq \sum_{(I_1, \dots, I_r) \in \mathcal{J}} \left(\frac{1-o(1)}{n}\right)^t = \binom{t}{t_1, \dots, t_r} \binom{m-Lt}{t} \left(\frac{1-o(1)}{n}\right)^t \\
&= (1+o(1)) \frac{\mu^t}{\prod_{i=1}^r t_i!}. \tag{15}
\end{aligned}$$

Inequalities (13) and (15) imply that (3) holds, completing the proof of Proposition 1.2.  $\blacksquare$

**Remark 2.4:** The assumption that  $G$  is an  $(n, d, \lambda)$  graph for some fixed  $d \geq 3$  and fixed  $\lambda$  was exploited solely in Claim 2.3. In fact, the proof holds whenever for some  $L = o(n)$  and  $g = \omega(1)$ ,  $g < L$ , the girth of  $G$  is at least  $g$  and  $\rho^g = o(1)$ . Suppose that  $d \geq 3$  but  $\lambda, d$  are no longer fixed. Recalling that  $\rho \leq \max\{\frac{\lambda}{d}, \frac{1}{\sqrt{d-1}}\}$ , the requirements of Proposition 1.2 may be replaced, for instance, with  $G$  being an  $(n, d, \lambda)$  graph of girth larger than  $g = \omega(1)$ , where  $d - \lambda = \omega(d/g)$ .

## 2.2 Proof of Theorem 1.1

To prove the theorem, we use the estimates given by Proposition 1.2 for the cases  $r = 1, 2$ , and apply a simple second moment argument. The assumptions of the theorem imply that for any two vertices  $u, v \in V$ , whose distance, as well as their distance from  $w_0$ , are all at least  $g$ , we have:

$$\Pr[X_u = t] = \frac{1}{et!} + o(1) , \quad (16)$$

$$\Pr[X_u = X_v = t] = \Pr[X_u = t]^2 + o(1) \text{ for every fixed } t , \quad (17)$$

where the two  $o(1)$ -terms tend to 0 as  $n \rightarrow \infty$ . Let  $g = g(n)$  be such that the girth of  $G$  is at least  $g$ , and in addition,  $g = o(\log n)$ . Let  $N_t$  denote the number of vertices which  $\widetilde{W}$  visits precisely  $t$  times; we wish to obtain an estimate on the probability that  $N_t = (1 + o(1))n/(et!)$ . As  $t$  is fixed, the effect of any  $o(n)$  positions along  $\widetilde{W}$  have on this value is negligible, and we may ignore the set of vertices whose distance from  $w_0$  is less than  $g$ . Therefore, let  $U$  denote the set of vertices whose distance from  $w_0$  is at least  $g$ , and define

$$N'_t = \sum_{u \in U} \mathbf{1}_{\{X_u = t\}} .$$

Since  $|U| \geq n - d(d-1)^{g-1} = (1 - o(1))n$ , we have:

$$\frac{|N_t - N'_t|}{n/(et!)} = o(1) , \quad (18)$$

and thus, showing that  $N'_t = (1 + o(1))n/(et!)$  almost surely will complete the proof. By (16),

$$\mathbb{E}N'_t = (1 + o(1))\frac{n}{et!} ,$$

and denoting by  $\delta(u, v)$  the distance between two vertices  $u, v$ , we deduce the following from (17):

$$\begin{aligned} \text{Var}(N'_t) &\leq \mathbb{E}N'_t + \sum_{u \in U} \sum_{v \in U} \left( \Pr[X_u = X_v = t] - \Pr[X_u = t] \Pr[X_v = t] \right) \\ &\leq \mathbb{E}N'_t + \left( \sum_{u \in U} \sum_{\substack{v \in U \\ \delta(u, v) < g}} \Pr[X_u = t] \right) + o(n^2) \\ &\leq (1 + o(n)) \mathbb{E}N'_t + o(n^2) = o(n^2) . \end{aligned} \quad (19)$$

Chebyshev's inequality now gives:

$$\Pr \left[ \left| N'_t - \frac{n}{et!} \right| = \Omega(n) \right] = O(\text{Var}(N'_t)/n^2) = o(1) ,$$

completing the proof of the theorem. ■

### 3 Multivariate Brun's Sieve with an estimated rate of convergence

Recall that the versions of Brun's Sieve stated in Section 2 (Theorem 2.1 and Theorem 2.2) do not specify the rate of convergence to the Poisson distribution, and furthermore, the inductive proof of the multivariate case (which appears in [15]) gives an undesirable extra dependence of the rate of convergence on the number of variables. We therefore prove the next version of Brun's Sieve, which follows directly from a multivariate version of the Bonferroni inequalities:

**Proposition 3.1.** *Let  $\mathcal{A}_i = \{A_{ij} : j \in [M_i]\}$ ,  $i \in [r]$ , denote  $r$  classes of events, and denote by  $X_i = \sum_{j=1}^{M_i} \mathbf{1}_{A_{ij}}$  the number of events in  $\mathcal{A}_i$  which occur. Suppose that for some integer  $T$  and some choice of  $\varepsilon, s, \mu_1, \dots, \mu_r > 0$  satisfying  $s > \mu$  and  $2\frac{\mu^s}{s!} < \varepsilon < (2re^\mu)^{-2}$ , where  $\mu = \max_i |\mu_i|$ , we have:*

$$\left| \frac{\mathbb{E}[\prod_{i=1}^r \binom{X_i}{t_i}]}{\prod_{i=1}^r \mu_i^{t_i} / t_i!} - 1 \right| \leq \varepsilon \text{ for all } t_1, \dots, t_r \in \{0, 1, \dots, r(T + 2s)\} . \quad (20)$$

Then:

$$\left| \frac{\Pr[\bigcap_{i=1}^r X_i = t_i]}{\prod_{i=1}^r \Pr[Z_i = t_i]} - 1 \right| \leq \varepsilon' \text{ for all } t_1, \dots, t_r \in \{0, \dots, T\} , \quad (21)$$

where  $\varepsilon' = 2 \exp(2 \sum_i \mu_i) \varepsilon + \sqrt{\varepsilon}$  and  $Z_1, \dots, Z_r$  are i.i.d.,  $Z_i \sim \text{Po}(\mu_i)$ .

*Proof.* We need the following known multivariate generalization of the Bonferroni inequalities:

**Theorem 3.2** ([12]). *Let  $\mathcal{A}_i = \{A_{ij} : j \in [M_i]\}$ ,  $i \in [r]$ , denote  $r$  classes of events, and let  $X_i = \sum_{j=1}^{M_i} \mathbf{1}_{A_{ij}}$  denote the number of events in  $\mathcal{A}_i$  which occur. Define:*

$$S^{(i_1, \dots, i_r)} = \mathbb{E}[\prod_j \binom{X_j}{i_j}] = \sum_{\substack{I_1 \subset [M_1] \\ |I_1|=i_1}} \dots \sum_{\substack{I_r \subset [M_r] \\ |I_r|=i_r}} \Pr[\bigcap_{j=1}^r \bigcap_{k \in I_j} A_{jk}] \quad (22)$$

The following holds for all non-negative integers  $m_1, \dots, m_r$ ,  $0 \leq m_j \leq M_j$ , and  $k \geq 0$ :

$$\Lambda(2k + 1) \leq \Pr[\bigcap_i X_i = m_i] \leq \Lambda(2k) , \text{ where:} \quad (23)$$

$$\Lambda(k) = \sum_{t=\sum m_j}^{(\sum m_j)+k} \sum_{\sum i_j=t} (-1)^{t-\sum m_j} \left( \prod_{j=1}^r \binom{i_j}{m_j} \right) S^{(i_1, \dots, i_r)} .$$

As the function  $f(x, k) = \sum_{l=0}^k x^l / l!$  satisfies  $|f(x, k) - e^x| \leq 2\frac{|x|^k}{k!}$  for all  $x$  with  $|x| \leq \frac{k+1}{2}$ , the assumption on  $s$  implies that

$$\left| \sum_{l=0}^k \frac{x^l}{l!} - e^x \right| \leq \varepsilon \text{ for all } k \geq 2s - 1 \text{ and } |x| \leq \mu . \quad (24)$$

Let  $m_1, \dots, m_r \in \{0, \dots, T\}$ , and set  $M = \sum_i m_i$  and

$$p = \Pr[\bigcap_i Z_i = m_i] = \prod_{i=1}^r e^{-\mu_i} \frac{\mu_i^{m_i}}{m_i!} . \quad (25)$$

Notice that  $r(T + 2s) \leq \min_i M_i$ , otherwise we would get the contradiction  $\varepsilon \geq 1$  from (20). According to the notations of Theorem 3.2, by (20) and the facts  $M = \sum_{i=1}^r m_i$  and  $m_i \leq T$ :

$$(1 - \varepsilon) \prod_j \frac{\mu_j^{i_j}}{i_j!} \leq S^{(i_1, \dots, i_r)} \leq (1 + \varepsilon) \prod_j \frac{\mu_j^{i_j}}{i_j!} \text{ for all } i_1, \dots, i_r \in \{0, \dots, M + 2rs\} .$$

Therefore, the following holds:

$$\begin{aligned} \Lambda(k) &= \sum_{t=M}^{M+k} \sum_{\sum i_j=t} (-1)^{t-M} \left( \prod_{j=1}^r \binom{i_j}{m_j} \right) S^{(i_1, \dots, i_r)} \\ &\leq \sum_{t=M}^{M+k} \sum_{\sum i_j=t} \left( \prod_{j=1}^r \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \right) ((-1)^{t-M} + \varepsilon) \\ &= \left( \prod_{j=1}^r \sum_{i_j=m_j}^{M+k} \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} (-1)^{i_j-m_j} \right) + \varepsilon \left( \prod_{j=1}^r \sum_{i_j=m_j}^{M+k} \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \right) \\ &\quad - \sum_{i_1=m_1}^{M+k} \dots \sum_{i_r=m_r}^{M+k} \mathbf{1}_{\sum i_j > M+k} \left( \prod_{j=1}^r \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \right) \left( (-1)^{(\sum i_j)-M} + \varepsilon \right) . \end{aligned} \quad (26)$$

Let  $E_1, E_2, E_3$  denote the final three expressions in (26), that is,  $\Lambda(k) \leq E_1 + E_2 - E_3$ . A similar calculation gives  $\Lambda(k) \geq E_1 - E_2 - E_3$  (with room to spare, as we could have replaced  $E_3$  by a smaller expression by replacing  $\varepsilon$  by  $-\varepsilon$ ). We therefore wish to provide bounds on  $E_1, E_2, E_3$ . For all  $k \geq 2s - 1$  we have:

$$\begin{aligned} \left| 1 - \frac{E_1}{p} \right| &= \left| 1 - \frac{1}{p} \prod_{j=1}^r \sum_{i_j=m_j}^{M+k} \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} (-1)^{i_j-m_j} \right| \\ &= \left| 1 - \frac{1}{p} \prod_{j=1}^r \frac{\mu_j^{m_j}}{m_j!} \sum_{l=0}^{M-m_j+k} \frac{(-\mu_j)^l}{l!} \right| \leq (1 + \varepsilon e^\mu)^r - 1 \leq e^{\sqrt{\varepsilon}/2} - 1 \leq \sqrt{\varepsilon}, \end{aligned} \quad (27)$$

where the first inequality is by (24), as  $\left| 1 - e^{\mu_j} \sum_{l=0}^k \frac{(-\mu_j)^l}{l!} \right| \leq \varepsilon e^\mu$ , and the second follows from the assumption that  $\varepsilon < (2re^\mu)^{-2}$ . Similarly, for  $k \geq 2s - 1$ :

$$E_2 = \varepsilon \prod_{j=1}^r \sum_{i_j=m_j}^{M+k} \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \leq \varepsilon \prod_{j=1}^r e^{\mu_j} \frac{\mu_j^{m_j}}{m_j!} = \varepsilon \exp\left(2 \sum_i \mu_i\right) p . \quad (28)$$

For the bound on  $|E_3|$ , recall that  $M = \sum_i m_i$ , and hence, if  $\sum_j i_j \geq M + 2rs$  we must have  $i_t \geq m_t + 2s$  for some  $t$ . Therefore, for all  $k \geq 2rs - 1$ :

$$\begin{aligned} |E_3| &= \left| \sum_{i_1=m_1}^{M+k} \dots \sum_{i_r=m_r}^{M+k} \mathbf{1}_{\sum i_j > M+k} \left( \prod_{j=1}^r \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \right) \left( (-1)^{(\sum i_j)-M} + \varepsilon \right) \right| \\ &\leq (1 + \varepsilon) \sum_{i_1=m_1}^{M+k} \dots \sum_{i_r=m_r}^{M+k} \mathbf{1}_{\sum i_j > M+k} \prod_{j=1}^r \binom{i_j}{m_j} \frac{\mu_j^{i_j}}{i_j!} \\ &\leq (1 + \varepsilon) \sum_{t=1}^r \sum_{i_t=m_t+2s}^{M+k} \binom{i_t}{m_t} \frac{\mu_t^{i_t}}{i_t!} \left( \prod_{j \neq t} e^{\mu_j} \frac{\mu_j^{i_j}}{i_j!} \right) \leq \varepsilon \exp\left(2 \sum_i \mu_i\right) p , \end{aligned} \quad (29)$$

where the last inequality is by the fact  $\frac{\mu^s}{s!} < \varepsilon/2$ , which implies that

$$(1 + \varepsilon) \sum_{i_t=m_t+2s}^{M+k} \binom{i_t}{m_t} \frac{\mu_t^{i_t}}{i_t!} \leq (1 + \varepsilon) \frac{\mu_t^{m_t}}{m_t!} \sum_{l \geq 2s} \frac{\mu_t^l}{l!} \leq (1 + \varepsilon) \frac{\mu_t^{m_t}}{m_t!} \cdot 2 \frac{\mu_t^{2s}}{(2s)!} \leq \frac{1 + \varepsilon}{2} \varepsilon^2 e^{\mu_t} \leq \frac{\varepsilon}{r} e^{\mu_t} .$$

Altogether, combining (27), (28) and (29) we get the following for  $k \geq 2rs - 1$ :

$$\left| \frac{\Lambda(k)}{p} - 1 \right| \leq \sqrt{\varepsilon} + 2 \exp\left(2 \sum_i \mu_i\right) \varepsilon = \varepsilon' .$$

The proof is completed by the fact that  $\Lambda(2rs) \leq \Pr[\cap_i X_i = m_i] \leq \Lambda(2rs - 1)$ . ■

## 4 A Poisson approximation for high-girth expanders

In this section, we prove Proposition 1.7 and its corollary, Theorem 1.5, which are analogous to Proposition 1.2 and Theorem 1.1, but also provide an estimate on the rate of convergence to the limiting distributions. This is imperative when looking at vertices which are visited  $t$  times, for  $t$  tending to  $\infty$  with  $n$ . The proof of Proposition 1.7 follows the ideas of the proof of Proposition 1.2, where instead of the simple version of Brun's Sieve, we use Proposition 3.1 proved in Section 3.

### 4.1 Proof of Proposition 1.7

Recall that  $g \geq c \log_{d-1} \log n$  for some fixed  $c > 6$ . We need the following definitions:

$$\tau = \min_t \left\{ \left| \tilde{P}_{uv}^{(k)} - \frac{1}{n} \right| \leq \frac{1}{n^2} \text{ for all } u, v \in V \text{ and } k \geq t \right\} . \quad (30)$$

$$T = \lfloor \log n \rfloor , \quad (31)$$

$$h = (\log n)^{3 - \frac{c}{2}} . \quad (32)$$

Recalling (10), for  $k = \Omega(\log n)$  we have  $\left| \tilde{P}_{uv}^{(k)} - \frac{1}{n} \right| \leq n^{-\Omega(1)}$ , giving

$$\tau = O(\log n) . \quad (33)$$

According to the notation of Proposition 3.1, set  $\mu_i = \mu$  for all  $i$ , let  $h$  play the role of  $\varepsilon$ , and define  $h'$  to be the analogue of  $\varepsilon'$ :

$$h' = 2e^{2r\mu} h + \sqrt{h} = (1 + o(1)) (\log n)^{\frac{6-c}{4}} . \quad (34)$$

It follows from Proposition 3.1 that, in order to show that

$$\left| \frac{\Pr[\cap_{i=1}^r X_i = t_i]}{\prod_{i=1}^r \Pr[Z = t_i]} - 1 \right| \leq O(h') \text{ for all } t_1, \dots, t_r \in \{0, \dots, T\} ,$$

it suffices to show that for some  $s$  satisfying  $s > \mu$  and  $2\frac{\mu^s}{s!} < h < (2re^\mu)^{-2}$  we have:

$$\left| \frac{\mathbb{E}[\prod_{i=1}^r \binom{X_i}{t_i}]}{\mu^t / \prod_{i=1}^r t_i!} - 1 \right| \leq O(h) \text{ for all } t_1, \dots, t_r \in \{0, 1, \dots, r(T + 2s)\} . \quad (35)$$

Substituting  $s = T$ , the requirements  $T > \mu$  and  $h < (2re^\mu)^{-2}$  immediately hold for a sufficiently large  $n$ , as  $T = \omega(1)$ ,  $h = o(1)$  and both  $\mu$  and  $r$  are fixed. The requirement  $2\frac{\mu^T}{T!} < h$  holds as well, since  $2\frac{\mu^T}{T!} = \exp(-(1-o(1))(\log n)(\log \log n))$ , and for a sufficiently large  $n$ , this term is clearly smaller than  $h = \exp(-O(\log \log n))$ . Therefore, proving (35) for  $s = T$  would complete the proof of the proposition, that is, we need to show that

$$\left| \frac{\mathbb{E}[\prod_{i=1}^r \binom{X_i}{t_i}]}{\mu^t / \prod_{i=1}^r t_i!} - 1 \right| \leq O(h) \text{ for all } t_1, \dots, t_r \in \{0, 1, \dots, 3rT\}. \quad (36)$$

Let  $t_1, \dots, t_r \in \{0, \dots, 3rT\}$ , and set

$$t = \sum_{i=1}^r t_i \leq 3r^2T = O(\log n).$$

Let  $(v_1, w_1, \dots, w_m)$  denote the path of the non-backtracking random walk  $\widetilde{W}$ , and as before, for all  $i \in [r]$  and  $j \in [m]$  let  $X_{ij}$  denote the indicator for the event that  $\widetilde{W}$  visits  $v_i$  in position  $j$ :

$$X_{ij} = \mathbf{1}_{\{w_j=v_i\}}, \quad X_i = \sum_{j=1}^m X_{ij}.$$

As in the proof of Proposition 1.2, we next define the collection  $\mathcal{I}_s$ , this time for  $L = \tau = O(\log n)$ . For all  $s \in \{0, \dots, t = \sum t_i\}$ , let  $\mathcal{I}_s$  denote the collection of  $r$ -tuples  $(I_1, \dots, I_r)$  of disjoint subsets of  $[m]$ ,  $|I_j| = t_j$ , so that there are precisely  $s$  consecutive elements of  $\cup_j I_j \cup \{0\}$  whose distance is less than  $\tau$ :

$$\mathcal{I}_s = \left\{ (I_1, \dots, I_r) : \begin{array}{l} \cup_j I_j = \{x_1, x_2, \dots, x_t\} \subset [m], \quad x_0 = 0, \quad |I_j| = t_j \text{ for all } j, \\ x_{i-1} < x_i \text{ for all } i, \text{ and } |\{1 \leq i \leq t : x_i - x_{i-1} < \tau\}| = s \end{array} \right\}. \quad (37)$$

The facts that the events  $X_{ij} = 1$  and  $X_{i'j} = 1$  are disjoint for  $i \neq i'$  implies that

$$\mathbb{E} \left[ \prod_{i=1}^r \binom{X_i}{t_i} \right] = \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr \left[ \bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1 \right]. \quad (38)$$

By definition (30), for all  $v_i, v_j$  and  $k \geq \tau$ ,  $\widetilde{P}_{v_i v_j}^{(k)} \leq n^{-1} + n^{-2}$ . The following claim estimates the sum of the probabilities  $\widetilde{P}_{v_i v_j}^{(k)}$  over all  $k < \tau$ .

**Claim 4.1.** *Let  $G$  be as above, and define  $M = \max_{i,j \in [r]} \sum_{k < \tau} \widetilde{P}_{v_i v_j}^{(k)}$ . Then  $M = O\left((\log n)^{1-\frac{\epsilon}{2}}\right)$ .*

*Proof.* Let  $v_i$  and  $v_j$  denote two (not necessarily distinct) elements of  $\{v_1, \dots, v_r\}$ . By the assumption on the pairwise distances of  $v_1, \dots, v_r$  and the girth of  $G$ ,  $\widetilde{P}_{v_i v_j}^{(k)} = 0$  for all  $k < g$ . It remains to estimate  $\sum_{k=g}^{\tau-1} \widetilde{P}_{v_i v_j}^{(k)}$ .

Set  $\ell = \lfloor \frac{g-1}{2} \rfloor$ , and consider  $U$ , the set of vertices of  $G$  whose distance from  $v_j$  is at most  $\ell$ . Since the girth of  $G$  is at least  $g$ , the induced subgraph of  $G$  on  $U$  is isomorphic to a  $d$ -regular tree. Next, examine a non-backtracking walk of length  $k \geq \ell$  from  $u$  to  $v$ ; crucially, since the walk cannot

backtrack, the last  $\ell$  vertices along the walk must form a path from a leaf of the above mentioned tree, up to its root. In each of the  $\ell$  steps along this path there is a probability of  $1 - \frac{1}{d-1}$  to stray from the path, hence  $\tilde{P}_{uv}^{(k)} \leq (d-1)^{-\ell}$ . Altogether,

$$\sum_{k=g}^{\tau-1} \tilde{P}_{v_i v_j}^{(k)} \leq \frac{\tau - g}{(d-1)^{\lfloor (g-1)/2 \rfloor}} = O\left((\log n)^{1-\frac{\epsilon}{2}}\right),$$

as required. ■

Letting  $i_1, \dots, i_s \in [m]$  denote the  $s$  indices of the  $x_i$ -s which satisfy  $|x_i - x_{i-1}| < \tau$  in the definition (37) of  $\mathcal{I}_s$ , and defining  $v(x_i) = v_j$ , where  $j$  is such that  $x_i \in I_j$ , we have:

$$\begin{aligned} & \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr\left[\bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1\right] \\ & \leq \binom{t}{t_1, \dots, t_r} \binom{m}{t-s} \binom{t}{s} \left(\frac{1+n^{-1}}{n}\right)^{t-s} \sum_{k_1=1}^{\tau-1} \dots \sum_{k_s=1}^{\tau-1} \prod_{j=1}^s \tilde{P}_{v(x_{i_{j-1}})v(x_{i_j})}^{(k_j)} \\ & \leq \binom{t}{t_r, \dots, t_1} \binom{m}{t-s} \binom{t}{s} \left(\frac{1+n^{-1}}{n}\right)^{t-s} M^s. \end{aligned} \quad (39)$$

Let  $\xi(s)$  denote the right hand side of (39). Recalling that  $m = \Theta(n)$  and  $t = O(\log n)$ , the following holds for all  $s < t$ :

$$\frac{\xi(s+1)}{\xi(s)} = \frac{(t-s)^2}{(m-t+s+1)(s+1)} \cdot \frac{n}{1+n^{-1}} \cdot M = O(t^2 M) = O\left((\log n)^{3-\frac{\epsilon}{2}}\right) = o(1),$$

where the last equality is by the fact that  $c > 6$ . Combining this with the fact that, as  $t = n^{o(1)}$ ,  $(1+n^{-1})^t = 1 + O(n^{-1+o(1)}) = 1 + o(h)$ , we get:

$$\begin{aligned} \sum_{s=0}^t \xi(s) & \leq \frac{\xi(0)}{1 - O(t^2 M)} = (1 + O(t^2 M)) \binom{t}{t_1, \dots, t_r} \binom{m}{t} \left(\frac{1+n^{-1}}{n}\right)^t \\ & \leq \left(1 + O\left((\log n)^{3-\frac{\epsilon}{2}}\right)\right) \frac{\mu^t}{\prod_{i=1}^r t_i!} \leq (1 + O(h)) \frac{\mu^t}{\prod_{i=1}^r t_i!}, \end{aligned}$$

and:

$$\mathbb{E}\left[\prod_{i=1}^r \binom{X_i}{t_i}\right] \leq \sum_{s=0}^t \sum_{(I_1, \dots, I_r) \in \mathcal{I}_s} \Pr\left[\bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1\right] \leq (1 + O(h)) \prod_{i=1}^r \frac{\mu^{t_i}}{t_i!}. \quad (40)$$

For the other direction, consider  $\mathcal{J}$ , the collection of all  $r$ -tuples of disjoint subsets of  $[m] \setminus [\tau]$ ,  $(I_1, \dots, I_r)$ , where  $|I_j| = t_j$  and the pairwise distances of the indices all exceed  $\tau$ :

$$\mathcal{J} = \left\{ (I_1, \dots, I_r) : \begin{array}{l} \bigcup_j I_j = \{x_1, x_2, \dots, x_t\} \subset \{\tau+1, \dots, m\}, \\ |I_j| = t_j \text{ for all } j \text{ and } x_{i+1} > x_i + \tau \text{ for all } i \end{array} \right\}. \quad (41)$$

By the definition (30) of  $\tau$ , and the fact that  $(1+n^{-1})^t = 1 + O(n^{-1+o(1)}) = 1 + o(h)$ ,

$$\begin{aligned} \mathbb{E}\left[\prod_{i=1}^r \binom{X_i}{t_i}\right] & \geq \sum_{(I_1, \dots, I_r) \in \mathcal{J}} \Pr\left[\bigcap_{i \in [r]} \bigcap_{j \in I_i} X_{ij} = 1\right] \geq \sum_{(I_1, \dots, I_r) \in \mathcal{J}} \left(\frac{1-n^{-1}}{n}\right)^t \\ & = \binom{t}{t_1, \dots, t_r} \binom{m-\tau t}{t} \left(\frac{1-n^{-1}}{n}\right)^t = \frac{(1-o(h))\mu^t}{\prod_{i=1}^r t_i!}. \end{aligned} \quad (42)$$

Combining (40) and (42), we obtain that (36) holds for all  $0 \leq t_1, \dots, t_r \leq 3rT$ , completing the proof.  $\blacksquare$

## 4.2 Proof of Theorem 1.5

The proof will follow from Proposition 1.7 using a second moment argument, in a manner analogous to Theorem 1.1. Let  $g = 10 \log_{d-1} \log n$ ; by the assumption on  $G$ , Proposition 1.7 implies that for any two vertices  $u, v \in V$ , whose distance, as well as their distance from  $w_0$ , are all at least  $g$ , we have:

$$\left| \frac{\Pr[X_u = t]}{1/(et!)} - 1 \right| = O\left(\frac{1}{\log n}\right) \text{ for all } t \leq \log n, \quad (43)$$

$$\left| \frac{\Pr[X_u = X_v = t]}{1/(et!)^2} - 1 \right| = O\left(\frac{1}{\log n}\right) \text{ for all } t \leq \log n. \quad (44)$$

Let  $t$  be some integer satisfying

$$t \leq \left(1 + c \frac{\log \log \log n}{\log \log n}\right) \frac{\log n}{\log \log n} \text{ for some } c < 1, \quad (45)$$

and let  $N_t$  denote the number of vertices which  $\widetilde{W}$  visits precisely  $t$  times. We wish to obtain an estimate on the probability that  $N_t = (1 + o(1))n/(et!)$ . The above choice of  $t$  implies that:

$$\frac{n}{et!} \geq \exp\left((1 - c - o(1)) \log n \frac{\log \log \log n}{\log \log n}\right) = \exp\left((1 - c)(\log n)^{1-o(1)}\right). \quad (46)$$

Hence, the effect of any  $(\log n)^{O(1)}$  positions along  $\widetilde{W}$  on this value is negligible, and we may ignore the set of vertices whose distance from  $w_0$  is less than  $g$ . Therefore, let  $U$  denote the set of vertices whose distance from  $w_0$  is at least  $g$ , let  $X_u$  ( $u \in U$ ) denote the number of visits which  $\widetilde{W}$  makes to  $u$ , and let  $N'_t = \sum_{u \in U} \mathbf{1}_{\{X_u = t\}}$ . According to this definition, we have:

$$\frac{|N_t - N'_t|}{n/(et!)} = \exp\left(- (1 - c)(\log n)^{1-o(1)}\right), \quad (47)$$

and it remains to determine the behavior of  $N'_t$ . By (43),

$$\left| \frac{\mathbb{E}N'_t}{n/(et!)} - 1 \right| = \left| \left( \sum_{u \in U} \frac{\Pr[X_u = t]}{n/(et!)} \right) - 1 \right| = O(1/\log n),$$

and we deduce from (46) that

$$\mathbb{E}N'_t = (1 - o(1)) \frac{n}{et!} = \Omega\left(\exp\left((1 - c)(\log n)^{1-o(1)}\right)\right).$$

Furthermore, denoting by  $\delta(u, v)$  the distance between two vertices  $u, v$ , the following holds:

$$\begin{aligned} \text{Var}(N'_t) &\leq \mathbb{E}N'_t + \sum_{u \in U} \sum_{v \in U} \left( \Pr[X_u = X_v = t] - \Pr[X_u = t] \Pr[X_v = t] \right) \\ &\leq \mathbb{E}N'_t + \left( \sum_{u \in U} \sum_{\substack{v \in U \\ \delta(u, v) < g}} \Pr[X_u = t] \right) + \left( \sum_{u \in U} \sum_{\substack{v \in U \\ \delta(u, v) \geq g}} O\left(\frac{1}{\log n}\right) \Pr[X_u = t]^2 \right) \\ &\leq \left(1 + (\log n)^{O(1)}\right) \mathbb{E}N'_t + O\left(\frac{(\mathbb{E}N'_t)^2}{\log n}\right) = O\left(\frac{(\mathbb{E}N'_t)^2}{\log n}\right). \end{aligned} \quad (48)$$

Set  $h = h(n) = \log \log \log n$ . Applying Chebyshev's inequality gives the following:

$$\Pr \left[ \left| \frac{N'_t}{n/(et!)} - 1 \right| \geq \frac{1}{h} \right] \leq \Pr \left[ |N'_t - \mathbb{E}N'_t| \geq \frac{1}{2h} \cdot \frac{n}{et!} \right] = O \left( \frac{h^2}{\log n} \right),$$

and summing this probability for all  $t$  in the range specified in (45) (containing  $O(\frac{\log n}{\log \log n})$  values) we deduce that with high probability:

$$\left| \frac{N'_t}{n/(et!)} - 1 \right| \leq \frac{1}{\log \log \log n} \text{ for all } t \leq \left( 1 + c \frac{\log \log \log n}{\log \log n} \right) \frac{\log n}{\log \log n}. \quad (49)$$

Recalling the relation between  $N'_t, N_t$  in (47), we obtain that when replacing  $N'_t$  by  $N_t$ , (49) holds as well. It remains to show that with high probability,  $N_t = 0$  for all  $t > t_0$ , where

$$t_0 = \left( 1 + c \frac{\log \log \log n}{\log \log n} \right) \frac{\log n}{\log \log n} \text{ for some } c > 1. \quad (50)$$

Let  $u$  be a vertex of  $G$ , and let  $X_u$  denote the number of visits that  $\widetilde{W}$  makes to  $u$ . Consider  $\widetilde{W}'$ , a non-backtracking random walk of length  $n$  on  $G$  starting from  $u$ . Proposition 1.7 (for the case of one variable  $v_1 = u$ ) implies that:

$$\Pr[X'_u = t] = \frac{1 + o(1)}{et!} \text{ for all } t \leq \log n,$$

where  $X'_u$  counts the number of visits that  $\widetilde{W}'$  makes to  $u$ . Clearly, the probability that  $X_u > t_0$  is bounded from above by the probability that  $X'_u \geq t_0$  (as we can always condition on the first visit to  $u$ ). Therefore:

$$\Pr[X_u > t_0] \leq \Pr[X'_u \geq t_0] = 1 - \sum_{l < t_0} \Pr[X'_u = l] \leq \frac{1 + o(1)}{et_0!}.$$

We deduce that the expected number of vertices with more than  $t_0$  visits satisfies:

$$\mathbb{E}|\{u \in V : X_u > t_0\}| \leq (1 + o(1)) \frac{n}{et_0!} = \exp \left( (1 - c)(\log n)^{1 - o(1)} \right) = o(1).$$

This completes the proof of the theorem. ■

## 5 Concluding remarks and open problems

- We have shown that the distribution of the number of visits at vertices made by a non-backtracking random walk of length  $n$  on  $G$ , a regular  $n$ -vertex expander of fixed degree and large girth, tends to a Poisson distribution with mean 1. Furthermore, if the girth is  $\Omega(\log \log n)$  we prove the following concentration result: with high probability, the number of vertices visited  $t$  times is  $(1 + o(1)) \frac{n}{et!}$  uniformly over all  $t \leq (1 - o(1)) \frac{\log n}{\log \log n}$ , and 0 for all  $t \geq (1 + o(1)) \frac{\log n}{\log \log n}$  (in fact, the threshold window we get is sharper by a factor of  $\frac{\log \log \log n}{\log \log n}$ ). In particular, we obtain an alternative proof for the typical maximal number of visits to a vertex in the above walk, and (slightly) improve upon the estimate of this maximum in [2].

- The above result implies that the distribution of the visits at vertices made by a non-backtracking random walk of length  $n$  on an  $n$ -vertex expander of high girth is asymptotically the same as the result of throwing  $n$  balls to  $n$  bins independently and uniformly at random.
- The main tool in the proof is an extended version of Brun’s Sieve, which includes error estimates and may be of independent interest. Combining this result with some additional ideas, we show that the variables counting the number of visits to vertices, which are sufficiently distant apart, are asymptotically independent Poisson variables. This implies the required result on the overall distribution of the number of visits at vertices.
- Theorem 1.5 characterizes the distribution of visits at vertices in non-backtracking random walk on a high-girth regular expander. For such a graph on  $n$  vertices, the values  $N_t/n$  converge to  $(1 + o(1))/(e^t)$ , where  $N_t$  is the number of vertices visited precisely  $t$  times in a walk of length  $n$  as above. Moreover, we show that the above convergence of  $\{N_t/n\}$  is uniform over all values of  $t$  up to roughly  $\frac{\log n}{\log \log n}$ , after which  $N_t$  is almost surely 0.

It seems interesting to investigate this distribution,  $(\frac{N_0}{n}, \frac{N_1}{n}, \dots, \frac{N_n}{n})$ , as a parameter of general vertex transitive graphs, and determine it for additional families of such graphs.

- Corollary 1.6 determines that the maximum number of visits to a vertex, made by a typical non-backtracking random walk of length  $n$  on a high-girth  $n$ -vertex regular expander, is  $(1 + o(1))\frac{\log n}{\log \log n}$  (with an improved error term compared to the results of [2]).

For which other families of  $d$ -regular graphs, with  $d \geq 3$ , is this maximum  $\Theta(\frac{\log n}{\log \log n})$ ?

- The “random setting”, where  $n$  balls are thrown to  $n$  bins, uniformly at random, results in a maximal load of  $(1 + o(1))\frac{\log n}{\log \log n}$ ; it seems plausible that this bound is the smallest maximal load possible for a non-backtracking walk on any regular graph of degree at least 3. Is it indeed true that for any  $n$ -vertex  $d$ -regular graph with  $d \geq 3$ , a non-backtracking random walk of length  $n$  visits some vertex at least  $(1 + o(1))\frac{\log n}{\log \log n}$  times almost surely?

**Acknowledgement** We would like to thank Itai Benjamini for useful discussions.

## References

- [1] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff, Random walks, universal traversal sequences, and the complexity of maze problems, in: Proc. of 20th FOCS (1979), pp. 218-223.
- [2] N. Alon, I. Benjamini, E. Lubetzky and S. Sodin, Non-backtracking random walks mix faster, to appear.
- [3] N. Alon and M. Capalbo, Optimal universal graphs with deterministic embedding, to appear.

- [4] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, Wiley, New York, 2000.
- [5] G. Barnes and U. Feige, Short random walks on graphs, Proc. of 25th STOC (1993), pp 728-737. Also in SIAM J. Disc. Math, 9(1), 19-28, 1996.
- [6] C.E. Bonferroni, Teoria statistica delle classi e calcolo delle probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8 (1936), pp. 3-62.
- [7] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol I, Wiley, 1968.
- [8] C. Godsil and G. Royle, *Algebraic Graph Theory*, volume 207 of Graduate Text in Mathematics, Springer, New York, 2001.
- [9] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, John Wiley and Sons, 2000.
- [10] M. Krivelevich and B. Sudakov, Pseudo-random graphs, More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer, 2006, 199-262.
- [11] 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.
- [12] R.M. Meyer, Note on a ‘Multivariate’ Form of Bonferroni’s Inequalities, The Annals of Mathematical Statistics, Vol. 40, No. 2 (1969), pp. 692-693.
- [13] O. Reingold, Undirected ST-connectivity in log-space, STOC 2005, pp. 376-385.
- [14] A. Sinclair, Improved Bounds for Mixing Rates of Markov Chains and Multicommodity Flow, Combinatorics, Probability and Computing 1 (1992), pp. 351-370.
- [15] N.C. Wormald, Models of random regular graphs, in: Surveys in Combinatorics, (ed. J.D. Lamb, D.A. Preece), London Mathematical Society Lecture Note Series, vol 276, pp. 239-298, Cambridge University Press, Cambridge, 1999.