Electronic Colloquium on Computational Complexity, Report No. 151 (2020)

Pseudobinomiality of the Sticky Random Walk

Venkatesan Guruswami* Vinayak Kumar†

Abstract

Random walks on expanders are a central and versatile tool in pseudorandomness. If an arbitrary half of the vertices of an expander graph are marked, known Chernoff bounds for expander walks imply that the number M of marked vertices visited in a long n-step random walk strongly concentrates around the expected $n/2$ value. Surprisingly, it was recently shown that the parity of M also has exponentially small bias.

Is there a common unification of these results? What other statistics about M resemble the binomial distribution (the Hamming weight of a random n-bit string)? To gain insight into such questions, we analyze a simpler model called the *sticky random walk*. This model is a natural stepping stone towards understanding expander random walks, and we also show that it is a necessary step. The sticky random walk starts with a random bit and then each subsequent bit independently equals the previous bit with probability $(1 + \lambda)/2$. Here λ is the proxy for the expander's (second largest) eigenvalue.

Using Krawtchouk expansion of functions, we derive several probabilistic results about the sticky random walk. We show an asymptotically tight $\Theta(\lambda)$ bound on the total variation distance between the (Hamming weight of the) sticky walk and the binomial distribution. We prove that the correlation between the majority and parity bit of the sticky walk is bounded by $O(n^{-1/4})$. This lends hope to unifying Chernoff bounds and parity concentration, as well as establishing other interesting statistical properties, of expander random walks.

1 Introduction

Expander graphs and random walks on their vertices are an essential and widely employed tool in pseudorandomness, and related areas like coding theory. In this paper, we are interested in a particular simple random walk model where we mark half of the vertices of an expander graph, and look at how many marked vertices an n-step random walk visits. This model has already been explored in many influential works. In particular, expander Chernoff bounds show that the number of marked vertices visited is concentrated around $n/2$ with exponential tails [3, 5]. In his recent breakthrough construction of ε-balanced codes, Ta-Shma [10] proved that the parity of the number of visited marked nodes has exponentially small bias. This fact is quite striking since there are distributions on n bits that are $(n-1)$-wise independent yet have fixed parity, so one might not expect a sensitive function like parity to exhibit such strong concentration for expander random walks in general.

*Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. Email: venkatg@cs.cmu.edu. Supported by NSF grants CCF-1814603 and CCF-1908125.

†Divison of Computing & Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125. Email: vmkumar@caltech.edu. Supported by Caltech's Summer Undergraduate Research Fellowships (SURF) Program and the Stephen Adelman Memorial Fellowship.

ISSN 1433-8092

Let $\boldsymbol{w}$ denote the bit string indicating which steps visit a marked node in an n-step expander random walk starting at a random node. The above two results show nontrivial properties that $\boldsymbol{w}$ shares with a purely random bit string. This naturally raises questions about what other statistical similarities might hold between $\boldsymbol{w}$ and purely random strings. For instance, are the parity bit and majority bit of $\boldsymbol{w}$ (almost) uncorrelated? This can be viewed as a unification of the above-mentioned concentration results for the high order bit and low order bit of the Hamming weight of $\boldsymbol{w}$. More generally, given some arbitrary symmetric property of $\boldsymbol{w}$ (i.e., one that only depends on its Hamming weight), how does its probability deviate in an expander random walk compared to a purely random string? What is the total variation distance (TVD) between the Hamming weight distribution of $\boldsymbol{w}$ and the binomial distribution? Surprisingly, to the best of our knowledge, these and many other similar questions relating to the "pseudobinomiality" properties of the weight of the sequence $\boldsymbol{w}$, seem unexplored.

Towards gaining insight into and making progress on such questions, we analyze a simpler distribution over the n-bit strings called the *sticky random walk* as a necessary stepping stone towards understanding the general expander walk. The sticky random walk $S(n, \lambda)$ is an n-step walk on a Markov chain with two states $0, 1$. The start state is chosen uniformly at random, and at each subsequent step, we stick to the same state with probability $\frac{1+\lambda}{2}$, and change states with probability $\frac{1-\lambda}{2}$. Let $\boldsymbol{s}$ be the n-bit string listing the states visited by $S(n, \lambda)$. The sticky random walk has served as a useful proxy for analyzing the actual expander random walk. Chernoff bounds for a slight variant of the sticky random walk imply similar bounds on the expander random walk, as explicitly pointed out by Kahale [5] and also revisited in the recent work by Rao and Regev [8] who obtained tighter bounds. Such a translation, however, only works for estimating the probability of monotone events. For non-monotone functions like parity, it is not known if one can deduce bounds for the expander random walk from their counterparts for the sticky random walk.

Nevertheless, understanding the simpler sticky random walk model is a meaningful first step towards understanding the general expander walk model. In fact, it is a *necessary* step, as there are expanders where the behavior of the random walk coincides with the sticky walk (i.e., the sequences $\boldsymbol{w}$ and $\boldsymbol{s}$ above are identically distributed). We make this (probably folklore) relationship between these two models explicit in Section 7 (see Theorem 1.4 below).

1.1 Our Results

We answer some of the questions proposed above in the case of the sticky random walk. Our main technique is to represent (a function related to) the probability density function of the Hamming weight of the sticky random walk in the basis of Krawtchouk functions. The Krawtchouk basis is a handy choice to compare this distribution with the binomial distribution. We show the following result on the TVD between these distributions in Section 4. In all our results we think of λ as being fixed and the length of the walk n to be growing.

Theorem 1.1. *The total variation distance between the weight distribution of the sticky random walk and the binomial distribution is* $\Theta(\lambda)$.

Note that the above shows both an upper and lower bound. For the upper bound, we show a stronger claim that an appropriately weighted ℓ_2 distance between the two distributions is bounded from above by $O(\lambda^2)$ for λ small enough (the above then follows by Cauchy-Schwarz). However, as $\lambda \to 1$, the ℓ_2 version of the bound fails to hold, and in fact we prove that this distance grows exponentially in n for $\lambda > 0.95$ (Section 4.2). For the lower bound, the event of the Hamming weight being in the range $n/2 \pm \sqrt{n}$ exhibits a $\Omega(\lambda)$ deviation in probability between these two distributions.

Our next result (in Section 5) shows that the parity and majority bits of the sticky random walk are uncorrelated (up to lower order terms).

Theorem 1.2. *The probability that the Hamming weight of an n-step sticky random walk is even and larger than $n/2$ is $\frac{1}{4} + o(1)$. The same result holds for the other three symmetric cases.*

The explicit lower order term that we derive is $O(n^{-1/4})$, but we believe the true error should be bounded by $O(n^{-1/2})$. The analysis hinges on an upper estimate for the probability that the sticky random walk has weight exactly $\lfloor n/2 \rfloor$, which is again obtained via the Krawtchouk expansion. We note that the arguments for the parity bias from [10] and the Chernoff bound from [3] are quite different. The above theorem hints that Krawtchouk functions might offer a more general tool that can unify these two arguments.

Next, we verify that the residues of the Hamming weight of the sticky walk with respect to any fixed modulus m are almost equidistributed (Section 6). This in particular shows that for any fixed ℓ the ℓ least significant bits of the (binary representation of the) Hamming weight of the sticky random walk are nearly uniformly distributed.

Theorem 1.3. *For any fixed $m \geq 2$, the total variation distance between the residues modulo m of the binomial random variable and the Hamming weight of the n-step sticky walk is at most $\exp(-\Omega_m(n))$.*

Recall that TaShma established such a result for the $m = 2$ case even for the expander random walk [10]. We execute a similar analysis, albeit only for the sticky walk, using m'th roots of unity (in the place ± 1) to track the bias. We suspect extending this analysis to adjacency matrices of expanders (instead of the 2×2 sticky walk transition matrix) can establish this equidistribution result for general expander walks, though we have not verified this.

Establishing the analogs of Theorems 1.1 and 1.2 for expander random walks, however, remain interesting challenges. It is our hope that this work will spur such results.

Finally, the following confirms that analyzing sticky random walks is necessary in order to establish the corresponding claim for expander random walks. We should note that the graph family constructed does not fit the standard definition of an expander family, as degree-boundedness of the graph is not enforced. This is not a significant issue as the analysis of random walks on graphs always proceed by abstracting only the spectral properties of the graph. Further, we believe sampling a sparse regular subgraph with a similar structure should yield a similar claim for a bounded-degree expander family.

Theorem 1.4. *There is a family of regular graphs whose nontrivial eigenvalues are bounded in magnitude by λ, half of whose vertices are marked, such that the bit string indicating which steps of a random walk visits a marked vertex has the same distribution as the sticky random walk.*

We conclude the introduction by mentioning a very interesting work of Bazzi on pseudobinomiality [1]. This work establishes an upper bound on the total variation distance between the binomial distribution and the weight distribution of a δ-biased random sequence, based on the entropy of the latter distribution. One can show that the sticky random walk sequence $\boldsymbol{s}$ is λ-biased (see Lemma 4.1). The bound on TVD in [1] is at best $O(\sqrt{\delta n})$ and is only meaningful when the bias is tiny, which isn't the case for the sticky random walk. Also, calculating the entropy of the weight distribution of $\boldsymbol{s}$ seems as hard, if not harder, than the Krawtchouk based calculations we use to directly bound the total variation distance in Section 4. In fact, it could be the case that the best approach to estimate the entropy of the weight distribution of the sticky random walk is via the entropy-difference bound [2] together with our bound on total variation distance.

1.2 Open Problems

Our results give rise to some immediate follow-up questions. Can the $O(\lambda)$ TVD bound between the sticky and binomial distribution be extended to give a non-trivial bound (that is bounded away from 1) for any fixed $\lambda < 1$? Theorem 1.3 showed that any fixed number of least significant bits of (the Hamming weight of) the sticky walk are near uniform, but can this result be extended to other bits, for example the middle bit? Are there other symmetric properties of the sticky walk that resemble purely random strings? The error term on the non-correlation result claimed in Theorem 1.2 is $O(n^{-1/4})$ but the true bound should probably be $O(n^{-1/2})$.

A more important family of questions relate to extending our results from the sticky walk to the general expander walk model. For instance, can a TVD upper bound between the expander walk and the binomial distribution be shown? Also, can the method of Krawtchouk functions used in Theorem 1.2 give insight towards unifying the Chernoff and parity bias results for expander random walks? In general, can we show distributions of various symmetric functions on expander walks are statistically close to the corresponding distributions on random strings? Moment generating function results from [5, 8] allow bounds or monotone symmetric functions to be lifted from the sticky walk to the expander walk, but no relationships are known for non-monotone functions like parity.

2 Preliminaries

2.1 Conventions and Notation

Asymptotics. In our asymptotic analysis, we will take λ to be constant and observe asymptotics as $n \to \infty$. Take o, O, ω, Ω to be the standard definitions. We will say $f \lesssim g$ to mean $f \leq Cg$ for some absolute constant C independent of n and λ. We also say $f = g + O(h)$ when we mean $|f - g| \leq O(h)$, and analogously for o, Ω, and ω. We also denote $\sim$ to be shorthand for $= (1 + o(1))$.
Miscellaneous. For a bit string s, we will denote $|s|$ to be the Hamming weight of s. $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2. We write $\mathrm{Ber}(p)$ to be the Bernoulli distribution on $\{0, 1\}$ where 1 has probability p and 0 has probability $1-p$, and write $\mathrm{Bin}(n, 1/2)$ to be the binomial distribution of $\sum_{i=1}^{n} b_i$ with independent choices of $b_i \sim \mathrm{Ber}(1/2)$. Let $\mathbb{1}_E$ represents the indicator variable of the event E. When written as a function, $\mathbb{1}_S(i)$ is 1 if $i \in S$, and is 0 otherwise. $[n] = \{1, 2, \ldots, n\}$ denotes the set of the first n positive integers, and $[n]_0 = \{0\} \cup [n]$. $\binom{[n]}{k}$ denotes the set of all size k subsets of $[n]$.

2.2 Definitions

In this paper, we will work with distributions on $[n]_0$ and see how close they are to one another. For this reason, we state the standard definitions of the ℓ_p distance between distributions

Definition 2.1 (ℓ_p Distance/TVD)**.** *Let Y and Z be distributions on $[n]_0$. For $p \geq 1$, we define* *the ℓ_p distance between Y and Z to be*

$$||Y - Z||_p = \left(\sum_{i=0}^{n} |Y(i) - Z(i)|^p\right)^{1/p}.$$

The total variation distance (TVD) between Y and Z is simply $\frac{1}{2}||Y - Z||_1$.

Of course the first type of distribution we should formally define is the sticky random walk itself.

Definition 2.2 (Sticky Random Walk)**.** *The sticky random walk $S(n, \lambda)$ is a distribution on n-bit strings s, where $\boldsymbol{s}_1 \sim Ber(1/2)$, and for $2 \leq i \leq n$ and $b \in \{0, 1\}$, we have $\Pr[\boldsymbol{s}_i = b | \boldsymbol{s}_{i-1} = b] = \frac{1+\lambda}{2}$. In cases where n and λ are evident, only S may be used to denote the distribution.*

Define a bit string to be *homogeneous* if it only consists of 1's or of 0's. Given a bit string, we define a *run* to be a homogeneous substring that isn't a proper substring of another homogeneous substring. Intuitively, the sticky walk can be seen as a Markov chain on two states, where you stick to the same state with probability $\frac{1+\lambda}{2}$ and transition to the other with probability $\frac{1-\lambda}{2}$. Thus the probability of a string is really only dependent on how many consecutive pair of bits are equal (equivalently the number of runs), rather than the precise value of the bits. Another thing to note is that λ can be seen as a parameter measuring the stickiness of a bit to its preceding one. Notice when $\lambda = 0$ there is no stickiness present, and the sticky walk is just n independent coin flips.

2.3 Krawtchouk Functions

To analyze this sticky random walk, we heavily rely on the basis of Krawtchouk functions. Hence we define these functions here and state some standard identities of these functions without proof.

Definition 2.3 (Krawtchouk Functions)**.** *The Krawtchouk function $K_k : [n]_0 \to \mathbb{R}$ is defined to be*

$$K_k(\ell) = \sum_{\substack{y \in \{0,1\}^n \\ |y| = k}} (-1)^{\alpha \cdot y}$$

for each integer $0 \leq \ell \leq n$ and an artbirary n-bit string α of Hamming weight ℓ (the specific choice of α does not matter due to symmetry).

It can be shown that these functions form an orthogonal basis of the functions mapping $[n]_0 \to \mathbb{R}$ with respect to the inner product

$$\langle f, g \rangle = \mathbb{E}_{\boldsymbol{b} \sim \text{Bin}(n,1/2)}[f(\boldsymbol{b})g(\boldsymbol{b})]. \tag{1}$$

From the definition it is not hard to show that (e.g. Section 2.2 of [9]).

$$\mathbb{E}_{\boldsymbol{b} \sim \text{Bin}(n,1/2)}[K_r(\boldsymbol{b})K_s(\boldsymbol{b})] = 0, \tag{2}$$

which shows that the Krawtchouk functions indeed form an orthogonal basis with respect to the inner product in (1). Furthermore, one can verify the following identities hold (see [9]).

$$\mathbb{E}_{\boldsymbol{b} \sim \text{Bin}(n,1/2)}[K_k(\boldsymbol{b})^2] = \binom{n}{k} \tag{3}$$

$$\binom{n}{\ell} K_k(\ell) = \binom{n}{k} K_\ell(k). \tag{4}$$

Identities (2) and (3) allow us to explicitly expand any function uniquely as a sum of Krawtchouk functions.

Proposition 2.4. *For function $f : [n]_0 \to \mathbb{R}$, there exists a unique expansion $f(\ell) = \sum_{k=0}^{n} \widehat{f}(k) K_k(\ell)$ where*

$$\widehat{f}(k) = \frac{1}{\binom{n}{k}} \mathbb{E}_{\boldsymbol{b} \sim Bin(n,1/2)}[f(\boldsymbol{b}) K_k(\boldsymbol{b})]$$

for each integer $0 \le k \le n$.

3 Using Krawtchouk Functions to Analyze the Sticky Walk

To help us analyze the sticky walk, we define the function $p : [n]_0 \to \mathbb{R}$ to be $p(\ell) = \frac{\Pr_{\boldsymbol{s} \sim S}[|\boldsymbol{s}|=\ell]}{\binom{n}{\ell} 2^{-n}}$, and look at the Krawtchouk expansion. Doing so results in the following lemma.

Lemma 3.1. *We have*

$$\widehat{p}(k) = \frac{1}{\binom{n}{k}} \mathbb{E}_{\boldsymbol{s} \sim S}[K_k(|\boldsymbol{s}|)].$$

Proof. We just apply Proposition 2.4 to get that

$$\begin{aligned}
\widehat{p}(k) &= \frac{1}{\binom{n}{k}} \mathbb{E}_{\boldsymbol{b} \sim \mathrm{Bin}(n,1/2)}[p(\boldsymbol{b}) K_k(\boldsymbol{b})] \\
&= \frac{1}{\binom{n}{k}} \sum_{b=0}^{n} \binom{n}{b} 2^{-n} p(\boldsymbol{b}) K_k(\boldsymbol{b}) \\
&= \frac{1}{\binom{n}{k}} \sum_{b=0}^{n} \Pr_{\boldsymbol{s} \sim S}[|\boldsymbol{s}| = b] K_k(\boldsymbol{b}) \\
&= \frac{1}{\binom{n}{k}} \mathbb{E}_{\boldsymbol{s} \sim S}[K_k(|\boldsymbol{s}|)].
\end{aligned}$$

□

What follows is a useful lemma that displays how Krawtchouk expanding $p(\ell)$ can help analyze probabilities of the sticky distribution.

Lemma 3.2. *For $\boldsymbol{s} \sim S(n, \lambda)$, we can evaluate*

$$\Pr[|\boldsymbol{s}| = \ell] = \frac{1}{2^n} \sum_{k=0}^{n} K_\ell(k) \mathbb{E}[K_k(|\boldsymbol{s}|)]$$

Proof. To estimate this probability, we can use Lemma 3.1 to find

$$\begin{aligned}\Pr[|\boldsymbol{s}| = \ell] &= \binom{n}{\ell} 2^{-n} p(\ell) \\ &= \frac{\binom{n}{\ell}}{2^n} \sum_{k=0}^{n} \widehat{p}(k) K_k(\ell) \\ &= \frac{1}{2^n} \sum_{k=0}^{n} \frac{\binom{n}{\ell} K_k(\ell) \mathbb{E}[K_k(|\boldsymbol{s}|)]}{\binom{n}{k}} \\ &= \frac{1}{2^n} \sum_{k=0}^{n} K_\ell(k) \mathbb{E}[K_k(|\boldsymbol{s}|)]\end{aligned}$$

where we used the reciprocity relation (4) at the end. □

Notice that these lemmas didn't depend on the specific distribution $S(n, \lambda)$, and so these lemmas are applicable for arbitrary distributions on $[n]_0$. For our purposes, the sticky walk versions will be used for our analysis done in the next sections.

4 Total Variation Distance Bounds

4.1 Upper Bounding the TVD Between the Sticky and Binomial Distribution

Upper bounding the TVD requires calculating the expectation of various sums and products of sticky walk random variables. We will abstract out these calculations in the following lemmas. Analogous expressions for the general expander walk model were also analyzed in Lemmas 3.3 and 4.2 in Rao and Regev [8]. In particular, [8] gives upper bounds on these quantities in the general expander walk, while we give exact values for the simpler sticky walk.

Lemma 4.1. *Let $\boldsymbol{s} \sim S(n, \lambda)$. For even-sized subsets $A \subset [n]$ where $a_1 < \cdots < a_m$ are the elements of A in increasing order, define* $\text{shift}(A) = \sum_{i=1}^{|A|/2} (a_{2i} - a_{2i-1})$ *. For any $A \subset [n]$, we have*

$$\mathbb{E}\left[\prod_{i \in A} (-1)^{\boldsymbol{s}_i}\right] = \begin{cases} 0 & |A| \text{ odd} \\ \lambda^{\text{shift}(A)} & |A| \text{ even} \end{cases}.$$

Proof. Since the sticky walk is a Markov chain where $(-1)^{\boldsymbol{s}_i}$ has the same sign as $(-1)^{\boldsymbol{s}_{i-1}}$ with probability $\frac{1+\lambda}{2}$, we can rewrite these random variables in terms of a product of independent random variables representing the transitions of the chain. In particular, we define a new random variable $u \in \{0,1\}^n$, where $\boldsymbol{u}_1 \sim \text{Ber}(1/2)$ and $\boldsymbol{u}_i \sim \text{Ber}\left(\frac{1-\lambda}{2}\right)$ for $2 \le i \le n$. One can easily check that $(-1)^{\boldsymbol{s}_i}$ is the same random variable as $\prod_{j=1}^{i} (-1)^{\boldsymbol{u}_j} = (-1)^{\sum_{j=1}^{i} \boldsymbol{u}_j}$. Hence we have

$$\mathbb{E}\left[\prod_{i \in A} (-1)^{\boldsymbol{s}_i}\right] = \mathbb{E}[(-1)^{\sum_{i \in A} \sum_{j=1}^{i} \boldsymbol{u}_j}] = \prod_{j=1}^{a_m} \mathbb{E}[(-1)^{\sum_{i \in A; i \ge j} \boldsymbol{u}_j}]$$

Note when $|A|$ is odd, the factor when $j = 1$ is 0 since $\mathbb{E}[(-1)^{\sum_{i \in A} \boldsymbol{u}_1}] = \mathbb{E}[(-1)^{\boldsymbol{u}_1}] = 0$. Hence the total expectation is 0. Otherwise, when $|A|$ is even, the $j = 1$ term is just $\mathbb{E}[(-1)^{|A|\boldsymbol{u}_1}] = 1$. For $j \ge 2$, one sees that if $A_j = \{i \in A : i \ge j\}$ is of odd cardinality, then $\mathbb{E}[(-1)^{\sum_{i \in A; i \ge j} \boldsymbol{u}_j}] = \mathbb{E}[(-1)^{\boldsymbol{u}_j}] = \lambda$, and is 1 otherwise. The set of j such that $|A_j|$ is odd is simply the integers in

$(a_1, a_2] \cup (a_3, a_4] \cup \cdots \cup (a_{m-1}, a_m]$, of which there are $\text{shift}(A)$. Consequently, upon multiplying all the j factors, we derive that the expectation is $\lambda^{\text{shift}(A)}$ for $|A|$ even. $\square$

Lemma 4.2. *For all nonnegative integers k and $\boldsymbol{s} \sim S(n, \lambda)$,*

- $\mathbb{E}[K_{2k}(|\boldsymbol{s}|)] = \sum_{m=k}^{n-k} \binom{m-1}{k-1}\binom{n-m}{k}\lambda^m$, *and*
- $\mathbb{E}[K_{2k+1}(|\boldsymbol{s}|)] = 0$.

Proof. Using the definition, we rewrite

$$K_k(|\boldsymbol{s}|) = \sum_{\substack{\alpha \in \{0,1\}^n \\ |\alpha| = k}} (-1)^{\sum_{i=1}^n \alpha_i \boldsymbol{s}_i} = \sum_{T \in \binom{[n]}{k}} (-1)^{\sum_{i \in T} \boldsymbol{s}_i} = \sum_{T \in \binom{[n]}{k}} \prod_{i \in T} (-1)^{\boldsymbol{s}_i}. \tag{5}$$

Then from Lemma 4.1, we have

$$\mathbb{E}[K_k(|\boldsymbol{s}|)] = \sum_{T \in \binom{[n]}{k}} \mathbb{E}\left[\prod_{i \in T} (-1)^{\boldsymbol{s}_i}\right] = \begin{cases} 0 & k \text{ odd} \\ \sum_{T \in \binom{[n]}{k}} \lambda^{\text{shift}(T)} & k \text{ even} \end{cases}$$

where f is defined as in Lemma 4.1. We now evaluate

$$\mathbb{E}[K_{2k}(|\boldsymbol{s}|)] = \sum_{T \in \binom{[n]}{2k}} \lambda^{\text{shift}(T)} = \sum_{m=k}^{n-k} \Bigg(\sum_{\substack{T \in \binom{[n]}{2k} \\ \text{shift}(T) = m}} 1 \Bigg) \lambda^m$$

Hence to prove the lemma, it suffices to show the number of subsets $T \in \binom{[n]}{2k}$ with $\text{shift}(T) = m$ is $\binom{m-1}{k-1}\binom{n-m}{k}$. Let $t_1 < t_2 < \cdots < t_{2k}$ be the elements of T. We claim the desired subsets T are in bijection with a k-tuple $(d_1, \ldots, d_k)$ of positive integers summing to m, paired with an $n-m$ letter word consisting of $n-m-k$ A's and k B's. To construct the pair from a subset T, we can set $d_i = t_{2i} - t_{2i-1}$ and the $n-m$ letter word to be $A^{t_1-1}CBA^{n-t_{2k}}$ where $C = \bigcirc_{i=1}^{k-1}(BA^{t_{2i+1}-t_{2i}-1})$ ($\circ$ is concatenation). Since $\text{shift}(T) = m$, the d_i are positive integers summing to m, and the word can easily be verified to have $n-m-k$ A's and k B's.

Now given a $(d_1, \ldots, d_k)$ and a word, we can construct a T with $\text{shift}(T) = m$ as follows. The word will be of the form $A^{a_1}BA^{a_2}B \cdots A^{a_k}BA^{a_{k+1}}$ where each $a_i \geq 0$. We then set $t_1 = a_1 + 1$, and assign $t_2, \ldots, t_{2k}$ inductively as follows.

$$t_{2i} = t_{2i-1} + d_i$$

$$t_{2i+1} = t_{2i} + a_{i+1} + 1$$

Finally we set $T = \{t_i\}_{i=1}^{2k}$. It can then be verified all $t_i \in [n]$ and $\text{shift}(T) = m$. In fact, one can check the two constructed maps are inverses of each other. Hence we have established the bijection and counting the number of such T is simply counting the number of k-tuples and word pairs, which by standard counting methods is $\binom{m-1}{k-1}\binom{n-m}{k}$. The desired result follows. $\square$

We are now ready to establish some upper bounds using the lemmas above. We first upper bound the weighted ℓ_2 distance between the sticky walk and binomial distribution by defining a suitable function and taking the Krawtchouk expansion.

Lemma 4.3. *Let $\boldsymbol{s} \sim S(n, \lambda)$ for $\lambda < 0.16$ and define $p(\ell) := \frac{\Pr_{\boldsymbol{s}\sim S}[|\boldsymbol{s}|=\ell]}{\binom{n}{\ell}2^{-n}}$. Then we have*

$$\mathbb{E}_{\boldsymbol{b}\sim Bin(n,1/2)}\left[(p(\boldsymbol{b})-1)^2\right] \leq O(\lambda^2)\ .$$

Proof. Let us write $p(\ell) = \sum_{k=0}^{n} \widehat{p}(k) K_k(\ell)$ where

$$\widehat{p}(k) = \frac{1}{\binom{n}{k}} \mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[p(\boldsymbol{b})K_k(\boldsymbol{b})] \tag{6}$$

from Proposition 2.4. From the definition, we can verify $K_0(\ell) = 1$ for all ℓ. Combining this fact with (6) implies that $\widehat{p}(0) = 1$ too. Hence,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[(p(\boldsymbol{b})-1)^2] &= \mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}\left[\left(\sum_{k=1}^{n} \widehat{p}(k)K_k(\boldsymbol{b})\right)^2\right] \\
&= \sum_{k=1}^{n} \widehat{p}(k)^2 \binom{n}{k} \\
&= \sum_{k=1}^{n} \frac{1}{\binom{n}{k}} \mathbb{E}[K_k(|\boldsymbol{s}|)]^2 \qquad (7)
\end{aligned}$$

where only the diagonal terms of the square survive due to the orthogonality relations (2) and (3). Adding the $k = 0$ term back in (7) will simply give the expression for $\mathbb{E}[p(\boldsymbol{b})^2]$ (and will be stated as a subsequent corollary).

From Lemma 4.2 and the generating function relation $(\frac{x}{1-x})^k = \sum_{m\geq k}\binom{m-1}{k-1}x^m$,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[(p(\boldsymbol{b})-1)^2] &= \sum_{1\leq k\leq n/2} \frac{1}{\binom{n}{2k}} \left(\sum_{m=k}^{n-k}\binom{m-1}{k-1}\binom{n-m}{k}\lambda^m\right)^2 \qquad (8) \\
&\leq \sum_{1\leq k\leq n/2} \frac{\binom{n}{k}^2}{\binom{n}{2k}} \left(\sum_{m=k}^{n-k}\binom{m-1}{k-1}\lambda^m\right)^2 \\
&\leq \sum_{1\leq k\leq n/2} \frac{\binom{n}{k}^2}{\binom{n}{2k}} \left(\frac{\lambda}{1-\lambda}\right)^{2k} \qquad (9)
\end{aligned}$$

With the following claim (whose proof will be deferred to the appendix) we can deduce a lower bound.

Claim 4.4. *For $1 \leq k \leq n/2$, we can bound $\frac{\binom{n}{k}^2}{\binom{n}{2k}} \leq \frac{e^3\sqrt{2}}{4\pi^2}\cdot 4^{2k}$.*

Hence by combining (9) and Claim 4.4, we conclude

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[(p(\boldsymbol{b})-1)^2] &\leq \frac{e^3\sqrt{2}}{4\pi^2}\sum_{1\leq k\leq n/2}\left(\frac{4\lambda}{1-\lambda}\right)^{2k} \\
&\leq \frac{4e^3\sqrt{2}}{\pi^2}\frac{\lambda^2}{(1-5\lambda)(1+3\lambda)} \\
&< \left(\frac{\lambda}{0.16}\right)^2,
\end{aligned}$$

which is a nontrivial $O(\lambda^2)$ bound that is strictly less than 1. ☐

Remembering our prior observation that the sum in Equation (7) is 1 less than $\mathbb{E}[p(\boldsymbol{b})^2]$, we immediately get the following corollary.

Corollary 4.5. *Let $\boldsymbol{s}, p$ be defined as in Lemma 4.3. We can then bound*

$$\mathbb{E}_{\boldsymbol{b}\sim Bin(n,1/2)}[p(\boldsymbol{b})^2] = \sum_{k=0}^{n} \frac{1}{\binom{n}{k}} \mathbb{E}[K_k(|\boldsymbol{s}|)]^2 \leq 1 + O(\lambda^2).$$

Going back to proving the main result of the section, the brunt of the work was actually done in Lemma 4.3. We can now simply apply convexity to establish the desired TVD upper bound between the sticky and binomial distribution.

Theorem 4.6. *Let $\boldsymbol{s} \sim S(n, \lambda)$ with $\lambda < 0.16$. We can then bound the TVD between $S(n,\lambda)$ and $Bin(n, 1/2)$,*

$$\frac{1}{2}\sum_{\ell=0}^{n} |\Pr[|\boldsymbol{s}| = \ell] - \binom{n}{\ell} 2^{-n}| \leq O(\lambda).$$

Proof. By convexity, and then Lemma 4.3, we get

$$\sum_{\ell=0}^{n} |\Pr[|\boldsymbol{s}| = \ell] - \binom{n}{\ell} 2^{-n}| = \mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[|p(\boldsymbol{b}) - 1|] \leq \sqrt{\mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[(p(\boldsymbol{b}) - 1)^2]} \leq O(\lambda)$$

for $\lambda < 0.16$. $\square$

4.2 Limitations to the Upper Bound Approach

Unfortunately, the weighted ℓ_2 distance studied in Lemma 4.3 blows up as λ approaches 1. Looking at the term in the sum of Equation (8) when $k = m = (1/2 - \varepsilon)n$ for some $\varepsilon > 0$ (will be specified later), and applying the bounds $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$ yields

$$\begin{aligned}
\sum_{1\leq k\leq n/2} \frac{1}{\binom{n}{2k}} \left(\sum_{m=k}^{n-k} \binom{m-1}{k-1}\binom{n-m}{k}\lambda^m\right)^2 &\geq \frac{1}{\binom{n}{(1-2\varepsilon)n}} \binom{(1/2+\varepsilon)n}{(1/2-\varepsilon)n}^2 \lambda^{(1/2-\varepsilon)n} \\
&\geq \binom{n}{2\varepsilon n}^{-1} \binom{(1/2+\varepsilon)n}{2\varepsilon n}^2 \lambda^{n/2} \\
&\geq \left(\frac{\lambda^{1/2}}{(8e\varepsilon)^{2\varepsilon}}\right)^n
\end{aligned}$$

For $\lambda \geq .94$ and $\varepsilon = .017$, we have $\lambda^{1/2} > .969$ and $(8e\varepsilon)^{2\varepsilon} < .967$. Thus

$$\mathbb{E}_{\boldsymbol{b}\sim \mathrm{Bin}(n,1/2)}[(p(\boldsymbol{b}) - 1)^2] > \left(\frac{.969}{.967}\right)^n$$

which grows exponentially in n. Consequently, in order to show an $O(\lambda)$ bound for all $\lambda < 1$, an approach different from using the weighted ℓ_2 distance is required. Nevertheless, for the values of λ in which this TVD bound holds, we can show the bound is tight.

4.3 Showing the TVD Bound is Tight

Consider $\boldsymbol{s} \sim S(n, \lambda)$ and $\boldsymbol{N} \sim \mathcal{N}(0,1)$. Set $\boldsymbol{Y}_i = (-1)^{\boldsymbol{s}_i}$ and let $\boldsymbol{Y} = \sum_{i=1}^n (-1)^{\boldsymbol{s}_i}$ be the ± 1 variant of the sticky distribution. Similarly, let $b_i \sim \mathrm{Ber}(1/2)$ for $1 \le i \le n$. Set $\boldsymbol{Z}_i = (-1)^{b_i}$ and let $\boldsymbol{Z} = \sum_{i=1}^n \boldsymbol{Z}_i$ be the usual ± 1 unbiased random walk. By the central limit theorem, we already know $\frac{\boldsymbol{Z}}{\sqrt{n}} \to \boldsymbol{N}$ in distribution as $n \to \infty$. We now state a similar result for $\boldsymbol{Y}$, and defer the proof to the appendix, as it requires lengthy calculation of the moments.

Lemma 4.7. *As $n \to \infty$, $\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}} \to \boldsymbol{N}$ in distribution.*

With this lemma, we have a good understanding of Y, and can show the tightness of the TVD bound. The suggestion to consider an event like $|\boldsymbol{Y}| \le O(\sqrt{n})$ was made by Salil Vadhan [11].

Theorem 4.8. *For $\boldsymbol{Y}$ and $\boldsymbol{Z}$ defined above as the ± 1 version of the $S(n, \lambda)$ walk and the n-step unbiased walk, respectively, we have*

$$|\Pr[|\boldsymbol{Z}| \le \sqrt{n}] - \Pr[|\boldsymbol{Y}| \le \sqrt{n}]| \ge \Omega(\lambda)$$

for $\lambda < 1$.

Proof. By Lemma 4.7, for $\varepsilon < \int_{\sqrt{\frac{1-\lambda}{1+\lambda}}}^1 e^{-x^2} dx$, there exists a constant $N(\varepsilon)$ such that for $n > N(\varepsilon)$,

$$\begin{aligned} |\Pr[|\boldsymbol{Z}| \le \sqrt{n}] - \Pr[|\boldsymbol{Y}| \le \sqrt{n}]| &= \left| \Pr\left[\frac{|\boldsymbol{Z}|}{\sqrt{n}} \le 1\right] - \Pr\left[|\boldsymbol{Y}|\sqrt{\frac{1-\lambda}{(1+\lambda)n}} \le \sqrt{\frac{1-\lambda}{1+\lambda}}\right] \right| \\ &\ge 2\int_{\sqrt{\frac{1-\lambda}{1+\lambda}}}^1 e^{-x^2} dx - \varepsilon \ge \int_{\sqrt{\frac{1-\lambda}{1+\lambda}}}^1 e^{-x^2} dx \ge e^{-1}\left(1 - \sqrt{\frac{1-\lambda}{1+\lambda}}\right) \ge \frac{\lambda}{2e} \end{aligned}$$

where the last step follows from the easily verifiable inequality $1 - \sqrt{\frac{1-x}{1+x}} \ge \frac{x}{2}$ for $0 \le x \le 1$. Hence we have demonstrated an event that gives an $\Omega(\lambda)$ gap between the distributions of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ for $\lambda < 1$. Since $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are just shifted and dilated versions of the Hamming weight of $S(n, \lambda)$ and $\mathrm{Bin}(n, 1/2)$, respectively, we can deduce the TVD is $\Theta(\lambda)$ for $\lambda < 0.16$ due to Theorem 4.6. □

5 Parity and Majority of the Sticky Walk are Almost Uncorrelated

Following the spirit of how Ta-Shma [10] showed the bias of the parity of an expander walk sampler is exponentially small, we show some parity events of the sticky random walk are very close in probability to the corresponding probability under the binomial distribution. An interesting question to consider is whether more refined events, such as whether the output is even and above expectation is close to $\frac{1}{4}$. We show such results for the sticky distribution. Note that from Theorem 4.6 we know the event probability in the sticky distribution will be within $O(\lambda)$ of $\frac{1}{4}$, but in this section, we derive an $o(1)$ error bound.

Let $\boldsymbol{s} \sim S(n, \lambda)$. We first demonstrate that similar to the binomial distribution, $\Pr[|\boldsymbol{s}| \text{ even}]$ and $\Pr[|\boldsymbol{s}| \ge n/2]$ are close to $1/2$. For the parity, we straightforwardly calculate

$$\Pr[|\boldsymbol{s}| \text{ even}] - \Pr[|\boldsymbol{s}| \text{ odd}] = \sum_{t=0}^n (-1)^t \Pr[|\boldsymbol{s}| = t] = \mathbb{E}[(-1)^{|\boldsymbol{s}|}] = \mathbb{E}\left[\prod_{i=1}^n (-1)^{\boldsymbol{s}_i}\right] = \lambda^{n/2} \cdot \mathbb{1}_{n \text{ even}} \tag{10}$$

using Lemma 4.1. Hence $|\Pr[|\boldsymbol{s}| \text{ even}] - \frac{1}{2}| \leq \frac{\lambda^{n/2}}{2}$. Interestingly enough, this $\lambda^{n/2}$ bias is also present in the parity bias calculation done by Ta-Shma for the expander walk in [10] (Section 3.2). Furthermore, this calculation (combined with our work in Section 7) shows that this error term cannot be improved to something smaller like λ^n.

Let a be a string and let $\overline{a}$ be the string formed by toggling every bit of a. Notice that $\Pr[\boldsymbol{s} = a] = \Pr[\boldsymbol{s} = \overline{a}]$ since the number of runs are the same in both strings. Summing over all strings of Hamming weight t, we get the symmetric relation $\Pr[|\boldsymbol{s}| = t] = \Pr[|\boldsymbol{s}| = n - t]$, which directly implies $\Pr[|\boldsymbol{s}| > n/2] = \Pr[|\boldsymbol{s}| < n/2]$. In the case n is odd, we immediately have $\Pr[|\boldsymbol{s}| > n/2] = 1/2$. When n is even, we have

$$\Pr[|\boldsymbol{s}| > n/2] = \frac{1}{2} - \frac{\Pr[|\boldsymbol{s}| = n/2]}{2}. \tag{11}$$

We now prove a lemma that will help us bound $\Pr[|\boldsymbol{s}| = n/2]$ in the sticky distribution.

Theorem 5.1. *For $\boldsymbol{s} \sim S(n, \lambda)$, $\lambda < 1/5$, and $\boldsymbol{b} \sim Bin(n, 1/2)$, we can bound*

$$\Pr[|\boldsymbol{s}| = \ell] \lesssim \sqrt{(1 + O(\lambda^2)) \Pr[b = \ell]}.$$

Proof. To estimate this probability, we can use Lemma 3.2) to find

$$\Pr[|\boldsymbol{s}| = \ell] = \frac{1}{2^n} \sum_{k=0}^{n} K_\ell(k) \mathbb{E}[K_k(|\boldsymbol{s}|)] \leq \frac{1}{2^n} \sqrt{\sum_{k=0}^{n} \binom{n}{k} K_\ell(k)^2} \sqrt{\sum_{k=0}^{n} \frac{\mathbb{E}[K_k(|\boldsymbol{s}|)]^2}{\binom{n}{k}}} \tag{12}$$

by Cauchy-Schwarz. Notice by Corollary 4.5,

$$\sqrt{\sum_{k=0}^{n} \frac{\mathbb{E}[K_k(|\boldsymbol{s}|)]^2}{\binom{n}{k}}} \lesssim \sqrt{1 + O(\lambda^2)}. \tag{13}$$

Finally, by using Equation (3) we get

$$\sqrt{\sum_{k=0}^{n} \binom{n}{k} K_\ell(k)^2} = \sqrt{\binom{n}{\ell} 2^n} \tag{14}$$

Combining (12), (13), and (14) yields

$$\Pr[|\boldsymbol{s}| = \ell] \lesssim \frac{1}{2^n} \sqrt{\binom{n}{\ell} 2^n (1 + O(\lambda^2))} = \sqrt{(1 + O(\lambda^2)) \Pr[b = \ell]}$$

as desired. □

Going back to (11), Theorem 5.1 with $\ell = n/2$ now allows us to deduce

$$\Pr[|\boldsymbol{s}| > n/2] = \frac{1}{2} + O(n^{-1/4}\sqrt{1 + \lambda^2}).$$

Thus, we have shown that $\Pr[|\boldsymbol{s}| \text{ even}]$ and $\Pr[|\boldsymbol{s}| > n/2]$ are near $1/2$, which are properties shared by purely random strings. However, we can go further and show a more refined equidistribution result. One can calculate for $\boldsymbol{b} \sim \text{Bin}(n, 1/2)$ that $\Pr[(|b| \text{ odd}) \wedge (|b| > n/2)] = \frac{1}{4} + O(n^{-1/2})$. We show an analogous result for the sticky walk, albeit with a worse $o(1)$ error term.

Theorem 5.2. *Let $\boldsymbol{s} \sim S(n, \lambda)$. For $a, b \in \{0, 1\}$, denote the event*

$$E_{ab} = (|\boldsymbol{s}| \equiv a \pmod 2) \wedge ((-1)^b |\boldsymbol{s}| > (-1)^b n/2)$$

and $p_{ab} = \Pr[E_{ab}]$. Then $p_{ab} = \frac{1}{4} + O(n^{-1/4}\sqrt{1+\lambda^2})$ for all a, b.

Proof. We show the result for p_{00} (symmetric arguments work for any p_{ab}). We split into cases when n is even and odd. When n is even, notice that due to the symmetry $\Pr[|\boldsymbol{s}| = t] = \Pr[|\boldsymbol{s}| = n - t]$, $p_{00} = p_{01}$ and $p_{10} = p_{11}$. Hence from (10) we have $\lambda^{n/2} \geq |p_{10} + p_{11} - p_{01} - p_{00}| = 2|p_{10} - p_{00}|$. From (11) and Theorem 5.1, we have $p_{10} + p_{00} = \frac{1}{2} + O(n^{-1/4}\sqrt{1+\lambda^2})$, and so we can conclude $p_{00} = \frac{1}{4} + O(n^{-1/4}\sqrt{1+\lambda^2})$ as desired.

When n is odd, a little more effort is required to exploit symmetry. Let S be the set of n-bit strings having a run of size ≥ 2 which doesn't contain the nth bit, and let T be the set of n-bit strings having a run of size ≥ 2 which doesn't contain the 1st bit. Consider the map $g : S \to T$ defined by toggling the last bit of the first run of size ≥ 2. It can be easily seen that g is a bijection. Note that for any $s \in S$, s and $g(s)$ have the same probability under the sticky walk distribution (the map preserves the number of runs in the string). Furthermore, f changes the parity of the input string. The image of all strings in S with even parity and Hamming weight $> n/2$ under g will be all strings in T with odd parity and Hamming weight $> n/2$, along with some rogue strings in T with Hamming weight $(n-1)/2$. If we isolate these rogue cases, we can deduce, due to the fact g preserves probability, that

$$\begin{aligned}\Pr[(\boldsymbol{s} \in S) \wedge E_{00}] &= \Pr[(\boldsymbol{s} \in T) \wedge E_{10}] + \Pr\left[(\boldsymbol{s} \in S) \wedge \left(|g(\boldsymbol{s})| = \frac{n-1}{2}\right) \wedge E_{00}\right] \\ &\leq \Pr[(\boldsymbol{s} \in T) \wedge E_{10}] + O(n^{-1/4}\sqrt{1+\lambda^2}) \end{aligned} \tag{15}$$

by Theorem 5.1 with $\ell = \frac{n-1}{2}$. Notice any n-bit string has probability measure at most $\frac{1}{2}\left(\frac{1+\lambda}{2}\right)^{n-1}$ in the sticky distribution. Furthermore, there are only $O(n)$ strings not contained in S (such strings must be of form $0^k A$ or $1^k A$ where A is a binary string that alternates 0s and 1s). Therefore we can bound

$$|\Pr[(\boldsymbol{s} \in S) \wedge E_{00}] - \Pr[E_{00}]| \leq \Pr[\boldsymbol{s} \notin S] \lesssim n\left(\frac{1+\lambda}{2}\right)^n. \tag{16}$$

An analogous argument with set T gives

$$|\Pr[(\boldsymbol{s} \in T) \wedge E_{01}] - \Pr[E_{01}]| \lesssim n\left(\frac{1+\lambda}{2}\right)^n. \tag{17}$$

Combining (15), (16), and (17) using an application of the triangle inequality allows us to bound

$$\begin{aligned}|p_{00} - p_{01}| &\leq |\Pr[E_{00}] - \Pr[(\boldsymbol{s} \in S) \wedge E_{00}]| + |\Pr[(\boldsymbol{s} \in S) \wedge E_{00}] - \Pr[(\boldsymbol{s} \in T) \wedge E_{01}]| \\ &\quad + |\Pr[(\boldsymbol{s} \in T) \wedge E_{01}] - \Pr[E_{01}]| \\ &\lesssim n^{-1/4}\sqrt{1+\lambda^2}.\end{aligned}$$

Since $p_{00} + p_{01} = 1/2$ for n odd, we can deduce $p_{00} = \frac{1}{4} + O(n^{-1/4}\sqrt{1+\lambda^2})$. □

One thing to note is that Theorem 5.1 is not tight (taking $\lambda = 0$ demonstrates this) solely due to the application of Cauchy-Schwarz in (12). Tightening this bound (at least for ℓ near $n/2$) will improve the error term in Theorem 5.2 as an immediate consequence.

6 Sticky Walk Modulo m is Close to Uniform

In [10], Ta-Shma provides an argument on how the parity of the expander walk is unbiased. In this section we use his method to show the sticky distribution modulo m is approximately uniform for any fixed m and large n.

Lemma 6.1. *Fix $\lambda < 1$ and $m \geq 2$. For any nontrivial m'th root of unity $\zeta \neq 1$ and $\boldsymbol{s} \sim S(n, \lambda)$, we have $|\mathbb{E}[\zeta^{|\boldsymbol{s}|}]| \leq \exp(-\Omega(n))$.*

Proof. One can verify $E[\zeta^{|\boldsymbol{s}|}] = \mathbf{1}^\intercal M^n \mathbf{1}$ where $\mathbf{1}$ is the unit vector with all coordinates equal and

$$M = \begin{pmatrix} \frac{1+\lambda}{2} & \frac{1-\lambda}{2} \\ \frac{1-\lambda}{2} & \frac{1+\lambda}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \zeta \end{pmatrix} = \begin{pmatrix} \frac{1+\lambda}{2} & \frac{1-\lambda}{2}\zeta \\ \frac{1-\lambda}{2} & \frac{1+\lambda}{2}\zeta \end{pmatrix}.$$

Let λ_1, λ_2 be the eigenvalues of M. If M is diagonalizable, we know $\mathbf{1}^\intercal M^n \mathbf{1} = c_1\lambda_1^n + c_2\lambda_2^n$, and if M is not diagonalizable we can write $\mathbf{1}^\intercal M^n \mathbf{1} = (c_1 + nc_2)\lambda_1^n$ for constants c_1 and c_2 (this can be seen by writing M in Jordan normal form). Thus in either case, it suffices to show the norm of both eigenvalues are strictly less than 1 to prove the lemma. From Gershgorin's circle theorem, we know the norms of the eigenvalues are at most $\frac{1+\lambda}{2} + \frac{1-\lambda}{2} = 1$.

Assume one eigenvalue has norm 1. Since the product of the eigenvalues is $\det(M) = \lambda\zeta$, we deduce the eigenvalues are $\lambda z, \zeta/z$ for some $|z| = 1$. Since $\lambda z + \zeta/z = \mathrm{Tr}(M) = \frac{1+\lambda}{2}(1+\zeta)$, we particularly know $\lambda z + \zeta/z$ and $1 + \zeta$ have the same argument angle.

If $\zeta \neq -1$, we must have the quotient of these two complex numbers be real. Consequently

$$0 = \frac{\lambda z + \zeta/z}{1+\zeta} - \overline{\left(\frac{\lambda z + \zeta/z}{1+\zeta}\right)} = \frac{\lambda z + \zeta/z}{1+\zeta} - \frac{\lambda/z + z/\zeta}{1 + 1/\zeta} = \frac{(\lambda - 1)(z - \zeta/z)}{1+\zeta}.$$

Since $\lambda \neq 1$, we have $z = \zeta/z$ and so the eigenvalues are $\lambda z, z$. The trace is $\frac{1}{2}(1+\lambda)(1+\zeta) = \frac{1}{2}(1+\lambda)(1+z^2)$, so

$$\lambda z + z = \frac{1}{2}(1+\lambda)(1+z^2) \iff z = \frac{1+z^2}{2} \iff z = 1.$$

However $\zeta = z^2 = 1$, a contradiction.

If $\zeta = -1$, then $\lambda z, -1/z$ are the eigenvalues. The trace is now zero, so $\lambda z = 1/z$. Since λ is a nonnegative real, taking norms of both sides implies $\lambda = 1$, a contradiction. Hence the eigenvalues are indeed less than 1. $\square$

Lemma 6.1 allows us to employ Lemma 4.2 in [7] (since all characters of $\mathbb{Z}/m\mathbb{Z}$ are $\chi(n) = \zeta^n$ for some m'th root of unity ζ) to deduce the following.

Theorem 6.2. *Let U_m be the uniform distribution over $[m]$, and let $S_m(n, \lambda)$ be the distribution of $|s| \pmod m$ over $[m]$, where $\boldsymbol{s} \sim S(n, \lambda)$. The ℓ_1 distance between these two distributions can then be bounded by*

$$||U_m - S_m(n, \lambda)||_1 \leq \exp(-\Omega(n))$$

where the implied constants only depend on m.

7 Relationship Between the Sticky Walk and Expander Walk

Define a λ-expander[1] to be a regular graph G such that all eigenvalues $\lambda_1 \leq \cdots \leq \lambda_m$ of the normalized adjacency matrix satisfy $|\lambda_i| \leq \lambda$ for $1 \leq i \leq m-1$. One of the main motivations to study the sticky random walk is because of its perceived close relationship with the distributions generated by expander walks. In particular, if $(v_1, \ldots, v_n)$ is a n-step expander walk on a λ-expander $G = (V, E)$, and $W \subset V$ with $|W| = |V|/2$, we believe the distribution $(\mathbb{1}_{v_1 \in W}, \ldots, \mathbb{1}_{v_n \in W})$ is linked to the sticky walk $S(n, \lambda)$. In this section, we demonstrate one direction of this relationship by explicitly constructing a λ-expander and vertex subset W of half the size such that the distribution of $(\mathbb{1}_{v_1 \in W}, \ldots, \mathbb{1}_{v_n \in W})$ is precisely $S(n, \lambda)$.

Theorem 7.1. *There exists λ-expander $G = (V, E)$ and vertex set $W \subset V$ with $|W| = |V|/2$ such that if $(v_1, \ldots, v_n)$ is a random walk on G, the random n-bit string $(\mathbb{1}_{v_1 \in W}, \ldots, \mathbb{1}_{v_n \in W}) \sim S(n, \lambda)$.*

Proof. For simplicity, assume λ is rational, and take integer m such that $(\frac{1-\lambda}{1+\lambda})m$ is an integer. Let C_0 and C_1 be two m-cliques with an added self-loop at each vertex. Construct the graph G by taking each vertex in C_0 and connect it to $(\frac{1-\lambda}{1+\lambda})m$ vertices in C_1 in a cyclic uniform manner to make this graph $\frac{2m}{1+\lambda}$-regular (i.e. if we arbitrarily number the vertices in C_0 and C_1 from 1 to m, just connect vertex i in C_0 with the C_1 vertices $i, i+1, \ldots, i + (\frac{1-\lambda}{1+\lambda})m - 1 \pmod{m}$). Note upon setting $W = C_1$, a random walk on this expander resembles the sticky random walk, because at each step, m edges will keep us in the same clique, and $(\frac{1-\lambda}{1+\lambda})m$ will move us to the other, which gives us a $\frac{1+\lambda}{2}$ chance of staying in the same clique and $\frac{1-\lambda}{2}$ chance of moving to the other. Our aim is to now show the eigenvalues of the normalized Laplacian, $\overline{L}$, are within λ of 1.

To do so, we will first show all eigenvalues of $\overline{L}$ are $\leq 1 + \lambda$, and then demonstrate $\overline{L}$ has second smallest eigenvalue $1 - \lambda$. Note that the second smallest eigenvalue is $\min_{v \perp \mathbf{1}, ||v||_2 = 1} v^\intercal \overline{L} v$, and the largest eigenvalue is $w^\intercal \overline{L} w$, where w is the corresponding normalized eigenvector of this largest eigenvalue. Since $\mathbf{1} \perp w$, it suffices to show that for $v \in \mathbb{R}^{2m}$ with $||v||_2 = 1$ and $v \perp \mathbf{1}$, we have $1 - \lambda \leq v^\intercal \overline{L} v \leq 1 + \lambda$. WLOG assume the first m rows/columns are the vertices in C_0 and the latter m are the vertices in C_1. Let $t = \frac{1-\lambda}{1+\lambda} m$ and let $v = (a_1, \ldots, a_m, b_1, \ldots, b_m)$. Recall from the quadratic form version of the Laplacian and by construction of G that

$$(m+t) v^\intercal \overline{L} v = \sum_{1 \leq i < j \leq m} (a_i - a_j)^2 + \sum_{1 \leq i < j \leq m} (b_i - b_j)^2 + \sum_{\substack{1 \leq i \leq m \\ 0 \leq j < t}} (a_i - b_{i+j})^2$$

where indices are taken modulo m. Since $v \perp \mathbf{1}$, we have

[1] Unlike standard definitions, we do not restrict the degree of each vertex to be constant. The purpose of this section is to show any analysis of expander walks cannot give better bounds than the sticky walk. Since analyses on expander walks are based off of the spectral properties of the graph rather than its degree, restriction of the degrees are unnecessary for our purposes.

$$
\begin{aligned}
0 &= \left(\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} b_i\right)^2 \\
&= 2m\left(\sum_{i=1}^{m} a_i^2 + \sum_{i=1}^{m} b_i^2\right) - \sum_{1\le i<j\le m}(a_i - a_j)^2 - \sum_{1\le i<j\le m}(b_i - b_j)^2 - \sum_{\substack{1\le i\le m\\ 0\le j<m}}(a_i - b_{i+j})^2 \\
&= 2m||v||_2^2 - (m+t)v^{\intercal}\overline{L}v - \sum_{\substack{1\le i\le m\\ t\le j<m}}(a_i - b_{i+j})^2 \\
v^{\intercal}\overline{L}v &= \frac{2m}{m+t} - \frac{1}{m+t}\sum_{\substack{1\le i\le m\\ t\le j<m}}(a_i - b_{i+j})^2 \qquad (18)
\end{aligned}
$$

For one side, we can trivially upper bound (18)

$$v^{\intercal}\overline{L}v \le \frac{2m}{m+t} = 1+\lambda.$$

For the lower bound of (18), we can use Cauchy-Schwarz to get

$$
\begin{aligned}
v^{\intercal}\overline{L}v &\ge \frac{2m}{m+t} - \frac{2}{m+t}\sum_{\substack{1\le i\le m\\ t\le j<m}}(a_i^2 + b_{i+j}^2) \\
&= \frac{2m}{m+t} - \frac{2m-2t}{m+t}||v||_2^2 \\
&= \frac{2m}{m+t} \\
&= 1-\lambda.
\end{aligned}
$$

Hence we can conclude the nonzero eigenvalues of $\overline{L}$ are within λ of 1. Thus, G is indeed a λ-expander that models a $S(n,\lambda)$ sticky walk. □

Note that this construction gives a family of λ-expanders: one for each m where $\left(\frac{1-\lambda}{1+\lambda}\right) m$ is an integer. In order to extend the above theorem to degree-bounded expanders, we believe replacing the cliques C_0 and C_1 with degree d expanders, and adding a random bi-regular bipartite graph with degree $\left(\frac{1-\lambda}{1+\lambda}\right) d$ between them will suffice, but we have not verified this.

Acknowledgments

V.G thanks Salil Vadhan, Vijay Bhattiprolu, Nicolas Resch, and Mahdi Cheraghchi for valuable discussions about sticky random walks at various points. Thanks to Salil also for suggesting the event underlying Theorem 4.8. V.K thanks Mayank Pandey for helpful discussions and pointing him to [4].

A Deferred Proofs

Claim 4.4. For $1 \le k \le n/2$, we can bound $\frac{\binom{n}{k}^2}{\binom{n}{2k}} \le \frac{e^3\sqrt{2}}{4\pi^2} \cdot 4^{2k}$.

Proof. We can use Stirling's approximation $\sqrt{2\pi}\left(\frac{n}{e}\right)^n \sqrt{n} \le n! \le e\left(\frac{n}{e}\right)^n \sqrt{n}$ to bound

$$\begin{aligned}
\frac{\binom{n}{k}^2}{\binom{n}{2k}} &= \frac{n!(2k)!(n-2k)!}{k!^2(n-k)!^2} \\
&\le \frac{e^3}{4\pi^2} \cdot \frac{n^n(2k)^{2k}(n-2k)^{n-2k}}{k^{2k}(n-k)^{2n-2k}} \sqrt{\frac{2n(n-2k)}{k(n-k)^2}} \\
&\le \frac{e^3}{4\pi^2} \cdot 2^{2k}\left(\frac{n}{n-k}\right)^n \left(\frac{n-2k}{n-k}\right)^{n-2k} \sqrt{\frac{2}{k}} \\
&\le \frac{e^3\sqrt{2}}{4\pi^2}\left(\frac{2n}{n-k}\right)^{2k} \\
&\le \frac{e^3\sqrt{2}}{4\pi^2} \cdot 4^{2k}
\end{aligned}$$

□

Lemma 4.7. As $n \to \infty$, $\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}} \to \boldsymbol{N}$ in distribution.

Proof. The idea is to use the method of moments (see Theorem 8.6 of [4]), which states the lemma can be deduced if all moments of Y approach the moments of $\boldsymbol{N}$ as $n \to \infty$. The odd moments of the Gaussian distribution is zero, and the $2k$'th moment is $(2k-1)!! := \frac{(2k)!}{2^k k!}$ for all $k \ge 1$ (see [6]). Clearly $\mathbb{E}\left[\left(\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}}\right)^{2k+1}\right] = 0$ by the symmetry of $\boldsymbol{Y}$.

It now remains to show that $\mathbb{E}\left[\left(\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}}\right)^{2k}\right] \to (2k-1)!!$ as $n \to \infty$. Let P_k denote the set of unordered partitions of k into positive parts, where we express each partition as a multiset of positive integers $A = [a_1, \dots, a_m]$ where $\sum_{i=1}^m a_i = k$. We define $g([a_1, \dots, a_m]) = \frac{(\sum_{i=1}^m a_i)!}{\prod_{i=1}^m a_i!}$. By expanding $(\sum_{i=1}^n \boldsymbol{Y}_i)^{2k}$, collecting terms with the same multiset of degrees, and taking exponents modulo 2,

$$\mathbb{E}[\boldsymbol{Y}^{2k}] = \mathbb{E}\left[\left(\sum_{i=1}^n \boldsymbol{Y}_i\right)^{2k}\right] = \sum_{P \in P_{2k}} g(P)\mathbb{E}\left[\sum_{T \in \binom{[n]}{|P|}} \prod_{i=1}^{|P|} \boldsymbol{Y}_{t_i}^{p_i}\right] = \sum_{P \in P_{2k}} g(P)\mathbb{E}\left[\sum_{T \in \binom{[n]}{|P|}} \prod_{i; 2 \nmid p_i} \boldsymbol{Y}_{t_i}\right] \quad (19)$$

where $P = [p_1, \dots, p_{|P|}]$ and $T = \{t_1, \dots, t_{|P|}\}$. Now let P_o and P_e be the multiset containing all odd and even elements of P (with multiplicity), respectively. Define $h([a_1, \dots, a_m])$ to be the number of ways to permute $(a_1, \dots, a_n)$ (e.g. $g([1, 2, 2, 6]) = 12$ since there are 12 ways of permuting $(1, 2, 2, 6)$, $g([2, 2, 2]) = 1$, and $g([1, 4, 5, 6]) = 24$). By Equation 5 and Lemma 4.2, we have

$$\mathbb{E}\left[\sum_{T \in \binom{[n]}{|P_o|}} \prod_{i \in T} \boldsymbol{Y}_i\right] = \mathbb{E}[K_{|P_o|}(|\boldsymbol{s}|)] = \sum_{m=|P_o|/2}^{n-|P_o|/2} \binom{m-1}{|P_o|/2-1}\binom{n-m}{|P_o|/2}\lambda^m.$$

With this, we can rewrite

$$\begin{aligned}\mathbb{E}\left[\sum_{T\in\binom{[n]}{|P|}}\prod_{i;2\nmid p_i}\boldsymbol{Y}_{t_i}\right] &= h(P_e)\binom{n-|P_o|}{|P_e|}\mathbb{E}\left[\sum_{T\in\binom{[n]}{|P_o|}}\prod_{i\in T}\boldsymbol{Y}_i\right] \\ &= h(P_e)\binom{n-|P_o|}{|P_e|}\sum_{m=|P_o|/2}^{n-|P_o|/2}\binom{m-1}{|P_o|/2-1}\binom{n-m}{|P_o|/2}\lambda^m \\ &\sim h(P_e)\frac{n^{|P_e|}}{(|P_e|)!}\sum_{m=|P_o|/2}^{n-|P_o|/2}\binom{m-1}{|P_o|/2-1}\frac{n^{|P_o|/2}}{(|P_o|/2)!}\lambda^m \\ &= n^{|P|-|P_o|/2}\cdot\frac{h(P_e)}{(|P_o|/2)!(|P_e|)!}\sum_{m=|P_o|/2}^{n-|P_o|/2}\binom{m-1}{|P_o|/2-1}\lambda^m \end{aligned} \tag{20}$$

and so by combining (19) and (20), we have

$$\mathbb{E}[Y^{2k}] \sim \sum_{P\in P_{2k}} n^{|P|-|P_o|/2}\cdot\frac{g(P)h(P_e)}{(|P_o|/2)!(|P_e|)!}\sum_{m=|P_o|/2}^{n-|P_o|/2}\binom{m-1}{|P_o|/2-1}\lambda^m. \tag{21}$$

We just have to look at the leading term of this sum, which is when $|P|-|P_o|/2$ is maximized (since $|P_{2k}|$, $g(P)$, and $h(P_e)$ don't depend on n). Note

$$|P|-\frac{|P_o|}{2} = (|P_o|+|P_e|)-\frac{|P_o|}{2} = \frac{|P_o|+2|P_e|}{2} \le \frac{\sum_{p\in P_o}p+\sum_{p\in P_e}p}{2} = k.$$

Hence the leading terms are when $|P|-|P_o|/2=k$, and these terms correspond to when all elements of P are 1 or 2. In particular, the leading terms correspond to the multisets P with $2r$ 1's and $k-r$ 2's for $0\le r\le k$. In these cases, we can calculate $|P|=k+r$, $|P_o|=2r, h(P_e)=1$ and $g(P)=\frac{(2k)!}{2^{k-r}}$ Hence from (21) we get

$$\begin{aligned}\mathbb{E}[Y^{2k}] &\sim \sum_{r=0}^{k}\frac{(2k)!}{2^{k-r}r!(k-r)!}\left(\sum_{m=r}^{n-r}\binom{m-1}{r-1}\lambda^m\right)n^k \\ &= (2k-1)!!\sum_{r=0}^{k}\binom{k}{r}2^r\left(\sum_{m=r}^{n-r}\binom{m-1}{r-1}\lambda^m\right)n^k.\end{aligned}$$

Consequently, we have the $2k$'th moment of $\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}}$ is

$$\sim (2k-1)!!\left(\frac{1-\lambda}{1+\lambda}\right)^k\sum_{r=0}^{k}\binom{k}{r}2^r\left(\sum_{m=r}^{n-r}\binom{m-1}{r-1}\lambda^m\right)$$

Taking $n\to\infty$ and evaluating the series with the well-known generating function identity $(\frac{x}{1-x})^r =$

$\sum_{m\geq r}\binom{m-1}{r-1}x^m$, we get that the $2k$'th moment approaches

$$
\begin{aligned}
(2k-1)!!\left(\frac{1-\lambda}{1+\lambda}\right)^k\sum_{r=0}^{k}\binom{k}{r}2^r\left(\sum_{m=r}^{\infty}\binom{m-1}{r-1}\lambda^m\right) &= (2k-1)!!\left(\frac{1-\lambda}{1+\lambda}\right)^k\sum_{r=0}^{k}\binom{k}{r}\left(\frac{2\lambda}{1-\lambda}\right)^r \\
&= (2k-1)!!\left(\frac{1-\lambda}{1+\lambda}\right)^k\left(1+\frac{2\lambda}{1-\lambda}\right)^k \\
&= (2k-1)!!.
\end{aligned}
$$

Hence by the method of moments $\boldsymbol{Y}\sqrt{\frac{1-\lambda}{(1+\lambda)n}} \to \boldsymbol{N}$ in distribution as $n \to \infty$. □

References

[1] Louay Bazzi. Entropy of weight distributions of small-bias spaces and pseudobinomiality. *Chic. J. Theor. Comput. Sci.*, 2015, 2015.

[2] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 2006.

[3] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998.

[4] A. Gut. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer New York, 2006.

[5] Nabil Kahale. Large deviation bounds for Markov chains. *Comb. Probab. Comput.*, 6(4):465–474, 1997.

[6] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, Boston, fourth edition, 2002.

[7] Anup Rao. An exposition of Bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.

[8] Shravas Rao and Oded Regev. A sharp tail bound for the expander random sampler. *CoRR*, abs/1703.10205, 2017.

[9] Alex Samorodnitsky. On the optimum of Delsarte's linear program. *J. Comb. Theory, Ser. A*, 96(2):261–287, 2001.

[10] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 238–251. ACM, 2017.

[11] Salil Vadhan. Personal communication. 2018.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il