Pseudorandomness of the Sticky Random Walk

Emile Timothy Anand

In Partial Fulfillment of the Requirements for the
Degree of
Bachelor of Science

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Submitted June 11, 2023

Emile Timothy Anand
ORCID: 0000-0003-2893-9469

ACKNOWLEDGEMENTS

I thank my advisor Professor Chris Umans for his invaluable advice and guidance on my academics and research. This thesis project would not have been possible without Professor Umans' support over the last three years.

I would also like to thank Professors Adam Wierman, Leonard Schulman, David Conlon, and Katy Bouman for their help and mentorship during my time at Caltech.

I acknowledge the countless individuals who have contributed to the development of the field of pseudorandomness and expander random walks. Their pioneering work and scholarly contributions have paved the way for this research, and I am humbled to be a part of this academic community.

Finally, this thesis is dedicated to my family. I am incredibly grateful to you, Mom, Dad, and Erin, for your support, love, and encouragement.

ABSTRACT

We extend the pseudorandomness of random walks on expander graphs using the sticky random walk. Building on the works of [Coh+22] and [GK21], [GV22] recently showed that expander random walks can fool all symmetric functions in total variation distance (TVD) upto an $O(\lambda(\frac{p}{\min f})^{O(p)})$ error in total variation distance, where λ is the second largest eigenvalue of the expander, p is the size of the arbitrary alphabet used to label the vertices, and $\min f = \min_{b\in[p]} f_b$, where f_b is the fraction of vertices labeled b in the graph. [GV22] conjectures that the dependency on the $(\frac{p}{\min f})^{O(p)}$ term is not tight.

In this paper, we resolve the conjecture in the affirmative for a family of expanders. We present a generalization of [GK21]'s sticky random walk for which [GV22] predicts a TVD upper bound of $O(\lambda p^{O(p)})$ using a Fourier-analytic approach. For this family of graphs, we use a combinatorial approach involving the Krawtchouk functions used in [GK21] to derive a strengthened TVD of $O(\lambda)$. Furthermore, we present equivalencies between the generalized sticky random walk, and, using linear-algebraic techniques, show that the generalized sticky random walk is an infinite family of expander graphs.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Preliminaries: Notation and Convention

This section describes the basic notation that is used throughout this report.

For any $n \in \mathbb{N}$, let $[n] = \{1, ..., n\}$ and $\mathbb{Z}_n = \{0, ..., n-1\}$. Let $[n]^k$ denote k copies of elements in $[n]$, and let $\mathbb{Z}_n^k$ denote k copies of elements in $\mathbb{Z}_n$. Furthermore, let $\binom{[n]}{k}$ denote the set of all k-sized subsets of $[n]$, which has cardinality $\binom{n}{k}$. For any n, let the Hamming cube of dimension n be a graph $G = (V, E)$ with vertex set $V = \{0,1\}^n$ and edge-set $E = \{(u, v) : u \text{ and } v \text{ differ exactly in one coordinate}\}$.

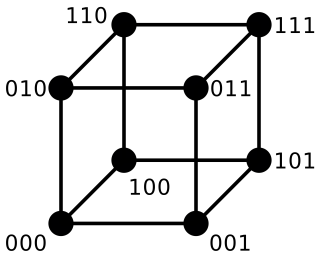

Figure 1: Hamming cube of dimension 3

For any n-bit strings x_1 and x_2, let $d_{\text{Hamming}}(x_1, x_2)$ denote the Hamming distance between x_1 and x_2, which is the minimal number of edges needed to reach x_2 from x_1 on the Hamming cube of dimension n. For any n-bit string s, let $|s|$ denote the Hamming weight (the Hamming distance between s and 0^n) of s. Intuitively, the Hamming weight of a a bit-string counts its number of 1's. Similarly, let $|s|_i$ denote the number of i's in s. We generalize the notion of counting the number of occurrences of any character $\chi \in \mathbb{Z}_p$ for $p \geq 2$ with the symmetric function $\Sigma(x) : \mathbb{Z}_p^n \to \mathbb{Z}_n^p$, where $\Sigma(x)$ is a vector that counts the number of occurrences of each $\chi \in \mathbb{Z}_p$. Specifically, for all $\chi \in \mathbb{Z}_p$ and for all $x \in \mathbb{Z}_p^n$, we write that $[\Sigma(x)]_\chi = |\{i \in x : x_i = \chi\}| = |x|_\chi$.

Let $\text{Ber}(q)$ denote the Bernoulli distribution on $\{0,1\}$, such that if $X \sim \text{Ber}(q)$, then $\Pr[X = 1] = q$ and $\Pr[X = 0] = 1 - q$. Next, let $\text{Bin}(n, 1/2)$ denote the binomial distribution of $\sum_{i=1}^n b_i$ with independent choices of $b_i \sim \text{Ber}(1/2)$. Let $\text{U}_p^n = \text{U}[\{0, \ldots, p-1\}]^n$ denote n samples of the uniform distribution on $\mathbb{Z}_p$, where

for each sample, each character is sampled with probability $1/p$. Then, $[\Sigma(\mathrm{U}_p^n)]_0$ reports the number of 0's in an n-bit sample from the uniform distribution on $\mathbb{Z}_p$.

Furthermore, we write that $x \in A$ if x is an element of A, and $x \in_U A$ if x is an element chosen uniformly randomly from A. Finally, we use $\overline{C}$ to denote the complement of a set $C \subseteq \Omega$, and for any two sets $A, B \subseteq \Omega$, we define their symmetric difference $A \Delta B$ as $(A \cap \overline{B}) \cup (B \cap \overline{A})$.

1.2 Pseudorandomness and derandomizing BPP

Is randomness a special computational resource? Randomness can be viewed as a resource for algorithms to employ as part of its procedure. Some well-known randomized algorithms include the Miller-Rabin protocol [Mil75][Rab80] for primality testing and Karger's algorithm [KS96] for finding the min-cut of a graph. The set of all decision problems that are solvable by randomized algorithms in polynomial time is captured in the large complexity class BPP.

Definition 1 (BPP)**.** *A language* L *is said to be in* BPP *(bounded-error probabilistic polynomial time) if and only if there exists a probabilistic polynomial-time (PPT) Turing machine* A *for which,* $x \in \mathrm{L}$ *implies that* $\Pr[\mathrm{A}(x) = 1] \geq 2/3$ *and* $x \notin \mathrm{L}$ *implies that* $\Pr[\mathrm{A}(x) = 1] < 1/3$*. In words,* $\mathrm{L} \in$ BPP *implies the existence of a PPT Turing machine that can correctly decide the membership of strings in the language with probability atleast* $2/3$*.*

BPP is a robust complexity class because the choice of error probability to be atmost $1/3$ is arbitrary. For an input $x \in \{0,1\}^n$, any error probability in the range of $[0, \frac{1}{2} - \frac{1}{\mathrm{poly(n)}})$ yields the same set of decision problems, since running the probabilistic algorithm polynomially many times and taking the majority result causes the error to decrease to 0 exponentially by the Chernoff bound [Vad13][Lez01].

Definition 2 (P)**.** *A language* L *is said to be in* P *if and only if there exists a deterministic polynomial-time Turing machine* A *for which* $x \in \mathrm{L}$ *implies that* $\mathrm{A}(x)$ *accepts and* $x \notin \mathrm{L}$ *implies that* $\mathrm{A}(x)$ *rejects.*

Clearly, all problems contained in P are also in BPP. However, it is not currently known whether there is a definite random advantage for all polynomial-time problems since the number of problems that have been thought to be in BPP *but not* P has decreased over the years. Can any problem in BPP be derandomized so that it is

solvable by a deterministic polynomial time algorithm? It is conjectured that BPP = P, and resolving it is a major **open problem** in theoretical computer science.

To digress, it is worthwhile to note the important milestone in 2004 by [AKS04] of derandomizing the language of prime numbers which showed PRIMES $\in$ P.

Derandomization is the process of transforming a randomized algorithm into a deterministic algorithm. Oftentimes, this happens by creating a vastly *different* algorithm. However, promising methods aim to replace a true source of random bits with an efficient deterministic source that amplifies a limited initial amount of randomness. Such sources are said to be pseudorandom sources if they can 'fool' tests computed by various families of functions.

Definition 3 (Fooling Tests)**.** *A test-function* T *is ϵ-fooled by a pseudorandom function $g : X \to [n]$ if the statistical distance between the distributions* $\mathrm{T}(g(X))$ *and $f(U)$, where U is the uniform distribution on $[n]$, is less than ϵ.*

Promising derandomization techniques include using pseudorandom generators (PRGs) and expander graphs. This thesis is focused on the pseudorandomness of random walks on expander graphs.

1.3 Expander Graphs

We first provide the necessary linear algebraic terminology for studying expander graphs. We consider a graph $G = (V, E)$, where $|V| = n$ and $|E| = m$. Then:

Definition 4 (Adjacency Matrix)**.** *The adjacency matrix A of G is an $n \times n$ matrix such that $A_{ij} = 1$ if $(i, j) \in E$, and 0 otherwise.*

Definition 5 (Degree Matrix)**.** *The degree of $v \in V$ (denoted d_v) is the number of nodes in G connected to v. The degree matrix D of G is an $n \times n$ matrix such that $D_{ij} = 1/d_i$ if $i = j$ and 0 otherwise.*

Definition 6 (Normalized Adjacency Matrix)**.** *The normalized adjacency matrix of G is the matrix $\mathcal{A} := D^{-1/2} A D^{-1/2}$, where A is the adjacency matrix from definition 4 and D is the degree matrix from definition 5.*

As a remark, since $\mathcal{A}$ is a real symmetric matrix, the spectral theorem implies that $\mathcal{A}$ has n real eigenvalues and n real eigenvectors.

Lemma 1.3.1. *Without loss of generality, let $\lambda_1 \geq \cdots \geq \lambda_n$ be the n eigenvalues of $\mathcal{A}$, where $\mathcal{A}$ is the normalized adjacency matrix (from definition 6) of a graph $G = (V, E)$. Then, $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$. We provide a proof of lemma 1.3.1 in A. We define $1 - \lambda_2$ as the spectral gap of graph G.*

Armed with the knowledge of the characteristics of eigenvalues of $\mathcal{A}$, we now study the (multiple) definitions of expander graphs. Recall that a graph G is connected if and only if $\lambda_2 < 1$. Since $\lambda_2 < 1$ enforces that G is connected, smaller values of λ_2 (or larger spectral gaps) correspond to stronger notions of connectivity.

Informally, an expander graph is a sparse graph with good connectivity properties. A graph is said to be *sparse* if it does not contain many edges, and is said to have good *connectivity properties* if the only way to disconnect a component from the graph is to remove many edges. These (seemingly contradictory) properties give rise to (indeed, explicit) constructions of expanders and the following definitions of expansion of a graph G.

Definition 7 (Sparsest Cut)**.** *Let $(S, V - S)$ be a partition of the vertex set V, and let $E(S, V - S) = \{(u, v) \in E : u \in S, v \in V - S\}$. Then, the sparsity of the cut is:*

$$\phi(S) := \frac{E(S, V - S)}{|E|} \cdot \left(\frac{|S| \cdot |V - S|}{|V|^2/2} \right)^{-1}$$

The sparsest cut of a graph G is then:

$$\phi(G) := \min_{S \subseteq V : S \neq \emptyset, S \neq V} \phi(S)$$

Intuitively, $\phi(G)$ is the minimal ratio between the fraction of edges that need to be removed from E to disconnect S and $V - S$ and the fraction of pairs of vertices that would be disconnected with such an edge removal.

The sparsest cut measure in definition 7 of a graph $G = (V, E)$ is very closely related to the measure of edge expansion provided below in definition 8.

Definition 8 (Edge Expansion)**.** *If G is a d-regular graph (each vertex has d neighbors), then the edge expansion of a cut $(S, V - S)$ is given by:*

$$h(S) := \frac{E(S, V - S)}{d \cdot \min\{|S|, |V - S|\}}$$

The edge expansion $h(G)$ of G is then:

$$h(G) := \min_{S \subseteq V : S \neq \emptyset, S \neq V} h(S)$$

Intuitively, $h(G)$ is the smallest ratio between the number of edges between two cuts of G and the number of edges incident on the smaller cut.

Definition 9 (Vertex Expansion)**.** *A graph G is said to be a (K, A)* vertex expander *if $\forall S \subseteq V$ where $|S| \leq K$, such that $|\{(u, v) \in E : u \in S\}| \geq A \cdot |S|$, as defined in [Vad13].*

Note that the notions of edge and vertex expansion in 8 and 9 are closely related.

The connectedness of a graph can also be measured by how well a random walk on the graph converges to the stationary distribution, as mentioned in [Vad13]. This is typically characterized by the second largest eigenvalue of the normalized adjacency matrix $\mathcal{A}$ provided in definition 6.

Definition 10 (Spectral Expansion)**.** *For $u = (1/n, \ldots, 1/n) \in \mathbb{R}^n$, the uniform distribution on $[n]$, and probability distributions $\pi \in [0, 1]^n$, the spectral expansion of a graph $G = (V, E)$ is $\lambda(G)$, which is characterized by:*

$$\lambda(G) := \lambda_2(\mathcal{A}(G)) := \min_{\pi} \frac{\|\pi M - u\|}{\|\pi - u\|} = \min_{x \perp u} \frac{\|xM\|}{\|x\|}$$

Informally, a lower $\lambda(G)$ indicates a stronger connectivity property of G.

These notions of expansion are agree are very strongly related. For instance, we direct the reader to Theorems 4.6 and 4.9 in [Vad13] which show an equivalence between vertex expansion and spectral expansion. We state them here (without proof) for convenience.

Theorem 4.6 of [Vad13]: (spectral expansion $\implies$ vertex expansion). If G is a regular graph with spectral expansion λ, for $\lambda \in [0, 1]$, then $\forall \alpha \in [0, 1]$, G is an $(n/2, 2 - \lambda)$ vertex expander.

Theorem 4.9 of [Vad13]: (vertex expansion $\implies$ spectral expansion). $\forall \delta > 0, D > 0$, if G is a D-regular $(n/2, 1 + \delta)$ vertex expander, then $\exists \lambda > 0$ such that G has spectral expansion λ. Here, we can take $\lambda = 1 - \Omega((\delta/D)^2)$.

Further, the celebrated Cheeger inequalities provide a relationship between edge expansion and spectral expansion. As before, we state the Cheeger inequality and direct the reader to the excellent expanders survey by Hoory, Linial, and Wigderson, [HLW06], for the proof of the theorem.

Definition 11 (Cheeger's Inequality)**.** *Let $G = (V, E)$ be an undirected d-regular graph and let $\lambda = \lambda(G)$ be the spectral expansion of G, given by the second*

smallest eigenvalue of its normalized adjacency matrix $\mathcal{A}(G)$. Further, let the edge expansion of the graph be denoted $h(G)$. Then, Cheeger's inequality states that:

$$\frac{1}{2}\lambda d \leq h(G) \leq d\sqrt{2\lambda}$$

Corollary 1.3.1.1. *As a consequence of Cheeger's inequality, we also see the relation between the sparsest cut measure and the edge expansion measure, given by: $\phi(G) \leq h(G) \leq 2\phi(G)$, where $G = (V, E)$ is a d-regular graph.*

Given these (largely equivalent) definitions of expander graphs, we observe that graphs with good expansion properties also have good connectivity properties. For instance, note the following lemma proven in [Tre11].

Lemma 1.3.2. *Let $G = (V, E)$ be a d-regular graph with edge expansion h. If any $\epsilon < h$ fraction of edges are removed from G, then the graph has a connected component that spans atleast $1 - \epsilon/2h$ fraction of the vertices. We include the proof of this lemma for convenience in A.*

1.3.1. Good Expander Graphs

We now turn to the question of what is a good expander graph. A family of constant degree expanders is a family of graphs $\{G_n\}_{n \geq d}$ where each G_n is a d-regular graph on n vertices, and there is an absolute constant $\lambda(G) \leq \lambda$ for each n. We next look at the following theorem by Alon and Boppana in [Fri03].

Lemma 1.3.3. *For every constant $d \in \mathbb{N}$, any d-regular graph $G = (V, E)$ satisfies $\lambda(G) \geq 2\sqrt{d-1}/d - o(1)$, where the $o(1)$ term vanishes as $n \to \infty$. We provide a proof of the Alon-Boppana bound in [Fri03] for convenience in A.*

It is fascinating to note that there exist several explicit constructions of d-regular graphs $G = (V, E)$ for which $\lambda(G) < 2\sqrt{d-1}/d$ (see [Lub17][BT11][SS96][RVW04] of expander graphs). Such graphs which saturate the Alon-Boppana bound in lemma 1.3.3 are called *Ramanujan* graphs. We provide some explicit examples of interesting expander graph constructions (most of which are number-theoretic), and direct the reader to sources for proofs of their expansion properties.

Construction 1: (Discrete Torus Expanders). The first known explicit construction of expanders is from Margulis' construction in [Mar75], where $G = (V, E)$

with vertex set $V = \mathbb{Z}_m^2$ and a vertex $(x, y) \in V$ is adjacent to the following vertices $(x \pm y, y), (x \pm (y+1), y), (x, y \pm (x+1)), (x, y \pm (x+1))$. Here, all operations are done modulo m. Further, $\{G_n\}_n$ is an expander family of constant degree 8 with expansion $\Omega(1)$.

Construction 2: (p-cycle with inverse chords). The next explicit construction of an expander we present from [GG81] is the graph $G = (V, E)$ with vertex set $V = \mathbb{Z}_p$ (for p prime), where the edges connect each node x with nodes $x+1, x-1$, and x^{-1}. As before, the arithmetic is mod p, and 0^{-1} is set to 0.

Construction 3: (Ramanujan graphs). We now provide an explicit construction of a Ramanujan graph as described in [LPS88]. Let $G = (V, E)$ be a graph where $V = \mathbb{F}_q \cup \{\infty\}$. Here, $\mathbb{F}_q$ is the finite field of prime order q such that $q \equiv 1 \pmod 4$ (and one extra node representing ∞). The edges in this graph connect each node $z \in V$ with all $z' \in V$ of form:

$$z' = \frac{(a_0 + ia_i)z + (a_2 + ia_3)}{(-a_2 + ia_3)z + (a_0 - ia_1)}$$

Here, $a_0, a_1, a_2, a_3 \in \mathbb{Z}$ such that $a_0^2 + a_1^2 + a_2^2 + a_3^2 = p, a_0$ is odd and positive, and a_1, a_2, a_3 are even, for some fixed prime $p \neq q$ such that $p \equiv 1 \pmod 4, q$ is a square modulo p and $i \in \mathbb{F}_q$ such that $i^2 = -1 \pmod q$. Needless to say, this construction is an optimal spectral expander as it saturates the Alon-Boppana bound in lemma 1.3.3 for strong expander graphs.

1.3.2. Properties of Expanders

Expanders have many unique properties (see [Rei05][RVW04][BT11]). A key property of expander graphs is the expander mixing lemma. In a random graph with constant degree d, the number of edges between any two sets S and T is approximately $\frac{d}{n}|S||T|$. Intuitively, expanders with a spectral expansion constant of $\lambda(G)$ closer to 0 display properties of random graphs, and mimic this property as well. We show this below.

Lemma 1.3.4 (Expander Mixing Lemma)**.** *Let $G = (V, E)$ with $|V| = n, |E| = m$ be a d-regular graph with spectral expansion constant λ. Then, for any subsets S, T of the vertex-set V, we have that:*

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \lambda\sqrt{|S||T|}$$

We defer the proof of this lemma to A.

Another property of expander graphs is that random walks on expanders mix quickly, which means that they converge very quickly to the uniform distribution. To formalize this notion, observe that the normalized adjacency matrix $D^{-1/2}AD^{-1/2}$ of a graph G is a homogeneous stochastic matrix (each row sums to 1 and the coefficients are positive). Therefore, we can define a random walk on an expander graph by selecting $X_0 \in_U V$ and letting $X_k \sim N(X_{k-1})$, where $\Pr[X_k = v] = (D^{-1/2}AD^{-1/2})_{X_k,v}$. For convenience, we denote $M := D^{-1/2}AD^{-1/2}$ the random-walk matrix of G.

Definition 12 (Irreducible Markov Chain)**.** *A homogeneous matrix is irreducible if and only if* $\exists n$ *such that* $\forall i, j \in V : \Pr[X_n = i | X_0 = j] > 0$.

Definition 13 (Invariant Probability Distribution)**.** *Let* $\pi \in \mathbb{R}^n$ *where* $\pi_i \geq 0$ *and* $\sum_{v \in V} \pi_v = 1$. *Then, a probability distribution on* V *is called an invariant or stationary probability distribution if* $\forall v \in V$*:*

$$\pi_v = \sum_{u \in V} \pi_u M_{u,v} \iff \pi = \pi M$$

Note that our homogeneous stochastic matrix M is irreducible and, thus, has an invariant probability distribution π. Since random walks on expanders mix very well, it must be the case that π is very close to the uniform distribution on $[n]$ (and it is, and we shall prove it shortly). For now, though, note that an exciting implication of this is that expander walks have good randomness properties not just for the final vertex in the sequence, but for the sequence of vertices itself, which displays characteristics of uniform independent samples of V.

Lemma 1.3.5 (Expander Hitting Lemma)**.** *Let* $G = (V, E)$ *with* $|V| = n, |E| = m$ *be a* d*-regular graph with spectral expansion constant* λ*. Then, for any* $B \subseteq V$ *such that* $|B| = (1 - \delta)n$*, the probability that a random walk* $X_1, X_2, \ldots, X_t$ *of* $t - 1$ *steps starting at a uniformly random vertex of* G *completely stays inside* B *is given by:*

$$\Pr[X_i \in B, \forall i \in [t]] \leq (1 - \delta(1 - \lambda))^{t-1}$$

We defer the proof of this lemma to A.

So, expander graphs are undirected spectral sparsifiers of the clique with high expansion properties, and they are among the most useful combinatorial objects in

theoretical computer science due to their numerous applications (some of which we shall present in the subsequent subsection). For a good expander graph that saturates the Alon-Boppana bound from [Fri03], the spectrum $(\lambda_1, \ldots, \lambda_n)$ approximates the spectrum of the complete graph, which makes expanders a sparsification of the clique. As described before, random walks on expanders mix very quickly. We state Gilman's [Gil98] result which uses the Chernoff bound to quantify the mixing rate.

Lemma 1.3.6 (Expander-Walk Chernoff Bound (Gilman))**.** *For graph $G = (V, E)$, let $X_0, \ldots, X_{t-1}$ denote a sequence of vertices obtained from a t-step -walk on an expander graph G. For any function $f : [n] \to \{0, 1\}$, let the stationary distribution of G be $\pi(f) := \lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} f(X_i)$. The expander-walk Chernoff bound [Gil98] states that $\forall \varepsilon > 0$,*

$$\Pr\left[\left|\frac{1}{t}\sum_{i=0}^{t-1} f(X_i) - \pi(f)\right| \geq \lambda + \varepsilon\right] \leq 2e^{-\Omega(\varepsilon^2 t)}$$

Equivalently, we can write that $\forall \varepsilon > 0$:

$$\Pr\left[\left|\frac{1}{t}\sum_{i=0}^{t-1} f(X_i) - \pi(f)\right| \geq \varepsilon\right] \leq 2e^{-\Omega((\lambda-\varepsilon)^2 t)}$$

We refer the reader to [HLW06][Kom+02][RR17] for a proof of the expander-walk Chernoff bound. An important direct consequence of the expander Chernoff bound is that the mixing time of a d-regular expander graph on n vertices is atmost $O(\log n)$.

The expander hitting lemma and the expander mixing lemma imply a number of combinatorial consequences for any d-regular graph $G = (V, E)$ where $|V| = n, |E| = m$. We state some of them below and refer the reader to [Vad13] for proofs.

Corollary 1.3.6.1. *The chromatic number $\chi(G)$ is atleast $(2 - \lambda)/(1 - \lambda)$.*

Corollary 1.3.6.2. *The diameter of G is $O(\log_{1/\lambda} n)$.*

Corollary 1.3.6.3. *The size of the largest independent set $\alpha(G)$ is atmost $\frac{n(1-\lambda)}{2-\lambda}$.*

1.3.3. Applications of Expanders

Expander graphs have intriguingly ubiquitous applications. They were studied for the purpose of constructing fault-tolerant networks in [INW94], where if a small

number of channels (edges) broke down, the system could be made to be still largely intact due to its good connectivity properties if it were modeled as an expander. More recently, they have been used in representation learning theoretic settings (see [DLV22]) to create graph neural networks that can propagate information to train models more efficiently. In coding theory, expander codes (created from linear bipartite expanders - see [Alo86]) are the only known construction (see [SS96]) of asymptotically good error-correcting codes which can be decoded in linear time when a constant fraction of symbols are in error.

More recent works that combine ideas from combinatorial topology and algebraic geometry have also led to the exciting study of high dimensional expanders (HDX) which are pure simplicial complexes (hypergraphs that are downwards closed under containment) where the 1-skeletons are spectral expanders and the links exhibit good expansion properties. We direct the reader to [Con19] and [GK23].

One of the most important applications of expanders (which is the topic of this thesis) is on derandomization and in pseudorandomness. Suppose that there is a randomized algorithm for a language L using n bits such that: If a string $x \in L$, then the algorithm accepts with probability 1. If a string $x \notin L$, then the algorithm rejects with probability atleast $1/2$. Our goal is to reduce the error probability of the algorithm.

If we repeat the algorithm t times then the error probability goes down to $1/2^t$, which is ideal. However, the number of random bits used by the algorithm is then equal to nt, which is very large. One work around is to "reuse the randomness by weakening our independent choices to correlated choices on an expander graph" [Gur20]. If we start at a random vertex in G (which is a random number in $\{0, \ldots, n\}$ which uses $\log n$ random bits) and pick random neighbors of v, then since a good expander has degree $d = O(1)$ and since we need $\log d = O(1)$ bits, we can continue this process till we pick t vertices overall, and the overall number of random bits that we would need would be equal to $\log n + O(t)$. Further, by the expander mixing lemma, for $t \gg O(\log n)$ the sequence of vertices will still be extremely close to uniformly random. Comparing with previous methods for error reduction, we present the table from [Vad13]:

Furthermore, the number of random bits required to take t independent samples from a function g is $O(t \log n)$, but sampling via expander random walks only necessitates $\log n + O(t)$ many random bits. So, a random walk on an expander graph provides a derandomized approximation for a random walk on a complete graph with self-

	Number of Repetitions	Number of Random Bits
Independent Repetitions	$O(t)$	$O(tn)$
Pairwise Independent Repetitions	$O(2^t)$	$O(t+n)$
Expander Walks	$O(t)$	$\log n + O(t)$

Table 1.1: Comparing with previous methods for error reduction

loops. This makes expander graphs invaluable in the field of pseudorandomness. Consider the t-step expander random walk which generates a sequence of vertices $v_0, \ldots, v_{t-1}$. We are then interested in the degree to which $(v_0, \ldots, v_{t-1})$ "fools" classes of test functions, where the definition of fooling is consistent with what is provided in definition 3.

A major goal of this work is to obtain tight error bounds in the approximation of true random bits with bits supplies from an expander random walk. From definition 3, an expander random walk fools a test-function f if $f(X_0, \ldots, X_{t-1})$ has approximately the same distribution regardless of whether the vertices are sampled from a random walk on an expander, or independently and uniformly at random (which is the same as a random walk on a complete graph with self-loops). The goal of this thesis seeks to find strong bounds on the extent to which expander random walks fool various classes of test functions f.

1.4 Random Walks on Expander Graphs

A recent line of work started in [Ta-17] which led to [GK21][CPT21][Coh+22][BCG20] [GV22] has shown that random walks on expanders with second largest eigenvalue λ fool various functions upto an $O(\lambda)$ error in total variational distance. For the rest of the introduction, we provide an overview of [GK21] and [GV22].

If an arbitrary half of vertices of an expander-graph $G = (V, E)$ with $|V| = n, |E| = m$ are marked, then the expander-walk Chernoff bound says that, for an n-step random walk, the number of marked vertices visited is strongly concentrated at $n/2$. In [Ta-17], Ta-Shma proved that the parity function is fooled by expander random walks by showing that the parity of the number of visited marked nodes has exponentially small bias. To analyze this more rigorously, we define the statistical distance (which is essential for our notion of fooling) as the total variation distance d_{TV} or TVD.

Definition 14 (Total Variation Distance (TVD))**.** *Given a measure space $(\Omega, \mathcal{F}, \mu)$ and a σ-algebra $\mathcal{A} \subseteq \mathcal{F}$, the total variational distance $d_{TV}(\mu_1, \mu_2)$ between prob-*

ability measures $\mu_1, \mu_2 : \mathcal{F} \to \mathbb{R}$ *is*

$$d_{TV}(\mu_1, \mu_2) = \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)|$$

Similarly, the ℓ_1 *distance between* μ_1 *and* μ_2 *is defined as* $d_{\ell_1}(\mu_1, \mu_2) = 2d_{TV}(\mu_1, \mu_2)$. *Additionally, for countable* Ω *and* $\mathcal{A} = 2^{\Omega}$, *we have* $d_{\ell_1}(\mu_1, \mu_2) = \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)|$. *Transitively,*

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)|$$

Recall [Gil98]'s result about the expander Chernoff bound in lemma 1.3.6. An alternative, but equivalent, formulation of the result is that if λ is small enough, then for every $B \subseteq V$, the number of times we visit B in a random walk of length t in an expander random walk is very close to the expected value, and the probability that the expander random walk deviates from the expected value of the number of times B is visited is very similar to the probability that t truly random variables deviate from the expected value. [Ta-17] compared the parity of the number of times an expander random walk fell into set B with the parity of the number of times a truly random variable's realization is contained in B and showed that the distributions were also almost identical.

Lemma 1.4.1 (Ta-Shma)**.** *Let* S_0 *be the parity of the number of times an expander random walk of length* t *hit* B *and* S_1 *be the parity of the number of times a uniformly random sequence of variables realized a value contained in* B. *Then, for* $\epsilon_0 = 0.8, \beta = 0.01$, *and small* λ *such that* $\epsilon_0 + 2\beta + 2\lambda < 0.9$*:*

$$\mathrm{TVD}(S_0, S_1) = (\epsilon_0 + 2\beta + 2\lambda)^{\lfloor t/2 \rfloor}$$

For convenience, we restate Ta-Shma's proof of lemma 1.4.1 in Chapter A.

Ta-Shma's breakthrough construction of optimal ϵ-balanced codes [Ta-17] which showed that expander random walk can fool the extremely sensitive parity function led to an exciting series of work in [GK21] [GV22] which generalizes [Ta-17]'s result to all symmetric functions and a broad class of test functions including permutation branching programs, read-only branching programs, AC^0, and decision trees.

1.5 Pseudorandomness against Symmetric Functions

The success in showing that expander random walks could fool the parity function gave rise to the question of what other functions could be fooled. One general candidate of functions are symmetric functions. Informally, even though an expander random walk mixes very well, each adjacent vertex in the random walk sequence is extremely correlated. Therefore, since symmetric functions loose all information about the order of vertices, it is unable to capture these local correlations and therefore cannot distinguish between an expander random walk and a uniform distribution.

[GK21] proved that the TVD between the weight distribution of the sticky random walk and the binomial distribution is $\Theta(\lambda)$, and goes on to prove results that sticky random walks fools the majority and parity function. We summarize some of their proof methodologies and describe our work which extends this in Chapter 2. [CPT21] and its follow-up work [Coh+22] study the arbitrary expander graph G with a balanced binary labeling on its vertices using Fourier analysis on $\mathbb{Z}_2^n$ to prove that expander random-walks fool all symmetric functions with $\Theta(\lambda)$ error, and conclude by showing that test-functions computed by AC^0 circuits are fooled by expander random walks with a constant spectral gap. Finally, in 2022, Golowich and Vadhan's breakthrough paper [GV22] on the pseudorandomness of expander graphs showed that, for any p-ary labeling of its vertices, expander random walks fool tests computed by all symmetric functions and permutation-branching programs upto an $\Omega(\lambda)$ and $O(\lambda p^{O(p)})$ error, using Fourier analysis on $\mathbb{Z}$. We summarize some of the proof methods of [GK21] and [GV22] below.

Definition 15 (Symmetric Functions)**.** *A function f on t variables is symmetric if $f(x_1, x_2, \ldots, x_t) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(t)})$, for any permutation $\pi \in S_t$.*

This motivates the work in [GK21] (Pseudobinomiality of the Sticky Random Walk) which studies whether expander random walks can fool symmetric functions. [GK21] studies the canonical expander graph, the sticky random walk on two vertices, which is a modified uniform-probability Markov chain with an additional probability, λ, of staying at the same vertex for the next time-step, since it a useful proxy for studying expander random walks.

We rigorously define the sticky random walks now.

Definition 16. *The Sticky Random Walk $S(n, \lambda)$ is a distribution on n-bit strings that represent n-step walks on a Markov chain with states $\{0, 1\}$ such that for each*

$s \sim S(n,\lambda), \Pr[s_{i+1} = b | s_i = b] = \frac{1+\lambda}{2}$, for $b \in \{0,1\}$, and $s_1 \sim Ber(1/2)$ such that $\Pr[s_1 = 0] = \Pr[s_1 = 1] = \frac{1}{2}$. As $\lambda \to 0$, the distribution of strings from the Markov chain converges to the distribution of n independent coin-flips.

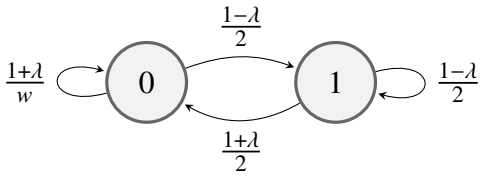

Figure 2: The Markov chain of the sticky random walk $S(n,\lambda)$.

1.5.1. Proof Methodology of Pseudobinomiality of the Sticky Random Walk

[GK21] defines the Krawtchouk functions as an orthogonal basis for functions from $\mathbb{Z}_{n+1} \to \mathbb{R}$. They use them to decompose the TVD into separate orthogonal terms which are individually bounded by the expected-value of the Krawtchouk function.

Definition 17 (Krawtchouk Functions)**.** *The Krawtchouk function $K_k : \mathbb{Z}_{n+1} \to \mathbb{R}$, for $\ell \in \mathbb{Z}_{n+1}$ and an arbitrary n-bit string α for $|\alpha| = \ell$, is*

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

Lemma 1.5.1 (Orthogonality of the Krawtchouk function)**.**

$$\langle K_r, K_s \rangle = \mathop{\mathbb{E}}_{b \sim Bin(n,\frac{1}{2})} [K_r(b) K_s(b)] = \begin{cases} 0, & r \neq s \\ \binom{n}{s}, & r = s \end{cases}$$

We provide a proof of lemma 1.5.1 in A. Since the Krawtchouk functions are orthogonal, any $f : \mathbb{Z}_{n+1} \to \mathbb{R}$ has an expansion $f(\ell) = \sum_{k=0}^{n} \hat{f}(k) K_k(\ell)$, where for $k \in \mathbb{Z}_{n+1}$, the expansion coefficient $\hat{f}(k)$ is $\hat{f}(k) = \frac{1}{\binom{n}{k}} \cdot \mathbb{E}_{b \in \mathrm{Bin}(n,\frac{1}{2})}[f(b) K_k(b)]$. [GK21] then defines a ratio function $p(\ell)$, where $p : \mathbb{Z}_{n+1} \to \mathbb{R}$, which is used to compute the total variation distance (TVD).

Corollary 1.5.1.1 (Krawtchouk invariance)**.** The Krawtchouk function is invariant against choices of α which satisfy $|\alpha| = \ell$. This statement is proven in [Sam98].

$$\mathop{\mathbb{E}}_{\substack{|L|=k \\ \text{Fixed } A \in \mathbb{Z}_p^n \\ |A|=\ell}} [(-1)^{A \cdot B}] = \mathop{\mathbb{E}}_{\substack{|L|=k \\ \text{Random } A' \in \mathbb{Z}_p^n \\ |A'|=\ell}} [(-1)^{A' \cdot B}]$$

Definition 18 (Probability ratio function)**.** *We let the injective probability ratio mapping* $p : \mathbb{Z}_{n+1} \mapsto \mathbb{R}$ *be defined as* $p(\ell) = \Pr_{s\sim S}[|s| = \ell]/(\binom{n}{\ell}2^{-n})$*. Intuitively,* $p(\ell)$ *is a ratio of the probability that a sticky random walk has a hamming weight of* ℓ *to the probability that n samples from the uniform distribution on* $\{0, 1\}$ *yield* ℓ *0s.*

Lemma 1.5.2. *For* $p(\ell) = \Pr_{s\sim S}[|s| = \ell]/(\binom{n}{\ell}2^{-n})$*, the expansion coefficient is* $\hat{p}(k) = 1/\binom{n}{k} \cdot \mathbb{E}_{s\sim S}[K_k(|s|)]$*. This implies for* $s \sim S(n, \lambda)$*, we must have that* $\Pr[|s| = \ell] = 1/2^n \cdot \sum_{k=0}^{n} K_\ell(k)\,\mathbb{E}[K_k(|s|)]$*. We refer the reader to [GK21] for the proof of this result.*

To compute the TVD between the binomial probability distribution and sticky random walk distribution of the Hamming weight of a string sampled, [GK21] uses the convexity of the expected value operator to show that:

$$\text{TVD} = \frac{1}{2}\sum_{\ell=0}^{n}\left|\Pr[|s| = \ell] - \binom{n}{\ell}2^{-n}\right| \leq \sqrt{\mathop{\mathbb{E}}_{b\sim\text{Bin}(n,1/2)}[(p(b)-1)^2]}$$

Following this, some algebraic manipulation leads to the following lemma.

Lemma 1.5.3. *The TVD of the hamming weights between the sticky random walk and the binomial distribution is:*

$$\text{TVD} \leq \sum_{k=1}^{n}\hat{p}(k)^2\binom{n}{k} = \sum_{k=1}^{n}\frac{1}{\binom{n}{k}}\,\mathbb{E}[K_k(|s|)]^2$$

To compute the expectation of the Krawtchouk function, [GK21] defines the shift of a subset.

Definition 19 (Shift of Subsets)**.** *For even-sized subsets* $A \subseteq [n]$*, where* $A = a_1 < \cdots < a_m$ *are the elements of A in increasing order, define* shift : $A \to \mathbb{R}$ *as:*

$$\text{shift(A)} = \sum_{i=1}^{|A|/2}(a_{2i} - a_{2i-1})$$

This definition is handy in lemmas 1.5.4 and 1.5.5 in [GK21].

Lemma 1.5.4. *For any* $A \subseteq [n]$*, we have that* $\mathbb{E}\left[\prod_{i\in A}(-1)^{s_i}\right] = \begin{cases} 0 & |A| \text{ odd} \\ \lambda^{\text{shift(A)}} & |A| \text{ even} \end{cases}$

The proof for the above lemma involves defining a new re-parameterized random variable to model transitions between the sticky random walk, using the definition of the expectation, and considering each term in the ensuing expression. We similarly refer the reader to [GK21] for a proof of this statement. This lemma then lends itself to lemma 1.5.5 which states the expectation of the Krawtchouk function.

Lemma 1.5.5. *The expectation of the Krawtchouk function is given by:*

$$\mathbb{E}[K_k(|s|)] = \begin{cases} 0 & k \text{ odd} \\ \sum_{T \in \binom{[n]}{k}} \lambda^{\text{shift(A)}} & k \text{ even} \end{cases} = \begin{cases} 0 & k \text{ odd} \\ \sum_{m=k/2}^{n-k/2} \binom{m-1}{k/2-1}\binom{n-m}{k/2}\lambda^m & k \text{ even} \end{cases}$$

The first equality in the above lemma follows by definition, and the second equality is shown by a counting argument which shows that the number of subsets $T \in \binom{[n]}{2k}$ with shift(T) = m is $\binom{m-1}{k-1}\binom{n-m}{k}$. Then, substituting the result of lemma 1.5.4 into lemma 1.5.3 using algebraic manipulation and some standard inequalities yields that TVD $\leq O(\lambda)$ when $\lambda < 0.16$.

This TVD bound shows that the Hamming weight counting function is fooled by the sticky random walk. [GK21] goes on to show that TVD $\geq \Omega(\lambda)$ using a calculation of the moments and the central limit theorem, which we do not consider further. We will revisit these proof methods when we summarize our proof for a TVD upper bound for a generalized q-ary sticky random walk.

[CPT21][Coh+22] then used Fourier analysis to expand this result. Specifically, they showed that test functions computed by AC^0 circuits and symmetric functions are fooled by the random walks on the full expander random walk, but only for balanced binary labelings. This culminated in the work of [GV22].

Proof Methodology of Pseudorandomness of Expander Random Walks for Symmetric Functions and Permutation Branching Programs

These works culminate in [GV22] which establishes that random walks on expanders where the vertices are labeled from an arbitrary alphabet $[p] = \{1, \ldots, p\}$ where f_b is the fraction of vertices labeled b for any $b \in [p]$, can fool symmetric functions (upto an $O(\lambda(p/\min_{b\in[p]})^{O(p)})$ error) and permutation branching programs. Specifically, we are interested in the result concerning symmetric function which we restate below.

Lemma 1.5.6. *Fooling symmetric functions (Corollary 4 of [GV22]). For all integers $t \geq 1$ and $p \geq 2$, let $G = (G_i)_{1 \leq i \leq t-1}$ be a sequence of λ-spectral expanders on a shared vertex set V with labeling $\mathrm{val} : V \to [p]$ that assigns each label $b \in [p]$ to f_b-fraction of the vertices. Then, for any label b, we have that the total variation distance between the number of b's seen in the expander random walk and the uniform distribution on $[p]$ has the following bound (where $[\Sigma(Z)_b]$ counts the number of occurrences of b in Z) is:*

$$\mathrm{TVD}([\Sigma(\mathrm{RW}^t_{\mathcal{G}})]_b, [\Sigma(U[d])]_b) \leq O\left(\left(\frac{p}{\min_{b \in [p]} f_b}\right)^{O(p)} \cdot \lambda\right)$$

We present some of the proof ideas here, while leaving some lemmas in Chapter A.

We first introduce the following notation from [GV22]. Consider a graph sequence $\mathcal{G} = (G_1, \ldots, G_{t-1})$ on a common set of vertices V, and let the random variable $\mathrm{RW}^t_{\mathcal{G}}$ denote the t-step random walk on V, where the i'th step is taken on graph G_i. Let the vertex labeling be $\mathrm{val} : V \to \mathbb{Z}_d$ such that $\mathrm{val}(v_0, \ldots, v_{t-1}) = (\mathrm{val}(v_0), \ldots, \mathrm{val}(v_{t-1}))$. Next, let $\mathcal{G}$ be a sequence of λ-expanders and $\mathcal{J}$ denote the complete graph with self-loops. Then, let $\Sigma : \mathbb{Z}_d^{\mathbb{Z}_{t-1}} \to \mathbb{Z}_{t+1}^{\mathbb{Z}_d}$ denote the histogram function where $(\Sigma a)_b = |\{i \in \mathbb{Z}_t : a_i = b\}|$. By noting that all symmetric functions factor through Σ, [GV22] studies the strongly generalized problem of bounding the TVD between $\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{J}}))$.

Lemma 1.5.7 (Theorem 18 of [GV22])**.** . *Fix regular graph sequences $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ on a common set of vertices where for all $i \neq u, 1 \leq u, t-1$, $G_i = G'_i$ with $\lambda(G) = \lambda(G') \leq 1/100$. For a fixed labeling binary labeling that assigns labels 0 and 1 to p_0 and p_1 fraction of vertices respectively, we have that for any $c > 0$,*

$$\sum_{j \in \mathbb{Z}_t : |j - p_1 t| \geq c} |\Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})) = (t-j, j)] - \Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})) = (t-j, j)]| \leq 4000 e^{-c^2/8t} \cdot \frac{\|G'_u - G_u\|}{t}$$

Intuitively, lemma 1.5.7 bounds the probability that a vertex labeled 1 is visited c more times than the expected number in $\mathcal{G}$, relative to $\mathcal{G}'$, by the normed difference $\|G_u - G'_u\|$ and a decreasing exponential function of c. Corollary 19 in [GV22] shows that lemma 1.5.7 yields a TVD between $\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and the binomial

distribution. Setting $\mathcal{G}' = J$ and letting $\mathcal{G}$ be a λ-spectral expander yields:

$$\sum_{j\in\mathbb{Z}_t:|j-p_1t|\geq c} |\Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})) = (t-j,j)] - \Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})) = (t-j,j)]| \leq 4000\lambda e^{-c^2/8t}$$

The proof of lemma 1.5.7 (theorem 18) relies on theorem 20 and lemma 23 from [GV22] which we state here and prove in Chapter A.

Lemma 1.5.8 (Theorem 20 of [GV22])**.** *Let the vector denote g the ℓ_1 norm of the difference in distributions we are considering:*

$$g = (\Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}) = (t-j,j))] - \Pr[\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j))])_{j\in\mathbb{Z}_t} \in \{-1,1\}^{\mathbb{Z}_t}$$

Next, extend g to the (countably) infinite (and normalized) vector g^{sr} for $s = \pm 1, 0 \leq r \leq \frac{1}{2}$, and $sg_j = 0$ for $j \notin \mathbb{Z}_{t+1}$, $g^{(sr)} = (e^{sr(j-p_1t}g_j)_{j\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$. Then,

$$\|g^{(sr)}\| \leq \|G_u - G'_u\| \cdot p_0p_1e^{2p_0p_1tr^2} \cdot \min\left\{44, \frac{22r^2}{(p_0p_1t)^{1/4}} + \frac{70}{(p_0p_1t)^{5/4}}\right\}$$

Lemma 1.5.9 (Lemma 23 of [GV22])**.** *For $k \geq 0$, let*

$$S^+_k = \{j \in \mathbb{Z}_t : k\sqrt{p_0p_1t} \leq j - p_1t \leq (k+1)\sqrt{p_0p_1t}\}$$

$$S^-_k = \{j \in \mathbb{Z}_t : (k+1)\sqrt{p_0p_1t} \leq j - p_1t \leq -k\sqrt{p_0p_1t}\}$$

Then, set $s = \pm 1, r > 0$. For $j \in \mathbb{Z}_t$, we have that

$$|g_j| \leq e^{-sr(j-p_1t)} \cdot \|g^{(sr)}\|$$

and for $p_0p_1t \geq 1$, we have that

$$\|g_{S^s_k}\|_1 \leq \sqrt{2}(p_0p_1t)^{1/4}e^{-rk\sqrt{p_0p_1t}} \cdot \|g^{(sr)}\|$$

[GV22] goes on to show similar bounds for tests computed by permutation branching programs, which extends Braverman's result [BCG20] for read-once branching programs.

1.6 Pseudorandomness against AC0 circuits

As a digression, the main results of [CPT21] and the follow-up work [Coh+22] are superseded in [GV22]. However, [Coh+22] does prove a result of expander random walks fooling test functions beyond symmetric functions using a Fourier analytic approach. Specifically, we present theorem 1.5 in [Coh+22] and refer the reader to the original paper for a proof.

Lemma 1.6.1. *For every function $f : \{0,1\}^t \to \{0,1\}$, let the total variation distance between f when applied to a uniform distribution and an expander random walk with spectral expansion constant λ be denoted by $\varepsilon_\lambda(f)$. Then, if f is computable by a size-s depth-d circuit, then $\varepsilon_\lambda(f) = O(\sqrt{\lambda} \cdot (\log s)^{2(d-1)})$. Further, $\varepsilon_\lambda(f) = O(\sqrt{\lambda} \cdot \mathrm{DT}(f)^2)$, where $\mathrm{DT}(f)$ denotes the decision tree complexity of f.*

Two consequences of lemma 1.6.1 in [Coh+22] are that 1) any test-function in the complexity class AC^0 is fooled by an expander random walk, any class of functions with a bounded Fourier tail ([Tal17]) are fooled by expander random walks. This leaves us at our current state of the pseudorandomness of expander random walks.

Chapter 2

PSEUDORANDOMNESS WITH *ARBITRARY* LABELS: STICKY RANDOM WALK

Recall theorem 18 of [GV22] (presented in lemma 1.5.6) which says that:

$$\mathrm{TVD}([\Sigma(\mathrm{RW}^t_{\mathcal{G}})]_b, [\Sigma(U[d])]_b) \leq O\left(\left(\frac{p}{\min_{b\in[p]} f_b}\right)^{O(p)} \cdot \lambda\right)$$

[GV22] asks whether the $(\frac{p}{\min_{b\in[p]} f_b})^{O(p)}$ dependence in the upper bound of the total variation distance is tight.

Contributions. In this paper, we answer in the negative for a family of graphs. Specifically, we present a family of generalized sticky random walks (where the alphabet size can be arbitrarily large), where we find that the optimal TVD is $O(\lambda)$, for $\lambda < 0.27$, whereas lemma 1.5.6 in [GV22] predicts a bound of $O(\lambda p^{2p})$, which provides evidence that the $(\frac{p}{\min_{b\in[p]} f_b})^{O(p)}$ factor is not tight. Further, [GK21] studied the sticky random walk because it was an "essential step" to understanding the full expander random walk - specifically, theorem 4 in [GK21] shows that every λ-parameterized sticky random walk is bijective with a corresponding expander graph. We extend their result in lemma 2.4.2 by showing that our generalized sticky random walk (parameterized by λ and p also correspond to expander graphs with a linearly-reduced spectral expansion of λp. We then show that our generalized sticky random walk reduces from [GK21]'s two-vertex sticky random walk in lemma 2.4.1. Finally, in Appendix B we provide a novel alternate treatment of the Krawtchouk functions into the complex domain which can be used to attain an $O(\lambda p^p)$ bound on the TVD.

	SRW	ERW	G-SRW	G-ERW	Method
[GK21]	✓	×	×	×	Krawtchouk Functions
[Coh+22]	✓	✓	×	×	Fourier on $\mathbb{Z}^d$
[GV22]	✓	✓	×	Almost	Fourier on $\mathbb{Z}$
Our work	✓	×	✓	×	Generalized Krawtchouk Functions

Table 2.1: **Comparing our work with previous works.** Here, SRW denotes sticky random walk, ERW denotes expander random walk, G- denotes generalized, and a ✓ at the corresponding location implies a result of an optimal $O(\lambda)$ bound.

2.1 Generalizing the Sticky Random Walk

We consider the case where the vertices of the sticky random walk (SRW) can be labeled with an arbitrary alphabet $\mathbb{Z}_p$, since a decomposition of the vertex-set $V = \{0, \ldots, p-1\} = V_1 \sqcup \ldots \sqcup V_q$ for $q \geq 2$, could allow us to model random walks where the probability of transitioning between different states is asymmetric, while allowing us to study the pseudorandomness of random walks on graphs with vertices with more complex labelings which has already been explored in [GV22]. This section generalizes the sticky random walk on p characters, and provides the context for bounding the total variation distance between the sticky random walk and U_p^n.

Definition 20 (The generalized sticky random walk)**.** The generalized sticky random walk $S(n, p, \lambda)$ is an n-step long, p-symbol walk on a Markov chain with p states $\mathbb{Z}_p$ labeled as vertices on a complete graph with self-loops J_p, where $s_0 \in_U \mathbb{Z}_p$ and at each subsequent step, we either stick to the same state with probability $\frac{1}{p} + (p-1)\lambda$, or change to any other state with uniform probability $\frac{1}{p} - \lambda$. So, for instance, comparing $S(n, 4, \lambda)$ to $S(n, 4, 0) = U_4^n$ (the uniform random walk on 4 vertices) yields the following Markov chain graphs:

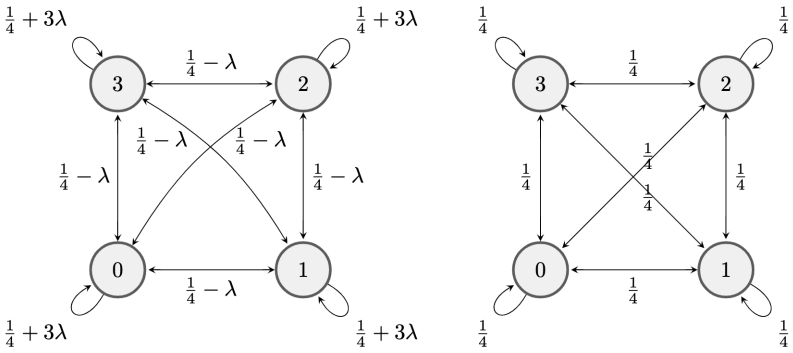

Figure 3: *Markov chains of the generalized sticky random walk on 4 states and the uniform distribution on* 4 *states. The figure on the left corresponds to $S(n, 4, \lambda)$, the λ-biased sticky random walk on four vertices, and the figure on the right corresponds $S(n, 4, 0) = U_4^n$, the unbiased random walk on four vertices.*

Proposition 1 (Probability Invariance Under Permutations)**.** *For any $x_1, x_2 \in \mathbb{Z}_p^n$, we have $Pr[x_1] = Pr[x_2]$ iff $|(i, i+1)|$ such that $(x_1)_i = (x_1)_{i+1}$ is equal to $|(j, j+1)|$ such that $(x_2)_j = (x_2)_{j+1}$. This can be shown by considering the products of conditional probabilities on each state. A slight weakening of this statement is that for any permutation π such that $\pi : \mathbb{Z}_p \to \mathbb{Z}_p$, we must have that $Pr[x_1 \ldots x_n] =$*

$Pr[\pi(x_1)...\pi(x_n)]$. *Consider the case of* $p = 2$. *Then, the lemma yields that* $Pr(x) = Pr(\bar{x})$, *which says that inverting the labels of a string from the sticky random walk does not change its probability. This proposition extends the same argument to all* $p \in \mathbb{N}$ *for* $p \geq 2$.

Proposition 2 (Krawtchouk Orthogonality)**.** The orthogonality of the Krawtchouk function $K_k(\ell)$ implies that for any function $f : \mathbb{Z}_{n+1} \to \mathbb{R}$, there exists a unique expansion $f(\ell) = \sum_{k=0}^{n} \hat{f}(k) K_k(\ell)$, where for $0 \leq k \leq n$,

$$\hat{f}(k) = \mathbb{E}\left[\binom{n}{k} f(b) K_k(b)\right]$$

Definition 21 (Probability ratio)**.** Let $q : \mathbb{Z}_{n+1} \to \mathbb{R}$, where $q(\ell) = \frac{\Pr_{s\sim S}[|s|_0=\ell]}{\binom{n}{\ell}(p-1)^{n-\ell}} p^n$. Intuitively, $q(\ell)$ is the ratio of the probability of getting a string with ℓ 0s from the generalized sticky random walk $S(n, \lambda, p)$ to the probability of getting a string with ℓ 0s from U_p^n.

Lemma 2.1.1. *[Krawtchouk coefficient of the probability ratio]* Expanding $q(\ell)$ through the Krawtchouk function expansion in Proposition 2s yields that:

$$\hat{q}(k) = \frac{1}{\binom{n}{k}(p-1)^{n-k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]$$

We provide a proof of this lemma in A.

Lemma 2.1.2. For $s \in S(n, p, \lambda)$, we have that

$$\Pr[|s|_0 = \ell] = \frac{1}{p^n} \sum_{k=0}^{n} K_\ell(k) \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]$$

We provide a proof of lemma 2.1.2 in A.

Therefore, we observe that to compute $\Pr[|s|_0 = \ell]$, it is imperative to calculate the expected value of the Krawtchouk function. Section 2.2 is devoted to computing $\mathbb{E}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]$.

2.2 Expected Value of the Krawtchouk Function

We first generalize the shift function from definition 19.

Definition 22 (Shift Function)**.** Given any set $T \subseteq [n]$ such that $|T| = k$, let $a_1 < ... < a_k$ be the elements of T in increasing order. Then, for any $c \in \mathbb{Z}_p$, let

$$\mathrm{shift}_c(T) = \sum_{i=0}^{\lfloor |k-c|/p \rfloor} (a_{c+ip} - a_{c+ip-1})$$

Then, for any c such that $k \mod p = -c$, and for any $d \in \mathbb{Z}_{n+1}$, let $\phi_c(d)$ denote the number of subsets of $[n]$ of size k such that $\text{shift}_c(T) = d$. Note that for any $t \leq 0, a_t = 0$.

Lemma 2.2.1. The expected value of the Krawtchouk function is given by:

$$\mathbb{E}[K_k(|s|_0)] = \begin{cases} (p-1)^{n-k} \sum_{d=k}^{n-k} \phi_0(d)\lambda^d, & \text{if } c = k \mod p \equiv 0 \\ 0, & \text{if } c = k \mod p \not\equiv 0 \end{cases}$$

We provide a proof of this lemma in A.

Lemma 2.2.2. For $c \in \mathbb{N}$ where $0 \leq c \leq p$, and for $d \in \mathbb{N}$ where $k \leq d \leq n-k$, the number of k-sized subsets of $[n]$ that satisfy $\text{shift}_c(T) = d$ *is:*

$$\phi_c(d) = \frac{1}{p^k} \sum_{\substack{T \in \binom{[n]}{k} \\ shift_c(T) = d}} 1 = \frac{1}{p^k} \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

We provide a proof of this lemma in A

Corollary 2.2.2.1. By combining the results from lemmas 2.2.1 and 2.2.2, we have that the expectation of the Krawtchouk function is:

$$\mathbb{E}[K_k(|s|_0)] = \begin{cases} \frac{1}{p^k} \sum_{d=k}^{n-k} \binom{d-1}{\lfloor \frac{k}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{k}{p-1} \rfloor} \lambda^d, & \text{if } k \mod p \equiv 0 \\ 0, & \text{if } k \mod p \not\equiv 0 \end{cases}$$

Thus, having computed $\mathbb{E}[K_k(|s|_0)]$, we are now prepared to upper-bound the total variation distance between $[\Sigma(S(n,p,\lambda))]_0$ and $[\Sigma(\mathrm{U}_p^n)]_0$.

2.3 Upper Bound for the Total Variation Distance

We devote this chapter to deriving an optimal upper bound of the total variation distance of $O(\lambda)$.

Lemma 2.3.1. The total variational distance between the generalized sticky random walk on p vertices and the uniform distribution on p states is given by:

$$\text{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) = \frac{1}{2} \mathop{\mathbb{E}}_{b \sim \mathrm{U}_p^n} [|q(b) - 1|]$$

We provide a proof of this lemma in A.

Corollary 2.3.1.1. The total variational distance between the generalized sticky random walk and the uniform distribution on p states has the following upper bound as a result of the convexity of the expected value operator:

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) \leq \frac{1}{2}\sqrt{\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2}$$

We then see how lemmas 2.1.1, 2.2.1, and 2.3.1 interact in the below lemma and proof methodology which utilizes the reciprocity relation of the Krawtchouk functions.

Lemma 2.3.2. For $k \leq n$ and for $b \sim [\Sigma(\mathrm{U}_p^n)]_0$, we have that

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 = \sum_{k=1}^{n} \frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}(p-1)^{n-k}}$$

Proof. We know that $q(b)$ has a unique Krawtchouk expansion, where the coefficients on each Krawtchouk basis are given by Proposition 2. So, we observe that:

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 = \mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}\left[\left(\sum_{k=0}^{n}\hat{q}(k)K_k(b)-1\right)^2\right]$$

Then, recall that $\hat{q}(k) = \frac{\mathbb{E}[K_k(|s|_0)]}{\binom{n}{k}(p-1)^{n-k}}$. So, $\hat{q}(0) = \frac{\mathbb{E}[K_0(|s|_0)]}{(p-1)^n} = \frac{1}{(p-1)^n}$. Similarly, by the definition of the Krawtchouk function and the reciprocity relation $\frac{K_k(\ell)}{\binom{n}{k}(p-1)^{n-k}} = \frac{K_s(\ell)}{\binom{n}{s}(p-1)^{n-s}}$, we have that $K_0(b) = K_n(b) = (p-1)^n$. Therefore, $\hat{q}(0)K_0(b) = 1$. Hence, the above equation simplifies to:

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 = \mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}\left[\left(\sum_{k=1}^{n}\hat{q}(k)K_k(b)\right)^2\right]$$

Since the generalized Krawtchouk functions are orthogonal (as proven in lemma 1.5.1), the product of the non-diagonal entries in the above term all evaluate to 0. Thus, counting the residuals, we have that the square of the summation is just the summation of the squared terms that it contains. Thus, exploiting the orthogonality of the generalized Krawtchouk functions and the linearity of the expectations, we

write that:

$$\begin{aligned}
\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 &= \mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}\left[\sum_{k=1}^{n}\hat{q}(k)^2K_k(b)^2\right] \\
&= \sum_{k=1}^{n}\hat{q}(k)^2\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[K_k(b)^2] \quad \text{(linearity of expectations)} \\
&= \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}^2(p-1)^{2n-2k}}\cdot\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[K_k(b)^2]
\end{aligned}$$

Finally, we use lemma 1.5.1 to write $\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[K_k(b)^2]$ as $\langle K_k, K_k\rangle = \binom{n}{k}$.

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 = \sum_{k=1}^{n}\frac{\binom{n}{k}\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}^2(p-1)^{2n-2k}} = \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}(p-1)^{2n-2k}}$$

□

Finally, we prove the total variation distance bound between the hamming weight distribution of the generalized sticky distribution and the uniform distribution.

Theorem 2.3.3. For $\lambda \leq 0.27$,

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) \leq O(\lambda)$$

Proof. Substituting the result of Corollary 2.3.1.1 into the equation derived in lemma 2.3.2, and scaling the indexes of the summation, we have that:

$$\begin{aligned}
\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2 &= \sum_{k=1}^{n/p}\frac{1}{\binom{n}{pk}(p-1)^{2n-2pk}}\left(\sum_{d=pk}^{n-pk}\frac{1}{p^k}\binom{d-1}{\lfloor\frac{kp}{p-1}\rfloor-1}\binom{n-d}{\lfloor\frac{kp}{p-1}\rfloor}\lambda^d\right)^2 \\
&= \frac{1}{p^{2k}}\sum_{k=1}^{n/p}\frac{1}{\binom{n}{pk}(p-1)^{2n-2pk}}\left(\sum_{d=pk}^{n-pk}\binom{d-1}{\lfloor\frac{k}{1-\frac{1}{p}}\rfloor-1}\binom{n-d}{\lfloor\frac{k}{1-\frac{1}{p}}\rfloor}\lambda^d\right)^2 \\
&\leq \frac{1}{p^{2k}}\sum_{k=1}^{n/p}\frac{\binom{n}{k}^2}{\binom{n}{pk}(p-1)^{2n-2pk}}\left(\sum_{d=pk}^{n-pk}\binom{d-1}{k-1}\lambda^d\right)^2 \\
&\leq \frac{1}{p^{2k}}\sum_{k=1}^{n/p}\frac{\binom{n}{k}^2}{\binom{n}{pk}}\left(\sum_{d=pk}^{n-pk}\binom{d-1}{k-1}\lambda^d\right)^2
\end{aligned}$$

Note the following generating function relation that $(\frac{x}{1-x})^k = \sum_{m \geq k} \binom{m-1}{k-1} x^m$. Then,

$$\begin{aligned}
\mathop{\mathbb{E}}_{b \sim [\Sigma(U_p^n)]_0} [q(b) - 1]^2 &\leq \frac{1}{p^{2k}} \sum_{k=1}^{n/p} \frac{\binom{n}{k}^2}{\binom{n}{pk}} \left(\frac{\lambda}{1-\lambda}\right)^{2k} \\
&\leq \frac{1}{p^{2k}} \sum_{k=1}^{n/p} \left(\frac{pk}{n}\right)^{pk} \left(\frac{en}{k}\right)^{2k} \left(\frac{\lambda}{1-\lambda}\right)^{2k} \quad \text{(We prove this in section A.0.1)} \\
&= \sum_{k=1}^{n/p} \left(\frac{pk}{n}\right)^{pk-2k} \left(\frac{e\lambda}{1-\lambda}\right)^{2k} \\
&\leq \sum_{k=1}^{n/p} \left(\frac{e\lambda}{1-\lambda}\right)^{2k} \leq O(\lambda^2), \quad \text{(for } \lambda \leq \frac{1}{1+e} \text{ by geometric sums)}
\end{aligned}$$

Therefore, for $\lambda \leq \frac{1}{1+e} \approx 0.27$, we have that the total variation distance is atmost $\sqrt{\mathop{\mathbb{E}}_{b \sim [\Sigma(U_p^n)]_0} [p(b) - 1]^2} \leq O(\lambda)$. □

Proof Strengths and Limitations:

When $\lambda > 0.27$, our proof method fails to provide the desired $O(\lambda)$ total variation distance since $\sum_{k=1}^{n/p} (\frac{e\lambda}{1-\lambda})^{2k}$ does not converge and goes to ∞. We do, however, reach a higher lower-bound on the radius of convergence ($\lambda \leq 0.27$) than [GK21]'s $\lambda \leq 0.16$ and [GV22]'s more general result but which is only valid for $\lambda < 0.01$ and a fixed p in their interpretation of their $O(\lambda p^{O(p)})$ result, whereas our methodology allows us to remove the dependency of p in the total variation distance (though only for the generalized sticky random walk). We conjecture that $\lambda \leq 0.27$ is not the optimal radius of convergence for the generalized sticky random walk and leave this as an open problem for future research directions in this topic to resolve.

2.4 Generalized Sticky Random Walk Markov Chain is an Expander

In this section, we present and prove some equivalence properties between the generalized families of sticky random walks and show that the generalized sticky random walk is also an expander graph.

Lemma 2.4.1. Let p be the number of vertices in the generalized sticky random walk. Then, for all $p \mod k \equiv 0$, $S(n, p, \lambda)$ reduces from $S(n, k, p\lambda(1 - \frac{1}{k}))$.

Proof. Consider a generalized sticky random walk which details an irreducible homogeneous Markov chain on p states where the probability of staying at the same state is $\frac{1}{p} + (p-1)\lambda$ and the probability of switching states is $\frac{1}{p} - \lambda$. Then, consider a 'grouped' random-walk, for a grouping of states $V = [p] = V_0 \sqcap \cdots \sqcap V_{k-1}$, where V_i contains an arbitrary selection of p/k vertices and where $p \mod k \equiv 0$. Then, $S(n, p, \lambda)$ details an n-step long random walk on the generalized sticky random walk. Then, we note that the probability that the current state of the grouped random walk stays at itself must be $(\frac{1}{k} + (p-1)\lambda) + (\frac{1}{p} - \lambda)(\frac{p}{k} - 1) = \frac{1}{k} + p\lambda(1 - \frac{1}{k})$. Here, the first term comes from the probability of any state E in the sticky random walk staying at itself, and the second term comes from the probability that any other vertex in the same group as E transitions to E. Since this is true for any of the grouped vertex, we conclude that our grouping of the random walk yields a sticky random walk on k vertices and bias $p\lambda(1 - \frac{1}{k})$, which completes the reduction. □

Lemma 2.4.2. Every generalized sticky random walk $S(n, p, \lambda)$ corresponds to a n-step long random walk on a $p\lambda$-spectral expander with p vertices.

Proof. We first extend Definition 45 of the sticky random walk matrix in [GV22]. For subsets $A, B \in [p]$, let $J_{A,B} \in \mathbb{R}^{A \times B}$ denote the matrix with all entries equal to $\frac{1}{|A|}$. Then, for $\lambda \in [0, 1]$, let $G_{\lambda,p} \in \mathbb{R}^{p \times p}$ denote the generalized sticky random walk matrix. Then, by definition, we write that:

$$G_{\lambda,p} = (1 - \lambda)\mathrm{J}_{V,V} + p\lambda \mathrm{I}_{\{n \times n\}}$$

Then, note that $\|I_{n \times n}\|_2$ is 1, as it acts as J on the orthogonal subspaces $\mathbb{R}$ of $\mathbb{R}^p$. Therefore, $\lambda(G_{\lambda,p}) \leq p\lambda$ since the eigenvector of $G_{\lambda,p}$ is orthogonal to $J_{V,V}$ and so $G_{\lambda,p}\nu = ((1 - \lambda)\mathrm{J}_{v,v} + p\lambda I_{\{n \times n\}})\nu = (1 - \lambda)\mathrm{J}_{v,v}\nu + p\lambda I_{\{n \times n\}}\nu = p\lambda\nu$. The opposite inequality comes from the fact that $\frac{n-1}{n}\vec{\mathbb{1}}_{\{c_0\}} - \frac{1}{n}\sum_{i=1}^{p-1} \vec{\mathbb{1}}_{\{c_i\}} \in \vec{\mathbb{1}}^{\perp}$ is an eigenvector of $G_{\lambda,p}$ with eigenvalue $p\lambda$, which proves the theorem. □

BIBLIOGRAPHY

[Ahm+19] AmirMahdi Ahmadinejad et al. "High-precision Estimation of Random Walks in Small Space". In: *CoRR* abs/1912.04524 (2019). arXiv: 1912.04524. URL: http://arxiv.org/abs/1912.04524.

[AKS04] Manindra Agarwal, Neeraj Kayal, and Nitin Saxena. "Primes is in P". In: *Annals of Mathematics* 160 (2004), pp. 781–793. URL: http://www.cse.iitk.ac.in/primality.pdf.

[Alo86] Noga Alon. *Eigenvalues and Expanders*. 1986.

[BCG20] Mark Braverman, Gil Cohen, and Sumegha Garg. "Pseudorandom Pseudo-distributions with Near-Optimal Error for Read-Once Branching Programs". In: *SIAM Journal on Computing* 49.5 (2020), STOC18-242-STOC18–299. DOI: 10.1137/18M1197734. eprint: https://doi.org/10.1137/18M1197734. URL: https://doi.org/10.1137/18M1197734.

[BT11] Avraham Ben-Aroya and Amnon Ta-Shma. "A Combinatorial Construction of Almost-Ramanujan Graphs Using the Zig-Zag Product". In: *SIAM Journal on Computing* 40.2 (2011), pp. 267–290. DOI: 10.1137/080732651. eprint: https://doi.org/10.1137/080732651. URL: https://doi.org/10.1137/080732651.

[Coh+22] Gil Cohen et al. "Expander Random Walks: The General Case and Limitations". In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Ed. by Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff. Vol. 229. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 43:1–43:18. ISBN: 978-3-95977-235-8. DOI: 10.4230/LIPIcs.ICALP.2022.43. URL: https://drops.dagstuhl.de/opus/volltexte/2022/16384.

[Con19] David Conlon. *Hypergraph expanders from Cayley graphs*. 2019. arXiv: 1709.10006 [math.CO].

[CPT21] 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. Virtual, Italy: Association for Computing Machinery, 2021, pp. 1643–1655. ISBN: 9781450380539. DOI: 10.1145/3406325.3451049. URL: https://doi.org/10.1145/3406325.3451049.

[DLV22] Andreea Deac, Marc Lackenby, and Petar Veličković. *Expander Graph Propagation*. 2022. arXiv: 2210.02997 [cs.LG].

[Fri03] J. Friedman. “Relative expanders or weakly relatively Ramanujan graphs”. In: (2003), pp. 19–35.

[GG81] O. Gabber and Z. Galil. “Explicit Constructions of Linear Size Superconcentrators”. In: (1981), pp. 407–420.

[Gil98] David Gillman. “A Chernoff Bound for Random Walks on Expander Graphs”. In: *SIAM Journal on Computing* 27.4 (1998), pp. 1203–1220. DOI: 10.1137/S0097539794268765. eprint: https://doi.org/10.1137/S0097539794268765. URL: https://doi.org/10.1137/S0097539794268765.

[GK21] Venkatesan Guruswami and Vinayak M. Kumar. “Pseudobinomiality of the Sticky Random Walk”. In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 48:1–48:19. ISBN: 978-3-95977-177-1. DOI: 10.4230/LIPIcs.ITCS.2021.48. URL: https://drops.dagstuhl.de/opus/volltexte/2021/13587.

[GK23] Roy Gotlib and Tali Kaufman. *No Where to Go But High: A Perspective on High Dimensional Expanders*. 2023. arXiv: 2304.10106 [math.CO].

[Gur20] Venkatesan Guruswami. “More Great Ideas in Theoretical Computer Science”. In: (2020).

[GV22] Louis Golowich and Salil Vadhan. “Pseudorandomness of Expander Random Walks for Symmetric Functions and Permutation Branching Programs”. In: *37th Computational Complexity Conference (CCC 2022)*. Ed. by Shachar Lovett. Vol. 234. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 27:1–27:13. ISBN: 978-3-95977-241-9. DOI: 10.4230/LIPIcs.CCC.2022.27. URL: https://drops.dagstuhl.de/opus/volltexte/2022/16589.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. “Expander graphs and their applications”. In: *Bulletin of the American Mathematical Society* 43(4):439–561 (2006).

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. “Pseudorandomness for Network Algorithms”. In: *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*. STOC ’94. Montreal, Quebec, Canada: Association for Computing Machinery, 1994, pp. 356–364. ISBN: 0897916638. DOI: 10.1145/195058.195190. URL: https://doi.org/10.1145/195058.195190.

[Kom+02] J. Komlós et al. *The regularity lemma and its applications in graph theory*. 2002. DOI: 10.1007/3-540-45878-6_3.

[KS96] David R. Karger and Clifford Stein. "A New Approach to the Minimum Cut Problem". In: *J. ACM* 43.4 (1996), pp. 601–640. ISSN: 0004-5411. DOI: 10.1145/234533.234534. URL: https://doi.org/10.1145/234533.234534.

[Lez01] Pascal Lezaud. "Chernoff and Berry-Esséen inequalities for Markov processes". en. In: *ESAIM: Probability and Statistics* 5 (2001), pp. 183–201. URL: http://www.numdam.org/item/PS_2001__5__183_0/.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. "Ramanujan graphs". In: (1988), pp. 261–277.

[Lub17] Alexander Lubotzky. *Ramanujan Graphs*. 2017. DOI: 10.48550/ARXIV.1711.06558. URL: https://arxiv.org/abs/1711.06558.

[Mar75] G.A. Margulis. "Explicit Construction of Concentrators". In: (1975), pp. 325–332.

[Mil75] Gary L. Miller. "Riemann's Hypothesis and Tests for Primality". In: *Proceedings of the Seventh Annual ACM Symposium on Theory of Computing*. STOC '75. Albuquerque, New Mexico, USA: Association for Computing Machinery, 1975, pp. 234–239. ISBN: 9781450374194. DOI: 10.1145/800116.803773. URL: https://doi.org/10.1145/800116.803773.

[Rab80] Michael Rabin. *Probabilistic Algorithm for Testing Primality*. 1980. URL: https://www.sciencedirect.com/science/article/pii/0022314X80900840.

[Rei05] Omer Reingold. "Undirected ST-Connectivity in Log-Space". In: *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '05. Baltimore, MD, USA: Association for Computing Machinery, 2005, pp. 376–385. ISBN: 1581139608. DOI: 10.1145/1060590.1060647. URL: https://doi.org/10.1145/1060590.1060647.

[RR17] Shravas Rao and Oded Regev. "A Sharp Tail Bound for the Expander Random Sampler". In: *arXiv e-prints*, arXiv:1703.10205 (Mar. 2017), arXiv:1703.10205. arXiv: 1703.10205 [math.PR].

[RVW04] Omer Reingold, Salil Vadhan, and Avi Wigderson. *Entropy waves, the zig-zag graph product, and new constant-degree*. June 2004. URL: https://arxiv.org/abs/math/0406038.

[Sam98] Alex Samorodnitsky. "On the optimum of Delsarte's linear program". In: *J. Combinatorial Theory, Ser. A* 96 (1998).

[SS96] M. Sipser and D.A. Spielman. "Expander codes". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1710–1722. DOI: 10.1109/18.556667.

[Ta-17] Amnon Ta-Shma. "Explicit, Almost Optimal, Epsilon-Balanced Codes". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. Montreal, Canada: Association for Computing Machinery, 2017, pp. 238–251. ISBN: 9781450345286. DOI: 10.1145/3055399.3055408. URL: https://doi.org/10.1145/3055399.3055408.

[Tal17] Avishay Tal. "Tight Bounds on the Fourier Spectrum of AC0". In: *32nd Computational Complexity Conference (CCC 2017)*. Ed. by Ryan O'Donnell. Vol. 79. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, 15:1–15:31. ISBN: 978-3-95977-040-8. DOI: 10.4230/LIPIcs.CCC.2017.15. URL: http://drops.dagstuhl.de/opus/volltexte/2017/7525.

[Tre11] Luca Trevisan. "Graph Partitioning and Expanders". In: (2011). URL: https://lucatrevisan.github.io/teaching/cs359g-11/lecture01.pdf.

[Vad13] Salil Vadhan. *Pseudorandomness*. Book On Demand Ltd, 2013.

Appendix A

PROOF OF INTRODUCTORY THEOREMS

Lemma 1.3.1: *Without loss of generality, let $\lambda_1 \geq \cdots \geq \lambda_n$ be the n eigenvalues of $\mathcal{A}$. Then, $1 = \lambda_1 \geq \cdots \geq \lambda_n = -1$.*

Proof. We show this by considering the Laplacian $\mathcal{L}$ of the graph G defined by:

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2}$$

We then show the equivalent statement that if $\lambda_1 \geq \cdots \geq \lambda_n$ are the n eigenvalues of $\mathcal{A}$, then

$$0 = \lambda_1 \geq \cdots \geq \lambda_n = -2$$

We first show that 0 is an eigenvalue of the Laplacian $\mathcal{L}$ with eigenvector $D^{1/2}\vec{\mathbf{1}}$, where $\vec{\mathbf{1}}$ is the all-ones vector of length n. Observe that:

$$\mathcal{L}(D^{1/2}\vec{\mathbf{1}}) = D^{-1/2}(I - D)D^{-1/2}D^{1/2}\vec{\mathbf{1}} = D^{-1/2}(I - D)\vec{\mathbf{1}} = 0$$

The last equality holds because $\vec{\mathbf{1}}$ is an eigenvector of $(I - A)$ which corresponds to an eigenvalue of 0. Thus, 0 is an eigenvalue of $\mathcal{L}$. To show that 0 is a minimal eigenvalue of $\mathcal{L}$, note that since $\mathcal{L}$ is symmetric and positive semidefinite (PSD), we must have that:

$$\lambda_1 := \inf_{x \in \mathbb{R}^n : \|x\| = 1} x^T \mathcal{L} x$$

So, for an arbitrary $v \in \mathbb{R}^n$, observe that:

$$\begin{aligned} v^T \mathcal{L} v &= v^T (1 - D^{-1/2} A D^{-1/2}) v \\ &= \sum_{i \in V} v_i^2 - \sum_{(i,j) \in E} \frac{2 v_i v_j}{\sqrt{d_i d_j}} \\ &= \sum_{(i,j) \in E} \left(\frac{v_i}{\sqrt{d_i}} - \frac{v_j}{\sqrt{d_j}} \right)^2 \\ &\geq 0 \end{aligned}$$

Thus, $\lambda_1 = \inf_{x \in \mathbb{R}^n : \|x\| = 1} x^T \mathcal{L} x \geq 0$.

For the other direction, we observe (by the PSD of $\mathcal{L}$) that $x^T(1-D^{-1/2}AD^{-1/2})x \geq 0$ implies $\lambda_n \leq 2$ as shown below:

$$-x^T D^{-1/2}AD^{-1/2}x \leq x^T x \implies x^T Ix - x^T D^{-1/2}AD^{-1/2}x \leq 2x^T x$$

Factorizing, we get that:

$$\frac{x^T(I - D^{-1/2}AD^{-1/2})x}{x^T x} := \frac{x^T \mathcal{L} x}{x^T x} \leq 2 \implies x^T \mathcal{L} x \leq 2$$

Thus, each eigenvalue is bounded above by 2. So, $\lambda_n \leq 2$, which proves the claim. □

Lemma 1.3.2: *Let $G = (V, E)$ be a d-regular graph with edge expansion h. If any $\epsilon < h$ fraction of edges are removed from G, then the graph has a connected component that spans atleast $1 - \epsilon/2h$ fraction of the vertices.*

Proof adapted from [Tre11]. Let $E' \subseteq E$ be an any subset of $\leq \epsilon|E| = \epsilon d|V|/2$ edges that an adversary may wish to remove from the graph.Let $C_1, \ldots, C_m$ be the connected components of the graph $(V, E - E')$ ordered (WLOG) such that $|C_1| \geq |C_2| \geq \cdots \geq |C_m|$. Observe that:

$$|E'| \geq \frac{1}{2} \sum_{i \neq j} E(C_i, C_j) = \frac{1}{2} \sum_i E(C_i, V - C_i)$$

Then, $|C_1| > |V|/2$, since $|C_1| \leq |V|/2$ would imply that $|E'| \geq \frac{1}{2} \sum_i dh|C_i| = dh|V|/2$, which is impossible if $h > \epsilon$. If $|C_1| \geq |V|/2$, we define $S := C_2 \cup \cdots \cup C_m$. Then, $|E'| \geq E(C_1, S) \geq dh|S|$, which implies $|S| \leq \frac{\epsilon}{2h}|V|$ and so $C_1 \geq (1 - \epsilon/2h)|V|$. This proves our claim. □

Lemma 1.3.3: *For every constant $d \in \mathbb{N}$, any d-regular graph $G = (V, E)$ satisfies $\lambda(G) \geq 2\sqrt{d-1}/d - o(1)$, where the $o(1)$ term vanishes as $n \to \infty$.*

Proof adapted from [Vad13]. Let G be a regular undirected graph and T_d be the infinite d-regular tree. For a graph H and $\ell \in \mathbb{N}$, let $p_\ell(H)$ denote the probability that if we choose a random vertex $v \in H$ and do a random walk of length 2ℓ, we end back at vertex v. Then, we first observe that $p_\ell(G) \geq p_\ell(T_d)$. Second, note that $p_\ell(T_d) \geq C_\ell(d-1)^\ell/d^{2\ell}$. Here, C_ℓ is the Catalan number, which equals the number of properly parenthesized strings of length 2ℓ. This last inequality holds because in the worst case, the random walk goes ℓ steps away from v and back, where at each vertex there are $d - 1$ branches possible in the d-regular tree. Here, there are atmost $d^{2\ell}$ possible choices of paths. Finally, given an exact path, the number of ways in which the vertex could get to a vertex ℓ steps away is exactly C_ℓ, the ℓth Catalan number.

Since the trace of the matrix $\mathcal{L}$ is the sum of its eigenvalues, the 2ℓ'th transition matrix has sum of eigenvalues exactly $np_\ell(G)$. We can bound this from above by $1 + (n-1)\lambda^{2\ell}$. Then, using the fact that $C_\ell = \binom{2\ell}{\ell}/(\ell+1)$, we get that $\lambda(G) \geq 2\sqrt{d-1}/d - o(1)$. □

Lemma 1.3.4 (Expander Mixing Lemma): *Let $G = (V, E)$ with $|V| = n, |E| = m$ be a d-regular graph with spectral expansion constant λ. Then, for any subsets S, T of the vertex-set V, we have that:*

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \lambda \sqrt{|S||T|}$$

Proof. Let $\mathbf{1}_S$ be the indicator vector of the set S and $\mathbf{1}_T$ be the indicator vector of the set T, where:

$$(\mathbf{1}_S)_j = \begin{cases} 0, & j \notin S \\ 1, & j \in S \end{cases} \quad \text{and} \quad (\mathbf{1}_T)_j = \begin{cases} 0, & j \notin T \\ 1, & j \in T \end{cases}$$

Let $u = (1/n, \ldots, 1/n)$. Let $\mathbf{1}_S = |S|u + v_1, \mathbf{1}_T = |T|u + v_2$. Then, observe that $\langle v_1, u \rangle = \langle v_2, u \rangle = 0$. Also:

$$\|v_1\| = \sqrt{(n - |S|)/n^2 + |S|(1 + n^2 - 2/n)} = \sqrt{1/n + |S| - 2|S|/n} \leq \sqrt{|S|}$$

$$\|v_2\| = \sqrt{(n - |T|)/n^2 + |T|(1 + n^2 - 2/n)} = \sqrt{1/n + |T| - 2|T|/n} \leq \sqrt{|T|}$$

Note that the number of edges between S and T is exactly $(\mathbf{1}_S)^T A \mathbf{1}_T$. Then:

$$\begin{aligned} E(S,T) = (\mathbf{1}_S)^T A \mathbf{1}_T &= (|S|u^T + v_1^T) A (|T|u + v_2) \\ &= \frac{d|S||T|}{n} + d|T| v_1^T u + |S| u^T A v_2 + v_1^T A v_2 \\ &= \frac{d|S||T|}{n} + v_1^T A v_2 \end{aligned}$$

So, subtracting, we have that:

$$\begin{aligned} \left| E(S,T) - \frac{d|S||T|}{n} \right| &= v_1^T A v_2 \\ &\leq \|v_1\| \|A v_2\| \quad \text{(Cauchy Schwartz)} \\ &\leq \|v_1\| \lambda \|v_2\| \\ &\leq \lambda \sqrt{|S|} \sqrt{|T|} \end{aligned}$$

This yields the claim. □

Lemma 1.3.5 [Expander Hitting Lemma]. *Let $G = (V, E)$ with $|V| = n, |E| = m$ be a d-regular graph with spectral expansion constant λ. Then, for any $B \subseteq V$ such that $|B| = (1-\delta)n$, the probability that a random walk $X_1, X_2, \ldots, X_t$ of $t-1$ steps starting at a uniformly random vertex of G completely stays inside B is given by:*

$$\Pr[X_i \in B, \forall i \in [t]] \leq (1 - \delta(1-\lambda))^{t-1}$$

Proof adapted from [Gur20]. Let M be the random walk matrix of the graph G and let P be the projection matrix that zeroes out all the indices not in B such that $P_{ij} = \mathbb{1}_{\{i=j, i\in B\}}$. So, the probability that all the t vertices of a random walk are in B is equal to:

$$\Pr[V_i \in B, \forall i \in [t]] = \|PMPM \ldots PMPu\|_1$$

Then, by Cauchy Schwartz (and since $P^2 = P$), we write that:

$$\|(PMP)^{t-1}Pu\|_1 \leq \sqrt{n}\|(PMP)^{t-1}Pu\|$$

We then bound the largest absolute value of the eigenvalues of PMP. Let $\nu = \max_{\|x\|=1} x^T PMPx$. Let $y = Px$. We then have that $\nu = y^T My$, where $y = y^{\|} + y^{\perp}$ such that $y^{\|} = \alpha u$ and $\langle y^{\perp}, y^{\|}\rangle = 0$. Note that $\|y^{\|}\|^2 + \|y^{\perp}\|^2 = \|y\|^2 \leq \|x\|^2 = 1$. Then, $y^{\|} = \alpha u$, where $\alpha = \langle y, \mathbf{1}\rangle$. Thus, $\|y^{\|}\|^2 = \frac{1}{n}\langle y, \mathbf{1}\rangle^2 \leq (1-\delta)\|y\|^2$. So:

$$\begin{aligned}
\nu = y^T My &= ({y^{\|}}^T + {y^{\perp}}^T)M(y^{\|} + y^{\perp}) \\
&\leq (1-\lambda)\|y^{\|}\|^2 + \lambda\|y\|^2 \\
&\leq 1 - \delta(1-\lambda)
\end{aligned}$$

So, it must then be the case that:

$$\begin{aligned}
\|(PMP)^{t-1}Pu\|_1 \leq \sqrt{n}\|(PMP)^{t-1}Pu\| &\leq \sqrt{n}(1-\delta(1-\lambda))^{t-1}\|Pu\| \\
&\leq \sqrt{1-\delta}(1-\delta(1-\lambda))^{t-1} \\
&\leq (1-\delta(1-\lambda))^{t-1}
\end{aligned}$$

This proves the claim. □

Lemma 1.4.1 [Ta-Shma]. *Let S_0 be the parity of the number of times an expander random walk of length t hit B and S_1 be the parity of the number of times a uniformly random sequence of variables realized a value contained in B. Then, for $\epsilon_0 = 0.8, \beta = 0.01$, and small λ such that $\epsilon_0 + 2\beta + 2\lambda < 0.9$:*

$$\mathrm{TVD}(S_0, S_1) = (\epsilon_0 + 2\beta + 2\lambda)^{\lfloor t/2 \rfloor}$$

Proof adapted from [Ta-17]. Let $\mathcal{V}$ be a vector space with $\dim(\mathcal{V}) = |V| = n'$ and identify an element $v \in V$ with a basis vector $\vec{v} \in \mathcal{V}$. We will use this basis to define G, the linear operator of a random walk in G. Now, let Υ be a distribution over $\{0,1\}^k$. For a linear test $\alpha \in \{0,1\}^k$, let:

$$\mathrm{Bias}_\alpha(\Upsilon) = \left| \Pr_{s \in \Upsilon}(\langle \alpha, s \rangle = 0) - \Pr_{s \in \Upsilon}[\langle a, s \rangle = 1] \right|$$

Then, $\mathrm{Bias}(\Upsilon) = \max_{\alpha \neq \emptyset} \mathrm{Bias}_\alpha(\Upsilon)$. We say that Υ is ϵ-based if $\mathrm{Bias}(\Upsilon) \leq \epsilon$. Now, let $\alpha = (\alpha_1, \ldots, \alpha_k) \in \{0,1\}^k$ be a test-set that maximizes $\mathrm{Bias}_\alpha(\Upsilon)$ and let S_0, S_1 be the corresponding partitions of $[n']$, such that $S_b = \{v \in [n'] : \langle Z(v) | \alpha \rangle = b\}$. Then, define Π_0 and Π_1, where Π_b is the projection on the vector space of $\mathrm{span}(\{\vec{v} | v \in S_b\})$. We let $\Pi = \Pi_0 - \Pi_1$. Then, sample a random walk on an expander. Let $p_{even}(S_1)$ be the probability that a sampled path $(v_0, \ldots, v_t)$ visits S_1 an even number of times, and $p_{odd}(S_1)$ for an odd number of times. Then, observe the following:

1) $\mathrm{Bias}_\alpha(\Upsilon) = |\mathbb{E}_{v_0,\ldots,v_t}[(-1)^{\oplus_{j=0}^t \langle Z(v_j), \alpha \rangle}]| = |p_{even}(S_1) - p_{odd}(S_1)|$.

2) $p_{even}(S_1) - p_{odd}(S_1) = \sum_{b_0,\ldots,b_t \in \{0,1\}} (-1)^{b_0 + \cdots + b)t} \mathbf{1}^\dagger \Pi_{b_t} G \ldots \Pi_{b_2} G \Pi_{b_1} G \Pi_{b_0} \mathbf{1}$ because paths that fall an even number of times into S_1 contribute 1 while the other paths contribute -1. Then, by the distributive law, this must be identically equal to $\mathbf{1}^\dagger (\sum_{b_t \in \{0,1\}} (-1)^{b_t} \Pi_{b_t}) G \ldots (\sum_{b_0 \in \{0,1\}} (-1)^{b_0} \Pi_{b_0}) \mathbf{1} = \mathbf{1}^\dagger (\Pi G)^t \Pi \mathbf{1}$.

3) Let $v \in \mathcal{V}$ such that $\|v\| = 1$ and $v = v^\perp + v^\parallel$. Then since $G v^\parallel = v^\parallel = \|v^\parallel\| \mathbf{1}$:

$$\begin{aligned} \|\Pi G \Pi G v\| &\leq \|\Pi G \Pi G v^\perp\| + \|\Pi G \Pi G v^\parallel\| \leq \|\Pi G (\Pi \mathbf{1})^\parallel\| + \|\Pi G (\Pi \mathbf{1})^\perp\| + \|G v^\perp\| \\ &\leq \|(\Pi \mathbf{1})^\parallel\| + 2\lambda = \frac{||S_0| - |S_1||}{n} + 2\lambda \leq \epsilon_0 + 2\beta + 2\lambda \end{aligned}$$

The last inequality arises from letting Υ_0 be ϵ_0-biased and noting that $n - n' \leq \beta n$.
4) $|p_{even}(S_1) - p_{odd}(S_1)| = |\mathbf{1}^\dagger (\Pi G)^t \Pi \mathbf{1}| \leq \|(\Pi G)^t\| \leq \|(\Pi G)^2\|^{\lfloor t/2 \rfloor} \leq (\epsilon_0 + 2\beta + 2\lambda)^{\lfloor t/2 \rfloor}$. For large enough t, this term goes to 0, which is identical to the Chernoff result for the uniform random variables. This proves the claim. □

Lemma 17. *Orthogonality of the Krawtchouk functions:*

$$\langle K_r, K_s \rangle = \mathop{\mathbb{E}}_{b \sim \mathrm{Bin}(n, \frac{1}{2})} [K_r(b) K_s(b)] = \begin{cases} 0 & \text{if } r \neq s \\ \binom{n}{s} & \text{if } r = s \end{cases}$$

Proof. We provide a probabilistic interpretation of the Krawtchouk function as demonstrated in [Sam98]. Fix $A \in \binom{[n]}{\ell}$, $B \in_U \binom{[n]}{s}$, and choose $C \in_U \binom{[n]}{r}$. Then,

$$K_s(\ell) = \binom{n}{s} \mathbb{E}[(-1)^{|A \cap B|}], \quad K_r(\ell) = \binom{n}{r} \mathbb{E}[(-1)^{|A \cap C|}]$$

The inner product of K_r and K_s is then:

$$\begin{aligned} \langle K_r, K_s \rangle &= \binom{n}{r}\binom{n}{s} \mathbb{E}[(-1)^{|A \cap B|}] \, \mathbb{E}[(-1)^{|A \cap C|}] \\ &= \binom{n}{r}\binom{n}{s} \mathbb{E}[(-1)^{|A \cap B|}(-1)^{|A \cap C|}] \quad \text{(A, B, and C are independent)} \end{aligned}$$

If $B \neq C$, then $\mathbb{E}[(-1)^{|A \cap B|}(-1)^{|A \cap C|} | B \neq C] = 0$, since for each $x \in B \Delta C$, we could either have $x \in A$ or $x \notin A$, which contributes a +1 or -1 (or vice versa) (respectively) to the expectation. So, the expected value must be 0. Conversely, if $B = C$, then it must be the case that $r = s$, which occurs with probability $\frac{1}{\binom{n}{s}}$. So, $\mathbb{E}[(-1)^{|A \cap B|}(-1)^{|A \cap C|}] = \mathbb{E}[(-1)^{2|A \cap B|}] = \frac{1}{\binom{n}{s}}$. This yields that

$$\langle K_r, K_s \rangle = \begin{cases} 0 & \text{if } r \neq s \\ \binom{n}{s} & \text{if } r = s \end{cases}$$

□

Lemma 1.5.6 [Fooling Symmetric Functions]. *For all integers $t \geq 1$ and $p \geq 2$, let $G = (G_i)_{1 \leq i \leq t-1}$ be a sequence of λ-spectral expanders on a shared vertex set V with labeling $\mathrm{val} : V \to [p]$ that assigns each label $b \in [p]$ to f_b-fraction of the vertices. Then, for any label b, we have that the total variation distance between the number of b's seen in the expander random walk and the uniform distribution on $[p]$ has the following bound (where $[\Sigma(Z)_b]$ counts the number of occurrences of b in Z) is:*

$$\mathrm{TVD}([\Sigma(\mathrm{RW}^t_{\mathcal{G}})]_b, [\Sigma(U[d])]_b) \leq O\left(\left(\frac{p}{\min_{b\in[p]} f_b}\right)^{O(p)} \cdot \lambda\right)$$

Proof adapted from [GV22]. Consider the case $p_0 p_1 t < 1$. Then, for $j \in \mathbb{Z}_t$, we have that:

$$|g_j| \leq 44 \cdot \frac{\|G_u - G'_u\|}{t} \cdot e^{2r^2 - sr(j - p_1 t)}$$

This expression is minimized when setting $r = \frac{1}{2}$ and $s = \mathrm{sgn}(j - p_1 t)$. Summing over all $j : |j - p_1 t| \geq c$ gives:

$$\sum_{j \in \mathbb{Z}_t : |j - p_1 t| \geq c} |g_j| \leq 88 \cdot \frac{\|G'_u - G_u\|}{t} \cdot \frac{e^{1/2 - c/2}}{1 - e^{-1/2}} \leq 4000 e^{-c^2/8t} \cdot \frac{\|G'_u - G_u\|}{t}$$

A similar method shows an equivalent result for $p_0 p_1 t \geq 1$, which thus proves lemma 1.5.7 (theorem 18).

The proof of lemma 1.5.8 (Theorem 20), in turn, gets to the heart of the proof methodologies used in the paper, since it involves a Fourier-analytic component. We will fulfil this by providing a high-level overview of the (auxiliary) proof of lemma 1.5.8. To do this, we re-tell [GV22]'s description of the Fourier group on $\mathbb{Z}$. Let $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ with ℓ^2 norm $\|f\| = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 \mathrm{d}\theta/2\pi}$ and use [Ahm+19]'s formulation of the singular-value approximation.

Let $\ell^2(\mathbb{Z})$ and $\ell^2(S^1)$ represent the subspaces of $\mathbb{C}^{\mathbb{Z}}$ and $\mathbb{C}^{S^1}$ (resp.) that contain all elements of finite ℓ^2 norm. Then, the Fourier transform of $\mathbb{Z}$ is the map $\mathcal{F} : \ell^2(\mathbb{Z}) \to \ell^2(S^1)$ such that the Fourier transform of $h \in \ell^2(\mathbb{Z})$, denoted $\mathcal{F}h = \hat{h} \in \ell^2(S^1)$, is given by $\hat{h}(\theta) = \sum_{j \in \mathbb{Z}} h_j e^{-i\theta j}$. It can also be expressed in terms of the Fourier characters $\chi_\theta = (e^{i\theta j})_{j \in \mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$ as $\hat{h}(\theta) = \chi^*_\theta h$. The central idea for lemma 1.5.8 is that the ℓ_2 norm is preserved under the Fourier transform, and so, $\|g^{(sr)}\| = \|\hat{g}^{(sr)}\|$. The last component needed for this proof is theorem 25 from [GV22] which is a

linear-algebraic result that bounds $|\hat{g}^{(sr)}|$ as follows:

$$|\hat{g}^{(rs}| \leq 4 \cdot \|G'_u - G_u\| \cdot p_0 p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right) \cdot e^{p_0 p_1 t(2r^2 - \theta^2/20)}$$

Then, to prove lemma 1.5.8, [GV22] uses the result from theorem 25 above.

$$\|g^{(sr)}\| = \|\hat{g}^{(sr)}\| = \sqrt{\int_{-\infty}^{\infty} |\hat{g}^{(sr)}(\theta)|^2 \frac{\mathrm{d}\theta}{2\pi}}$$

$$\leq 4\sqrt{2}\|G'_u - G_u\| p_0 p_1 e^{2p_0 p_1 t r^2} \left(4r^2 \sqrt{\int_{-\pi}^{\pi} e^{-p_0 p_1 t\theta^2/10} \frac{\mathrm{d}\theta}{2\pi}} + \sqrt{\int_{-\pi}^{\pi} \frac{9}{4}\theta^4 e^{-p_0 p_1 t\theta^2/10} \frac{\mathrm{d}\theta}{2\pi}}\right)$$

From here, bounding the inequality further with different standard results yields the result for 1.5.8. The big-picture idea of this proof is the notion that the ℓ_2 norm is preserved under the Fourier transform which relates the quantity to the ℓ_1 norm which can also, in turn, be bounded. □

Lemma 2.1.1. [Krawtchouk coefficient of the probability ratio] *Expanding $q(\ell)$ through the Krawtchouk function expansion in Proposition 3.2 yields that:*

$$\hat{q}(k) = \frac{1}{\binom{n}{k}(p-1)^{n-k}} \mathop{\mathbb{E}}_{s \sim S(n,p,\lambda)} [K_k(|s|_0)]$$

Proof. Writing the expected value of $K_k(|s|_0)$ using $q(b)$ and $K_k(b)$, we get:

$$\begin{aligned}
\hat{q}(k) &= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \sum_{b=0}^{n} \binom{n}{b} \frac{(p-1)^{n-b}}{p^n} q(b) K_k(b) \\
&= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \sum_{b=0}^{n} \Pr_{s \sim S(n,p,\lambda)} [|s|_0 = b] K_k(b) \quad \text{(substituting } q(b)\text{)} \\
&= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \mathop{\mathbb{E}}_{s \sim S(n,p,\lambda)} [K_k(|s|_0)] \quad \text{(by definition of } \mathop{\mathbb{E}}_{s \in S(n,p,\lambda)} [K_k(|s|_0)]\text{)}
\end{aligned}$$

□

Lemma 2.1.2. *For $s \in S(n, p, \lambda)$, we have that*

$$\Pr[|s|_0 = \ell] = \frac{1}{p^n} \sum_{k=0}^{n} K_\ell(k) \mathop{\mathbb{E}}_{s \sim S(n,p,\lambda)} [K_k(|s|_0)]$$

Proof. Writing out $\Pr[|s|_0 = \ell]$ in terms of the probability ratio $q(\ell)$, we get:

$$\begin{aligned}
\Pr[|s|_0 = \ell] &= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} q(\ell) \\
&= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} \sum_{k=0}^{n} \hat{q}(k) K_k(\ell) \quad \text{(Krawtchouk expansion of } q(\ell)) \\
&= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} \sum_{k=0}^{n} \frac{K_k(\ell)}{\binom{n}{k}(p-1)^{n-k}} \mathbb{E}[K_k(|s|_0)] \quad \text{(Lemma 3.1)} \\
&= \frac{1}{p^n} \sum_{k=0}^{n} \frac{\binom{n}{\ell}}{\binom{n}{k}} \frac{(p-1)^{n-\ell}}{(p-1)^{n-k}} \mathbb{E}[K_k(|s|_0)] K_k(\ell) \\
&= \frac{1}{p^n} \sum_{k=0}^{n} K_\ell(k) \mathop{\mathbb{E}}_{s \sim S(n,p,\lambda)} [K_k(|s|_0)] \quad \text{(By the reciprocity relation)}
\end{aligned}$$

□

Lemma 2.2.1 . *The expected value of the Krawtchouk function is given by:*

$$\mathbb{E}[K_k(|s|_0)] = \begin{cases} (p-1)^{n-k} \sum\limits_{d=k}^{n-k} \phi_0(d) \lambda^d, & \text{if } c = k \mod p \equiv 0 \\ 0, & \text{if } c = k \mod p \not\equiv 0 \end{cases}$$

Proof. By the formula of expected values, we have that:

$$\begin{aligned}
\mathbb{E}[K_k(|s|_0)] &= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_2^n \\ |y|_0 = k}} (-1)^{y \cdot s} \\
&= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_2^n \\ |y|_0 = k}} (-1)^{\sum_{i=1}^{n} y_i \cdot s_i} \\
&= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_2^n \\ |y|_0 = k}} \prod_{i=1}^{n} (-1)^{y_i \cdot s_i}
\end{aligned}$$

We note then that the dot-product on the exponent of (-1) only takes the summation of the element-wise product of α and y for positions on y that are strictly non-zero. Therefore, we can rewrite the summation by considering the indices corresponding to locations of non-zeros in y, and instead take the summation of the dot-product along these indices. So, for $T = \{a_1 < \ldots < a_{n-k}\}$, we have that:

$$\mathbb{E}[K_k(|s|_0)] = \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{T \in \binom{[n]}{n-k}} \prod_{i \in T} (-1)^{s_i}$$

Further, choosing a $T \in \binom{[n]}{n-k}$ implies a choice of $\overline{T} = \binom{[n]}{k} = [n] \setminus T$. Hence, the summation reduces to:

$$\begin{aligned}\mathbb{E}[K_k(|s|_0)] &= \sum_{s\sim S(n,p,\lambda)} \Pr[s] \sum_{T\in\binom{[n]}{k}} \prod_{i\in\overline{T}} (-1)^{s_i} \\ &= \sum_{T\in\binom{[n]}{k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)} \left[\prod_{i\in\overline{T}} (-1)^{s_i}\right] \quad \text{(definition of expectations)}\end{aligned}$$

Next, observe that the sticky random walk is a Markov chain where $(-1)^{s_i} = (-1)^{s_{i-1}}$ with probability $1/p + (p-1)\lambda)$. So, we can instead model the transitions of strings from the sticky random walk as random variables u, where u_1 is uniformly distributed in $\mathbb{Z}_p$ and for $i \geq 2$, u_i is uniformly distributed on $(1-\lambda)U[\mathbb{Z}_p] + \lambda \cdot \mathbb{1}_0$. To provide an intuition for this refactorization, $(1-\lambda)U[\mathbb{Z}_p]$ is the 'base' probability of switching to any vertex and λ is the additional probability of staying on the same vertex.

Then, since $s_i = \sum_{i\in T} \sum_{j=1}^{i} u_j$, we write that:

$$\begin{aligned}\mathop{\mathbb{E}}_{s\in S(n,p,\lambda)} \left[\prod_{i\in\overline{T}} (-1)^{s_i}\right] &= \mathop{\mathbb{E}}_{s\in S(n,p,\lambda)} \left[(-1)^{\sum_{i\in\overline{T}} \sum_{j=1}^{i} u_j}\right] \\ &= \prod_{j=1}^{a_{n-k}} \mathop{\mathbb{E}}_{s\in S(n,p,\lambda)} \left[(-1)^{\sum_{i\in\overline{T}; i\geq j} u_j}\right] \quad \text{(independence of } u_j\text{'s)}\end{aligned}$$

When $j = 1$, we get that:

$$\begin{aligned}\mathbb{E}\left[(-1)^{\sum_{i\in\overline{T}; i\geq 1} u_1}\right] &= \mathbb{E}[(-1)^{|\overline{T}|u_1}] \\ &= \begin{cases} 1, & \text{if } |\overline{T}| \mod p \equiv 0 \\ 0, & \text{otherwise} \end{cases}\end{aligned}$$

Conversely, when $j \geq 2$, let $T_j = \{i \in \overline{T}; i \geq j\}$. Then,

$$\begin{aligned}\mathbb{E}\left[(-1)^{\sum_{i\in\overline{T}; i\geq j} u_j}\right] &= \mathbb{E}[(-1)^{|T_j|u_j}] \\ &= \begin{cases} 1, & \text{if } |T_j| \mod p \equiv 0 \\ \mathbb{E}[(-1)^{u_j}], & \text{otherwise} \end{cases}\end{aligned}$$

Next, observe that for $j \geq 2$, $\mathbb{E}[(-1)^{u_j}] = \lambda$.

Proof. To show this, we write the expression for u_j, $j \geq 2$ in the exponent and take the expected value of $(-1)^{u_j}$.

$$\begin{aligned}
\mathbb{E}[(-1)^{u_j}] &= \mathbb{E}[(-1)^{(1-\lambda)U[\mathbb{Z}_p]+\lambda\cdot\mathbb{1}_0}] \quad \text{(since } u_j \sim (1-\lambda)U[\mathbb{Z}_p] + \lambda\cdot\mathbb{1}_0\text{)} \\
&= \sum_{k=0}^{p-1}(-1)^k \Pr[u_j = k] \quad \text{(definition of expectation)} \\
&= (-1)^0 \Pr[u_j = 0] + (-1)^1 \Pr[u_j = 1] + \ldots + (-1)^{p-1}\Pr[u_j = p-1] \\
&= \left(\frac{1}{p} + \lambda\left(\frac{p-1}{p}\right)\right) - \left(\frac{1}{p} - \frac{\lambda}{p}\right) + \ldots + (-1)^{p-1}\left(\frac{1}{p} - \frac{\lambda}{p}\right) \\
&= \frac{1}{p}\sum_{k=0}^{p-1}(-1)^k - \frac{\lambda}{p}\sum_{k=0}^{p-1}(-1)^k + \lambda(-1)^0 = \lambda
\end{aligned}$$

□

Then, for $k \mod p \equiv 0$, we have that:

$$\begin{aligned}
\mathbb{E}[K_k(|s|_0)] &= \sum_{T\in\binom{[n]}{k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}\left[\prod_{i\in\overline{T}}(-1)^{s_i}\right] \\
&= \sum_{T\in\binom{[n]}{k}} \prod_{i\in\overline{T}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[(-1)^{s_i}] \quad \text{(independence of } (-1)^{s_i}\text{)} \\
&= \sum_{T\in\binom{[n]}{k}} \prod_{j=1}^{a_{n-k}} \lambda \quad \text{(since } |T| = a_{n-k}\text{)} \\
&= \sum_{T\in\binom{[n]}{k}} \lambda^{a_{n-k}}
\end{aligned}$$

We then parameterize the summation over every possible value of the shift of T (for $k \mod p \equiv 0$), where the shift function is given in Definition 22.

$$\begin{aligned}
\mathbb{E}[K_k(|s|_0)] &= \sum_{d=k}^{n-k}\left(\sum_{\substack{T\in\binom{[n]}{k} \\ \text{shift}_0(\text{T})=\text{d}}} 1\right)\lambda^d \\
&= \sum_{d=k}^{n-k}\phi_0(d)\lambda^d
\end{aligned}$$

This yields the claim. □

Lemma 2.2.2. *For $c \in \mathbb{N}$ where $0 \leq c \leq p$, and for $d \in \mathbb{N}$ where $k \leq d \leq n-k$, the number of k-sized subsets of $[n]$ that satisfy $\text{shift}_c(T) = d$* is:

$$\phi_c(d) = \frac{1}{p^k} \sum_{\substack{T \in \binom{[n]}{k} \\ \text{shift}_c(T)=d}} 1 = \frac{1}{p^k} \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

Proof. To determine $\phi_c(d)$, we count the total number of ways to choose $a_1 < a_2 < ... < a_k$ such that the lengths of the intervals $(a_c - a_{c-1}) + (a_{c+p} - a_{c+p-1}) + (a_{c+2p} - a_{c+2p-1}) + ... = d$, where for any $j \leq 0$, $a_j = 0$. To do this, we combine each element-wise interval (a_{c+ip}, a_{c+ip-1}) to form a contiguous interval of length d (starting from $a_{c-1} = 0$). The remaining contiguous region that excludes these intervals must then have a length of $n - d$. We then abstract the number of ways to count $a_1 < ... < a_k$ by counting the number of intervals that have a length of d when combined, such that the remaining intervals have a length $n - d$. From a length of $d - 1$ (accounting for $a_0 = 0$), we need to select intervals that form a length of $\lfloor |k - c|/(p - 1) \rfloor - 1$ since they represent the number of choices of elements of T that are index-separated by p. Similarly, from a length of $n - d$, we need to select intervals that form a length of $\lfloor |k - c|/(p - 1) \rfloor$ possible intervals, since they represent every other element of T. This second constraint is to ensure that the total length of the intervals chosen is exactly n. Finally, we divide by the maximum number of repetitions to prevent duplicates, which is p^k. Hence, we write that:

$$\sum_{\substack{T \in \binom{[n]}{k} \\ \text{shift}_c(T)=d}} 1 = \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

Therefore,

$$\phi_c(d) = \frac{1}{p^k} \sum_{\substack{T \in \binom{[n]}{k} \\ \text{shift}_c(T)=d}} 1 = \frac{1}{p^k} \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

□

Lemma 2.3.1. *The total variational distance between the generalized sticky random walk on p vertices and the uniform distribution on p states is given by:*

$$\text{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\text{U}_p^n)]_0) = \frac{1}{2} \mathop{\mathbb{E}}_{b \sim \text{U}_p^n} [|q(b) - 1|]$$

Proof. We write the expression of the total variation distance between the n-step sticky random walk on p states and n-samples from the uniform distribution on p states. This then yields:

$$\begin{aligned}
\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) &= \frac{1}{2}\sum_{\ell=0}^{n}\left|\Pr[|s|_0 = \ell] - \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n}\right| \\
&= \frac{1}{2}\sum_{\ell=0}^{n}\left|\binom{n}{\ell}q(\ell)\frac{(p-1)^{n-\ell}}{p^n} - \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n}\right| \\
&= \frac{1}{2}\sum_{\ell=0}^{n}\left|\frac{\binom{n}{\ell}}{p^n}(p-1)^{n-\ell}(q(\ell)-1)\right| \\
&= \frac{1}{2}\sum_{\ell=0}^{n}|\Pr[\ell](q(\ell)-1)| \\
&= \frac{1}{2}\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[|q(b)-1|]
\end{aligned}$$

□

Lemma A.0.1. For $1 \le k \le \frac{n}{p}$, we can bound $\frac{\binom{n}{k}^2}{\binom{n}{pk}} \le (\frac{ne}{k})^{2k} \cdot (\frac{pk}{n})^{pk}$.

Proof. We use the bound that $(\frac{n}{k})^k \le \binom{n}{k} < (\frac{ne}{k})^k$. This gives us that $\frac{\binom{n}{k}^2}{\binom{n}{pk}} \le \frac{(\frac{ne}{k})^{2k}}{(\frac{n}{pk})^{pk}}$. Simplifying it yields that $(\frac{ne}{k})^{2k} \cdot (\frac{pk}{n})^{pk}$, which proves the claim. □

Appendix B

PROOF OF THEOREMS IN CHAPTER 2

In this section of the Appendix, we present a generalization of the Krawtchouk polynomials into the complex domain that yields an elegant method of analysis for bounding the total variation distance of $[\Sigma(S(n,p,\lambda))]_0$ and $[\Sigma(\mathrm{U}_n^p)]_0$, but only for when p is a prime number. For the generalized sticky random walk, this method yields an upper-bound on $\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_n^p)]_0)$ of $O(\lambda p^{O(p)})$, which matches the upper-bound derived in Corollary 4 of [GV22] through Fourier-analytic means, which the body of our paper shows to be suboptimal when the size of the alphabet used is allow to increase asymptotically. Nonetheless, we include a proof of this generalization to present our novel treatment of the Krawtchouk function.

Definition 23. Given any set $S_k \subseteq [n]$ with cardinality $|S_k| = k$, let v_{S_k} denote a bit-string in $\{0,1\}^n$ such that for each $i \in S_k$, $(v_{S_k})_i = 0$ and for each $i \in [n] \setminus S_k$, $(v_{S_k})_i = 1$. Similarly, given the context of a prior prime number p, let w_{S_k} denote a string in $\mathbb{Z}_p^n$ where for each $i \in S_k$, $(w_{S_k})_i = 0$, and for each $i \in [n] \setminus S_k$, $(w_{S_k})_i \in [p-1]$.

Definition 24. Let D denote the "zero distribution" where for any $\ell \in [n]$, $\Pr[D = \ell]$ denotes the probability that a string $s \in \mathbb{Z}_p^n$ has ℓ zeros. Specifically, $\Pr[D = \ell] = \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n}$, since there are $\binom{n}{\ell}$ ways to select the locations for the ℓ 0s in a string of length n, and $(p-1)^{n-\ell}$ to populate the other $n-k$ locations with characters from $[p-1]$.

Definition 25. For $p \geq 2$, let ω_p denote the p^{th} primitive root of unity if it satisfies $(\omega_p)^p = 1$, and if there does not exist $q \in \mathbb{N}$ where $q < p$ such that $(\omega_p)^q = 1$. Specifically, the multiplicative order of the p^{th} primitive root of unity must be p. Then, for $1 \leq k < p$, we must have that

$$\sum_{j=0}^{p-1} (\omega_p)^{kj} = \omega_p^{k\cdot 0} + \omega_p^{k\cdot 1} + \ldots + \omega_p^{k\cdot(p-1)} = 0$$

Proof. Given $\omega_p^p = 1$, it is clear that for $k < p$, $(\omega_p^k)^p = 1$. Thus, $(\omega_p^k)p - 1 = 0 = (\omega_p^k - 1)(\omega_p^{k\cdot 0} + \omega_p^{k\cdot 1} + \ldots + \omega_p^{k\cdot(p-1)})$. Since $\omega_p^k \neq 1$ as ω_p is a primitive root of unity, we must have that $\omega_p^{k\cdot 0} + \omega_p^{k\cdot 1} + \ldots + \omega_p^{k\cdot(p-1)} = 0$, which proves the claim. □

Definition 26. For all $p \in \mathbb{N}$, consider ω_p, the p-th principal root of unity. Then, the generalized Krawtchouk function $K_k(\ell)$, for any α where α has ℓ 0s and $n-\ell$ 1s, is defined as:

$$K_k(\ell) = \sum_{\substack{y \in \mathbb{Z}_p^n \\ |y|_0 = k}} (\omega_p)^{\alpha \cdot y}$$

Lemma B.0.1. Orthogonality of the generalized Krawtchouk function: The generalized Krawtchouk functions form an orthogonal basis of the functions mapping $\mathbb{Z}_{n+1} \to \mathbb{R}$ (for the distribution D as described in Definition 24 with respect to the inner product $\langle f, g \rangle = \mathop{\mathbb{E}}_{b \sim D}[f(b)g(b)]$.

$$\langle K_r, K_s \rangle = \begin{cases} 0, & \text{if } r \neq s \\ \binom{n}{r}(p-1)^{n-r}, & \text{if } r = s \end{cases}$$

Proof. In a similar vein to the proof of lemma 1.5.1, we start by providing a probabilistic interpretation of the generalized Krawtchouk function. Fix $A \in \binom{[n]}{\ell}$, and choose $B \in_U \binom{[n]}{r}$ and $C \in_U \binom{[n]}{s}$. This is equivalent to fixing a string $v_{A_\ell} \in \mathbb{Z}_p^n$ where $|v_{A_\ell}|_0 = \ell$ and $(v_{A_\ell})_t = 0$ for all $t \in A$, and randomly choosing $w_{B_r}, w_{C_s} \in \mathbb{Z}_p^n$ where $|w_{B_r}|_0 = r$ and $|w_{C_s}|_0 = s$. Then,

$$K_s(\ell) = \binom{n}{s}(p-1)^{n-s}\mathop{\mathbb{E}}_{B_r}[\omega_p^{\langle v_{A_\ell}, w_{B_r}\rangle}], \quad K_r(\ell) = \binom{n}{r}(p-1)^{n-r}\mathop{\mathbb{E}}_{C_s}[\omega_p^{\langle v_{A_\ell}, w_{C_s}\rangle}]$$

The inner product of K_r and K_s (with respect to the probability distribution D) is then:

$$\begin{aligned}
\langle K_r, K_s \rangle &= \sum_{\ell=0}^{n} \Pr[\ell] K_r(\ell)\overline{K_s(\ell)} \\
&= \sum_{\ell=0}^{n} \binom{n}{\ell}\binom{n}{r}\binom{n}{s}\frac{(p-1)^{n-\ell}}{p^n}(p-1)^{2n-s-r}\mathop{\mathbb{E}}_{B_r}[\omega_p^{\langle v_{A_\ell}, w_{B_r}\rangle}]\mathop{\mathbb{E}}_{C_s}[\overline{\omega}_p^{\langle v_{A_\ell}, w_{C_s}\rangle}] \\
&= \binom{n}{r}\binom{n}{s}\frac{(p-1)^{2n-s-r}}{p^n}\sum_{\ell=0}^{n}\binom{n}{\ell}(p-1)^{n-\ell}\mathop{\mathbb{E}}_{B_r, C_s}[\omega_p^{\langle v_{A_\ell}, w_{B_r}\rangle - \langle v_{A_\ell}, w_{C_s}\rangle}] \\
&= \binom{n}{r}\binom{n}{s}\frac{(p-1)^{2n-s-r}}{p^n}\sum_{\ell=0}^{n}\binom{n}{\ell}(p-1)^{n-\ell}\mathop{\mathbb{E}}_{B_r, C_s}[\omega_p^{\langle v_{A_\ell}, w_{B_r} - w_{C_s}\rangle \mod p}]
\end{aligned}$$

The second-last line follows by the independence of A, B, C, and the last line in the derivation follows by the bilinearity of the inner product. If $r \neq s$, then $B_r \neq C_s$ and $\mathbb{E}[\omega_p^{\langle v_{A_\ell}, w_{B_r} - w_{C_s}\rangle \mod p}] = \mathbb{E}[\omega_p^{\langle v_{A_\ell}, w_{B_r}\rangle \mod p}]$ since the distribution of $B - C$ is uniformly random (as are the distributions of B and C). Therefore, the distribution

of $B - C$ must be indistinguishable from the distribution of B under the expectation operator. Additionally, $\mathbb{E}[\omega_p^{\langle A,B\rangle \mod p}] = \sum_{j=0}^{p} \omega_p^{kj} = 0$. If $B = C$, (which is only when $r = s$), the associated probability of it occurring must be $\frac{1}{\binom{n}{r}(p-1)^{n-r}}$.

In this case, $\langle K_r, K_s\rangle = \binom{n}{s}\binom{n}{r}(p-1)^{2n-r-s}\,\mathbb{E}[\omega^{\langle A,0\rangle}] = \binom{n}{s}(p-1)^{n-r}$.

$$\mathop{\mathbb{E}}_{b\sim D}[K_r(b)K_s(b)] = \begin{cases} 0, & \text{if } r \neq s \\ \binom{n}{r}(p-1)^{n-r}, & \text{if } r = s \end{cases}$$

□

Lemma B.0.2. The generalized Krawtchouk function $K_k(\ell)$ is invariant against choices of $\alpha \in \{0,1\}^n$, where $|\alpha|_0 = \ell$. Specifically, we have that for a fixed $B \in \mathbb{Z}_p^n$ where $|B|_0 = k$, that

$$\mathop{\mathbb{E}}_{\substack{\text{Fixed } A\in\mathbb{Z}_2^n \\ |A|_0=\ell}}[\omega_p^{\langle A,B\rangle}] = \mathop{\mathbb{E}}_{\substack{\text{Random } A'\in\mathbb{Z}_2^n \\ |A'|_0=\ell}}[\omega_p^{\langle A',B\rangle}]$$

Proof. The problem is equivalent to showing that

$$\omega_p^{\langle A,B\rangle} \cdot \frac{1}{\mathop{\mathbb{E}}_{A'}[\omega_p^{\langle A',B\rangle}]} = \omega_p^{\langle A,B\rangle} \cdot \mathop{\mathbb{E}}_{A'}\left[\frac{1}{\omega_p^{\langle A',B\rangle}}\right] \stackrel{?}{=} 1$$

By the independence of A, A', and B, the problem reduces to showing that

$$\mathop{\mathbb{E}}_{A'}\left[\frac{\omega_p^{\langle A,B\rangle}}{\omega_p^{\langle A',B\rangle}}\right] = \mathop{\mathbb{E}}_{A'}[\omega_p^{\langle A-A',B\rangle}] \stackrel{?}{=} 1$$

Since A' is uniformly random in $\mathbb{Z}_2^n$, $A - A'$ must also be uniformly random for a fixed A. So, the problem reduces further to showing (under the same conditions of B and A') that $\mathbb{E}[\omega_p^{\langle A',B\rangle}] \stackrel{?}{=} 1$. Since $\sum_{j=0}^{p} \omega_p^{kj} = 0$ from definition 24, we must have that $\sum_{j=0}^{p-1} \omega_p^{kj} = 1$. Thus, $\mathbb{E}[\omega_p^{\langle A',B\rangle}] = \sum_{j=0}^{p-1} \omega_p^{kj} = 1$, which proves the claim. □

Corollary B.0.2.1. The orthogonality of the generalized Krawtchouk function implies a reciprocity relation:

$$\frac{K_k(\ell)}{\binom{n}{k}(p-1)^{n-k}} = \frac{K_s(\ell)}{\binom{n}{s}(p-1)^{n-s}}$$

We now provide a brief calculation for the expectation of the generalized Krawtchouk function, since it is necessary for bounding the total variational distance between $[\Sigma(S(n,p,\lambda))]_0$ and $[\Sigma(U_p^n)]_0$.

Definition 27. Given any set $T \subseteq [n]$ such that $|T| = k$, let $a_1 < ... < a_k$ be the elements of T in increasing order. Then, for any $c \in \mathbb{Z}_p$, let

$$\text{shift}_c(T) = \sum_{i=0}^{\lfloor |k-c|/p \rfloor} (a_{c+ip} - a_{c+ip-1})$$

Then, for any c such that $k \mod p = -c$, and for any $d \in \mathbb{Z}_{n+1}$, let $\phi_c(d)$ denote the number of subsets of $[n]$ of size k such that $\text{shift}_c(T) = d$. Note that for any $t \leq 0, a_t = 0$.

Lemma B.0.3. The expectation of the Krawtchouk function is*:*

$$\mathbb{E}[K_k(|s|_0)] = \begin{cases} (p-1)^{n-k} \sum_{d=k}^{n-k} \phi_0(d)\lambda^d, & c = k \mod p \equiv 0 \\ 0, & c = k \mod p \not\equiv 0 \end{cases}$$

Proof.

$$\begin{aligned} \mathbb{E}[K_k(|s|_0)] &= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_p^n \\ |y|_0 = k}} \omega_p^{y \cdot s} \\ &= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_p^n \\ |y|_0 = k}} \omega_p^{\sum_{i=1}^{n} y_i \cdot s_i} \\ &= \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{\substack{y \in \mathbb{Z}_p^n \\ |y|_0 = k}} \prod_{i=1}^{n} \omega_p^{y_i \cdot s_i} \end{aligned}$$

The dot-product on the exponent of ω only takes the summation of the element-wise product of α and y for positions on y that are non-zero. We can rewrite this summation by considering the location of non-zero terms in y. So, for $T = \{a_1 < ... < a_{n-k}\}$:

$$\mathbb{E}[K_k(|s|_0)] = \sum_{s \sim S(n,p,\lambda)} \Pr[s] \sum_{T \in \binom{[n]}{n-k}} \sum_{\beta \in [p-1]^{n-k}} \prod_{i \in T} \omega_p^{\beta_i s_i}$$

Further, choosing a $T \in \binom{[n]}{n-k}$ implies a choice of $\overline{T} = \binom{[n]}{k} = [n] \setminus T$. Hence, the summation reduces to:

$$\begin{aligned}\mathbb{E}[K_k(|s|_0)] &= \sum_{s\sim S(n,p,\lambda)} \Pr[s] \sum_{T\in\binom{[n]}{k}} \sum_{\beta\in[p-1]^{n-k}} \prod_{i\in\overline{T}} \omega_p^{\beta_i s_i} \\ &= \sum_{T\in\binom{[n]}{k}} \sum_{\beta\in[p-1]^{n-k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)} \left[\prod_{i\in\overline{T}} \omega_p^{\beta_i s_i} \right]\end{aligned}$$

Next, observe that the sticky random walk is a Markov chain where $\omega_p^{s_i} = \omega_p^{s_{i-1}}$ with probability $\frac{1}{p} + (p-1)\lambda$. We can instead model the transitions of strings from the sticky random walk as random variables u, where u_1 is uniformly distributed in $\mathbb{Z}_p$ and for $i \geq 2$, u_i is uniformly distributed on $(1-\lambda)U[\mathbb{Z}_p] + \lambda \cdot \mathbb{1}_0$. Intuitively, λ is the additional probability of staying on the same vertex. So, for each $s \sim S$, we refactor s to $\tilde{s} = \{\tilde{s}_1, ..., \tilde{s}_n\}$, where $\tilde{s}_i = \mathbb{1}\{s_i \neq 0\}$. Note then that since β_i is uniformly random in $[p-1]$, that $\beta_i s_i \mod p$, and therefore $\beta_i \tilde{s}_i \mod p$ must also be uniformly random in $[p-1]$. Therefore, $\beta_i \tilde{s}_i$ and $\beta_i s_i$ are both uniformly random in $[p-1]$ and have the same distributions. Hence, we write that:

$$\begin{aligned}\mathop{\mathbb{E}}_{\beta,s} \left[\prod_{i\in\overline{T}} \omega_p^{\beta_i s_i} \right] &= \mathop{\mathbb{E}}_{\beta,s} \left[\prod_{i\in\overline{T}} \omega_p^{\beta_i \tilde{s}_i} \right] \\ &= \mathop{\mathbb{E}}_{\beta,s} \left[\omega_p^{\sum_{i\in\overline{T}} \beta_i \sum_{j=1}^{i} u_j} \right] \\ &= \prod_{j=1}^{a_{n-k}} \mathop{\mathbb{E}}_{\beta,s} \left[\omega_p^{\sum_{i\in\overline{T};i\geq j} \beta_i u_j} \right]\end{aligned}$$

Since u_j is random, the distribution of $\beta_i u_j \mod p$ must also be random, and therefore indistinguishable from the distribution of u_j. Therefore, $\beta_i u_j \mod p$ and $u_j \mod p$ are invariant under the expectation of its exponentiation under ω_p. So:

$$\mathop{\mathbb{E}}_{\beta,s} \left[\prod_{i\in\overline{T}} \omega_p^{\beta_i s_i} \right] = \prod_{j=1}^{a_{n-k}} \mathbb{E} \left[\omega_p^{\sum_{i\in\overline{T};i\geq j} u_j} \right]$$

When $j = 1$, the above definition directly implies that:

$$\mathbb{E} \left[\omega_p^{\sum_{i\in\overline{T};i\geq 1} u_1} \right] = \mathbb{E}[\omega_p^{|\overline{T}|u_1}] = \begin{cases} 1, & \text{if } |\overline{T}| \mod p \equiv 0 \\ 0, & \text{otherwise} \end{cases}$$

Conversely, when $j \geq 2$, let $T_j = \{i \in \overline{T}; i \geq j\}$. Then,

$$\mathbb{E}\left[\omega_p^{\sum_{i\in\overline{T};i\geq j} u_j}\right] = \mathbb{E}[\omega_p^{|T_j|u_j}] = \begin{cases} 1, & \text{if } |T_j| \mod p \equiv 0 \\ \mathbb{E}[\omega_p^{u_j}], & \text{otherwise} \end{cases}$$

Next, observe that for $j \geq 2$, $\mathbb{E}[\omega_p^{u_j}] = \lambda$.

Proof. We write u_j in terms of our refactoring and take the expectation:

$$\begin{aligned}
\mathbb{E}[\omega_p^{u_j}] = \mathbb{E}[\omega_p^{(1-\lambda)U[\mathbb{Z}_p]+\lambda\cdot\mathbb{1}_0}] &= \sum_{k=0}^{p-1} \omega_p^k \Pr[u_j = k] \\
&= \omega_p^0 \Pr[u_j = 0] + \omega_p^1 \Pr[u_j = 1] + \ldots + \omega_p^{p-1} \Pr[u_j = p-1] \\
&= \omega_p^0\left(\frac{1}{p} + \lambda\left(\frac{p-1}{p}\right)\right) + \omega_p^1\left(\frac{1}{p} - \frac{\lambda}{p}\right) + \ldots + \omega_p^{p-1}\left(\frac{1}{p} - \frac{\lambda}{p}\right) \\
&= \frac{1}{p}\sum_{k=0}^{p-1} \omega_p^k - \frac{\lambda}{p}\sum_{k=0}^{p-1} \omega_p^k + \lambda\omega_p^0 \\
&= \lambda
\end{aligned}$$

□

Then, for $k \mod p \equiv 0$, we have that

$$\begin{aligned}
\mathbb{E}[K_k(|s|_0)] &= \sum_{T\in\binom{[n]}{k}} \sum_{\beta\in[p-1]^{n-k}} \mathop{\mathbb{E}}_{s\sim S}\left[\prod_{i\in\overline{T}} \omega_p^{\beta_i s_i}\right] \\
&= (p-1)^{n-k} \sum_{T\in\binom{[n]}{k}} \mathop{\mathbb{E}}_{\substack{s\sim S \\ \beta\in[p-1]^{n-k}}}\left[\prod_{i\in\overline{T}} \omega_p^{\beta_i s_i}\right] \\
&= (p-1)^{n-k} \sum_{T\in\binom{[n]}{k}} \prod_{j=1}^{a_{n-k}} \lambda \\
&= (p-1)^{n-k} \sum_{T\in\binom{[n]}{k}} \lambda^{a_{n-k}}
\end{aligned}$$

We then parameterize the summation over every possible value of the shift of T (for $k \mod p \equiv 0$):

$$\mathbb{E}[K_k(|s|_0)] = (p-1)^{n-k} \sum_{d=k}^{n-k} \left(\sum_{\substack{T\in\binom{[n]}{k} \\ \text{shift}_0(\text{T})=\text{d}}} 1 \right) \lambda^d = (p-1)^{n-k} \sum_{d=k}^{n-k} \phi_0(d)\lambda^d$$

This yields the claim.

□

Lemma B.0.4. For $c \in \mathbb{N}$ where $0 \leq c \leq p$, and for $d \in \mathbb{N}$ where $k \leq d \leq n-k$, the number of k-sized subsets of $[n]$ that satisfy $\text{shift}_c(T) = d$ is:

$$\phi_c(d) = \sum_{\substack{T \in \binom{[n]}{k} \\ \text{shift}_c(T)=d}} 1 = \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

Proof. To determine $\phi_c(d)$, we count the total number of ways to choose $a_1 < a_2 < ... < a_k$ such that $(a_c - a_{c-1}) + (a_{c+p} - a_{c+p-1}) + (a_{c+2p} - a_{c+2p-1}) + ... = d$, where for any $j \leq 0$, $a_j = 0$. To do this, we combine each element-wise interval (a_{c+ip}, a_{c+ip-1}) to form a contiguous interval of length d (starting from $a_{c-1} = 0$). The remaining contiguous region that excludes these intervals must then have a length of $n - d$. We then abstract the number of ways to count $a_1 < ... < a_k$ by counting the number of intervals that have a length of d when combined, such that the remaining intervals have a length $n - d$.

From a length of $d - 1$ (accounting for $a_0 = 0$), we need to select intervals that form a length of $\lfloor |k-c|/(p-1) \rfloor - 1$ since they represent the number of choices of elements of T that are index-separated by p. Similarly, from a length of $n - d$, we need to select intervals that form a length of $\lfloor |k-c|/(p-1) \rfloor$ possible intervals, since they represent every other element of T. This second constraint is to ensure that the total length of the intervals chosen is exactly n. Hence, we write that:

$$\phi_c(d) = \sum_{\substack{T \in \binom{[n]}{k} \\ \text{shift}_c(T)=d}} 1 = \binom{d-1}{\lfloor \frac{|k-c|}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{|k-c|}{p-1} \rfloor}$$

□

Corollary B.0.4.1. By combining the results from lemmas B.0.3 and B.0.4, we have that the expectation of the Krawtchouk function is*:*

$$\mathbb{E}[K_k(|s|_0)] = \begin{cases} (p-1)^{n-k} \sum_{d=k}^{n-k} \binom{d-1}{\lfloor \frac{k}{p-1} \rfloor - 1} \binom{n-d}{\lfloor \frac{k}{p-1} \rfloor} \lambda^d, & k \bmod p \equiv 0 \\ 0, & k \bmod p \not\equiv 0 \end{cases}$$

Proposition 3. The orthogonality of the generalized Krawtchouk function $K_k(\ell)$ in B.0.1 implies that for any function $f : \mathbb{Z}_{n+1} \to \mathbb{R}$, there exists a unique expansion $f(\ell) = \sum_{k=0}^{n} \hat{f}(k) K_k(\ell)$, where for $0 \leq k \leq n$,

$$\hat{f}(k) = \frac{\mathop{\mathbb{E}}_{b \sim [\Sigma(U_p^n)]_0} [f(b) K_k(b)]}{\binom{n}{k}(p-1)^{n-k}}$$

Definition 28. Let $q : \mathbb{Z}_{n+1} \to \mathbb{R}$, where $q(\ell) = \frac{\Pr_{s\sim S(n,p,\lambda)}[|s|_0=\ell]}{\binom{n}{\ell}(p-1)^{n-\ell}}p^n$. *Intuitively, $q(\ell)$ is the ratio of the probability of getting a string with ℓ 0s from the sticky random walk $S(n,\lambda,p)$ to the probability of getting a string with ℓ 0s from the uniformly random distribution.*

Lemma B.0.5. Expanding $q(\ell)$ through the generalized Krawtchouk function yields that:

$$\hat{q}(k) = \frac{1}{\binom{n}{k}(p-1)^{n-k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]$$

Proof.

$$\begin{aligned}
\hat{q}(k) &= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \sum_{b=0}^{n} \binom{n}{b} \frac{(p-1)^{n-b}}{p^n} q(b) K_k(b) \\
&= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \sum_{b=0}^{n} \Pr_{s\sim S(n,p,\lambda)}[|s|_0 = b] K_k(b) \\
&= \frac{1}{\binom{n}{k}(p-1)^{n-k}} \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]
\end{aligned}$$

□

Lemma B.0.6. For $s \in S(n,p,\lambda)$, we have that

$$\Pr[|s|_0 = \ell] = \frac{1}{p^n} \sum_{k=0}^{n} K_\ell(k) \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)]$$

Proof.

$$\begin{aligned}
\Pr[|s|_0 = \ell] &= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} q(\ell) \\
&= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} \sum_{k=0}^{n} \hat{q}(k) K_k(\ell) \\
&= \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n} \sum_{k=0}^{n} \frac{K_k(\ell)}{\binom{n}{k}(p-1)^{n-k}} \mathbb{E}[K_k(|s|_0)] \\
&= \frac{1}{p^n} \sum_{k=0}^{n} \frac{\binom{n}{\ell}}{\binom{n}{k}} \frac{(p-1)^{n-\ell}}{(p-1)^{n-k}} \mathbb{E}[K_k(|s|_0)] K_k(\ell) \\
&= \frac{1}{p^n} \sum_{k=0}^{n} K_\ell(k) \mathop{\mathbb{E}}_{s\sim S(n,p,\lambda)}[K_k(|s|_0)] \quad \text{(By the reciprocity relation in B.0.2.1}
\end{aligned}$$

□

Lemma B.0.7. The total variational distance between the n-step generalized sticky random walk on p vertices and the n-round uniform distribution on p states is:

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(U_p^n)]_0) = \frac{1}{2} \mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[|p(b)-1|]$$

Proof.

$$\begin{aligned}
\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(U_p^n)]_0) &= \frac{1}{2}\sum_{\ell=0}^{n}\left|\Pr[|s|_0=\ell] - \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n}\right| \\
&= \frac{1}{2}\sum_{\ell=0}^{n}\left|\binom{n}{\ell}q(\ell)\frac{(p-1)^{n-\ell}}{p^n} - \frac{\binom{n}{\ell}(p-1)^{n-\ell}}{p^n}\right| \\
&= \frac{1}{2}\sum_{\ell=0}^{n}\left|\frac{\binom{n}{\ell}}{p^n}(p-1)^{n-\ell}(q(\ell)-1)\right| \\
&= \frac{1}{2}\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[|q(b)-1|]
\end{aligned}$$

□

Claim. *The total variational distance between the generalized sticky random walk and the multinomial distribution has the following upper bound as a result of convexity:*

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) \le \frac{1}{2}\sqrt{\mathop{\mathbb{E}}_{b\sim[\Sigma(\mathrm{U}_p^n)]_0}[q(b)-1]^2}$$

Lemma B.0.8. For $k \le n$ and for $b \sim [\Sigma(U_p^n)]_0$, we have that

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 = \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}(p-1)^{n-k}}$$

Proof.

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 = \mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}\left[\left(\sum_{k=0}^{n}\hat{q}(k)K_k(b)-1\right)^2\right]$$

Recall that $\hat{q}(k) = \frac{\mathbb{E}[K_k(|s|_0)]}{\binom{n}{k}(p-1)^{n-k}}$. So, $\hat{q}(0) = \frac{\mathbb{E}[K_0(|s|_0)]}{(p-1)^n} = \frac{1}{(p-1)^n}$. Similarly, by the definition of the Krawtchouk function and the reciprocity relation $\frac{K_k(\ell)}{\binom{n}{k}(p-1)^{n-k}} =$

$\frac{K_s(\ell)}{\binom{n}{s}(p-1)^{n-s}}$, we have that $K_0(b) = K_n(b) = (p-1)^n$. Therefore, $\hat{q}(0)K_0(b) = 1$. Thus, the above equation simplifies to:

$$\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 = \mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}\left[\left(\sum_{k=1}^{n}\hat{q}(k)K_k(b)\right)^2\right]$$

Since the generalized Krawtchouk functions are orthogonal (as proven in B.0.1), the non-diagonal products evaluate to 0. So, the square of the summation is just the summation of the squared terms that it contains. Thus, exploiting the orthogonality of the generalized Krawtchouk functions and the linearity of the expectations, we write:

$$\begin{aligned}\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 &= \mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}\left[\sum_{k=1}^{n}\hat{q}(k)^2K_k(b)^2\right]\\ &= \sum_{k=1}^{n}\hat{q}(k)^2\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[K_k(b)^2]\\ &= \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}^2(p-1)^{2n-2k}}\cdot\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[K_k(b)^2]\end{aligned}$$

Finally, we use lemma B.0.1 to write $\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[K_k(b)^2]$ as $\langle K_k, K_k\rangle = \binom{n}{k}(p-1)^{n-k}$.

$$\begin{aligned}\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 &= \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}^2(p-1)^{2n-2k}}\binom{n}{k}(p-1)^{n-k}\\ &= \sum_{k=1}^{n}\frac{\mathbb{E}[K_k(|s|_0)]^2}{\binom{n}{k}(p-1)^{n-k}}\end{aligned}$$

□

Theorem B.0.9. *For $\lambda \le \frac{1}{1+e}$,*

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) \le \sqrt{\mathop{\mathbb{E}}_{b\sim M}[p(b)-1]^2} \le O(\lambda p^p)$$

Proof. Substituting the result of B.0.6 into the equation derived in B.0.8, and scaling the indexes of the summation, we have that:

$$\begin{aligned}\mathop{\mathbb{E}}_{b\sim[\Sigma(U_p^n)]_0}[q(b)-1]^2 &= \sum_{k=1}^{n/p}\frac{1}{\binom{n}{pk}(p-1)^{n-pk}}\left((p-1)^{n-pk}\sum_{d=pk}^{n-pk}\binom{d-1}{\lfloor\frac{kp}{p-1}\rfloor-1}\binom{n-d}{\lfloor\frac{kp}{p-1}\rfloor}\lambda^d\right)^2\\ &= \sum_{k=1}^{n/p}\frac{(p-1)^{n-pk}}{\binom{n}{pk}}\left(\sum_{d=pk}^{n-pk}\binom{d-1}{\lfloor\frac{k}{1-\frac{1}{p}}\rfloor-1}\binom{n-d}{\lfloor\frac{k}{1-\frac{1}{p}}\rfloor}\lambda^d\right)^2\\ &\le \sum_{k=1}^{n/p}(p-1)^{n-pk}\frac{\binom{n}{k}^2}{\binom{n}{pk}}\left(\sum_{d=pk}^{n-pk}\binom{d-1}{k-1}\lambda^d\right)^2\end{aligned}$$

Note the following generating function relation that $(\frac{x}{1-x})^k = \sum_{m \geq k} \binom{m-1}{k-1} x^m$. Then,

$$\begin{aligned}
\mathop{\mathbb{E}}_{b \sim [\Sigma(U_p^n)]_0} [q(b) - 1]^2 &\leq \sum_{k=1}^{n/p} (p-1)^{n-pk} \frac{\binom{n}{k}^2}{\binom{n}{pk}} \left(\frac{\lambda}{1-\lambda}\right)^{2k} \\
&\leq \sum_{k=1}^{n/p} (p-1)^{n-pk} \left(\frac{pk}{n}\right)^{pk} \left(\frac{en}{k}\right)^{2k} \left(\frac{\lambda}{1-\lambda}\right)^{2k} \qquad \text{(From claim A.0.1)} \\
&= \sum_{k=1}^{n/p} (p-1)^{n-pk} \left(\frac{pk}{n}\right)^{pk-2k} \left(\frac{pe\lambda}{1-\lambda}\right)^{2k} \\
&\leq p^{2p} \sum_{k=1}^{n/p} \left(\frac{e\lambda}{1-\lambda}\right)^{2k} \\
&\leq p^{2p} O(\lambda^2), \quad \text{for } \lambda \leq \frac{1}{1+e}
\end{aligned}$$

Therefore, for $\lambda \leq \frac{1}{1+e}$, we have that

$$\mathrm{TVD}([\Sigma(S(n,p,\lambda))]_0, [\Sigma(\mathrm{U}_p^n)]_0) \leq \sqrt{\mathop{\mathbb{E}}_{b \sim M} [p(b) - 1]^2} \leq O(\lambda p^{O(p)})$$

□

This method shows that the total variation distance between the n-step generalized sticky random walk on p vertices and the n-ary samples from the uniform distribution on $\mathbb{Z}_p$ is $O(\lambda p^{O(p)})$, which matches the more general result predicted in [GV22].