Electronic Colloquium on Computational Complexity, Report No. 163 (2020)

Expander Random Walks: A Fourier-Analytic Approach

Gil Cohen* Noam Peri Amnon Ta-Shma†

November 5, 2020

Abstract

In this work we ask the following basic question: assume the vertices of an expander graph are labelled by $0, 1$. What "test" functions $f : \{0,1\}^t \to \{0,1\}$ cannot distinguish t independent samples from those obtained by a random walk? The expander hitting property [AKS87] is captured by the AND test function, whereas the fundamental expander Chernoff bound [Gil98a, Hea08] is about test functions indicating whether the weight is close to the mean. In fact, it is known that all threshold functions are fooled by a random walk [KV86]. Recently, it was shown that even the highly sensitive PARITY function is fooled by a random walk [TS17].

We focus on balanced labels. Our first main result is proving that *all* symmetric functions are fooled by a random walk. Put differently, we prove a central limit theorem (CLT) for expander random walks *with respect to the total variation distance*, significantly strengthening the classic CLT for Markov Chains that is established with respect to the Kolmogorov distance [KV86]. Our approach significantly deviates from prior works. We first study how well a Fourier character χ_S is fooled by a random walk as a function of S. Then, given a test function f, we expand f in the Fourier basis and combine the above with known results on the Fourier spectrum of f.

We also proceed further and consider general test functions - not necessarily symmetric. As our approach is Fourier analytic, it is general enough to analyze such versatile test functions. For our second result, we prove that random walks on sufficiently good expander graphs fool tests functions computed by $\mathbf{AC}^0$ circuits, read-once branching programs, and functions with bounded query complexity.

*Department of Computer Science, Tel Aviv University, Tel Aviv, Israel. The research leading to these results has received funding from the Israel Science Foundation (grant number 1569/18) and from the Azrieli Faculty Fellowship. Email: gil@tauex.tau.ac.il.

†Department of Computer Science, Tel Aviv University, Tel Aviv, Israel. The research leading to these results has received funding from the Israel Science Foundation (grant number 952/18). Email: amnon@tauex.tau.ac.il.

ISSN 1433-8092

Contents

1 Introduction

Expander graphs are among the most useful combinatorial objects in theoretical computer science. They are pivotal to fundamental works in derandomization [INW94, Rei05], complexity theory [Val76, AKS87, Din07] and coding theory [SS96, KMRZS17, TS17] to name a few. Informally, expanders are sparse undirected graphs that have many desirable pseudorandom properties. A formal definition can be given in several equivalent ways,[1] and here we consider the algebraic definition. An undirected graph $G = (V, E)$ is a λ-*spectral expander* if the second largest eigenvalue of its normalized adjacency matrix is bounded above by λ. For simplicity, we only consider d-regular graphs. In this case, M is also the random walk matrix of G.

In their seminal works, [LPS88, Mar88] proved the existence of Ramanujan graphs, i.e., an infinite family of d-regular λ-spectral expanders with number of vertices n going to infinity, and $\lambda \leq 2\frac{\sqrt{d-1}}{d}$. This relation between the degree d and λ is essentially tight as follows by the Alon and Boppana bound (see [Alo86, Nil91]). Explicit constructions of expander graphs–Ramanujan or otherwise–attracted a significant attention, e.g., [GG81, AGM87, Ajt94, BL06, RVW00, BATS11] and more recently [Coh16] (extending on [MSS15, MSS18]) and [MOP20]. Many works in the literature have studied and utilized the pseudorandom properties of expanders, and we refer the reader to excellent expositions on expander graphs [HLW06, Tre17] and to Chapter 4 of [Vad12]. See also [Lub12] for applications to pure mathematics.

Expanders can be thought of as spectral sparsifiers of the clique. Specifically, let $\mathbf{J}$ be the normalized adjacency matrix of the n-vertex complete graph (with self-loops). That is, $\mathbf{J}$ is the $n \times n$ matrix with all entries equal to $\frac{1}{n}$. One can express the normalized adjacency matrix M of G by $M = (1-\lambda)\mathbf{J} + \lambda E$ for some operator E with spectral norm at most 1. As such, one can hope to substitute a sample of two *independent* vertices with the cheaper process of sampling an edge from an expander and using its two (highly correlated) end-points. This is captured, e.g., by the expander mixing lemma [AC88]. This idea also appears in many derandomization results, e.g., [INW94, AEL95, RRV99, Rei05, RV05, BCG20], to name a few.

A natural and useful generalization of the above idea is to consider not just an edge but rather a length $t-1$ random walk (where the length is measured in edges) on the expander as a replacement to t independent samples of vertices. For concreteness, consider a labelling $\mathsf{val} : V \to \{0, 1\}$ of the vertices by 0 and 1 with mean $\mu = \mathbf{E}[\mathsf{val}(V)]$. Indeed, quite a lot is known:

- The basic hitting property of expanders [AKS87, Kah95] states that for every set $A \subset V$, a length $t-1$ random walk is contained in A with probability at most $(\mu+\lambda)^t$. For $\lambda \ll \mu$, this bound is close to μ^t–the probability of the event with respect to t independent samples. Note that the expander hitting property corresponds to a random walk "fooling" the AND function, that is, for every λ-spectral expander and every labelling val as above, the AND function cannot distinguish with good

[1] In certain regime of parameters, the equivalence breaks.

probability labels obtained by t independent samples from labels obtained by taking a length $t-1$ random walk.

- To give another example, the fundamental expander Chernoff bound [AKS87, Gil98a, Hea08] states that the number of vertices on a random walk residing in A is highly concentrated around μ. Observe that the expander Chernoff bound corresponds to fooling functions $f_\tau : \{0,1\}^t \to \{0,1\}$ indicating whether the normalized Hamming weight of the input is concentrated around μ, more precisely, $f_\tau(x_1, \ldots, x_t) = 1$ if and only if $\frac{1}{t}\sum_{i=1}^{t} x_i \in [\mu - \tau, \mu + \tau]$.
- In fact, it is also known that all threshold functions are fooled by a random walk [KV86, Lez01, Klo17] and we explain this in more detail later.
- It was shown that the highly sensitive PARITY function is fooled by a random walk on expanders (this was noted by Alon in 1993, Wigderson and Rozenman in 2004 and [TS17] where the result appears).

However, it is clear that sometimes a long random walk *is not* a good replacement to independent samples. To see this, suppose G is a λ-spectral expander for some constant λ, that has a cut $A \subset V$ with $|A| = \frac{|V|}{2}$ and $|E(A, \overline{A})| \geq \mu|A|$ for $\mu \geq \frac{1}{2} + \widetilde{\Omega}(\lambda)$. Such graphs exist, e.g., the graph constructed in [GK20, Section 7] is such. If we sample t independent vertices $(v_1, \ldots, v_t)$ from the graph, we expect (v_i, v_{i+1}) to cross the cut about half the time, and by the Chernoff bound the actual number of cut crossings is highly concentrated around the mean. In contrast, when we take a random walk on the graph we expect to cross the cut a μ-fraction of the time, and intuitively the number of cut crossings should be concentrated around μ.[2] Thus, the simple test function that counts the number of times we cross the cut and apply a threshold at $\frac{1}{2} + \tau$ for some $\tau = \widetilde{\Theta}(\lambda)$ should distinguish with probability close to 1 between a random walk and independent samples. This brings to the front a natural and fundamental question:

What test functions does a random walk on an expander fool?

1.1 Our contribution

To give a formal description of our contribution, we set some notation. First, we are mainly concerned with balanced labelling functions $\mathsf{val} : V \to \{0,1\}$, that is, $\mu = \mathbf{E}[\mathsf{val}(V)] = \frac{1}{2}$, or equivalently, with balanced cuts. The reason being is that we are trying to focus our attention on the dependencies across the vertices of a random walk. Setting $\mu = 0$ (and working with regular graphs) allows us to do so as the label of every vertex on a random walk is marginally unbiased. Of course, the case $\mu \neq 0$ is very interesting as well though we leave it for future research.

[2] To show such a concentration one needs to prove a Chernoff bound for a walk on the corresponding directed line graph.

We compare two distributions on the set $\{0,1\}^t$. The first "ideal" distribution is obtained by sampling independently and uniformly at random t vertices $v_1, \ldots, v_t$ and returning $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. As we assume val is balanced, this is simply the uniform distribution over $\{0,1\}^t$ which we denote by U_t. The second distribution is obtained by taking a length $t-1$ random walk on the graph, namely, we sample v_1 uniformly at random from V, and then for $i = 2, 3, \ldots, t$, we sample v_i uniformly at random from the set of neighbors of v_{i-1}. We then return $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. Denote

$$\mathcal{E}_{G,\mathsf{val}}(f) = \left| \mathbf{E}\, f(\mathrm{RW}_{G,\mathsf{val}}) - \mathbf{E}\, f(U_t) \right|.$$

Informally, $\mathcal{E}_{G,\mathsf{val}}(f)$ measures the distinguishability between these two distributions as observed by the test function f on the graph G with respect to the labelling val.

We wish to have a result that holds uniformly on all λ-spectral expanders (on any number of vertices) and for every balanced labelling. We denote by $\mathcal{E}_\lambda(f)$ the supremum of $\mathcal{E}_{G,\mathsf{val}}(f)$ over all λ-spectral expanders G, on any number of vertices $|V| = n$, and all balanced labelling functions $\mathsf{val} : V \to \{0,1\}$. We say that a random walk on λ-spectral expanders ε-fools f if $\mathcal{E}_\lambda(f) \le \varepsilon$.

1.1.1 Random walks fool all symmetric functions

As discussed above, it is known that several symmetric functions are fooled by a random walk, and each teaches us a different aspect of the pseudorandom nature of expander graphs. For example, f_τ is concerned with concentration around the mean whereas the majority function focuses on the symmetry around the mean (recall, val is a balanced labelling). The fact that PARITY is fooled by a random walk is somewhat surprising as PARITY is as far as can be from being monotone, put differently, it is highly sensitive.

Our first main result states that *all symmetric functions* are fooled by a random walk.

Theorem 1.1. *For every symmetric function $f : \{0,1\}^t \to \{0,1\}$,*

$$\mathcal{E}_\lambda(f) = O(\lambda \cdot \log^{3/2}(1/\lambda)).$$

We remark that the requirement that f is symmetric is important as we already saw before that the test that counts the number of times we cross a cut distinguishes between independent samples and the random walk samples. Indeed, the number of times we cross a cut depends on the order of zeroes and ones in the sequence and is not a symmetric function.

A different perspective one can take on Theorem 1.1 is that it establishes a *central limit theorem* for random walks with respect to the *total variation distance*. We turn to elaborate on this but first introduce a convenient notation, specialized to symmetric test functions. When focusing on symmetric functions it is more natural to consider not $\mathrm{RW}_{G,\mathsf{val}}$ and U_t as above, but rather the two distributions on $\{0, 1, \ldots, t\}$ obtained by taking the weights of the two corresponding bit strings. More explicitly, we define $\Sigma\mathrm{Ind}_t$ to be the distribution obtained by sampling t independent vertices $v_1, \ldots, v_t$ and returning

$\mathsf{val}(v_1)+\cdots+\mathsf{val}(v_t)$. The distribution $\Sigma\mathrm{RW}_t$ is defined as the sum of $\mathsf{val}(v_1)+\cdots+\mathsf{val}(v_t)$ where $(v_1,\ldots,v_t) \sim \mathrm{RW}_{G,\mathsf{val}}$. Note that we suppress the dependence on G, val from the notation as they will be clear from context. Instead, we focus our attention on t.

What is known about $\Sigma\mathrm{Ind}_t$? First, the Chernoff bound [Che52] tells us that $\Sigma\mathrm{Ind}_t$ is highly concentrated around the mean, where the probability to be c standard deviations away from the mean is about $2^{-\Omega(c^2)}$ small. This implies that there is very low weight on the tails, but does not tell us much about the center, where almost all of the probability mass resides. In particular, the Chernoff bound does not rule out the possibility that all the weight lies on the mean. Further, it gives no information about, say, how symmetric is the distribution around its mean.

The central limit theorem (CLT) guarantees that $\Sigma\mathrm{Ind}_t$ converges to the normal distribution $\mathcal{N}_t$ (with the same mean $t/2$ and variance $t/4$). The convergence of the CLT is with respect to the Kolmogorov distance (see Definition 3.5). That means that the cumulative distribution function (CDF) of $\Sigma\mathrm{Ind}_t$ converges point-wise to the CDF of the normal distribution. Equivalently, it means that every threshold test function cannot distinguish independent samples from the normal distribution. The Berry-Esseen Theorem specifies the rate of convergence, and when, e.g., we sum random variables with bounded first three moments (as in our case where the values are Boolean) the distance between t independent samples and the normal distribution is in the order of $t^{-1/2}$ with respect to the Kolmogorov distance. To summarize, the Chernoff bound guarantees tails have low-weight, the CLT tells us how the weight is distributed around the mean, and the Berry-Esseen theorem bounds the rate of convergence.

We next ask what is known about $\Sigma\mathrm{RW}_t$? In particular, what can we say about the weight of tails, and what can we say about the distribution around its mean. The expander Chernoff bound [Gil98a] states that when the spectral gap $1-\lambda$ is non-trivial, the probability to be c standard deviations away from the mean is still about $2^{-\Omega(c^2)}$ small. The proof was simplified in [Hea08]. Possibly less known by the CS community is that the CLT and the Berry-Esseen Theorem were also shown to hold for random walks on expanders. The CLT was first shown for expanders by Kipnis and Varadhan [KV86] and their work was later vastly generalized (see, e.g., [Lez01, Klo17]). That work shows, e.g., that:

Theorem 1.2. *(Based on, e.g., [Klo17, Thm C]) Let $G=(V,E)$ be a λ-spectral expander, and assume λ is bounded away from 1. Let $\mathsf{val}: V \to \{0,1\}$ with $\mathbf{E}[\mathsf{val}(V)] = \frac{1}{2}$. Then,*

$$\| \Sigma\mathrm{RW}_t - \Sigma\mathrm{Ind}_t \|_{\mathrm{KOL}} = O\left(\frac{1}{\sqrt{t}}\right). \tag{1.1}$$

We remark that by the Berry-Esseen theorem for independent random variables we know that

$$\| \Sigma\mathrm{Ind}_t - \mathcal{N}_t \|_{\mathrm{KOL}} = O\left(\frac{1}{\sqrt{t}}\right),$$

where $\mathcal{N}_t$ is the normal distribution with the appropriate mean and variance (the mean and variance depend on t). It therefore follows that Equation (1.1) is equivalent to

$\| \Sigma\mathrm{RW}_t - \mathcal{N}_t \|_{\mathrm{KOL}} = O\left(\frac{1}{\sqrt{t}}\right)$.

A natural question is whether the convergence can be strengthened to the stronger total variation distance, and this question applies both to the possible convergence of $\Sigma\mathrm{Ind}_t$ to $\mathcal{N}_t$ and of $\Sigma\mathrm{RW}_t$ to $\Sigma\mathrm{Ind}_t$.

The first question was heavily studied in Probability. A representative case is the question on the rate of point-wise convergence of $\Sigma\mathrm{Ind}_t$ to $\mathcal{N}_t$, i.e., how well the appropriate normal distribution $\mathcal{N}_t$ approximates the probability $\Sigma\mathrm{Ind}_t$ gets a specific outcome $m \in [t]$. The bottom line is that when val is distributed over $\{0, 1\}$ the rate of convergence is $o(\frac{1}{\sqrt{t}})$ (see, e.g., [Tao15, Theorem 7]). The fact that the error is $o(\frac{1}{\sqrt{t}})$ rather than $O(\frac{1}{\sqrt{t}})$ is crucial, and, in particular, implies convergence in the TVD with error $o(1)$ (because the probability mass outside $[-c\sqrt{t}, c\sqrt{t}]$ is tiny $2^{-\Omega(c^2)}$ and therefore almost all of the action takes place on an interval of length $O(\sqrt{t}))$.

The same question can be asked with respect to the random variables $\Sigma\mathrm{RW}_t$ and $\Sigma\mathrm{Ind}_t$. The answer in this case is given by Theorem 1.1 that can be equivalently stated as:

Theorem 1.3 (Theorem 1.1; equivalent statement)**.** *Let $G = (V, E)$ be a λ-spectral expander, and assume λ is bounded away from 1. Let $\mathsf{val} : V \to \{0, 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = \frac{1}{2}$. Then,*

$$\| \Sigma\mathrm{RW}_t - \Sigma\mathrm{Ind}_t \|_{TVD} = O(\lambda \cdot \log^{3/2}(1/\lambda)).$$

This is because the total variation distance between $\Sigma\mathrm{RW}_t$ and $\Sigma\mathrm{Ind}_t$ is the same as the best distinguishing probability a test on $\Sigma\mathrm{RW}_t$ and $\Sigma\mathrm{Ind}_t$ can achieve, which amounts to the best distinguishing probability a symmteric function can achieve on RW_t and Ind_t. Thus, while the Kolmogorov distance amounts to fooling all *threshold* functions, total variation distance amounts to fooling all *symmetric* functions.

To conclude the section we remark that sometimes our results give better bounds even for *threshold* functions. For example, from Theorem 1.2 one can infer than $\mathcal{E}_\lambda(\mathsf{MAJ}_t) \leq O\left(\frac{1}{\sqrt{t}}\right)$ with the constant factor independent of λ. However, in Theorem 4.6 we show a similar result but with the constant going to zero together with λ, namely:

Theorem 1.4. *For every $\lambda \in [0, 1]$ and $t \in \mathbb{N}$,*

$$\mathcal{E}_\lambda(\mathsf{MAJ}_t) \leq O\left(\frac{\lambda^2}{\sqrt{t}}\right).$$

We do not know whether the bound should decay with t for general symmetric functions and leave this as an open problem.

1.1.2 Beyond symmetric functions

We proceed even further and consider general test functions - not necessarily symmetric. We take a complexity-oriented perspective and instead of analyzing specific test functions, we consider natural complexity classes. In particular, we analyze tests that are computable by $\mathbf{AC}^0$ circuits, various types of read once branching programs, and functions with bounded query complexity. As we discuss in Section 2, our approach for proving Theorem 1.1 is Fourier-analytic. As such it is general enough to allow us to analyze such versatile tests functions as well, deviating significantly from prior works both in terms of techniques and results. Moreover, it allows us to utilize known results on the Fourier spectrum of the above-mentioned classes [RSV13, Tal17, CHRT18]. Our second main result is summarized in the following theorem.

Theorem 1.5. *For every function $f : \{0,1\}^t \to \{0,1\}$ the following holds.*

1. *If f is computable by a size-s depth-d circuit then $\mathcal{E}_\lambda(f) = O(\sqrt{\lambda} \cdot (\log s)^{2(d-1)})$.*
2. *If f is computable by (any order) width-w ROBP P, then $\mathcal{E}_\lambda(f) = O(\sqrt{\lambda} \cdot (\log t)^{2w})$. Moreover, if P is a permutation ROBP, $\mathcal{E}_\lambda(f) = O(\sqrt{\lambda} \cdot w^4)$.*
3. *$\mathcal{E}_\lambda(f) = O(\sqrt{\lambda} \cdot \mathrm{DT}(f)^2)$, where $\mathrm{DT}(f)$ denotes the decision tree complexity of f.*

Theorem 1.5 implies that every test function in $\mathbf{AC}^0$ cannot distinguish t independent labels from labels obtained by a random walk on a λ-spectral expander provided λ is taken sufficiently small poly-logarithmic in t. This result can be thought of as an analog of Braverman's celebrated result [Bra10] (see also [Tal17]) that studies the pseudorandomness of k-wise independent distributions with respect to $\mathbf{AC}^0$ test functions. As an example, the Tribes function is fooled by a random walk provided $\lambda = O((\log t)^{-8})$. It is well-known that the decision tree complexity is polynomially-related to other complexity measures such as the randomized and quantum decision tree measures, the certificate query complexity, and the approximate real degree of a function (see, e.g., [BDW02] for further details). Moreover, in a recent breakthrough, Huang resolved the sensitivity conjecture to the affirmative, implying that the decision tree complexity is polynomially-related to the sensitivity of the function [Hua19]. Thus, by Theorem 1.5, every test function f with a bound b on any of these measures cannot distinguish independently sampled labels from labels obtained by a random walk on a λ-spectral expander provided that $\lambda \le b^{-c}$, where c is some universal constant.

1.2 Related work

Very recently, Guruswami and Kumar [GK20], in an independent work, studied the following problem. We are given a distribution $Y_1, \dots, Y_t$, where Y_1 is uniform over $\{0,1\}$, and Y_{i+1} equals Y_i with probability $\frac{1+\lambda}{2}$ and equals $1 - Y_i$ otherwise. They showed that the number of times we see a 1 converge *in total variation distance* to the independent case. They also showed this distribution can emerge from a random walk over some λ-expander $G = (V, E)$ and some balanced coloring of the vertices. A major open problem they raise is whether the same is true for any random walk on a λ-expander, which is answered in the affirmative in this paper.

The techniques Guruswami and Kumar use have a lot in common with our techniques. They use the Krawchuck polynomials (that also appear in our study) and analyze the probability the walk hits the set exactly w times (which corresponds to the weight function $\mathbf{1}_w$ that we study in Section 4.4). The main difference between their work and ours is that we study the problem for an arbitrary λ-spectral expander, rather than the specific λ-sticky walk they analyze. Our approach for that is to use the Fourier representation, and analyze the basis functions (i.e., characters, or equivalently, parities) using the analysis of parity on random walks. This analysis was first done in the unpublished works by Alon in 1993 and Wigderson and Rozenman in 2004, and later in [TS17]. We explain our technique in Section 2.

We already mentioned that there has been a lot of work on CLT and Berry-Esseen on Markov chains, see e.g., [KV86, Lez01, Klo17]. We also give in Section 6 a new analysis for the CLT and Berry-Esseen theorem on expander graphs. This analysis build on the techniques of [Gil98b, Hea08, Bec15] to analyze the characteristic function of $\Sigma\mathrm{RW}_t$, and then uses the framework suggested in [Klo17] to finish the proof.

2 Proof overview

As mentioned, our approach for proving Theorem 1.1 and Theorem 1.5 is Fourier analytic. That is, we first analyze $\mathcal{E}_\lambda$ on Fourier characters (namely, parity functions). Then, we invoke known results on the Fourier expansion of the function under consideration. This leads us to study a new Fourier tail we dub the *Random Walk Fourier tail*, or the λ-*tail*. To this end, it is more convenient to discuss test functions of the form $f : \{\pm1\}^t \to \{\pm1\}$. Parity functions are then given by $\chi_S(x) = \prod_{i \in S} x_i$ for $S \subseteq [t]$. Before giving the formal definition and results, in Section 2.1 we consider some examples to gain intuition. In particular, we find it instructive to proceed by analyzing parities according to their degree.

2.1 Toy examples: the first few parities

Degree 1. To start with, consider degree 1, namely, a dictator function $\mathsf{Dict}_i(x) = x_i$ for some $i \in [t]$. As we assume G is regular and val balanced, the marginal distribution of the i^{th} vertex on a random walk is uniform over V, and so $\mathcal{E}_\lambda(\mathsf{Dict}_i) = 0$.

Degree 2. Consider a function that is the parity of two of its input bits $f(x) = x_{i_1}x_{i_2}$ for some $i_1 < i_2$. We already know that the i_1^{st} vertex is uniformly distributed over V. Intuitively, the larger the distance $\Delta = i_2 - i_1$ is, the less correlated is the i_2^{nd} vertex on the path to the i_1^{st} vertex. Fully aligned with this intuition, it can be easily shown that $\mathcal{E}_\lambda(x_{i_1}x_{i_2}) \leq \lambda^\Delta$. Indeed, one can think of a length Δ random walk on a λ-spectral expander as picking a random edge (i.e., a random walk of length 1) on a λ^Δ-spectral expander.

Degree 3. Moving on to degree 3, consider the test function $f(x) = x_{i_1}x_{i_2}x_{i_3}$ with $i_1 < i_2 < i_3$. Denote $\Delta_1 = i_2 - i_1$ and $\Delta_2 = i_3 - i_2$. Here one may root for one of several (conflicting) intuitive arguments. First, one might argue that if one of Δ_1, Δ_2

is small then two of the bits are highly correlated. Being cautious regarding to how correlations behave on a random walk, one might suspect that this results in an overall high correlation. By that logic, $\mathcal{E}_\lambda(f) \approx \lambda^{\min(\Delta_1,\Delta_2)}$. On the other hand, one might argue that if one of Δ_1, Δ_2 is large, regardless of the other, then the far away vertex gained a "large amount of independence", resulting in an overall low distinguishability. Thus, $\mathcal{E}_\lambda(f) \approx \lambda^{\max(\Delta_1,\Delta_2)}$.

Perhaps surprisingly, we show that $\mathcal{E}_\lambda(f) \approx \lambda^{\Delta_1+\Delta_2} = \lambda^{i_3-i_1}$. That is to say, it is only the "effective" path's length–the distance between the first and last observed vertices on the path–that is taken into account, independent of the location of the middle vertex. To see why this is the case, recall that for a λ-spectral expander with a random walk matrix M, it holds that $M = (1-\lambda)\mathbf{J} + \lambda E$. Thus, one can intuitively think of a step on a λ-spectral expander as follows: With probability $1-\lambda$ sample uniformly at random a vertex, completely ignoring the current vertex we are at and the edge structure of the graph, and with probability λ sample a vertex adversarially. We stress that this intuition is not accurate as E is an operator that is not necessarily a random walk matrix of any graph.

With this in mind, consider the random walk from the first to the second vertex. With probability $1-\lambda^{\Delta_1}$ we completely decouple the first vertex from the second, and hence from the entire remaining part of the path. As the first vertex is marginally uniform (recall $\mu = 0$), the parity of the three bits is unbiased. Similarly, with probability $1-\lambda^{\Delta_2}$, the third vertex is independent from the first two. As we think of these events as independent, it is only with probability $\lambda^{\Delta_1+\Delta_2}$ that (adversarial) correlations may appear.

Degree 4. Generalizing the above notation in the natural way, for a degree 4 parity test function, our analysis shows that $\mathcal{E}_\lambda(f) \le \lambda^{\Delta_1+\Delta_3}$. This might be somewhat counter-intuitive. Indeed, one might expect that $\chi_{\{1,2,3,4\}}$ will be harder to fool than $\chi_{\{1,2,t-1,t\}}$ as in the latter case, the first pair of vertices is "far away" from the second pair and so the two pairs should be less correlated compared to the corresponding pairs in $\chi_{\{1,2,3,4\}}$. This, however, is not how correlations on a random walk behave.

To intuitively see why $\mathcal{E}_\lambda(f) \le \lambda^{\Delta_1+\Delta_3}$ note, as in the previous example, that with probability $1-\lambda^{\Delta_1}$ the first vertex is "cut" from the remaining part of the path. Similarly, with probability $1-\lambda^{\Delta_3}$ the fourth vertex is independent of the rest, and so it is only with probability $\lambda^{\Delta_1+\Delta_3}$ that the first and fourth vertices are not independent from the other vertices. We turn to give a formal proof of this fact mainly served as a warm-up for the proof of the general case (see Section 4.1).

Claim 2.1. *Let $S = \{s_1, s_2, s_3, s_4\}$ with $s_1 < s_2 < s_3 < s_4$. For $i = 1,2,3$, denote $\Delta_i = s_{i+1} - s_i$. Then, $\mathcal{E}_\lambda(\chi_S(x)) \le \lambda^{\Delta_1+\Delta_3}$ where $\chi_S(x) = \prod_{i \in S} x_i$.*

Proof. Let $\mathbf{1}$ be the normalized length-n unit vector, that is, every entry of $\mathbf{1}$ equals to $\frac{1}{\sqrt{n}}$. Take G to be any regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ balanced. We slightly abuse notation and denote by G the random walk matrix for G. Let P be the

$V \times V$ diagonal matrix with entry (v, v) equals to $\mathsf{val}(v)$. We first observe that

$$\begin{aligned}\mathcal{E}_{G,\mathsf{val}}(\chi_S) &= \left|\,\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]\right| \\ &= \left|\mathbf{1}^T(PG^{\Delta_3})(PG^{\Delta_2})(PG^{\Delta_1})P\mathbf{1}\right|.\end{aligned}$$

As mentioned, we can write $G = (1-\lambda)\mathbf{J}+\lambda E$ for some bounded operator $\|E\| \leq 1$. More generally, for every $i = 1, 2, 3$ we have that $G^{\Delta_i} = (1 - \lambda^{\Delta_i})\mathbf{J} + \lambda^{\Delta_i} E_i$ with $\|E_i\| \leq 1$. Thus, we can express the right hand side of the above equation as a summation of 8 terms where in each term, we replace each of G^{Δ_i} by either $(1 - \lambda^{\Delta_i})\mathbf{J}$ or $\lambda^{\Delta_i} E_i$. Not all 8 summands contribute to the sum. Indeed, if we replace G^{Δ_1} by $(1 - \lambda^{\Delta_1})\mathbf{J}$ then

$$\begin{aligned}\mathbf{1}^T(PG^{\Delta_3})(PG^{\Delta_2})(P(1-\lambda^{\Delta_1})\mathbf{J})P\mathbf{1} &= \mathbf{1}^T(PG^{\Delta_3})(PG^{\Delta_2})(P(1-\lambda^{\Delta_1})\mathbf{1}\mathbf{1}^T)P\mathbf{1} \\ &= \mathbf{1}^T(PG^{\Delta_3})(PG^{\Delta_2})P(1-\lambda^{\Delta_1})\mathbf{1}(\mathbf{1}^T P\mathbf{1}),\end{aligned}$$

which equals 0 as $\mathbf{1}^T P\mathbf{1} = \mathbf{E}[\mathsf{val}(V)] = 0$. Similarly, to get a nonzero contribution we must take $\lambda^{\Delta_3}E_3$ for G^{Δ_3}. Thus, there are only two contributing summands correspond to the sequences we denote by EJE and EEE. As for the first sequence,

$$\begin{aligned}\left|\mathbf{1}^T(P\lambda^{\Delta_3}E_3)(P(1-\lambda^{\Delta_2})\mathbf{J})(P\lambda^{\Delta_1}E_1)P\mathbf{1}\right| &\leq \|(P\lambda^{\Delta_3}E_3)(P(1-\lambda^{\Delta_2})\mathbf{J})(P\lambda^{\Delta_1}E_1)P\|_2 \\ &\leq (1-\lambda^{\Delta_2})\lambda^{\Delta_1+\Delta_3},\end{aligned}$$

where we used the fact that $\|P\|_2 \leq 1$. Similarly, for the EEE sequence we get a bound of $\lambda^{\Delta_1+\Delta_2+\Delta_3}$. The proof follows by adding the bounds corresponding to the two summands. $\square$

For degrees higher than 4, another reason an E/J sequence does not contribute is the existence of two consecutive J symbols. This is the main "saving" one capitalize on in high degrees (see Section 4.1).

2.2 The general framework

The general framework that we develop for bounding $\mathcal{E}_\lambda(f)$ for a given function f (not necessarily symmetric) is as follows. First, expand f in the Fourier basis and note that

$$\mathcal{E}_{G,\mathsf{val}}(f) \leq \left|\sum_{\substack{S\subseteq[t]\\ S\neq\emptyset}} \widehat{f}(S)\,\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]\right|.$$

We stress that we do not ignore the cancellations that may occur, namely, we work with the absolute value of the sum rather than with the sum of absolute values. This is crucial for our proof of Theorem 1.1 which, indeed, is very delicate. For each character χ_S we follow the steps of Claim 2.1 and express $\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]$ algebraically. As in previous works (e.g., [RVW00, RV05, TS17]), we replace a step G of the graph with $(1-\lambda)\mathbf{J}+\lambda E$, and view E as low-order noise. In previous works one often argues about norms of short sub-sequences, e.g., [RVW00, RV05] look at the norm of two steps while [TS17] look at longer length (but still short) sub-sequences. Instead, here we expand the whole product in full and take into account the structure of the set S in the parity χ_S under consideration.

This is the gist of our general Fourier-analytic framework for analyzing expander random walks. We turn to give some more details on the proof of Theorems 1.1 and 1.5 which falls into this framework.

2.3 Analyzing symmetric functions

For the proof of Theorem 1.1 we consider all weight-indicating test functions. For every $w \in \{0, 1, \ldots, t\}$ let $f_w : \{\pm 1\}^t \to \{0, 1\}$ be defined by $f_w(x) = 1$ if and only if x is of Hamming weight w. To analyze all symmetric functions, it suffices to analyze the weight-indicating functions. In fact, note that one is only interested in $w \in [\frac{t}{2} - c\sqrt{t}, \frac{t}{2} + c\sqrt{t}]$ for some parameter c, as the remaining weights can be handled via the expander Chernoff bound.

Fix w in this range. For the proof of Theorem 1.1, for each $S \subseteq [t]$ we collect the $2^{|S|}$ summands obtained by expanding $\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]$, namely, the summands that correspond to the E/J sequences. We then "switch perspective" and for every fixed such E/J sequence calculate contributions to it from all sets S, taking into account the Fourier spectrum of f_w. The analysis is very delicate. Remarkably, all the pieces fall in place and give the result. As a warm-up in Section 4.3 we prove that the $\mathsf{MAJORITY}$ function is fooled by a random walk. Although this is a known result, our proof is based on completely different techniques.

2.4 The λ-tail

For sets of size $|S| \geq 5$, the bound on $\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]$ is getting more and more cumbersome. For symmetric functions, using a very delicate argument, we are able to work with a very tight bound. However, for the non-symmetric functions under consideration, it is possible and much cleaner to work with a looser bound that is more amendable for analysis. In the following, for a set A, denote by $\binom{A}{\geq k}$ the set of all subsets of A of size at least k.

Definition 2.2. *For an integer $t \geq 1$ define the map $\mathbf{\Delta} : \binom{[t]}{\geq 2} \to \mathbb{N}$ as follows. Let $S \subseteq [t]$, of size $k \geq 2$, and denote $S = \{s_1, \ldots, s_k\}$ where $s_1 < \cdots < s_k$. For $i \in [k-1]$ write $\Delta_i = s_{i+1} - s_i$. For $k = 2$ we define $\mathbf{\Delta}(S) = \Delta_1$, for $k = 3$ define $\mathbf{\Delta}(S) = \Delta_1 + \Delta_2$, and for $k \geq 4$,*

$$\mathbf{\Delta}(S) = \sum_{i=1}^{k-2} \min(\Delta_i, \Delta_{i+1}). \tag{2.1}$$

Using ideas similar to those in Claim 2.1, we prove.

Proposition 2.3. *For every $\lambda \in [0, 1]$, $t \in \mathbb{N}$ and $S \subseteq [t]$ a subset of size $|S| \geq 2$, it holds that*

$$\mathcal{E}_\lambda(\chi_S) \leq 2^{|S|} \cdot \lambda^{\mathbf{\Delta}(S)/2}.$$

We refer the reader to Proposition 4.2 for a stronger statement. Proposition 2.3 naturally leads us to the study of what we call the λ-*tail*.

Definition 2.4 (The λ-tail)**.** *Let $f : \{\pm 1\}^t \to \{\pm 1\}$. For $\lambda \in [0, 1]$ and $k \in \{2, 3, \dots, t\}$, we define*

$$\mathbf{\Lambda}_{\lambda,k}(f) = \sum_{\substack{S \subseteq [t] \\ |S|=k}} |\widehat{f}(S)| \cdot \lambda^{\mathbf{\Delta}(S)}. \tag{2.2}$$

The λ-tail of f is defined by $\mathbf{\Lambda}_\lambda(f) = \sum_{k=2}^{t} \mathbf{\Lambda}_{\lambda,k}(f)$.

In Claim 5.2 we prove that $\mathcal{E}_\lambda(f) \le 4\mathbf{\Lambda}_{2\sqrt{\lambda}}(f)$, and so, to analyze how well random walks fool a given function, it suffices to bound its λ-tail. In Claim 5.3 we bound the λ-tail of functions with a decaying $\mathcal{L}_1$ tail. This then allows us to invoke [RSV13, Tal17, CHRT18] and deduce Theorem 1.5.

2.5 Remarks and future work

We conclude this section with several remarks and open problems that follow from our work.

1. It is an interesting problem that we leave for future work to consider also unbalanced labelling. Namely, a labelling $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = \mu \neq 0$.

2. Can the $\operatorname{poly}\log \frac{1}{\lambda}$ factor in Theorem 1.1 be improved?

3. Can one obtain a bound as in Theorem 1.4, namely decaying as $t \to \infty$ and $\lambda \to 0$, for all threshold functions? For all symmetric functions?

4. Our results on non symmetric functions follow by applying known bounds on the $\mathcal{L}_1$ Fourier tail of the function of interest together with Claim 5.3 that relates the $\mathcal{L}_1$ decay to the λ-tail. Typically, bounds on the $\mathcal{L}_1$ (and $\mathcal{L}_2$) tails are obtained by using random restrictions. An interesting problem is to prove stronger results than those obtained in Theorem 1.5 by directly analyzing (perhaps suitable variants of) random restrictions with respect to the λ-tail. Indeed, we note that the $\mathcal{L}_1$ tail and the λ-tail can behave very differently. To see this, consider any function $f : \{\pm 1\}^t \to \{\pm 1\}$ that is determined by $x_a, x_{2a}, x_{3a}, \dots$ for a parameter $a \gg \log t$. Then, $\mathbf{\Lambda}_{\lambda,k}(f) \le \binom{t}{k}\lambda^{ak} \ll \frac{1}{t}$ for a sufficiently small constant λ, and so $\mathbf{\Lambda}_\lambda(f) \ll 1$. On the other hand, a typical such function does not have a nontrivial decaying $\mathcal{L}_1$ Fourier tail.

3 Preliminaries

We let $[n]$ denote the set $\{1, \dots, n\}$. We let $\mathbf{1} \in \mathbb{R}^n$ denote the normalized all 1s vector, i.e., $\mathbf{1} = \frac{1}{\sqrt{n}} \cdot (1, \dots, 1)^T \in \mathbb{R}^n$. We let $\mathbf{J} = \mathbf{1}\mathbf{1}^T$.Throughout the paper, we make use of the following well known inequalities about binomial coefficients.

Claim 3.1. *Let $0 < \lambda < 1$, and integers $r \geq 0$, $a \geq b \geq 1$. Then,*

1. $(\frac{a}{b})^b \leqslant \binom{a}{b} \leqslant (\frac{ea}{b})^b$,
2. $\sum_{i=r}^{\infty} \binom{i}{r} \lambda^i = \frac{\lambda^r}{(1-\lambda)^{r+1}}$.

The first item is a well-known estimate. For completeness we prove the second item.

Proof of Claim 3.1, Item (2). Denote $S_r = \sum_{i=r}^{\infty} \binom{i}{r} \lambda^i$. Let $h(x) = \sum_{n=0}^{\infty} x^n$ and $H(x) = \frac{1}{1-x}$. Within the domain $x \in (0,1)$ it holds that $H(x) = h(x)$. Notice that,

$$h(\lambda) = \sum_{n=0}^{\infty} \lambda^n = S_0.$$

Fix $\tau < 1$. The series defining $h(x)$, and all its derivatives, uniformly converges in $[0,\tau]$. Thus we can derive underneath the summation sign, i.e, for every $k \in \mathbb{N}$, and every $x \in [0,\tau]$, $h^{(k)}(x) = H^{(k)}(x)$. Furthermore,

$$h^{(k)}(x) = \sum_{n=0}^{\infty} (x^n)^{(k)} = \sum_{n=k}^{\infty} k! \binom{n}{k} x^{n-k},$$

and $H^{(k)}(x) = k!(1-x)^{-(k+1)}$. Together this gives

$$\begin{aligned} S_k &= \frac{\lambda^k}{k!} h^{(k)}(\lambda) = \frac{\lambda^k}{k!} H^{(k)}(\lambda) \\ &= \frac{\lambda^k}{k!} \cdot \frac{k!}{(1-\lambda)^{k+1}} = \frac{\lambda^k}{(1-\lambda)^{k+1}}. \end{aligned}$$

□

3.1 Fourier analysis

Consider the space of functions $f : \{\pm 1\}^t \to \mathbb{R}$, along with the inner product $\langle f, g \rangle = 2^{-t} \sum_{x \in \{\pm 1\}^t} f(x) g(x)$. It is a well-known fact that the set $\{\chi_S \mid S \subseteq [t]\}$, where $\chi_S = \prod_{i \in S} x_i$, forms an orthonormal basis with respect to this inner product, which is called the Fourier basis. Thus every function $f : \{\pm 1\}^t \to \mathbb{R}$ can be uniquely represented as $f(x) = \sum_{S \subseteq [t]} \widehat{f}(S) \chi_S(x)$, where $\widehat{f}(S) \in \mathbb{R}$.

A technical tool that we use in our proof is the noise operator. The definitions and following claims appear in [O'D14].

Definition 3.2. *Let $\rho \in [-1,1]$. For a fixed $x \in \{\pm 1\}^t$ we write $y \sim N_\rho(x)$ to denote the random string y that is drawn as follows: for each $i \in [t]$ independently,*

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2}, \\ -x_i & \text{with probability } \frac{1-\rho}{2}. \end{cases}$$

Definition 3.3. *Let $\rho \in [-1,1]$. The noise operator T_ρ is the linear operator on functions $\{\pm 1\}^t \to \mathbb{R}$, defined as:*

$$T_\rho f(x) \quad = \underset{y\sim N_\rho(x)}{\mathbf{E}} f(y)$$

The fact that the operator is linear follows directly from the linearity of the expectation.

Notice that $T_1(f) = f$ whereas $T_0(f)$ in the constant function $T_0(f) = \mathbf{E} f$. We make use of the following lemma.

Lemma 3.4. *For every function $f : \{\pm 1\}^t \to \mathbb{R}$ it holds that:*

$$T_\rho f(x) \quad = \sum_{S\in[t]} \widehat{f}(S)\rho^{|S|}\chi_S(x).$$

3.2 Distances between probability distributions

Definition 3.5. *Let P, Q be a pair of (not necessarily discrete) distributions over $\mathbb{R}$. Let $\mathcal{B}$ denote the class of Borel sets. We define*

$$\begin{aligned}\mathrm{d_{TV}}(P,Q) &= \sup\nolimits_{A\in\mathcal{B}}\left(P(A) - Q(A)\right),\\ \mathrm{d_{Kol}}(P,Q) &= \sup\nolimits_{x\in\mathbb{R}, I=(-\infty,x]}\left(P(I) - Q(I)\right).\end{aligned}$$

We call $\mathrm{d_{TV}}$ the total variation distance *and $\mathrm{d_{Kol}}$ the* Kolmogorov-Smirnov distance*.*

Definition 3.6. *Let $\{P_n\}_{n\in\mathbb{N}}, Q$ be distributions over $\mathbb{R}$. Then,*

- *(Weak convergence) We write $P_n \Rightarrow Q$ if for every $x_0 \in \mathbb{R}$, $\lim_{n\to\infty} P_n(x_0) = Q(x_0)$.*
- *(Kolmogorov convergence) We write $P_n \Rightarrow_{\mathrm{Kol}} Q$ if $\lim_{n\to\infty} \mathrm{d_{Kol}}(P_n, Q) = 0$.*
- *(TV convergence) We write $P_n \Rightarrow_{\mathrm{TV}} Q$ if $\lim_{n\to\infty} \mathrm{d_{TV}}(P_n, Q) = 0$.*

We note that the TV convergence implies Kolmogorov convergence which, in turn, implies weak convergence. In this language, the CLT and the Berry Esseen theorems state the following.

Theorem 3.7 (CLT for independent distributions)**.** *Suppose X_i are i.i.d. and marginally uniform on $\{\pm 1\}$. Let $S_n = \sum_{i=1}^n X_i$. Then, $\frac{S_n}{\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$. Furthermore, the Berry-Esseen Theorem states that*

$$\mathrm{d_{Kol}}\left(\frac{S_n}{\sqrt{n}}, \mathcal{N}(0,1)\right) \leq \frac{3}{\sqrt{n}}.$$

4 Random walks fool all symmetric test functions

Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling of the vertices of G, that is, $\mathbf{E}[\mathsf{val}(V)] = 0$. Let $t \geq 1$ be a natural number. We recall from the introduction that we want to compare two distributions on $\{\pm 1\}^t$.

- Note that U_t–the uniform distribution over $\{\pm 1\}^t$–is the distribution obtained by sampling t vertices $v_1, \ldots, v_t$ uniformly and independently at random from V and outputting the ordered tuple $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$.
- $\mathrm{RW}_{G,\mathsf{val}}$ is the distribution obtained by sampling a random length $t-1$ path $v_1, \ldots, v_t$ over G and outputting the ordered tuple $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. Equivalently, sample v_1 uniformly at random from V. Then, for $i = 2, 3, \ldots, t$, sample v_i uniformly at random from the neighbors of v_{i-1}.

Let $f : \{\pm 1\}^t \to \{\pm 1\}$ be any test function. Expand f in the Fourier basis,

$$f(x) = \sum_{S \subseteq [t]} \widehat{f}(S) \chi_S(x),$$

where $\chi_S(x) = \prod_{i \in S} x_i$. We have the following easy lemma.

Lemma 4.1. *Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling of the vertices of G. Then, for every function $f : \{\pm 1\}^t \to \mathbb{R}$,*

$$\mathcal{E}_{G,\mathsf{val}}(f) \leq \sum_{\substack{S \subseteq T \\ S \neq \emptyset}} |\widehat{f}(S)| \mathcal{E}_{G,\mathsf{val}}(\chi_S)$$

Proof. As $\mathbf{E}[f(U_t)] = \widehat{f}(\emptyset)$,

$$\begin{aligned} \mathcal{E}_{G,\mathsf{val}}(f) &= |\mathbf{E}\, f(\mathrm{RW}_{G,\mathsf{val}}) - \mathbf{E}\, f(U_t)| \\ &= \Big| \sum_{\substack{S \subseteq T \\ S \neq \emptyset}} \widehat{f}(S)\, \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \Big|. \end{aligned}$$

Since val is balanced, $\mathbf{E}[\chi_S(U_t)] = 0$, and so $\mathcal{E}_{G,\mathsf{val}}(\chi_S) = |\,\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]|$. The proof follows by the triangle inequality. $\square$

Lemma 4.1 motivates us to consider parity test functions. This is the content of the following section.

4.1 Parities test functions

In this section we analyze to what extent expander random walks fool parity tests functions. In particular, we prove Proposition 2.3. In fact, we prove a stronger statement. We start by introducing some notation. For an integer $k \geq 2$, we define the family $\mathcal{F}_k$ of subsets of $[k-1]$ that, informally, consists of all subsets for which at least one of every two consecutive elements participate in the set. We also require the "end points" $1, k-1$ to participate in the set. Formally, we define

$$\mathcal{F}_k = \{I \subseteq [k-1] \mid \{1, k-1\} \subseteq I \text{ and } \forall j \in [k-2] \ \{j, j+1\} \cap I \neq \emptyset\}. \tag{4.1}$$

So, for example, $\mathcal{F}_6$ consists of the elements $\{1,3,5\}$, $\{1,2,4,5\}$ as well as of all subsets of $[5]$ that contain any one of these two elements, namely, $\{1,2,3,5\}$, $\{1,3,4,5\}$ and $\{1,2,3,4,5\}$. We extend the definition in the natural way to $k = 0,1$ by setting $\mathcal{F}_0 = \mathcal{F}_1 = \emptyset$.

Let $S \subseteq [t]$ be a set of cardinality $|S| = k \geq 1$. Write $S = \{s_1, \ldots, s_k\}$ with $s_1 < s_2 < \cdots < s_k$. Set $s_0 = 0$ and $s_{k+1} = t+1$. For $i = 0, 1, \ldots, k$, we denote by $\Delta_i(S) = s_{i+1} - s_i$. When the set S is clear from context, we write Δ_i for short. With these notations, we prove.

Proposition 4.2. *Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling of the vertices of G, that is, $\mathbf{E}[\mathsf{val}(V)] = 0$. Then, for every integers $1 \leq k \leq t$ and every subset $S \subseteq [t]$ of size k,*

$$\mathcal{E}_{G,\mathsf{val}}(\chi_S) \leq \sum_{I \in \mathcal{F}_k} \lambda^{\sum_{j \in I} \Delta_j(S)}.$$

For example, for a set S of size $|S| = 6$,

$$\begin{aligned}\mathcal{E}_{G,\mathsf{val}}(\chi_S) \leq &\lambda^{\Delta_1+\Delta_3+\Delta_5} + \lambda^{\Delta_1+\Delta_2+\Delta_4+\Delta_5} + \lambda^{\Delta_1+\Delta_2+\Delta_3+\Delta_5} + \\ &\lambda^{\Delta_1+\Delta_3+\Delta_4+\Delta_5} + \lambda^{\Delta_1+\Delta_2+\Delta_3+\Delta_4+\Delta_5}.\end{aligned}$$

Before proving Proposition 4.2, we remark that for sets of size $|S| = 1$, the sum is taken over the empty index set $\mathcal{F}_1$ and so, by the standard convention, the sum equals to 0. We also observe that Proposition 2.3 follows by Proposition 4.2. To see this, note that for every $I \in \mathcal{F}_k$,

$$2\sum_{i \in I} \Delta_i \geq \sum_{i=1}^{k-2} \min(\Delta_i, \Delta_{i+1}). \tag{4.2}$$

Indeed, if we define δ_i to be the corresponding indicator for $i \in I$, namely, $\delta_i = 1$ if $i \in I$ and $\delta_i = 0$ otherwise, we see that

$$2\sum_{i \in I} \Delta_i \geq \sum_{i=1}^{k-2} \delta_i \Delta_i + \delta_{i+1}\Delta_{i+1}.$$

Equation (4.2) follows since $\delta_i \Delta_i + \delta_{i+1}\Delta_{i+1} \geq \min(\Delta_i, \Delta_{i+1})$ as indeed, for every $i \in [k-2]$, at least one of $i, i+1$ is in I. Now, recall that in Equation (2.1), the right hand side of Equation (4.2) was denoted by $\mathbf{\Delta}(S)$. As $|\mathcal{F}_k| \leq 2^{k-1}$, Proposition 2.3 follows by Proposition 4.2. We turn to prove Proposition 4.2.

Proof of Proposition 4.2. Consider any nonempty set $S \subseteq [t]$ of size $|S| = k$. As $\mathbf{E}[\chi_S(U_t)] = 0$, we have that

$$\mathcal{E}_{G,\mathsf{val}}(\chi_S) = \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})].$$

We wish to express the right hand side algebraically. Let $n = |V|$ and identify V with $[n]$ in an arbitrary way. Let P be a $n \times n$ diagonal matrix with $\mathsf{val}(v)$ on the diagonal in row v. We slightly abuse notation and denote the random walk matrix (that is, the normalized adjacency matrix) of G also by G. Define $\delta_i = 1$ if $i \in S$ and $\delta_i = 0$ otherwise. Observe that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^{t} P^{\delta_i} G \right) \mathbf{1}.$$

Indeed, informally, at the i'th step we take a random step using G and then, depending on i being an element of I or not, we multiply by P or by I, respectively. Thus, we can write

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^{k-1} P G^{\Delta_i} \right) P\mathbf{1}, \tag{4.3}$$

where we have used the regularity of G, namely, $G\mathbf{1} = \mathbf{1}$.

Next, we use the spectral decomposition of the expander graph G. As G is a λ-spectral expander we know that $G = \mathbf{J} + \lambda E$ where $\| E \| \le 1$ [3]. Similarly, As G^ℓ is a λ^ℓ-spectral expander we have that $G^\ell = \mathbf{J} + \lambda^\ell E_\ell$ where $\| E_\ell \| \le 1$. Thus,

$$\prod_{i=1}^{k-1} P G^{\Delta_i} = \sum_{I \subseteq [k-1]} \prod_{i=1}^{k-1} P B_i(I), \tag{4.4}$$

where

$$B_i(I) = \begin{cases} \lambda^{\Delta_i} E_{\Delta_i} & i \in I; \\ \mathbf{J} & \text{otherwise.} \end{cases}$$

For $I \subseteq [k-1]$ let

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{k-1} P B_i(I) \right) P\mathbf{1}.$$

Equations (4.3) and (4.4) imply that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \sum_{I \subseteq [k-1]} e_I. \tag{4.5}$$

Not all subsets $I \subseteq [k-1]$ contribute non-zero values e_I to the sum. Indeed, if $k-1 \notin I$

[3] Note that this is slightly different than the decomposition $G = (1-\lambda)\mathbf{J} + \lambda E$ that was used in the introduction.

then $B_{k-1}(I) = \mathbf{J}$ and so

$$\begin{aligned} e_I &= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) (P\mathbf{J})P\mathbf{1} \\ &= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) (P\mathbf{1}\mathbf{1}^T)P\mathbf{1} \\ &= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) P\mathbf{1}(\mathbf{1}^T P\mathbf{1}). \end{aligned}$$

As $\mathbf{1}^T P\mathbf{1} = \mathbf{E}[\mathsf{val}(V)] = 0$, we have that $e_I = 0$. Similarly $e_I = 0$ for I not containing 1. Moreover, if $j, j+1$ are both not contained in I for some $j \in [k-2]$ then

$$\begin{aligned} e_I &= \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I) \right) (PB_j(I))(PB_{j+1}(I)) \left(\prod_{i=j+2}^{k-2} PB_i(I) \right) P\mathbf{1} \\ &= \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I) \right) (P\mathbf{J})(P\mathbf{J}) \left(\prod_{i=j+2}^{k-2} PB_i(I) \right) P\mathbf{1}. \end{aligned}$$

However,

$$(P\mathbf{J})(P\mathbf{J}) = (P\mathbf{1}\mathbf{1}^T)(P\mathbf{1}\mathbf{1}^T) = P\mathbf{1}(\mathbf{1}^T p\mathbf{1})\mathbf{1}^T = 0.$$

Thus, any subset $I \subseteq [k-1]$ that may contribute to the sum in Equation (4.5) is contained in $\mathcal{F}_k$ as defined in Equation (4.1). Using that $\|P\| \le 1$ and the submultiplicativity of the euclidean norm, for every $I \in \mathcal{F}_k$ we have that

$$\begin{aligned} e_I &= \mathbf{1}^T \left(\prod_{i=1}^{k-1} PB_i(I) \right) P\mathbf{1} \\ &\le \prod_{i=1}^{k-1} \|PB_i(I)\| \\ &\le \prod_{i \in I} \|B_i(I)\|. \end{aligned}$$

Recall that for every $i \in I$, $B_i(I) = \lambda^{\Delta_i} E_{\Delta_i}$ and that $\|E_{\Delta_i}\| \le 1$. Thus,

$$\prod_{i \in I} \|B_i(I)\| \le \prod_{i \in I} \lambda^{\Delta_i},$$

which concludes the proof. □

4.2 Symmetric test functions

Given a symmetric function $f : \{\pm 1\}^t \to \mathbb{R}$ and $k \in [t]$ we slightly abuse notation and denote by $\widehat{f}(k) = |\widehat{f}([k])|$. For analyzing the random walk with respect to symmetric test functions, we define for every integer $k \in \{0, 1, \ldots, t\}$,

$$\beta_k = \sum_{\substack{S \subseteq [t] \\ |S|=k}} \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]. \tag{4.6}$$

Note that β_k is independent of the choice of test function. However, for symmetric tests functions, these quantities will appear in the analysis, and so we begin by analyzing them. Indeed, a straightforward corollary of Lemma 4.1 is the following

Corollary 4.3. *Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling of the vertices of G. Then, for every symmetric function $f : \{\pm 1\}^t \to \mathbb{R}$,*

$$\mathcal{E}_{G,\mathsf{val}}(f) \leq \sum_{k=2}^{t} \widehat{f}(k) |\beta_k|.$$

The main technical work in this section is proving the bound on $|\beta_k|$ as given by the following lemma.

Lemma 4.4. *Let G be a regular λ-spectral expander. Then, for every $k \in \{0, 1, \ldots, t\}$, it holds that*

$$|\beta_k| \;\leq\; 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda} \right)^{\lceil \frac{k}{2} \rceil} \tag{4.7}$$

To prove Lemma 4.4, we first prove the following claim.

Claim 4.5. *Let G be a regular λ-spectral expander. Then, for every $k \in \{0, 1, \ldots, t\}$, it holds that*

$$|\beta_k| \leq 2^k \sum_{m=\lceil \frac{k}{2} \rceil}^{t - \lfloor \frac{k}{2} \rfloor} \binom{m-1}{\lceil \frac{k}{2} \rceil - 1} \binom{t-m}{k - \lceil \frac{k}{2} \rceil} \lambda^m.$$

Proof of Claim 4.5. By Proposition 4.2, we have that

$$\begin{aligned} |\beta_k| &\leq \sum_{\substack{S \subseteq [t] \\ |S|=k}} \sum_{I \in \mathcal{F}_k} \lambda^{\sum_{j \in I} \Delta_j(S)} \\ &= \sum_{I \in \mathcal{F}_k} \sum_{\substack{S \subseteq [t] \\ |S|=k}} \lambda^{\sum_{j \in I} \Delta_j(S)}. \end{aligned}$$

Note that for every $S \subseteq [t]$ of size $|S| = k$ and every $j \in \{0, 1, \ldots, k\}$, $\Delta_j(S) \geq 1$. Moreover, for every such S, $\sum_{j=0}^{k} \Delta_j(S) = t + 1$. The encoding $S \mapsto (\Delta_0, \ldots, \Delta_k)$ is a bijection between cardinality k subsets of $[t]$ and partitions of $\{1, \ldots, t+1\}$ into $k+1$ non-empty intervals. Fix $I \in \mathcal{F}_k$ and let $\beta_k(I)$ denote the contribution of I to the above sum, that is, $\beta_k(I) = \sum_{S \subseteq [t] : |S|=k} \lambda^{\sum_{j \in I} \Delta_j(S)}$. Every cardinality k subset S contributes λ^m

to the sum, where m is the sum of lengths of the intervals indexed by $I \in \mathcal{F}_k$. To bound $\beta_k(I)$ we find for every $m \leqslant t$ the number of cardinality k sets S that contribute λ^m. Note that this is precisely the number of ways to choose positive integers $m_1, \ldots, m_{|I|}$ and $n_1, \ldots, n_{k+1-|I|}$ such that $\sum m_j = m$ and $\sum n_j = t - m + 1$, which is $\binom{m-1}{|I|-1}\binom{t-m}{k-|I|}$. Therefore,

$$\beta_k(I) \;=\; \sum_{m=0}^{t} \binom{m-1}{|I|-1}\binom{t-m}{k-|I|}\lambda^m.$$

Note that $\beta_k(I)$ depends only on the cardinality of I and not on I itself. Moreover, $\beta_k(I)$ is monotonically decreasing in $|I|$. To see this, notice that when $I \subseteq I'$ then for every $S \subseteq [t]$, $\sum_{j\in I} \Delta_j(S) \leq \sum_{j\in I'} \Delta_j(S)$, and therefore $\beta_k(I') \leq \beta_k(I)$ (because $\lambda \leq 1$). Thus, if I^* is a minimal cardinality set $\mathcal{F}_k$, then

$$\begin{aligned} |\beta_k| \;&\leq\; \sum_{I\in\mathcal{F}_k} \beta_k(I) \leq \sum_{I\in\mathcal{F}_k} \beta_k(I^*) \\ &\leq\; 2^k \sum_{m=0}^{t} \binom{m-1}{|I^*|-1}\binom{t-m}{k-|I^*|}\lambda^m \end{aligned}$$

The lemma follows by noting that for every $I \in \mathcal{F}_k$ we have $|I| \geq \lceil \frac{k}{2} \rceil$. $\square$

We turn to prove Lemma 4.4.

Proof of Lemma 4.4. The case $k = 1$ readily follows as, by Proposition 4.2 and the remark following it, $\beta_1 = 0$. By Claim 4.5,

$$\begin{aligned} |\beta_k| &\leq 2^k \sum_{m=\lceil \frac{k}{2}\rceil}^{t-\lfloor \frac{k}{2}\rfloor} \binom{m-1}{\lceil \frac{k}{2}\rceil - 1}\binom{t-m}{k-\lceil \frac{k}{2}\rceil}\lambda^m \\ &\leq 2^k \binom{t-1}{k-\lceil \frac{k}{2}\rceil} \cdot \lambda \cdot \sum_{m=\lceil \frac{k}{2}\rceil}^{t-\lfloor \frac{k}{2}\rfloor} \binom{m-1}{\lceil \frac{k}{2}\rceil - 1}\lambda^{m-1} \\ &\leq 2^k \binom{t-1}{k-\lceil \frac{k}{2}\rceil} \cdot \lambda \cdot \sum_{i=\lceil \frac{k}{2}\rceil-1}^{\infty} \binom{i}{\lceil \frac{k}{2}\rceil - 1}\lambda^{i}. \end{aligned} \tag{4.8}$$

By Claim 3.1,

$$\sum_{i=\lceil \frac{k}{2}\rceil-1}^{\infty} \binom{i}{\lceil \frac{k}{2}\rceil - 1}\lambda^{i} = \frac{\lambda^{\lceil \frac{k}{2}\rceil-1}}{(1-\lambda)^{\lceil \frac{k}{2}\rceil}}.$$

Substituting to Equation (4.8) concludes the proof. $\square$

4.3 Warm-up: analyzing the majority function

As a warm-up for the proof of Theorem 1.1, in this section we use the machinery developed in Section 4, namely, Proposition 4.2, Corollary 4.3, and Lemma 4.4, to prove that random walks fool the majority function. Recall that $\mathsf{MAJ}_t : \{\pm 1\}^t \to \{\pm 1\}$ on input $x = (x_1, \ldots, x_t) \in \{\pm 1\}^t$ is defined by $\mathsf{MAJ}_t(x) = 1$ if $\sum x_i \geq 0$ and $\mathsf{MAJ}_t(x) = -1$ otherwise. When t is clear from context, we omit it from the subscript. More generally, for $w \in [t]$ we define the w threshold function $\mathsf{Th}_w : \{\pm 1\}^t \to \{\pm 1\}$ by $\mathsf{Th}_w(x) = 1$ if $|\{x_i | x_i = 1\}| \geq w$ and $\mathsf{Th}_w(x) = -1$ otherwise.

Theorem 4.6. *There exists a universal constant c_{MAJ} such that for every $0 < \lambda < \frac{1}{4c_{\mathsf{MAJ}}^2}$ and every $t \in \mathbb{N}$*

$$\mathcal{E}_\lambda(\mathsf{MAJ}_t) \leq 2c_{\mathsf{MAJ}}^3 \left(\frac{\lambda}{1-\lambda}\right)^2 \cdot \frac{1}{\sqrt{t}}.$$

As explained in the introduction it is known [KV86, Lez01] that the distribution of the sum when taking t independent distributions and when taking a random walk, is $O(\frac{1}{\sqrt{t}})$ close in the Kolmogorov distance. This means that the two distributions look the same when the test function can be an arbitrary threshold function. More formally:

Theorem 4.7 (follows, e.g., from [Klo17], Theorem C)**.** *For every $t \in \mathbb{N}$ and $w \in [t]$,*

$$\mathcal{E}_\lambda(\mathsf{Th}_w) = O\left(\frac{1}{\sqrt{t}}\right).$$

Theorem 4.7 from [KV86] is more general than Theorem 4.6 that we prove in this section. However, the proof techniques are completely different. Theorem 4.7 holds only against threshold tests, while the proof of Theorem 4.6 builds upon the behaviour of the parity function that is far away from being threshold. We prove Theorem 4.6 as we believe it is a good warm-up exercise towards the more delicate calculations of Section 4.4. Indeed, in Section 4.4 we show this allows proving CLT convergence in the stronger total variation distance, rather than in the weaker Kolmagorov distance as in Theorem 4.7.

Proof of Theorem 4.6. Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling of V. The Fourier coefficients of the MAJ functions are well-known (see, e.g., [O'D14], Theorem 5.19). Let $S \subseteq [t]$ with $|S| = k$. Then, for k even $\widehat{\mathsf{MAJ}}(S) = 0$; otherwise,

$$\widehat{\mathsf{MAJ}}(S) = (-1)^{\frac{k-1}{2}} \frac{\binom{\frac{t-1}{2}}{\frac{k-1}{2}}}{\binom{t-1}{k-1}} \frac{2}{2^t} \binom{t-1}{\frac{t-1}{2}}. \tag{4.9}$$

Using Lemma 4.4, Equation (4.9) and the standard estimates

$$\frac{2}{2^t}\binom{t-1}{\frac{t-1}{2}} \le \frac{c_1}{\sqrt{t}},$$
$$\binom{\frac{t-1}{2}}{\frac{k-1}{2}} \le \left(e\cdot\frac{t-1}{k-1}\right)^{\frac{k-1}{2}},$$
$$\binom{t-1}{k-1} \ge \left(\frac{t-1}{k-1}\right)^{k-1},$$

where c_1 is some absolute constant (see Claim 3.1) we get for every odd $k \in [t]$,

$$\begin{aligned}
|\widehat{\mathsf{MAJ}}(k)\cdot\beta_k| &\le \frac{\binom{\frac{t-1}{2}}{\frac{k-1}{2}}}{\binom{t-1}{k-1}}\frac{2}{2^t}\binom{t-1}{\frac{t-1}{2}}\cdot 2^k\binom{t-1}{\frac{k-1}{2}}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k+1}{2}}\\
&\le \left(e\cdot\frac{t-1}{k-1}\right)^{\frac{k-1}{2}}\cdot\frac{c_1}{\sqrt{t}}\cdot\left(\frac{k-1}{t-1}\right)^{k-1}\cdot 2^k\cdot\left(2e\frac{t-1}{k-1}\right)^{\frac{k-1}{2}}\cdot\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k+1}{2}}\\
&= \frac{c_1}{\sqrt{t}}\cdot e^{k-1}\cdot 2^{k+\frac{k-1}{2}}\cdot\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k+1}{2}}\\
&\leqslant \frac{{c_2}^k}{\sqrt{t}}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k+1}{2}}.
\end{aligned}$$

Corollary 4.3 then implies that

$$\begin{aligned}
\mathcal{E}_{G,\mathsf{val}}(\mathsf{MAJ}) &\le \sum_{\substack{k=3\\k\text{ odd}}}^{t}|\widehat{\mathsf{MAJ}}(k)\cdot\beta_k|\\
&\le \sum_{\substack{k=3\\k\text{ odd}}}^{t}\frac{{c_2}^k}{\sqrt{t}}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k+1}{2}}\\
&= \frac{c_2^3}{\sqrt{t}}\left(\frac{\lambda}{1-\lambda}\right)^2\cdot\sum_{i=0}^{\infty}\left(c_2^2\frac{\lambda}{1-\lambda}\right)^i. \qquad (4.10)
\end{aligned}$$

Set $c_{\mathsf{MAJ}} = \max(c_2, 1)$. As $\lambda \le \frac{1}{4c_{\mathsf{MAJ}}^2}$ we have that $\lambda \le \frac{1}{4}$ and so

$$c_2^2\frac{\lambda}{1-\lambda} \le \frac{4}{3}\cdot c_2^2\lambda \le \frac{4}{3}\cdot c_{\mathsf{MAJ}}^2\lambda \le \frac{1}{3}.$$

Thus, the sum in Equation (4.10) is bounded by $\frac{3}{2}$, and we conclude that

$$\mathcal{E}_{G,\mathsf{val}}(\mathsf{MAJ}) \le \frac{2c_{\mathsf{MAJ}}^3}{\sqrt{t}}\left(\frac{\lambda}{1-\lambda}\right)^2.$$

$\square$

4.4 Weight indicator functions

For integers t and $w \in \{0, 1, \dots, t\}$ let $\mathbf{1}_w : \{\pm 1\}^t \to \{0, 1\}$ be the function indicating whether the weight of the input is w. That is, $\mathbf{1}_w(x_1, \dots, x_t) = 1$ if $\sum_i x_i = 1$ and $\mathbf{1}_w(x_1, \dots, x_t) = 0$ otherwise. In this section we prove

Theorem 4.8. *There exists universal constants $0 < \gamma \le 1 \le c$ such that the following holds. Let $1 \leqslant \sigma_0 \in \mathbb{R}$ and $0 \le b \le \sigma_0 \sqrt{t}$ an integer. Set $w = \frac{t+b}{2}$. Then, for any $\lambda \le \frac{\gamma}{\sigma_0^2}$ it holds that*

$$\mathcal{E}_\lambda(\mathbf{1}_w) \le \frac{c\lambda}{1-\lambda} \cdot \frac{\sigma_0^2}{\sqrt{t}}.$$

We analyze the weight indicator function in a similar way to the majority function, except that we need to work harder to express the Fourier coefficients of the weight function, and, more importantly, the analysis is more delicate as the weight indicator function is not anti-symmetric and therefore has Fourier mass on even layers. This section is organized as follows: in Section 4.4.1 we compute the Fourier coefficients of $\mathbf{1}_w$ and in Section 4.4.2 we prove Theorem 4.8.

4.4.1 The Fourier coefficients of $\mathbf{1}_w$

In this section we compute the Fourier coefficients of the weight indicator function. While this calculation is certainly known, we could not find a reference and so we develop it here.

Lemma 4.9.

$$\widehat{\mathbf{1}_w}(k) = \begin{cases} \frac{1}{2^t} \frac{\binom{t}{w}}{\binom{t}{k}} \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} (-1)^l \binom{t-w}{l} \binom{2w-t}{k-2l} & w \geqslant \frac{t}{2} \\ \frac{1}{2^t} \frac{\binom{t}{w}}{\binom{t}{k}} \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} (-1)^{k-l} \binom{w}{l} \binom{t-2w}{k-2l} & w \leqslant \frac{t}{2} \end{cases}$$

Proof. We compute the quantity $T_\rho \mathbf{1}_w(1, \dots, 1)$ in two ways. First, we apply Lemma 3.4 to $\mathbf{1}_w$ on input $x_0 = (1, \dots, 1)$, where we think of ρ as a formal symbol. Using the fact that $\mathbf{1}_w$ is symmetric, we get that

$$(T_\rho \mathbf{1}_w)(x_0) = \sum_{S \subseteq [t]} \widehat{\mathbf{1}_w}(S) \rho^{|S|} \chi_S(x_0) = \sum_{k=0}^{t} \binom{t}{k} \widehat{\mathbf{1}_w}(k) \rho^k. \tag{4.11}$$

On the other hand, by direct computation

$$\begin{aligned} (T_\rho \mathbf{1}_w)(x_0) &= \mathop{\mathbf{E}}_{y_i \sim \mathrm{Ber}(\frac{1+\rho}{2})} \mathbf{1}_w(y) \\ &= \frac{1}{2^t} \binom{t}{w} (1+\rho)^w (1-\rho)^{t-w} \\ &= \begin{cases} \frac{1}{2^t} \binom{t}{w} (1-\rho^2)^{t-w} (1+\rho)^{2w-t} & w \geqslant \frac{t}{2} \\ \frac{1}{2^t} \binom{t}{w} (1-\rho^2)^{w} (1-\rho)^{t-2w} & w \leqslant \frac{t}{2}. \end{cases} \end{aligned} \tag{4.12}$$

Equations (4.11) and (4.12) give two equal polynomial expressions in the formal symbol ρ and so their corresponding coefficients are equal. Comparing the coefficient of ρ^k, we learn that when $w \geqslant \frac{t}{2}$

$$\binom{t}{k}\widehat{\mathbf{1}_w}(k) \;=\; \frac{1}{2^t}\binom{t}{w}\sum_{\ell=0}^{\lfloor \frac{k}{2} \rfloor}(-1)^{\ell}\binom{t-w}{\ell}\binom{2w-t}{k-2\ell},$$

and for $w < \frac{t}{2}$

$$\binom{t}{k}\widehat{\mathbf{1}_w}(k) \;=\; \frac{1}{2^t}\binom{t}{w}\sum_{\ell=0}^{\lfloor \frac{k}{2} \rfloor}(-1)^{k-\ell}\binom{w}{\ell}\binom{t-2w}{k-2\ell}.$$

□

Using Lemma 4.9 we turn to bound the magnitude of the weight indicator Fourier coefficients.

Claim 4.10. *Let $1 \leqslant \sigma_0 \in \mathbb{R}$ and $0 \leq b \leq \sigma_0\sqrt{t}$ an integer. Then there is some constant $c_2 > 0$ such that for $w = \frac{t+b}{2}$ and $w' = t - w$ it holds that*

$$|\widehat{\mathbf{1}_w}(k)|, |\widehat{\mathbf{1}_{w'}}(k)| \;\leq\; \frac{k\cdot(c_2\sigma_0)^k}{\sqrt{t}}\left(\frac{k}{t}\right)^{k/2}$$

Proof. By symmetry, it is enough bound $|\widehat{\mathbf{1}_w}(k)|$. Denote $\sigma = \frac{b}{\sqrt{t}}$, and note that $\sigma \leq \sigma_0$. By Lemma 4.9,

$$\begin{aligned}
|\widehat{\mathbf{1}_w}(k)| &= \left|\frac{1}{2^t}\frac{\binom{t}{\frac{t+\sigma\sqrt{t}}{2}}}{\binom{t}{k}}\sum_{\ell=0}^{\lfloor \frac{k}{2} \rfloor}(-1)^{\ell}\binom{\frac{t-\sigma\sqrt{t}}{2}}{\ell}\binom{\sigma\sqrt{t}}{k-2\ell}\right| \\
&\leq \frac{1}{2^t}\frac{\binom{t}{\frac{t}{2}}}{\binom{t}{k}}\cdot\sum_{\ell=0}^{\lfloor \frac{k}{2} \rfloor}\binom{\frac{t-\sigma\sqrt{t}}{2}}{\ell}\binom{\sigma\sqrt{t}}{k-2\ell} \\
&\leq \frac{c_1 k}{\sqrt{t}}\left(\frac{k}{t}\right)^{k}\cdot\max_{\ell}\binom{\frac{t-\sigma\sqrt{t}}{2}}{\ell}\binom{\sigma\sqrt{t}}{k-2\ell}. \qquad (4.13)
\end{aligned}$$

However,

$$\begin{aligned}
\binom{\frac{t-\sigma\sqrt{t}}{2}}{\ell}\binom{\sigma\sqrt{t}}{k-2\ell} &\leq \binom{t}{\ell}\binom{\sigma_0\sqrt{t}}{k-2\ell} \\
&\leq \left(\frac{et}{\ell}\right)^{\ell}\left(\frac{e\sigma_0\sqrt{t}}{k-2\ell}\right)^{k-2\ell} \\
&\leq e^{k-\ell}t^{k/2}\frac{1}{\ell^{\ell}}\cdot\left(\frac{\sigma_0}{k-2\ell}\right)^{k-2\ell} \\
&\leq t^{k/2}\,(e\sigma_0)^k\,\frac{1}{\ell^{\ell}}\frac{1}{(k-2\ell)^{k-2\ell}} \\
&\leq t^{k/2}(e\sigma_0)^k\cdot\max_{1<\ell<\frac{k}{2}}\left(\frac{1}{\ell^{\ell}}\frac{1}{(k-2\ell)^{k-2\ell}}\right). \qquad (4.14)
\end{aligned}$$

We have the following claim whose proof we defer.

Claim 4.11. *Let* $1 < \ell < \frac{k}{2}$. *Then*

$$\ell^\ell \cdot (k-2\ell)^{k-2\ell} \geq \frac{k^{k/2}}{3^k}.$$

Claim 4.11 together with Equations (4.13) and (4.14) imply that

$$\begin{aligned}
|\widehat{\mathbf{1}_w}(k)| &\leq \frac{c_1 k}{\sqrt{t}} \left(\frac{k}{t}\right)^k t^{k/2} (e\sigma_0)^k \frac{3^k}{k^{k/2}} \\
&\leq \frac{k}{\sqrt{t}} \left(\frac{k}{t}\right)^{k/2} (3ec_1\sigma_0)^k.
\end{aligned}$$

The proof follows by taking $c_2 = 3ec_1$. □

We complete the proof of Claim 4.10 by proving Claim 4.11.

Proof of Claim 4.11. Define $f(x) = x^x \cdot (k-2x)^{k-2x} = e^{x\ln(x)+(k-2x)\ln(k-2x)}$. Note that

$$f(1) = (k-2)^{k-2}$$
$$f\left(\frac{k}{2}-1\right) = 2^2 \left(\frac{k-2}{2}\right)^{\frac{k-2}{2}} \geq \frac{k^{k/2}}{3^k}.$$

The extreme points of $f(x)$ in the interior of the domain are the extreme points of $g(x) = \ln(f(x)) = x\ln(x) + (k-2x)\ln(k-2x)$. If x is an extreme point in the interior then $g'(x) = 0$. As,

$$g'(x) = \ln(x) + 1 - 2(\ln(k-2x)+1) = \ln x - 2\ln(k-2x) - 1,$$

the point x can be an interior extreme point only if $\ln(x) - 2\ln(k-2x) = 1$, i.e., $x = e(k-2x)^2$. The two solutions to the equation $4ex^2 - (4ek+1)x + ek^2 = 0$ are given by

$$\begin{aligned}
x_{1,2} &= \frac{4ek+1 \pm \sqrt{(4ek+1)^2 - 16e^2k^2}}{8e} \\
&= \frac{4ek+1 \pm \sqrt{1+8ek}}{8e} \\
&= \frac{k}{2} + \frac{1 \pm \sqrt{1+8ek}}{8e}.
\end{aligned}$$

The positive solution lays outside of the domain. Thus, the only possible extremum in the interior is

$$x_1 = \frac{k}{2} + \frac{1-\sqrt{1+8ek}}{8e} \tag{4.15}$$

Notice that $x_1 = e(k-2x_1)^2$. Denote by $a = x_1$, $b = (k-2x_1)$, thus $k = 2a+b$ and $a = eb^2$. The candidate extremum is

$$\begin{aligned} g(x_1) &= a\ln a + b\ln b \\ &= a\ln(eb^2) + b\ln b \\ &= a(2\ln(b)+1) + b\ln b \\ &= \ln(b)(2a+b) + a = k\ln(b) + a. \end{aligned}$$

And so $f(x_1) = e^{g(x_1)} = b^k \cdot e^a \geq b^k$. We therefore want to lower bound $b = k - 2x_1$. By Equation (4.15) we have that

$$x_1 \leq \frac{k}{2} + \frac{1-\sqrt{8ek}}{8e} = \frac{k}{2} - \sqrt{\frac{k}{8e}} + \frac{1}{8e}.$$

It follows that

$$b = k - 2x_1 \geq \sqrt{\frac{k}{2e}} - \frac{1}{4e}.$$

As $k \geq 1$, we have that

$$b \geq \sqrt{k}\left(\frac{1}{\sqrt{2e}} - \frac{1}{4e}\right) \geq \frac{\sqrt{k}}{3},$$

and so $f(x_1) \geq b^k = \frac{k^{k/2}}{3^k}$ as desired. □

4.4.2 Proof of Theorem 4.8

We are now ready to prove Theorem 4.8.

Proof of Theorem 4.8. Let G be a regular λ-spectral expander and let $\mathsf{val} : V \to \{\pm 1\}$ be a balanced labelling. By Corollary 4.3,

$$\mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) \leq \sum_{k=2}^{t} |\widehat{\mathbf{1}_w}(k)\beta_k|.$$

Using Claim 4.10 to upper bound $|\widehat{\mathbf{1}_{wW}}|$ and Lemma 4.4 to bound $|\beta_k|$ we get,

$$\begin{aligned} \mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) &\leq \sum_{k=2}^{t} |\widehat{\mathbf{1}_w}(k)| \cdot |\beta_k| \\ &\leq \sum_{k=2}^{t} k\frac{(c_2\sigma_0)^k}{\sqrt{t}}\left(\frac{k}{t}\right)^{k/2} \cdot 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor}\left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \\ &\leq \sum_{k=2}^{t} k\frac{(c_2\sigma_0)^k}{\sqrt{t}}\left(\frac{k}{t}\right)^{k/2} \cdot 2^k \left(\frac{2et}{k}\right)^{k/2}\left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \\ &= \sum_{k=2}^{t} \frac{(3c_2\sigma_0)^k}{\sqrt{t}}\left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \\ &\leq \frac{(3c_2\sigma_0)^2}{\sqrt{t}} \cdot \frac{\lambda}{1-\lambda} \cdot \sum_{i=0}^{\infty}\left(3c_2\sigma_0\sqrt{\frac{\lambda}{1-\lambda}}\right)^i. \end{aligned}$$

The proof follows by taking $3c_2\sigma_0\sqrt{\frac{\lambda}{1-\lambda}} < \frac{1}{2}$.

$\square$

4.5 Proof of Theorem 1.3

For convenience we restate the theorem here.

Theorem (Theorem 1.3; restated)**.** *Let $G = (V, E)$ be a λ-spectral expander. Let $\mathsf{val} : V \to \{0, 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = \frac{1}{2}$. Then,*

$$\| \, \Sigma\mathrm{RW}_t - \Sigma\mathrm{Ind}_t \, \|_{TVD} = O(\lambda \cdot \log^{3/2}(1/\lambda)).$$

Proof. Note that it suffices to prove the theorem only for $\lambda < \lambda_0$, where $\lambda_0 < 1$ is some constant. Indeed, this can be incorporated to the hidden constant factor in the big-O notation that appears in the bound. We have that,

$$\| \, \Sigma\mathrm{RW}_t - \Sigma\mathrm{Ind}_t \, \|_{\mathrm{TVD}} = \frac{1}{2}\sum_{w=0}^{t} \left|\mathbf{Pr}[\Sigma\mathrm{Ind}_t = w] - \mathbf{Pr}[\Sigma\mathrm{RW}_t = w]\right| .$$

Let $\sigma_0 \in \mathbb{R}$ be a parameter to be chosen later. For the proof, we split the domain into two different intervals, based on σ_0. The central interval $I_C = \left\{w \mid |w - \frac{t}{2}| \leqslant \frac{\sigma_0}{2}\sqrt{t}\right\}$ and the tails, $I_T = \left\{w \mid |w - \frac{t}{2}| > \frac{\sigma_0}{2}\sqrt{t}\right\}$. First notice that both $\Sigma\mathrm{RW}_t$, $\Sigma\mathrm{Ind}_t$ have a very small probability to enter the tails region. Indeed, by the Chernoff bound,

$$\begin{aligned}
\mathbf{Pr}[\Sigma\mathrm{Ind}_t \in I_T] &= \mathbf{Pr}\left[\Sigma\mathrm{Ind}_t \geqslant \left(1 + \frac{\sigma_0}{\sqrt{t}}\right)\frac{t}{2}\right] + \mathbf{Pr}\left[\Sigma\mathrm{Ind}_t \leqslant \left(1 - \frac{\sigma_0}{\sqrt{t}}\right)\frac{t}{2}\right] \\
&\leqslant 2\exp\left(-\left(\frac{\sigma_0}{\sqrt{t}}\right)^2 \frac{t}{6}\right) \\
&= 2e^{-\sigma_0^2/6}.
\end{aligned}$$

By the Chernoff bound for expander walks, and assuming $\lambda_0 \leq \frac{1}{2}$,

$$\begin{aligned}
\mathbf{Pr}[|\Sigma\mathrm{RW}_t \in I_T] &= \mathbf{Pr}\left[\left|\Sigma\mathrm{RW}_t - \frac{t}{2}\right| \geqslant \frac{\sigma_0}{2\sqrt{t}} \cdot t\right] \\
&< 2\exp\left(-\frac{(1-\lambda)t}{4}\left(\frac{\sigma_0}{2\sqrt{t}}\right)^2\right) \\
&= 2e^{-\sigma_0^2/32}.
\end{aligned}$$

Combining those two results we get a bound on the total variation distance in the tails region.

$$\begin{aligned}
\sum_{w \in I_T} |\, \mathbf{Pr}[\Sigma\mathrm{Ind}_t = w] - \mathbf{Pr}[\Sigma\mathrm{RW}_t = w]| &\leqslant \sum_{w \in I_T} \mathbf{Pr}[\Sigma\mathrm{Ind}_t = w] + \mathbf{Pr}[\Sigma\mathrm{RW}_t = w]| \\
&\leqslant 4e^{-\sigma_0^2/32}. \qquad (4.16)
\end{aligned}$$

In the central interval, we invoke Theorem 4.8 to obtain

$$\sum_{w\in I_C} |\,\mathbf{Pr}[\Sigma\mathrm{Ind}_t = w] - \mathbf{Pr}[\Sigma\mathrm{RW}_t = w]| \leqslant \sum_{w\in I_C} \mathcal{E}_\lambda(\mathbf{1}_w) \leqslant \sum_{w\in I_C} c\lambda\frac{\sigma_0^2}{\sqrt{t}} = c\lambda\sigma_0^3, \tag{4.17}$$

where c is the constant that appears in the statement of Theorem 4.8. Set $\sigma_0 = \sqrt{32\ln\frac{1}{\lambda}}$. Note that by choosing λ_0 sufficiently small so that $32\lambda_0\ln\frac{1}{\lambda_0} \leq \gamma$, where γ is the constant from Theorem 4.8, this meets the requirement of Theorem 4.8. Thus, we obtain that the bound in Equation (4.16) evaluates to 4λ and the bound in Equation (4.17) is $O(\lambda\log^{3/2}\frac{1}{\lambda})$. Combining both bounds concludes the proof. □

5 Beyond symmetric functions

Several natural computational classes such as $\mathbf{AC}^0$ circuits, read-once branching programs of various forms and functions with bounded query complexity are known to have bounded Fourier tails. In many cases, such tails are key to our understanding of these classes.

Definition 5.1. *For an integer $t \geq 1$ and $b \geq 1$, we denote by $\mathcal{L}_1^t(b)$ the family of functions $f : \{\pm1\}^t \to \{\pm1\}$ that satisfy*

$$L_{1,k}(f) \triangleq \sum_{\substack{S\subseteq[t]\\|S|=k}} |\widehat{f}(S)| \leq b^k.$$

When t is clear from context we omit it and write $\mathcal{L}_1(b)$. Most works consider the L_2 norm. In the following we focus on the L_1 norm as it is known that a bound on the L_2 norm implies a bound on the L_1 norm [Tal17]. Thus, the class of functions with L_1 bounded Fourier tails is richer. We turn to give some examples.

Bounded-depth circuits. The class of bounded-depth circuits has been widely studied. The seminal work by Linial, Mansour and Nisan [LMN93] gives a bound on the L_2 Fourier tail for this class. Tal [Tal17] obtained an improved result by showing that a function computed by a depth-d size-s circuit is contained in $\mathcal{L}_1(b)$ for $b = O(\log^{d-1} s)$.

Read-once branching programs. The class of ROBP is of wide interest, motivated mostly by the study of the **BPL** vs. **L** problem. Reingold, Steinke and Vadhan [RSV13] proved that any function $f : \{\pm1\}^t \to \{\pm1\}$ computed by a width-w *permutation* ROBP is in $\mathcal{L}_1(2w^2)$. They further conjectured a bound for general ROBP. Their conjecture was settled by Chattopadhyay *et al.* [CHRT18], who proved that any function $f : \{\pm1\}^t \to \{\pm1\}$ computed by a width-w ROBP is in $\mathcal{L}_1(b)$ for $b = O(\log^w n)$. Both results hold in the more general setting where the bits can be read in any (predetermined) order.

Query complexity measures. Denote by $\mathrm{DT}(f)$ the decision tree complexity of f. It is easy to show that $L_{1,k}(f) \le \mathrm{DT}(f)^k$ and so the class of functions with decision tree complexity d is in $\mathcal{L}_1(d)$. As mentioned in the introduction, it is well-known that the decision tree complexity is polynomially-related to other complexity measures such as the randomized and quantum decision tree measures, the certificate query complexity (namely, nondeterministic query complexity), the approximate real degree of a function, and most recently also to the sensitivity of a function [Hua19]. Thus, every function with a bound b on any one of these measures is in $\mathcal{L}_1(b^c)$ for some universal constant $c \ge 1$.

In this section we prove Theorem 1.5. To start with, we prove

Claim 5.2. *For every $\lambda \in [0,1]$ and function $f : \{\pm 1\}^t \to \{\pm 1\}$,*

$$\mathcal{E}_\lambda(f) \le 4\mathbf{\Lambda}_{2\sqrt{\lambda}}(f).$$

Proof. Let $G = (V,E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a balanced function. As $\mathbf{E}\, f(U_t) = \widehat{f}(\emptyset)$,

$$\begin{aligned} \mathcal{E}_{G,\mathsf{val}}(f) &= |\,\mathbf{E}\, f(\mathrm{RW}_{G,\mathsf{val}}) - \mathbf{E}\, f(U_t)| \\ &\le \sum_{\emptyset \ne S \subseteq [t]} |\widehat{f}(S)||\,\mathbf{E}\,\chi_S(\mathrm{RW}_{G,\mathsf{val}})| \\ &\le \sum_{\substack{S \subseteq [t] \\ |S| \ge 2}} |\widehat{f}(S)| \cdot 2^{|S|}\lambda^{\mathbf{\Delta}(S)/2}. \end{aligned}$$

where the last inequality follows by Proposition 2.3. Now, $\mathbf{\Delta}(S) \ge |S| - 2$ and so

$$\mathcal{E}_{G,\mathsf{val}}(f) \le 4 \sum_{\substack{S \subseteq [t] \\ |S| \ge 2}} |\widehat{f}(S)| \cdot (2\sqrt{\lambda})^{\mathbf{\Delta}(S)},$$

proving the corollary. □

The proof of Theorem 1.5 readily follows by the above-mentioned results [RSV13, Tal17, CHRT18] and the following claim.

Claim 5.3. *There exists a universal constant $c \ge 1$ such that the following holds. For every function $f : \{\pm 1\}^t \to \{\pm 1\}$ in $\mathcal{L}_1(b)$ and $\varepsilon > 0$, it holds that $\mathcal{E}_\lambda(f) \le \varepsilon$ provided $\lambda \le \frac{\varepsilon^2}{cb^4}$.*

Proof. Let $G = (V,E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a balanced function. By Claim 5.2,

$$\begin{aligned} \mathcal{E}_{G,\mathsf{val}}(f) &\le 4\mathbf{\Lambda}_{2\sqrt{\lambda}}(f) \\ &= 4 \sum_{\substack{S \subseteq [t] \\ |S| \ge 2}} |\widehat{f}(S)| \cdot (2\sqrt{\lambda})^{\mathbf{\Delta}(S)}. \end{aligned}$$

Consider a set S of size $|S| = k$. Recall that for $k = 2, 3$ we have that $\mathbf{\Delta}(S) \geq k - 1$, and that for $k \geq 4$, it holds that $\mathbf{\Delta}(S) \geq k - 2$. Bounding the above sum according to the set size, we get

$$\begin{aligned}\mathcal{E}_{G,\mathsf{val}}(f) &\leq 4\left((2\sqrt{\lambda})b^2 + (2\sqrt{\lambda})^2 b^3 + \sum_{k=4}^{t}(2\sqrt{\lambda})^{k-2} b^k\right) \\ &\leq 8\sqrt{\lambda}b^2 + 16\lambda b^3 + \frac{16\lambda b^4}{1 - 2\sqrt{\lambda}b}.\end{aligned}$$

It is straightforward to verify that the above is bounded by ε for a sufficiently large constant c. □

6 A CS proof for the expander CLT

In a chain of works the Berry-Esseen theorem was proved for very general stochastic processes, including, as a very special case, random walks on graphs. We give another proof, specalized to random walks on graphs, in this section.

We adopt the approach of Kloeckner [Klo17] and reduce the Berry-Esseen theorem for $\Sigma\mathrm{RW}_t$ to estimates on the distance between the characteristic functions of $\Sigma\mathrm{RW}_t$ and $\Sigma\mathrm{Ind}_t$, and then use a lemma (e.g., from [Fel71]) that relates the Kolmogorov distance to the distance between characteristic functions. Kloeckner [Klo17] prove the distance between the characteristic functions in great generality. Instead, we do it specifically for random walks on graphs. We adopt a variant of Gilman [Gil98b] (with the simplification and notation suggested by Heally [Hea08] and Beck [Bec15]) to achieve this task.

The proof in this section reveals the similarity between the Expander Chernoff bound that studies the negligible part of the distribution, and the Expander CLT that studies the non-negligible part of the distribution, and put both proofs on the same ground, using the same tools and philosophy. It also highlights the phenomenon that already exist in the Chernoff bound and the CLT for independent processes, where the moment generating function is used to bound the negligible part, and the related characteristic function is used for studying the non-negligible part.

The result proved in this specific section is known in the mathematical community and in more generality. Nevertheless we made the choice to include the proof in the paper because we believe the new proof and our exposition make it more accessible (and intuitive!) to the CS audience. We also believe that such a proof has a potential to interact better with possible CS applications.

6.1 The reduction from Berry-Esseen to the characteristic function

The moment generating function of X is the function $M_x : \mathbb{R} \to \mathbb{R}$ defined by $M_X(t) = \mathbf{E}[e^{tX}]$. The characteristic function of X is the function $\varphi_x : \mathbb{R} \to \mathbb{C}$ defined by $\varphi_X(t) =$

$\mathbf{E}[e^{itX}]$. I.e., φ_X is the Fourier transform of the density function of X. Let $G = (V, E)$ be a λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a balanced labelling. We define three distributions:

- N - the normal distribution with mean 0 and variance 1. The CDF is denoted by $\Phi(x)$. The characteristic polynomial of N is $\gamma(b) = e^{-b^2/2}$.
- B_t - the distribution obtained by taking t random samples V_i from V, letting $Y_i = \mathsf{val}(V_i)$ and $B_t = \frac{1}{\sqrt{t}} \sum_{i=1}^{t} Y_i$. We denote the characteristic polynomial by $\beta(b)$. $\beta(b)$ is easy to describe:
$$\begin{aligned}\beta(b) &= \mathbf{E}\left[e^{ib\frac{1}{\sqrt{t}}\sum Y_i}\right] = \prod_i \mathbf{E}\left[e^{\frac{ib}{\sqrt{t}}Y_i}\right] \\ &= \prod_i \frac{e^{\frac{ib}{\sqrt{t}}} + e^{\frac{-ib}{\sqrt{t}}}}{2} = \left(\cos\left(\frac{b}{\sqrt{t}}\right)\right)^t.\end{aligned}$$
- R_t - the distribution obtained by taking a length $t-1$ random walk over G, visiting vertices $(v_1, \ldots, v_t)$ and letting $R_t = \frac{1}{\sqrt{t}} \sum_{j=1}^{t} \mathsf{val}(v_j)$. We denote the characteristic function of R_t by $\rho(b)$.

Our main technical claim is that the characteristic functions of R_t and B_t are close. Specifically,

Lemma 6.1. *Suppose $\lambda \le \frac{1}{16}$ and $b \in [-B, B]$ for $B = \frac{\pi}{3}\sqrt{t}$. Then,*
$$|\rho(b) - \beta(b)| \le 16\lambda \cdot \frac{b^2}{t} \cdot e^{-\frac{b^2}{16}}.$$

With that we claim:

Theorem 6.2. *(Berry Esseen-for random walk on expander graphs)*
$$\| R_t - N \|_{\text{Kol}} \le O\left(\frac{1}{\sqrt{t}}\right).$$

Proof. We apply [Fel71, Lemma 2, Chapter XVI.3] to get:

$$\begin{aligned}\| R_t - N \|_{\text{Kol}} = \| R - \Phi \|_\infty &\le \frac{1}{\pi}\int_{-B}^{B} \left|\frac{\rho(x) - \gamma(x)}{x}\right| dx + \frac{24}{\sqrt{2\pi}\pi}\frac{1}{B} \\ &\le \Theta\left(\frac{1}{B}\right) + \int_{-B}^{B} \left|\frac{\rho(x) - \beta(x)}{x}\right| dx + \int_{-B}^{B} \left|\frac{\beta(x) - \gamma(x)}{x}\right| dx.\end{aligned}$$

It is well known that $\int_{-B}^{B} \left|\frac{\beta(x)-\gamma(x)}{x}\right| dx = O(\frac{1}{B})$ because the binomial distribution converges in CDF to the normal distribution (see [Fel71], page 543, in the proof of Theorem 1 in Section XVI.5). Also, from Lemma 6.1 we see:
$$\int_{-B}^{B} \left|\frac{\rho(x) - \beta(x)}{x}\right| dx \le \frac{16\lambda}{t} \int_0^B x e^{-x^2/16} dx \le \frac{2^7\lambda}{t},$$
because $\int_0^B x e^{-x^2/16} = (-8e^{-x^2/16})|_0^B \le 8$. The two bounds together conclude the proof. $\square$

6.2 Flow

Our ultimate goal is to prove Lemma 6.1. We can define P_b to be a diagonal matrix with

$$P_b[v,v] = e^{\mathsf{val}(v)\cdot\frac{b}{\sqrt{t}}\cdot i}.$$

Then, as in previous sections we can express

$$\rho(b) \quad = \quad \mathbf{E}\, e^{ibR_t} = \mathbf{1}^t (P_b G)^{t-1} P_b \mathbf{1}.$$

P_b and G are linear operators over $\mathbb{R}^V$. We let $\mathbf{V} = \left\{\sum_{v\in V} \alpha_v \cdot v \mid v \in V\right\}$. We identify $\mathbb{R}^V$ with $\mathbf{V}$ by associating $f : V \to \mathbb{R}$ with the vector $\sum_{v\in V} f(v)\cdot v$. We decompose $\mathbf{V} = \mathbf{V}_0 \oplus \mathbf{V}_1$ where $\mathbf{V}_0 = \mathrm{Span}\,\{\mathbf{1}\}$ is the "parallel" space and $\mathbf{V}_1 = \mathbf{V}_0^{\perp}$ is the space perpendicular to it. Following Gillman, Heally and Beck, and using the notation and claims of Beck, we want to measure the amount of "flow" from $\mathbf{V}_0$ to $\mathbf{V}_1$ in a linear operator $T : \mathbf{V} \to \mathbf{V}$. We therefore define:

Definition 6.3. *(The flow matrix) Given $T : \mathbf{V} \to \mathbf{V}$ let $T_{i,j} : \mathbf{V} \to \mathbf{V}$ be*

$$T_{i,j} = \Pi_i T \Pi_j,$$

where Π_k ($k \in \{0,1\}$) is the projection operator on $\mathbf{V}_k$. We define the flow matrix *of T to be the 2×2 matrix $\widetilde{T}$ defined by*

$$\widetilde{T}[i,j] \quad = \quad \|\, T_{i,j} \,\|\,.$$

$\widetilde{T}$ measures the amount of flow from $\mathbf{V}_0$ (or $\mathbf{V}_1$) to $\mathbf{V}_0$ and $\mathbf{V}_1$.

It is a standard check that

$$\widetilde{G} \quad \le_{ew} \quad \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix},$$

where all entries except for the right bottom are equality. Also,

Claim 6.4. $\widetilde{P_b} \le_{ew} \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & 1 \end{pmatrix}$.

Proof. For $\widetilde{P_b}[0,0]$,

$$\mathbf{1}^t P_b \mathbf{1} \quad = \quad \frac{1}{t}\sum_v e^{\mathsf{val}(v)\cdot\frac{b}{\sqrt{t}}\cdot i} \quad = \quad \frac{1}{2}\left(e^{\frac{b}{\sqrt{t}}i} + e^{-\frac{b}{\sqrt{t}}i}\right) = \cos\left(\frac{b}{\sqrt{t}}\right) = \cos(\theta).$$

As P_b is unitary,

$$\widetilde{P_b}[0,0]^2 + \widetilde{P_b}[0,1]^2 = \|\,\Pi_0(P_b\mathbf{1})\,\|^2 + \|\,(\Pi_1 P_b \mathbf{1})\,\|^2 = \|\,P_b\mathbf{1}\,\| = \|\,\mathbf{1}\,\| = 1,$$

which gives

$$\widetilde{P_b}[0,1] \quad = \quad \widetilde{P_b}[1,0] \quad = \quad \sqrt{1-\widetilde{P_b}[0,0]^2}.$$

Finally, clearly, $\widetilde{P_b}[1,1] \le \|\,U\,\| = 1$ as U is unitary. Notice that the entries except for the right bottom are equalities. $\square$

Furthermore, say $M_1 \leq_{ew} M_2$, where M_1, M_2 are 2×2 matrices over $\mathbb{R}$, if $M_1[i,j] \leq M_2[i,j]$ for all i, j, where $\leq_{eq}$ stands for entry-wise inequality. Then,

Fact 6.5. *Let $A, B : \mathbf{V} \to \mathbf{V}$ be arbitrary linear operators. $\widetilde{AB} \leq_{ew} \widetilde{A} \cdot \widetilde{B}$, where the product on the RHS is 2×2 matrix multiplication.*

Proof. We remind the reader that $\widetilde{AB}[i,j] = \| \Pi_i AB \Pi_j \|$, where $\Pi_0 = \mathbf{1}\mathbf{1}^t$ the projection matrix on $\mathbf{1}$ and $\Pi_1 = I - \Pi_0$.

$$\begin{aligned}
\widetilde{AB}[i,j] &= \| \Pi_i AB \Pi_j \| \\
&= \| \Pi_i A(\Pi_0 + \Pi_1) B \Pi_j \| \\
&\leq \| \Pi_i A \Pi_0 B \Pi_j \| + \| \Pi_i A \Pi_1 B \Pi_j \| \\
&\leq \| \Pi_i A \Pi_0 \| \cdot \| \Pi_0 B \Pi_j \| + \| \Pi_i A \Pi_1 \| \cdot \| \Pi_1 B \Pi_j \| \\
&= \widetilde{A}[i,0] \cdot \widetilde{B}[0,j] + \widetilde{A}[i,1] \cdot \widetilde{B}[1,j] \\
&= (\widetilde{A} \cdot \widetilde{B})[i,j].
\end{aligned}$$

□

With that we are ready to prove Lemma 6.1.

6.3 Approximating the characteristic function

Proof of Lemma 6.1. As before P_b is the diagonal matrix with $P_b[v,v] = e^{\mathsf{val}(v) \cdot \frac{b}{\sqrt{t}} \cdot i}$. Let $\theta = \frac{b}{\sqrt{t}}$, $\theta \in [-\theta_0, \theta_0]$ for $\theta_0 = \frac{\pi}{3}$. Then,

$$\begin{aligned}
\rho(b) &= \mathbf{E}\, e^{ibY} = \mathbf{1}^t (P_b G)^{t-1} P_b \mathbf{1} \\
\beta(b) &= (\cos\theta)^t
\end{aligned}$$

By the above discussion about flow matrices we have:

$$\widetilde{GP_b} \leq_{ew} \widetilde{G} \cdot \widetilde{P_b} \leq_{ew} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \lambda \sin\theta & \lambda \end{pmatrix}.$$

Our main technical tool is:

Lemma 6.6. *(The characteristic distance lemma) If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{ew} 0$ and $d < a$ then*

$$|A^t[0,0] - (A[0,0])^t| \leq \left(a + \frac{bc}{a-d}\right)^{t-2} \frac{abc}{a-d}.$$

We prove the lemma in Section 6.4. With that:

$$
\begin{aligned}
|\rho(b)-\beta(b)| &= |\mathbf{1}^t(GP_b)^t\mathbf{1}-(\cos\theta)^t| \\
&= |\widetilde{(GP_b)}^t[0,0]-(\widetilde{GP_b}[0,0])^t| \\
&\le \left(\cos\theta+\frac{\lambda\sin^2\theta}{\cos\theta-\lambda}\right)^{t-2}\cdot\frac{\lambda\sin^2\theta\cos\theta}{\cos\theta-\lambda} \\
&\le (\cos\theta+4\lambda\sin^2\theta)^{t-2}\cdot 4\lambda\sin^2\theta\cos\theta \\
&\le 16\lambda\theta^2(\cos\theta+4\lambda\sin^2\theta)^t,
\end{aligned}
$$

where we have used $\cos\theta \ge \frac{1}{2}$, $\lambda \le \frac{1}{4}$ and $\sin\theta \le \theta$. Now,

$$
\begin{aligned}
\cos\theta+4\lambda\sin^2\theta &\le \cos\theta+\frac{\sin^2\theta}{4}=\sqrt{1-\sin^2\theta}+\frac{\sin^2\theta}{4} \\
&\le 1-\frac{\sin^2\theta}{2}+\frac{\sin^2\theta}{4}=1-\frac{\sin^2\theta}{4} \\
&\le 1-\frac{1}{4}\cdot(\frac{\theta}{2})^2=1-\frac{\theta^2}{16}
\end{aligned}
$$

where we have used $\lambda \le 1/16$ and $\sin x \ge x-\frac{x^3}{6} \ge \frac{x}{2}$ for $x \le 1$. Altogether,

$$
\begin{aligned}
|\rho(b)-\beta(b)| &\le 16\lambda\theta^2(\cos\theta+4\lambda\sin^2\theta)^t \\
&\le 16\lambda\theta^2\left(1-\frac{b^2}{16t}\right)^t \\
&\le \frac{16\lambda}{t}b^2e^{-\frac{b^2}{16}}.
\end{aligned}
$$

□

6.4 The characteristic distance lemma

Lemma 6.7. *If* $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \ge_{ew} 0$ *with* $a \ge 1$ *and* $d<1$ *then* $A^t[0,0] \le (a+\frac{bc}{1-d})^t$.

Proof. By induction on t. The cases $t=0,1$ are clear. Assume for $1,\ldots,t$ and let us prove for $t+1$.

$$
A^t[0,0] = A^{t-1}[0,0]A[0,0]+\sum_{j=0}^{t-2}A^j[0,0]\cdot A[0,1]\cdot A[1,0]\cdot A[1,1]^{t-j-2},
$$

where j goes over the last time the path was at 0 ($0 \le j \le t-2$). As $A[i,j] \ge 0$ and $a=A[0,0] \ge 1$ we see that $A^t[0,0] \ge A^{t-1}[0,0]$. Hence,

$$
\begin{aligned}
A^t[0,0] &\le A^{t-1}[0,0]a+\sum_{j=0}^{t-2}A^{t-1}[0,0]\cdot bc\cdot d^{t-j-2} \\
&\le A^{t-1}[0,0]\left(a+bc\sum_{j=0}^{\infty}d^j\right) \\
&\le A^{t-1}[0,0]\left(a+\frac{bc}{1-d}\right).
\end{aligned}
$$

$\square$

Lemma 6.8. *If* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{ew} 0$ *with* $a \geq 1$ *and* $d < 1$ *then*

$$\left|A^t[0,0] - (A[0,0])^t\right| \leq \left(a + \frac{bc}{1-d}\right)^t.$$

Proof. Notice that the only path from 0 to 0 that never leaves 0 gives the value a^t. Also, all paths give non-negative contribution. Thus, to bound $|A^t[0,0] - (A[0,0])^t|$ we need to sum over all paths that sometimes leave 0. Doing an analysis similar to Lemma 6.7 we get:

$$\begin{aligned} |A^t[0,0] - (A[0,0])^t| &\leq \sum_{j=0}^{t-2} A^{t-2}[0,0] \cdot bc \cdot d^{t-j-2} \\ &\leq A^{t-2}[0,0] bc \sum_{j=0}^{\infty} d^j \\ &\leq \left(a + \frac{bc}{1-d}\right)^{t-2} \cdot \frac{bc}{1-d}. \end{aligned}$$

$\square$

We can extend the lemma to the case where $A[0,0] > 0$ but not necessarily greater than 1, and $d < a$, as follows:

Lemma 6.9. *If* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{ew} 0$ *and* $d < a$ *then*

$$\left|A^t[0,0] - (A[0,0])^t\right| \leq \left(a + \frac{bc}{a-d}\right)^{t-2} \frac{abc}{a-d}.$$

Proof. we have $A = a \cdot \begin{pmatrix} 1 & \frac{b}{a} \\ \frac{c}{a} & \frac{d}{a} \end{pmatrix}$. Hence

$$\begin{aligned} |A^t[0,0] - (A[0,0])^t| &\leq a^t \left(1 + \frac{bc}{a^2\left(1 - \frac{d}{a}\right)}\right)^{t-2} \frac{bc}{a^2(1 - \frac{d}{a})} \\ &= a^t \left(1 + \frac{bc}{a(a-d)}\right)^{t-2} \frac{bc}{a(a-d)} \\ &= \left(a + \frac{bc}{a-d}\right)^{t-2} \frac{abc}{a-d}. \end{aligned}$$

$\square$

7 Acknowledgments

We thank Venkat Guruswami and Vinayak Kumar for discussing their results with us. We thank Ron Peled for discussions on the CLT in TVD, and for pointing us to results on local convergence for independent processes.

References

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.

[AEL95] Noga Alon, Jeff Edmonds, and Michael Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519. IEEE, 1995.

[AGM87] N. Alon, Z. Galil, and V. D. Milman. Better expanders and superconcentrators. *J. Algorithms*, 8(3):337–347, 1987.

[Ajt94] M. Ajtai. Recursive construction for 3-regular expanders. *Combinatorica*, 14(4):379–416, 1994.

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 132–140, 1987.

[Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[BATS11] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-Ramanujan graphs using the zig-zag product. *SIAM J. Comput.*, 40(2):267–290, 2011.

[BCG20] Mark Braverman, Gil Cohen, and Sumegha Garg. Pseudorandom pseudo-distributions with near-optimal error for read-once branching programs. *SIAM Journal on Computing*, 49(5):STOC18–242–STOC18–299, 2020.

[BDW02] Harry Buhrman and Ronald De Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[Bec15] Christopher Beck. Chernoff bounds for expander walks. https://www.ias.edu/video/csdm/2015/0310-ChristopherBeck, 2015. Video of a lecture at IAS.

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5):Art. 28, 10, 2010.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[CHRT18] Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 363–375. ACM, New York, 2018.

[Coh16] Michael B. Cohen. Ramanujan graphs in polynomial time. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*, pages 276–281. IEEE Computer Soc., Los Alamitos, CA, 2016.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf, 2019. Version 5 January 11, 2019.

[Fel71] William Feller. *An Introduction to Probability theory and its application Vol II*. John Wiley and Sons, 1971.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[GG81] Ofer Gabber and Zvi Galil. Explicit constructions of linear-sized superconcentrators. *J. Comput. System Sci.*, 22(3):407–420, 1981. Special issued dedicated to Michael Machtey.

[Gil98a] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[Gil98b] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GK20] Venkatesan Guruswami and Vinayak Kumar. Pseudobinomiality of the sticky random walk. Technical report, ECCC TR20-151, 2020.

[Hea08] Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17(1):3–37, 2008.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[Hua19] Hao Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Ann. of Math. (2)*, 190(3):949–955, 2019.

[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*, pages 356–364, 1994.

[Kah95] Nabil Kahale. Eigenvalues and expansion of regular graphs. *J. Assoc. Comput. Mach.*, 42(5):1091–1106, 1995.

[Klo17] Benoît Kloeckner. Effective limit theorems for Markov chains with a spectral gap. *arXiv preprint arXiv:1703.09623*, 2017.

[KMRZS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):Art. 11, 42, 2017.

[KV86] Claude Kipnis and SR Srinivasa Varadhan. Central limit theorem for additive functionals of reversible markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[Lez01] Pascal Lezaud. Chernoff and Berry-Esséen inequalities for markov processes. *ESAIM: Probability and Statistics*, 5:183–201, 2001.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. Assoc. Comput. Mach.*, 40(3):607–620, 1993.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Lub12] Alexander Lubotzky. Expander graphs in pure and applied mathematics. *Bull. Amer. Math. Soc. (N.S.)*, 49(1):113–162, 2012.

[Mar88] Grigorii Aleksandrovich Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy peredachi informatsii*, 24(1):51–60, 1988.

[MOP20] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-ramanujan graphs of every degree. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 510–523, 2020.

[MSS15] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *Ann. of Math. (2)*, 182(1):307–325, 2015.

[MSS18] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV: Bipartite Ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018.

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Rei05] Omer Reingold. Undirected ST-connectivity in log-space. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 376–385. ACM, New York, 2005.

[RRV99] Ran Raz, Omer Reingold, and Salil Vadhan. Error reduction for extractors. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 191–201. IEEE, 1999.

[RSV13] Omer Reingold, Thomas Steinke, and Salil Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *Approximation, randomization, and combinatorial optimization*, volume 8096 of *Lecture Notes in Comput. Sci.*, pages 655–670. Springer, Heidelberg, 2013.

[RV05] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 436–447. Springer, 2005.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13. IEEE, 2000.

[SS96] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of ac0. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[Tao15] Terence Tao. What's new: 275a, notes 5. variants of the central limit theorem. https://terrytao.wordpress.com/2015/11/19/275a-notes-5-variants-of-the-central-limit-theorem/#more-8566, 2015.

[Tre17] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley, https://people. eecs. berkeley. edu/˜ luca/books/expanders-2016. pdf*, 2017.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251. ACM, New York, 2017.

[Vad12] Salil P Vadhan. *Pseudorandomness*, volume 7. Now, 2012.

[Val76] Leslie G. Valiant. Graph-theoretic properties in computational complexity. *J. Comput. System Sci.*, 13(3):278–285, 1976.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il