



# A New Berry-Esseen Theorem for Expander Walks\*

Louis Golowich<sup>†</sup>  
UC Berkeley  
Berkeley, CA USA

[lgolowich@berkeley.edu](mailto:lgolowich@berkeley.edu)

May 3, 2023

## Abstract

We prove that the sum of  $t$  boolean-valued random variables sampled by a random walk on a regular expander converges in total variation distance to a discrete normal distribution at a rate of  $O(\lambda/t^{1/2-o(1)})$ , where  $\lambda$  is the second largest eigenvalue of the random walk matrix in absolute value. To the best of our knowledge, among known Berry-Esseen bounds for Markov chains, our result is the first to show convergence in total variation distance, and is also the first to incorporate a linear dependence on expansion  $\lambda$ . In contrast, prior Markov chain Berry-Esseen bounds showed a convergence rate of  $O(1/\sqrt{t})$  in weaker metrics such as Kolmogorov distance.

Our result also improves upon prior work in the pseudorandomness literature, which showed that the total variation distance is  $O(\lambda)$  when the approximating distribution is taken to be a binomial distribution. We achieve the faster  $O(\lambda/t^{1/2-o(1)})$  convergence rate by generalizing the binomial distribution to discrete normals of arbitrary variance. We specifically construct discrete normals using a random walk on an appropriate 2-state Markov chain. Our bound can therefore be viewed as a regularity lemma that reduces the study of arbitrary expanders to a small class of particularly simple expanders.

---

\*To appear in the 55th Annual ACM Symposium on Theory of Computing (STOC 2023). These results first appeared in the author's undergraduate thesis [[Gol22](#)].

<sup>†</sup>Currently supported by an NSF Graduate Fellowship. This work was done while the author was at Harvard University.

# 1 Introduction

The Berry-Esseen Theorem [Ber41, Ess42] is the quantitative version of the central limit theorem, and states that the sum of  $t$  sufficiently independent random variables converges to a normal distribution at a rate of  $O(1/\sqrt{t})$ . In this paper, we prove a new Berry-Esseen theorem for random walks on expander graphs. Specifically, we show that the sum of  $t$  boolean-valued random variables sampled using a random walk on a regular expander graph converges in total variation distance to an appropriate discrete normal distribution at a rate of  $O(\lambda/t^{1/2-o(1)})$ , where  $\lambda$  is the second largest eigenvalue of the random walk matrix in absolute value. This bound recovers the  $O(1/\sqrt{t})$  convergence rate in the classical Berry-Esseen theorem (up to a  $t^{o(1)}$  error), and simultaneously incorporates a linear dependence on  $\lambda$ . To the best of our knowledge, prior known Berry-Esseen theorems for Markov chains with spectral gap  $1 - \lambda$  did not achieve the linear dependence on  $\lambda$  in our bound. Furthermore, our bound applies to total variation distance, whereas prior Markov chain Berry-Esseen bounds only considered weaker metrics such as Kolmogorov distance.

## 1.1 Main result

This section describes our main result, and interprets it both as a Berry-Esseen theorem and as a regularity lemma for expander walks. The formal result statement is given in Section 3.

In words, our main result says that if  $G$  is a  $\lambda$ -spectral expander graph with vertex labeling  $\text{val} : V(G) \rightarrow \{0, 1\}$ , then the sum  $\sum_{i \in [t]} \text{val}(\text{RW}_G^t)_i$  of the labels from a length- $t$  random walk on  $G$  converges in total variation distance to an appropriate discrete normal distribution  $\mathcal{N}_{\sigma^2}^t$  at a rate of almost  $O(\lambda/\sqrt{t})$ .

**Theorem 1** (Main result; informal statement of Theorem 14). *Let  $G$  be a regular  $\lambda$ -spectral expander graph with vertex labeling  $\text{val} : V(G) \rightarrow \{0, 1\}$ , where  $\lambda$  is less than some sufficiently small constant. Then for all  $t \in \mathbb{N}$ ,*

$$d_{\text{TV}} \left( \sum_{i \in [t]} \text{val}(\text{RW}_G^t)_i, \mathcal{N}_{\sigma^2}^t \right) \leq \frac{\lambda}{t^{1/2-o_p(1)}} \quad (1)$$

for an appropriate discrete normal distribution  $\mathcal{N}_{\sigma^2}^t$ , where the  $o_p(1)$  term above approaches 0 as  $t \rightarrow \infty$  for every fixed value of the weights  $p_b = |\text{val}^{-1}(b)|/n$  of the labeling  $\text{val} : V(G) \rightarrow \{0, 1\}$ .

The parameter  $\sigma^2$  above denotes the *asymptotic variance* so that

$$\sigma^2 = \lim_{t \rightarrow \infty} \frac{\text{Var}(\sum \text{val}(\text{RW}_G^t))}{t} = \lim_{t \rightarrow \infty} \frac{\text{Var}(\mathcal{N}_{\sigma^2}^t)}{t},$$

where we denote  $\sum \text{val}(\text{RW}_G^t) = \sum_{i \in [t]} \text{val}(\text{RW}_G^t)_i$ .

For intuition, observe that when  $\lambda = 0$ , then  $G$  is a complete graph with self loops so that  $\sum \text{val}(\text{RW}_G^t) = \text{Bin}(t, p_1)$ , and the right hand side of (1) vanishes. Thus we must have  $\mathcal{N}_{\sigma^2}^t = \text{Bin}(t, p_1)$  with  $\sigma^2 = p_0 p_1$ . Therefore our discrete normal  $\mathcal{N}_{\sigma^2}^t$  provides a generalization of the binomial distribution to arbitrary variances for any given  $t, p$ .

We prove Theorem 1 for a general class of discrete normals satisfying a set of axioms (see Section 2.5). We then present a concrete instantiation of such a discrete normal family satisfying these axioms, given by letting  $\mathcal{N}_{\sigma^2}^t$  be the distribution  $\sum \text{val}(\text{RW}_{G_{\nu, p}}^t)$  for an appropriate  $\nu = \nu(p, \sigma^2)$ ,

where  $G_{\nu,p}$  denotes the  $\nu$ -sticky,  $p$ -biased random walk. This walk is the 2-state<sup>1</sup> Markov chain with stationary distribution  $p$  that moves from state  $b \in \{0, 1\}$  to state  $b' \in \{0, 1\}$  with probability  $(1 - \nu) \cdot p_{b'} + \nu \cdot \mathbb{1}_{b=b'}$ .

Our Berry-Esseen result can also be viewed as a regularity lemma for expander walks. The  $\nu$ -sticky walk has expansion  $\lambda(G_{\nu,p}) = \nu$ , and therefore provides a canonical “simplest” example of a  $\nu$ -spectral expander. Indeed, Guruswami and Kumar [GK21] introduced  $G_{\nu,p}$  for the case  $p_0 = p_1 = 1/2$  as a model for studying expander walks, but with no formal way to derive results about general expanders from  $G_{\nu,p}$ . Our bound in Theorem 1 with  $\mathcal{N}_{\sigma^2}^t = \Sigma \text{val}(\text{RW}_{G_{\nu(p,\sigma^2),p}}^t)$  shows that the distribution of the sum of the labels from a  $t$ -step random walk on an arbitrary  $\lambda$ -spectral expander graph  $G$  is approximated by the analogous distribution for  $G_{\nu,p}$ , up to a total variation error that vanishes as  $\lambda \rightarrow 0$  or as  $t \rightarrow \infty$ . That is, our bound reduces the study of the random walk on an arbitrary expander  $G$  to the study of a much simpler sticky walk  $G_{\nu,p}$ .

Though we do not present any direct applications of our Berry-Esseen bound, similar prior results have found various applications in computer science. Prior limit theorems that approximate sums of random variables in total variation distance have yielded applications to learning theory and algorithmic game theory [Das08, DP09, DDS12, DDO<sup>+</sup>13]. At a high level, these applications make use of the regularity lemma interpretation described above. Specifically, they use appropriate limit theorems to obtain concise descriptions of target distributions. The Berry-Esseen theorem also has applications to pseudorandomness, e.g. [MZ10, GMRZ11]. Further connections of our work to pseudorandomness are discussed in Section 1.3 below.

The following two sections place our results in context within the literature. While our main result is a Berry-Esseen bound (or viewed alternatively, a regularity lemma), our proof techniques evolved from a line of work on the pseudorandomness of expander walks [GK21, CPTS21, CMP<sup>+</sup>22, GV22a], and our work also has implications for this area. We therefore first compare our result to prior Berry-Esseen bounds, and then describe the pseudorandomness implications.

## 1.2 Comparison to prior Berry-Esseen bounds

Recall that the classic Berry-Esseen Theorem [Ber41, Ess42] states that the sum of  $t$  independent random variables with bounded 2nd and 3rd moments converges in Kolmogorov distance to a normal distribution, with the error decaying at a rate of  $O(1/\sqrt{t})$ . Markov chain Berry-Esseen theorems show that the same rate of convergence holds when the variables may be correlated according to a Markov chain. Various versions of these results have been shown under different convergence metrics and conditions on the variables and the Markov chain.

In particular, there are known Markov chain Berry-Esseen theorems that show convergence in Kolmogorov distance to a normal distribution at a rate of  $O(1/\sqrt{t})$  (e.g. [Bol80, Bol82, HP10, Klo19]). Some of these results (e.g. [Bol80, Klo19]) do apply to Markov chains with a discrete state space, as in our setting. Yet convergence in Kolmogorov distance is weaker than convergence in total variation distance. Indeed, no sequence of discrete random variables can converge to a (continuous) normal distribution in total variation distance.

Berry-Esseen bounds in total variation distance for discrete random variables have been shown when the variables are not correlated according to a Markov chain. In this case, the approximating distribution is taken to be some discrete analogue of the normal distribution, such as a compound

---

<sup>1</sup>To express  $G_{\nu,p}$  as a regular graph, we let the graph have arbitrarily large vertex set  $V = V_0 \sqcup V_1$ , but specify that all  $|V_b| = p_b|V|$  vertices within each  $V_b$  are interchangeable (see Section 2.4).

Poisson [BX99], translated Poisson [Rö07], binomial [Rö06], and a histogram discretization of the continuous normal [Fan14]. While these works primarily consider the iid case, some of them, such as Röllin [Rö07, Rö06] as well as Barbour and Xia [BX99], also provide total variation Berry-Esseen bounds for integer-valued random variables that may have some dependencies among the variables. However, the permitted dependencies among the variables in such prior work are not general enough to apply in our setting of random walks on expanders.

In contrast, our expander walk Berry-Esseen bound applies when the variables  $\text{val}(\text{RW}_G^t)_i$  are correlated according to a random walk on a  $\lambda$ -spectral expander  $G$ , which is a Markov chain with spectral gap  $1 - \lambda$ . Our total variation bound of  $\lambda/t^{1/2-o(1)}$  for the rate of convergence unifies the ordinary  $O(1/\sqrt{t})$  Berry-Esseen convergence rate (up to a  $t^{o(1)}$  loss) with the linear dependence on spectral expansion  $\lambda$  from [CMP<sup>+</sup>22, GV22a] (see below). As described above, to the best of our knowledge, no such bound of almost  $\lambda/\sqrt{t}$  was previously known, even for the weaker Kolmogorov distance metric.

### 1.3 Implications for the pseudorandomness of expander walks

Our Berry-Esseen bound contributes to a recent line of work [GK21, CPTS21, CMP<sup>+</sup>22, GV22a] studying the extent to which random walks on expander graphs fool symmetric functions. This problem was motivated as a generalization of the observation that expander walks fool the parity function, which plays a central role in Ta-Shma’s breakthrough construction of almost-optimal  $\epsilon$ -balanced codes [TS17].

Let  $G$  be a  $\lambda$ -spectral expander with vertex labeling  $\text{val} : V(G) \rightarrow \{0, 1\}$  that assigns label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the vertices. The length- $t$  random walk on  $G$  is said to  $\epsilon$ -fool a given function  $f$  on  $\{0, 1\}^t$  if

$$d_{\text{TV}}(f(\text{val}(\text{RW}_G^t)), f(\text{val}(\text{RW}_J^t))) \leq \epsilon,$$

where  $J$  denotes the complete graph with self loops on vertex set  $V(G)$  (and we extend  $\text{val}$  to act on sequences component-wise).

The recent work [GK21, CPTS21, CMP<sup>+</sup>22, GV22a] studied the problem of bounding the extent to which expander walks fool symmetric functions  $f$ . As a general symmetric function on  $\{0, 1\}^t$  only depends on the sum of the  $t$  input bits, it is sufficient to consider  $f(a) = \sum_{i \in [t]} a_i$ , and then to bound  $d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \Sigma \text{val}(\text{RW}_J^t))$ . Here by definition  $\Sigma \text{val}(\text{RW}_J^t) = \text{Bin}(t, p_1)$  is simply the binomial distribution.

Cohen et al. [CMP<sup>+</sup>22] and Golowich and Vadhan [GV22a] showed that for every  $\lambda$ -spectral expander  $G$ , the random walk  $O(\lambda)$ -fools all symmetric functions, that is,

$$d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \text{Bin}(t, p_1)) \leq O(\lambda). \tag{2}$$

They also showed that this bound is tight, in the sense that there exist  $\lambda$ -spectral expanders  $G$  for which the left hand side above is  $\Omega(\lambda)$ . Specifically, Golowich and Vadhan [GV22a] show that (2) is tight for the sticky walk  $G = G_{\lambda,p}$  because as  $t \rightarrow \infty$ , the central limit theorem implies that  $\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)$  and  $\text{Bin}(t, p_1)$  converge in Kolmogorov distance to normal distributions whose variances have ratio  $1 + \Theta(\lambda)$ . The main idea of our Berry-Esseen result is to leverage this insight to improve the bound in (2) to  $O(\lambda/t^{1/2-o_p(1)})$ , by replacing the binomial approximating distribution with a discrete normal of appropriate variance.

By improving upon the  $O(\lambda)$  bound in (2), our Berry-Esseen bound in Theorem 1 also helps characterize which symmetric functions  $f$  are  $\epsilon$ -fooled by expander walks for  $\epsilon \ll O(\lambda)$ . Specifically,

our result implies that for all  $\epsilon \geq \lambda/t^{1/2-o_p(1)}$ , a symmetric function  $f$  is  $O(\epsilon)$ -fooled by the random walk on an arbitrary  $\lambda$ -spectral expander if and only if  $f$  is  $O(\epsilon)$ -fooled by the sticky walk  $G_{\nu,p}$  for an appropriate choice of  $\nu$ .

This result helps explain the previously known fact that certain symmetric functions are  $\lambda^{O(1)}/\sqrt{t}$ -fooled by expander walks. For instance, Cohen et al. [CPTS21, CMP<sup>+</sup>22] showed that expander walks  $O_p(\lambda/\sqrt{t})$ -fool all the indicator functions  $f(a) = \mathbb{1}_{\sum_i a_i=j}$  for  $0 \leq j \leq t$ , and  $O_p(\lambda^2/\sqrt{t})$ -fool the threshold function  $f(a) = \mathbb{1}_{\sum_i a_i \geq p_1 t}$ . Our Berry-Esseen result shows that (slightly weaker versions of) these bounds for general  $G$  are implied by the respective bounds on the sticky walk.

Our Berry-Esseen result does *not* explain why some symmetric functions, such as the parity function, are  $e^{-\Omega(t)}$ -fooled by expander walks. Rather, this exponentially small error for the parity function follows from the more general fact that the distribution  $\Sigma \text{val}(\text{RW}_G^t)$  is smooth, in the sense that it has rapidly decaying Fourier tails; see Lemma 4 (shown implicitly in [GV22a]) and the surrounding discussion.

*Remark.* In recent independent and concurrent work (posted after the submission of the undergraduate thesis containing our results [Gol22], and after the submission of this paper to STOC 2023, but before the posting of this paper online), Chiclana and Peres [CP22] showed a local central limit theorem for expander walks, which in particular implies that  $\Sigma \text{val}(\text{RW}_G^t)$  converges in total variation distance to an appropriate discrete normal distribution as  $t \rightarrow \infty$  for any fixed graph  $G$ . However, Chiclana and Peres [CP22] do not obtain a bound on the rate of this convergence, whereas our main result bounds the total variation distance by  $O(\lambda/t^{1/2-o_p(1)})$  uniformly over all graphs  $G$ .

## 1.4 Open questions

Our results lead to the following questions.

- Can the  $t^{o_p(1)}$  factor in our bound (1) be removed?
- As a regularity lemma, our result states that for every  $\lambda$ -spectral expander  $G$ , there exists a 2-state Markov chain  $G'$  such that  $\Sigma \text{val}(\text{RW}_G^t)$  is approximated by  $\Sigma \text{val}(\text{RW}_{G'}^t)$  up to a  $O(\lambda/t^{1/2-o_p(1)})$  total variation error. Can better approximations be achieved by  $k$ -state Markov chains  $G'$  for  $k > 2$ ?
- Are there alternative constructions of discrete normal distributions for which our bound holds? As described in Section 1.1, we prove Theorem 1 for any distribution  $\mathcal{N}_{\sigma^2}^t$  satisfying a set of axioms, and then provide an instantiation of such a distribution  $\mathcal{N}_{\sigma^2}^t$  using the sticky walk. While this instantiation has the advantage of providing a regularity lemma, the sticky walk is itself nontrivial to analyze, as was the focus of Guruswami and Kumar [GK21]. Therefore additional constructions of discrete normals satisfying our axioms would also be of interest.
- Can Theorem 1 be extended to more general labelings  $\text{val} : V(G) \rightarrow [d]$  for  $d > 2$ ? Specifically, Golowich and Vadhan [GV22a] show a generalization of (2) for such  $d$ -ary labelings, so does an analogous generalization of Theorem 1 hold?

## 1.5 Organization

The remainder of this paper is organized as follows. Section 2 describes necessary background and preliminary results. Section 3 provides the formal statement of our main Berry-Esseen bound.

Section 4 outlines the proof of the main result, and Section 5 presents the complete proof.

## 2 Preliminaries

In this section, we describe the necessary background to present our Berry-Esseen bound. Section 2.1 provides the basic notation and problem setup. In Section 2.2, we describe the result of Golowich and Vadhan [GV22a] that bounds how the distribution  $\Sigma \text{val}(\text{RW}_G^t)$  changes when the expanders at some steps in  $\mathcal{G}$  are changed. Our proofs rely on this bound, while our main result strengthens certain implications of it.

In Sections 2.3–2.6, we describe the notion of asymptotic variance as well as some basic properties, and we introduce the family of discrete normals  $\mathcal{N}_{\sigma^2}^t$  that we use to approximate  $\Sigma \text{val}(\text{RW}_G^t)$  for a  $\lambda$ -spectral expander  $G$ . The proofs of results in these sections are standard or follow directly from prior work, and for completeness are provided in Appendix A.

### 2.1 Notation and problem setup

This section introduces the basic notation and problem setup of this paper.

We use the following notation throughout. For  $N \in \mathbb{N}$ , let  $[N] = \{0, \dots, N-1\}$ . For a matrix  $A \in \mathbb{R}^{N \times N}$ , the spectral norm of  $A$  is defined to be  $\|A\| = \max_{x \in \mathbb{R}^N \setminus \{0\}} \|Ax\|/\|x\|$ . A matrix  $W \in [0, 1]^{N \times N}$  is a random walk matrix on  $N$  vertices if the columns of  $W$  sum to 1, so that  $W_{j,i}$  denotes the transition probability from vertex  $i$  to vertex  $j$ . When the dimension  $N$  is clear from context, let  $I \in \mathbb{R}^{N \times N}$  denote the identity matrix. Let  $\vec{1} \in \mathbb{R}^N$  denote the unit vector with all entries equal to  $1/\sqrt{N}$ , and let  $J = \vec{1}\vec{1}^\top \in \mathbb{R}^{N \times N}$  denote the matrix with all entries equal to  $1/N$ . Therefore  $J$  is the random walk matrix for the  $N$ -vertex complete graph with self-loops.

For a regular digraph  $G = (V, E)$  on  $n$  vertices, the spectral expansion is defined as

$$\lambda(G) = \|G|_{\vec{1}^\perp}\| = \max_{x \perp \vec{1}} \frac{\|Gx\|}{\|x\|},$$

where by abuse of notation  $G \in \mathbb{R}^{V \times V}$  also denotes the random walk matrix of the graph  $G$ , so that  $G_{v',v} = w_G(v, v')/\text{deg}_G(v)$ .

Given  $t \in \mathbb{N}$  and a sequence of random walk matrices  $\mathcal{W} = (W_1, \dots, W_{t-1})$  on shared vertex set  $V$ , let  $\text{RW}_{\mathcal{W}}^t$  denote the probability distribution over  $V^t$  obtained by taking a  $t$ -step random walk on  $V$ , where the  $i$ th step is taken according to the transition probabilities in  $W_i$ . Formally, to sample  $(v_0, \dots, v_{t-1}) \sim \text{RW}_{\mathcal{W}}^t$ , the initial vertex  $v_0 \in V$  is chosen uniformly at random, and then for  $1 \leq i \leq t-1$  the vertex  $v_i$  is sampled given  $v_{i-1}$  according to  $\Pr[v_i = v] = (W_i)_{v, v_{i-1}}$ . If all  $W_i$  equal some matrix  $W$ , we let  $\text{RW}_W^t = \text{RW}_{\mathcal{W}}^t$ .

Let  $G = (V, E)$  be a  $\lambda$ -spectral expander with some vertex labeling  $\text{val} : V \rightarrow \{0, 1\}$ , and let  $p_b = |\text{val}^{-1}(b)|/n$ , so that  $p = (p_0, p_1)$  gives the probability distribution of the label of a uniformly random vertex. For  $t \in \mathbb{N}$ , we extend the label function component-wise to  $\text{val} : V^t \rightarrow \{0, 1\}^t$ . Let  $\Sigma \text{val}(\text{RW}_G^t) = \sum_{i \in [t]} \text{val}(\text{RW}_G^t)_i$  denote the sum of the labels from a length- $t$  random walk on  $G$ .

Our goal in this paper is to study the distribution  $\Sigma \text{val}(\text{RW}_G^t)$ . Specifically, our main result shows that  $\Sigma \text{val}(\text{RW}_G^t)$  is approximated by an appropriate discrete normal distribution up to a  $O(\lambda/t^{1/2 - o_p(1)})$  error in total variation distance.

## 2.2 Expander walks $O(\lambda)$ -fool symmetric functions

As described in Section 1.3, our Berry-Esseen result can be viewed as a strengthening of the result of [CMP<sup>+</sup>22, GV22a] that random walks on  $\lambda$ -spectral expanders  $O(\lambda)$ -fool symmetric functions. Golowich and Vadhan [GV22a] in fact showed the following more general result, which we apply in our proof.

**Theorem 2** ([GV22a]). *Fix integers  $t \geq 1$  and  $1 \leq u \leq t - 1$ . Let  $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$  and  $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$  be sequences of regular graphs on a shared vertex set  $V$  such that for all  $i \neq u$  we have  $G_i = G'_i$  with  $\lambda(G_i) = \lambda(G'_i) \leq 1/100$ . Fix a labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the vertices. Then for every  $c \geq 0$ ,*

$$\begin{aligned} & \sum_{j \in [t+1] : |j - p_1 t| \geq c} \left| \Pr[\Sigma \text{val}(\text{RW}_{\mathcal{G}'}^t) = j] - \Pr[\Sigma \text{val}(\text{RW}_{\mathcal{G}}^t) = j] \right| \\ & \leq 4000 \cdot \frac{\|G'_u - G_u\| \cdot e^{-c^2/8t}}{t}. \end{aligned} \tag{3}$$

The corollary below follows by changing the graphs in the sequence  $\mathcal{G}$  to  $J$  one at a time, and applying Theorem 2 at each step.

**Corollary 3** ([GV22a]). *Fix an integer  $t \geq 1$ . Let  $\lambda \leq 1/100$ , and let  $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$  be a sequence of regular  $\lambda$ -spectral expanders on shared vertex set  $V$ . Fix a labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the vertices. Then for every  $c \geq 0$ ,*

$$\begin{aligned} & \sum_{j \in [t+1] : |j - p_1 t| \geq c} \left| \Pr[\Sigma \text{val}(\text{RW}_{\mathcal{G}}^t) = j] - \Pr[\Sigma \text{val}(\text{RW}_J^t) = j] \right| \\ & \leq 4000 \cdot \lambda \cdot e^{-c^2/8t}. \end{aligned}$$

Letting  $c = 0$  in Corollary 3 gives the total variation bound (2) described in Section 1.3, which shows that expander walks  $O(\lambda)$ -fool symmetric functions. The  $c > 0$  case of Theorem 2 and Corollary 3 unites this total variation bound with a tail bound, which strengthens the expander walk Chernoff bound as described in [GV22a]. We apply this tail bound in our proofs to show that we may focus on bounding components of the relevant distributions that lie near the mean.

As described in Section 4 and Section 5, the main idea in the proof of our Berry-Esseen bound is to induct on  $t$  using Theorem 2. Specifically, Theorem 2 implies that if we split a length- $t$  random walk on  $G$  into a sequence of  $\ell$  length- $t/\ell$  independent walks on  $G$  by replacing  $\ell - 1$  of the steps on  $G$  with steps on  $J$ , then the distribution of the sum of the labels is preserved up to a  $O(\ell \cdot \lambda/t)$  total variation error. Thus setting  $\ell = \sqrt{t}$ , we reduce the study of  $\Sigma \text{val}(\text{RW}_G^t)$  to  $\Sigma \text{val}(\text{RW}_G^{\sqrt{t}})$ , from which we apply induction.

Our proof will also use the following “smoothness” lemma that is shown implicitly by Golowich and Vadhan [GV22a] in their proof of Theorem 2.

**Lemma 4.** *For  $\lambda \leq 1/100$ , let  $G = (V, E)$  be a  $\lambda$ -spectral expander with vertex labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the vertices. Then for all  $-\pi \leq \theta \leq \pi$ ,*

$$|\mathbb{E}[e^{-i\theta \Sigma \text{val}(\text{RW}_G^t)}]| \leq e^{-p_0 p_1 t \theta^2 / 20}.$$

Lemma 4 implies that the distribution of  $\Sigma \text{val}(\text{RW}_G^t)$  has rapidly decaying Fourier tails. For instance, setting  $\theta = \pi$  in this lemma recovers the previously known fact that expander walks  $e^{-\Omega_p(t)}$ -fool the parity function.

### 2.3 Asymptotic variance

This section describes asymptotic variance. The results below are standard; proofs are provided in Appendix A for completeness.

**Definition 5.** For a sequence  $X = (X^t)^{t \in \mathbb{N}}$  of probability distributions over  $\mathbb{R}$ , the **asymptotic variance** of  $X$  is

$$\sigma^2(X) = \lim_{t \rightarrow \infty} \frac{1}{t} \text{Var}(X^t).$$

The following formula for the asymptotic variance  $\sigma^2(\Sigma \text{val}(\text{RW}_G^t))$  is well known; it for instance is a special case of the definition of asymptotic variance in Kloeckner [Klo19].

**Lemma 6.** Let  $G = (V, E)$  be a regular graph with labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the  $n$  vertices. Viewing  $\text{val}$  and  $\text{val} - p_1 : V \rightarrow \mathbb{R}$  as vectors in  $\mathbb{R}^V$ , then

$$\sigma^2(\Sigma \text{val}(\text{RW}_G^t)) = p_0 p_1 + 2 \sum_{i=1}^{\infty} \frac{1}{n} (\text{val} - p_1)^\top G^i \text{val}.$$

**Corollary 7.** For a  $\lambda$ -spectral expander  $G$ ,

$$|\sigma^2(\Sigma \text{val}(\text{RW}_G^t)) - p_0 p_1| \leq \frac{2}{1 - \lambda} \cdot \lambda \cdot p_0 p_1.$$

For a  $\lambda$ -spectral expander  $G$ , the following lemma shows a  $O(\lambda p_0 p_1 / t)$  bound on the rate of convergence of  $\text{Var}(\Sigma \text{val}(\text{RW}_G^t)) / t$  to  $\sigma^2(\Sigma \text{val}(\text{RW}_G^t))$ .

**Lemma 8.** Let  $G = (V, E)$  be a regular  $\lambda$ -spectral expander with labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the  $n$  vertices. Then

$$\left| \frac{1}{t} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) - \sigma^2(\Sigma \text{val}(\text{RW}_G^t)) \right| \leq \frac{2}{(1 - \lambda)^2} \cdot \frac{\lambda}{t} \cdot p_0 p_1.$$

### 2.4 Sticky random walk

In this section, we introduce the  $\lambda$ -sticky,  $p$ -biased random walk, which is a particularly simple random walk with spectral expansion  $\lambda$  and label weights  $p$ . We will use the sticky walk to construct discrete normal distributions, and therefore our Berry-Esseen bound can be viewed as a regularity lemma that reduces the study of arbitrary expanders to the much simpler class of sticky walks.

The special case of the sticky walk on  $|V| = 2$  vertices with  $p_0 = p_1 = 1/2$  was studied extensively by Guruswami and Kumar [GK21]. Here we describe a more general sticky walk for arbitrary  $p$ .

**Definition 9.** Fix a vertex set  $V = V_0 \sqcup V_1$  with labeling  $\text{val} : V \rightarrow \{0, 1\}$  given by  $\text{val}(v) = b$  for  $v \in V_b$ , so that  $p_0 = |V_0|/|V|$  and  $p_1 = |V_1|/|V|$ . For subsets  $A, B \subseteq V$ , let  $J_{A,B} \in \mathbb{R}^{A \times B}$  denote the matrix with all entries equal to  $1/|A|$ . For  $0 \leq \lambda \leq 1$ , define the  **$\lambda$ -sticky,  $p$ -biased random walk matrix**  $G_{\lambda,p} \in \mathbb{R}^{V \times V}$  by

$$G_{\lambda,p} = \begin{pmatrix} (p_0 + p_1 \lambda) J_{V_0, V_0} & (p_0 - p_0 \lambda) J_{V_0, V_1} \\ (p_1 - p_1 \lambda) J_{V_1, V_0} & (p_1 + p_0 \lambda) J_{V_1, V_1} \end{pmatrix}.$$



That is,  $G_{\lambda,p}$  treats all vertices within  $V_b$  identically for each  $b = 0, 1$ , and if  $(v, v')$  represents a 1-step random walk on  $G_{\lambda,p}$ , then the transition probabilities are

$$\begin{aligned}\Pr[v' \in V_0 | v \in V_0] &= p_0 + p_1\lambda = (1 - \lambda)p_0 + \lambda \\ \Pr[v' \in V_0 | v \in V_1] &= p_0 - p_0\lambda = (1 - \lambda)p_0 \\ \Pr[v' \in V_1 | v \in V_0] &= p_1 - p_1\lambda = (1 - \lambda)p_1 \\ \Pr[v' \in V_1 | v \in V_1] &= p_1 + p_0\lambda = (1 - \lambda)p_1 + \lambda.\end{aligned}$$

We show that the  $\lambda$ -sticky random walk is indeed a  $\lambda$ -spectral expander.

**Lemma 10.**  $\lambda(G_{\lambda,p}) = \lambda$ .

*Proof.* By definition

$$G_{\lambda,p} = (1 - \lambda)J_{V,V} + \lambda W$$

for

$$W = \begin{pmatrix} J_{V_0,V_0} & 0 \\ 0 & J_{V_1,V_1} \end{pmatrix}.$$

We have  $\|W\| = 1$ , as  $W$  acts as  $J$  on the orthogonal subspaces  $\mathbb{R}^{V_0}$  and  $\mathbb{R}^{V_1}$  of  $\mathbb{R}^V$ . Thus  $\lambda(G_{\lambda,p}) \leq \lambda$ . The opposite inequality follows from the fact that  $p_1\mathbb{1}_{V_0} - p_0\mathbb{1}_{V_1} \in \bar{1}^\perp$  is an eigenvector of  $G_{\lambda,p}$  with eigenvalue  $\lambda$ , as is evident from the decomposition of  $G_{\lambda,p}$  above.  $\square$

Below, we compute the asymptotic variance of the sticky walk.

**Lemma 11.** Define  $G_{\lambda,p}$  and  $\text{val}$  as in Definition 9. Then

$$\sigma^2(\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)) = p_0p_1 \cdot \frac{1 + \lambda}{1 - \lambda}.$$

*Proof.* View  $\text{val}$  and  $\text{val} - p_1 : V \rightarrow \mathbb{R}$  as vectors in  $\mathbb{R}^V$ . Then  $\text{val} - p_1 \in \bar{1}^\perp$  is an eigenvector of  $G_{\lambda,p} = (1 - \lambda)J_{V,V} + \lambda \begin{pmatrix} J_{V_0,V_0} & 0 \\ 0 & J_{V_1,V_1} \end{pmatrix}$  with eigenvalue  $\lambda$ , so by Lemma 6,

$$\begin{aligned}\sigma^2(\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)) &= p_0p_1 + 2 \sum_{i=1}^{\infty} \frac{1}{n} (\text{val} - p_1)^\top G^i (\text{val} - p_1) \\ &= p_0p_1 + 2 \sum_{i=1}^{\infty} \lambda^i \cdot \frac{\|\text{val} - p_1\|^2}{n} \\ &= p_0p_1 + 2 \cdot \frac{\lambda}{1 - \lambda} \cdot p_0p_1 \\ &= p_0p_1 \cdot \frac{1 + \lambda}{1 - \lambda}.\end{aligned}$$

$\square$

## 2.5 Definition of discrete normal families

To formally state our Berry-Esseen bound, we must define the family of discrete normal approximating distributions  $\mathcal{N}_{\sigma^2}^t$ . We take an axiomatic approach here, where we define below the set of conditions that this family of distributions must satisfy for our Berry-Esseen proof. We then prove that such a family can be constructed using the sticky random walk.

**Definition 12.** For  $p = (p_0, p_1)$ ,  $\sigma^2 > 0$ , and  $c = (c_1, c_2, c_3) \in \mathbb{R}_+^3$ , a  $(p, \sigma^2, c)$ -**discrete normal family** is a family<sup>2</sup>  $(\mathcal{N}_{\sigma^2}^t)_{t \in \mathbb{N}}$  of probability distributions over  $\mathbb{Z}$  such that the following conditions hold for all  $t \in \mathbb{N}$ :

1.  $\mathbb{E}[\mathcal{N}_{\sigma^2}^t] = p_1 t$ .
2.  $|\text{Var}(\mathcal{N}_{\sigma^2}^t) - \sigma^2 t| \leq c_1 \cdot |\sigma^2 - p_0 p_1|$ .
3.  $\mathcal{N}_{\sigma^2}^1 = \text{Bern}(p)$ .
4. For all positive integers  $\ell \leq t$  and  $t_0, \dots, t_{\ell-1}$  such that  $\sum_{i \in [\ell]} t_i = t$ , then

$$d_{\text{TV}} \left( \sum_{i \in [\ell]} \mathcal{N}_{\sigma^2}^{t_i}, \mathcal{N}_{\sigma^2}^t \right) \leq \frac{c_2}{2} \cdot \frac{(\ell - 1) \cdot |\sigma^2 / p_0 p_1 - 1|}{t},$$

where the variables  $\mathcal{N}_{\sigma^2}^{t_i}$  in the sum above are independent.

5. For all  $a \geq 0$ ,

$$\sum_{j \in \mathbb{Z}: |j - p_1 t| \geq a} |\Pr[\mathcal{N}_{\sigma^2}^t = j] - \Pr[\text{Bin}(t, p) = j]| \leq c_3 \cdot |\sigma^2 / p_0 p_1 - 1| \cdot e^{-a^2 / 8t}.$$

6. The characteristic function of  $\mathcal{N}_{\sigma^2}^t$  satisfies  $|\mathbb{E}[e^{-i\theta \mathcal{N}_{\sigma^2}^t}]| \leq e^{-p_0 p_1 t \theta^2 / 20}$ .

Note that the constants 8 and 20 in the exponents above could be replaced with generic constants  $c_4$  and  $c_5$ , but we leave them as explicit values to simplify notation.

In Definition 12, it is helpful to think of  $p$  and  $c$  as fixed constants. For given  $p, c$ , we want to allow  $\sigma^2$  to vary within a neighborhood of  $p_0 p_1$ , in order to obtain discrete normals parametrized by mean  $p_1 t$  and variance  $\approx \sigma^2 t$ . Proposition 13 below shows that such a family exists, and is given by  $\Sigma \text{val}(\text{RW}_{G_{\lambda, p}}^t)$  for the sticky random walk  $G_{\lambda, p}$  (see Section 2.4) with  $\lambda = (\sigma^2 - p_0 p_1) / (\sigma^2 + p_0 p_1)$ .

Although Definition 12 provides a nonstandard definition of an integer-valued discrete normal, all of the conditions are natural. Condition 1 and condition 2 simply specify that  $\mathcal{N}_{\sigma^2}^t$  has expectation  $p_1 t$  and variance approximately  $\sigma^2 t$ . Condition 3 specifies that the  $t = 1$  case  $\mathcal{N}_{\sigma^2}^1$  takes values in  $\{0, 1\}$ , as is for instance satisfied by a binomial distribution. Condition 4 specifies that the sum of discrete normals is approximately a discrete normal, just as the sum of continuous normals is a normal distribution. Condition 5 specifies that the tails of  $\mathcal{N}_{\sigma^2}^t$  approach the tails of the binomial, which is natural as the binomial is a canonical discrete normal. Condition 6 specifies that the Fourier tails of the discrete normal decay rapidly, as is again the case for continuous normals.

<sup>2</sup>In this section, we typically treat  $p$  and  $c$  as fixed constants, so we exclude them in the notation  $\mathcal{N}_{\sigma^2}^t$  for readability.

## 2.6 Construction of discrete normal families

This section describes our construction of discrete normal families using the sticky walk.

Note that while Definition 9 only defined the sticky random walk for  $0 \leq \lambda \leq 1$ , the definition extends naturally to all  $-\min\{p_0/p_1, p_1/p_0\} \leq \lambda \leq 1$ . Many properties from the  $\lambda \geq 0$  case extend to the  $\lambda < 0$  case. In particular, for all  $-\min\{p_0/p_1, p_1/p_0\} \leq \lambda \leq 1$ , the  $\lambda$ -sticky walk has spectral expansion  $\lambda(G_{\lambda,p}) = |\lambda|$ , as can be seen from Lemma 10 along with the fact that  $G_{-\lambda,p} = 2J - G_{\lambda,p}$ . Our proof of Lemma 11 also still holds for all  $-\min\{p_0/p_1, p_1/p_0\} \leq \lambda \leq 1$ .

For  $\lambda < 0$ , the  $\lambda$ -sticky walk can be thought of as a “ $(-\lambda)$ -jumpy” walk, as in this case the probabilities are skewed towards jumping to the opposite set  $V_{1-b} = \text{val}^{-1}(1-b)$  from  $V_b = \text{val}^{-1}(b)$ . However, there is an unfortunate complication arising in this interpretation, which explains why we cannot let  $\lambda$  descend all the way to  $-1$  when  $p \neq (1/2, 1/2)$ . Recall that  $G_{\lambda,p} = (1-\lambda)J + \lambda \begin{pmatrix} J_{V_0, V_0} & 0 \\ 0 & J_{V_1, V_1} \end{pmatrix}$ , so that the  $\lambda$ -sticky random walk can be interpreted as going to a random vertex in  $V$  with probability  $1-\lambda$ , and remaining in the current set  $V_b$  with probability  $\lambda$ . This interpretation does not extend to  $\lambda < 0$ . A more natural “ $\mu$ -jumpy” random walk, which is defined for all  $0 \leq \mu = -\lambda \leq 1$ , would be given by the random walk matrix  $(1-\mu)J + \mu \begin{pmatrix} 0 & J_{V_0, V_1} \\ J_{V_1, V_0} & 0 \end{pmatrix}$ , which corresponds to going to a random vertex in  $V$  with probability  $1-\mu$ , and jumping to the opposite set  $V_{1-b}$  from  $V_b$  with probability  $\mu$ . Unfortunately, this notion of a jumpy random walk does not come from a regular graph when  $p \neq (1/2, 1/2)$ , so we do not use it here.

The  $\lambda$ -sticky,  $p$ -biased random walk cannot be extended to  $\lambda < -\min\{p_0/p_1, p_1/p_0\}$  without having negative values in the random walk matrix  $G_{\lambda,p}$ . For the purpose of our proofs, such negative probabilities do not seem to pose any fundamental issue, and we may in fact be able to consider  $G_{\lambda,p}$  for all  $-1 \leq \lambda \leq 1$ . However, to avoid confusion, we do not pursue this generalization.

**Proposition 13.** *For every  $p = (p_0, p_1)$  and every*

$$\frac{1 - \min\{\frac{p_0}{p_1}, \frac{p_1}{p_0}, \frac{1}{100}\}}{1 + \min\{\frac{p_0}{p_1}, \frac{p_1}{p_0}, \frac{1}{100}\}} \cdot p_0 p_1 \leq \sigma^2 \leq \frac{1 + \frac{1}{100}}{1 - \frac{1}{100}} \cdot p_0 p_1,$$

*there exists a  $(p, \sigma^2, c)$ -discrete normal family  $(\mathcal{N}_{\sigma^2}^t)_{t \in \mathbb{N}}$  with*

$$\begin{aligned} c_1 &= 2 \\ c_2 &= 2020 \\ c_3 &= 2020. \end{aligned}$$

*Specifically, such a family is given by*

$$\mathcal{N}_{\sigma^2}^t = \Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t),$$

*where  $G_{\lambda,p}$  denotes the  $\lambda$ -sticky,  $p$ -biased random walk with  $\lambda = (\sigma^2 - p_0 p_1) / (\sigma^2 + p_0 p_1)$ .*

The proof of Proposition 13 consists of straightforward applications of results that were stated above; it is provided in Appendix A.

### 3 Statement of Berry-Esseen bound

We are now ready to formally state our main Berry-Esseen result.

**Theorem 14.** *Fix  $p = (p_0, p_1)$ . For  $\lambda \leq 1/100$ , let  $G = (V, E)$  be a  $\lambda$ -spectral expander with labeling  $\text{val} : V \rightarrow \{0, 1\}$  that assigns each label  $b \in \{0, 1\}$  to  $p_b$ -fraction of the vertices. Let  $\sigma^2 = \sigma^2(\Sigma \text{val}(\text{RW}_G^t))$  denote the asymptotic variance. For some  $c \in \mathbb{R}_+^3$ , let  $(\mathcal{N}_{\sigma^2}^t)_{t \in \mathbb{N}}$  be a  $(p, \sigma^2, c)$ -discrete normal family. Then for all  $t \in \mathbb{N}$ ,*

$$d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \mathcal{N}_{\sigma^2}^t) \leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2},$$

where

$$\begin{aligned} \eta_1 &= 140 \\ \eta_2 &= 140 + 3 \cdot \log \left( \frac{2^{28} + 2^{10}c_1 + 2^{18}c_3}{(p_0p_1)^{7/2}} + 3c_2 \right). \end{aligned}$$

We give explicit constants  $\eta_1, \eta_2$  for the sake of completeness, but we do not attempt to optimize these constants.

Recall that Corollary 3 shown by [GV22a] implies that  $d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \Sigma \text{val}(\text{RW}_J^t)) = O(\lambda)$ , while condition 5 of Definition 12 along with Corollary 7 implies that  $d_{\text{TV}}(\mathcal{N}_{\sigma^2}^t, \Sigma \text{val}(\text{RW}_J^t)) = O(\lambda)$ . Thus the results of Golowich and Vadhan [GV22a] imply that  $d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \mathcal{N}_{\sigma^2}^t) = O(\lambda)$ . Theorem 14 improves this bound to  $O(\lambda/t^{1/2-o(1)})$ , at the cost of a worse dependence on the label weights  $p$ .

### 4 Proof overview

This section outlines the proof of Theorem 14. The full proof is given in Section 5 below. At a high level, we follow the standard proof of the central limit theorem, in that we prove our bound using a Taylor approximation of the characteristic function of  $\Sigma \text{val}(\text{RW}_G^t)$ . However, we bound the third moment of this random variable using induction on the walk length  $t$ , which allows us to obtain the desired linear dependence on the spectral expansion  $\lambda$  in our bound. The  $(1 + \log t)^{O(\log \log t)} = t^{o(1)}$  factor arises from a poly log  $t$  loss at each step of the induction. It is an interesting question whether a tighter analysis could remove this factor.

We now present the inductive argument. Throughout this section, we take the label weighting  $p$  to be a fixed constant. Our goal is to prove that for a sufficiently large constant  $\eta$ ,

$$d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \mathcal{N}_{\sigma^2}^t) \leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta \log \log t}. \quad (4)$$

We will prove this inequality by induction over  $t$ . The base case of  $t = 1$  is immediate from the definition of  $\mathcal{N}_{\sigma^2}^1$ . For the inductive step, assume that (4) holds for walks of length  $\sqrt{t}$  (assuming for simplicity that  $t$  is a perfect square to avoid rounding issues).

To begin, we split the length- $t$  walk into  $\sqrt{t}$  walks of length  $\sqrt{t}$ . The sum  $\sum_{k \in [\sqrt{t}]} \Sigma \text{val}(\text{RW}_G^{\sqrt{t}})$  of  $\sqrt{t}$  independent variables can be expressed as  $\Sigma \text{val}(\text{RW}_{\mathcal{G}'}^t)$  where  $\mathcal{G}' = (G'_1, \dots, G'_{\sqrt{t}})$  consists

of  $\sqrt{t} - 1$  evenly spaced copies of  $J$  among  $t - \sqrt{t}$  copies of  $G$ . Thus because  $\|G - J\| = \lambda(G) \leq \lambda$ , Theorem 2 implies that

$$d_{\text{TV}} \left( \Sigma \text{val}(\text{RW}_G^t), \sum_{k \in [\sqrt{t}]} \Sigma \text{val}(\text{RW}_G^{\sqrt{t}}) \right) = O \left( \frac{\lambda}{\sqrt{t}} \right).$$

Similarly, Definition 12 implies that

$$d_{\text{TV}} \left( \mathcal{N}_{\sigma^2}^t, \sum_{k \in [\sqrt{t}]} \mathcal{N}_{\sigma^2}^{\sqrt{t}} \right) = O \left( \frac{\lambda}{\sqrt{t}} \right).$$

Thus by the triangle inequality, to show (4), it is sufficient to show that

$$d_{\text{TV}} \left( \sum_{k \in [\sqrt{t}]} \Sigma \text{val}(\text{RW}_G^{\sqrt{t}}), \sum_{k \in [\sqrt{t}]} \mathcal{N}_{\sigma^2}^{\sqrt{t}} \right) \leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta \log \log t - \Omega(1)}.$$

By the tail bounds in Theorem 2 and Definition 12, almost all of the mass in the probability distributions above is contained within an interval of length roughly  $O(\sqrt{t})$  around the mean  $p_1 t$ , so by Cauchy-Schwartz it is in fact sufficient to show the  $\ell_2$ -bound

$$d_{\ell_2} \left( \sum_{k \in [\sqrt{t}]} \Sigma \text{val}(\text{RW}_G^{\sqrt{t}}), \sum_{k \in [\sqrt{t}]} \mathcal{N}_{\sigma^2}^{\sqrt{t}} \right) \leq \frac{\lambda}{t^{3/4}} \cdot (\log t)^{\eta \log \log t - \Omega(1)}. \quad (5)$$

We show (5) by taking the Fourier transform, and then proving the  $\ell_2$ -bound using the inductive hypothesis. Specifically, denote the probability distributions of  $\Sigma \text{val}(\text{RW}_G^{\sqrt{t}})$  and  $\mathcal{N}_{\sigma^2}^{\sqrt{t}}$  by

$$\begin{aligned} (h_G^{\sqrt{t}})_j &= \Pr \left[ \Sigma \text{val}(\text{RW}_G^{\sqrt{t}}) = j \right] \\ (n_G^{\sqrt{t}})_j &= \Pr \left[ \mathcal{N}_{\sigma^2}^{\sqrt{t}} = j \right], \end{aligned}$$

and denote their centered characteristic functions by

$$\begin{aligned} \hat{h}_G^{\sqrt{t}}(\theta) &= \mathbb{E}[e^{-i\theta(\Sigma \text{val}(\text{RW}_G^{\sqrt{t}}) - p_1 \sqrt{t})}] \\ \hat{n}_G^{\sqrt{t}}(\theta) &= \mathbb{E}[e^{-i\theta(\mathcal{N}_{\sigma^2}^{\sqrt{t}} - p_1 \sqrt{t})}]. \end{aligned}$$

Then the left hand side of (5) equals  $\|(\hat{h}_G^{\sqrt{t}})^{\sqrt{t}} - (\hat{n}_G^{\sqrt{t}})^{\sqrt{t}}\|$  because the Fourier transform preserves  $\ell_2$ -norms. We will bound  $\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}}$  using a Taylor expansion around  $\theta = 0$ . By definition,

$$\begin{aligned} (\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}})(0) &= 0 \\ \frac{d}{d\theta}(\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}})(0) &= 0 \\ \left| \frac{d^2}{d\theta^2}(\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}})(0) \right| &= |\text{Var}(\Sigma \text{val}(\text{RW}_G^{\sqrt{t}})) - \text{Var}(\mathcal{N}_{\sigma^2}^{\sqrt{t}})| = O(\lambda) \\ \left| \frac{d^3}{d\theta^3}(\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}})(\theta) \right| &\leq \sum_{j \in \mathbb{Z}} |j - p_1 \sqrt{t}|^3 \cdot |(h_G^{\sqrt{t}})_j - (n_G^{\sqrt{t}})_j|, \end{aligned}$$

where the bound in the third line above holds by Lemma 8, Corollary 7, and Definition 12. The key point in the proof is now to bound the third derivative above using the inductive hypothesis

$$\|h_G^{\sqrt{t}} - n_G^{\sqrt{t}}\|_1 \leq \frac{\lambda}{t^{1/4}} \cdot (1 + \log \sqrt{t})^{\eta \log \log \sqrt{t}}$$

Combining this inequality with the tail bounds in Theorem 2 and Definition 12, which imply that  $h_G^{\sqrt{t}} - n_G^{\sqrt{t}}$  is mostly supported within an interval of length  $\tilde{O}(t^{1/4})$  around  $p_1 \sqrt{t}$ , gives that

$$\begin{aligned} \left| \frac{d^3}{d\theta^3} (\hat{h}_G^{\sqrt{t}} - \hat{n}_G^{\sqrt{t}})(\theta) \right| &\leq \tilde{O}(t^{1/4})^3 \cdot \|h_G^{\sqrt{t}} - n_G^{\sqrt{t}}\|_1 \\ &\leq \lambda \sqrt{t} \cdot (1 + \log t)^{\eta \log \log \sqrt{t} + O(1)}. \end{aligned}$$

Thus we have the Taylor approximation

$$|\hat{h}_G^{\sqrt{t}}(\theta) - \hat{n}_G^{\sqrt{t}}(\theta)| \leq \lambda \cdot (|\theta|^2 + \sqrt{t} \cdot |\theta|^3) \cdot (1 + \log t)^{\eta \log \log \sqrt{t} + O(1)}.$$

Now by the Fourier tail bounds in Lemma 4 and in Definition 12, both  $|\hat{h}_G^{\sqrt{t}}(\theta)|$  and  $|\hat{n}_G^{\sqrt{t}}(\theta)|$  are bounded above by  $e^{-\Omega(\sqrt{t} \cdot \theta^2)}$ . Therefore

$$\begin{aligned} \left| \left( \hat{h}_G^{\sqrt{t}}(\theta) \right)^{\sqrt{t}} - \left( \hat{n}_G^{\sqrt{t}}(\theta) \right)^{\sqrt{t}} \right| &= \left| \sum_{k=0}^{\sqrt{t}-1} \left( \hat{n}_G^{\sqrt{t}}(\theta) \right)^k \cdot \left( \hat{h}_G^{\sqrt{t}}(\theta) - \hat{n}_G^{\sqrt{t}}(\theta) \right) \cdot \left( \hat{h}_G^{\sqrt{t}}(\theta) \right)^{\sqrt{t}-k-1} \right| \\ &\leq \sqrt{t} \cdot |\hat{h}_G^{\sqrt{t}}(\theta) - \hat{n}_G^{\sqrt{t}}(\theta)| \cdot e^{-\Omega(t \cdot \theta^2)} \\ &\leq \frac{\lambda}{\sqrt{t}} \cdot (t|\theta|^2 + t^{3/2}|\theta|^3) \cdot e^{-\Omega(t \cdot \theta^2)} \cdot (1 + \log t)^{\eta \log \log \sqrt{t} + O(1)}. \end{aligned}$$

Squaring the bound above and integrating over  $-\pi < \theta \leq \pi$  then gives the desired  $\ell_2$ -bound (5). Intuitively, for  $|\theta| \leq \tilde{O}(1/\sqrt{t})$  the right hand side above is bounded by  $\lambda/\sqrt{t} \cdot (1 + \log t)^{\eta \log \log \sqrt{t} + O(1)}$ , while for  $|\theta| \gg 1/\sqrt{t}$  the right hand side above decays rapidly. Therefore the  $\ell_2$ -norm of this function is bounded by

$$\begin{aligned} \left\| \left( \hat{h}_G^{\sqrt{t}} \right)^{\sqrt{t}} - \left( \hat{n}_G^{\sqrt{t}} \right)^{\sqrt{t}} \right\| &\leq \sqrt{\tilde{O} \left( \frac{1}{\sqrt{t}} \right) \cdot \left( \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta \log \log \sqrt{t} + O(1)} \right)^2} \\ &\leq \frac{\lambda}{t^{3/4}} \cdot (1 + \log t)^{\eta \log \log t - \Omega(1)}, \end{aligned}$$

where the final inequality above assumes that  $\eta$  is sufficiently large so that  $\eta \log 2$  dominates the  $O(1)$  constant in the exponent. Then (5) follows because the left hand side above equals the left hand side of (5), as the Fourier transform interchanges convolution and multiplication, and preserves  $\ell_2$ -norms. Thus (4) holds, completing the inductive step.

## 5 Proof of Berry-Esseen bound

This section presents the full proof of Theorem 14.

We first introduce some notation. Fix a graph  $G$ , and set  $\sigma^2 = \sigma^2(\Sigma \text{val}(\text{RW}_G^t))$  to be the asymptotic variance. For every  $t \in \mathbb{N}$ , define vectors  $h_G^t, n_G^t \in [0, 1]^{\mathbb{Z}} \subseteq \mathbb{R}^{\mathbb{Z}}$  by

$$\begin{aligned}(h_G^t)_j &= \Pr[\Sigma \text{val}(\text{RW}_G^t) = j] \\ (n_G^t)_j &= \Pr[\mathcal{N}_{\sigma^2}^t = j].\end{aligned}$$

That is,  $h_G^t$  is the probability distribution of  $\Sigma \text{val}(\text{RW}_G^t)$ , and  $n_G^t$  is the probability distribution of the discrete normal  $\mathcal{N}_{\sigma^2}^t$  with the same mean and asymptotic variance as  $\Sigma \text{val}(\text{RW}_G^t)$ . For a given  $t$ , Lemma 8, Corollary 7, and condition 2 of Definition 12 imply that

$$|\text{Var}(\Sigma \text{val}(\text{RW}_G^t)) - \text{Var}(\mathcal{N}_{\sigma^2}^t)| \leq \left( \frac{2}{(1-\lambda)^2} + \frac{2c_1}{1-\lambda} \right) \cdot \lambda \cdot p_0 p_1. \quad (6)$$

For a probability distribution  $f^t \in [0, 1]^{\mathbb{Z}}$  with mean  $\sum_{j \in \mathbb{Z}} j f_j^t = p_1 t$  (e.g.  $f^t = h_G^t$  or  $n_G^t$ ), in this proof we define the **centered Fourier transform**

$$\hat{f}^t(\theta) = \sum_{j \in \mathbb{Z}} e^{-i\theta(j-p_1 t)} f_j^t.$$

The centered Fourier transform is by definition equal to  $e^{i\theta p_1 t}$  times the ordinary Fourier transform. Therefore in particular, the absolute values of the centered and ordinary Fourier transforms agree, so the centered Fourier transform preserves  $\ell_2$ -norms like the ordinary Fourier transform. Furthermore, recall that the sum of independent random variables with distributions  $f^{t_1}$  and  $f^{t_2}$  has distribution given by the convolution  $f^{t_1} * f^{t_2}$ . As with the ordinary Fourier transform, the centered Fourier transform changes convolution into multiplication, that is,  $\widehat{f^{t_1} * f^{t_2}} = \hat{f}^{t_1} \cdot \hat{f}^{t_2}$ . Note that we cannot instead center the distribution  $f^t$  and then take the ordinary Fourier transform because if  $p_1 t \notin \mathbb{Z}$ , then the centered version of  $f^t$  no longer takes values in  $\mathbb{Z}$ .

*Proof of Theorem 14.* We will prove by induction that for all  $t \in \mathbb{N}$ , we have the desired inequality

$$\|h_G^t - n_G^t\|_1 \leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2}. \quad (7)$$

For the base case of our induction, when  $t = 1$ , then  $\Sigma \text{val}(\text{RW}_G^1)$  is simply a Bernoulli distribution with parameter  $p$ , and by condition 3 of Definition 12,  $\mathcal{N}_{\sigma^2}^1$  is also a Bernoulli with parameter  $p$ . Thus  $h_G^1 = n_G^1$ , so (7) holds for  $t = 1$ .

For the inductive step, fix  $t \geq 1$ . We will prove that the inequality (7) holds for  $t$  assuming that it holds for all  $u < t$ . To begin, we reduce our problem of studying a random walk on  $G$  of length  $t$  to studying  $\ell \approx \sqrt{t}$  independent random walks on  $G$  of length approximately  $\sqrt{t}$ . We will then apply the inductive hypothesis on these shorter random walks. Formally, choose  $\ell \in \mathbb{N}$  and  $t_0, \dots, t_{\ell-1} \in \mathbb{N}$  such that  $\sum_{k \in [\ell]} t_k = t$ , and such that  $\ell$  and all  $t_k$  differ from  $\sqrt{t}$  by less than 1. In the case that  $t = 2$ , we specify that  $\ell = 2$  and  $t_0 = t_1 = 1$ , so that for all  $t \geq 2$  we have all  $t_k \leq \min\{\sqrt{t} + 1, t - 1\}$ .

Let  $S = \{j \in \mathbb{Z} : |j - p_1 t| \leq 8\sqrt{t \log t}\}$ . The distributions  $h_G^t$  and  $n_G^t$  are mostly supported on  $S$ , so the main point of our induction is to bound the  $\ell_1$ -norm of the restriction  $(h_G^t - n_G^t)_S$ . We will separately apply tail bounds to bound  $\|(h_G^t - n_G^t)_{\mathbb{Z} \setminus S}\|_1$ . Specifically, we begin with the following lemma.

**Lemma 15.**

$$\|h_G^t - n_G^t\|_1 \leq \left(8000 + \frac{2(c_2 + c_3)}{1 - \lambda}\right) \cdot \frac{\lambda}{\sqrt{t}} + \left\| \left( \begin{array}{c} * \\ k \in [\ell] \end{array} h_G^{t_k} - \begin{array}{c} * \\ k \in [\ell] \end{array} n_G^{t_k} \right)_S \right\|_1.$$

*Proof.* Let  $b^t \in [0, 1]^{\mathbb{Z}}$  denote the binomial distribution with parameters  $t, p$ . Then

$$\begin{aligned} \|h_G^t - n_G^t\|_1 &\leq \left\| \left( \begin{array}{c} h_G^t - * \\ k \in [\ell] \end{array} h_G^{t_k} \right)_S \right\|_1 + \left\| \left( \begin{array}{c} * \\ k \in [\ell] \end{array} h_G^{t_k} - \begin{array}{c} * \\ k \in [\ell] \end{array} n_G^{t_k} \right)_S \right\|_1 + \left\| \left( \begin{array}{c} * \\ k \in [\ell] \end{array} n_G^{t_k} - n_G^t \right)_S \right\|_1 \\ &\quad + \|(h_G^t - b^t)_{\mathbb{Z} \setminus S}\|_1 + \|(b^t - n_G^t)_{\mathbb{Z} \setminus S}\|_1. \end{aligned}$$

By definition  $*_{k \in [\ell]} h_G^{t_k}$  is the distribution of the sum  $\sum_{k \in [\ell]} \Sigma \text{val}(\text{RW}_G^{t_k})$  of  $\ell$  independent variables, which can be expressed as  $\Sigma \text{val}(\text{RW}_{\mathcal{G}'}^t)$  where  $\mathcal{G}' = (G'_1, \dots, G'_{t-1})$  consists of  $\ell - 1$  copies of  $J$  and  $t - \ell$  copies of  $G$ . Thus because  $\|G - J\| = \lambda(G) \leq \lambda$ , Theorem 2 implies that

$$\begin{aligned} \left\| h_G^t - \begin{array}{c} * \\ k \in [\ell] \end{array} h_G^{t_k} \right\|_1 &= 2 \cdot d_{\text{TV}}(\Sigma \text{val}(\text{RW}_G^t), \Sigma \text{val}(\text{RW}_{\mathcal{G}'}^t)) \\ &\leq 4000 \cdot \frac{\lambda}{t} \cdot (\ell - 1) \\ &\leq 4000 \cdot \frac{\lambda}{\sqrt{t}}. \end{aligned}$$

Similarly,  $*_{k \in [\ell]} n_G^{t_k}$  is the distribution of the sum  $\sum_{k \in [\ell]} \mathcal{N}_{\sigma^2}^{t_k}$  of  $\ell$  independent variables, so by condition 4 of Definition 12 along with Corollary 7,

$$\begin{aligned} \left\| \begin{array}{c} * \\ k \in [\ell] \end{array} n_G^{t_k} - n_G^t \right\|_1 &= 2 \cdot d_{\text{TV}} \left( \sum_{k \in [\ell]} \mathcal{N}_{\sigma^2}^{t_k}, \mathcal{N}_{\sigma^2}^t \right) \\ &\leq c_2 \cdot \frac{(\ell - 1) \cdot |\sigma^2 / p_0 p_1 - 1|}{t} \\ &\leq c_2 \cdot \frac{\sqrt{t} \cdot \frac{2}{1 - \lambda} \cdot \lambda}{t} \\ &= \frac{2c_2}{1 - \lambda} \cdot \frac{\lambda}{\sqrt{t}}. \end{aligned}$$

Furthermore, Theorem 2 implies that

$$\|(h_G^t - b^t)_{\mathbb{Z} \setminus S}\|_1 \leq 4000 \cdot \lambda \cdot \frac{1}{t^8},$$

and condition 5 of Definition 12 with Corollary 7 implies that

$$\|(b^t - n_G^t)_{\mathbb{Z} \setminus S}\|_1 \leq c_3 \cdot \frac{2\lambda}{1 - \lambda} \cdot \frac{1}{t^8}.$$

Combining the above inequalities gives the desired inequality

$$\|h_G^t - n_G^t\|_1 \leq \left(8000 + \frac{2(c_2 + c_3)}{1 - \lambda}\right) \cdot \frac{\lambda}{\sqrt{t}} + \left\| \left( \begin{array}{c} * \\ k \in [\ell] \end{array} h_G^{t_k} - \begin{array}{c} * \\ k \in [\ell] \end{array} n_G^{t_k} \right)_S \right\|_1.$$

□



Thus our problem is reduced to bounding  $\left\| \left( \ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k} \right)_S \right\|_1$ . For this purpose, we will bound the  $\ell_2$ -norm of  $\ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k}$  by bounding its centered Fourier coefficients. We will then obtain an  $\ell_1$ -bound from the  $\ell_2$ -bound by applying the Cauchy-Schwartz inequality.

By definition, the centered Fourier transform of  $\ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k}$  is given by  $\prod_{k \in \ell} \hat{h}_G^{t_k}(\theta) - \prod_{k \in \ell} \hat{n}_G^{t_k}(\theta)$ . Thus we will first bound the difference between  $\hat{h}_G^{t_k}(\theta)$  and  $\hat{n}_G^{t_k}(\theta)$  in the following lemma by applying the inductive hypothesis, which is the key technical step in our proof.

**Lemma 16.** *For every  $1 \leq u \leq t - 1$ , we have*

$$|\hat{h}_G^u(\theta) - \hat{n}_G^u(\theta)| \leq \lambda \cdot \left( \frac{C_1}{2} \cdot |\theta|^2 + \frac{C_2}{6} \cdot u \cdot |\theta|^3 \right) \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2}$$

for constants

$$C_1 = \left( \frac{2}{(1 - \lambda)^2} + \frac{2c_1}{1 - \lambda} \right) \cdot p_0 p_1$$

$$C_2 = 2^9 + 2^8 \left( 4000 + \frac{2c_3}{1 - \lambda} \right).$$

*Proof.* If  $u = 1$ , then the desired result follows directly from the fact that  $h_G^1 = n_G^1$  by condition 3 of Definition 12. Thus we may assume that  $u \geq 2$ .

Recall that for  $u \in \mathbb{N}$ , if  $f^u \in [0, 1]^{\mathbb{Z}} \subseteq \mathbb{R}^{\mathbb{Z}}$  is the probability distribution for a random variable  $F^u$  with mean  $\mathbb{E}[F^u] = p_1 u$ , then the centered Fourier transform  $\hat{f}^u$  satisfies

$$\begin{aligned} \hat{f}^u(0) &= 1 \\ \frac{d\hat{f}^u}{d\theta}(0) &= -i(\mathbb{E}[F^u] - p_1 u) = 0 \\ \frac{d^2\hat{f}^u}{d\theta^2}(0) &= -\mathbb{E}[(F^u - p_1 u)^2] = -\text{Var}(F^u) \\ \frac{d^3\hat{f}^u}{d\theta^3}(\theta) &= \sum_{j \in \mathbb{Z}} i e^{-i\theta(j - p_1 u)} \cdot (j - p_1 u)^3 f_j^u. \end{aligned}$$

Thus for every  $2 \leq u \leq t - 1$ , we have

$$\begin{aligned} (\hat{h}_G^u - \hat{n}_G^u)(0) &= 0 \\ \frac{d}{d\theta}(\hat{h}_G^u - \hat{n}_G^u)(0) &= 0 \\ \left| \frac{d^2}{d\theta^2}(\hat{h}_G^u - \hat{n}_G^u)(0) \right| &= |\text{Var}(\Sigma \text{val}(\text{RW}_G^u)) - \text{Var}(\mathcal{N}_{\sigma^2}^u)| \leq C_1 \cdot \lambda, \end{aligned} \tag{8}$$

where the final equality above is given by (6). The key step in the proof is now to bound the third derivative  $\frac{d^3}{d\theta^3}(\hat{h}_G^u - \hat{n}_G^u)(\theta)$  using the inductive hypothesis. To begin, by the above expression for

the third derivative,

$$\begin{aligned}
& \left| \frac{d^3}{d\theta^3} (\hat{h}_G^u - \hat{n}_G^u)(\theta) \right| \\
& \leq \sum_{j \in \mathbb{Z}} |j - p_1 u|^3 \cdot |(h_G^u)_j - (n_G^u)_j| \\
& \leq (8\sqrt{u \log u})^3 \cdot \|h_G^u - n_G^u\|_1 + \sum_{j \in \mathbb{Z}: |j - p_1 u| > 8\sqrt{u \log u}} |j - p_1 u|^3 \cdot |(h_G^u)_j - (n_G^u)_j|.
\end{aligned} \tag{9}$$

By the inductive hypothesis,

$$\begin{aligned}
(8\sqrt{u \log u})^3 \cdot \|h_G^u - n_G^u\|_1 & \leq (8\sqrt{u \log u})^3 \cdot \frac{\lambda}{\sqrt{u}} \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2} \\
& = 2^9 \cdot \lambda u \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2}.
\end{aligned}$$

To bound the second term on the right hand side of (9), let  $b^u \in [0, 1]^{\mathbb{Z}}$  denote the binomial distribution with parameters  $u, p$ . Then by Theorem 2 and by condition 5 of Definition 12 along with Corollary 7,

$$\begin{aligned}
& \sum_{j \in \mathbb{Z}: |j - p_1 u| > 8\sqrt{u \log u}} |j - p_1 u|^3 \cdot |(h_G^u)_j - (n_G^u)_j| \\
& \leq \sum_{a=8\sqrt{u \log u}}^{\infty} (a+1)^3 \sum_{j \in \mathbb{Z}: |j - p_1 u| > a} (|(h_G^u)_j - b_j^u| + |b_j^u - (n_G^u)_j|) \\
& \leq \sum_{a=8\sqrt{u \log u}}^{\infty} (a+1)^3 \left( 4000 \cdot \lambda \cdot e^{-a^2/8u} + c_3 \cdot \frac{2\lambda}{1-\lambda} \cdot e^{-a^2/8u} \right) \\
& \leq \left( 4000 + \frac{2c_3}{1-\lambda} \right) \cdot \lambda \cdot \int_{a=8\sqrt{u \log u} - 1}^{\infty} (2a)^3 e^{-a^2/8u} da \\
& \leq \left( 4000 + \frac{2c_3}{1-\lambda} \right) \cdot \lambda \cdot 2^8 u^2 \int_{q=6 \log u}^{\infty} q e^{-q} dq \\
& = \left( 4000 + \frac{2c_3}{1-\lambda} \right) \cdot \lambda \cdot 2^8 u^2 \cdot \frac{6 \log u + 1}{u^6} \\
& \leq \left( 4000 + \frac{2c_3}{1-\lambda} \right) \cdot \lambda \cdot 2^8
\end{aligned}$$

where the third inequality above holds because  $a^3 e^{-a^2/8u}$  is maximized at  $\sqrt{12u} < 8\sqrt{u \log u} - 1$ , the fourth inequality uses the fact that  $8\sqrt{u \log u} - 1 \geq 7\sqrt{u \log u}$  and then applies the substitution  $q = a^2/8u$ , the equality follows by applying integration by parts, and the final inequality holds because  $u \geq 2$  by assumption. Thus (9) becomes

$$\begin{aligned}
\left| \frac{d^3}{d\theta^3} (\hat{h}_G^u - \hat{n}_G^u)(\theta) \right| & \leq 2^9 \cdot \lambda u \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2} + 2^8 \left( 4000 + \frac{2c_3}{1-\lambda} \right) \cdot \lambda \\
& \leq C_2 \cdot \lambda u \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2}.
\end{aligned}$$

Combining the above inequality with (8) and expanding the Taylor approximation at  $\theta = 0$  gives

$$\begin{aligned} |\hat{h}_G^u(\theta) - \hat{n}_G^u(\theta)| &\leq \frac{C_1}{2} \cdot \lambda \cdot |\theta|^2 + \frac{C_2}{6} \cdot \lambda u \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2} \cdot |\theta|^3 \\ &\leq \lambda \cdot \left( \frac{C_1}{2} \cdot |\theta|^2 + \frac{C_2}{6} \cdot u \cdot |\theta|^3 \right) \cdot (1 + \log u)^{\eta_1 \log \log u + \eta_2 + 3/2}. \end{aligned}$$

□

We apply Lemma 16 below to bound the Fourier coefficients of  $\ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k}$ .

**Lemma 17.**

$$\begin{aligned} \left| \prod_{k \in [\ell]} \hat{h}_G^{t_k}(\theta) - \prod_{k \in [\ell]} \hat{n}_G^{t_k}(\theta) \right| &\leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \frac{(1 + \log t)^{6 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \\ &\quad \cdot \left( C_1 \cdot t \cdot |\theta|^2 + \frac{C_2}{6} \cdot t^{3/2} \cdot |\theta|^3 \right) \cdot e^{-p_0 p_1 t \theta^2 / 60}. \end{aligned}$$

*Proof.* By Lemma 4,

$$|\hat{h}_G^u(\theta)| \leq e^{-p_0 p_1 u \theta^2 / 20}.$$

The same bound holds for  $|\hat{n}_G^u(\theta)|$  by condition 6 of Definition 12. Thus

$$\begin{aligned} &\left| \prod_{k \in [\ell]} \hat{h}_G^{t_k}(\theta) - \prod_{k \in [\ell]} \hat{n}_G^{t_k}(\theta) \right| \\ &= \left| \sum_{k \in [\ell]} \left( \prod_{k'=0}^{k-1} \hat{n}_G^{t_{k'}}(\theta) \right) \cdot (\hat{h}_G^{t_k}(\theta) - \hat{n}_G^{t_k}(\theta)) \cdot \left( \prod_{k'=k+1}^{\ell-1} \hat{h}_G^{t_{k'}}(\theta) \right) \right| \\ &\leq \sum_{k \in [\ell]} |\hat{h}_G^{t_k}(\theta) - \hat{n}_G^{t_k}(\theta)| \cdot e^{-p_0 p_1 (t - t_k) \theta^2 / 20} \\ &\leq \sum_{k \in [\ell]} |\hat{h}_G^{t_k}(\theta) - \hat{n}_G^{t_k}(\theta)| \cdot e^{-p_0 p_1 t \theta^2 / 60}, \end{aligned}$$

where the final inequality above holds because  $t - t_k \geq t/3$  by definition. Applying Lemma 16 with  $u = t_k$  for  $k \in [\ell]$  to bound the sum above gives

$$\begin{aligned} &\sum_{k \in [\ell]} |\hat{h}_G^{t_k}(\theta) - \hat{n}_G^{t_k}(\theta)| \\ &\leq \lambda \cdot \left( \frac{C_1}{2} \cdot \ell \cdot |\theta|^2 + \frac{C_2}{6} \cdot t \cdot |\theta|^3 \right) \cdot \left( 1 + \log(\sqrt{t} + 1) \right)^{\eta_1 \log \log \min\{\sqrt{t} + 1, t - 1\} + \eta_2 + 3/2} \\ &\leq \lambda \cdot \left( C_1 \cdot \sqrt{t} \cdot |\theta|^2 + \frac{C_2}{6} \cdot t \cdot |\theta|^3 \right) \cdot \left( \frac{1 + \log t}{\sqrt{2}} \right)^{\eta_1 (\log \log t - 1/20) + \eta_2 + 6} \\ &\leq \lambda \cdot \left( C_1 \cdot \sqrt{t} \cdot |\theta|^2 + \frac{C_2}{6} \cdot t \cdot |\theta|^3 \right) \cdot \frac{(1 + \log t)^{6 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2}. \end{aligned}$$

where the first inequality above holds because all  $k \in [\ell]$  have  $t_k \leq \min\{\sqrt{t} + 1, t - 1\}$  by definition, and the second inequality above holds because  $\ell \leq \sqrt{t} + 1 \leq 2\sqrt{t}$ , and  $t \geq 2$  so that  $1 + \log(\sqrt{t} + 1) \leq ((1 + \log t)/\sqrt{2})^4$  and  $\log \log \min\{\sqrt{t} + 1, t - 1\} \leq \log \log t - 1/20$ .

The desired result now follows by combining the two inequalities above.  $\square$

The following lemma squares and integrates the Fourier coefficient bound in Lemma 17 to bound the  $\ell_2$ -norm of  $\ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k}$ .

**Lemma 18.**

$$\left\| \ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k} \right\| \leq t^{-1/4} \cdot \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \frac{(1 + \log t)^{6 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \cdot \frac{70 \cdot (C_1 + C_2)}{(p_0 p_1)^{7/2}}.$$

*Proof.* Squaring and integrating the bound in Lemma 17 gives

$$\begin{aligned} \left\| \ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k} \right\| &= \sqrt{\int_{\theta=-\pi}^{\pi} \left| \prod_{k \in [\ell]} \hat{h}_G^{t_k}(\theta) - \prod_{k \in [\ell]} \hat{n}_G^{t_k}(\theta) \right|^2 \frac{d\theta}{2\pi}} \\ &\leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \frac{(1 + \log t)^{6 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \\ &\quad \cdot \sqrt{2C_1^2 \int_{\theta=-\pi}^{\pi} t^2 \theta^4 e^{-p_0 p_1 t \theta^2/30} \frac{d\theta}{2\pi} + \frac{C_2^2}{18} \int_{\theta=-\pi}^{\pi} t^3 \theta^6 e^{-p_0 p_1 t \theta^2/30} \frac{d\theta}{2\pi}}. \end{aligned}$$

Substituting  $q = \sqrt{p_0 p_1 t/30} \cdot \theta$  in the integrals in the right hand side above gives

$$\begin{aligned} &\sqrt{2C_1^2 \int_{\theta=-\pi}^{\pi} t^2 \theta^4 e^{-p_0 p_1 t \theta^2/30} \frac{d\theta}{2\pi} + \frac{C_2^2}{18} \int_{\theta=-\pi}^{\pi} t^3 \theta^6 e^{-p_0 p_1 t \theta^2/30} \frac{d\theta}{2\pi}} \\ &\leq \sqrt{2C_1^2 \left(\frac{30}{p_0 p_1}\right)^{5/2} \cdot t^{-1/2} \cdot \int_{q=-\infty}^{\infty} q^4 e^{-q^2} \frac{dq}{2\pi} + \frac{C_2^2}{18} \left(\frac{30}{p_0 p_1}\right)^{7/2} \cdot t^{-1/2} \cdot \int_{q=-\infty}^{\infty} q^6 e^{-q^2} \frac{dq}{2\pi}} \\ &\leq t^{-1/4} \cdot \frac{70 \cdot (C_1 + C_2)}{(p_0 p_1)^{7/2}}. \end{aligned}$$

The two inequalities above imply the desired bound.  $\square$

To complete the proof, we apply the Cauchy-Schwartz inequality to the bound in Lemma 18 to obtain an  $\ell_1$ -bound, which combines with Lemma 15 to imply the theorem statement. Specifically, by Lemma 18 and Cauchy-Schwartz,

$$\begin{aligned} &\left\| \left( \ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k} \right) \right\|_{S^1} \\ &\leq \sqrt{16\sqrt{t \log t} + 1} \cdot \left\| \ast_{k \in [\ell]} h_G^{t_k} - \ast_{k \in [\ell]} n_G^{t_k} \right\| \\ &\leq \frac{\lambda}{\sqrt{t}} \cdot (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \frac{(1 + \log t)^{7 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \cdot \frac{300 \cdot (C_1 + C_2)}{(p_0 p_1)^{7/2}}. \end{aligned}$$

Applying this inequality with Lemma 15 gives

$$\begin{aligned} & \|h_G^t - n_G^t\|_1 \\ & \leq \frac{\lambda}{\sqrt{t}} \cdot \left( 8000 + \frac{2(c_2 + c_3)}{1 - \lambda} + (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \frac{(1 + \log t)^{7 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \cdot \frac{300 \cdot (C_1 + C_2)}{(p_0 p_1)^{7/2}} \right) \\ & \leq \frac{\lambda}{\sqrt{t}} (1 + \log t)^{\eta_1 \log \log t + \eta_2} \cdot \left( \frac{8000 + 2(c_2 + c_3)/(1 - \lambda)}{(1 + \log 2)^{\eta_1 \log \log t + \eta_2}} + \frac{(1 + \log t)^{7 - \eta_1/20}}{2^{(\eta_2 + 6 - \eta_1)/2}} \cdot \frac{300 \cdot (C_1 + C_2)}{(p_0 p_1)^{7/2}} \right). \end{aligned}$$

Thus  $\|h_G^t - n_G^t\|_1 \leq (\lambda/\sqrt{t})(1 + \log t)^{\eta_1 \log \log t + \eta_2}$  when

$$\begin{aligned} \eta_1 &= 140 \\ \eta_2 &= 140 + \frac{2}{\log 2} \cdot \log \max \left\{ 8000 + 3(c_2 + c_3), \frac{300(C_1 + C_2)}{(p_0 p_1)^{7/2}} \right\} \\ &\leq 140 + 3 \cdot \log \left( \frac{2^{28} + 2^{10}c_1 + 2^{18}c_3}{(p_0 p_1)^{7/2}} + 3c_2 \right), \end{aligned}$$

where we have applied the fact that  $C_1 + C_2 \leq 2^{20} + 3c_1 + 2^{10}c_3$  by definition. Thus we have shown that (7) holds, completing the inductive step of the proof.  $\square$

## 6 Acknowledgments

The author thanks Salil Vadhan for numerous helpful discussions and suggestions, and thanks Venkat Guruswami for helping to motivate this work and providing feedback.

This work was primarily done while the author was at Harvard University. The author is currently supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE 2146752.

## References

- [Ber41] Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.
- [Bol80] E. Bolthausen. The Berry-Esseen theorem for functionals of discrete Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 54(1):59–73, January 1980.
- [Bol82] E. Bolthausen. The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 60(3):283–289, September 1982.
- [BX99] Andrew D. Barbour and Aihua Xia. Poisson perturbations. *ESAIM: Probability and Statistics*, 3:131–150, 1999.

- [CMP<sup>+</sup>22] Gil Cohen, Dor Minzer, Shir Peleg, Aaron Potechin, and Amnon Ta-Shma. Expander Random Walks: The General Case and Limitations. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 43:1–43:18, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISSN: 1868-8969.
- [CP22] Rafael Chiclana and Yuval Peres. A local central limit theorem for random walks on expander graphs, December 2022. arXiv:2212.00958 [math].
- [CPTS21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a Fourier-analytic approach. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1643–1655, New York, NY, USA, June 2021. Association for Computing Machinery.
- [Das08] Constantinos Daskalakis. An Efficient PTAS for Two-Strategy Anonymous Games. In Christos Papadimitriou and Shuzhong Zhang, editors, *Internet and Network Economics*, Lecture Notes in Computer Science, pages 186–197, Berlin, Heidelberg, 2008. Springer.
- [DDO<sup>+</sup>13] Constantinos Daskalakis, Ilias Diakonikolas, Ryan O’Donnell, Rocco A. Servedio, and Li-Yang Tan. Learning Sums of Independent Integer Random Variables. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 217–226, October 2013. ISSN: 0272-5428.
- [DDS12] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco A. Servedio. Learning poisson binomial distributions. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, STOC ’12, pages 709–728, New York, NY, USA, May 2012. Association for Computing Machinery.
- [DP09] Constantinos Daskalakis and Christos H. Papadimitriou. On oblivious PTAS’s for nash equilibrium. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, STOC ’09, pages 75–84, New York, NY, USA, May 2009. Association for Computing Machinery.
- [Ess42] Carl-Gustav Esseen. On the Liapunov limit error in the theory of probability. *Ark. Mat. Astr. Fys.*, 28:1–19, 1942.
- [Fan14] Xiao Fang. Discretized normal approximation by Stein’s method. *Bernoulli*, 20(3):1404–1431, August 2014. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- [GK21] Venkatesan Guruswami and Vinayak M. Kumar. Pseudobinomiality of the Sticky Random Walk. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:19, Dagstuhl, Germany, 2021. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- [GMRZ11] Parikshit Gopalan, Raghu Meka, Omer Reingold, and David Zuckerman. Pseudorandom generators for combinatorial shapes. In *Proceedings of the forty-third annual ACM*

*symposium on Theory of computing*, STOC '11, pages 253–262, New York, NY, USA, June 2011. Association for Computing Machinery.

- [Gol22] Louis Golowich. *Random Walks on Expander Graphs*. Senior Thesis, Harvard University, Cambridge, MA, March 2022.
- [GV22a] Louis Golowich and Salil Vadhan. Pseudorandomness of Expander Random Walks for Symmetric Functions and Permutation Branching Programs. In Shachar Lovett, editor, *37th Computational Complexity Conference (CCC 2022)*, volume 234 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 27:1–27:13, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISSN: 1868-8969.
- [GV22b] Louis Golowich and Salil Vadhan. Pseudorandomness of Expander Random Walks for Symmetric Functions and Permutation Branching Programs. *Electronic Colloquium on Computational Complexity*, 2022.
- [HP10] Loïc Hervé and Françoise Pène. The Nagaev-Guivarc’h method via the Keller-Liverani theorem. *Bulletin de la Société mathématique de France*, 138(3):415–489, 2010.
- [Klo19] Benoît Kloeckner. Effective Berry–Esseen and concentration bounds for Markov chains with a spectral gap. *The Annals of Applied Probability*, 29(3):1778–1807, June 2019.
- [MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the forty-second ACM symposium on Theory of computing*, STOC '10, pages 427–436, New York, NY, USA, June 2010. Association for Computing Machinery.
- [Rö06] Adrian Röllin. Symmetric and centered binomial approximation of sums of locally dependent random variables. *Electronic Journal of Probability*, 13, September 2006.
- [Rö07] Adrian Röllin. Translated Poisson approximation using exchangeable pair couplings. *The Annals of Applied Probability*, 17(5-6), October 2007. arXiv: math/0607781.
- [TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, Montreal Canada, June 2017. ACM.

## A Omitted proofs

This section presents the proofs we omitted in Section 2. These proofs are standard or follow directly from prior work, but we include them here for completeness.

*Proof of Lemma 4.* This result was implicitly shown by Golowich and Vadhan [GV22a] in the proof of Theorem 2. Specifically, define  $F \in \mathbb{R}^{V \times V}$  by  $F = J + (I - J)/10$ , and define  $P_\theta^{(0)} \in \mathbb{C}^{V \times V}$  to be the diagonal matrix with  $(P_\theta^{(0)})_{v,v} = e^{-i\theta(\text{val}(v) - p_1)}$ . Then the proof of Lemma 26 in [GV22b] (the full version of [GV22a]) shows that

$$|\mathbb{E}[e^{-i\theta \Sigma \text{val}(\text{RW}_G^t)}]| = |\vec{1}^\top (GP_\theta^{(0)})^t \vec{1}|.$$

Now

$$\begin{aligned} |\vec{1}^\top (GP_\theta^{(0)})^t \vec{1}| &= |\vec{1}^\top (F^{-1}GF^{-1} \cdot FP_\theta^{(0)}F)^t \vec{1}| \\ &\leq (\|F^{-1}GF^{-1}\| \cdot \|FP_\theta^{(0)}F\|)^t. \end{aligned}$$

Because  $\lambda(G) \leq 1/100$  and  $F^{-1} = J + 10(I - J)$ , it follows that  $\|F^{-1}GF^{-1}\| \leq 1$ , so

$$|\mathbb{E}[e^{-i\theta\Sigma \text{val}(\text{RW}_G^t)}]| \leq \|FP_\theta^{(0)}F\|^t.$$

Lemma 28 in [GV22b] implies that  $\|FP_\theta^{(0)}F\| \leq e^{-p_0p_1\theta^2/20}$ , so we obtain the desired inequality

$$|\mathbb{E}[e^{-i\theta\Sigma \text{val}(\text{RW}_G^t)}]| \leq e^{-p_0p_1t\theta^2/20}.$$

□

*Proof of Corollary 7.* By Lemma 6,

$$\begin{aligned} |\sigma^2(\Sigma \text{val}(\text{RW}_G^t)) - p_0p_1| &\leq 2 \sum_{i=1}^{\infty} \frac{1}{n} |(\text{val} - p_1)^\top G^i (\text{val} - p_1)| \\ &\leq 2 \sum_{i=1}^{\infty} \lambda^i \cdot \frac{\|\text{val} - p_1\|^2}{n} \\ &= \frac{2}{1 - \lambda} \cdot \lambda \cdot p_0p_1, \end{aligned}$$

where the first inequality above holds because  $(\text{val} - p_1)^\top G^i \vec{1} = (\text{val} - p_1)^\top \vec{1} = 0$  so that  $(\text{val} - p_1)^\top G^i \text{val} = (\text{val} - p_1)^\top G^i (\text{val} - p_1)$ , and the second inequality holds because  $\lambda(G) = \lambda$  and  $\text{val} - p_1 \in \vec{1}^\perp$ . □

In the (standard) proof below of Lemma 8, by bounding the rate of convergence to the expression in Lemma 6, we also implicitly prove Lemma 6.

*Proof of Lemma 8.* By definition

$$\begin{aligned} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) &= \mathbb{E}[(\Sigma \text{val}(\text{RW}_G^t))^2] - \mathbb{E}[\Sigma \text{val}(\text{RW}_G^t)]^2 \\ &= \sum_{i \in [t]} \sum_{i' \in [t]} \mathbb{E}[\text{val}(\text{RW}_G^t)_i \cdot \text{val}(\text{RW}_G^t)_{i'}] - (p_1t)^2. \end{aligned}$$

The sum on the right hand side above can be simplified as follows. For all  $i = i' \in [t]$ , then  $\mathbb{E}[\text{val}(\text{RW}_G^t)_i \cdot \text{val}(\text{RW}_G^t)_{i'}] = p_1$ . Otherwise, if  $i < i'$ , then  $\mathbb{E}[\text{val}(\text{RW}_G^t)_i \cdot \text{val}(\text{RW}_G^t)_{i'}] = \mathbb{E}[\text{val}(\text{RW}_G^t)_0 \cdot \text{val}(\text{RW}_G^t)_{i'-i}]$ , with an analogous equality if  $i > i'$ . Thus

$$\begin{aligned} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) &= p_1t + 2 \sum_{\ell=1}^{t-1} (t - \ell) \mathbb{E}[\text{val}(\text{RW}_G^t)_0 \cdot \text{val}(\text{RW}_G^t)_\ell] - (p_1t)^2 \\ &= (p_1 - p_1^2)t + 2 \sum_{\ell=1}^{t-1} (t - \ell) \mathbb{E}[\text{val}(\text{RW}_G^t)_0 \cdot \text{val}(\text{RW}_G^t)_\ell - p_1^2] \\ &= p_0p_1t + 2 \sum_{\ell=1}^{t-1} (t - \ell) \mathbb{E}[\text{val}(\text{RW}_G^t)_0 \cdot (\text{val}(\text{RW}_G^t)_\ell - p_1)]. \end{aligned}$$



For  $v \in V$ , conditioned on the event that  $(\text{RW}_G^t)_0 = v$ , then the distribution of  $(\text{RW}_G^t)_\ell$  is given by  $G^\ell \mathbf{1}_v$ . Thus

$$\mathbb{E}[\text{val}(\text{RW}_G^t)_\ell - p_1 \mid (\text{RW}_G^t)_0 = v] = (\text{val} - p_1)^\top G^\ell \mathbf{1}_v,$$

so

$$\begin{aligned} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) &= p_0 p_1 t + 2 \sum_{\ell=1}^{t-1} (t - \ell) \sum_{v \in V} \frac{1}{n} \text{val}(v) \cdot (\text{val} - p_1)^\top G^\ell \mathbf{1}_v \\ &= p_0 p_1 t + 2 \sum_{\ell=1}^{t-1} (t - \ell) \frac{1}{n} (\text{val} - p_1)^\top G^\ell \text{val}. \end{aligned}$$

The difference between  $1/t$  times the above expression and the asymptotic variance as given in Lemma 6 is

$$\begin{aligned} \sigma^2(\Sigma \text{val}(\text{RW}_G^t)) - \frac{1}{t} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) \\ = 2 \sum_{i=1}^{t-1} \frac{i}{t} \cdot \frac{1}{n} (\text{val} - p_1)^\top G^i \text{val} + 2 \sum_{i=t}^{\infty} \frac{1}{n} (\text{val} - p_1)^\top G^i \text{val}. \end{aligned}$$

Because  $G$  is a  $\lambda$ -spectral expander and  $\text{val} - p_1 \in \mathbb{R}^V$  is orthogonal to  $\vec{1}$ ,

$$\begin{aligned} \left| \frac{1}{n} (\text{val} - p_1)^\top G^i \text{val} \right| &= \left| \frac{1}{n} (\text{val} - p_1)^\top G^i (\text{val} - p_1) \right| \\ &= \left| \frac{1}{\sqrt{n}} (\text{val} - p_1)^\top G^i \frac{1}{\sqrt{n}} (\text{val} - p_1) \right| \\ &\leq \left\| \frac{1}{\sqrt{n}} (\text{val} - p_1)^\top \right\| \lambda^i \left\| \frac{1}{\sqrt{n}} (\text{val} - p_1) \right\| \\ &= \lambda^i \cdot p_0 p_1. \end{aligned}$$

Thus

$$\begin{aligned} \left| \sigma^2(\Sigma \text{val}(\text{RW}_G^t)) - \frac{1}{t} \text{Var}(\Sigma \text{val}(\text{RW}_G^t)) \right| &\leq 2 \sum_{i=1}^{t-1} \frac{i}{t} \cdot \lambda^i \cdot p_0 p_1 + 2 \sum_{i=t}^{\infty} \lambda^i \cdot p_0 p_1 \\ &\leq 2 \sum_{i=1}^{\infty} \frac{i}{t} \cdot \lambda^i \cdot p_0 p_1 \\ &= \frac{2}{(1 - \lambda)^2} \cdot \frac{\lambda}{t} \cdot p_0 p_1. \end{aligned}$$

□

*Proof of Proposition 13.* By Lemma 11,  $\sigma^2 = \sigma^2(\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t))$ , that is, the choice of  $\lambda$  is such that  $\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)$  has asymptotic variance  $\sigma^2 = p_0 p_1 (1 + \lambda) / (1 - \lambda)$ . Also by definition,  $\sigma^2 - p_0 p_1 = p_0 p_1 \cdot 2\lambda / (1 - \lambda)$ , and  $-\min\{p_0/p_1, p_1/p_0, 1/100\} \leq \lambda \leq 1/100$ . We apply these facts below to show that  $\mathcal{N}_{\sigma^2}^t = \Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)$  satisfies each of the conditions in Definition 12:

1. By definition  $\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t) = p_1 t$ .

2. By Lemma 8,

$$\begin{aligned}
|\text{Var}(\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t)) - \sigma^2 t| &\leq \frac{2}{(1 - |\lambda|)^2} \cdot |\lambda| \cdot p_0 p_1 \\
&= \frac{2}{(1 - |\lambda|)^2} \cdot \frac{|\sigma^2 - p_0 p_1| \cdot (1 - \lambda)}{2} \\
&\leq \frac{101/100}{(99/100)^2} \cdot |\sigma^2 - p_0 p_1|.
\end{aligned}$$

3. By definition  $\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^1) = \text{Bern}(p)$ .

4. By definition, the distribution of  $\sum_{i \in [\ell]} \Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^{t_k})$  is equal to the distribution of  $\Sigma \text{val}(\text{RW}_G^t)$  for a sequence  $G = (G_i)_{1 \leq i \leq t-1}$  of expander graphs,  $\ell - 1$  of which are  $J$  and  $t - \ell$  of which are  $G_{\lambda,p}$ . Thus because  $\|G_{\lambda,p} - J\| = \lambda(G_{\lambda,p}) = |\lambda|$ , Theorem 2 implies that

$$\begin{aligned}
2 \cdot d_{\text{TV}} \left( \sum_{i \in [\ell]} \Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^{t_k}), \Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t) \right) &\leq 4000 \cdot |\lambda| \cdot \frac{\ell - 1}{t} \\
&\leq 2000(1 - \lambda) \cdot |\sigma^2/p_0 p_1 - 1| \cdot \frac{\ell - 1}{t} \\
&\leq 2020 \cdot |\sigma^2/p_0 p_1 - 1| \cdot \frac{\ell - 1}{t}.
\end{aligned}$$

5. By Theorem 2,

$$\begin{aligned}
&\sum_{j \in \mathbb{Z}: |j - p_1 t| \geq a} |\Pr[\Sigma \text{val}(\text{RW}_{G_{\lambda,p}}^t) = (t - j, j)] - \Pr[\text{Bin}(t, p) = j]| \\
&\leq 4000 \cdot |\lambda| \cdot e^{-a^2/8t} \\
&\leq 2020 \cdot |\sigma^2/p_0 p_1 - 1| \cdot e^{-a^2/8t}.
\end{aligned}$$

6. This fact follows directly from Lemma 4.

□