Electronic Colloquium on Computational Complexity, Report No. 196 (2025)

Ultra-Sparse Expanders and the Free Method

Gil Cohen* Gal Maor†

November 27, 2025

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.

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.

Our results also extend to expansion in general irregular graphs. Moreover, as a byproduct, we give two alternative proofs of the Marcus–Spielman–Srivastava existence theorem for one-sided Ramanujan graphs.

*Tel Aviv University. gil@tauex.tau.ac.il. Supported by ERC starting grant 949499 and by the Israel Science Foundation grant 2989/24.

†Tel Aviv University. galmaor@mail.tau.ac.il. Supported by ERC starting grant 949499.

ISSN 1433-8092

Contents

1 Introduction

Expanders are sparse yet highly connected, well-mixing graphs. Their sparsity makes them "cost-effective", while their expansion properties allow them—up to some error—to be treated as if they were complete graphs, which are ideal but too expensive to use. They have been studied extensively for decades, with construction techniques ranging from combinatorial [RVW00, BL06, BATS11] and group-theoretic [LPS88, Mar82] to analytic [MSS15]. They have a myriad of applications in coding theory [SS94, AEL95, TS17], computational complexity [INW94, Din07, Rei08], and beyond. For background, see, e.g., the survey [HLW06], Chapter 4 of [Vad12], or the book manuscript [Spi25].

In this paper we ask for the optimal expansion when the edge budget is just above connectivity—the almost-tree regime. Specifically, for graphs with average degree $2+\varepsilon$, what is the "Ramanujan bound"—how does spectral expansion depend on ε? Such *ultra-sparse* graphs are necessarily irregular, so to understand the full landscape we use both the normalized and unnormalized (adjacency-based) notions of expansion. For completeness, we first briefly recall the standard definitions.

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 refer to λ_2 as the *one-sided spectral expansion* of G, and to $\lambda_1 - \lambda_2$ as the *one-sided spectral gap*. The quantity $\lambda \triangleq \max\{|\lambda_2|, |\lambda_n|\}$ is the (two-sided) *spectral expansion* of G. In the normalized setting, let D be the diagonal degree matrix and let $W \triangleq AD^{-1}$ be 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* of G, $1-\omega_2$ is the *one-sided normalized spectral gap*, and $\omega \triangleq \max\{|\omega_2|, |\omega_n|\}$ is the (two-sided) *normalized spectral expansion*.

Clearly, a 2-regular graph is not a good expander, as the only connected 2-regular graph is a cycle, whose spectral gap vanishes as the number of vertices grows. It is well-known that 3-regular graphs can achieve normalized spectral expansion of $\frac{\sqrt{8}}{3}$. The exact value is of secondary importance in most applications; what matters is that it is bounded away from 1. More generally, for every integer $d \geq 3$ the best possible (asymptotic) normalized spectral expansion for d-regular graphs, $\frac{2\sqrt{d-1}}{d}$ [Nil91], is attained by the so-called *Ramanujan graphs*. Thus, as one increases the degree d (making the graph more expensive to use), the spectral expansion improves, making the graph a better expander.

1.1 Ultra-sparse expanders

Given that the chief utility of expanders is having normalized spectral expansion bounded away from 1, we ask how sparse a graph can be while still keeping this quantity bounded away from 1—even if it exceeds the "regular-graph barrier" $\frac{\sqrt{8}}{3} \approx 0.94$—and we aim to characterize the precise sparsity–expansion tradeoff in this regime.

Connectivity is the minimal prerequisite for expansion, and since any connected n-vertex

graph has average degree at least $2-\frac{2}{n}$, we focus on graphs with average degree $2+\varepsilon$, where $\varepsilon > 0$ controls the sparsity. As mentioned, our goal is to understand how the achievable expansion depends on ε—that is, to determine the "Ramanujan bound" for ultra-sparse graphs with parameter ε.

To minimize degree variance, we primarily consider graphs whose degrees lie in $\{2, 3\}$. High-degree vertices can, in some applications, impose a disproportionate workload on a single node. Although hub-and-spoke graphs—e.g., the wheel graph (see Section 5.4)—can be even more amortization-efficient than 3-regular graphs, we avoid them for this reason. Our techniques, however, extend readily to broader degree distributions, and so we prioritize simplicity to clarify the landscape of expanders in this largely unexplored regime. As we will see, while this work advances our understanding of this setting, it also leaves basic questions open and introduces techniques of independent interest.

As mentioned, ultra-sparse graphs are inherently irregular; accordingly, the *unnormalized* spectral expansion reflects different properties than the *normalized* spectral expansion. The normalized spectrum reflects random walk behavior, whereas the unnormalized spectrum is better suited to analyzing edge expansion and expander–mixing–lemma–type bounds—each accounting for degree heterogeneity in different ways (see [AZ24] for a survey). Therefore, in this work we also study the unnormalized spectral expansion of ultra-sparse graphs. In general, for irregular graphs, the unnormalized spectral expansion has been less studied than the normalized one, in part because irregularity can obscure its usefulness. When the degree variance is small—particularly in ultra-sparse graphs with degrees in $\{2, 3\}$—the parameter nevertheless reveals meaningful aspects of the graph's combinatorial structure.

There are two general frameworks for constructing Ramanujan graphs. The first is based on Cayley graphs of carefully chosen non-abelian groups, originating in the seminal works of Lubotzky, Phillips, and Sarnak [LPS88], and Margulis [Mar82]. Departing from this decades-old, deep group-theoretic approach, Marcus, Spielman, and Srivastava (MSS) initiated a new, more analytic framework, proving the existence of one-sided Ramanujan graphs in a celebrated sequence of papers [MSS15, MSS18, MSS22]. The construction originating in this line of work, which draws inspiration from free probability theory, was later made explicit by Cohen [Coh16].

The group-theoretic approach seems ill-suited to our goal of constructing ultra-sparse graphs, as much of the appeal of Cayley graphs lies in their high degree of symmetry—which, in particular, enforces regularity. That said, there are ways to generate irregular graphs from groups, for instance by considering Schreier graphs instead of Cayley graphs. We leave this direction for future investigation.

The MSS approach to Ramanujan graphs analyzes the union of d perfect matchings that are, in an appropriate sense, uncorrelated, and it targets one-sided expansion. Although this framework does not directly fit our setting, it will serve as our main point of departure.

1.2 Five models of ultra-sparse graphs

As hinted above, the MSS approach studies a distribution over graphs and proves that at least one graph in its support is (one-sided) Ramanujan. Specifically, in the setting of d-regular graphs, they examined the distribution—or model—induced by the union of d perfect matchings. This raises a natural question: what model is appropriate for generating ultra-sparse graphs? We propose five natural models.

Two-and-a-fraction matchings. The first is perhaps closest in spirit to the MSS distribution. In this model, denoted suggestively as $\mathcal{M} + \mathcal{M} + \varepsilon\mathcal{M}$, or $(2+\varepsilon)\mathcal{M}$ for short, we consider the union of two perfect matchings and a third ε-matching—that is, a partial matching covering an ε-fraction of the vertices. Note that at the extremes $\varepsilon = 0$ and $\varepsilon = 1$, one recovers the MSS distributions for 2-regular and 3-regular graphs, respectively.

Three equally sized partial matchings. An alternative is the union of three β-matchings, where $\beta = \frac{2+\varepsilon}{3}$. This distribution is denoted by $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$. Upon brief reflection, the reader may lean toward the $(2+\varepsilon)\mathcal{M}$ distribution, as the alternative appears to struggle even with basic connectivity. Indeed, the expected fraction of isolated vertices in the $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ model is $\left(\frac{1-\varepsilon}{3}\right)^3 \approx \frac{1}{27}$ for small ε. Nonetheless, we explore this distribution as well. Our goal is not to pick the "best" model, but to better understand each of them.

Cycle plus partial matching. In light of the connectivity issues of the previous model, one may reconsider the use of perfect matchings altogether. In particular, we propose a model in which one takes the union of a cycle passing through all vertices and an ε-partial matching. We denote this model by $\mathcal{C} + \varepsilon\mathcal{M}$. In a certain sense, this distribution seems to improve upon both previous models: the presence of the cycle guarantees connectivity, whereas the first two perfect matchings in $(2+\varepsilon)\mathcal{M}$ do not.

It is also worth recalling a well-known explicit construction [Lub11] of 3-regular expanders formed by the union of a cycle and a matching, defined as follows: the vertex set is identified with a finite field of prime order, and the neighbors of a vertex—represented by a field element x—are $x+1$, $x-1$ (forming the cycle), and x^{-1} (providing the matching). Another similar model was recently analyzed from a property testing perspective in [DG25].

Exploring this model naturally leads to questions about the standard setting of regular graphs. For instance, what can be said about the model formed by the union of $\frac{d}{2}$ cycles, and how does it compare to the MSS model, which considers the union of d perfect matchings? More broadly, if one views the cycle as the optimal 2-regular expander, it is natural to ask what can be said about models formed by the union of several Ramanujan graphs. In addition to our focus on ultra-sparse expanders, we investigate these questions (see

Section 2.4). As a side result, we give an alternative proof of the MSS existence theorem for one-sided Ramanujan graphs, using a distribution generated by unions of cycles rather than unions of perfect matchings.

The configuration model. In this model a perfect matching is selected over a set of "half-edges". Specifically, each vertex v designated to have degree d_v (in the case of ultra-sparse graphs, d_v is either 2 or 3, depending on v) contributes d_v half-edges to a global pool. A perfect matching is then chosen on this pool, thereby determining the edges of the graph.

The configuration model is widely studied in the context of bounding the spectrum of random graphs (e.g. [Bor20]), as well as in the analysis of graph operators such as the zig-zag product [RVW00, CCM24]. Unlike the MSS model—which considers the union of d perfect matchings—the configuration model is naturally suited for irregular graphs, as it easily accommodates varying degrees across vertices. As a byproduct of our study of this model, we obtain a second alternative proof of the MSS result in the standard regular setting.

Subdividing 3-regular graphs. The notion of subdivision—replacing each edge by a path by inserting degree-2 vertices—goes back at least to Kuratowski [Kur30], who characterized planar graphs via subgraphs homeomorphic to K_5 and $K_{3,3}$. With this in mind, a natural (and final) model suggests itself: start from any 3-regular Ramanujan graph (or, more generally, any d-regular Ramanujan graph) and replace each edge by a long path, inserting many degree-2 vertices along it. As the paths get longer, the resulting graph's average degree drifts down toward 2 from above; by choosing the path length appropriately, we can make the average degree as close to $2 + \varepsilon$ as we like. Intuitively, the edges are "stretched" into paths, and we retain the global structure of the original graph while operating in the near–degree-two regime.

One advantage of this fifth model is explicitness: if the base graph is explicit, then the subdivided ultra-sparse graph is explicit as well; likewise for strong explicitness.

2 Our Results

In this section, we present our results. We find the five ultra-sparse graph models introduced above to be natural and compelling. As noted earlier, this work advances their study while raising more questions than it answers, in both the normalized and unnormalized settings.

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 2.1 (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*

$$\frac{2\sqrt{\bar{d}-1}}{\bar{d}}, \tag{1}$$

where $\bar{d}$ is the average degree.

Our proof is based on the configuration model (rather than the perfect matchings model used in [MSS18]) 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 2.2 (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*

$$\frac{2\sqrt{1+\varepsilon}}{2+\varepsilon} = 1 - \frac{\varepsilon^2}{8} + O(\varepsilon^3). \tag{2}$$

At first glance, the bound in Theorem 2.1 resembles the Ramanujan bound for d-regular graphs. Moreover, empirical evidence suggests that the typical behavior of random configuration graphs is indeed captured by Equation (1) (and therefore by Equation (2) for ultra-sparse graphs). 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. 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 2.1 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 2.2 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 2.3 (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 to be a 3-regular Ramanujan graph, the guaranteed spectral expansion for the resulting ultra-sparse graph 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 2.2. In particular, the syntactically Ramanujan-like bound in Theorem 2.1 is not tight.

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 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. This leads to the following open problem.

Open Problem 2.4. *What is the optimal normalized spectral gap for graphs with degrees 2 and 3 and average degree $2 + \varepsilon$?*

We conjecture that the true answer to Open Problem 2.4 is $1 - C \cdot \varepsilon^2 - o(\varepsilon^2)$ for some constant C.

We also remark that applying Theorem 2.3 starting with d-regular Ramanujan graphs with increasing degree d improves the spectral gap. In the limit, since

$$\lim_{d\to\infty} \frac{d}{2(d-2)}\arccos\left(\frac{2\sqrt{d-1}}{d}\right) = \frac{\pi}{4},$$

Theorem 2.3 yields ultra-sparse graphs with degrees 2 and d whose spectral expansion is

$$\cos\left(\tfrac{\pi}{4}\,\varepsilon\right) \;=\; 1 - \frac{\pi^2}{32}\,\varepsilon^2 + O(\varepsilon^4) \approx 1 - 0.3084\varepsilon^2.$$

It is instructive to place our bound in the broader context of expansion for irregular graphs, beyond the ultra-sparse regime. As noted, Theorem 2.1 syntactically resembles the Ramanujan bound for d-regular graphs. However, optimality in the irregular setting is subtler. Lower bounds analogous to Alon–Boppana are known: for example, Chung [Chu16] lower-bounded the normalized spectral expansion of an irregular graph by

$$\frac{2\sum_i d_i\sqrt{d_i - 1}}{\sum_i d_i^2}. \tag{3}$$

This bound is not known to be tight in general, and Theorem 2.1 does not match it—the bounds drift further apart as the degree variance increases. Moreover, as discussed above, one can beat Equation (3) for certain degree sequences.

For the ultra-sparse regime with degrees in $\{2, 3\}$, Equation (3) yields

$$\frac{4 + (6\sqrt{2} - 4)\varepsilon}{4 + 5\varepsilon} = 1 - \frac{9 - 6\sqrt{2}}{4}\,\varepsilon + O(\varepsilon^2), \tag{4}$$

leaving a gap relative to our bounds–not only in the constants but also in the asymptotic dependence on ε. A more general question than Open Problem 2.4 is the following, concerning the best expansion for a given degree sequence.

Open Problem 2.5. What is the optimal normalized spectral expansion for a graph with degree sequence $(d_1, \ldots, d_n)$?

Unlike the d-regular case—where random graphs are expected to be nearly optimal (cf. Friedman's theorem [Fri08, Bor20])—imposing structure on a non-regular graph can substantially improve expansion. We discuss this difference in Section 5.4. However, because our focus is on ultra-sparse graphs, we defer a systematic study of irregular-graph expansion to future work, and we expect the techniques developed here to be useful in that setting.

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 2.6 (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*

$$\lambda = \frac{\sinh((k+1)\xi) - (d-1)\sinh((k-1)\xi)}{\sinh(\xi)}. \tag{5}$$

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 2.6 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. A celebrated example is Ramsey graphs. In the paper that inaugurated the probabilistic method, Erdős [Erd47] gave a one-line proof that for $k = (2+o(1))\log n$ there exist n-vertex graphs that are k-Ramsey. Many decades later—and despite significant advances relying on fairly involved techniques (see [FW81, BRSW12, Coh21, CZ19, Li23] and references therein)—we still do not have an explicit construction matching Erdős's bound. Similar stories abound: in coding theory, for the distance–rate tradeoff captured by the Gilbert–Varshamov bound [Gil52, Var57] and in complexity theory, for pseudorandom generators and randomness extractors.

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.

The underlying philosophy is that such constructions *should* exist, even though the probabilistic method does not provide them. It is important to note, however, that this is not guaranteed—and in generic, unstructured settings it is seldom true. Belief in the existence of explicit constructions typically stems from the structure of the problem—structure not exploited by the probabilistic method itself. Thus its quantitative prediction is a guide—often trustworthy in natural settings—but still only a prediction. The method's power lies in:

(i) **Ease of use.** Benchmarks are derived quickly, especially compared with the extreme difficulty of constructing explicit objects that even approach them.

(ii) **Evidence.** Across many examples over the decades the predictions have proved reliable; explicit constructions often come close (though they rarely achieve them).

(iii) **Scope.** The method applies across a wide range of combinatorial problems.

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 situate the method relative to the three merits of the probabilistic method discussed earlier:

(i) **Ease of use.** The free method is likewise fairly straightforward to apply. Much as the probabilistic method has standard techniques, we provide practical procedures

requiring no background in free probability and often reducing to solving a scalar equation. We hope future work will extend them.

(ii) **Verification.** Being new, the method lacks accumulated evidence of correctness. However—and this is a distinctive advantage—its predictions are easy to *empirically verify* by experiments. Indeed, just as the probabilistic method typically asserts that a random object attains the benchmark, the free method appears to capture typical behavior observable in efficient experiments.

(iii) **Scope.** The free method seems quite general within graph-theoretic frameworks: it applies across various models while relying only on their spectral descriptions.

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:

$$\mathcal{G}_G(x) \;=\; \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x-\lambda_i}.$$

This encodes exactly the spectrum—and nothing else. In this paper we restrict $\mathcal{G}_G$ to the domain (λ_1,∞), following the convention in *finite* free probability. By contrast, classical free probability treats the Cauchy transform as an analytic function on an open subset of $\mathbb{C}\setminus\{\lambda_i\}$, most commonly the open upper half-plane. Our use of the free method draws on both perspectives.

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

$$(t-1)\,\mathcal{G}_G'(x)+t\,\mathcal{G}_G(x)^2 \;=\; 0. \tag{6}$$

2. The prediction is given by

$$t\cdot x_0 \;-\; \frac{t-1}{\mathcal{G}_G(x_0)}. \tag{7}$$

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 Deriving the classical $2\sqrt{d-1}$ bound using the free method

Let us illustrate the above procedure of the free method in the classical setting of d-regular graphs, constructed as the union of $t = d$ perfect matchings. The adjacency matrix M of a perfect matching has half of its eigenvalues equal to 1 and half equal to -1. Thus, the Cauchy transform of a perfect matching is given by

$$\mathcal{G}_M(x) = \frac{1}{2}\left(\frac{1}{x-1} + \frac{1}{x+1}\right) = \frac{x}{x^2-1},$$

with derivative $\mathcal{G}'_M(x) = -\frac{x^2+1}{(x^2-1)^2}$. Solving Equation (6) for this case yields the positive solution $x_0 = \sqrt{d-1}$. Plugging this into Equation (7), noting that $\mathcal{G}_M(\sqrt{d-1}) = \frac{\sqrt{d-1}}{d-2}$, we obtain the prediction, denoted here as

$$\mathsf{pred}(d) = d\sqrt{d-1} - \frac{d-1}{\mathcal{G}_M(\sqrt{d-1})} = 2\sqrt{d-1},$$

which, of course, matches the well-known and tight spectral bound for d-regular graphs.

A worth-noting observation is that this prediction is independent of n, the number of vertices. This reflects the fact that the free method draws on tools from free probability, where the underlying operators are infinite-dimensional. In this sense, the prediction treats n as if it were infinite.

2.3.2 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 (6) 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 (7) and expressing the result in terms of ε yields the prediction

$$\begin{aligned}\mathsf{pred}_{3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)}(\varepsilon) &= \sqrt{1 + \sqrt{\frac{4-\varepsilon^2}{3}} \cdot \frac{\sqrt{3(4-\varepsilon^2)} - 3\varepsilon}{2(1-\varepsilon)}} \\ &= \sqrt{3 + \sqrt{12}} + \frac{\sqrt{3+\sqrt{12}}}{4+\sqrt{12}}\,\varepsilon + O(\varepsilon^2). \end{aligned} \tag{8}$$

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.

2.3.3 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 2.1. Its application gives

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

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

$$\mathsf{pred}_{\mathcal{C}+\varepsilon\mathcal{M}}(\varepsilon) = \mathsf{pred}_{(2+\varepsilon)\mathcal{M}}(\varepsilon) = \sqrt{5} + \sqrt{\frac{4}{5\phi}}\,\sqrt{\varepsilon}, \tag{10}$$

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 [Hoo05] states that if a graph G has average degree

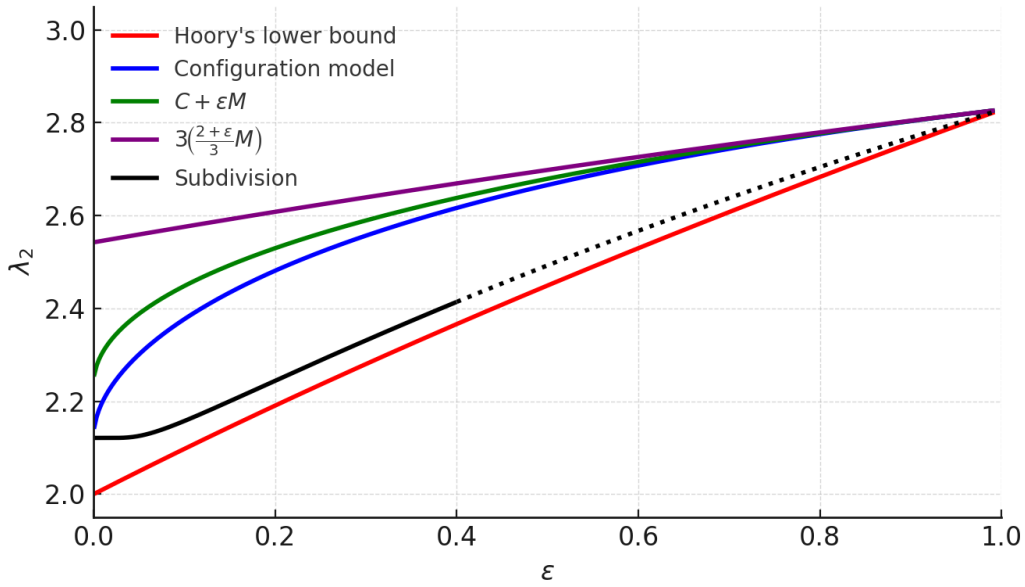


Figure 1: Comparison of the predictions for the bound on λ_2 for the conf, $\mathcal{C} + \varepsilon\mathcal{M}$ and $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ models. Hoory's lower bound of $2\sqrt{\bar{d}-1}$ is displayed in red, and the subdivision model of a 3-regular Ramanujan graph ($\lambda = 2\sqrt{2}$) 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 (5) for real $k < 2$ as a dotted line (see full analysis in Section 3.1).

at least $\bar{d}$ even after deleting any ball of radius r, then

$$\lambda_2(G) \;\geq\; 2\sqrt{\bar{d}-1}\left(1 - c\,\frac{\log r}{r}\right), \tag{11}$$

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 $\varepsilon = 1$.

2.4 Union of Ramanujan graphs and cycles

In the $\mathcal{C} + \varepsilon\mathcal{M}$ model, we considered the union of a full cycle—one that traverses all vertices—with a partial matching. The role of the cycle is to ensure connectivity even before the partial matching is applied. This motivates the natural idea of replacing perfect matchings with cycles, even in the classical setting of d-regular graphs. In particular, for d even, we define the model $\frac{d}{2}\mathcal{C}$, in which a graph is formed by taking the union of $\frac{d}{2}$ full cycles, and ask how it compares to the more standard model—studied by MSS—where the graph is formed as the union of d perfect matchings, denoted by $d\mathcal{M}$.

More generally, if one views a cycle as an optimal 2-regular expander, it is natural to extend this idea by considering any c-regular Ramanujan graph G. One can then ask about the model obtained by taking the union of $\frac{d}{c}$ independent copies of G, assuming for simplicity that c divides the target degree d. Does this distribution—naturally denoted by $\frac{d}{c}G$—contain a d-regular Ramanujan graph, regardless of the specific identity of G?

Starting with the more general question, our main result in this section is that an almost-Ramanujan graph always exists in the support of $\frac{d}{c}G$.

Theorem 2.7 (Union of Ramanujan Graphs)**.** *Let G be a c-regular one-sided Ramanujan graph of girth g. Then, in the support of the distribution $\frac{d}{c}G$, there exists a graph with one-sided spectral expansion*

$$2\sqrt{d-1} + O(\sqrt{d}) \cdot 2^{-\Omega(g)}.$$

A key ingredient in the proof of Theorem 2.7 is the *adapter* trick, introduced in [CM23] and refined in Section 7. This technique effectively replaces the (unknown) spectrum of G with the Kesten–McKay distribution, incurring a small error that depends on the girth of G. To apply it, one needs both that G is Ramanujan and a lower bound on its girth. This shows that the Free Method can be applied with only this information, at the cost of a small error term. However, a recurring theme is that additional structural information about G enables a tighter analysis. For example, in the $\frac{d}{2}\mathcal{C}$ model—where G is a cycle—the spectrum is fully explicit, allowing us to prove Theorem 2.8 with no error term, yielding another proof of the MSS result.

Theorem 2.8 (Union of Cycles)**.** *In the support of the distribution $\frac{d}{2}\mathcal{C}$, there exists a one-sided Ramanujan graph.*

3 Ultra-Sparse Expanders via Subdivision

In this section we prove our results on ultra-sparse expanders in the subdivision model (our fifth model). We first establish Theorem 2.3 in Section 3.1, and then prove the slightly more involved Theorem 2.6 in Section 3.2.

3.1 The normalized case

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 [XZC16].

Lemma 3.1 (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 $j = 1, \ldots, k-1$.

Proof. For every edge $\{u, v\}$ in G, we denote the simple path replacing it by $(x_0, x_1, \ldots, x_k)$, identifying x_0 with u and x_1 with v. Let (r, y) be an eigenpair of $\mathbf{W}_k$ such that $\mathbf{W}_k y = ry$. Fix a subdivided edge $u = x_0, x_1, \ldots, x_k = v$ and write $a := y_{x_0}$ and $b := y_{x_k}$. At each internal vertex of degree 2, the eigen-equation is

$$(\mathbf{W}_k y)_{x_i} = \tfrac{1}{2}(y_{x_{i-1}} + y_{x_{i+1}}) = r\, y_{x_i}.$$

Set $r = \cos\theta$. Similarly to the eigenvalues of the cycle graph, the solution takes the form

$$y_{x_i} = \alpha \sin(i\theta) + \beta \cos(i\theta),$$

for some values of α, β. Imposing the two endpoint values $y_{x_0} = a$ and $y_{x_k} = b$ determines *uniquely* the whole sequence. By setting $i = 0$ one gets $\beta = a$, and by $i = k$ we have that $\alpha = \frac{b - a\cos(\theta k)}{\sin(\theta k)}$. This gives the overall formula

$$\boxed{y_{x_i} = \frac{\sin((k-i)\theta)}{\sin(k\theta)}\, a + \frac{\sin(i\theta)}{\sin(k\theta)}\, b} \qquad (i = 0, 1, \ldots, k),$$

and in particular

$$y_{x_1} = \frac{\sin((k-1)\theta)}{\sin(k\theta)}\, a + \frac{\sin\theta}{\sin(k\theta)}\, b.$$

We now look at the eigen-equation at an original vertex u (degree d). Denoting by $N(u)$ the neighbors of u in G,

$$r \cdot y_u = (\mathbf{W}_k y)_u = \frac{1}{d} \sum_{v \in N(u)} y_{x_1}^{(u \to v)} = \frac{\sin((k-1)\theta)}{\sin(k\theta)}\, a + \frac{\sin\theta}{\sin(k\theta)} \cdot \frac{1}{d} \sum_{v \in N(u)} y_v.$$

Rearranging,

$$\frac{1}{d}\sum_{v\in N(u)} y_v = \frac{r\sin(k\theta)-\sin((k-1)\theta)}{\sin\theta} y_u.$$

Notice that the LHS translates to the form of the original random walk matrix: if g is a vector such that $\mathbf{W}g = \omega g$ and we set $y_u = g_u$, we have that (setting again $r = \cos\theta$)

$$\omega = \frac{\cos\theta\sin(k\theta)-\sin((k-1)\theta)}{\sin\theta} = \cos(k\theta). \tag{12}$$

This concludes the first part: given any eigenpair (ω, g) of W, pick θ with $\cos(k\theta) = \omega$ and define y along each subdivided edge by the boxed formula; then $(\cos\theta, y)$ is an eigenpair of $\mathbf{W}_k$.

The above resulted in kn eigenvalues, while the graph G_k is of size

$$n + (k-1)|E| = \left(1 + \frac{1}{2}(k-1)d\right) n.$$

For the remainder of the eigenvalues r, set

$$y^e_{x_i} = \alpha_e \sin(i\theta), \quad y_u = y_v = 0,$$

which satisfies the internal-vertex equation with $r = \cos\theta = \cos(\pi j/k)$. At every original vertex u, the eigen-equation reads

$$(\mathbf{W}_k y)_u = \frac{1}{d}\sum_{e\ni u} y^{(e)}_{x_1} = \frac{\sin\theta}{d}\sum_{e\ni u}\alpha_e = r\, y_u = 0,$$

so it holds precisely when the edge-weights obey $\sum_{e\ni u}\alpha_e = 0$ for all original u. This results in n constraints for the $|E| = \frac{dn}{2}$ variables, providing (together with the $k-1$ choices for j) the remaining $\left(\frac{d}{2}-1\right)(k-1)n$ eigenvalues. □

Corollary 3.2. *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*

$$\omega_d(\varepsilon) \;=\; \cos\left(\frac{d\,\varepsilon}{2(d-2)}\,\arccos\left(\omega(G)\right)\right). \tag{13}$$

In particular:

1. *If $d = 3$ and G is Ramanujan, then*

$$\omega_3(\varepsilon) \;=\; \cos\left(\frac{3}{2}\arccos\left(\tfrac{\sqrt{8}}{3}\right)\varepsilon\right) \;\approx\; 1 - 0.12993\,\varepsilon^2.$$

2. *As $d \to \infty$ (with G Ramanujan), since* $\arccos\left(\frac{2\sqrt{d-1}}{d}\right) \to \frac{\pi}{2}$, *we obtain*

$$\omega_d(\varepsilon) \longrightarrow \cos\left(\tfrac{\pi}{4}\,\varepsilon\right) = 1 - \frac{\pi^2}{32}\,\varepsilon^2 + O(\varepsilon^4).$$

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 3.1, 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 2.3 yields Equation (13).

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 (13) gives the stated limit and its Taylor expansion. □

3.2 The unnormalized case

In this section we turn to the unnormalized (adjacency) spectrum. Unlike the normalized case, the resulting expression does not admit a simple closed form. Nevertheless, the formula we obtain is exact and informative, and it facilitates direct comparison with the other models.

Lemma 3.3 (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*

$$\lambda = \frac{\sin((k+1)\theta) - (d-1)\,\sin((k-1)\theta)}{\sin\theta}, \tag{14}$$

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

$$\lambda = \frac{\sinh((k+1)\xi) - (d-1)\sinh((k-1)\xi)}{\sinh(\xi)}, \tag{15}$$

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

Notice that Item 3 in the theorem statement provides the only potential eigenvalue larger than 2.

Proof. The proof of Items 1 and 2 follows the same lines as the proof of Lemma 3.1. Assuming $\mu > 2$ we denote $\mu = 2\cosh\xi$ and define the vector y by

$$y_{x_i} = \alpha \sinh(i\xi) + \beta \cosh(i\xi).$$

The equivalent of Equation (12) then becomes

$$\lambda = \frac{2\cosh(\xi)\sinh(k\xi) - d \cdot \sinh((k-1)\xi)}{\sinh(\xi)}, \tag{16}$$

which by trigonometric identities is equivalent to Equation (15). The proofs of the other items are the same as in Lemma 3.1. For the negative case, the proof follows by defining the vector using

$$y_{x_i} = (-1)^i \left(\alpha \sinh(i\xi) + \beta \cosh(i\xi)\right),$$

or by assuming k is even and using the fact that for bipartite graphs the spectrum is symmetric.

□

Sanity check at $k = 2$. As explained in the opening of Section 3.1, we will be interested in ultra-sparse expanders, and hence in large values of k. However, Lemma 3.3 holds for any k, including $k = 2$. This is the simplest subdivision application, where a single vertex is added in the middle of every edge, resulting in a $(d, 2)$-biregular graph, where the original vertices form one side of the graph and the new vertices form the other. Equation (15) simplifies in this case to

$$\lambda = \frac{\sinh(3\xi) - (d-1)\sinh(\xi)}{\sinh(\xi)} = 4\cosh^2(\xi) - d.$$

Recall from Lemma 3.3 that the resulting eigenvalue of G_k is $\mu = \pm 2\cosh\xi$, therefore the above equation simplifies even further to

$$\lambda = \mu^2 - d.$$

Assuming the original graph was Ramanujan, one can solve the above by assigning $\lambda = 2\sqrt{d-1}$ to get the positive eigenvalue

$$\mu = \sqrt{d + 2\sqrt{d-1}} = \sqrt{d-1} + 1,$$

which is exactly the spectrum of the universal cover of this graph: the biregular infinite tree of degrees d and 2 (recall that the spectrum of the infinite (a,b)-biregular tree is $\sqrt{a-1}+\sqrt{b-1}$).

An explicit (rather cumbersome) solution is also feasible for $k=3$, but this is not the case for general k as Equation (15) is not invertible even for a fixed $\lambda = 2\sqrt{d-1}$. Nonetheless, we are able to 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 3.4. *All eigenvalues μ of $\mathbf{A}_k$ with $\mu > 2$ satisfy*

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

Proof. Another way to write Equation (15) is

$$\lambda = \frac{e^{k\xi}\left(e^{\xi}-(d-1)e^{-\xi}\right)+e^{-k\xi}\left((d-1)e^{\xi}-e^{-\xi}\right)}{e^{\xi}-e^{-\xi}}. \tag{17}$$

Note that λ is an eigenvalue of the original graph, and therefore independent of k. Because the first term in the numerator grows like $e^{k\xi}$, a fixed left-hand side forces us to choose ξ so that the multiplicative factor

$$A(\xi) \triangleq e^{\xi}-(d-1)e^{-\xi}$$

is small. This happens near the unique zero of $A(\xi)$, namely $\xi = \frac{1}{2}\ln(d-1)$. Accordingly, set

$$\xi^* = \tfrac{1}{2}\ln(d-1)+x, \qquad x = x(k)\to 0 \text{ as } k\to\infty.$$

A short calculation gives

$$A(\xi^*) = \sqrt{d-1}\,e^{x} - \sqrt{d-1}\,e^{-x} \;=\; 2\sqrt{d-1}\,\sinh x \;=\; 2\sqrt{d-1}\,x\,(1+o(1)),$$

and

$$\begin{aligned} e^{\xi^*}-e^{-\xi^*} \;&=\; 2\sinh\left(\frac{1}{2}\ln(d-1)+x\right) \\ &=\; \frac{d-2}{\sqrt{d-1}} \;+\; \left(\sqrt{d-1}+\frac{1}{\sqrt{d-1}}\right)x\,(1+o(1)). \end{aligned}$$

Since $x = o(1)$, the denominator is

$$e^{\xi^*}-e^{-\xi^*} \;=\; \frac{d-2}{\sqrt{d-1}}\,(1+o(1)).$$

Moreover,

$$e^{k\xi^*} \;=\; (d-1)^{k/2}\, e^{kx} \;=\; (d-1)^{k/2}\,(1+o(1)),$$

because with the scaling we obtain below, $kx \to 0$.

Plugging these approximations into Equation (17) (with the term having $e^{-k\xi}$ ignored for large k) yields

$$\begin{aligned}\lambda \;&=\; \frac{(d-1)^{k/2}\,(1+o(1))\cdot 2\sqrt{d-1}\,x\,(1+o(1))}{(d-2)/\sqrt{d-1}\,(1+o(1))}\\ &=\; \frac{2(d-1)^{k/2+1}}{d-2}\,x\,(1+o(1)),\end{aligned}$$

so

$$x \;=\; \frac{\lambda}{2}\cdot\frac{d-2}{(d-1)^{k/2+1}}\,(1+o(1)).$$

Finally, the corresponding eigenvalue is $\mu = 2\cosh\xi^*$, and

$$\begin{aligned}\mu &= \sqrt{d-1}\,e^{x} + \frac{1}{\sqrt{d-1}}\,e^{-x}\\ &= \sqrt{d-1} + \frac{1}{\sqrt{d-1}} + x\left(\sqrt{d-1} - \frac{1}{\sqrt{d-1}}\right)(1+o(1)),\end{aligned}$$

whence

$$\begin{aligned}\mu &= \sqrt{d-1} + \frac{1}{\sqrt{d-1}} + (1+o(1))\,\frac{\lambda}{2}\cdot\frac{(d-2)}{(d-1)^{k/2+1}}\cdot\frac{d-2}{\sqrt{d-1}}\\ &= \sqrt{d-1} + \frac{1}{\sqrt{d-1}} + (1+o_k(1))\,\frac{\lambda}{2}\cdot\frac{(d-2)^2}{\left(\sqrt{d-1}\right)^{k+3}},\end{aligned}$$

as claimed. □

4 Interlacing Families and Free Convolutions

In this section we review definitions and known results in free probability and interlacing families. 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 and 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 [MSS18], where it appeared as the main tool enabling us to analyze the expected characteristic polynomial over a distribution, and deduce an existence result.

Lemma 4.1. *Suppose $\mathbf{A}_1, \ldots, \mathbf{A}_t$ are symmetric $m \times m$ matrices and let $(\mathbf{P}_i)_{i\in[t]}$ 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 $k \leq m$ there exist permutation matrices $(\mathbf{R}_i)_{i\in[t]}$ such that*

$$\lambda_k(\mathbf{A}_\mathbf{R}) \leq \alpha_k\left(\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x(\mathbf{A}_\mathbf{P})\right). \tag{18}$$

The multiplicative analog of Lemma 4.1 was proved in [CM23, Theorem 6.5 and Lemma 6.3]. A simplified version, suitable for our needs, is stated below.

Lemma 4.2. *Suppose $\mathbf{A}, \mathbf{B}$ are symmetric $m \times m$ matrices and $\mathbf{P}$ is a uniformly random $m \times m$ permutation matrix. Then, for every $k \leq m$ there exists a permutation matrix $\mathbf{S}$ such that*

$$\lambda_k\left(\mathbf{ASBS}^\mathsf{T}\right) \leq \alpha_k\left(\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{APBP}^\mathsf{T}\right)\right). \tag{19}$$

Denote the group of $m \times m$ orthogonal matrices by $\mathcal{O}(m)$. The *Haar distribution* is the unique distribution over $\mathcal{O}(m)$ which is invariant under multiplication (from the right or from the left) with any orthogonal matrix. We call a matrix drawn from this distribution a *Haar random matrix*.

Definition 4.3 (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*

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

and

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

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.

Although Definition 4.3 involves the matrices $\mathbf{A}$ and $\mathbf{B}$, it depends in fact only on $a(x)$ and $b(x)$, due to the properties of the Haar measure.[1] It is important to note that, as proven in [MSS22], both $(a \boxplus b)(x)$ and $(a \boxtimes b)(x)$ are real-rooted. Interestingly, there are

[1] It is easily seen that the convolution is well defined for any real-rooted polynomials $a(x)$ and $b(x)$ by choosing $\mathbf{A}, \mathbf{B}$ to be diagonal matrices with their respective roots on the diagonal.

explicit formulas for both $(a \boxplus b)(x)$ and $(a \boxtimes b)(x)$ as functions of the coefficients of $a(x)$ and $b(x)$. We refer the reader to [MSS22] for more details.

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 *permutations* matrices - which do have such interpretation - to Haar random matrices. To this end we state the following lemma (which appeared first in [MSS18] for the additive case, and later on in [CM23] 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 4.4 (*Quadrature*; Corollary 4.9 from [MSS18] and Lemma 2.3 from [CM23])**.** *Let $\mathbf{A}, \mathbf{B}$ be real $m \times m$ symmetric matrices such that $\mathbf{A1} = a\mathbf{1}$ and $\mathbf{B1} = b\mathbf{1}$. 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,*

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{A} + \mathbf{P}\mathbf{B}\mathbf{P}^{\mathsf{T}}\right) = (x - (a+b))\left(p_{\mathbf{A}} \boxplus p_{\mathbf{B}}\right)(x), \tag{20}$$

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{A}\mathbf{P}\mathbf{B}\mathbf{P}^{\mathsf{T}}\right) = (x - ab)\left(p_{\mathbf{A}} \boxtimes p_{\mathbf{B}}\right)(x). \tag{21}$$

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 [NS06, Remark 2.19]):

$$\mathcal{G}_\mu(x) = \sum_{r=0}^{\infty} \frac{m_r(\mu)}{x^{r+1}}. \tag{22}$$

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

$$\mathcal{M}_\mu(x) = x\mathcal{G}_\mu(x) - 1. \tag{23}$$

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 $\mathcal{G}_\mu$, and, in a similar manner, we define $\mathcal{N}_\mu(y)$ as the inverse of $\mathcal{M}_\mu$. 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 4.5. *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 4.4 if applicable, and similar notations for the other transforms.

Example 4.6 (Perfect matching)**.** *Let $\mathbf{M}$ be the adjacency matrix of a perfect matching of dimension $2m$. Then we have its transforms (defined for $x > 1$ and $y > 0$):*

$$\mathcal{G}_\mathbf{M}(x) = \frac{1}{2m}\left(\frac{m}{x+1} + \frac{m}{x-1}\right) = \frac{x}{x^2-1}, \tag{24}$$

$$\mathcal{K}_\mathbf{M}(y) = \frac{1+\sqrt{4y^2+1}}{2y}. \tag{25}$$

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 [MSS22, Theorem 1.11] and [CCM24, Claim 4.12], and stated below.

Lemma 4.7 (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 4.7 and Claim 4.5 is the following.

Corollary 4.8. *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 4.9. *[CCM24, Claim 4.12] Let $\mathbf{B}$ be a symmetric matrix such that $\mathbf{B1} = b\mathbf{1}$ 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.

4.2 The Kesten-McKay distribution

The probability measure of the Kesten-McKay distribution with parameter d is given by

$$\mu_{\mathsf{km}}(t) = \begin{cases} \dfrac{d\sqrt{4(d-1)-t^2}}{2\pi(d^2-t^2)}, & \text{for } |t| \leq 2\sqrt{d-1}; \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

Note that we suppress the parameter d from the notation when clear from context. A special case of Kesten-McKay is $d = 2$ where it is referred to as the *arcsin distribution*:

$$\mu_{\mathbf{arc}}(t) = \frac{1}{\pi}\frac{1}{\sqrt{4-t^2}}, \tag{27}$$

supported on $(-2, 2)$. The following is a well-known fact from free probability theory (see, e.g., [NS06, Chapter 12]).

Claim 4.10. *The Cauchy and $\mathcal{K}$-transforms of the Kesten-McKay distribution with parameter d are given by*

$$\mathcal{G}_{\mu_{\mathsf{km}}}(x) = \frac{d\sqrt{x^2-4(d-1)} - x(d-2)}{2(x^2-d^2)}.$$

$$\mathcal{K}_{\mu_{\mathsf{km}}}(y) = \frac{2-d+d\sqrt{1+4y^2}}{2y}.$$

In particular

$$\mathcal{G}_{\mu_{\mathbf{arc}}}(x) = \frac{1}{x^2-4}, \qquad \mathcal{K}_{\mu_{\mathbf{arc}}}(y) = \frac{\sqrt{4y^2+1}}{y}.$$

For ease of readability we denote the Cauchy transform of the Kesten-McKay distribution by $\mathcal{G}_{\mathsf{km}}$. Lastly, the following claim, relating the moments of a d-regular graph with the moments of the Kesten-McKay distribution with parameter d is implicit in [McK81].

Claim 4.11. *Let G be a d-regular graph with girth g. Then, for every $0 \leq r < g$, $m_r(G) = m_r(\mu_{\mathsf{km}})$.*

5 Ultra-Sparse Expanders in the Configuration Model

In this section we prove Theorem 2.1, which immediately implies Corollary 2.2. We begin by recalling the configuration model in Section 5.1. Then, before turning to our main argument in Section 5.3, we give an alternative proof of the celebrated MSS theorem using the configuration model—rather than the standard union-of-d-matchings approach. This appears in Section 5.2 and serves as a warm-up for the techniques used in our main proof.

The next paragraphs, up to Section 5.2, are intended as an intuitive starting point and are not required for following the formal proofs; they also highlight the connection to free probability and its potential role here.

Assume we have a matrix $\mathbf{A}$, with eigenvalue distribution $\mu_{\mathbf{A}}$. Assume also that we have a matrix $\mathbf{P}$ which is a random projection to an α-fraction of the space. What should we expect the eigenvalue distribution of $\mathbf{PAP}$ to be? Assuming no connections between the randomness of $\mathbf{P}$ and the eigenvectors of $\mathbf{A}$, it is not surprising that free probability has an answer to this question. However, the nature of this answer is remarkable, and is encompassed in *Free Projection* and convolution semigroups, concepts in free probability introduced in [NS96]. Explained in free probabilistic terms, we have two elements a and p, where p has expectation α and satisfies $p^2 = p$ (which defines a projection), and a has probability distribution μ. The compression of a by p is the element pap. If a and p are free, we call the resulting distribution the *free α-projection of μ* and denote it by μ_α. The following surprising result from [NS96] is that, up to scaling, the free α-projection is equivalent to the summation of $1/\alpha$ free elements with the same distribution.

Theorem 5.1. *Let $a_1, \ldots, a_d$ be free identically distributed random variables, each with distribution μ. Then, their sum has the same distribution (up to scaling) as the free $1/d$-projection of μ. That is,*

$$\mu^{\boxplus d}(t) = \mu_{1/d}(dt). \tag{28}$$

The rich structure of free probability theory gives us this surprising characteristic, which has no equivalent in classical probability–that is, there is no similar statement for sums of independent random variables. Notice that although $\mu^{\boxplus d}$ was defined for an integer d only, the projection is not restricted to simple fractions of the form $1/d$. What Theorem 5.1 inspired in [NS06, Chapter 14] is the idea of defining $\mu^{\boxplus t}$ for any *real* $t \geq 1$, being consistent with the original convolution definition.

When μ is the ± 1 distribution associated with a perfect matching (as in Example 4.6), $\mu^{\boxplus d}$ is the analogue of summing d free matchings, which underlies the original MSS proof. In turn, Theorem 5.1 yields an alternative existence proof of Ramanujan graphs via the

configuration model, leveraging its matrix structure. The observation above extends to any $\alpha \in (0,1)$—in particular, to $\alpha = 1/\bar{d}$, where $\bar{d}$ is the graph's average degree.

5.1 Rotation maps and the configuration model

Let $G = (V, E)$ be a d-regular undirected graph on n vertices, and assume that each vertex v assigns distinct labels from $\{1, \ldots, d\}$ to its d incident edges, so that every label appears exactly once. Note that each edge receives two labels—one from each of its endpoints—and these labels need not coincide. Let $v[i]$ denote the i-th neighbor of v according to this labeling.

Definition 5.2 (Edge rotation map)**.** *Let G be a graph labeled as above. The* edge rotation map *(or simply* rotation map*) of G, denoted $\mathrm{Rot}_G : V \times [d] \to V \times [d]$, is defined by*

$$\mathrm{Rot}_G(v, i) = (u, j) \iff v[i] = u \ \wedge \ u[j] = v.$$

In other words, $\mathrm{Rot}_G(v, i) = (u, j)$ if the i-th neighbor of v is u, and the j-th neighbor of u is v. Observe that Rot_G is an involution. We note that, unlike in some other random graph models, self-loops are allowed: it is possible that $\mathrm{Rot}_G(v, i) = (v, j)$ for some vertex v and indices $i, j \in [d]$, in which case v is its own i-th and j-th neighbor. We also define the associated *rotation matrix* $\dot{\mathbf{G}}$, which is an $N \times N$ Boolean matrix with $N = nd$, where $\dot{\mathbf{G}}_{(v,i),(u,j)} = 1$ if and only if $\mathrm{Rot}_G(v, i) = (u, j)$.

We now define the *configuration model* for sampling a random d-regular graph on n vertices, where $n, d \in \mathbb{N}$ and at least one of them is even. One way to visualize the process is to imagine that each vertex is initially connected to d "half-edges", and the graph is formed by pairing up the N half-edges at random. This definition will later be extended to the non-regular setting.

Definition 5.3 (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*.*

The model can also be described more formally in matrix form. To describe this, let $\mathbf{U}$ be the $N \times n$ *up* matrix, defined by its action on vectors $x \in \mathbb{R}^n$ as $\mathbf{U}x = x \otimes \mathbf{1}_d$, where $\otimes$ denotes the tensor product and $\mathbf{1}_d$ is the all-ones vector of dimension d. Let $\mathbf{M}$ denote the adjacency matrix of an arbitrary perfect matching on N vertices (recall that $N = nd$). Then, sampling a graph G from the configuration model is equivalent to choosing a random $N \times N$ permutation matrix $\mathbf{P}$ and setting

$$\mathbf{A_P} = \mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\mathbf{U}. \tag{29}$$

Note that the middle operator $\mathbf{PMP}^\mathsf{T}$ is precisely $\dot{\mathbf{G}}$, which encodes the rotation map Rot_G. We denote the d-regular n-vertex graph whose adjacency matrix is $\mathbf{A}_\mathbf{P}$ by $G_\mathbf{P}$.

5.2 Alternate MSS proof via finite free projection

In this section, we present our alternative proof of the MSS result using the configuration model. We make use of the notation introduced in Section 5.1 throughout.

Theorem 5.4. *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*

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

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 $\frac{1}{d}$. 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

$$\begin{aligned}\chi_x\left(\mathbf{M}\right) &= (x-1)^{\frac{N}{2}}(x+1)^{\frac{N}{2}},\\ \chi_x\left(\mathbf{V}\right) &= x^{n(d-1)}(x-d)^n.\end{aligned}$$

Thus, a direct calculation yields

$$\mathcal{N}_\mathbf{M}(y) = \sqrt{\frac{y+1}{y}}, \quad \mathcal{N}_\mathbf{V}(y) = \frac{dy+1}{y}, \tag{30}$$

where $\mathcal{N}$-transform is as defined in Section 4.1.

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 4.2, it is enough to prove that

$$\alpha_2 \triangleq \alpha_2\left(\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\right)\right) < 2\sqrt{d-1}, \tag{31}$$

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

root. By Lemma 4.4,

$$\mathop{\mathbf{E}}_{\mathbf{P}} \chi_x\left(\mathbf{U}\mathbf{U}^{\mathsf{T}}\mathbf{P}\mathbf{M}\mathbf{P}^{\mathsf{T}}\right) = (x-d)\left(p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}\right)(x), \tag{32}$$

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

$$\mathcal{N}_{p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}}(y) \leq \frac{y}{y+1}\mathcal{N}_{p_{\mathbf{M}}}(y)\mathcal{N}_{p_{\mathbf{V}}}(y) < \frac{dy+1}{\sqrt{(y(y+1))}}, \tag{33}$$

where the first inequality follows by Lemma 4.7, and the second, sharp, inequality follows by Claim 4.9 and the calculations of $\mathcal{N}_{\mathbf{M}}(y), \mathcal{N}_{\mathbf{V}}(y)$ above.

By Claim 4.5, $\alpha_2 \leq \mathcal{N}_{p_{\mathbf{V}} \boxtimes p_{\mathbf{M}}}(y)$ for any $y > 0$. Choosing the point $y_0 = \frac{1}{d-2}$, we get by the above inequality that

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

□

5.3 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 of Equation (29) 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

$$\mathbf{W_P} = \mathbf{C}^\mathsf{T}\mathbf{P}\mathbf{M}\mathbf{P}^\mathsf{T}\mathbf{C}, \tag{34}$$

where $\mathbf{M}$ is the adjacency matrix of a perfect matching on the N half-edges, as in the regular configuration model. This is best illustrated by an example.

Example. For a 3-vertex graph with degrees $2, 3, 3$ we have

$$\mathbf{C} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}. \tag{35}$$

With this, we are ready to prove Theorem 2.1. The proof follows similar lines to that of Theorem 5.4, and so we refrain from repeating it in full, instead focusing on the differences.

Proof of Theorem 2.1. 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 5.4, we arrive at the analogue of Equation (33),

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

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

5.4 Structure vs. randomness in non-regular graphs

The bound of Theorem 2.1, expressed in Equation (1), has a strong syntactic resemblance to the Ramanujan bound for the non-regular setting. As this is known to be optimal in the regular case, an immediate question to be asked is whether Theorem 2.1 is the optimal possible expansion for a particular sequence of degrees. We answer this question with a simple counterexample.

Consider a graph on $m+1$ vertices, where m vertices have degree 3 and a single vertex is of degree m. An example of such a graph is the *Wheel* graph $\mathcal{W}_m$, consisting of a cycle graph of size m and an additional vertex v_0 connected to all the vertices of the cycle denoted $v_1, \ldots, v_m$. See Figure 2a for a visualisation of the graph.

It is intuitive that the spectrum of the random walk matrix of $\mathcal{W}_m$ should approach $\frac{2}{3}$, as within two steps one reaches a uniformly random vertex with probability $\frac{1}{3}$. This is formalized in the following claim.

Claim 5.5. *The nontrivial eigenvalues of the normalized adjacency matrix of $\mathcal{W}_m$ are*

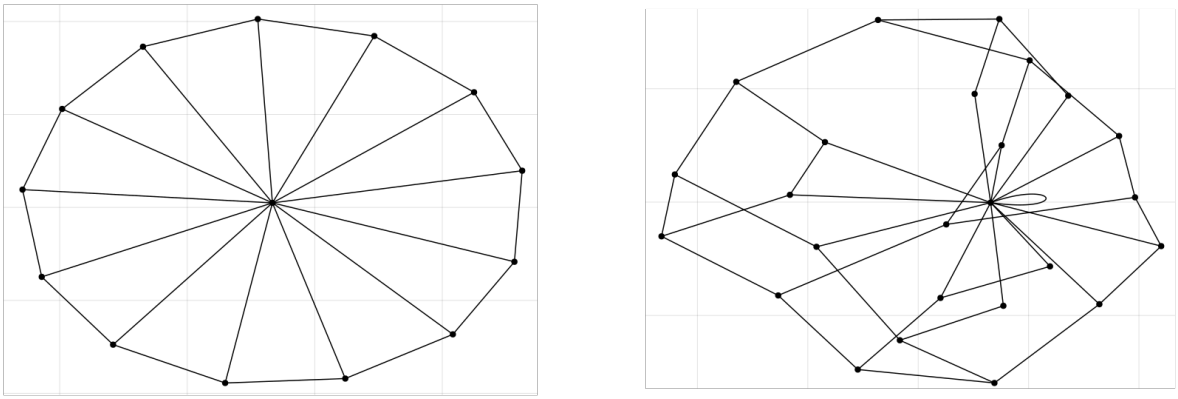
$$\omega_j(\mathcal{W}_m) = \frac{2}{3}\cos\left(\frac{2\pi j}{m}\right),$$

for $j = 1, \ldots, m$.

Proof sketch. We prove by constructing the eigenvectors $(x_j)_{j=1}^{m/2}$ and $(y_j)_{j=1}^{m/2}$ corresponding to the eigenvalues. Let $x_j(v_0) = y_j(v_0) = 0$ and $x_j(v_\ell) = \cos\left(\frac{2\pi j\ell}{m}\right)$, $y_j(v_\ell) = \sin\left(\frac{2\pi j\ell}{m}\right)$. It is easy to verify that these are indeed m distinct eigenvectors of $\mathcal{W}_m$ with the desired eigenvalues. □

An immediate corollary of the above is that $\omega(\mathcal{W}_m) \leq \frac{2}{3}$, and it approaches $\frac{2}{3}$ as m grows. However, $\bar{d} = \frac{m+3m}{m+1} \approx 4$, and Theorem 2.1 gives us the promise of a graph G with $\omega(G) \leq \frac{\sqrt{3}}{2} \approx 0.866$, which is significantly worse, while Chung's bound, Equation (3), only tells us that $\omega(G) \geq \frac{2}{\sqrt{m}}$.

Observing random graphs picked in the configuration model shows that Equation (1) *does* give a reliable description of the spectral expansion of these graphs. In particular, as m grows, random graphs in the configuration model approach $\frac{\sqrt{3}}{2}$ with high probability (even though graphs looking like $\mathcal{W}_m$ have nonzero probability to be drawn). While in the regular case, we expect almost all graphs to be close to optimal by Friedman's theorem [Fri08, BC19], we have that in the non-regular case the expected behavior is different than the optimal, and adding structure may improve expansion properties. This idea, together with Open Problem 2.5, leaves more questions than answers in this setting and is left for future research.



(a) The graph $\mathcal{W}_{13}$.

(b) Random configuration graph with the degree sequence of $\mathcal{W}_{24}$.

Figure 2: Visualisation of $\mathcal{W}_m$ and random configuration graph. Note that multiple edges are allowed in the configuration model and are omitted from the drawing.

6 The Free Method

The analysis in Section 5.3 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 [AZ24] for an up to date survey of the latter).

However, the analysis of Section 5.3 does not translate to the adjacency case: when using the "up matrix" $\mathbf{C}$ as in Equation (35) without normalization–that is, when the nonzero entries of the up matrix are 1–the same proof technique does not work, as $\mathbf{1}$ is not an eigenvector of all the matrices involved (and so Lemma 4.4 could not be applied).

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 [MSS18] (and later used in [MO20, CM23] and others), and exemplified here in Section 5.3. Others are the strong convergence approaches of Bordenave and Collins [BC19] and Chen et al. [CGVTvH]. 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 2.1), or a high probability result as in [BC19, CGVTvH], 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 [NS06]).

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]:

Procedure 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).

The flow of the above recipe in the basic case that $\mu_a = \mu_b = \mu$ is illustrated in Figure 3. Item 3 is the most famous theorem in free probability theory, commonly referred to as *Voiculescu's Theorem* and stated in terms of the $\mathcal{R}$-transform ([NS06, Theorem 12.7]).

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 $c = ab$ –

[2]The proof is standard in the free probability literature, and can be found in [NS06, Chapter 12].

where the $\mathcal{G}$ and $\mathcal{K}$ transforms are replaced with $\mathcal{M}, \mathcal{N}$ respectively, and Item 3 is replaced with $\mathcal{N}_{\mu_c}(y) \leq \frac{y}{y+1} \cdot \mathcal{N}_{\mu_a}(y) \cdot \mathcal{N}_{\mu_b}(y)$.

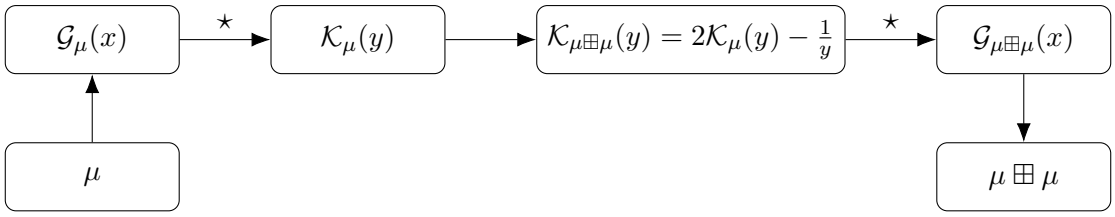


Figure 3: Procedure 1 in the basic case of a sum of free random variables, given that the distribution of each of them is μ. Transitions marked with $\star$ involve inverting a function.

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 may not be available. We handle a particular case of this issue in Section 7, where we are able to begin the analysis directly from Item 2. For the rest of the discussion here, though, we assume that the distributions are simple enough and are given in closed form.

The second caveat 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:

Procedure 2. Bounding the free additive convolution

- Perform Items 1 to 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 [BES20] or develop numerical techniques for this task (see [CY23, ON12] and the subordination technique presented in [Spe19]). 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 6.1. *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:

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

---

1. Find the solution x_0 of the equation

$$\frac{t-1}{t}\mathcal{G}'_\mu(x) = -\mathcal{G}_\mu(x)^2. \tag{36}$$

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 6.1 do we need to invert a function. Its usage is illustrated in Figure 4. As was shown in Section 2.3.1, this procedure can be applied to easily deduce the Ramanujan bound of $2\sqrt{d-1}$.

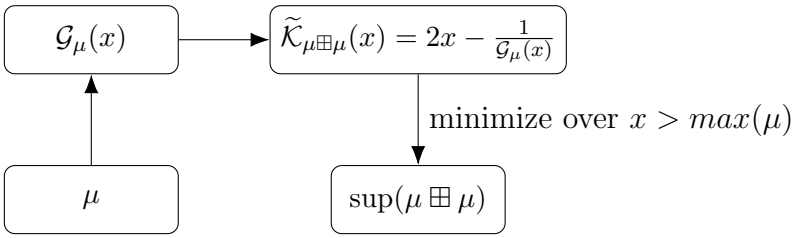


Figure 4: Examplified use of the bounding technique of Lemma 6.1, in the case $t = 2$. Unlike the recipe depicted in Figure 3, no inversions are involved.

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

$$\mathcal{K}'_\mu(y) = \frac{1}{\mathcal{G}'_\mu(\mathcal{K}_\mu(y))}. \tag{37}$$

As shown, we look for the minimal value of

$$\mathcal{K}_{\mu_t}(y) = t \cdot \mathcal{K}_\mu(y) - \frac{t-1}{y} \tag{38}$$

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 (38)

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

Plugging in Equation (37) 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 (36). The minimal value of Equation (38) 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 6.1 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 6.1 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 6.2. *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*

$$\widetilde{\mathcal{K}}_{\mu_1 \boxplus \mu_2}(x) \triangleq \mathcal{K}_{\mu_1}(\mathcal{G}_{\mu_2}(x)) + x - \frac{1}{\mathcal{G}_{\mu_2}(x)}. \tag{39}$$

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

$$\widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x) \triangleq \frac{\mathcal{M}_{\mu_2}(x)}{\mathcal{M}_{\mu_2}(x)+1} \cdot \mathcal{N}_{\mu_1}(\mathcal{M}_{\mu_2}(x)) \cdot x. \tag{40}$$

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.

6.2.1 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, we follow the analysis of Section 5.3, however replacing the *normalized* matrix $\mathbf{C}$ in Equation (34) with the simpler "up matrix" $\mathbf{U}$ with every nonzero entry being 1 (similar to Equation (29) in the regular case). The convolution in this case 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 (30) we have that $\mathcal{N}_{\mu_1}(y) = \sqrt{\frac{y+1}{y}}$. Thus, applying Lemma 6.2 gives us that $\sup(\mu_1 \boxtimes \mu_2)$ is found by minimizing over $x > 3$ the function

$$\widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x) = x \cdot \sqrt{\frac{\mathcal{M}_{\mu_2}(x)}{\mathcal{M}_{\mu_2}(x)+1}} = x \cdot \sqrt{\frac{x\mathcal{G}_{\mu_2}(x)-1}{x\mathcal{G}_{\mu_2}(x)}}. \tag{41}$$

This leads us to the main result of this section.

Lemma 6.3 (Small–ε bound for $\mathsf{pred}_{\mathsf{conf}}(\varepsilon)$)**.**

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

Before proving the small ε approximation, we first handle the minimization of Equation (41) in the main technical lemma below, and then deduce the small–ε bound using it.

Lemma 6.4. *Fix $\varepsilon \in (0,1)$ and let $\widetilde{\mathcal{N}}(x) \triangleq \widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x)$ be as defined in Equation (41). The function $\widetilde{\mathcal{N}}(x)$ has a minimum at x^*, which is the unique root in $(3,\infty)$ of*

$$x^3 - 8x^2 + (18+c)\,x + 2c^2 - 12c, \tag{42}$$

where $c = \dfrac{6}{2+\varepsilon}$, and the minimum $P_{\min}(\varepsilon) \triangleq \min_{x>3} \widetilde{\mathcal{N}}(x)$ is

$$P_{\min}(\varepsilon) = \sqrt{\frac{(x^*)^2(x^*-c)}{(x^*)^2 - 4x^* + 6 - c}}.$$

See Figure 5 for an illustration of this function's behavior.

Proof. We begin with an algebraic simplification. With c as defined, we get

$$\mathcal{M}_{\mu_2}(x) = \frac{(2+\varepsilon)x - 6}{(2+\varepsilon)(x-2)(x-3)} = \frac{x-c}{(x-2)(x-3)}.$$

Hence, for $x > 3$,

$$f(x) \triangleq \widetilde{\mathcal{N}}(x)^2 = \frac{x^2(x-c)}{x^2 - 4x + 6 - c}.$$

In order to write $f(x) = f_N(x)/f_D(x)$ as a rational function, we write

$$f_N(x) = x^2(x-c) \quad \text{and} \quad f_D(x) = x^2 - 4x + 6 - c. \tag{43}$$

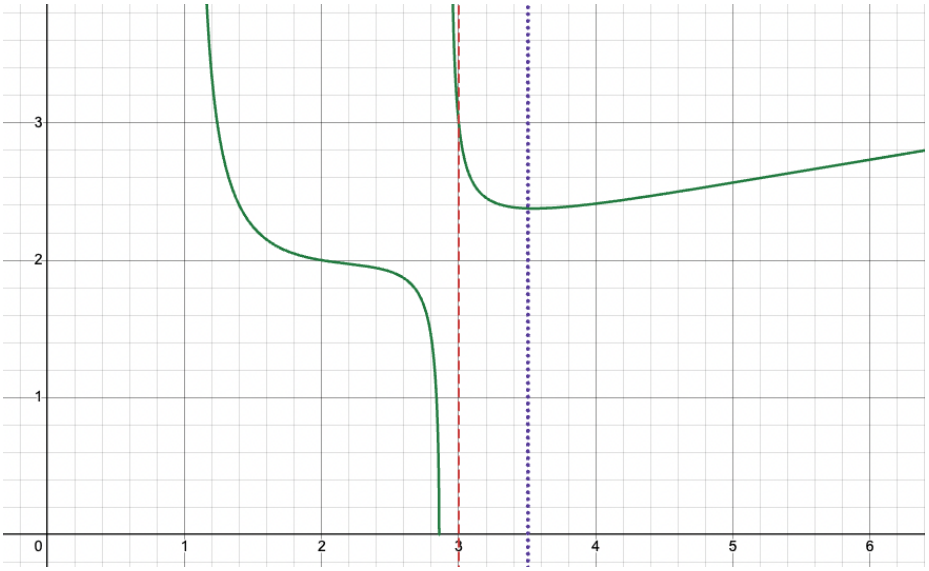


Figure 5: The function $\widetilde{\mathcal{N}}_{\mu_1 \boxtimes \mu_2}(x)$ of Equation (41) drawn in solid green for $\varepsilon = 0.1$. The dotted purple vertical line is the minimizer x^*.

Differentiating we get

$$f'(x) = \frac{(3x^2 - 2cx) f_D(x) - f_N(x)(2x - 4)}{f_D(x)^2}.$$

Setting $f'(x) = 0$ yields

$$x^3 - 8x^2 + (18 + c)\, x + 2c^2 - 12c = 0. \tag{44}$$

Evaluating f at x^* gives the minimum

$$P_{\min}(\varepsilon) = \sqrt{\frac{(x^*)^2 (x^* - c)}{(x^*)^2 - 4x^* + 6 - c}}$$

completing the proof. □

Sanity check at $\varepsilon = 1$. In this case $c = \frac{6}{2+1} = 2$, and Equation (44) factors as

$$x^3 - 8x^2 + 20x - 16 = (x - 4)(x - 2)^2,$$

with the only root at $x > 3$ being $x^* = 4$. Consequently,

$$P_{\min}(1) = \sqrt{\frac{4^2(4 - 2)}{4^2 - 16 + 6 - 2}} = 2\sqrt{2},$$

which coincides with the spectral radius of the adjacency operator on the infinite 3-regular tree, as expected.

Explicit formula for $\mathsf{pred}_{\mathsf{conf}}(\varepsilon)$. Set $t = x - \frac{8}{3}$ to bring Equation (42) into the form

$$t^3 + p\,t + q = 0,$$

where $p = c - \frac{10}{3}$ and $q = 2c^2 - \frac{28}{3}\,c + \frac{272}{27}$. By Cardano's formula, we have the solution

$$t^* = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}} \;+\; \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}},$$

where $\Delta = (q/2)^2 + (p/3)^3$. Evaluating $\widetilde{\mathcal{N}}(x)$ at $x^* = t^* + \frac{8}{3}$ gives

$$\widetilde{\mathcal{N}}(x^*) = \sqrt{\frac{(x^*)^2(x^* - c)}{(x^*)^2 - 4x^* + 6 - c}} = \sqrt{\frac{9t^{*2} + 3(16 - 2c)t^* + 16(4 - c)}{6t^* + 4}}.$$

Although we have a closed-form expression, its complexity prompts us to take an alternative route to understand its value for small ε. We use Lemma 6.4 as stated, and prove Lemma 6.3 using the approximation for small ε earlier in the process. This method will find itself useful in proofs for the other models as well.

Proof of Lemma 6.3. Let x^* be the unique minimizer of $\widetilde{\mathcal{N}}(x)$ on $(3, \infty)$ and recall $P_{\min}(\varepsilon) = \widetilde{\mathcal{N}}(x^*)$.

Write $c = \dfrac{6}{2 + \varepsilon} = 3 - \delta$ with $\delta = \frac{3}{2}\varepsilon + O(\varepsilon^2)$, and set $x = 3 + u$. In this notation Equation (44) becomes

$$u^3 + u^2 - \delta\,u - 3\delta + 2\delta^2 = 0. \tag{45}$$

By comparing the leading terms, it holds that $u = \Theta(\delta^{1/2})$. Let $u = a\,\delta^{1/2} + b\,\delta + O(\delta^{3/2})$ and substitute into Equation (45). Comparing the coefficients of δ and $\delta^{3/2}$ yields

$$a^2 = 3, \quad a^3 + 2ab - a = 0 \quad \Longrightarrow \quad a = \sqrt{3}, \quad b = -1.$$

Hence

$$x^* = 3 + u = 3 + \frac{3}{\sqrt{2}}\varepsilon^{1/2} - \frac{3}{2}\varepsilon + O(\varepsilon^{3/2}). \tag{46}$$

Next, compute $P_{\min}(\varepsilon)^2$ from its definition

$$P_{\min}(\varepsilon)^2 = f(x^*) = \frac{x^{*2}(x^* - c)}{x^{*2} - 4x^* + 6 - c}.$$

With $f_N(x), f_D(x)$ defined as in Equation (43), at the point x^* we have by substituting

$s = \delta^{1/2}$:

$$f_N(x^*) = 9\sqrt{3}s + 18s^2 + \frac{27}{2\sqrt{3}}s^3 + O(s^4),$$
$$f_D(x^*) = 2\sqrt{3}s + 2s^2 - \frac{1}{\sqrt{3}}s^3 + O(s^4).$$

We will use the following identity for estimating a rational function as a power series. Let

$$A(s) = a_0 + a_1 s + a_2 s^2 + O(s^3), \qquad B(s) = b_0 + b_1 s + b_2 s^2 + O(s^3),$$

with $b_0 \neq 0$. Then

$$\frac{A(s)}{B(s)} = \frac{a_0}{b_0} + \frac{a_1 b_0 - a_0 b_1}{b_0^2}\, s + \frac{a_2 b_0^2 - a_1 b_0 b_1 + a_0(b_1^2 - b_0 b_2)}{b_0^3}\, s^2 + O(s^3).$$

Used with $A(s) = \frac{f_N(x^*)}{s}$ and $B(s) = \frac{f_D(x^*)}{s}$ we have

$$a_0 = 9\sqrt{3}, \quad a_1 = 18, \quad a_2 = \frac{27}{2\sqrt{3}}, \quad b_0 = 2\sqrt{3}, \quad b_1 = 2, \quad b_2 = -\frac{1}{\sqrt{3}}$$

resulting in

$$\frac{f_N(x^*)}{f_D(x^*)} = \frac{9}{2} + \frac{3\sqrt{3}}{2}\, s + \frac{3}{2}\, s^2 + O(s^3).$$

Substitute back we get $s = \delta^{1/2} = \sqrt{\frac{3}{2}}\, \varepsilon^{1/2} + O(\varepsilon^{3/2})$ and $s^2 = \delta = \frac{3}{2}\, \varepsilon + O(\varepsilon^2)$, to obtain

$$P_{\min}(\varepsilon)^2 = \frac{9}{2} + \frac{9\sqrt{2}}{4}\, \varepsilon^{1/2} + \frac{9}{4}\, \varepsilon + O(\varepsilon^{3/2}).$$

Taking the square root by binomial expansion yields

$$P_{\min}(\varepsilon) = \frac{3}{\sqrt{2}} + \frac{3}{4}\, \varepsilon^{1/2} + \frac{9\sqrt{2}}{32}\, \varepsilon + O(\varepsilon^{3/2}),$$

as claimed. □

6.2.2 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

In order to apply the method on other models, we shall first recall their definitions to come up with the correct distributions and convolutions required for the analysis. This in turn will achieve a full analytic solution for the $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ model, and a tight approximation for small ε for $\mathcal{C} + \varepsilon\mathcal{M}$ and $(2+\varepsilon)\mathcal{M}$.

Given a parameter $\rho \in (0, 1)$, we define the ρ-matching as a matching of a ρ fraction of the vertices. We assume for simplicity that ρn is an integer, however the n parameter plays no part in the (infinite) free analysis. The distribution μ_ρ of its eigenvalues gives 0 with

probability $1-\rho$, and otherwise ± 1 with equal probabilities. By definition, the Cauchy transform of its adjacency matrix is

$$\mathcal{G}_\rho(x) = (1-\rho)\cdot\frac{1}{x} + \frac{\rho}{2}\cdot\frac{1}{x-1} + \frac{\rho}{2}\cdot\frac{1}{x+1}. \tag{47}$$

A special case of Equation (47) is the case of $\rho = 1$, which is the perfect matching Cauchy transform from Equation (24). This enables the proof, using the techniques of Section 6.1, of the two following predictions:

Lemma 6.5. *For the* $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ *model we have the benchmark*

$$\begin{aligned}\mathsf{pred}_{3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)}(\varepsilon) &= \sqrt{1+\sqrt{\frac{4-\varepsilon^2}{3}}}\cdot\frac{\sqrt{3(4-\varepsilon^2)}-3\varepsilon}{2(1-\varepsilon)} \\ &= A + B\varepsilon + O(\varepsilon^2),\end{aligned} \tag{48}$$

where $A = \sqrt{3+\sqrt{12}} \approx 2.54$ *and* $B = \frac{\sqrt{3+\sqrt{12}}}{4+\sqrt{12}} \approx 0.34$.

Lemma 6.6. *For small* ε*, we can bound the predictions for both the* $\mathcal{C}+\varepsilon\mathcal{M}$ *and* $(2+\varepsilon)\mathcal{M}$ *models by*

$$\mathsf{pred}_{\mathcal{C}+\varepsilon\mathcal{M}}(\varepsilon) = \mathsf{pred}_{(2+\varepsilon)\mathcal{M}}(\varepsilon) < \sqrt{5} + \sqrt{\frac{4}{5\phi}}\sqrt{\varepsilon}, \tag{49}$$

where $\phi = \frac{1+\sqrt{5}}{2}$ *is the golden ratio.*

While $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ may seem less natural combinatorially than other models, its analytical treatment turns out to be the most straightforward due to the ability to use Lemma 6.1. The derivation of an explicit benchmark for $\mathcal{C}+\varepsilon\mathcal{M}$ and $(2+\varepsilon)\mathcal{M}$ turns out to be infeasible; however, techniques similar to those used in Section 6.2.1 enable us to deduce a good approximation for it for small ε.

Proof of Lemma 6.5. In the $3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)$ model, we wish to apply Lemma 6.1 using the Cauchy-transform $\mathcal{G}_\beta(x)$, as defined in Equation (47), for $\beta = \frac{2+\varepsilon}{3}$. Equation (36) of Procedure 3 then takes the form

$$2\mathcal{G}_\beta(x)' + 3\mathcal{G}_\beta(x)^2 = 0,$$

whose positive solution can be computed to be

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

Plugging this into the rest of Procedure 3, we have that $y_0 = \frac{\beta+\sqrt{\beta(4-3\beta)}}{x_0\sqrt{\beta(4-3\beta)}}$ and overall

$$\begin{aligned}\mathsf{pred}_{3\left(\frac{2+\varepsilon}{3}\mathcal{M}\right)}(\varepsilon) &= 3\cdot x_0 - \frac{2}{y_0} \\ &= \sqrt{1+\sqrt{\beta(4-3\beta)}}\frac{3\beta+\sqrt{\beta(4-3\beta)}}{\beta+\sqrt{\beta(4-3\beta)}} \\ &= \sqrt{1+\sqrt{\frac{4-\varepsilon^2}{3}}}\cdot\frac{\sqrt{3(4-\varepsilon^2)}-3\varepsilon}{2(1-\varepsilon)}.\end{aligned}$$

□

Remark 6.7. As μ_ρ has support size 3, it *is* possible to invert $\mathcal{G}_\rho(x)$, and a solution using Procedure 2 may technically be possible. However, this turns out to be an extremely cumbersome (and not very informative) process, as the inverse of Equation (47) can be shown to be (after solving the cubic equation $\mathcal{G}_\rho(\mathcal{K}_\rho(y)) = y$):

$$\mathcal{K}_\rho(y) = 2\cdot\sqrt{\frac{1}{3}+\frac{1}{9y^2}}\cdot\cos\left(\frac{1}{3}\cdot\arccos\left(\left(\frac{3\left(-\frac{2}{27y^3}+\frac{2-3\rho}{3y}\right)}{2\left(-1-\frac{1}{3y^2}\right)}\right)\cdot\sqrt{\frac{3}{1+\frac{1}{3y^2}}}\right)\right)+\frac{1}{3y}, \quad (50)$$

and a solution would require finding the minimum value of $3\mathcal{K}_\beta(y) - \frac{2}{y}$. This shows that Procedure 3 has an advantage even in cases of invertible Cauchy transforms.

As discussed in Section 1.2, the $\mathcal{C}+\varepsilon\mathcal{M}$, and $(2+\varepsilon)\mathcal{M}$ models are different, in particular in the sense that $\mathcal{C}+\varepsilon\mathcal{M}$ is guaranteed to be a connected graph while $(2+\varepsilon)\mathcal{M}$ is not. In spectral terms, these would result in the *normalized* second eigenvalue being either 1 in the disconnected case or smaller otherwise. However, in the unnormalized regime (the adjacency matrix eigenvalues), these models behave similarly when observed in experiments, a behavior which is reflected in our free analysis. The latter is due to the fact that the eigenvalue distribution of the cycle graph on n vertices is given by (see e.g. [HLW06])

$$\mathsf{spec}\left(\mathbf{C}_n\right) = \left\{2\cos\left(\frac{2\pi}{n}j\right) : j = 0,1,\ldots,n-1\right\} \quad (51)$$

which, as $n\to\infty$, converges to the arcsin distribution

$$\mu_{\mathbf{arc}}(t) = \frac{1}{\pi}\frac{1}{\sqrt{4-t^2}}.$$

It is well known that $\mu_{\mathbf{arc}}$ is the free convolution of two Rademacher distributions (the uniform distribution on ±1), which are exactly the distributions induced by two perfect matchings (see [NS06, Chapter 12]).

Proof of Lemma 6.6. In order to apply the technique of Lemma 6.2, we use $\mu_{\mathbf{arc}}$ in the role of μ_1, the distribution easier to work with (and for which we have a $\mathcal{K}$-transform) and μ_ε in the role of μ_2. Recall from Claim 4.10 that

$$\mathcal{K}_C(y) = \frac{\sqrt{1+4y^2}}{y}, \tag{52}$$

which also equals $2\mathcal{K}_{\mathcal{M}}(y)$, where $\mathcal{M}$ stands for a perfect matching. Applying Lemma 6.2 we get

$$\widetilde{\mathcal{K}}(x) = \mathcal{K}_C(\mathcal{G}_\varepsilon(x)) + x - \frac{1}{\mathcal{G}_\varepsilon(x)} = \frac{x}{x^2+\varepsilon-1}\left(\varepsilon + \sqrt{T(x)}\right), \tag{53}$$

where

$$T(x) \triangleq x^4 + 2x^2 + \frac{4(\varepsilon-1)^2}{x^2} + 8\varepsilon - 7.$$

While finding the minimizer x^* for Equation (53) is not feasible analytically, a bound for small ε is possible. To do so, we set $x = 1+b$, where $b = b(\varepsilon)$. We notice that for $\varepsilon = 0$, one can simplify the function and get that $\widetilde{\mathcal{K}}(x) = \sqrt{x^2+1}$, having its minimum value $\sqrt{5}$ achieved at $x = 1$. For this reason, we expect b to approach 0 as $\varepsilon \to 0$.

In a manner similar to the proof of Lemma 6.4, we write $\widetilde{\mathcal{K}}(x) = \frac{f_N(x)}{f_D(x)}$, where $f_N(x) = x\left(\varepsilon + \sqrt{T(x)}\right)$ and $f_D(x) = 2b + b^2 + \varepsilon$. We start with approximating $T(x)$. Under the above assumption that b is small, we use the approximation

$$x^{-2} = 1 - 2b + 3b^2 - 4b^3 \pm O(b^4)$$

which we get by expanding $\frac{1}{1-(-b)} = 1 - b + b^2 - b^3 \pm O(b^4)$. This gives

$$T(1+b) = -16\varepsilon^2 b^3 + 12\varepsilon^2 b^2 - 8\varepsilon^2 b + 4\varepsilon^2 + 32\varepsilon b^3 - 24\varepsilon b^2 + 16\varepsilon b + b^4 - 12b^3 + 20b^2.$$

Denote $t = \sqrt{\varepsilon}$. In order to find the min-value for $\widetilde{\mathcal{K}}(1+b)$, we set $b = ct$ and try to minimize for the constant c, which would yield a bound on the actual minimum (note that the result is, numerically, a very good estimation of the actual minimum). The above formula for $T(x)$ becomes

$$T(1+b) = 20c^2t^2 + \left(16c - 12c^3\right)t^3 + O(t^4).$$

Taking the upper approximation $\sqrt{1+y} < 1 + \frac{y}{2}$ we have that

$$\sqrt{T(1+b)} < 2\sqrt{5}ct + \frac{\sqrt{5}}{5}(4-3c^2)t^2,$$

and the numerator becomes

$$\begin{aligned} f_N(x) &= (1+b)\left(\varepsilon + \sqrt{T(1+b)}\right) \\ &< 2\sqrt{5}ct + \frac{\sqrt{5}}{5}(4-3c^2)t^2 + t^2 + 2\sqrt{5}c^2t^2 + O(t^3) \\ &= 2\sqrt{5}ct + \left(1 + \frac{4}{\sqrt{5}} + \frac{7}{\sqrt{5}}c^2\right)t^2 + O(t^3), \end{aligned}$$

while we have the denominator

$$f_D(x) = 2ct + (c^2+1)t^2.$$

Therefore, ignoring terms of order t^3 in the numerator, we can write

$$\widetilde{\mathcal{K}}(1+ct) < \frac{A+Bt}{C+Dt},$$

where $A = 2\sqrt{5}c$, $B = (1 + \frac{4}{\sqrt{5}} + \frac{7}{\sqrt{5}}c^2)$, $C = 2c$ and $D = c^2+1$. We use the following formula for writing this as a power series in t:

$$\frac{A+Bt}{C+Dt} = \frac{A}{C} + \left(\frac{B}{C} - \frac{AD}{C^2}\right)t + O(t^2),$$

giving us

$$\widetilde{\mathcal{K}}(1+ct) < \sqrt{5} + \left(\frac{c}{\sqrt{5}} + \frac{5-\sqrt{5}}{10c}\right)t + O(t^2).$$

This is minimized when picking $c = \sqrt{\frac{2}{1+\sqrt{5}}} = \sqrt{\frac{1}{\phi}}$, $\phi = \frac{1+\sqrt{5}}{2}$ being the golden ratio. This results in the overall bound of

$$\widetilde{\mathcal{K}}(1+c\sqrt{\varepsilon}) < \sqrt{5} + \sqrt{\frac{4}{5\phi}}\sqrt{\varepsilon}.$$

By Lemma 6.2, we have the desired bound for $\mathsf{pred}_{\mathcal{C}+\varepsilon\mathcal{M}}(\varepsilon)$ and $\mathsf{pred}_{(2+\varepsilon)\mathcal{M}}(\varepsilon)$. □

6.3 Does the free method admit a combinatorial meaning?

The free method is analytic in nature and effectively "forgets" the underlying combinatorial structure. Indeed, the input to the method is the spectrum of a graph, encoded via the Cauchy transform; thus, much of the graph's structural information (as captured by the eigenvectors) is lost. That said, unlike most combinatorial and linear-algebraic techniques used to study expanders, the free method incorporates the *entire* spectrum of the graph, rather than focusing solely on its spectral expansion.

A natural question is whether the free method, particularly Equation (6), has a combinatorial meaning. Although a complete combinatorial interpretation remains elusive—if it exists at all—we offer a reformulation that is more combinatorial in spirit, and believe it may shed some light on the question.

Let $\chi_G(x)$ denote the characteristic polynomial of an undirected graph G. Observe that

$$\mathcal{G}_G(x) = \frac{\dot{\chi}_G(x)}{\chi_G(x)},$$

where we use Newton's dot notation to denote derivatives with respect to x. With this in mind, Equation (6) can be rewritten as

$$(t-1)\frac{\ddot{\chi}_G\chi_G - (\dot{\chi}_G)^2}{\chi_G^2} + t\left(\frac{\dot{\chi}_G}{\chi_G}\right)^2 = 0.$$

Rearranging, we obtain the equivalent form

$$(t-1)\frac{\ddot{\chi}_G}{\dot{\chi}_G} + \frac{\dot{\chi}_G}{\chi_G} = 0.$$

Here, the second term is simply the Cauchy transform $\mathcal{G}_G(x)$. The first term also resembles a Cauchy transform—but what is the object whose characteristic polynomial is given by $\dot{\chi}_G(x)$?

Strictly speaking, there need not exist a graph whose characteristic polynomial is $\dot{\chi}_G(x)$. However, using cofactor expansion or Jacobi's formula for matrix derivatives, one can derive the well-known identity

$$\dot{\chi}_G(x) = \sum_{v \in V} \chi_{G-v}(x),$$

where $G - v$ denotes the graph obtained by removing the vertex v from G. That is, $\dot{\chi}_G(x)$ is not the characteristic polynomial of a graph in the usual sense, but rather the *sum* of characteristic polynomials of graphs closely related to G—each formed by a simple combinatorial operation: vertex deletion. One can view this as a kind of "derivative" of the graph G, which we denote here as ∂G. With this perspective, and with slight abuse of notation, Equation (6) takes the form

$$(t-1)\mathcal{G}_G(x) + \mathcal{G}_{\partial G}(x) = 0. \tag{54}$$

Although this does not fully resolve the question of a combinatorial interpretation of the method, Equation (54) is, in some sense, more combinatorial than its analytically equivalent form Equation (6). Indeed, it avoids analytic operations such as differentiation or squaring of the Cauchy transform, and instead involves a simple linear combination of Cauchy transforms of graphs—or more precisely, of graph sums—where the "derivative"

used to define ∂G is interpreted purely at the combinatorial level.

7 Sum of Free Graphs

In this section, we prove that, by suitably permuting the vertices, the sum of Ramanujan graphs can be made nearly Ramanujan. This implies Theorem 2.7.[3]

Proposition 7.1. *Let G be a d-regular Ramanujan graph ($d \geq 3$) on n vertices with girth g and adjacency matrix $\mathbf{A}$. Then, there exist permutation matrices $\mathbf{P}_1, \ldots, \mathbf{P}_t$ such that*

$$\lambda_2\left(\sum_{i=1}^{t} \mathbf{P}_i \mathbf{A} \mathbf{P}_i^\intercal\right) < 2\sqrt{2t-1} + O\left(\sqrt{dt}\right) \cdot 2^{-\Omega(g)}.$$

Remark 7.2. The statement of Proposition 7.1 is phrased this way for simplicity. The proof shows that an analogous claim holds for sums of graphs with different degrees: for example, summing two Ramanujan graphs of degrees d_1 and d_2—after permuting the second—can yield an almost-Ramanujan graph of degree $d_1 + d_2$. Moreover, the proof extends to expanders that are not Ramanujan. We omit the details.

Proposition 7.1 directly implies that summing $d/2$ cycle graphs, after a suitable permutation, yields a d-regular graph whose second eigenvalue is bounded by

$$2\sqrt{d-1} + O(d)\, 2^{-\Omega(n)}.$$

However, as we show in Section 7.1, when summing cycles we can in fact match the Ramanujan bound exactly, with no error term. This is because for cycle graphs we know—and can readily work with—the entire spectrum (not only the expansion and girth required in Proposition 7.1), and we exploit this in our analysis.

The main ingredient in proving Proposition 7.1 is the following lemma.

Lemma 7.3. *Let G be a d-regular Ramanujan graph on n vertices with girth g and adjacency matrix $\mathbf{A}$. Then,*

$$\mathsf{maxroot}\left(p_{\mathbf{A}}(x)^{\boxplus t}\right) \leq 2\sqrt{t \cdot d - 1} + O(\sqrt{dt}) \cdot 2^{-\Omega(g)}, \tag{55}$$

where $p_{\mathbf{A}}(x)$ is as defined in Section 4.

Before proving Lemma 7.3, we show the straightforward way to deduce our main theorem from it.

[3] Note the change in notation relative to Theorem 2.7: there, d denotes the degree of the resulting graph.

Proof of Proposition 7.1. By Lemma 4.1, it is enough to prove that

$$\alpha_2\left(\mathop{\mathbf{E}}_{\mathbf{P}}\chi_x\left(\mathbf{A}_{\mathbf{P}}\right)\right) \le 2\sqrt{t\cdot d-1}+O(\sqrt{dt})\cdot 2^{-\Omega(g)},$$

where $\mathbf{A}_{\mathbf{P}}=\sum_{i=1}^{t}\mathbf{P}_i\mathbf{A}_i\mathbf{P}_i^{\mathsf{T}}$. By Lemma 4.4 (Equation (20)), we know that

$$\mathop{\mathbf{E}}_{\mathbf{P}}\chi_x\left(\mathbf{A}_{\mathbf{P}}\right)=(x-t\cdot d)p_{\mathbf{A}}^{\boxplus t}(x),$$

and by Lemma 7.3 the proof is complete. □

For proving Lemma 7.3, we start with noting, via Corollary 4.8, that for every $y>0$,

$$\mathsf{maxroot}\left(p_{\mathbf{A}}(x)^{\boxplus t}\right)\le t\cdot\mathcal{K}_{p_{\mathbf{A}}}(y)-\frac{t-1}{y}. \tag{56}$$

However, to bound the RHS of the latter equation, we need to take an analytic view of the $\mathcal{G}$ and $\mathcal{K}$ transforms, following the idea of the *adapter* defined by [CM23, Section 4.1] and simplified here.

Let $a\ge 0$ be a real number, μ a probability measure supported on $[-a,a]$. Let $\mathbf{A}$ be an $n\times n$ real symmetric matrix whose eigenvalues are $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n$. Denote $\lambda=\max(|\lambda_2|,|\lambda_n|)$ and set $b=\max(a,\lambda)$. Let $p_{\mathbf{A}}(x)$ be the polynomial that satisfies $\chi_x\left(\mathbf{A}\right)=(x-\lambda_1)p_{\mathbf{A}}(x)$ as in the notations set in Section 4. We denote the r-th moment of $\mathbf{A}$ as $m_r(\mathbf{A})=\frac{1}{n}\sum_{i=1}^{n}\lambda_i^r$.

Assume that

$$m_r(\mu)=m_r(\mathbf{A})\quad\text{for}\quad r=0,1,\ldots,h, \tag{57}$$

that is, the first h moments of μ and $\mathbf{A}$ match.

Claim 7.4. *With the notation and under the assumptions above, for every $x>b$,*

$$\varepsilon(x)\triangleq\mathcal{G}_{p_{\mathbf{A}}}(x)-\mathcal{G}_{\mu}(x)\le\frac{2}{x-b}\left(\frac{b}{x}\right)^{h+1}.$$

Proof. By Equation (22) we have that

$$\begin{aligned}
\mathcal{G}_{p_{\mathbf{A}}}(x) - \mathcal{G}_{\mu}(x) &= \sum_{r=0}^{\infty} \frac{m_r(p_{\mathbf{A}}) - m_r(\mu)}{x^{r+1}} \\
&= \sum_{r=0}^{h} \frac{m_r(p_{\mathbf{A}}) - m_r(\mu)}{x^{r+1}} + \sum_{r=h+1}^{\infty} \frac{m_r(p_{\mathbf{A}}) - m_r(\mu)}{x^{r+1}} \\
&= \sum_{r=0}^{h} \frac{m_r(p_{\mathbf{A}}) - m_r(\mathbf{A})}{x^{r+1}} + \sum_{r=h+1}^{\infty} \frac{m_r(p_{\mathbf{A}}) - m_r(\mu)}{x^{r+1}} \\
&< \sum_{r=h+1}^{\infty} \frac{m_r(p_{\mathbf{A}}) - m_r(\mu)}{x^{r+1}},
\end{aligned} \tag{58}$$

where in the last inequality we used the fact that any moment of $\mathbf{A}$ is by definition larger than that of $p_{\mathbf{A}}$.

For bounding the resulting sum, we use the triangle inequality on the numerator for each summand and get:

$$|m_r(\mu) - m_r(p_{\mathbf{A}})| \leq a^r + \lambda^r \leq 2b^r.$$

Summing all the terms we have that for $x > b$,

$$\sum_{r=h+1}^{\infty} \frac{m_r(\mu) - m_r(p_{\mathbf{A}})}{x^{r+1}} \leq \sum_{r=h+1}^{\infty} \frac{|m_r(\mu) - m_r(p_{\mathbf{A}})|}{x^{r+1}} \leq \frac{2}{x} \cdot \sum_{r=h+1}^{\infty} \left(\frac{b}{x}\right)^r = \frac{2}{x-b}\left(\frac{b}{x}\right)^{h+1}, \tag{59}$$

completing the proof. □

Claim 7.4 tells us that when μ and $\mathbf{A}$ agree on the low moments, the corresponding $\mathcal{G}$-transforms are close for sufficiently large x and h. However, as hinted in Equation (56), in order to prove Lemma 7.3 we will need to use closeness of their inverses, the $\mathcal{K}$-transforms. To this end, we cite the following claim from [CM23]:

Claim 7.5 ([CM23, Claim 4.6])**.** *Let $f, g : (c, \infty) \to \mathbb{R}$ be differentiable, strictly decreasing convex functions. Note that $f^{-1} : \mathrm{Im}(f) \to (c, \infty)$ and $g^{-1} : \mathrm{Im}(g) \to (c, \infty)$ are well-defined functions. Let $\varepsilon : (c, \infty) \to \mathbb{R}$ be a function such that for every $x > c$, $|f(x) - g(x)| \leq \varepsilon(x)$. Then, for every $y \in \mathrm{Im}(f) \cap \mathrm{Im}(g)$,*

$$\left|f^{-1}(y) - g^{-1}(y)\right| \leq \max\left(\frac{\varepsilon(f^{-1}(y))}{|g'(f^{-1}(y))|}, \frac{\varepsilon(g^{-1}(y))}{|f'(g^{-1}(y))|}\right).$$

We shall prove Lemma 7.3 by using Claim 7.5, where the $\mathcal{G}$-transforms play the roles of f, g and the $\mathcal{K}$-transforms are their inverses.

Proof of Lemma 7.3. By Corollary 4.8, for every $y > 0$ it holds that

$$\mathsf{maxroot}\left(p_{\mathbf{A}}(x)^{\boxplus t}\right) \leq t \cdot \mathcal{K}_{p_{\mathbf{A}}}(y) - \frac{t-1}{y}. \tag{60}$$

Let $\nu_d = \mu_{\mathsf{km}}^{(d)}$ be the Kesten-McKay distribution with parameter d. By straightforward calculation,

$$t\mathcal{K}_{\nu_d}(y) - \frac{d-1}{y} = \frac{2 + dt\left(\sqrt{1+4y^2} - 1\right)}{2y}, \tag{61}$$

where the RHS achieves its minimum value $2\sqrt{td-1}$ at $y_t = \frac{\sqrt{td-1}}{td-2}$. Using Equations (60) and (61), it remains to show that:

$$\mathcal{K}_{p_{\mathbf{A}}}(y_t) \leq \mathcal{K}_{\nu_d}(y_t) + O\left(\sqrt{d}\right) \cdot 2^{-\Omega(g)}. \tag{62}$$

Let

$$x_t = \mathcal{K}_{\nu_d}(y_t) = \frac{(t+1)d - 2}{\sqrt{td-1}}, \tag{63}$$

and let x'_t be such that $\mathcal{G}_{p_{\mathbf{A}}}(x'_t) = y_t$. As the $\mathcal{K}$-transforms are the inverses of the $\mathcal{G}$-transforms, Equation (62) is equivalent to

$$x'_t - x_t \leq O\left(\sqrt{d/t}\right) \cdot 2^{-\Omega(g)}, \tag{64}$$

and it is enough to prove for case that $x'_t > x_t$. Equation (64) has the correct form of Claim 7.5, without the absolute value (as we care only for one side of the inequality). We will proceed by following Claim 7.5 and bounding

$$\max\left(\frac{\varepsilon(x'_t)}{\left|\mathcal{G}'_{p_{\mathbf{A}}}(x'_t)\right|}, \frac{\varepsilon(x_t)}{\left|\mathcal{G}'_{\nu_d}(x_t)\right|}\right)$$

for $x'_t > x_t$. By Claim 4.11, for every $h < g$,

$$m_r(\nu_d) = m_r(\mathbf{A}) \quad \text{for} \quad r = 0, 1, \ldots, h,$$

and so we can invoke Claim 7.4 and get that the numerator is bounded (plugging $b = \lambda \triangleq 2\sqrt{d-1}$) by

$$\varepsilon(x_t) \leq \frac{2}{x_t - \lambda}\left(\frac{\lambda}{x_t}\right)^{g+1} < \frac{2}{(\beta_t - 1)\lambda}\beta_t^{-g},$$

where $\beta_t = \frac{x_t}{\lambda} > \max\left(\frac{\sqrt{t}}{2}, \beta\right)$, and $\beta = \frac{3}{\sqrt{8}}$.

By definition of the Cauchy transform we have that $\mathcal{G}_{p_{\mathbf{A}}}(x) < \frac{1}{x-\lambda}$, and by straightforward calculation x'_t cannot be larger than 2λ. Also by definition of the Cauchy transorm

and λ it holds that for every $x > \lambda$ it holds that $|\mathcal{G}'_{\nu_d}(x)| > \frac{1}{(x+\lambda)^2}$ and $|\mathcal{G}'_{p_{\mathbf{A}}}(x)| > \frac{1}{(x+\lambda)^2}$. Overall we get that

$$\min\left(|\mathcal{G}'_{\nu_d}(x_t)|, |\mathcal{G}'_{p_{\mathbf{A}}}(x'_t)|\right) > \frac{1}{(x'_t+\lambda)^2} > \frac{1}{9\lambda^2}. \tag{65}$$

Combining the above, we have that

$$x'_t - x_t \leq \frac{18\lambda^2}{(\beta_t - 1)\lambda}\beta^{-g} = O\left(\sqrt{d/t}\right) \cdot 2^{-\Omega(g)},$$

as desired. □

7.1 Sum of free cycles

As noted, for cycle graphs we can prove a stronger statement than that implied by Proposition 7.1 by exploiting the full spectrum of the cycle, not just its spectral radius and girth. As in previous sections, we denote by $\mathbf{C}_n$ the adjacency matrix of a cycle graph of size n, and let $\mu_{\mathbf{arc}} = \nu_2$ be the arcsin distribution (the Kesten-McKay distribution with parameter $d = 2$).

We denote the graph $G_{\mathbf{P}}$ to be the (n, d)-graph whose adjacency matrix is

$$G_{\mathbf{P}} = \sum_{i=1}^{d/2} \mathbf{P}_i \mathbf{C}_n \mathbf{P}_i^{\mathsf{T}}, \tag{66}$$

where $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_{d/2}$ are $n \times n$ permutation matrices. We prove here the following (implying Theorem 2.8):

Proposition 7.6. *There exist permutation matrices $\mathbf{P}_1, \ldots, \mathbf{P}_{d/2}$ such that*

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

Given the proof of Lemma 7.3, the following lemma is the remaining piece for proving Proposition 7.6. It can be seen as replacing the error term from Claim 7.4 with 0.

Lemma 7.7. *Let $x_t = \frac{2t}{\sqrt{2t-1}}$ is as in Equation (63). For every $n > 16\pi^3$ and every $t \geq 2$,*

$$\mathcal{G}_{p_{\mathbf{C}_n}}(x_t) < \mathcal{G}_{\mu_{\mathbf{arc}}}(x_t).$$

This will conclude the proof of Proposition 7.6 given that $\mu_{\mathbf{arc}}^{\boxplus \frac{d}{2}} = \mu_{\mathsf{km}}^{(d)}$. Note that a more general statement, for every n and every $x > 2$, holds, however not necessary for our needs.

Proof of Lemma 7.7. We assume for simplicity that $n = 2k$ (the odd case is similar), and

denote $\mathcal{G}(x) \triangleq \mathcal{G}_{p_{\mathbf{C}_n}}(x)$. We have already seen (Equation (51)) that

$$\operatorname{spec}(\mathbf{C}_n) = \left\{ 2\cos\left(\frac{2\pi}{n} j\right) : j = 0, 1, \ldots, n-1 \right\}, \tag{67}$$

and we know that $\mu_{\mathbf{arc}}(s) = \frac{1}{\pi}\frac{1}{\sqrt{4-s^2}}$ and is supported on $(-2, 2)$. We use the following notation:

$$\begin{aligned} f_x(\theta) &= \frac{1}{x - 2\cos\theta} \\ \theta_j &= j\frac{\pi}{k} \\ a_i &= f_x(\theta_{k-i}) \text{ such that } a_0 = \frac{1}{x+2} \text{ and } a_k = \frac{1}{x-2}. \end{aligned}$$

By definition of the Cauchy transform and noticing the multiplicities in Equation (67) we have that

$$\mathcal{G}(x) = \frac{1}{n-1}\left(a_0 + 2\sum_{i=1}^{k-1} a_i\right). \tag{68}$$

We also know by definition that

$$\mathcal{G}_{\mu_{\mathbf{arc}}}(x) = \int_{-2}^{2} \frac{1}{x-t}\mu_{\mathbf{arc}}(t)dt = \frac{1}{\pi}\int_0^{\pi} f_x(\theta)d\theta, \tag{69}$$

where the last equality is a simple change of variables. The trapezoidal rule for integration tells us that the function

$$T_n(x) = \frac{1}{n}\left(a_0 + 2\sum_{i=1}^{k-1} a_i + a_k\right) \tag{70}$$

is a $\frac{\pi^3|E|}{12k^2}$-approximation of $\mathcal{G}_{\mu_{\mathbf{arc}}}(x)$, where $E = f_x''(\zeta)$ for some $\zeta \in [0, \pi]$. It can be shown by straightforward calculation that for $x_t = \frac{2t}{\sqrt{2t-1}}$, $|f_{x_t}''(\zeta)| < \frac{48}{t}$, and hence $T_n(x_t)$ approximates $\mathcal{G}_{\mu_{\mathbf{arc}}}(x_t)$ by at most $\frac{16\pi^3}{tn^2}$. Therefore the proof will be concluded by showing that $T_n(x_t) - \mathcal{G}(x_t)$ is large enough.

By Equations (68) and (70) we have

$$T_n(x) - \mathcal{G}(x) = \frac{1}{n}\left(a_k - \mathcal{G}(x)\right).$$

Notice that $a_i < \frac{1}{x}$ for $i < \frac{k}{2}$ and $a_i \le a_k$ for all i. Therefore $\mathcal{G}(x) < \frac{1}{2}\left(a_k + \frac{1}{x}\right)$. Plugging in x_t we see that $T_n(x_t) - \mathcal{G}(x_t) \ge \frac{1}{tn}$, concluding the proof. □

References

[AEL95] 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.

[AZ24] 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.

[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.

[BC19] Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Annals of Mathematics*, 190(3):811–875, 2019.

[BES20] Zhigang Bao, László Erdős, and Kevin Schnelli. On the support of the free additive convolution. *Journal d'Analyse Mathématique*, 142(1):323–348, 2020.

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

[Bor20] 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.

[BRSW12] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction. *Annals of Mathematics*, pages 1483–1543, 2012.

[CCM24] 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)*, pages 1470–1499. IEEE, 2024.

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

[Chu16] Fan Chung. A generalized Alon-Boppana bound and weak Ramanujan graphs. *the electronic journal of combinatorics*, pages P3–4, 2016.

[CM23] Gil Cohen and Gal Maor. Random walks on rotating expanders. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 971–984, 2023.

[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.

[Coh21] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. *SIAM J. Comput.*, 50(3):STOC16–30–STOC16–67, 2021.

[CY23] Alice Cortinovis and Lexing Ying. Computing free convolutions via contour integrals. *arXiv preprint arXiv:2305.01819*, 2023.

[CZ19] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. *Ann. of Math. (2)*, 189(3):653–705, 2019.

[DG25] Irit Dinur and Oded Goldreich. Expansion without connectivity: A property testing perspective. 2025.

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

[Erd47] Paul Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[Fri08] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.

[FW81] Peter Frankl and Richard M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981.

[Gil52] Edgar N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31(3):504–522, May 1952.

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

[Hoo05] 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.

[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.

[Kur30] Kazimierz Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematicae*, 15:271–283, 1930.

[Li23] Xin Li. Two source extractors for asymptotically optimal entropy, and (many) more. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science—FOCS 2023*, pages 1271–1281. IEEE Computer Soc., Los Alamitos, CA, [2023] ©2023.

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

[Lub11] Alexander Lubotzky. Finite simple groups of Lie type as expanders. *Journal of the European Mathematical Society*, 13(5):1331–1341, 2011.

[Mar82] G. A. Margulis. Explicit constructions of graphs without short cycles and low density codes. *Combinatorica*, 2(1):71–78, 1982.

[McK81] Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[MO20] 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.

[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 Journal on Computing*, 47(6):2488–2509, 2018.

[MSS22] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Finite free convolutions of polynomials. *Probab. Theory Related Fields*, 182(3-4):807–848, 2022.

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

[NS96] Alexandru Nica and Roland Speicher. On the multiplication of free n-tuples of noncommutative random variables. *American Journal of Mathematics*, 118(4):799–832, 1996.

[NS06] 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.

[ON12] Sheehan Olver and Raj Rao Nadakuditi. Numerical computation of convolutions in free probability theory. *arXiv preprint arXiv:1203.1958*, 2012.

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):Art. 17, 24, 2008.

[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.

[Spe19] Roland Speicher. Lecture notes on free probability theory. *arXiv preprint arXiv:1908.08125*, 2019.

[Spi25] Daniel A. Spielman. Spectral and algebraic graph theory. Unpublished book manuscript, 2025.

[SS94] 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.

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

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1–3):1–336, 2012.

[Var57] R. R. Varshamov. Estimate of the number of signals in error-correcting codes. *Doklady Akademii Nauk SSSR*, 117:739–741, 1957. English translation reprinted in I. F. Blake (ed.), *Algebraic Coding Theory: History and Development*, Dowden, Hutchinson & Ross, 1973, pp. 68–71.

[XZC16] Pinchen Xie, Zhongzhi Zhang, and Francesc Comellas. The normalized laplacian spectrum of subdivisions of a graph. *Applied Mathematics and Computation*, 286:250–256, 2016.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il