Ultra-Sparse Expanders and the Free Method

Gil Cohen
Tel Aviv University, Israel

Gal Maor
Tel Aviv University, Israel

Abstract

In this paper we ask how much expansion one can retain with almost no edges beyond connectivity. Concretely, for graphs of average degree $2+\varepsilon$, what is the "Ramanujan bound" – how does spectral expansion scale with ε? We compare five ultra–sparse graph models – including the configuration model, subdivision of regular expanders, and the union of a cycle with a partial matching – and analyze each under the normalized or unnormalized notions of expansion. In the normalized setting, we prove bounds that are essentially optimal, determining the correct asymptotic dependence on ε. Our results extend to expansion in general irregular graphs.

For some models we prove rigorous bounds – primarily via finite free probability – while others remain beyond our current techniques. To bridge this gap, we introduce the *Free Method*, which produces quantitative predictions without proving existence – analogous to the probabilistic method, which certifies existence without providing an explicit construction. These predictions align with experiments. We expect the free method to be useful more broadly in graph-theoretic settings.

2012 ACM Subject Classification Mathematics of computing → Spectra of graphs; Theory of computation → Expander graphs and randomness extractors

Keywords and phrases Spectral expanders, free probability

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.49

Category RANDOM

Related Version *Full Version*: `https://eccc.weizmann.ac.il/report/2025/196/`

Funding Funded by ERC starting grant 949499 and by the Israel Science Foundation grant 2989/24.

1 Introduction

Expanders are sparse, highly connected graphs whose expansion allows them, up to some error, to behave like complete graphs at far lower cost. They have been studied for decades, with constructions ranging from combinatorial [31, 5, 4] and group-theoretic [20, 25] to analytic [22]. Their applications include coding theory [33, 2, 36], computational complexity [17, 13, 30], and more. For background, see [16], Chapter 4 of [37], or [35].

We study the optimal expansion when the edge budget is barely above connectivity. Specifically, for graphs of average degree $2+\varepsilon$, what is the "Ramanujan bound" – how does spectral expansion depend on ε? Such *ultra-sparse* graphs are necessarily irregular, as are many graphs arising in applications, including sparse codes [21, 32]. We therefore consider both normalized and unnormalized (adjacency-based) expansion.

Let G be an undirected graph on n vertices with adjacency matrix A and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We call λ_2 the *one-sided spectral expansion*, $\lambda_1 - \lambda_2$ the *one-sided spectral gap*, and $\lambda \stackrel{\triangle}{=} \max\{|\lambda_2|, |\lambda_n|\}$ the (two-sided) *spectral expansion*. In the normalized setting, let D be the diagonal degree matrix and $W \stackrel{\triangle}{=} AD^{-1}$ the random walk matrix, with eigenvalues $1 = \omega_1 \geq \omega_2 \geq \cdots \geq \omega_n$. Here ω_2 is the *one-sided normalized spectral expansion*, $1-\omega_2$ the corresponding gap, and $\omega \stackrel{\triangle}{=} \max\{|\omega_2|, |\omega_n|\}$ the (two-sided) *normalized spectral expansion*.

A connected 2-regular graph is a cycle, whose spectral gap vanishes with its size. By contrast, 3-regular graphs can attain normalized spectral expansion $\frac{\sqrt{8}}{3}$, bounded away from 1. More generally, the optimal asymptotic value for d-regular graphs is $\frac{2\sqrt{d-1}}{d}$ [28], attained by Ramanujan graphs. Thus, greater degree buys better expansion.

1.1 Ultra-sparse expanders

Since the main utility of expanders is normalized spectral expansion bounded away from 1, we ask how sparse a graph can be while retaining this property, even beyond the regular-graph barrier $\frac{\sqrt{8}}{3} \approx 0.94$. Our goal is to characterize the sparsity–expansion tradeoff in this regime.

Any connected n-vertex graph has average degree at least $2 - \frac{2}{n}$, so we focus on average degree $2+\varepsilon$, where $\varepsilon > 0$ controls the sparsity. We seek the corresponding Ramanujan bound as a function of ε.

To keep degree variance small, we mainly study graphs with degrees in $\{2, 3\}$. High-degree vertices may impose disproportionate load, so we exclude hub-and-spoke examples such as the wheel graph, even though they can be more amortization-efficient. Our methods extend to broader degree distributions. In the normalized setting we obtain essentially optimal bounds, while the unnormalized setting remains more open.

For irregular graphs, normalized and unnormalized expansion capture different phenomena. The normalized spectrum governs random walks, while the unnormalized spectrum is better suited to edge expansion and expander-mixing-type bounds, with different treatments of degree heterogeneity; see [1]. Although less studied for irregular graphs, the unnormalized spectrum remains informative when the degree variance is small, as in the $\{2, 3\}$ regime.

There are two main frameworks for constructing Ramanujan graphs. The first uses highly symmetric Cayley graphs of carefully chosen non-abelian groups, beginning with Lubotzky, Phillips, and Sarnak [20] and Margulis [25]. The second, initiated by Marcus, Spielman, and Srivastava, is analytic and draws on free probability [22, 23, 24]; Cohen later made this approach explicit [11].

The group-theoretic framework appears poorly suited to ultra-sparse graphs because Cayley graphs are regular. Irregular group-based constructions, for example via Schreier graphs, remain an interesting direction.

The MSS framework studies unions of d suitably uncorrelated perfect matchings and targets one-sided expansion. Although it does not directly match our setting, it is our main point of departure.

1.2 Five models of ultra-sparse graphs

The MSS approach studies a graph distribution and proves that some graph in its support is Ramanujan. For d-regular graphs, the distribution is the union of d perfect matchings. We ask which analogous models are natural in the ultra-sparse regime, and consider five.

Two-and-a-fraction matchings

Our first model is the union of two perfect matchings and an ε-matching, denoted $\mathcal{M}+\mathcal{M}+\varepsilon\mathcal{M}$ or $(2+\varepsilon)\mathcal{M}$. The partial matching covers an ε-fraction of the vertices. At $\varepsilon = 0$ and $\varepsilon = 1$, this recovers the MSS models for 2- and 3-regular graphs.

Three equally sized partial matchings

Alternatively, take three β-matchings with $\beta = \frac{2+\varepsilon}{3}$, denoted $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$. This model has a basic connectivity drawback: for small ε, the expected fraction of isolated vertices is $\left(\frac{1-\varepsilon}{3}\right)^3 \approx \frac{1}{27}$. We nevertheless study it to compare the models rather than select a single "best" one.

Cycle plus partial matching

To guarantee connectivity, we also study the union of a spanning cycle and an ε-partial matching, denoted $\mathcal{C} + \varepsilon\mathcal{M}$. Unlike $(2+\varepsilon)\mathcal{M}$, this model is connected before the partial matching is added.

A related explicit 3-regular construction [19] joins each field element x to $x+1$, $x-1$, and x^{-1}, forming a cycle plus a matching. A similar model was recently studied from a property-testing perspective [14].

The configuration model

Here each vertex v contributes d_v half-edges, and a uniformly chosen perfect matching on all half-edges determines the graph. In our ultra-sparse setting, $d_v \in \{2, 3\}$.

The configuration model is standard in spectral random graph theory [6] and in the analysis of graph operators such as the zig-zag product [31, 9]. Unlike a union of perfect matchings, it naturally supports varying degrees. Our analysis also gives another proof of the MSS result in the regular case.

Subdividing 3-regular graphs

Finally, start from a 3-regular Ramanujan graph, or more generally a d-regular one, and replace every edge by a long path. This subdivision operation, classical since Kuratowski [18], inserts degree-2 vertices and drives the average degree toward 2. Choosing the path length appropriately yields average degree $2+\varepsilon$ while preserving the base graph's global structure.

This model is explicit whenever the base graph is explicit, and strongly explicit whenever the base graph is strongly explicit.

2 Our Results

In this section, we present our results on the five ultra-sparse graph models introduced above. In the normalized setting, we essentially settle the question of the optimal ultra-sparse expanders. In the unnormalized setting, we introduce a new framework for analyzing the problem, while many basic questions remain open.

As an additional application, the full version proves that suitable unions of Ramanujan graphs remain nearly Ramanujan, and that suitable unions of cycles attain the Ramanujan bound.

2.1 Ultra-sparse expanders in the normalized setting

We begin by presenting our results in the normalized setting. Our first result uses the configuration model (the fourth model) and applies to general irregular graphs – indeed, to any prescribed degree sequence.

▶ **Theorem 1** (Irregular graphs via the configuration model). *Let $(d_1, \ldots, d_n)$ be a sequence of degrees such that $\sum_{i=1}^{n} d_i$ is even. Then there exists a graph with this degree sequence whose one-sided normalized spectral expansion is bounded above by $\frac{2\sqrt{\bar{d}-1}}{\bar{d}}$, where $\bar{d}$ is the average degree.*

Our proof is based on the configuration model (rather than the perfect matchings model used in [23]) and is inspired by the notion of *free projections* from classical free probability theory, which we discuss in Section 5. As a direct corollary we obtain the following result for ultra-sparse graphs.

▶ **Corollary 2** (Ultra-sparse expanders via the configuration model). *For every $\varepsilon > 0$ and for infinitely many values of n, there exists a graph on n vertices with degrees 2 or 3, with average degree $2 + \varepsilon$, whose one-sided normalized spectral expansion is bounded by $\frac{2\sqrt{1+\varepsilon}}{2+\varepsilon} = 1 - \frac{\varepsilon^2}{8} + O(\varepsilon^3)$.*

At first glance, the bound in Theorem 1 resembles the Ramanujan bound for d-regular graphs. Moreover, empirical evidence suggests that the typical behavior of random configuration graphs is indeed captured by Theorem 1. However, optimality in the irregular setting is more delicate, and the bound is not generally tight. For example, consider the wheel graph, formed from a cycle plus one additional vertex connected to all cycle vertices (we elaborate on this example in the full version of this paper). The normalized spectral expansion of the wheel approaches $\frac{2}{3}$, since within two steps one reaches a uniformly random vertex with probability $\frac{1}{3}$. Yet its average degree approaches 4 as the number of vertices grows, so Theorem 1 guarantees only a graph G with $\omega(G) \leq \frac{\sqrt{3}}{2} \approx 0.866$, which is significantly worse.

This gap suggests that even for graphs with degrees 2 and 3 – where degree variance is kept to a minimum – Corollary 2 may not be tight. Indeed, our second result shows that the subdivision model (our fifth model) does yield better expansion. For a d-regular graph G and parameter $\varepsilon > 0$, let G_ε denote the graph obtained by subdividing each edge of G into paths of equal length, so that the resulting graph has average degree $2 + \varepsilon$.

▶ **Theorem 3** (Ultra-sparse expanders via subdivision). *Let G be a d-regular graph with normalized spectral expansion ω, and let $\varepsilon > 0$. Then the normalized spectral expansion of G_ε is $\cos\left(\frac{d}{2(d-2)} \arccos(\omega)\varepsilon\right)$.*

Specializing to $d = 3$ and taking G be a 3-regular Ramanujan graph, the guaranteed spectral expansion for the resulting ultra-sparse graph is the following.

▶ **Corollary 4.** *Let G to be a 3-regular Ramanujan graph. The normalized spectral expansion of G_ε is $\cos\left(\frac{3}{2}\arccos\left(\frac{\sqrt{8}}{3}\right)\varepsilon\right) \approx 1 - 0.1299\varepsilon^2$.*

This improves upon the bound obtained via the configuration model in Corollary 2. In particular, the syntactically Ramanujan-like bound in Theorem 1 is not optimal.

It is, perhaps, surprising that the configuration model yields weaker expanders. In the configuration model, the degree-2 vertices are incorporated *while* the expander is being constructed, whereas in the subdivision approach they are added only after the 3-regular Ramanujan base graph has been fixed – and in an indifferent, uniform manner: every edge is replaced by a path of the same length. However, the Ramanujan property only guarantees a large spectral gap; it conveys little about local edge densities, so some regions can still be denser than others. Thus, it can be advantageous to distribute the degree-2 vertices nonuniformly, taking into account these density variances.

On the other hand, the expected time spent "stuck" on a path – which weakens the spectral expansion – is quadratic in its length. Since the total length summed over all paths is fixed as a function of ε, it is therefore beneficial to make all paths the same length. Still, one might hope to exploit a 3-regular Ramanujan base graph with substantial variation in local edge density by distributing the subdivision lengths nonuniformly, potentially yielding an even larger spectral gap.

The gap is, however, bounded, in a concrete way that shows both Corollaries 2 and 4 are asymptotically optimal in the dependence on ε. The aforementioned quadratic time to escape a path is levaraged to obtain the following lower bound (proved in the full version), serving as an asymptotic "Alon-Boppana" bound for ultra-sparse graphs.

▶ **Theorem 5.** *Let G be a graph with all vertices of degrees 2 or 3 and average degree $2+\varepsilon$. Then the normalized spectral expansion of G is at least $1-O(\varepsilon^2)$.*

This leads to the following open problem.

▶ **Open Problem 6.** *What is the optimal normalized spectral gap for graphs with degrees 2 and 3 and average degree $2+\varepsilon$?*

Corollary 4, together with Theorem 5, prove that the true answer to Open Problem 6 is $C\varepsilon^2 + o\left(\varepsilon^2\right)$ for some constant $C \geq 0.1299$.

2.2 Ultra-sparse expanders in the unnormalized setting

In this section we present our results in the unnormalized setting. We begin with the subdivision model, presenting our result for ultra-sparse graphs in the unnormalized setting, and then turn to the other four models in the same setting. Although we are not yet able to prove the existence of ultra-sparse graphs for these other models in the unnormalized setting, we introduce the *Free Method*, which yields quantitative *predictions* for the attainable spectral expansion as a function of ε in each case. While nonrigorous, these predictions align with empirical observations. We view the Free Method, discussed in Section 6, as a main contribution alongside our formal results: it lets us "peek" at the expected behavior before an existence proof is in hand.

▶ **Theorem 7** (Ultra-sparse expanders via subdivision)**.** *Let G be a d-regular graph with spectral expansion λ, and let $\varepsilon > 0$. Then the spectral expansion of G_ε is obtained as follows: find the unique positive ξ solving*

$$\lambda = \frac{\sinh((k+1)\xi) - (d-1)\sinh((k-1)\xi)}{\sinh(\xi)}. \tag{1}$$

where $k = \left\lceil \frac{(d-2)(2+\varepsilon)}{d\varepsilon} \right\rceil$. Then the spectral expansion μ is given by $2\cosh\xi$.

Although Theorem 7 gives the exact one-sided spectral expansion, it is not straightforward to read off its dependence on ε (a numerical derivation for μ for the case where G is a 3-regular Ramanujan graph is shown in Figure 1). However, by expanding about $\varepsilon = 0^+$, we show that $\mu = \sqrt{d-1} + \frac{1}{\sqrt{d-1}} + (1+o_k(1))\frac{\lambda}{2}\cdot\frac{(d-2)^2}{\left(\sqrt{d-1}\right)^{k+3}}$. Observe the *exponential* decay of the error in $k \approx 1/\varepsilon$, in contrast to the normalized setting, where the decay in ε is only quadratic – even for the same model.

We now switch gears to the other four models and the free method we use to analyze them.

2.3 The free method

We view the free method as an analogue of the probabilistic method, and so we begin by discussing the latter. The probabilistic method is a powerful tool: it proves the *existence* of an object without revealing what it looks like. The probabilistic method offers no insight into the structure of the object or how to find it explicitly. The value associated with the guaranteed object often serves as a benchmark (a quantitative prediction) for researchers aiming to find explicit constructions – e.g., the dependence of k on n for Ramsey graphs, or the distance–rate tradeoff in coding theory.

However, in many graph-theoretic settings even proving existence is technically formidable. Consider, for example, the value $2\sqrt{d-1}$. Proving the existence of graphs that attain this spectral expansion is notoriously difficult. Nonetheless, the value itself – serving as a benchmark – is both informative and insightful.

The free method, in a sense, takes things one step further. It is designed for situations where even the *mere existence* of an object with the desired properties is too difficult to prove, and thus the optimal value cannot be established via the same route as in the probabilistic method (namely, by proving existence). We reiterate that the main takeaway from the probabilistic method is often the *value* that serves as a benchmark for explicit constructions. When it comes to devising such constructions, the existence proof is typically unused, as it generally ignores precisely the structure required for explicit constructions. The free method provides *only* this key information.

More concretely, the free method targets graph-theoretic problems by abstracting away everything except the spectrum. Given a problem on finite graphs, we first *model* it by an infinite-dimensional object that captures the essential spectral features of the original problem. We then apply tools from free probability (powerful yet simple to use) to solve the infinite-dimensional problem, and the resulting quantity serves as a prediction for the finite case. We denote such predicted values by $\mathsf{pred}(\cdot)$ throughout the paper.

This perspective generalizes the "universal cover" viewpoint, but without requiring identification of a specific cover or combinatorial structure – only the spectral data.

To summarize, the philosophy of the probabilistic method is to prove existence without providing an explicit construction, while also yielding a quantitative benchmark – often the most valuable piece of information when constructing explicit objects. In contrast, the free method aims to derive this benchmark while entirely sidestepping a proof of existence, let alone an explicit construction. Because it makes weaker guarantees, the free method applies in much more challenging settings – ubiquitous in spectral graph theory – where the probabilistic method often falls short due to intricate eigenvalue behavior and the problem's inherent noncommutativity.

To set the stage for a more formal discussion, we fix notation and recall the *Cauchy transform* of an undirected graph G on n vertices: $\mathcal{G}_G(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x-\lambda_i}$.

Consider the $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ model, the union of three ρ-partial matchings with $\rho = \frac{2+\varepsilon}{3}$. In settings where one unions t independent samples from the same distribution (here $t = 3$), the first procedure we present in the free method prescribes:

1. Find the unique solution $x_0 > \lambda_1$ to

$$(t-1)\,\mathcal{G}'_G(x) + t\,\mathcal{G}_G(x)^2 \;=\; 0. \tag{2}$$

2. The prediction is given by

$$t \cdot x_0 \;-\; \frac{t-1}{\mathcal{G}_G(x_0)}. \tag{3}$$

At this point we have not justified this procedure (let alone the existence and uniqueness of $x_0 > \lambda_1$). The point is that the resulting procedure is quite general and easy to apply. In Section 6 we explain the rationale and derivation; here we focus solely on applying the method.

Before applying the free method to ultra-sparse graphs, we warm up by rederiving the Ramanujan bound $2\sqrt{d-1}$ using the recipe above.

2.3.1 Predictions for the three partial matchings model via the free method

We now apply the free method to ultra-sparse expanders, in particular to the $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ model. Concretely, we take $t = 3$ independent copies of a ρ-partial matching with $\rho = \frac{2+\varepsilon}{3}$. Such a matching has a $(1-\rho)$-fraction of its eigenvalues equal to 0, with the remaining eigenvalues split evenly between $+1$ and -1. The corresponding Cauchy transform, denoted $\mathcal{G}_\rho$, is therefore $\mathcal{G}_\rho(x) = \frac{x^2-1+\rho}{x(x^2-1)}$. Equation (2) now becomes $2\,\mathcal{G}'_\rho(x) + 3\,\mathcal{G}_\rho(x)^2 = 0$, whose unique solution for $x > 1$ is

$$x_0(\rho) = \sqrt{1 + \sqrt{\rho(4-3\rho)}}\,.$$

Plugging this to Equation (3) and expressing the result in terms of ε yields the prediction

$$\begin{aligned}\mathsf{pred}_{3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)}(\varepsilon) &= \sqrt{1 + \sqrt{\frac{4-\varepsilon^2}{3}} \cdot \frac{\sqrt{3(4-\varepsilon^2)} - 3\varepsilon}{2(1-\varepsilon)}} \\ &= \sqrt{3+\sqrt{12}} + \frac{\sqrt{3+\sqrt{12}}}{4+\sqrt{12}}\,\varepsilon + O(\varepsilon^2).\end{aligned} \tag{4}$$

On the other extreme, when $\varepsilon = 1$, we recover the $2\sqrt{2}$ bound for 3-regular Ramanujan graphs. In fact, the Taylor expansion to the left of $\varepsilon = 1$ is $2\sqrt{2} - \frac{1-\varepsilon}{3\sqrt{2}} + O\left((1-\varepsilon)^2\right)$. This ε–dependence is hard to infer purely from experiments or by combinatorial means alone, and it matches experimental results.

2.3.2 Predictions for the remaining models

In Section 6–where we also explain the rationale behind the method–we apply the free method to the remaining models using techniques analogous to the procedure introduced above. For the configuration model, we apply a procedure to bound the spectrum of the product of two matrices, one of which is PSD, which has common ground with the proof of Theorem 1. Its application gives

$$\mathsf{pred}_{\mathsf{conf}}(\varepsilon) = \frac{3}{\sqrt{2}} + \frac{3}{4}\,\varepsilon^{1/2} + \frac{9\sqrt{2}}{32}\,\varepsilon + O(\varepsilon^{3/2}). \tag{5}$$

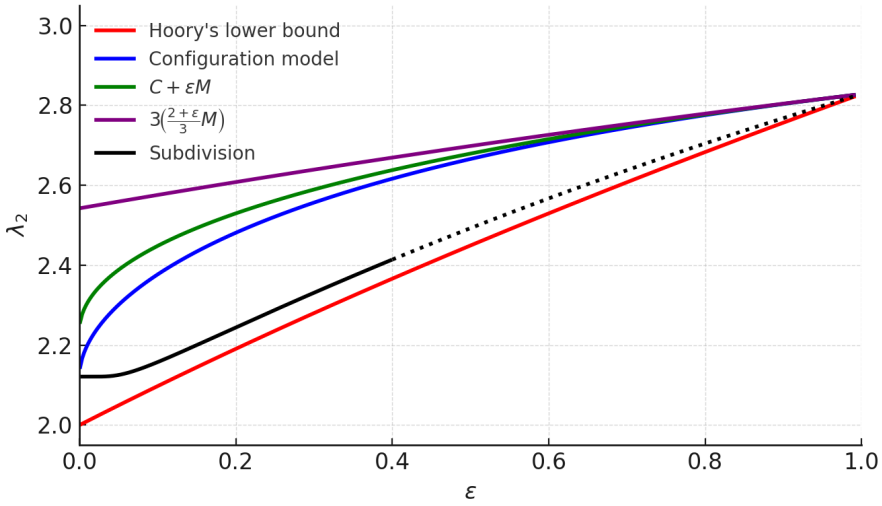

Figure 1 Comparison of the predictions for the bound on λ_2 for the conf, $\mathcal{C} + \varepsilon\mathcal{M}$ and $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ models. Hoory's lower bound of $2\sqrt{\bar{d}-1}$ is displayed in red, and the subdivision model of a 3-regular Ramanujan graph ($\lambda = 2\sqrt{2}$) appears in black. Note that, as the number of subdivisions k is an integer, the largest possible value for ε in this model is 0.4. For clarity, however, we also plot the solution to Equation (1) for real $k < 2$ as a dotted line.

Interestingly, although $(2+\varepsilon)\mathcal{M}$ and $\mathcal{C} + \varepsilon\mathcal{M}$ are quite different as finite-graph models, their behavior coincides under the free-method analysis (based on classical free probability). Using an additive analog of the multiplicative procedure used for the configuration model yields the prediction

$$\mathsf{pred}_{\mathcal{C}+\varepsilon\mathcal{M}}(\varepsilon) = \mathsf{pred}_{(2+\varepsilon)\mathcal{M}}(\varepsilon) = \sqrt{5} + \sqrt{\frac{4}{5\phi}}\,\sqrt{\varepsilon}, \tag{6}$$

where $\phi = (1+\sqrt{5})/2$ is the golden ratio. The analysis uses a variant of the procedure used for $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$, applied to two distinct distributions.

For comparison, Hoory's lower bound [15] states that if a graph G has average degree at least $\bar{d}$ even after deleting any ball of radius r, then

$$\lambda_2(G) \;\geq\; 2\sqrt{\bar{d}-1}\left(1 - c\,\frac{\log r}{r}\right), \tag{7}$$

for a universal constant c. In our unbounded-size models the r-dependent error should be negligible, so we compare to $2\sqrt{\bar{d}-1} = 2\sqrt{1+\varepsilon}$. Together with our predictions, this yields the comparison shown in Figure 1. As the figure suggests, there is an ordering of the models, and the curves coincide only at $\varepsilon = 1$.

3 Ultra-Sparse Expanders via Subdivision

Let G be a d-regular graph and denote its adjacency matrix by $\mathbf{A}$ and random walk matrix by $\mathbf{W}$. Let G_k be the k-subdivision of G, which is defined by replacing every edge of G with a path of length k, effectively adding $k-1$ new vertices of degree 2. While we will focus on analyzing the spectrum of the corresponding matrices $\mathbf{A}_k$ and $\mathbf{W}_k$ of the resulting graph, we clarify already at this point that this process yields a graph of average degree $2+\varepsilon$ by

choosing $k \approx \frac{2(d-2)}{d\varepsilon}$, tying it to our motivation. Nevertheless, the result is more general. The analysis provided here is elementary (i.e., it does not rely on finite free probability), and is inspired by the proofs in [38].

▶ **Lemma 8** (Subdivision and the random walk spectrum). *Let $k \geq 2$, and let G_k be as defined above. Then for every eigenvalue ω of $\mathbf{W}$ and every $t = 0, 1, \ldots, k-1$, $r_t = \cos\left(\frac{\arccos(\omega)+2\pi t}{k}\right)$ is an eigenvalue of $\mathbf{W}_k$. The rest of the eigenvalues are of the form $r_j = \cos\left(\frac{\pi j}{k}\right)$ for $j = 1, \ldots, k-1$.*

The proof uses elementary means, and appears in the full version of this paper.

▶ **Corollary 9.** *Let G be a d-regular graph with normalized one-sided spectral expansion $\omega(G)$, and let S be its subdivision with average degree $2+\varepsilon$. Then*

$$\omega_d(\varepsilon) \;=\; \cos\left(\frac{d\,\varepsilon}{2(d-2)}\,\arccos\left(\omega(G)\right)\right). \tag{8}$$

In particular:

1. *If $d = 3$ and G is Ramanujan, then $\omega_3(\varepsilon) \;=\; \cos\left(\frac{3}{2}\arccos\left(\frac{\sqrt{8}}{3}\right)\varepsilon\right) \;\approx\; 1 - 0.12993\,\varepsilon^2$.*
2. *As $d \to \infty$ (with G Ramanujan), since $\arccos\left(\frac{2\sqrt{d-1}}{d}\right) \to \frac{\pi}{2}$, we obtain $\omega_d(\varepsilon) \longrightarrow \cos\left(\frac{\pi}{4}\,\varepsilon\right) \;=\; 1 - \frac{\pi^2}{32}\,\varepsilon^2 + O(\varepsilon^4)$.*

Proof. Let $k = \left\lceil \frac{(d-2)(2+\varepsilon)}{d\,\varepsilon} \right\rceil$ be the subdivision parameter so that the resulting graph has average degree $2+\varepsilon$. By Lemma 8, the normalized expansion after k-subdivision is $\cos\left(\frac{\arccos(\omega(G))}{k}\right)$. Using the relation between k and ε established in Theorem 3 yields Equation (8).

For (1), put $\omega(G) = \frac{2\sqrt{2}}{3}$ (Ramanujan) and set $\phi \triangleq \arccos(\omega(G)) \approx 0.33984$, giving $\omega_3(\varepsilon) \;=\; \cos\left(\frac{3}{2}\phi\,\varepsilon\right) \;=\; 1 - \frac{(3\phi)^2}{8}\,\varepsilon^2 + O(\varepsilon^3) \;\approx\; 1 - 0.12993\,\varepsilon^2$. For (2), observe that for Ramanujan d-regular graphs, $\omega(G) = \frac{2\sqrt{d-1}}{d}$ and hence $\arccos(\omega(G)) \to \frac{\pi}{2}$ as $d \to \infty$. Plugging into Equation (8) gives the stated limit and its Taylor expansion. ◀

For the unnormalized case, we have the following:

▶ **Lemma 10** (Subdivision and the adjacency matrix spectrum). *Let $k \geq 2$, and let G_k be as defined above. Then every eigenvalue μ of $\mathbf{A}_k$ satisfies one of the conditions below.*

1. *$\mu = 2\cos(\pi j/k)$ for $j = 1, \ldots, k-1$.*
2. *$\mu = 2\cos\theta$ for real θ, satisfying*

$$\lambda = \frac{\sin((k+1)\theta) - (d-1)\,\sin((k-1)\theta)}{\sin\theta}, \tag{9}$$

   where λ is an eigenvalue of $\mathbf{A}$.
3. *$\mu = \pm 2\cosh\xi$ for real $\xi > 0$, satisfying*

$$\lambda = \frac{\sinh((k+1)\xi) - (d-1)\sinh((k-1)\xi)}{\sinh(\xi)}, \tag{10}$$

   where λ is an eigenvalue of $\mathbf{A}$.

Explicit solutions are feasible for $k = 2, 3$, but this is not the case for general k. Nonetheless, We give a sharp estimate of the result for large k, which is the regime of interest in our paper (as $k \approx 1/\varepsilon$ in our desired model).

▶ **Corollary 11.** *All eigenvalues μ of $\mathbf{A}_k$ with $\mu > 2$ satisfy*

$$\mu \;=\; \sqrt{d-1} + \frac{1}{\sqrt{d-1}} \;+\; (1+o_k(1))\,\frac{\lambda}{2}\cdot\frac{(d-2)^2}{\left(\sqrt{d-1}\right)^{k+3}},$$

where λ is an eigenvalue of $\mathbf{A}$.

4 Interlacing Families and Free Convolutions

In this section we review definitions and known results in free probability and interlacing families of polynomials. We use the following standard notation: for a symmetric matrix $\mathbf{A}$, $\chi_x(\mathbf{A})$ is the characteristic polynomial of $\mathbf{A}$ with variable x. We denote by $\lambda_k(\mathbf{A})$ the k-th largest eigenvalue of $\mathbf{A}$ (as $\mathbf{A}$ is symmetric, all eigenvalues are real and hence can be ordered as above). For every symmetric matrix $\mathbf{B}$ for which the all-ones vector $\mathbf{1}$ is an eigenvector and $\mathbf{B1} = b\mathbf{1}$, we denote by $p_{\mathbf{B}}(x)$ the polynomial satisfying $(x-b)p_{\mathbf{B}}(x) = \chi_x(\mathbf{B})$.

For a real-rooted polynomial $p(x)$, we denote by $\alpha_k(p)$ the k-th largest root of $p(x)$. Given a distribution P over polynomials, we denote by $\mathbf{E}_{p\sim P}[p(x)]$ the *expected* polynomial over this distribution, where the expectation is taken in coefficient space, namely, for every k, the coefficient of x^k in $\mathbf{E}_{p\sim P}[p(x)]$ is the expectation over coefficients corresponding to x^k in $p(x)$ drawn according to P.

The following lemma is a simplified version of a more general statement from [23], where it appeared as the main tool enabling us to analyze the expected characteristic polynomial over a distribution, and deduce an existence result.

▶ **Lemma 12.** *Let $\mathbf{A}_1,\ldots,\mathbf{A}_t$ be symmetric $m\times m$ matrices and $(\mathbf{P}_i)_{i\in[t]}$ be uniformly random $m\times m$ permutation matrices. Let $\mathbf{A}_{\mathbf{P}} = \sum_{i=1}^t \mathbf{P}_i\mathbf{A}_i\mathbf{P}_i^{\mathsf{T}}$. Then, for every $k\le m$ there exist permutation matrices $(\mathbf{R}_i)_{i\in[t]}$ such that*

$$\lambda_k(\mathbf{A}_{\mathbf{R}}) \le \alpha_k\left(\mathop{\mathbf{E}}_{\mathbf{P}}\chi_x(\mathbf{A}_{\mathbf{P}})\right). \tag{11}$$

The multiplicative analog of Lemma 12 was proved in [10]. A simplified version, suitable for our needs, is stated below [1].

▶ **Lemma 13.** *Let $\mathbf{M},\mathbf{N}$ be symmetric $m\times m$ matrices such that $\mathbf{M}$ is positive semidefinite, and let $\mathbf{P}$ be a uniformly random $m\times m$ permutation matrix. Then, for every $k\le m$ there exists a permutation matrix $\mathbf{S}$ such that*

$$\lambda_k\left(\mathbf{MSNS}^{\mathsf{T}}\right) \le \alpha_k\left(\mathop{\mathbf{E}}_{\mathbf{P}}\chi_x\left(\mathbf{MPNP}^{\mathsf{T}}\right)\right). \tag{12}$$

▶ **Definition 14** (Additive and multiplicative convolutions)**.** *Let $\mathbf{A},\mathbf{B}$ be real symmetric matrices of dimension m, with characteristic polynomials $a(x) = \chi_x(\mathbf{A})$ and $b(x) = \chi_x(\mathbf{B})$. The* additive convolution *$a\boxplus_m b$ and the* multiplicative convolution *$a\boxtimes_m b$ are the polynomials defined by*

$$(a\boxplus_m b)(x) = \mathop{\mathbf{E}}_{\mathbf{Q}}\chi_x\left(\mathbf{A}+\mathbf{QBQ}^{\mathsf{T}}\right),$$

[1] Lemma 13 was proved in [10, Theorem 6.5 and Lemma 6.3] in more generality (for a sequence of permutation matrices, as in Lemma 12). It is phrased therein, however, only for the case $\mathbf{M} = \mathbf{A}^2$ and $\mathbf{N} = \mathbf{A}$ for a symmetric matrix $\mathbf{A}$.

and

$$(a \boxtimes_m b)(x) = \mathop{\mathbf{E}}_{\mathbf{Q}} \chi_x \left(\mathbf{A}\mathbf{Q}\mathbf{B}\mathbf{Q}^{\mathsf{T}}\right),$$

where $\mathbf{Q}$ *is a Haar random orthogonal matrix. We will use* $\boxplus$ *and* $\boxtimes$ *instead of* $\boxplus_m$ *and* $\boxtimes_m$ *for simplifying notation when the dimension is clear from the context.*

Working with graph matrices, an issue with the above definition is that the Haar measure does not have a meaningful combinatorial interpretation. Therefore, we need a way to relate *permutation* matrices - which do have such interpretation - to Haar random matrices. To this end we state the following lemma (which appeared first in [23] for the additive case, and later on in [10] for the multiplicative case, whose proof uses similar ideas) which is described as a *Quadrature* result, translating an infinite (continuous) measure space to a finite one.

▶ **Lemma 15** (*Quadrature*; Corollary 4.9 from [23] and Lemma 2.3 from [10])**.** *Let* $\mathbf{A}, \mathbf{B}$ *be real* $m \times m$ *symmetric matrices such that* $\mathbf{A}\mathbf{1} = a\mathbf{1}$ *and* $\mathbf{B}\mathbf{1} = b\mathbf{1}$*. Denote by* $p_{\mathbf{A}}, p_{\mathbf{B}}$ *the polynomials satisfying* $\chi_x(\mathbf{A}) = (x-a)p_{\mathbf{A}}(x)$*,* $\chi_x(\mathbf{B}) = (x-b)p_{\mathbf{B}}(x)$*. Let* $\mathbf{P}$ *be a uniformly random* $m \times m$ *permutation matrix. Then,*

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x \left(\mathbf{A} + \mathbf{P}\mathbf{B}\mathbf{P}^{\mathsf{T}}\right) = (x - (a+b)) \left(p_{\mathbf{A}} \boxplus p_{\mathbf{B}}\right)(x), \tag{13}$$

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x \left(\mathbf{A}\mathbf{P}\mathbf{B}\mathbf{P}^{\mathsf{T}}\right) = (x - ab) \left(p_{\mathbf{A}} \boxtimes p_{\mathbf{B}}\right)(x). \tag{14}$$

4.1 Transforms

Let μ be a compactly supported probability distribution over $\mathbb{R}$, and let $a = \sup(\mu)$. The *Cauchy transform* of μ is defined on the domain (a, ∞) as the function

$$\mathcal{G}_\mu(x) = \int_{\mathbb{R}} \frac{1}{x-t} \mu(t) dt.$$

We remark that in many settings it is instructive to study the Cauchy transform as a function whose domain is $\mathbb{C}^+$. However, we will consider the Cauchy transform as a function on $\mathbb{R}$. The Cauchy transform is also related to the moments of a distribution. If $m_r(\mu)$ is the r-th moment of μ, then for every $x > a$ we have (see [27, Remark 2.19]):

$$\mathcal{G}_\mu(x) = \sum_{r=0}^{\infty} \frac{m_r(\mu)}{x^{r+1}}. \tag{15}$$

Let $p(x)$ be a degree m real-rooted polynomial with roots $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m$. To $p(x)$ we associate the uniform distribution over its roots, and therefore its Cauchy transform (defined for $x > \alpha_1$) is

$$\mathcal{G}_p(x) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{x - \alpha_i} = \frac{1}{m} \cdot \frac{p'(x)}{p(x)}.$$

Accompanying the Cauchy transform of μ is the *$\mathcal{M}$-transform* (or moment transform), which is defined by $\mathcal{M}_\mu(x) = x\mathcal{G}_\mu(x) - 1$. It is easy to see that $\mathcal{G}_\mu(x)$ is monotonically decreasing within its domain, and thus invertible. Denote the range of $\mathcal{G}_\mu(x)$ by $(0, L^+)$, that is, $L^+ = \lim_{x \to +a} \mathcal{G}_\mu(x)$ (for a polynomial $p(x)$, $L^+ = \infty$). With this in mind, we

define $\mathcal{K}_\mu : (0, L^+) \to (a, \infty)$ as the inverse of $\mathcal{G}_\mu$. In cases where $\mathcal{M}_\mu(x)$ is monotonically decreasing in the same domain, we can define $\mathcal{N}_\mu(y)$ as the inverse of $\mathcal{M}_\mu$ in a similar manner. By definition, for every $y \in (0, L^+)$, both $\mathcal{K}_\mu(y)$ and $\mathcal{N}_\mu(y)$ provide an upper bound on a, the supremum of the support of μ. In case of a polynomial $p(x)$, this bounds the largest root of $p(x)$, which we state for later reference in the following claim:

▷ Claim 16. Let $p(x)$ be a degree m real-rooted polynomial with roots $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m$. Then for every $y \in (0, \infty)$, both $\alpha_1 \leq \mathcal{K}_p(y)$ and $\alpha_1 \leq \mathcal{N}_p(y)$.

For a symmetric matrix $\mathbf{A}$, we use the notation $\mathcal{G}_\mathbf{A}$ for $\mathcal{G}_{\chi_x(\mathbf{A})}$, and $\mathcal{G}_{p_\mathbf{A}}$ for the Cauchy transform of $p_\mathbf{A}$ as used in Lemma 15 if applicable, and similar notations for the other transforms.

The key feature of the $\mathcal{K}$ and the $\mathcal{N}$ transforms is that they behave very well under additive and multiplicative convolutions, respectively, as was shown in [24, Theorem 1.11] and [9, Claim 4.12], and stated below.

▶ **Lemma 17** (Convolution bounds). *For real-rooted polynomials $p(x)$ and $q(x)$ of degree m, and for any $y > 0$,*

$$\mathcal{K}_{p \boxplus_m q}(y) \leq \mathcal{K}_p(y) + \mathcal{K}_q(y) - \frac{1}{y}.$$

When one of the polynomials has only non-negative roots, then for every $y > 0$,

$$\mathcal{N}_{p \boxtimes_m q}(y) \leq \frac{y}{y+1} \cdot \mathcal{N}_p(y) \cdot \mathcal{N}_q(y).$$

Both inequalities are strict in case both polynomials have at least two distinct roots.

An immediate corollary of Lemma 17 and Claim 16 is the following.

▶ **Corollary 18.** *For every real-rooted polynomial $p(x)$, $t \in \mathbb{N}$ and $y > 0$,* $\mathsf{maxroot}\left(p^{\boxplus t}(x)\right) \leq t \cdot \mathcal{K}_p(y) - \frac{t-1}{y}$.

▷ Claim 19. [9, Claim 4.12] Let $\mathbf{B}$ be a symmetric matrix such that $\mathbf{B}\mathbf{1} = b\mathbf{1}$ and b is its largest eigenvalue. Then, for every $x > b$, $\mathcal{G}_{p_\mathbf{B}}(x) \leq \mathcal{G}_\mathbf{B}(x)$ and $\mathcal{M}_{p_\mathbf{B}}(x) \leq \mathcal{M}_\mathbf{B}(x)$, and for every $y > 0$, $\mathcal{K}_{p_\mathbf{B}}(y) \leq \mathcal{K}_\mathbf{B}(y)$ and $\mathcal{N}_{p_\mathbf{B}}(y) \leq \mathcal{N}_\mathbf{B}(y)$.
When $\mathbf{B} \neq b\mathbf{I}$, all inequalities are sharp.

5 Ultra-Sparse Expanders in the Configuration Model

In this section we prove Theorem 1, which immediately implies Corollary 2.

▶ **Definition 20** (Configuration model for regular graphs). *A random undirected d-regular graph on n vertices is said to be sampled by the* configuration model *if it is generated by choosing a uniformly random perfect matching on the nd* half-edges.

5.1 Alternate MSS proof via finite free projection

In this section, we present our alternative proof of the MSS result using the configuration model.

▶ **Theorem 21.** *For every $d \geq 2, n \in \mathbb{N}$ such that $N = nd$ is even, there exists an $N \times N$ permutation matrix $\mathbf{P}$ such that the d-regular n-vertex graph $G_\mathbf{P}$ satisfies*

$$\lambda_2(G_\mathbf{P}) < 2\sqrt{d-1}.$$

Proof. We begin with the useful observation that

$$\mathbf{V} \triangleq \mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{J}_d \otimes \mathbf{I}_n,$$

where $\mathbf{J}_d$ denotes the $d \times d$ matrix with all entries equal to 1. Note that $\mathbf{J}_d$ is the random walk matrix (i.e., the normalized adjacency matrix) of the complete graph on d vertices with self-loops. Observe that

$$\chi_x(\mathbf{M}) = (x-1)^{\frac{N}{2}}(x+1)^{\frac{N}{2}},$$
$$\chi_x(\mathbf{V}) = x^{n(d-1)}(x-d)^n,$$

yielding the $\mathcal{M}$-transforms

$$\mathcal{M}_{\mathbf{M}}(x) = \frac{1}{x^2-1}, \quad \mathcal{M}_{\mathbf{V}}(x) = \frac{1}{x-d}. \tag{16}$$

As both are decreasing in their respective domains $(1,\infty)$, (d,∞), we can calculate their $\mathcal{N}$-transforms (see Section 4.1) and get

$$\mathcal{N}_{\mathbf{M}}(y) = \sqrt{\frac{y+1}{y}}, \quad \mathcal{N}_{\mathbf{V}}(y) = \frac{dy+1}{y}. \tag{17}$$

Now, for any matrix $\mathbf{P}$,

$$x^{n(d-1)}\chi_x\left(\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\mathbf{U}\right) = \chi_x\left(\mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\right),$$

since characteristic polynomials are invariant under cyclic rotations – up to a multiplicative factor of x, depending on the difference in dimensions. By Lemma 13, it is enough to prove that

$$\alpha_2 \triangleq \alpha_2\left(\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\right)\right) < 2\sqrt{d-1}, \tag{18}$$

where recall that for a real-rooted polynomial $p(x)$, we let $\alpha_i(p(x))$ denote its i-th largest root. By Lemma 15,

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\right) = (x-d)\left(p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}\right)(x), \tag{19}$$

where $p_{\mathbf{V}}(x)$ and $p_{\mathbf{M}}(x)$ are the polynomials defined by $\chi_x(\mathbf{V}) = (x-d)p_{\mathbf{V}}(x)$ and $\chi_x(\mathbf{M}) = (x-1)p_{\mathbf{M}}(x)$, respectively. Therefore,

$$\mathcal{N}_{p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}}(y) \le \frac{y}{y+1}\mathcal{N}_{p_{\mathbf{M}}}(y)\mathcal{N}_{p_{\mathbf{V}}}(y) < \frac{dy+1}{\sqrt{(y(y+1))}}, \tag{20}$$

where the first inequality follows by Lemma 17, and the second, sharp, inequality follows by Claim 19 and the calculations of $\mathcal{N}_{\mathbf{M}}(y), \mathcal{N}_{\mathbf{V}}(y)$ above.

By Claim 16, $\alpha_2 \le \mathcal{N}_{p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}}(y)$ for any $y > 0$. Choosing the point $y_0 = \frac{1}{d-2}$, we get by the above inequality that

$$\alpha_2 \le \mathcal{N}_{p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}}(y_0) < \frac{dy_0+1}{\sqrt{y_0(y_0+1)}} = 2\sqrt{d-1}. \qquad \blacktriangleleft$$

5.2 Non-regular graphs in the configuration model

In this section, we extend our alternative proof of MSS's theorem, presented in the previous section, to the setting of irregular graphs. The configuration model is well-suited for handling vertices with varying degrees, and the required changes are minor, mostly involving appropriate normalizations.

Before proceeding, we recall the definition of the random walk matrix for irregular graphs. First, recall that the random walk matrix of a d-regular graph G, with adjacency matrix $\mathbf{A}$, is given by $\mathbf{W} = \frac{1}{d}\mathbf{A}$, and we denote its eigenvalues by $1 = \omega_1 \geq \omega_2 \geq \cdots \geq \omega_n \geq -1$. The *normalized spectral expansion* of G, denoted $\omega(G)$, is defined as $\max(|\omega_2|, |\omega_n|) \in [0,1]$. When G is *not* regular, the random walk matrix is defined as $\mathbf{W} = \mathbf{A}\mathbf{D}^{-1}$, where $\mathbf{D}$ is the diagonal matrix of vertex degrees (we assume no isolated vertices). This generalizes the above definition for regular graphs. The parameter $\omega(G)$ governs the convergence rate of random walks on G. It is easy to see that the eigenvalues of $\mathbf{W}$ are the same as those of the symmetric matrix $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. Since we are primarily interested in the eigenvalues, we will treat this symmetric matrix as the random walk matrix in what follows.

To derive the analogue in the irregular case, we introduce the following notation. Let $d_1, \ldots, d_n$ be a degree sequence such that $N = \sum_{i=1}^{n} d_i$ is even. Define $j_k = \sum_{i=1}^{k} d_i$ as the sum of the first k degrees, and let $J_k \subset [N]$ be the set of indices

$$J_k = \{j_{k-1} + 1, j_{k-1} + 2, \ldots, j_{k-1} + d_k\}.$$

The analogue of the matrix $\mathbf{U}$ in the non-regular case is the $N \times n$ matrix $\mathbf{C}$ defined by $\mathbf{C}e_k = \frac{1}{\sqrt{d_k}}\mathbf{1}_{J_k}$ for every $k \in [n]$. Using this notation, we obtain

$$\mathbf{W}_{\mathbf{P}} = \mathbf{C}^{\mathsf{T}}\mathbf{P}\mathbf{M}\mathbf{P}^{\mathsf{T}}\mathbf{C}, \tag{21}$$

where $\mathbf{M}$ is the adjacency matrix of a perfect matching on the N half-edges, as in the regular configuration model. With this, we are ready to prove Theorem 1. The proof follows similar lines to that of Theorem 21, and so we refrain from repeating it in full, instead focusing on the differences.

Proof of Theorem 1. First, observe that $\mathbf{C}\mathbf{C}^{\mathsf{T}}$ is the block-diagonal matrix, which we denote by $\bar{\mathbf{V}}$, consisting of n blocks of the form $\frac{1}{d_i}\mathbf{J}_{d_i}$. Note that

$$\chi_x\left(\bar{\mathbf{V}}\right) = (x-1)^n x^{N-n}.$$

A straightforward calculation gives $\mathcal{N}_{\bar{\mathbf{V}}}(y) = \frac{\bar{d}y+1}{\bar{d}y}$. Following the same steps as in the proof of Theorem 21, we arrive at the analogue of Equation (20),

$$\mathcal{N}_{p_{\bar{\mathbf{V}}} \boxtimes p_{\mathbf{M}}}(y) \leq \frac{y}{y+1} \cdot \mathcal{N}_{p_{\mathbf{M}}}(y) \cdot \mathcal{N}_{p_{\bar{\mathbf{V}}}}(y) < \frac{\bar{d}y+1}{\bar{d}\sqrt{y(y+1)}}.$$

The proof follows as the right-hand side attains its minimum value $\frac{2\sqrt{\bar{d}-1}}{\bar{d}}$ (at $y_0 = \frac{1}{\bar{d}-2}$). ◀

6 The Free Method

The analysis in Section 5.2 deals with non-regular graphs directly for the normalized spectrum (of the matrix $\mathbf{W}$), which governs the convergence of random walks (among other characteristics). This raises the question of what can be said about the adjacency spectrum of $\mathbf{A}$, which is of importance when analyzing various generalizations of the expander mixing lemma (see [1] for an up to date survey of the latter).

There are several frameworks for using results from free probability to deduce analysis on graph distributions. One of these is the *finite free* approach using interlacing families of polynomials, pioneered in [23] (and later used in [26, 10] and others), and exemplified here in Section 5.2. Others are the strong convergence approaches of Bordenave and Collins [7] and Chen et al. [8]. All these approaches can be looked at as a two-step process:

1. A reduction from a combinatorial question to a setting of free random variables.
2. Solution in the free setting.

The reduction of Item 1 can yield either an existence proof (as in the MSS approach used here in Theorem 1), or a high probability result as in [7, 8], reduced to the solution of Item 2. The latter, however, is an elusive topic which did not get nearly enough attention in the literature, and we overview it next. We then analyze specific examples in the free regime and verify that the results well represent the behavior of random graphs.

6.1 Free method procedures

In this section, we present a couple of procedures within the framework of the free method. We begin with a brief overview of techniques used in free probability for computing distributions of sums and products of free random variables. In our context, we think of these as the eigenvalue distributions of matrices, for which the eigenspaces are maximally decoupled (for example, by rotating one matrix using a Haar unitary). However, in the free probability literature this context is not necessary, as there are clean self-contained definitions of free random variables expressed in terms of joint moments (for an excellent textbook on the topic, we refer the reader to [27]).

Let a and b be *free* random variables, and denote their distributions by μ_a and μ_b. We wish to find the distribution μ_c of the variable $c = a + b$, by which we define the free additive convolution of distributions $\mu_c = \mu_a \boxplus \mu_b$. This represents, in matrix terms, the eigenvalue distribution of $\mathbf{A} + \mathbf{Q}^\mathsf{T}\mathbf{B}\mathbf{Q}$ where $\mathbf{Q}$ is a Haar unitary and $\mathbf{A}, \mathbf{B}$ are square matrices with eigenvalue distributions μ_a, μ_b respectively. The recipe doing so is the following, provided here without proof[2]:

Algorithm 1. Free additive convolution.

1. Calculate the Cauchy transforms $\mathcal{G}_{\mu_a}(x)$ and $\mathcal{G}_{\mu_b}(x)$ as defined in Section 4.1.
2. Find the compositional inverses, denoted by $\mathcal{K}_{\mu_a}(y)$ and $\mathcal{K}_{\mu_b}(y)$ (as defined in Section 4.1).
3. Calculate $\mathcal{K}_{\mu_c}(y) = \mathcal{K}_{\mu_a}(y) + \mathcal{K}_{\mu_b}(y) - \frac{1}{y}$.
4. Invert $\mathcal{K}_{\mu_c}(y)$ to get $\mathcal{G}_{\mu_c}(x)$.
5. Calculate μ_c from $\mathcal{G}_{\mu_c}(x)$ using the Stieltjes inversion formula (details omitted).

Note that when at least one of the distributions is supported on non-negative reals only (in the context of matrices, at least one of them is PSD) there is a multiplicative analog of the above recipe – that is, the calculation of the distribution $\mu_c = \mu_a \boxtimes \mu_b$ for $c = ab$ – where the $\mathcal{G}$ and $\mathcal{K}$ transforms are replaced with $\mathcal{M}$, $\mathcal{N}$ respectively, and Item 3 is replaced with $\mathcal{N}_{\mu_c}(y) \leq \frac{y}{y+1} \cdot \mathcal{N}_{\mu_a}(y) \cdot \mathcal{N}_{\mu_b}(y)$.

The vast majority of the literature focuses on proofs of the above, and its generalizations. However, explicit applications of it as stated are not very common, due to a few clear caveats. The first is that for Item 1 we need to have explicit analytic formulas for μ_a, μ_b, which

[2] The proof is standard in the free probability literature, and can be found in [27, Chapter 12].

may not be available. The second is that inverting a function is a complicated, sometimes infeasible, task. Given a distribution μ supported on only 3 values, inverting $\mathcal{G}_\mu$ involves solving a cubic equation, which may be possible but leads to a function that is hard to analyze. For more complicated distributions, we know the task to be infeasible, making Items 2 and 4 mere theoretical ideas and not practical tools.

A solution to the inversion problem of Item 4 is implicit in the work of Marcus, Spielman and Srivastava. Once we relax our goal of having an explicit formula for μ_c and wish to only *bound* its support, the only necessary part is knowing $\mathcal{K}_{\mu_c}(y)$. This is due to the observation (discussed here in Section 4.1) that every value of this function is an upper bound on the support. Therefore Items 4 and 5 can be replaced with minimizing $\mathcal{K}_{\mu_c}(y)$ over $y > 0$, which is typically a simple analytic task, leading to the following simpler procedure:

Algorithm 2. Bounding the free additive convolution.

- Perform Items 1–3 of Procedure 1 to get $\mathcal{K}_{\mu_c}(y)$.
- Output $\min_{y>0} \mathcal{K}_{\mu_c}(y)$.

Although this procedure is simpler, we are still left with the case of Item 2 not being feasible, which is the issue we wish to solve next in Section 6.1.1. Note that other works have looked into relaxing the requirement to get a closed formula for μ_c, to rather learn characteristics of its support [3] or develop numerical techniques for this task (see [12, 29] and the subordination technique presented in [34]). To the best of our knowledge, although the technique presented here is elementary, this work is the first to formally define it.

6.1.1 Sums of identical spectra

The procedure given in the next lemma solves the inversion issue of Item 2 for the case that the random variables have an equal distribution μ. We use $\mu^{\boxplus t}$ to denote the additive convolution of μ with itself $t-1$ times.

▶ **Lemma 22.** *Let μ be a compactly supported distribution on $\mathbb{R}$, and $t \in \mathbb{N}$. Then the supremum of the support of $\mu^{\boxplus t}$ is achieved by minimizing*

$$\widetilde{\mathcal{K}_t}(x) \triangleq tx - \frac{t-1}{\mathcal{G}_\mu(x)}$$

over $\{x > \max(\mu)\}$, or equivalently applying the following procedure:

Algorithm 3. Bounding the support of $\mu^{\boxplus t}$.

1. Find the solution x_0 of the equation
$$\frac{t-1}{t}\mathcal{G}'_\mu(x) = -\mathcal{G}_\mu(x)^2. \tag{22}$$
2. Evaluate $y_0 \triangleq \mathcal{G}_\mu(x_0)$.
3. Output $t \cdot x_0 - \frac{t-1}{y_0}$.

Notice that at no point in Lemma 22 do we need to invert a function.

Proof. By definition of inverse under composition, $\mathcal{G}_\mu(\mathcal{K}_\mu(y)) = y$. Differentiating both sides and rearranging gives

$$\mathcal{K}'_\mu(y) = \frac{1}{\mathcal{G}'_\mu(\mathcal{K}_\mu(y))}. \tag{23}$$

As shown, we look for the minimal value of

$$\mathcal{K}_{\mu_t}(y) = t \cdot \mathcal{K}_\mu(y) - \frac{t-1}{y} \tag{24}$$

achieved at a point y_0 satisfying $\mathcal{K}'_{\mu_t}(y_0) = 0$. Note that such a minimum always exists, as $\mathcal{G}_\mu(x)$ decays like $\frac{1}{x}$. Therefore by Equation (24)

$$t \cdot \mathcal{K}'_\mu(y_0) = -\frac{t-1}{y_0^2}.$$

Plugging in Equation (23) one concludes

$$\frac{t}{\mathcal{G}'_\mu(\mathcal{K}_\mu(y_0))} = -\frac{t-1}{y_0^2}.$$

Let x_0 be the value such that $y_0 = \mathcal{G}_\mu(x_0)$. Then we get from the above that

$$\frac{t-1}{t}\mathcal{G}'_\mu(x_0) = -\mathcal{G}(x_0)^2,$$

as required in Equation (22). The minimal value of Equation (24) is thus

$$\mathcal{K}_{\mu_t}(y_0) = t \cdot \mathcal{K}_\mu(y_0) - \frac{t-1}{y_0} = t \cdot x_0 - \frac{t-1}{y_0}.$$ ◀

6.1.2 Sums and products of two distinct spectra

The procedure of Lemma 22 is very easy to use, but does not cover cases where we need to perform an additive convolution of different distributions μ_1 and μ_2. A straightforward variation of Lemma 22 shows that when one distribution has a simple Cauchy transform $\mathcal{G}_{\mu_1}(x)$ with an explicit inverse $\mathcal{K}_{\mu_1}(y)$, the supremum of the support of $\mu_1 \boxplus \mu_2$ can be found using the following lemma.

▶ **Lemma 23.** *Let μ_1, μ_2 be compactly supported distributions on $\mathbb{R}$. The supremum of the support of $\mu_1 \boxplus \mu_2$ is achieved by minimizing over $\{x > \max(\mu_2)\}$ the function*

$$\widetilde{\mathcal{K}}_{\mu_1 \boxplus \mu_2}(x) \triangleq \mathcal{K}_{\mu_1}(\mathcal{G}_{\mu_2}(x)) + x - \frac{1}{\mathcal{G}_{\mu_2}(x)}. \tag{25}$$

If either μ_1 or μ_2 is supported on $\mathbb{R}^+$ only, the supremum of the support of $\mu_1 \boxtimes \mu_2$ is achieved by minimizing over $\{x > \max(\mu_2)\}$ the function

$$\widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x) \triangleq \frac{\mathcal{M}_{\mu_2}(x)}{\mathcal{M}_{\mu_2}(x)+1} \cdot \mathcal{N}_{\mu_1}(\mathcal{M}_{\mu_2}(x)) \cdot x. \tag{26}$$

Proof. We will prove the additive case, the multiplicative follows similar arguments. By Item 3 of Procedure 1, we need to minimize

$$\mathcal{K}_{\mu_1 \boxplus \mu_2}(y) = \mathcal{K}_{\mu_1}(y) + \mathcal{K}_{\mu_2}(y) - \frac{1}{y}.$$

Set $y = \mathcal{G}_{\mu_2}(x)$, and we get the equation

$$\widetilde{\mathcal{K}}_{\mu_1 \boxplus \mu_2}(x) = \mathcal{K}_{\mu_1}(\mathcal{G}_{\mu_2}(x)) + x - \frac{1}{\mathcal{G}_{\mu_2}(x)},$$

completing the proof. ◀

6.2 Application to ultra-sparse graphs

As previewed in Section 2.2, we predict spectral bounds for our graph models by analyzing the spectra of their constituent pieces and how they combine. Our prediction for the resulting graph is the free convolution of the component spectral measures – additive ($\boxplus$) or multiplicative ($\boxtimes$), depending on the model. For each model we apply the appropriate theorem from Section 6.1 and express the bound as a function of the sparsity ε, denoted $\mathsf{pred}_{\mathsf{model}}(\varepsilon)$, where we substitute the suitable model in the subscript.

Free method predictions for the configuration model

As mentioned above, in order to apply the techniques of Section 6.1 to random graph models, one needs to understand the distributions involved. For the configuration model, the convolution is of the eigenvalue distribution of the matrix $\mathbf{M}$, which we already know to be the uniform distribution on ± 1 and we denote by μ_1, and the one of $\mathbf{V} = \mathbf{U}\mathbf{U}^{\mathsf{T}}$ which we denote by μ_2. The latter's form is defined by the distribution of degrees. In the case of a graph with degrees 2 and 3 and average degree $2+\varepsilon$, a straightforward calculation shows its Cauchy transform to be

$$\mathcal{G}_{\mu_2}(x) = \frac{1}{2+\varepsilon}\left(\frac{1+\varepsilon}{x} + \frac{\varepsilon}{x-3} + \frac{1-\varepsilon}{x-2}\right).$$

By Equation (17) we have that $\mathcal{N}_{\mu_1}(y) = \sqrt{\frac{y+1}{y}}$. Thus, applying Lemma 23 gives us that $\sup(\mu_1 \boxtimes \mu_2)$ is found by minimizing over $x > 3$ the function

$$\widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x) = x \cdot \sqrt{\frac{\mathcal{M}_{\mu_2}(x)}{\mathcal{M}_{\mu_2}(x)+1}} = x \cdot \sqrt{\frac{x\mathcal{G}_{\mu_2}(x)-1}{x\mathcal{G}_{\mu_2}(x)}}. \tag{27}$$

This leads us to the main result of this section.

▶ **Lemma 24** (Small–ε bound for $\mathsf{pred}_{\mathsf{conf}}(\varepsilon)$)**.**

$$\mathsf{pred}_{\mathsf{conf}}(\varepsilon) = \frac{3}{\sqrt{2}} + \frac{3}{4}\varepsilon^{1/2} + \frac{9\sqrt{2}}{32}\varepsilon - O(\varepsilon^{3/2}).$$

The proof uses elementary means and appears in the full version of this paper.

Free method predictions for the $\mathcal{C} + \varepsilon\mathcal{M}$, $(2+\varepsilon)\mathcal{M}$ and $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ models appear in the full version of the paper.

References

1 Aida Abiad and Sjanne Zeijlemaker. A unified framework for the expander mixing lemma for irregular graphs and its applications. *Linear Algebra and its Applications*, 702:19–45, 2024.

2 Noga Alon, Jeff Edmonds, and Michael Luby. Linear time erasure codes with nearly optimal recovery. In *36th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 512–519, 1995.

3 Zhigang Bao, László Erdős, and Kevin Schnelli. On the support of the free additive convolution. *Journal d'Analyse Mathématique*, 142(1):323–348, 2020.

4 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. doi:10.1137/080732651.

5 Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006. doi:10.1007/s00493-006-0029-7.

6 Charles Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(6):1393–1439, 2020. doi:10.24033/asens.2450.

7 Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Annals of Mathematics*, 190(3):811–875, 2019.

8 Chi-Fang Chen, Jorge Garza-Vargas, Joel A Tropp, and Ramon van Handel. A new approach to strong convergence. *Annals of Mathematics*, 2024.

9 Gil Cohen, Itay Cohen, and Gal Maor. Tight bounds for the zig-zag product. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, 2024.

10 Gil Cohen and Gal Maor. Random walks on rotating expanders. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, 2023.

11 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. doi:10.1109/FOCS.2016.37.

12 Alice Cortinovis and Lexing Ying. Computing free convolutions via contour integrals. *arXiv preprint*, 2023. doi:10.48550/arXiv.2305.01819.

13 Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007. doi:10.1145/1236457.1236459.

14 Irit Dinur and Oded Goldreich. Expansion without connectivity: A property testing perspective. *ECCC*, 2025.

15 Shlomo Hoory. A lower bound on the spectral radius of the universal cover of a graph. *Journal of Combinatorial Theory, Series B*, 93(1):33–43, 2005. doi:10.1016/J.JCTB.2004.06.001.

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

17 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. doi:10.1145/195058.195190.

18 Kazimierz Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematicae*, 15:271–283, 1930. doi:10.4064/fm-15-1-271-283.

19 Alexander Lubotzky. Finite simple groups of Lie type as expanders. *Journal of the European Mathematical Society*, 13(5):1331–1341, 2011.

20 Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. doi:10.1007/BF02126799.

21 Michael G Luby, Michael Mitzenmacher, Mohammad Amin Shokrollahi, and Daniel A Spielman. Improved low-density parity-check codes using irregular graphs. *IEEE Transactions on information Theory*, 47(2), 2002. doi:10.1109/18.910576.

22 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. doi:10.4007/annals.2015.182.1.7.

23 Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families IV: Bipartite ramanujan graphs of all sizes. *SIAM Journal on Computing*, 47(6):2488–2509, 2018. doi:10.1137/16M106176X.

24 Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Finite free convolutions of polynomials. *Probab. Theory Related Fields*, 182(3-4):807–848, 2022. doi:10.1007/s00440-021-01105-w.

25 G. A. Margulis. Explicit constructions of graphs without short cycles and low density codes. *Combinatorica*, 2(1):71–78, 1982. doi:10.1007/BF02579283.

26 Sidhanth Mohanty and Ryan O'Donnell. X-Ramanujan graphs. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms*, pages 1226–1243. SIAM, Philadelphia, PA, 2020. doi:10.1137/1.9781611975994.75.

27 Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006. doi:10.1017/CBO9780511735127.

28 A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991. doi:10.1016/0012-365X(91)90112-F.

29 Sheehan Olver and Raj Rao Nadakuditi. Numerical computation of convolutions in free probability theory. *arXiv preprint*, 2012. arXiv:1203.1958.

30 Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):Art. 17, 24, 2008. doi:10.1145/1391289.1391291.

31 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. doi:10.1109/SFCS.2000.892006.

32 Thomas J Richardson, Mohammad Amin Shokrollahi, and Rüdiger L Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE transactions on information theory*, 47(2), 2002. doi:10.1109/18.910578.

33 Michael Sipser and Daniel A. Spielman. Expander codes. In *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 566–576, 1994. doi:10.1109/SFCS.1994.365734.

34 Roland Speicher. Lecture notes on free probability theory. *arXiv preprint*, 2019. arXiv:1908.08125.

35 Daniel A. Spielman. Spectral and algebraic graph theory. Unpublished book manuscript, 2025. URL: http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.

36 Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017. doi:10.1145/3055399.3055408.

37 Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/0400000010.

38 Pinchen Xie, Zhongzhi Zhang, and Francesc Comellas. The normalized laplacian spectrum of subdivisions of a graph. *Applied Mathematics and Computation*, 286:250–256, 2016. doi:10.1016/J.AMC.2016.04.033.