Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges

Roberto Imbuzeiro Oliveira*

February 10, 2010

Abstract

Consider any random graph model where potential edges appear independently, with possibly different probabilities, and assume that the minimum expected degree is $\omega(\ln n)$. We prove that the adjacency matrix and the Laplacian of that random graph are concentrated around the corresponding matrices of the weighted graph whose edge weights are the probabilities in the random model.

We apply this result to two different settings. In bond percolation, we show that, whenever the minimum expected degree in the random model is not too small, the Laplacian of the percolated graph is typically close to that of the original graph. As a corollary, we improve upon a bound for the spectral gap of the percolated graph due to Chung and Horn.

We also consider *inhomogeneous random graphs* with average degree $\gg \ln n$. In this case we show that the adjacency matrix of the random graph can be approximated (in a suitable sense) by an integral operator defined in terms of the attachment kernel κ.

Our main proof tool might be of independent interest: a new concentration inequality for *matrix martingales* that generalizes Freedman's inequality for the standard scalar setting.

1 Introduction

Much of probabilistic combinatorics deals with questions of the following type:

Question 1.1 *Given a probability distribution over "large" combinatorial objects X and a real-valued parameter $P = P(X)$ defined over such objects, does there exists a typical value $P^{\rm typ}$ such that $P(X)$ is very likely to be close to $P^{\rm typ}$?*

Starting with the seminal work of Shamir and Spencer [57] on the chromatic number of $G_{n,p}$, many answers to instances of the above question have been obtained via *concentration*

* `rob.oliv@gmail.com` IMPA, Rio de Janeiro, RJ, Brazil, 22430-040. Work supported by a *Bolsa de Produtividade em Pesquisa* and a *Projeto Universal* from CNPq, Brazil.

inequalities, and developments in the two fields have often gone hand in hand; see [46, 7] and the references therein for many examples.

In this paper we introduce a new concentration inequality for *random Hermitian matrices* in order to address a variant of Question 1.1. Our combinatorial objects consist of *random graphs with independent edges*. These are random graphs where the events "ij is an edge" (with ij varying over all unordered pairs of vertices) are independent, but not necessarily identically distributed. The new twist is that the "parameters" for which we prove concentration are the *adjacency matrix* and the *graph Laplacian* of the resulting graph (defined in Section 2.3).

We briefly recall why these two matrices are important. Many (real-valued) parameters of a graph can be computed and/or estimated from these two matrices, including the diameter, distances between distinct subsets, discrepancy-like properties, path congestion, chromatic number and the mixing time for random walk; see e.g. [22] for a compendium of these results, [43, 24, 25, 23] for the relationship between the two matrices and "pseudo-random" properties of graphs and [4, 33, 30] for algorithmic applications. Given these facts, our main Theorem (stated below) sheds some light on the typical properties of the corresponding random graph models.

Theorem 1.1 (Loosely stated) *Let $\mathsf{G}_{\mathbf{p}}$ be a random graph on vertex set $[n]$ where each potential edge ij, $1 \leq i \leq j \leq n$ appears with probability $\mathbf{p}(i,j)$. Let $A_{\mathbf{p}}$ and $\mathcal{L}_{\mathbf{p}}$ be the adjacency matrix and graph Laplacian of $\mathsf{G}_{\mathbf{p}}$ and $A_{\mathbf{p}}^{\mathrm{typ}}$ and $\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}$ be the adjacency matrix and Laplacian of the weighted graph $\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}$ where ij has weight $\mathbf{p}(i,j)$ for each pair ij. Define d, Δ as the minimum and maximal weighted degrees in $\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}$. Then there exists a universal constant $C > 0$ such that if $\Delta \geq C \ln n$,*

$$\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| = O\left(\sqrt{\Delta \ln n}\right) \text{ with high probability}$$

and, if $d \geq C \ln n$,

$$\|\mathcal{L}_{\mathbf{p}} - \mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}\| = O\left(\sqrt{\frac{\ln n}{d}}\right) \text{ with high probability.}$$

A more precise quantitative statement of Theorem 1.1 is given in Section 3 below.

Theorem 1.1 is related to several known results about the standard Erdös-Rényi graph $G_{n,p}$ (the special case where $\mathbf{p}(i,j) = p$ for $i \neq j$). We will show in Section 4 that the kind of matrix concentration we prove here is implicit in the literature and that the standard notion of *quasi-randomness* for dense graphs [24, 43] can be reformulated in terms of concentration of the adjacency matrix around the "typical matrix" for the corresponding $G_{n,p}$ model. There is also a relationship between concentration of the Laplacian and quasi-randomness for given degree sequences [25, 25, 24] which is briefly discussed in Section 4.1.

For the special cases just described, the bounds obtained from Theorem 1.1 for the Laplacian are qualitatively sharp, in the sense that they becomes trivial at roughly the same point where one cannot expect concentration to hold. However, more specialized (and much more complex) approaches yield improved bounds [33, 29, 38]. In some sense, this is due to the fact that the typical adjacency matrices and Laplacians for such random graph models turn out to be very degenerate: one of the eigenvalues of each matrix has multiplicity $n-1$, and the other eigenvalue is well separated from the first.

The cases where this does *not* happen turn out to be more interesting. For instance, consider the case of *bond percolation* with a parameter $p \in (0,1)$ on an arbitrary n-vertex graph G. That is, we consider a random subgraph G_p of G that is obtained by retaining each edge of G independently with probability p. Let A be the adjacency matrix and $\mathcal{L}$ be the Laplacian of G (respectively). We will show that when the minimum expected degree in G_p is $\omega(\ln n)$, the adjacency matrix and Laplacian of G_p are close to pA and $\mathcal{L}$ (respectively); therefore, any estimate for G derived from $\mathcal{L}$ continues to hold (at least approximately) for the random subgraph. A simple corollary of our Theorem is a bound for the spectral gap of G_p that improves upon a recent result of Chung and Horn [26], derived via much more complicated methods.

We then turn to the general model of *inhomogeneous random graphs*. These are built from a set of points $X_1, \ldots, X_n$ that are uniformly distributed over $[0,1]$. The probability that i and j are connected in the random graph is $p\kappa(X_i, X_j)$, where $\kappa : M \times M \to \mathbb{R}_+$ is a symmetric function (called a *kernel*) and p is a parameter that controls the density of the resulting graph. Under some technical conditions, we will show that, for $p = \omega(\ln n/n)$, the adjacency matrix of the random graph will correspond to a kind of discretization of an integral operator T_κ defined in terms of κ. Theorem 1.1 takes care of the key step where we show that the adjacency matrix is concentrated around a deterministic matrix; the rest of the argument consists of proving that the latter matrix is an approximation of T_κ in some suitable sense. The end result implies that the random graph and the kernel κ are close in a metric that is stronger than the *cut metric* from the literature on graph limits [16, 17, 49, 13]. Our results also imply that the eigenvalue distributions and the eigenvectors of the adjacency matrix of the random graph model are closely related to those of T_κ.

1.1 A new concentration inequality

The main result, Theorem 3.1, is a straightforward consequence of a new concentration inequality for *random matrices*. Our result bounds the fluctuations $Z - \mathbb{E}[Z]$ of certain random $d \times d$ Hermitian matrices Z from their mean (defined entrywise), as measured by *largest eigenvalue* $\lambda_{\max}(Z - \mathbb{E}[Z])$ and the *spectral norm* $\|Z - \mathbb{E}[Z]\|$.

Not much is known in general about such inequalities. This is in sharp contrast with the scalar case, where there are several remarkable inequalities and many techniques to prove them [46, 19, 7]. The concentration results for random matrices that have been proven correspond to relatively old developments in the scalar case, such as the standard bounds due to Chernoff [20, 2] and Hoeffding [39, 21], as well as Khintchine's inequality [50, 56]. Accordingly, the new concentration result we introduce in this paper is a matrix analogue of *Freedman's inequality* for martingale sequences [34], which dates back to the 1970's. Here is a precise statement. [Measurability and conditional expectations are defined entrywise; see Section 2.4 for this and other definitions.]

Theorem 1.2 (Freedman's Inequality for Matrix Martingales) *Let*

$$0 = Z_0, Z_1, \ldots, Z_n$$

be a sequence of random $d \times d$ Hermitian matrices that forms a martingale sequence with respect to the filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$ (that is, for each $1 \le i \le n$ Z_i is $\mathcal{F}_i$-measurable and $\mathbb{E}\left[Z_i \mid \mathcal{F}_{i-1}\right] = Z_{i-1}$). Suppose further than $\|Z_i - Z_{i-1}\| \le M$ almost surely for each $1 \le i \le n$ and define:

$$W_n \equiv \sum_{i=1}^{n} \mathbb{E}\left[(Z_i - Z_{i-1})^2 \mid \mathcal{F}_{i-1}\right].$$

Then for all $t, \sigma > 0$:

$$\mathbb{P}\left(\lambda_{\max}(Z_n) \ge t,\, \lambda_{\max}(W_n) \le \sigma^2\right) \le d\, e^{-\frac{t^2}{8\sigma^2 + 4Mt}}.$$

Compared with Freedman's original bound, Theorem 1.2 has worse constants in the exponent and an extra d factor (which is necessary; cf. Section 8), but the two bounds are otherwise of the same form. In this paper we only need a version of Theorem 1.2 for independent sums (cf. Remark 7.1 and Corollary 7.1), but the martingale inequality is not any harder to prove.

The proof of Theorem 1.2 follows a methodology first proposed by Ahlswede and Winter [2]. These authors proved a version of the Chernoff bound for matrices which has had a very strong impact on the development of Quantum Information Theory [31, 32, 60]. Christofides and Markström [21] used the same method to obtain a version of Hoeffding's inequality for matrix martingales.

Theorem 1.3 ([21], in abridged form) *In the setting of Theorem 1.2, replace the assumption on $\|Z_i - Z_{i-1}\|$ by the assumption that there exist $0 \le r_i \le 1$ such that $\lambda_{\max}(Z_i - Z_{i-1}) \le 1 - r_i$ and $\lambda_{\max}(Z_{i-1} - Z_i) \le r_i$. Then for all $t > 0$,*

$$\mathbb{P}\left(\lambda_{\max}(Z_n) \ge t\right) \le d\, e^{-nH_{R/n}\left(\frac{R+t}{n}\right)}$$

where $R = \sum_{i=1}^{n} r_i/n$ and for $x, r \in [0, 1]$

$$H_r(x) \equiv x \ln\left(\frac{x}{p}\right) + (1 - x) \ln\left(\frac{1 - x}{1 - p}\right).$$

As we will see in Remark 3.1, this bound would not suffice for our applications. Roughly speaking, our Theorem is better because the variance term in our bound is the largest value of a sum of matrices, not the sum of the largest eigenvalues. In that respect, Theorem 1.2 is closer to an influential bound obtained by Rudelson [56] via certain inequalities from non-commutative probability [50]. The Ahlswede-Winter approach we adopt here has the advantage of requiring no such unfamiliar tools.[1]

Theorem 1.2 should also be contrasted with other ways for controlling eigenvalues and eigenvectors of random matrices. One of them is the "trace method" [26, 28, 38, 35, 36] which consists of analyzing traces of high powers of the matrices under consideration. This method can be very sharp, but it is also quite complex and we will see that we obtain better bounds in one context (but not all contexts) where the trace method has been applied. A more recent way of bounding eigenvalues and eigenvectors is in some sense based on bounding "discrepancies" [33, 29]. This is better than our bound when the technique applies (see e.g. the comments in Section 4.1), but our main applications seem to be beyond the reach of this methodology.

Finally, we note that our result is not quite comparable concentration bounds of Alon, Krivelevich and Vu [6] for the largest eigenvalues of a random symmetric matrix. Our bound is poorer than theirs when applied to k-th largest eigenvalue for any fixed k, but their bound quickly deteriorates when k grows, whereas our result bounds the maximal deviation of all eigenvalues simultaneously (cf. Corollary 3.1), as well as the deviation of eigenspaces (cf. Corollary 3.2). Moreover, their result cannot be used to determine the typical value of each eigenvalue.

1.2 Organization

The remainder of the paper is organized as follows. After the preliminary Section 2, we prove the main concentration result in Section 3. As a test case, we apply our results to the Erdös-Rényi random graph Section 4 where the connection with quasi-randomness is also discussed. Bond percolation is discussed in Section 5. The more complicated case of inhomogeneous random graphs is treated in Section 6, where we also compare our results to what is known about graph limits. The new concentration inequality is proven in Section 7. Some final remarks are made in Section 8. The Appendix contains two simple results on the perturbation theory of compact operators for which we did not find adequate references.

[1] There is now a proof of Rudelson's bound along the lines of the Ahlswede-Winter method; see [53] for details and for further discussion on the difference between the three bounds.

2 Preliminaries

2.1 Matrix notation

The space of $d_r \times d_c$ matrices with real (resp. complex) entries will be denoted by $\mathbb{R}^{d_r\times d_c}$ (resp. $\mathbb{C}^{d_r\times d_c}$). Moreover, for $A \in \mathbb{C}^{d_r\times d_c}$, $A^* \in \mathbb{C}^{d_c\times d_r}$ is the conjugate transpose of A. We will identify $\mathbb{R}^d$ (resp. $\mathbb{C}^d$) with the space $\mathbb{R}^{d\times 1}$ (resp. $\mathbb{C}^{d\times 1}$) of column vectors, so that the inner product of $v, w \in \mathbb{R}^d$ is w^*v. $\|\cdot\|$ denotes both the Euclean norm on $\mathbb{R}^d$ or $\mathbb{R}^d$ and the *spectral radius* norm induced on $\mathbb{C}^{d\times d'}$:

$$\|A\| \equiv \sup_{v\in\mathbb{C}^d, \|v\|=1} \|Av\|, A \in \mathbb{C}^{d\times d'}.$$

$\mathbb{C}^{d\times d}_{\mathrm{Herm}}$ is the space of $d\times d$ *Hermitian matrices*, which are the $A \in \mathbb{C}^{d\times d}$ with $A^* = A$. $\mathbb{R}^{d\times d}_{\mathrm{Herm}}$ is similarly defined; one could of course speak of *symmetric* matrices in this case and use $A^\dagger$ instead of A^*, but we will keep notation consistent.

The spectral theorem implies that for any $A \in \mathbb{C}^{d\times d}_{\mathrm{Herm}}$ there exist real numbers $\lambda_0(A) \leq \lambda_1(A) \leq \cdots \leq \lambda_{d-1}(A)$ and orthonormal vectors $\psi_0, \dots, \psi_{d-1}$ (the eigenvalues and eigenvectors of A, respectively) with:

$$A \equiv \sum_{i=0}^{d-1} \lambda_i(A)\, \psi_i\psi_i^*.$$

The *spectrum of* A is the set $\mathrm{spec}(A)$ of all $\lambda_i(A)$. The above formula implies that for $A \in \mathbb{C}^{d\times d}_{\mathrm{Herm}}$:

$$\|A\| = \max_{0\leq i\leq d-1} |\lambda_i(A)| = \max_{v\in\mathbb{C}^d, \|v\|=1} v^*Av.$$

For $A \in \mathbb{R}^{d\times d}_{\mathrm{Herm}}$, the eigenvectors of A are all real and one only needs to maximize over $v \in \mathbb{R}^d$ in the above formula to compute $\|A\|$.

We also note an equivalent statement of the spectral theorem as:

$$A = \sum_{\alpha\in\mathrm{spec}(A)} \alpha\, \Pi_\alpha,$$

where the $\{\Pi_\alpha\}_{\alpha\in\mathrm{spec}(A)}$ are projections with orthogonal ranges and $\sum_{\alpha\in\mathrm{spec}(A)} \Pi_\alpha = I_d$, the $d\times d$ identity matrix. The *multiplicity* of $\alpha \in \mathrm{spec}(A)$ is the dimension of the range of the corresponding Π_α; this is equal to the number of $0 \leq i \leq d-1$ with $\lambda_i(A) = \alpha$.

2.2 Integral operators on $L^2([0,1])$ and spectral theory

In Section 6 we will compare adjacency matrices with certain integral operators on $L^2([0,1])$. The spectral theory of these and other compact operators is a classical topic in Functional

Analysis and we refer to [55, 45] for all the results we review in this Section.

We will work with the space $L^2([0,1])$ of real measurable functions that are square-integrable with respect to Lebesgue measure. This space has a natural inner product

$$(f,g)_{L^2} \equiv \int_0^1 f(x)\, g(x)\, dx \quad (f,g \in L^2([0,1]))$$

and an associated norm $\|f\|_{L^2}^2 \equiv (f,f)_{L^2}$ with respect to which it is a real Hilbert space.

Given a function $\eta \in L^2([0,1]^2)$ (the latter space being defined similarly to $L^2([0,1])$), one can define a linear operator on $L^2([0,1])$ by the formula:

$$T_\eta : f(\cdot) \in L^2([0,1]) \mapsto (T_\eta f)(\cdot) \equiv \int_0^1 \eta(\cdot, y)\, f(y)\, dy. \tag{2.1}$$

The "$L^2 \to L^2$" norm of a linear operator V from $L^2([0,1])$ to itself is given by:

$$\|V\|_{L^2 \to L^2} \equiv \sup_{f \in L^2([0,1]) \setminus \{0\}} \frac{\|Vf\|_{L^2}}{\|f\|_{L^2}}.$$

It is an exercise to show via the Cauchy Schwartz inequality that:

$$\|T_\eta\|_{L^2 \to L^2}^2 \leq \int_{[0,1]^2} \eta^2(x,y)\, dx\, dy. \tag{2.2}$$

Moreover, if $\eta' : [0,1]^2 \to \mathbb{R}$ is also square integrable, $T_\eta - T_{\eta'}$ equals $T_{\eta-\eta'}$.

Assume that $\eta(x,y) = \eta(y,x)$ for almost every $(x,y) \in [0,1]$ (i.e. η is symmetric). In that case the operator T_η is a *compact, self adjoint* linear operator on the Hilbert space $L^2([0,1])$.

Let us recall what these properties imply. Let T be a bounded, compact, self-adjoint operator on the Hilbert space $\mathcal{H}$. Then there exists a finite or countable set $S \subset \mathbb{R}$ and a family $\{P_\alpha : \alpha \in S\}$ of orthogonal projection operators on $\mathcal{H}$ with orthogonal ranges such that:

$$T = \sum_{\alpha \in S} \alpha\, P_\alpha \text{ and } \mathrm{Id}_{\mathcal{H}} = \text{ identity operator on } \mathcal{H} = \sum_{\alpha \in S} P_\alpha.$$

Moreover, either S is finite and contains 0, or S is a countable, bounded subset of $\mathbb{R}$ with 0 as its only accumulation point. Finally, all P_α for $\alpha \neq 0$ are finitely dimensional; the *multiplicity* of α is precisely the dimension of the range of P_α. The spectrum of T is the set $\mathrm{spec}(T_\eta) = S \cup \{0\}$.

2.3 Concepts from Graph Theory

For our purposes a graph $G = (V,E)$ consists of a finite set V of vertices and a set E of edges, which are subsets of size 1 (loops) or 2 of V (we do not allow for parallel edges). Unless otherwise noted, we will assume that $V = [n]$ for some integer $n \geq 2$, where $[n] \equiv \{1, 2, \ldots, n\}$. We will write edges as pairs ij (allowing for $i = j$), but we make no distinction between ij and ji. We will also write $i \sim_G j$ to mean that $ij \in E$. The *degree* $\mathrm{d}_G(i)$ of a vertex i is the number of $1 \leq j \leq n$ such that $ij \in E$.

Assume that $V = [n]$. The *adjacency matrix* of G is the $n \times n$ matrix $A = A_G$ such that, for all $1 \leq i, j \leq n$, the (i, j)-th entry of A is 1 if $ij \in E$ and 0 otherwise. The *Laplacian* $\mathcal{L} = \mathcal{L}_G$ of G is the matrix:

$$\mathcal{L}_G = I_n - T_G \, A_G \, T_G$$

where T is the $n \times n$ diagonal matrix whose (i, i)-th entry is $\mathrm{d}_G(i)^{-1/2}$ if $\mathrm{d}_G(i) \neq 0$, or 0 if $\mathrm{d}_G(i) = 0$. We also let

$$\lambda(G) \equiv \min\{\lambda_1(\mathcal{L}), 2 - \lambda_{d-1}(\mathcal{L})\}$$

denote the *spectral gap* of G.

We will also consider *weighted graphs*, which correspond to a graph $H = (V', E')$ where a positive weight $w_e > 0$ is assigned to each edge $e \in E$. This is the same as defining a symmetric function $w : (V')^2 \to [0, +\infty)$ (i.e. $w(i, j) = w(j, i) \geq 0$ for all $i, j \in V$) and setting $E' = \{\{i, j\} \,:\, w(i, j) > 0\}$. In this case, the *degree* of $i \in V'$ is defined as

$$\mathrm{d}_H(i) \equiv \sum_{j=1}^{n} w(i, j).$$

Assume $V' = [m]$. The adjacency matrix of such an H is the $m \times m$ matrix A_H where for each $1 \leq i, j \leq m$ the (i, j)-th entry of A_H is $w(i, j)$. The Laplacian $\mathcal{L}_H$ is defined as

$$\mathcal{L}_H \equiv I_m - T_H A_H T_H,$$

where T_H is defined as before, but with the new notion of degree. The definition of $\lambda(H)$ is the same as for unweighted graphs.

2.4 Probability with matrices

We will be dealing with random Hermitian matrices throughout the paper. Following common practice, we will always assume that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in the background where all random variables are defined.

Call a map $X : \Omega \to \mathbb{C}^{d\times d}_{\mathrm{Herm}}$ a random $d \times d$ Hermitian matrix (or a $\mathbb{C}^{d\times d}_{\mathrm{Herm}}$-valued random variable) if for each $1 \leq i, j \leq n$, the function $X(i, j) : \Omega \to \mathbb{C}^{d\times d}$ corresponding to the (i, j)-th entry of X is $\mathcal{F}$-measurable. We say that X is *integrable* if all of these entries are and let $\mathbb{E}[X]$ be the matrix whose (i, j)-th entry is $\mathbb{E}[X(i, j)]$. Conditional expectations with respect to a sub σ-field are also defined entrywise.

If the entries are also square-integrable, one can define the variance by the usual formula,

$$\mathbb{V}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

The standard identity $\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$ also holds in this setting.

We will need two easily checked properties of matrix (conditional) expectations, valid for all integrable random $d \times d$ Hermitian matrices X and Y and any sub-σ-field $\mathcal{G} \subset \mathcal{F}$:

$$\begin{aligned} [\text{Tr } \textbf{and } \mathbb{E}[\dots] \textbf{ commute}]\ \mathrm{Tr}(\mathbb{E}[X]) &= \mathbb{E}[\mathrm{Tr}(X)]. && (2.3) \\ [\textbf{Conditioning}] \text{ If } Y \text{ is } \mathcal{G}\text{-measurable, } \mathbb{E}[XY \mid \mathcal{G}] &= \mathbb{E}[X \mid \mathcal{G}]\,Y \\ \text{and } \mathbb{E}[XY] &= \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\,Y]. && (2.4) \end{aligned}$$

3 Concentration of graph matrices

In this section we state and prove our main result, Theorem 1.1.

Given $n \in \mathbb{N}\backslash\{0,1\}$, let $\mathbf{p} : [n]^2 \to [0,1]$ be symmetric: $\mathbf{p}(i,j) = \mathbf{p}(j,i)$ for all $1 \le i,j \le n$. Define independent $0/1$ random variables $\{I_{ij} : 1 \le i \le j \le n\}$ with

$$\mathbb{P}(I_{ij} = 1) = 1 - \mathbb{P}(I_{ij} = 0) = \mathbf{p}(i,j).$$

We also define $I_{ji} = I_{ij}$ for $j > i$.

Define a random unweighted graph $\mathsf{G}_{\mathbf{p}}$ with vertex set $[n]$ and edge set

$$E_{\mathbf{p}} \equiv \{ij : 1 \le i \le j \le n,\ I_{ij} = 1\}.$$

Let $A_{\mathbf{p}}$ and $\mathcal{L}_{\mathbf{p}}$ be the adjacency matrix and Laplacian of the graph $\mathsf{G}_{\mathbf{p}}$. We will compare these to the corresponding matrices $A_{\mathbf{p}}^{\mathrm{typ}}$, $\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}$ of the weighted graph $\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}$ defined by the function $\mathbf{p}$.

The following is a more precise statement of Theorem 1.1.

Theorem 3.1 (Existence of typical graph matrices) *For any constant $c > 0$ there exists another constant $C = C(c) > 0$, independent of n or $\mathbf{p}$, such that the following holds. Let $d \equiv \min_{i \in [n]} d_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)$, $\Delta \equiv \max_{i \in [n]} d_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)$. If $\Delta > C \ln n$, then for all $n^{-c} \le \delta \le 1/2$,*

$$\mathbb{P}\left(\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| \le 4\sqrt{\Delta\, \ln(n/\delta)}\right) \ge 1 - \delta.$$

Moreover, if $d \geq C \ln n$, then for the same range of δ:

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbf{p}} - \mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}\| \leq 14\sqrt{\frac{\ln(4n/\delta)}{d}}\right) \geq 1 - \delta.$$

We will quickly derive some corollaries before we prove Theorem 3.1.

Let $B_1, B_2 \in \mathbb{R}^{n\times n}_{\mathrm{Herm}}$. Standard eigenvalue interlacing inequalities [40] imply:

$$\max_{i\in\{0,\ldots,n-1\}} |\lambda_i(B_1) - \lambda_i(B_2)| \leq \|B_1 - B_2\|. \tag{3.1}$$

This immediately implies that:

Corollary 3.1 *In the setting of Theorem 3.1,*

$$\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| \leq 4\sqrt{\Delta\ \ln(n/\delta)} \Rightarrow \forall 0 \leq i \leq n-1,\ |\lambda_i(A_{\mathbf{p}}) - \lambda_i(A_{\mathbf{p}}^{\mathrm{typ}})| \leq 4\sqrt{\Delta\ \ln(n/\delta)}.$$

Therefore, the RHS holds with probability $\geq 1-\delta$ for any $n^{-c} < \delta < 1/2$ if $\Delta \geq C \ln n$. Similarly,

$$\|\mathcal{L}_{\mathbf{p}} - \mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}\| \leq 14\sqrt{\frac{\ln(4n/\delta)}{d}} \Rightarrow \forall 0 \leq i \leq n-1,\ |\lambda_i(\mathcal{L}_{\mathbf{p}}) - \lambda_i(\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}})| \leq 14\sqrt{\frac{\ln(4n/\delta)}{d}},$$

and the RHS holds with probability $\geq 1-\delta$ for all δ as above if $d \geq C \ln n$.

Now consider some $B \in \mathbb{R}^{n\times n}_{\mathrm{Herm}}$ and, for $a < b$ real, let $\Pi_{a,b}(B)$ be the orthogonal projector onto the space spanned by the eigenvectors of B corresponding to eigenvalues in $[a,b]$. The following corollary is a consequence of Lemma A.2 in the Appendix, as all operators on a finite-dimensional Hilbert space are compact.

Corollary 3.2 *Given some $\gamma > 0$, let $N_\gamma(A_{\mathbf{p}}^{\mathrm{typ}})$ be the set of all pairs $a < b$ such $a + \gamma < b - \gamma$ and $A_{\mathbf{p}}^{\mathrm{typ}}$ has no eigenvalues in $(a-\gamma, a+\gamma) \cup (b-\gamma, b+\gamma)$. Then for $\gamma > 4\sqrt{\Delta\ \ln(n/\delta)}$,*

$$\begin{aligned}&\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| \leq 4\sqrt{\Delta\ \ln(n/\delta)}\\ &\quad\Rightarrow \forall (a,b) \in N_\gamma(A_{\mathbf{p}}^{\mathrm{typ}}),\ \|\Pi_{a,b}(A_{\mathbf{p}}) - \Pi_{a,b}(A_{\mathbf{p}}^{\mathrm{typ}})\| \leq \left(\frac{4(b-a+2\gamma)}{\pi(\gamma^2 - \gamma\sqrt{\Delta\ \ln(n/\delta)})}\right)\sqrt{\Delta\ \ln(n/\delta)}.\end{aligned}$$

In particular, the RHS holds with probability $\geq 1-\delta$ for any $n^{-c} < \delta < 1/2$.

Define $N_\gamma(\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}})$ similarly. Then for $\gamma > 14\sqrt{\ln(4n/\delta)/d}$,

$$\|\mathcal{L}_{\mathbf{p}} - \mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}\| \le 14\sqrt{\frac{\ln(4n/\delta)}{d}}$$
$$\Rightarrow \forall (a,b) \in N_\gamma(A_{\mathbf{p}}^{\mathrm{typ}})\, \|\Pi_{a,b}(\mathcal{L}_{\mathbf{p}}) - \Pi_{a,b}(\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}})\| \le \left(\frac{14(b-a+2\gamma)}{\pi(\gamma^2 - \gamma\sqrt{\frac{\ln(4n/\delta)}{d}})}\right)\sqrt{\frac{\ln(4n/\delta)}{d}}.$$

In particular, the RHS holds with probability $\geq 1-\delta$ for any $n^{-c} < \delta < 1/2$.

The upshot is that for any range of eigenvalues of $A_{\mathbf{p}}^{\mathrm{typ}}$ (resp. $\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}$) that are well-separated from the rest of the spectrum, the projection onto the corresponding eigenvectors of A (resp. $\mathcal{L}$) will be typically close to that of $A_{\mathbf{p}}^{\mathrm{typ}}$ (resp. $\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}$)[2]. We will see when dealing with inhomogeneous random graphs that the separation conditions demanded by the corollary are satisfied in non-trivial cases.

3.1 Proof of the concentration result

We now prove Theorem 3.1.

Proof: [of Theorem 3.1] Let $\{\mathbf{e}_i\}_{i=1}^n$ be the canonical basis for $\mathbb{R}^n$. For each $1 \le i,j \le n$, define a corresponding matrix A_{ij}:

$$A_{ij} \equiv \begin{cases} \mathbf{e}_i\mathbf{e}_j^* + \mathbf{e}_j\mathbf{e}_i^*, & i \neq j; \\ \mathbf{e}_i\mathbf{e}_i^*, & i = j. \end{cases} \in \mathbb{R}^{n\times n}_{\mathrm{Herm}}. \tag{3.2}$$

One can check that $A_{\mathbf{p}} = \sum_{1\le i\le j\le n} I_{ij}\, A_{ij}$ and $A_{\mathbf{p}}^{\mathrm{typ}} = \sum_{1\le i\le j\le n} \mathbf{p}(i,j) A_{ij}$. Therefore,

$$A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}} = \sum_{1\le i\le j\le n} X_{ij} \text{ where } X_{ij} \equiv (I_{ij} - \mathbf{p}(i,j))\, A_{ij},\ 1 \le i \le j \le n.$$

We wish to apply Theorem 1.2 (or rather, Corollary 7.1 in Section 7) to the above sum. To do this, we first notice that the random matrices X_{ij}, which take values in $\mathbb{C}^{n\times n}_{\mathrm{Herm}}$, are independent (since the I_{ij} are) and have mean zero (since $\mathbb{E}[I_{ij}] = \mathbf{p}(i,j)$). Moreover,

$$\|X_{ij}\| \le \|A_{ij}\| = 1$$

as the eigenvalues of A_{ij} are always contained in the set $\{1,0,-1\}$. Thus the assumptions of the Corollary apply with $M = 1$, but we still need to compute the sum of the variances. For

[2] Of course, there is not much one can do near eigenvalue degeneracies, where eigenvectors are typically unstable.

this, fix some pair ij and note that:

$$\mathbb{E}\left[X_{ij}^2\right] = \mathbb{E}\left[(I_{ij} - \mathbf{p}(i,j))^2 A_{ij}^2\right] = \mathbf{p}(i,j)(1-\mathbf{p}(i,j))A_{ij}^2$$

and a computation reveals that

$$A_{ij}^2 = \begin{cases} \mathbf{e}_i\mathbf{e}_i^* + \mathbf{e}_j\mathbf{e}_j^*, & i \neq j \\ \mathbf{e}_i\mathbf{e}_i^*, & i = j. \end{cases} \tag{3.3}$$

Therefore,

$$\begin{aligned} \sum_{i\leq j} \mathbb{E}\left[X_{ij}^2\right] &= \sum_i \mathbf{p}(i,i)(1-\mathbf{p}(i,i))\mathbf{e}_i\mathbf{e}_i^* + \sum_{i<j} \mathbf{p}(i,j)(1-\mathbf{p}(i,j))(\mathbf{e}_i\mathbf{e}_i^* + \mathbf{e}_j\mathbf{e}_j^*) \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n \mathbf{p}(i,j)(1-\mathbf{p}(i,j))\right) \mathbf{e}_i\mathbf{e}_i^*. \end{aligned}$$

This is a diagonal matrix and its largest eigenvalue is at most

$$\max_{i\in[n]} \left(\sum_{j=1}^n \mathbf{p}(i,j)(1-\mathbf{p}(i,j))\right) \leq \max_{i\in[n]} \sum_{j=1}^n \mathbf{p}(i,j) = \Delta.$$

One can now apply Corollary 7.1 with $\sigma^2 = \Delta$ and $M = 1$ to obtain:

$$\forall t > 0,\ \mathbb{P}\left(\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| \geq t\right) \leq 2n\, e^{-\frac{t^2}{8\Delta+4t}}. \tag{3.4}$$

Now let $c > 0$ be given and assume $n^{-c} \leq \delta \leq 1/2$. Then it is clear that there exists a $C = C(c)$ independent of n and $\mathbf{p}$ such that whenever $\Delta \geq C \ln n$,

$$t = 4\sqrt{\Delta \ln(2n/\delta)} \leq 2\Delta.$$

Plugging this t into (3.4) yields:

$$\mathbb{P}\left(\|A_{\mathbf{p}} - A_{\mathbf{p}}^{\mathrm{typ}}\| \geq 4\sqrt{\Delta\ln(2n/\delta)}\right) \leq 2ne^{-\frac{t^2}{16\Delta}} = 2n\,e^{-\frac{16\Delta\ln(2n/\delta)}{16\Delta}} = \delta.$$

This proves the first inequality in Theorem 3.1.

In order to prove the second inequality, we again fix $n^{-c} \leq \delta \leq 1/2$. Our first task is to control the vertex degrees in $\mathsf{G}_{\mathbf{p}}$. Notice that for each $1 \leq i \leq n$, $\mathrm{d}_{\mathsf{G}_{\mathbf{p}}}(i) = \sum_{j=1}^n I_{ij}$ is a sum of independent indicator random variables and the mean of that sum is $\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i) \geq d$. Standard Chernoff bounds [7] (or the case $d = 1$ of our own Corollary 7.1!) imply that there exists a value of $C = C(c)$ such that for $d \geq C \ln n$,

$$\forall i \in [n],\, \mathbb{P}\left(\left|\frac{d_{\mathsf{G}_{\mathbf{p}}}(i)}{d_{\mathsf{G}^{\mathrm{typ}}_{\mathbf{p}}}(i)} - 1\right| > 4\sqrt{\frac{\ln(4n/\delta)}{d}}\right) \leq \delta/2n.$$

Thus with probability $\geq 1 - \delta/2$ one has that

$$\forall i \in [n], \left|\frac{d_{\mathsf{G}_{\mathbf{p}}}(i)}{d_{\mathsf{G}^{\mathrm{typ}}_{\mathbf{p}}}(i)} - 1\right| \leq 4\sqrt{\frac{\ln(4n/\delta)}{d}}. \tag{3.5}$$

We will use this inequality to compare the matrices

$$T = \text{ diagonal with } \mathrm{d}_{\mathsf{G}_{\mathbf{p}}}(i)^{-1/2} \text{ at the } (i,i)\text{th position}$$

and

$$T_{\mathrm{typ}} = \text{ diagonal with } \mathrm{d}_{\mathsf{G}^{\mathrm{typ}}_{\mathbf{p}}}(i)^{-1/2} \text{ at the } (i,i)\text{th position}$$

By increasing C if necessary (and recalling that $\delta > n^{-c}$, $d > C \ln n$), we can ensure that the RHS of (3.5) is at most $3/4$. By the Mean Value Theorem for any $x \in [-3/4, 3/4]$:

$$|\sqrt{1+x} - 1| \leq \left(\sup_{\theta \in [-3/4, 3/4]} \frac{1}{2\sqrt{1+\theta}}\right) |x| = x.$$

Applying this to

$$x \equiv \frac{d_{\mathsf{G}_{\mathbf{p}}}(i)}{d_{\mathsf{G}^{\mathrm{typ}}_{\mathbf{p}}}(i)} - 1$$

yields that:

$$\begin{aligned} \|TT^{-1}_{\mathrm{typ}} - I\| &= \max_{1 \leq i \leq n} \left|\frac{\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}}(i)}}{\sqrt{\mathrm{d}_{\mathsf{G}^{\mathrm{typ}}_{\mathbf{p}}}(i)}} - 1\right| \\ &\leq 4\sqrt{\frac{\ln(4n/\delta)}{d}} \text{with probability } \geq 1 - \delta/2. \end{aligned} \tag{3.6}$$

We now wish compare $\mathcal{L}_{\mathbf{p}} = I - TA_{\mathbf{p}}T$ to $\mathcal{L}^{\mathrm{typ}}_{\mathbf{p}} = I - T_{\mathrm{typ}}A^{\mathrm{typ}}_{\mathbf{p}}T_{\mathrm{typ}}$. Introduce an intermediate operator:

$$\mathcal{M} \equiv I - T_{\mathrm{typ}} A_{\mathbf{p}} T_{\mathrm{typ}}. \tag{3.7}$$

A calculation reveals that:

$$\mathcal{M} = I - (TT^{-1}_{\mathrm{typ}})(I - \mathcal{L}_{\mathbf{p}})(TT^{-1}_{\mathrm{typ}})$$

The spectrum of any Laplacian lies in $[0,2]$ [22]; this implies $\|I - \mathcal{L}_{\mathbf{p}}\| \leq 1$. Using this in conjunction with (3.6) yields:

$$\begin{aligned}
\|\mathcal{M} - \mathcal{L}_{\mathbf{p}}\| &= \|(TT_{\text{typ}}^{-1})(I - \mathcal{L}_{\mathbf{p}})(TT_{\text{typ}}^{-1}) - (I - \mathcal{L}_{\mathbf{p}})\| \\
&\leq \|(TT_{\text{typ}}^{-1} - I)(I - \mathcal{L}_{\mathbf{p}})(TT_{\text{typ}}^{-1})\| \\
&\quad + \|(I - \mathcal{L}_{\mathbf{p}})(TT_{\text{typ}}^{-1})\| \\
\text{(use “}\|ABC\| \leq \|A\|\|B\|\|C\|\text{”)} \quad &\leq \|TT_{\text{typ}}^{-1} - I\| \, \|I - \mathcal{L}_{\mathbf{p}}\| \, \|TT_{\text{typ}}^{-1}\| \\
&\quad + \|I - \mathcal{L}_{\mathbf{p}}\| \, \|TT_{\text{typ}}^{-1} - I\| \\
&\leq 4\sqrt{\frac{\ln(4n/\delta)}{d}} \left(1 + 4\sqrt{\frac{\ln(4n/\delta)}{d}}\right) + 4\sqrt{\frac{\ln(4n/\delta)}{d}} \\
&\leq 10\sqrt{\frac{\ln(4n/\delta)}{d}} \text{ with probability } \geq 1 - \delta/2,
\end{aligned}$$

where again we increase C if necessary to ensure that $d \geq C \ln n$ and $\delta > n^{-c}$ imply the desired bound.

To finish the proof, we must show that $\|\mathcal{M} - \mathcal{L}_{\mathbf{p}}^{\text{typ}}\| \leq 4\sqrt{\ln(4n/\delta)/d}$ with probability $\geq 1 - \delta/2$. For this we will use the concentration result, Corollary 7.1. One can write:

$$\mathcal{L}_{\mathbf{p}}^{\text{typ}} - \mathcal{M} = \sum_{i \leq j} T_{\text{typ}} X_{ij} T_{\text{typ}}$$

where the X_{ij} are the same matrices from the first part of the proof (cf. (3.1)). Again we have a sum of mean-0 independent random matrices, in this case:

$$Y_{ij} \equiv T_{\text{typ}} X_{ij} T_{\text{typ}} = (I_{ij} - \mathbf{p}(i,j)) \frac{A_{ij}}{\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(j)}}, \text{ with } A_{ij} \text{ as in (3.2).}$$

In all possible cases, the eigenvalues of Y_{ij} are contained in the set:

$$\left\{ \frac{\pm(1 - \mathbf{p}(i,j))}{\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(j)}}, \frac{\pm\mathbf{p}(i,j)}{\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(j)}}, 0 \right\}$$

and therefore

$$\|Y_{ij}\| \leq 1/\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\text{typ}}}(j)} \leq 1/d.$$

The sum of variances is:

$$\begin{aligned}
\sum_{i\leq j}\mathbb{E}\left[Y_{ij}^2\right] &= \sum_{i\leq j}\mathbb{E}\left[(I_{ij}-\mathbf{p}(i,j))^2\right]\left(\frac{A_{ij}}{\sqrt{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(j)}}\right)^2 \\
\text{(use (3.3))} &= \sum_{i<j}\mathbf{p}(i,j)(1-\mathbf{p}(i,j))\,\frac{\mathbf{e}_i\mathbf{e}_i^*+\mathbf{e}_j\mathbf{e}_j^*}{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(j)} \\
&\quad+\sum_{i}\mathbf{p}(i,i)(1-\mathbf{p}(i,i))\,\frac{\mathbf{e}_i\mathbf{e}_i^*}{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)^2} \\
&= \sum_{i=1}^{n}\frac{1}{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)}\left(\sum_{j=1}^{n}\frac{\mathbf{p}(i,j)(1-\mathbf{p}(i,j))}{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(j)}\right)\mathbf{e}_i\mathbf{e}_i^*
\end{aligned}$$

Again we have a diagonal matrix. Its (i,i)-th entry is at most:

$$\frac{1}{\mathrm{d}_{\mathsf{G}_{\mathbf{p}}^{\mathrm{typ}}}(i)}\left(\sum_{j=1}^{n}\frac{\mathbf{p}(i,j)}{d}\right)=\frac{1}{d}.$$

We may thus apply Corollary 7.1 to $\sum_{ij}Y_{ij}$ with $M=\sigma^2=1/d$ to obtain:

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}-\mathcal{M}\|\geq t\right)\leq 2n\,e^{-\frac{t^2\,d}{8+4t}}.$$

To finish the proof, we take:

$$t=4\sqrt{\frac{\ln(4n/\delta)}{d}}.$$

We have already ensured that $t\leq 3/4\leq 2$. This implies

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}}-\mathcal{M}\|\geq 4\sqrt{\frac{\ln(4n/\delta)}{d}}\right)\leq 2n\,e^{-\frac{16\ln(4n/\delta)}{16}}\leq\frac{\delta}{2}.$$

This was precisely the required bound. □

Remark 3.1 (Comparing concentration bounds) *We now explain why the Hoeffding bound of Christofides and Markström [21] is insufficient for our purposes. In the case of the adjacency matrix, the random sum we deal with is* $\sum_{ij}(I_{ij}-\mathbf{p}(i,j))A_{ij}$. *We observed above that* A_{ij} *has eigenvalues* 1, -1 *and* 0, *hence we would have to take* $r_i=1/2$ *in order to apply Theorem 1.3 to* $(A_{\mathbf{p}}-A_{\mathbf{p}}^{\mathrm{typ}})/2$. *A simple calculation shows that the exponent in that bound would be of the order* $-t^2/\binom{n}{2}$ *for small enough* t, *which is much worse than the* $-t^2/\Delta$ *behavior we obtain. Our improvement comes from the fact that our "variance" term is the largest eigenvalue of a sum,* not *the sum of largest eigenvalues. Similar comments apply to the concentration of the*

Laplacian.

4 The Erdös-Rényi graph and quasi-randomness

As a first illustration of Theorem 3.1, we apply our results to the Erdös-Rényi graphs. Our bounds are suboptimal in this very special case, but the stronger results in [38, 33] require more difficult arguments that do not seem to generalize to other cases of bond percolation (cf. Section 5). Moreover, our result correctly predicts the range of p for which one can expect concentration of the adjacency matrix.

We then connect concentration to the theory of *quasi-randomness for dense graphs* [23] showing that, in a certain sense, quasi-randomness is equivalent to concentration of the adjacency matrix.

While we will not dwell on this point, a similar connection could be presented between *random graphs with given expected degrees* [28] and concentration of the Laplacian. Our bounds are also suboptimal in this setting, as attested by a recent preprint of Coja-Oghlan and Lanka [29].

4.1 Concentration for the Erdös-Rényi graph

For $0 < p < 1$, the Erdös-Rényi graph $G_{n,p}$ [9, 7] is the special case of the model $\mathsf{G}_{\mathbf{p}}$ in Section 3 where $\mathbf{p}(i,j) = p$ for $i \neq j$ and $\mathbf{p}(i,i) = 0$ for $i = j$. Notice that in this case $A_{\mathbf{p}}^{\mathrm{typ}} = p(\mathbf{1}_n\mathbf{1}_n^* - I_n)$ where $\mathbf{1}_n \in \mathbb{R}^n$ is the all-ones vector and I_n is the $n \times n$ identity matrix. Moreover, $\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}} = I_n - \mathbf{1}_n\mathbf{1}_n^*/n$

The following result is immediate from Theorem 3.1

Proposition 4.1 *There exists $C > 0$ such that for all $n \in \mathbb{N}$, $n^{-2} < \delta < 1/2$ and $p \in (0,1)$ with $p(n-1) \geq C \ln n$, if $A_{n,p}$ be the adjacency matrix and $\mathcal{L}_{n,p}$ the Laplacian of the Erdös-Rényi graph $G_{n,p}$, then*

$$\mathbb{P}\left(\|A_{n,p} - p(\mathbf{1}_n\mathbf{1}_n^* - I_n)\| \leq 4\sqrt{p(n-1)\ln(n/\delta)}\right) \geq 1-\delta$$

$$\mathbb{P}\left(\|\mathcal{L}_{n,p} - (I_n - \mathbf{1}_n\mathbf{1}_n^*/n)\| \leq 14\sqrt{\frac{\ln(4n/\delta)}{p(n-1)}}\right) \geq 1-\delta$$

This result is qualitatively sharp in the sense that one cannot expect that the Laplacian concentrates when $pn \ll \ln n$. To see this, recall that the multiplicity of 0 in the spectrum of $\mathcal{L}_{n,p}$ is the number of connected components of $G_{n,p}$ (this is a deterministic statement; cf. [22]). If $pn \leq \ln n$, the probability of there being 2 or more components is bounded away from 0 [9].

But if 0 has multiplicity ≥ 2, (3.1) implies that $\|\mathcal{L}_{n,p} - (I_n - \mathbf{1}_n\mathbf{1}_n^*/n)\| \geq 1$, therefore $\mathcal{L}_{n,p}$ is far from the "typical Laplacian" with positive probability.

Quantitatively, the bounds in Proposition 4.1 can be improved. We quickly sketch the argument for the adjacency matrix, which is implicit in the work of Feige and Ofek [33]. A key idea is that, since the typical adjacency matrix $p(\mathbf{1}_n\mathbf{1}_n^* - I_n)$ has one very large eigenvalue and lots of small ones, the same should hold for $A_{n,p}$.

One can use the reasoning in [33, Lemma 2.1] to show that, for $pn = \Omega(\ln n)$ the dominant eigenvector of $A_{n,p}$ is always close to $\mathbf{1}_n/\sqrt{n}$. Moreover, the largest eigenvalue is $pn + O\left(\sqrt{pn}\right)$ and all other are of the order $O\left(\sqrt{pn}\right)$ [33, 38]. This shows that, with probability $\geq 1 - \delta/2$

$$\|A_{n,p} - p\mathbf{1}_n\mathbf{1}_n^*\| = O\left(\sqrt{pn}\right),$$

and this results in

$$\|A_{n,p} - p(\mathbf{1}_n\mathbf{1}_n^* - I_n)\| = O\left(\sqrt{pn}\right) \text{ with probability } \geq 1 - \delta$$

because $\|pI_n\| = O(1)$.

4.2 Quasi-randomness as concentration of the adjacency matrix

We now point out that the idea of concentration of the adjacency matrix is implicit in the theory of *dense quasi-random graphs*

This theory was initiated by Chung, Graham and Wilson [23]. Their surprising discovery was that several properties that a Erdös-Rényi random graph is very likely to have are in fact equivalent.

More precisely, let $\{G_m\}_{m\in\mathbb{N}}$ be a sequence of graphs, each G_m having n_m vertices and adjacency matrix A_m. Assume that $n_m \to +\infty$ when $m \to +\infty$ and that $p > 0$ is fixed. The following statements (among others) are equivalent [23, 43]. [The asymptotic notation refers to $m \to +\infty$.]

- **[Q1]** There exists a $s \geq 4$ such that for all $0 \leq k \leq \binom{s}{2}$, G_m contains more than $p^{-k}(1 - p)^{\binom{s}{2}-k} n_m^s$ induced labeled copies of each graph on s vertices and k edges.

- **[Q2]** G_m has $\geq (1 + o(1))pn_m^2/2$ edges and $\leq (1 + o(1))(pn_m)^4$ labeled copies of the four-cycle C_4.

- **[Q3]** G_m has $\geq (1 + o(1))pn_m^2/2$ edges, the largest eigenvalue of A_m is $(1 + o(1))pn$ and all other eigenvalues of A_m are $o(n)$ in absolute value.

- **[Q4]** $\max_{S\subset V_m} |e(S) - p|S|^2/2| = o\left(n_m^2\right)$ where $e(S)$ is the number of edges of G_m inside S and V_m is the vertex set of G_m.

We now provide a characterization of quasi-randomness in terms of "concentration" of the adjacency matrix. Let

$$A_m^{\rm typ} \equiv p\,(\mathbf{1}_{n_m}\mathbf{1}_{n_m}^* - I_{n_m}),$$

where $\mathbf{1}_{n_m} \in \mathbb{R}^{n_m}$ is (again) the all-ones vector and I_{n_m} is the $n_m \times n_m$ adjacency matrix. This is the same matrix that appears in Proposition 4.1.

The following result shows that a sequence of graphs is quasi-random if and only if the adjacency matrices of the graphs are sufficiently close to $A_m^{\rm typ}$.

Proposition 4.2 *A sequence $\{G_m\}_m$ of graphs as above satisfies properties* **[Q1]**-**[Q4]** *above if and only if:*

$$[\mathbf{P1}]\|A_m - A_m^{\rm typ}\| = o\,(n)\,.$$

Proof: [of Proposition 4.2] We will show that **[P1]** is equivalent to **[Q3]** in the previous list.

[P1]$\Rightarrow$**[Q3]** : The eigenvalues of $A_m^{\rm typ}$ are $p(n_m - 1)$ (with multiplicity 1) and $-p$ (with multiplicity $n_m - 1$).

We use inequality (3.1) above to deduce that:

$$|\lambda_{n-1}(A_m) - pn_m| = |\lambda_{n-1}(A_m) - \lambda_{n-1}(A_m^{\rm typ})| + O\,(1) = o\,(n_m)$$

and for $0 \leq i \leq n_m - 2$:

$$|\lambda_i(A_m)| = |\lambda_i(A_m) + p| + O\,(1) = |\lambda_i(A_m) - \lambda_i(A_m^{\rm typ})| + O\,(1) = o\,(n_m)\,.$$

Moreover, the number of edges in G_m is:

$$\begin{aligned}\frac{1}{2}\,\mathbf{1}_{n_m}^* A_m \mathbf{1}_{n_m} &\geq \mathbf{1}_{n_m}^* A_m^{\rm typ}\mathbf{1}_{n_m} - \frac{1}{2}\,\mathbf{1}^*(A_m - A_m^{\rm typ})\,\mathbf{1} \\ &= \frac{pn_m(n_m-1)}{2} - \|\mathbf{1}_{n_m}\|^2\|A_m - A_{K_{n_m}}\| \\ &= \frac{pn_m^2}{2} - o\left(n_m^2\right).\end{aligned}$$

[Q3]$\Rightarrow$**[P1]**: It is immediate from **[Q3]** that A_m is $o\,(n)$-close to a rank-one operator: if $\psi_{\max}$ is the (normalized) eigenvector corresponding to the largest eigenvalue $\lambda_{\max}(A_m)$, then:

$$\|A_m - \lambda_{\max}(A_m)\,\psi_{\max}\psi_{\max}^*\| = \max_{0\leq i\leq n_m-1}|\lambda_i(A_m)| = o\,(n_m)\,.$$

By **[Q3]** we also know that:

$$\|\lambda_{\max}(A_m)\,\psi_{\max}\psi_{\max}^* - pn_m\,\psi_{\max}\psi_{\max}^*\| = |\lambda_{\max}(A_m) - pn_m| = o\,(n)\,.$$

It is shown in the proof of Fact 7 in [23] that, under **[Q3]**, $\psi_{\max}$ is $o(1)$-close to $\mathbf{1}_{n_m}/\sqrt{n_m}$. Thus we see that:

$$\|pn_m\,\psi_{\max}\psi_{\max}^* - p\mathbf{1}_{n_m}\,\mathbf{1}_{n_m}^*\| = o(n_m)\,.$$

Finally, we notice that

$$A_m^{\mathrm{typ}} = p\mathbf{1}_{n_m}\,\mathbf{1}_{n_m}^* - pI,$$

hence

$$\|A_m^{\mathrm{typ}} - p\mathbf{1}_{n_m}\,\mathbf{1}_{n_m}^*\| = O(1)\,.$$

Putting all the inequalities together implies the desired result. □

5 Application to bond percolation

In the previous section we discussed a random graph model where the typical Laplacian and adjacency matrices had one "special" eigenvalue with multiplicity 1 and $n-1$ "trivial" eigenvalues. In this setting, proving concentration of the adjacency matrix (say) essentially amounted to showing that one eigenvector was close to what it should be while the other eigenvalues clustered around the degenerate eigenvalue of the typical case.

We now consider a class of models for which one cannot expect this strategy to work. Let $p \in (0,1)$ and $G = (V,E)$ be an arbitrary unweighted graph on vertex set $V = [n]$. Consider the random subgraph G_p of G that is obtained via by deleting each edge of G independently with probability $1-p$. This model of *bond percolation* has received much attention in recent years, with a special focus the emergence of a giant component [10, 37, 3, 27, 52, 15]. Much less seems to be known about the spectrum of G_p [26].

In this section we apply our general Theorem, Theorem 3.1, in order to answer the following question: how large does p need to be in order for the graph matrices to concentrate? Clearly, this must occur way after the percolation threshold.

Bond percolation is a special case of the random model $\mathsf{G}_{\mathbf{p}}$ in Section 3. To see this, one only needs to define:

$$\mathbf{p}(i,j) = \begin{cases} p & \text{if } ij \in E, \\ 0 & \text{if not.} \end{cases}$$

A computation shows that the "typical matrices" for this choice of $\mathbf{p}$ are:

$$A_{\mathbf{p}}^{\mathrm{typ}} = pA_G,\ \text{where } A_G \text{ is the adjacency matrix of } G;$$

$$\mathcal{L}_{\mathbf{p}}^{\mathrm{typ}} = \mathcal{L}_G,\ \text{where } \mathcal{L}_G \text{ is the Laplacian of } G.$$

Moreover, the parameters d, Δ appearing in Theorem 3.1 are pd_G and $p\Delta_G$, where d_G (resp. Δ_G) is the minimum (resp. maximal) degree in G.

The following result is a direct corollary of Theorem 3.1.

Theorem 5.1 *For each $c > 0$ there exists a $C > 0$ such that the following holds. Suppose that G, p and G_p are as above and $pd_G \geq C \ln n$. Then:*

$$\mathbb{P}\left(\|A_{G_p} - pA_G\| \leq 4\sqrt{p\Delta_G \ln(n/\delta)}\right) \geq 1 - \delta$$

and

$$\mathbb{P}\left(\|\mathcal{L}_{G_p} - \mathcal{L}_G\| \leq 14\sqrt{\frac{\ln(4n/\delta)}{pd_G}}\right) \geq 1 - \delta,$$

where A_{G_p} and $\mathcal{L}_{G_p}$ are the adjacency matrix and Laplacian of G_p (resp.)

One can of course derive corollaries about eigenvectors and eigenvectors following Corollaries 3.1 and 3.2. For instance, suppose that:

$$\gamma > 14\sqrt{\frac{\ln(4n/\delta)}{pd_G}}$$

Then the following holds with probability $1 - \delta$: for each $0 \leq i \leq n-1$ such that the interval $(\lambda_i(\mathcal{L}_G) - 2\gamma, \lambda_i(\mathcal{L}_G) + 2\gamma)$ contains no eigenvalues of $\mathcal{L}_G$ other than $\lambda_i(\mathcal{L})$, $\lambda_i(\mathcal{L}_{G_p})$ has multiplicity 1 in the spectrum of $\mathcal{L}_{G_p}$ and moreover, the corresponding normalized eigenvectors ψ, ψ_p of $\mathcal{L}_G$ and $\mathcal{L}_{G_p}$ (resp.) satisfy:

$$\|\psi_p\psi_p^* - \psi\psi^*\| \leq \frac{4}{\pi}\,\frac{\sqrt{\frac{\ln(4n/\delta)}{pd_G}}}{\gamma - \sqrt{\frac{\ln(4n/\delta)}{pd_G}}}$$

with probability $\geq 1 - \delta$. This implies:

$$1 - (\psi^*\psi_p)^2 \leq \frac{4}{\pi}\,\frac{\sqrt{\frac{\ln(4n/\delta)}{pd_G}}}{\gamma - \sqrt{\frac{\ln(4n/\delta)}{pd_G}}}$$

for the same eigenvectors, which implies that ψ_p is close to ψ or $-\psi$. A similar result for the eigenspace projectors could be derived even if $\lambda_i(G)$ had higher multiplicity. It seems quite remarkable that one can approximately obtain the eigenvectors or eigenspaces of G from a (potentially very sparse) subgraph G_p.

We also note that the threshold for Laplacian concentration is indeed $pd_G = \Theta(\ln n)$, as shown in Section 4.1 in the special case of the Erdös-Rényi random graph $G_{n,p}$.

The following simple corollary is also of interest.

Corollary 5.1 *There exist $C, C' > 0$ such that, if $pd_G \geq C \ln n$, then with probability $1 - 1/n^2$,*

$$|\lambda(G) - \lambda(G_p)| \leq C' \sqrt{\frac{\ln n}{pd_G}}.$$

We have singled out this bound in order to compare it with a recent bound of Chung and Horn [26]. These authors proved that, with high probability,

$$\lambda(G_p) \geq \lambda(G) - O\left(\sqrt{\frac{\ln n}{pd_G}} + \frac{(\ln n)^{3/2}}{pd_G(\ln \ln n)^{3/2}}\right).$$

Our bound is better for all values of n and pd_G, most dramatically for $\ln n \ll pd_G \ll \ln^{3/2-\epsilon} n$, in which case their bound is vacuous while ours is non-trivial.

6 Application to inhomogeneous random graphs

In this section we consider a more complex random graph model that is defined in terms of an *attachment kernel* κ, a *density parameter* $0 < p < 1$ and a set of points $X_1, X_2, \ldots, X_n$.

More precisely, let $\kappa : [0,1]^2 \to \mathbb{R}_+ \cup \{0\}$ be a measurable function that is symmetric in the sense that for all $x, y \in [0,1]$, $\kappa(x,y) = \kappa(y,x)$. Pick some vector $X_{1:n} \equiv (X_1, \ldots, X_n)$ of points in $[0,1]$. Now consider the following weight function $\mathbf{p} : [n]^2 \to [0,1]$:

$$\mathbf{p}(i,j) \equiv \max\{p\kappa(X_i, X_j), 1\}, \ (i,j) \in [n]^2. \tag{6.1}$$

One can define a random graph $\mathsf{G}_\mathbf{p}$ as in Section 3 with the above weight function; we call this graph $G_{n,p,\kappa}$, the *inhomogeneous random graph* on n vertices, density parameter p and attachment kernel κ (the dependency on $X_{1:n}$ is implicit in this nomenclature). The adjacency matrix of this random graph will be denoted by $A_{n,p,\kappa}$

Our goal in this section will be to prove that, up to some error terms that are small with high probability, the adjacency matrix $A_{n,\kappa,p}/pn$ of $G_{n,p,\kappa}$ will be related to the integral operator on $L^2([0,1])$ that is defined by κ.

$$\begin{array}{rcccl} T_\kappa & : & L^2([0,1]) & \to & L^2([0,1]) \\ & & f(\cdot) & \mapsto & \int_0^1 \kappa(\cdot, y)\, f(y)\, dy \end{array} \tag{6.2}$$

Similar results for the Laplacian of $G_{n,p,\kappa}$ are discussed in Section 8.

6.1 Some history of the model

The phrase "inhomogeneous random graph" comes from a paper by Bollobas, Janson and Riordan [11] where the above model was studied in the range $p = \Theta(1/n)$ with background spaces more general than $[0,1]$. Their goal was to study the structure of connected components in the general model, in analogy with the well-known Erdös-Rényi phase transition at $p = 1/n$ [7].

A related random graph model generating dense graphs ($p = 1$) was introduced in [49] and studied in [16]. This model is related to the beautiful theory of *graph limits* where the space of graphs is "completed" into the space of *graphons*, which are non-negative, symmetric functions like κ above, with the further restriction that $\kappa \leq 1$. There is a fairly complete correspondence between the properties of sequences of graphs that are *convergent* in terms of normalized subgraph counts and the corresponding limiting graphon. Conversely, the sequence of random graphs correponding to a given graphon κ converges to that same graphon. The *cut metric* that defines graph convergence will be further discussed in Section 6.3 below.

The connection between the convergent graph sequences and inhomogeneous random graphs was noted in [10, 12], where the authors studied bond percolation over a convergent sequence of graphs and found the critical probability for existence of a giant component. Other papers [13, 14] have focused on the relationship between convergence of subgraph counts vs. convergence in the *cut metric* (see below) for sparse graphs, a topic that is far from completely elucidated. In what follows we will show that our random graphs converge to the corresponding kernel in a stronger metric.

6.2 The precise result

We will use the following technical assumption.

Assumption 6.1 $\kappa : [0,1]^2 \to \mathbb{R}_+ \cup \{0\}$ *is a symmetric measurable function with*

$$K \equiv \sup_{(x,y)\in[0,1]^2} \kappa(x,y) < +\infty.$$

Moreover, the points $X_1, X_2, \dots, X_n$ are random i.i.d. uniform over $[0,1]$.

Let $\overline{X}_1 \leq \overline{X}_2 \leq \cdots \leq \overline{X}_n$ be the ordered sequence of the X_i; ie. $\overline{X}_1$ is the minimum of the X_i, $\overline{X}_2$ is the second smallest element and so on (ties are broken arbitrarily). Let σ_n be a permutation such that $X_i = \overline{X}_{\sigma_n(i)}$ for each $1 \leq i \leq n$, chosen in a measurable manner. Associate with the graph $G_{n,p,\kappa}$ a symmetric, non-negative function from $[0,1]^2$ to $\mathbb{R}_+ \cup \{0\}$:

$$\mathcal{G}_{n,p,\kappa} \equiv \frac{1}{p} \sum_{ij\in E(G_{n,p,\kappa})} \chi_{\left(\frac{\sigma_n(i)-1}{n},\frac{\sigma_n(i)}{n}\right]\times\left(\frac{\sigma_n(j)-1}{n},\frac{\sigma_n(j)}{n}\right]},$$

where χ_S is the indicator function of the set S and $E(G_{n,p,\kappa})$ is the edge set of $G_{n,p,\kappa}$. Notice that $\mathcal{G}_{n,p,\kappa}$ defines a bounded linear operator on $L^2([0,1])$ via a formula similar to (6.2):

$$(T_{\mathcal{G}_{p,n,\kappa}} f)(\cdot) \equiv \int_0^1 \mathcal{G}_{p,n,\kappa}(\cdot, y)\, f(y)\, dy \quad (f \in L^2([0,1])).$$

Let $\{\mathbf{e}_i\}_{i=1}^n$ be the canonical basis of $\mathbb{R}^n$. Let us consider two linear operators (both of which depend on σ_n defined previously:

$$\begin{array}{rclcl} H_n & : & \mathbb{R}^n & \to & L^2([0,1]) \\ & & \psi = \sum_{i=1}^n \psi(i)\,\mathbf{e}_i & \mapsto & \sum_{i=1}^n \sqrt{n}\,\psi(i)\,\chi_{\left(\frac{\sigma_n(i)-1}{n}, \frac{\sigma_n(i)}{n}\right]} \end{array}$$

$$\begin{array}{rclcl} E_n & : & L^2([0,1]) & \to & \mathbb{R}^n \\ & & f & \mapsto & \sum_{i=1}^n \left(\sqrt{n} \int_{\frac{\sigma_n(i)-1}{n}}^{\frac{\sigma_n(i)}{n}} f\, dx\right) \mathbf{e}_i \end{array}$$

and note that $T_{\mathcal{G}_{n,p,\kappa}} = H_n A_{n,p,\kappa} E_n / pn$.

Finally, let $\mathrm{spec}(T_\kappa)$ be the *spectrum* of the operator T_κ in (6.2) (see Section 2.2 to recall what the spectrum is).

Theorem 6.1 (proven in Section 6.4) *There exist universal constants $c, C > 0$ such that the following holds under Assumption 6.1. Given $\epsilon > 0$, suppose there exists a L-Lipschitz function κ_ϵ that also takes values in $[0, K]$ and which is ϵ-close to κ in the $L^2([0,1]^2)$ norm. Define:*

$$\theta = \theta(\kappa, \epsilon, L, K, n, p) \equiv 2\epsilon + c(L+K)\left(\frac{\ln n}{n}\right)^{1/4} + \sqrt{\frac{K \ln n}{pn}},$$

and assume $pn \geq C \ln n$ and $p \leq 1/K$. Then there exists an event $\mathcal{E}$ with probability $\mathbb{P}(\mathcal{E}) \geq 1 - n^{-2}$ such that, inside $\mathcal{E}$, the following properties hold:

1. *The $n \times n$ matrices $A_{n,p,\kappa}$ and $E_n T_\kappa H_n$ satisfy:*

$$\left\| \frac{A_{n,p,\kappa}}{pn} - E_n T_\kappa H_n \right\| \leq \theta;$$

2. *The integral operators $T_{\mathcal{G}_{n,p,\kappa}}$ and T_κ satisfy:*

$$\left\| T_{\mathcal{G}_{n,p,\kappa}} - T_\kappa \right\|_{2\to 2} \leq \theta;$$

3. *Given $S \subset \mathbb{R}$, let $m_{A_{n,\kappa,p}/pn}(S)$ be the sum of the multiplicities of all eigenvalues of $A_{n,\kappa,p}$*

that lie in S and define $m_{T_\kappa}(S)$ similarly. Then if $\inf_{s\in S}|s|>\theta$,

$$m_{A_{n,\kappa,p}/pn}(S)\leq m_{T_\kappa}(S^\theta) \text{ and } m_{T_\kappa}(S)\leq m_{A_{n,\kappa,p}/pn}(S^\theta)$$

where $S^\theta\equiv\{x\in\mathbb{R}\,:\,\exists s\in S,\,|x-s|\leq\theta\}$.

4. *Consider each pair (α,γ) where $\alpha\in\mathrm{spec}(T_\kappa)$ and $\gamma>\theta$ is such that $(\alpha-2\gamma,\alpha+2\gamma)$ contains no eigenvalue of T_κ other than α itself. Let P_α be the orthogonal projection in $L^2([0,1])$ onto the eigenspace of α in $L^2([0,1])$ and consider the orthogonal projection $\Pi_{(\alpha-\gamma)pn,(\alpha+\gamma)pn}(A_{n,p,\kappa})$ in $\mathbb{C}^n$ over the span of the eigenvectors of $A_{n,p,\kappa}$ corresponding to eigenvalues in $[(\alpha-\gamma)pn,(\alpha+\gamma)pn]$. Then:*

$$\|\Pi_{(\alpha-\gamma)pn,(\alpha+\gamma)pn}(A_{n,p,\kappa})-E_nP_\alpha H_n\|\leq\frac{4\theta}{\pi(\gamma-\theta)}.$$

This Theorem implies that, up to error terms that are small with high probability, $A_{n,p,\kappa}/pn$ is defined solely in terms of the kernel function κ, up to a permutation of coordinates. It also implies that, statistical parlance, it implies that the non-zero eigenvalues of $A_{n,p,\kappa}$ are strongly consistent estimators of the non-zero eigenvalues of T_κ when $n\to+\infty$ and $p=p(n)$ $pn/\ln n\to+\infty$.

Both of these assertions hinge on the fact that Lipschitz functions are dense in $L^2([0,1]^2)$. Unfortunately, our error bounds are not independent of κ, as quality of the approximation by Lipschitz functions, measured by the size of the Lipschitz constant for a given approximation error ϵ, may vary with κ. This is in contrast with approximation in the *cut norm*, which we now discuss.

6.3 Convergence in the operator and cut metrics

6.3.1 The cut norm and the cut metric

Any function $\eta\in L^1([0,1]^2)$ determines a bounded linear operator $\tilde{T}_\eta:L^\infty([0,1])\to L^1([0,1])$ via the formula that we already used to define T_κ and $T_{\mathcal{G}_{p,n,\kappa}}$:

$$\tilde{T}_\eta f(\cdot)\equiv\int_0^1\eta(\cdot,y)\,f(y)\,dy. \tag{6.3}$$

The cut norm of η is the $L^\infty\to L^1$ norm of $\tilde{T}_\eta$:

$$\|\eta\|_{\mathrm{cut}}\equiv\sup\left\{\left|\int_0^1(T_\eta f)(x)g(y)\,dx\right|\,:\,f,g\in L^\infty([0,1]),\|f\|_{L^\infty}\leq1,\|g\|_{L^\infty}\leq1\right\}. \tag{6.4}$$

One can check that $\|\eta\|_{\text{cut}} \leq \|\eta\|_{L^1}$ always. This definition of $\|\eta\|_{\text{cut}}$ is natural from the point of view of Functional Analysis; a more "combinatorial" definition,

$$\|\eta\|_{\text{cut},2} \equiv \sup\left\{\left|\int_{A\times B} \eta(x,y)\,dx\,dy\right| \;:\; A, B \subset [0,1] \text{ measurable}\right\}$$

is equivalent to the previous one in the sense that:

$$\frac{1}{4}\|\eta\|_{\text{cut}} \leq \|\eta\|_{\text{cut},2} \leq \|\eta\|_{\text{cut}}.$$

Now assume that G_1 and G_2 are graphs with common vertex set $[n]$ and adjacency matrices A_{G_1}, A_{G_2}. Define:

$$\kappa_{G_i,p} \equiv \sum_{1\leq i,j\leq n\,:\,ih\in E(G_i)} \chi_{\left(\frac{i-1}{n},\frac{i}{n}\right]\times\left(\frac{j-1}{n},\frac{j}{n}\right]}.$$

Then one sees that:

$$\|\kappa_{G_1,p} - \kappa_{G_2,p}\|_{\text{cut},2} = \frac{\max_{S,V\subset[n]}\left|\sum_{(i,j)\in S\times W}(A_{G_1}(i,j) - A_{G_2}(i,j))\right|}{pn^2}$$

is the normalized *cut norm* of $A_G - A_H$ [49].

Thus the cut norm on $L^1([0,1]^2)$ induces a distance on graphs. Notice, however, that this distance might be positive even though G and H are isomorphic. This motivates the following definition: given two kernels $\kappa, \kappa'' \in L^1([0,1]^2)$, say that κ'' is a rearrangement of κ ($\kappa'' \approx \kappa$) is there exists a measure-preserving bijection $\tau : [0,1] \to [0,1]$ such that $\kappa(x,y) = \kappa(\tau(x),\tau(y))$ for almost every $(x,y) \in [0,1]^2$. The *cut metric* assigng to each pair κ, κ' of kernels a distance:

$$d_{\text{cut}}(\kappa,\kappa') \equiv \inf\{\|\kappa'' - \kappa'\|_{\text{cut}} \;:\; \kappa'' \approx \kappa\}.$$

Notice that the cut metric does not distinguish between (the kernels of) isomorphic graphs.

6.3.2 The operator norm and the operator metric

The metric d_{cut} yields a criterion for convergence of graph sequences. In the dense case $p = \Theta(1)$, this implies the convergence of normalized subgraph counts and also gives a criterion for testable graph properties [49, 16]. As mentioned above, much less is understood about the case $p = o(1)$ (see however the conjectures of Bollobás and Riordan [13, Section 5.2]).

Theorem 6.1 is mostly concerned with the eigenvalues and the eigenvectors of the adjacency matrix $A_{n,p,\kappa}$. Unfortunately, in general we do not even know how to control the eigenvalues of $A_{n,p,\kappa}$ in terms of the cut norm alone. For *bounded* kernels ($p = \Theta(1)$), this is easy enough (see [17, Theorem 6.6]), but there are difficulties in extending this to the sparse case. This does seem to be a serious problem, as related difficulties appear in [13] when the authors attempt to relate the convergence of subgraph counts to cut metric convergence. [Estimating the eigenvalues is related to counting cycles in the corresponding graph or graphon.]

Luckily, a stronger notion of convergence implied by the $L^2 \to L^2$ norm suffices for our purpose, and it is precisely this notion that we achieve via our methods.

We need some definitions in order to properly state this. Given $\eta \in L^2([0,1]^2)$, define a bounded linear operator T_η from $L^2([0,1])$ to itself via the formula in Section 2.2; this is the same as (6.3), except that the domain and range of $\tilde{T}_\eta$ are different. The operator or "$L^2 \to L^2$" norm of η is the $L^2 \to L^2$ norm of T_η, also defined in Section 2.2:

$$\|\eta\|_{\rm op} \equiv \|T_\eta\|_{L^2 \to L^2}.$$

From (6.4) we see that that $\|\eta\|_{\rm op} \geq \|\eta\|_{\rm cut}$ whenever η is square-integrable.

In analogy with the cut metric, one can also define an *operator (pseudo-)metric* on square-integrable kernels via the formula:

$$d_{\rm op}(\kappa, \kappa') \equiv \inf\{\|\kappa'' - \kappa'\|_{\rm op} \,:\, \kappa'' \approx \kappa\}.$$

One can show via our results that when Assumption 6.1 holds, $n \gg 1$ and $p \gg \ln n/n$, then the kernel determined by $G_{n,p,\kappa}$ – which is equivalent to $\mathcal{G}_{n,p,\kappa}$ in Theorem 6.1 – converges in the $d_{\rm op}$ metric to κ. We omit the details.

A drawback of $d_{\rm op}$ is that it lacks a corresponding (weak or strong) regularity lemma, which would allow one to approximate up to error ϵ any (say bounded) kernel κ by simple functions taking at most $m = m(\epsilon, \|\kappa\|_{L^\infty})$ values. Indeed, this is precisely why the bound in Theorem 6.1 depends on κ.

6.4 Proof of Theorem 6.1

The proof will consist of several steps.

6.4.1 The relationship between $T_{\mathcal{G}_{n,p,\kappa}}$ and $A_{\mathcal{G}_{n,p,\kappa}}$

For $f, g \in L^2([0,1])$, define $(f,g)_{L^2} \equiv \int_0^1 f(x)g(x)\,dx$ and $\|f\|_{L^2}^2 \equiv (f,f)_{L^2}$.

The following facts can be easily checked (proof omitted).

$$\forall f \in L^2([0,1]),\ \forall \psi \in \mathbb{R}^n,\ (f, H_n\psi)_{L^2} = (E_n f)^*\psi \text{ (i.e. } E_n \text{ is the adjoint of } H_n); \tag{6.5}$$

$$\forall \psi, \phi \in \mathbb{R}^n,\ (H_n\psi, H_n\phi)_{L^2} = \psi^*\phi \text{ (i.e. } H_n \text{ is an isometry)'} \tag{6.6}$$

$$\forall f \in L^2,\ \|E_n f\| \leq \|f\|_{L^2} \text{ (i.e. } E_n \text{ has operator norm at most 1);} \tag{6.7}$$

$$E_n H_n = I_n, \text{ the identity operator on } \mathbb{R}^n; \tag{6.8}$$

$$H_n E_n = \Pi_n, \text{ the projection onto the span of } \left\{ \left(\frac{i-1}{n}, \frac{i}{n} \right] \right\}_{i=1}^{n}; \text{ and} \tag{6.9}$$

$$T_{\mathcal{G}_{n,p,\kappa}} = \frac{1}{pn} H_n\, A_{n,p,\kappa}\, E_n, \text{ as seen above.} \tag{6.10}$$

Let us now relate the non-zero eigenvalues and eigenvectors of $A_{n,p,\kappa}$ with those of $T_{\mathcal{G}_{n,p,\kappa}}$. Write:

$$A_{n,p,\kappa} = \sum_{\alpha : \alpha pn \in \mathrm{spec}(A_{n,p,\kappa})} (\alpha pn)\, \Pi_\alpha$$

where each Π_α the projection onto the eigenspace corresponding to αpn. By (6.10),

$$T_{\mathcal{G}_{n,p,\kappa}} = \sum_{\alpha : \alpha pn \in \mathrm{spec}(A_{n,p,\kappa})} \alpha\, H_n\, \Pi_\alpha E_n.$$

Claim 6.1 *The operators $H_n\, \Pi_\alpha E_n$ are orthogonal projections with orthogonal ranges. Therefore, the non-zero eigenvalues of $T_{\mathcal{G}_{n,p,\kappa}}$ are the numbers $\alpha \neq 0$ with $\alpha pn \in \mathrm{spec}(A_{n,p,\kappa})$. Moreover, for each such α, $H_n\, \Pi_\alpha E_n$ is the projection onto the corresponding eigenspace of $T_{\mathcal{G}_{n,p,\kappa}}$.*

Proof: [of the Claim] First notice that for each α:

$$(H_n\, \Pi_\alpha E_n)^2 = H_n\, \Pi_\alpha E_n H_n \Pi_\alpha E_n = H_n\, \Pi_\alpha E_n.$$

because $E_n H_n = I_n$ (eqn. (6.8)) and $\Pi_\alpha^2 = \Pi_\alpha$. One can also check that for all $f, g \in L^2([0,1])$,

$$(f, H_n\, \Pi_\alpha E_n g)_{L^2} = (H_n f)^* (\Pi_\alpha E_n g) = (\Pi_\alpha H_n f)^* (E_n g) = (H_n\, \Pi_\alpha E_n f, g)_{L^2},$$

where we used (6.5) for the first and third equalities and the fact that $\Pi_\alpha = \Pi_\alpha^*$ for the second one. It follows that $H_n\, \Pi_\alpha E_n$ is a self-adjoint operator on L^2 that equals its square; this means that it is an orthogonal projection onto its range.

To see that these ranges are orthogonal for distinct α, notice that the range of $H_n\, \Pi_\alpha E_n$ is the set of all vectors of the form $H_n \psi$ where ψ belongs to the range of Π_α and is therefore an eigenvector of $A_{n,p,\kappa}$ with eigenvalue αpn. But eigenvectors of $A_{n,p,\kappa}$ with distinct eigenvalues are orthogonal, hence their images under H_n are orthogonal in L^2 (by (6.6)).

The other assertions follow directly. □

6.4.2 The concentration argument

Let us introduce a matrix $\overline{A}_{n,p,\kappa}$ whose (i,j)-th entry is $p\kappa(X_i, X_j)$, $1 \leq i \leq j \leq n$. Conditioning on the realization of the $X_1, \ldots, X_j$, our random graph model has independent edges with respective probabilities $\mathbf{p}(i,j) = p\kappa(X_i, X_j)$ and $\overline{A}_{n,p,\kappa}$ is precisely the typical adjacency matrix $A^{\mathrm{typ}}_{\mathbf{p}}$ in this setting. We deduce from Theorem 3.1 that there exists a constant $C > 0$ independent of n, κ and $X_1, \dots, X_n$, such that if $\Delta = \Delta(X_1, \dots, X_n)$ is as in that Theorem and $\Delta \geq C \ln n$,

$$\mathbb{P}\left(\|A_{n,p,\kappa} - \overline{A}_{n,p,\kappa}\| \geq 4\sqrt{\Delta \ln(2n^2)} \mid X_1, \dots, X_n\right) \leq \frac{1}{2n^2},$$

In our setting we always have

$$\Delta = \max_{1 \leq i \leq n} \sum_{j=1}^{n} p\kappa(X_i, X_j) \leq Kpn$$

where K is the quantity in Assumption 6.1. Therefore,

$$\mathbb{P}\left(\|A_{n,p,\kappa} - \overline{A}_{n,p,\kappa}\| \geq 4\sqrt{Kpn \ln(2n^2)}\right) \leq \frac{1}{2n^2}.$$

Let

$$\overline{T} \equiv H_n \overline{A}_{n,p,\kappa} E_n / pn = \sum_{1 \leq i,j \leq n} \kappa(X_i, X_j) \chi_{\left(\frac{\sigma_n(i)-1}{n}, \frac{\sigma_n(i)}{n}\right] \times \left(\frac{\sigma_n(j)-1}{n}, \frac{\sigma_n(j)}{n}\right]} \tag{6.11}$$

Since H_n is an isometry (by (6.6)) and E_n has norm at most 1 (by (6.7)),

$$\|\overline{T} - T_{\mathcal{G}_{n,p,\kappa}}\|_{L^2 \to L^2} = \frac{1}{pn} \|H_n(\overline{A}_{n,p,\kappa} - A_{n,p,\kappa})E_n\| \leq 4\sqrt{\frac{K \ln(2n^2)}{pn}}$$

with probability $\geq 1 - 1/2n^2$.

6.4.3 Nearing the end of the argument

We will show in Lemma 6.1 below that there exists a universal $c > 0$ such that for any $\epsilon > 0$

$$\mathbb{P}\left(\|\overline{T} - T_\kappa\| \leq 2\epsilon + c(L+K)(\ln n/n)^{1/4}\right) \geq 1 - \frac{1}{2n^2}. \tag{6.12}$$

Increasing c if necessary, this implies that, with probability $\geq 1 - n^{-2}$

$$\|T_{\mathcal{G}_{n,p,\kappa}} - T_\kappa\|_{L^2 \to L^2} \leq \|T_{\mathcal{G}_{n,p,\kappa}} - \overline{T}\|_{L^2 \to L^2} + \|\overline{T} - T_\kappa\|_{L^2 \to L^2} \leq \theta$$

for θ as in the Theorem. This proves the second assertion in the Theorem. To prove the first one, first notice that, since $E_n H_n = I_n$ (cf. (6.8)),

$$E_n T_{\mathcal{G}_{n,p,\kappa}} H_n = \frac{1}{pn} (E_n H_n) A_{n,p,\kappa} (E_n H_n) = \frac{A_{n,p,\kappa}}{pn}.$$

Now use again the fact that E_n and H_n have norm 1 to deduce:

$$\|\frac{A_{n,p,\kappa}}{pn} - E_n T_\kappa H_n\| \leq \|T_{\mathcal{G}_{n,p,\kappa}} - T_\kappa\|_{L^2\to L^2} \leq \theta.$$

The other two assertions follow from the perturbation lemmas provided in the Appendix. More precisely, recall from Claim 6.1 that the eigenvalues of $T_{\mathcal{G}_{n,p,\kappa}}$ are either 0 or equal to some $\alpha \neq 0$ with $\alpha pn \in \mathrm{spec}(A_{n,\kappa,p})$. Assertion 3 follows from Lemma A.1 applied to $T_{\mathcal{G}_{n,p,\kappa}}$ and T_κ.

As for Assertion 4, we recall from Claim 6.1 that whenever $\beta pn \in \mathrm{spec}(A_{n,p,\kappa})$ with corresponding eigenspace projection Π_β the corresponding eigenspace of $T_{\mathcal{G}_{n,\kappa,p}}$ is $H_n \Pi_\beta E_n$. This implies that:

$$H_n \Pi_{(\alpha-\gamma)pn,(\alpha+\gamma)pn}(A_{n,p,\kappa}) E_n$$

is the projection onto the eigenspaces of $T_{\mathcal{G}_{n,p,\kappa}}$ corresponding to eigenvalues between $\alpha-\gamma$ and $\alpha+\gamma$. One can apply Lemma A.2 with $\epsilon = \theta$ and $b - \gamma = a + \gamma = \alpha$ to deduce that, whenever α is as in assertion 4 and $\|T_{\mathcal{G}_{n,p,\kappa}} - T_\kappa\| \leq \theta$,

$$\|H_n \Pi_{(\alpha-\gamma)pn,(\alpha+\gamma)pn}(A_{n,p,\kappa}) E_n - P_\alpha\|_{L^2\to L^2} \leq \frac{4\theta}{\pi(\gamma-\theta)}.$$

Multiplying both operators above by E_n on the left and by H_n on the right, using that H_n and E_n have norm ≤ 1 and that $E_n H_n = I_n$, we see that:

$$\|\Pi_{(\alpha-\gamma)pn,(\alpha+\gamma)pn}(A_{n,p,\kappa}) - E_n P_\alpha H_n\| \leq \frac{4\theta}{\pi(\gamma-\theta)}.$$

This finishes the proof modulo inequality (6.12), which is the subject of Lemma 6.1 below.

6.4.4 Approximating T_κ

Lemma 6.1 *Under Assumption 6.1, suppose $\epsilon > 0$ is given and $\kappa_\epsilon : [0,1]^2 \to \mathbb{R}_+ \cup \{0\}$ is a L-Lipschitz symmetric function, with values between* 0 *and K, such that*

$$\int_0^1 \int_0^1 (\kappa(x,y) - \kappa_\epsilon(x,y))^2 \, dx \, dy \leq \epsilon^2.$$

Then the following holds with probability $\geq 1/2n^2$:

$$\|T_\kappa - \overline{T}\|_{L^2\to L^2} \leq 2\epsilon + c(L+K)\left(\frac{\ln n}{n}\right)^{1/4},$$

where $c > 0$ is universal.

Proof: Define:

$$\widehat{T} \equiv \sum_{1 \le i,j \le n} \kappa_\epsilon(X_i, X_j) \chi_{\left(\frac{\sigma_n(i)-1}{n}, \frac{\sigma_n(i)}{n}\right] \times \left(\frac{\sigma_n(j)-1}{n}, \frac{\sigma_n(j)}{n}\right]}.$$

We will bound:

$$\|T_\kappa - \overline{T}\|_{L^2 \to L^2} \le \|T_\kappa - T_{\kappa_\epsilon}\|_{L^2 \to L^2} + \|\overline{T} - \widehat{T}\|_{L^2 \to L^2} + \|T_{\kappa_\epsilon} - \widehat{T}\|_{L^2 \to L^2}. \tag{6.13}$$

By the results in Section 2.2, one can bound the first term in the RHS by:

$$\|T_\kappa - T_{\kappa_\epsilon}\|^2_{L^2 \to L^2} = \|T_{\kappa - \kappa_\epsilon}\|^2_{L^2 \to L^2} \le \int_0^1 \int_0^1 (\kappa(x,y) - \kappa_\epsilon(x,y))^2 \, dx \, dy \le \epsilon^2.$$

For the second term, we observe that $\overline{T} - \widehat{T}$ is of the form T_η for η taking the values $\kappa(X_i, X_j) - \kappa_\epsilon(X_i, X_j)$ on squares of area $1/n^2$. We deduce from the results in Section 2.2 that:

$$\|\overline{T} - \widehat{T}\|^2_{L^2 \to L^2} \le \frac{1}{n^2} \sum_{i,j=1}^n (\kappa(X_i, X_j) - \kappa_\epsilon(X_i, X_j))^2. \tag{6.14}$$

The expected value of the RHS is:

$$\int_0^1 \int_0^1 (\kappa(x,y) - \kappa_\epsilon(x,y))^2 \, dx \, dy \le \epsilon^2$$

Moreover, the random variables X_i are independent and replacing X_i by some other $X_i' \in [0,1]$ can change the value of the sum in the RHS of (6.14) by at most K^2/n (as each term is bounded by K and only n terms involve X_i). Azuma's inequality [7] implies:

$$\mathbb{P}\left(\frac{1}{n^2} \sum_{i,j=1}^n (\kappa(X_i, X_j) - \kappa_\epsilon(X_i, X_j))^2 \ge \epsilon + t \right) \le e^{-nt^2/2K^4}.$$

Therefore, with probability $\ge 1 - 1/4n^2$ we have:

$$\|\overline{T} - \widehat{T}\|_{L^2 \to L^2} \le \sqrt{\epsilon^2 + K^2 \sqrt{\frac{2\ln(4n^2)}{n}}} \le \epsilon + c\,K \left(\frac{\ln n}{n}\right)^{1/4},$$

where $c > 0$ is some universal constant. We deduce:

$$\|T_\kappa - T_{\kappa_\epsilon}\|_{L^2 \to L^2} + \|\overline{T} - \widehat{T}\|_{L^2 \to L^2} \le 2\epsilon + c\,K \left(\frac{\ln n}{n}\right)^{1/4} \tag{6.15}$$

with probability $\ge 1 - 1/4n^2$.

To finish the proof, we must bound the third term in (6.13). To do this, we notice that:

$$\widehat{T} - T_{\kappa_\epsilon} = T_\eta$$

where

$$\eta \equiv \sum_{1\leq i,j\leq n} (\kappa_\epsilon(X_i, X_j) - \kappa_\epsilon)\, \chi_{\left(\frac{\sigma_n(i)-1}{n}, \frac{\sigma_n(i)}{n}\right] \times \left(\frac{\sigma_n(j)-1}{n}, \frac{\sigma_n(j)}{n}\right]}.$$

Using the definition of σ_n from Section 6.2, one can rewrite this as:

$$\eta(x,y) = \sum_{1\leq i,j\leq n} \left(\kappa_\epsilon(\overline{X}_i, \overline{X}_j) - \kappa_\epsilon(x,y)\right) \chi_{\left(\frac{i-1}{n}, \frac{i}{n}\right] \times \left(\frac{j-1}{n}, \frac{j}{n}\right]}(x,y).$$

Recall that κ_ϵ is L_ϵ-Lipschitz and therefore,

$$\begin{aligned}
\forall (x,y) \in \left(\frac{i-1}{n}, \frac{i}{n}\right] \times \left(\frac{j-1}{n}, \frac{j}{n}\right], |\kappa_\epsilon(\overline{X}_i, \overline{X}_j) - \kappa_\epsilon(x,y)| &\leq \\
\leq 2L_\epsilon/n + |\kappa_\epsilon(\overline{X}_i, \overline{X}_j) - \kappa_\epsilon(i/n, j/n)| &\leq \\
\leq 2L_\epsilon/n + L_\epsilon|\overline{X}_i - i/n| + L_\epsilon|\overline{X}_j - j/n|&.
\end{aligned}$$

Integrating η^2, we find that:

$$\begin{aligned}
\int_{[0,1]^2} \eta^2 &\leq \frac{1}{n^2} \sum_{i,j=1}^{n} (2L_\epsilon/n + L_\epsilon|\overline{X}_i - i/n| + L_\epsilon|\overline{X}_j - j/n|)^2 \leq \\
&[\text{use } (a+b+c^2 \leq 3(a^2+b^2+c^2))] \leq \frac{12L_\epsilon^2}{n^2} + 6L_\epsilon^2 \max_{1\leq i\leq n} (\overline{X}_i - i/n)^2.
\end{aligned}$$

A simple calculation using e.g. Massart's version of the Dvoretsky-Kiefer-Wolfowitz inequality [51] reveals that the last term is $\leq c^2 \ln n/n$ ($c > 0$ universal) with probability $\geq 1 - 1/4n^2$. We deduce that:

$$\|\widehat{T} - T_{\kappa_\epsilon}\|_{L^2\to L^2} = \|T_\eta\|_{L^2\to L^2} \leq \sqrt{\int_{[0,1]^2} \eta^2} \leq \frac{2\sqrt{3}L}{n} + cL\sqrt{\frac{6\ln n}{n}}$$

with probability $\geq 1 - 1/4n^2$. Combining this with (6.15) and replacing $c > 0$ with a larger universal constant if necessary finishes the proof. $\square$

7 Freedman's inequality for matrix martingales

In this Section we prove our new concentration inequality, Theorem 1.2. We begin with some preliminaries from matrix analysis.

7.1 Preliminaries from matrix analysis

7.1.1 The positive semi-definite order

Matrix inequalities for the positive semi-definite order will be essential in our proof.

Given $A \in \mathbb{C}^{d\times d}_{\text{Herm}}$, say that $A \succeq 0$ if A is positive semi-definite, which is the same as saying that all eigenvalues of A are non-negative, or that $v^* A v \geq 0$ for all $v \in \mathbb{C}^d$. We will also write $A \preceq B$ (for $B \in \mathbb{C}^{d\times d}_{\text{Herm}}$) if $B - A \succeq 0$. Notice that $A \preceq \xi I$ for some $\xi \in \mathbb{R}$ iff $\lambda_{\max}(A) \leq \xi$.

We will need four other properties of the partial order "$\preceq$". The first three are easily checked and we omit their proofs:

$$\text{The set } \{(A,B) \in \left(\mathbb{C}^{d\times d}_{\text{Herm}}\right)^2 \,:\, A \preceq B\} \text{ is closed in the product topology.} \tag{7.1}$$

$$\forall \{A_i\}_{i=1}^k, \{B_i\}_{i=1}^k \subset \mathbb{C}^{d\times d}_{\text{Herm}} \,:\, \text{“}\forall 1 \leq i \leq k,\, A_i \preceq B_i\text{”} \Rightarrow \text{“}\sum_{i=1}^k A_i \preceq \sum_{i=1}^k B_i\text{”}. \tag{7.2}$$

$$\forall A, B \in \mathbb{C}^{d\times d}_{\text{Herm}} \,:\, \text{“}A \succeq 0\text{”} \Rightarrow \text{“}\lambda_{\max}(A+B) \geq \lambda_{\max}(B)\text{”}. \tag{7.3}$$

The fourth one is slightly less standard.

$$\forall A, B, C \in \mathbb{C}^{d\times d}_{\text{Herm}},\, (A \succeq 0 \wedge C - B \succeq 0) \Rightarrow \operatorname{Tr}(AB) \leq \operatorname{Tr}(AC). \tag{7.4}$$

To prove (7.4), notice that for for A, B, C as above,

$$\operatorname{Tr}(A(C-B)) = \operatorname{Tr}((C-B)^{1/2} A (C-B)^{1/2})$$

where $(C-B)^{1/2} \in \mathbb{C}^{d\times d}_{\text{Herm}}$ is the (also positive semi-definite) square root of $C - B$. Then notice that for any $v \in \mathbb{C}^n$,

$$v^*(C-B)^{1/2} A (C-B)^{1/2} v = [(C-B)^{1/2} v]^* A [(C-B)^{1/2} v] \geq 0$$

since $A \succeq 0$. This implies that $(C-B)^{1/2} A (C-B)^{1/2}$ must be positive semi-definite, hence its trace is non-negative: $\operatorname{Tr}(A(C-B)) \geq 0$, which is equivalent to (7.4) by linearity.

7.1.2 Conditional expectations are monotone

We will also need the following property that relates expectations to the positive semi-definite order. Let X, Y be integrable, random $d\times d$ Hermitian matrices defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then:

$$\text{If } X \preceq Y \text{ almost surely, then } \mathbb{E}[X \mid \mathcal{G}] \preceq \mathbb{E}[Y \mid \mathcal{G}] \text{ almost surely.} \tag{7.5}$$

To see this, it suffices to see that for all $v \in \mathbb{C}^d$, $v^*Xv \le v^*Yv$ and therefore $\mathbb{E}[v^*Xv \mid \mathcal{G}] \le \mathbb{E}[v^*Yv \mid \mathcal{G}]$. However, our definition of $\mathbb{E}[\cdot \mid \mathcal{G}]$ for matrices (cf. Section 2.4) implies that $\mathbb{E}[v^*Xv \mid \mathcal{G}] = v^*\mathbb{E}[X \mid \mathcal{G}]\, v$ and $\mathbb{E}[v^*Yv \mid \mathcal{G}] = v^*\mathbb{E}[Y \mid \mathcal{G}]\, v$. Therefore, if

$$X \preceq Y \text{ almost surely } \Rightarrow \forall v \in \mathbb{C}^d,\ \text{“}v^*\mathbb{E}[X \mid \mathcal{G}]\, v \le v^*\mathbb{E}[Y \mid \mathcal{G}]\, v \text{ almost surely”}.$$

Now let $Q \subset \mathbb{C}^d$ be dense and countable. Note that for all $A \in \mathbb{C}^{d\times d}_{\text{Herm}}$, $A \preceq 0$ if and only if $v^*Av \ge 0$ for all $v \in Q$.

$$\mathbb{E}[X \mid \mathcal{G}] \preceq \mathbb{E}[Y \mid \mathcal{G}] \text{ a.s.} \Leftrightarrow \mathbb{P}\left(\forall v \in Q,\ v^*\mathbb{E}[X \mid \mathcal{G}]\, v \le v^*\mathbb{E}[Y \mid \mathcal{G}]\, v\right) = 1$$

and the RHS follows from $X \preceq Y$ by the previous implication (since Q is countable).

7.1.3 Matrix functions and matrix exponentials

If $f : \mathbb{C} \to \mathbb{C}$ given by a power series $f(x) = \sum_{i=1}^{\infty} c_i x^i$ that converges for all $x \in \mathbb{C}$, one may define:

$$f(A) \equiv \sum_{i=1}^{\infty} c_i A^i, A \in \mathbb{C}^{d\times d},$$

which can be shown to converge for all A. $f(A)$ is Hermitian whenever $A \in \mathbb{C}^{d\times d}_{\text{Herm}}$ and the coefficients c_i belong to $\mathbb{R}$. In that case, the eigenvalues of $f(A)$ are given by $f(\lambda_i(A))$ for $0 \le i \le d-1$, with the same eigenvectors as A. In particular, $f(A) \preceq \xi I$ for some $\xi \in \mathbb{R}$ iff $f(\lambda_i(A)) \le \xi$ for each $0 \le i \le d-1$. Moreover, for all $s \ge 0$,

$$\exp(s\lambda_{\max}(A)) = \lambda_{\max}(\exp(sA)) \le \operatorname{Tr}(\exp(sA)). \tag{7.6}$$

We need one more result from matrix analysis, called the *Golden Thompson inequality.*

$$\forall d \in \{1, 2, 3, \dots\}, \forall A, B \in \mathbb{C}^{d\times d}_{\text{Herm}} : \operatorname{Tr}(e^{A+B}) \le \operatorname{Tr}(e^A e^B). \tag{7.7}$$

This inequality is fundamental in adapting the standard proofs of concentration to the matrix setting [2, 21, 53].

7.2 The proof

We begin with two simple Lemmas.

Lemma 7.1 *For any matrix $C \in \mathbb{C}^{d\times d}_{\text{Herm}}$ and $k \in \mathbb{N}\backslash\{0,1\}$, $C^k \preceq \|C\|_2^{k-2} C^2$.*

Proof: $\|C\|_2^{k-2}C^2 - C^k$ has the same eigenvectors as C and its eigenvalues are given by

$$\|C\|_2^{k-2}\lambda_i(C)^2 - \lambda_i(C)^k = (\|C\|_2^{k-2} - \lambda_i(C)^{k-2})\lambda_i(C)^2.$$

This is always ≥ 0 because $\|C\|_2 = \max_{1\leq i\leq d}|\lambda_i(C)|$. $\Box$

Lemma 7.2 *For any matrix $C \in \mathbb{C}^{d\times d}_{\mathrm{Herm}}$ with $\|C\|_2 \leq 1$, $e^C \preceq I + C + C^2$.*

Proof: The previous lemma implies that $C^i \preceq C^2$ for all $i \geq 2$. Property (7.2) of "$\preceq$" implies that for any k,

$$I + C + \sum_{i=2}^{k} \frac{C^i}{i!} \preceq I + C + \left(\sum_{i=2}^{k} \frac{1}{i!}\right) C^2 \preceq I + C + C^2.$$

Now let $k \nearrow +\infty$ and use (7.1). $\Box$

The next step is an exponential inequality for martingales.

Lemma 7.3 (Exponential inequality for martingales) *Let Z_n, W_n be as in Theorem 1.2 with $M = 1$. Then for all $s \in [0, 1/2]$ and all deterministic $C \in \mathbb{C}^{d\times d}_{\mathrm{Herm}}$,*

$$\mathbb{E}\left[\mathrm{Tr}\left[\exp\left(sZ_n - 2s^2W_n + C\right)\right]\right] \leq \mathrm{Tr}\left[\exp\left(C\right)\right].$$

Proof: Set $X_n \equiv Z_n - Z_{n-1}$ and $\Delta_n \equiv \mathbb{E}\left[X_n^2 \mid \mathcal{F}_{n-1}\right]$. We use Golden Thompson (7.7) to deduce that:

$$\mathrm{Tr}(e^{sZ_n - 2s^2W_n + C}) \leq \mathrm{Tr}(e^{sX_n - 2s^2\Delta_n} e^{sZ_{n-1} - 2s^2W_{n-1} + C}).$$

Taking conditional expectations, we see that:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Tr}(e^{sZ_n - s^2W_n + C}) \mid \mathcal{F}_{n-1}\right] &\leq \mathbb{E}\left[\mathrm{Tr}(e^{sDX_n - 2s^2\Delta_n} e^{sZ_{n-1} - 2s^2W_{n-1} + C}) \mid \mathcal{F}_{n-1}\right] \\
&= \mathrm{Tr}(\mathbb{E}\left[e^{sX_n - 2s^2\Delta_n} \mid \mathcal{F}_{n-1}\right] e^{sZ_{n-1} - 2s^2W_{n-1} + C}).
\end{aligned}$$

Here the equality is a result of Tr and expected values commuting (2.3), as well as noting that $e^{sX_{n-1} - 2s^2W_{n-1} + C}$ is $\mathcal{F}_{n-1}$-measurable and then applying (2.4) to the conditional expectation.

We now make the following claim.

Claim 7.1 $\mathbb{E}\left[e^{sX_n - 2s^2\Delta_n} \mid \mathcal{F}_{n-1}\right] \preceq I$.

This will imply (via monotonicity of the trace (7.4)) that:

$$\mathbb{E}\left[\mathrm{Tr}(e^{sZ_n - 2s^2W_n + C}) \mid \mathcal{F}_{n-1}\right] \leq \mathrm{Tr}(e^{sZ_{n-1} - 2s^2W_{n-1} + C}),$$

hence

$$\mathbb{E}\left[\mathrm{Tr}(e^{sZ_n - 2s^2W_n + C})\right] \leq \mathbb{E}\left[\mathrm{Tr}(e^{sZ_{n-1} - 2s^2W_{n-1} + C})\right]$$

and the Lemma follows from this via induction in n.

To prove the claim, we first note that for $|s| \leq 1/2$,

$$\|sX_n - 2s^2\Delta_n\|_2 \leq \frac{\|X_n\|_2 + \|\Delta_n\|_2}{2} \leq 1$$

by the assumption that $\|X_n\|_2 \leq 1$. We now apply Lemma 7.2 with $C = sX_n - s^2\Delta_n$ and the monotonicity of conditional expectations (7.4) to obtain:

$$\mathbb{E}\left[e^{sX_n - 2s^2\Delta_n} \mid \mathcal{F}_{n-1}\right] \preceq \mathbb{E}\left[I + sX_n - 2s^2\Delta_n + s^2X_n^2 - 2s^3X_n\Delta_n - 2s^3X_n\Delta_n + 4s^4\Delta_n^2 \mid \mathcal{F}_{n-1}\right].$$

$\Delta_n = \mathbb{E}\left[X_n^2 \mid \mathcal{F}_{n-1}\right]$ is $\mathcal{F}_{n-1}$-measurable and the martingale property implies $\mathbb{E}\left[X_n \mid \mathcal{F}_{n-1}\right] = 0$. Via equation (2.4), this implies $\mathbb{E}\left[\Delta_n X_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[X_n\Delta_n \mid \mathcal{F}_{n-1}\right] = 0$ almost surely. This means that the RHS above is a.s. equal to:

$$I - s^2\Delta_n + 4s^4\Delta_n^2.$$

Now notice that the eigenvalues of $-s^2\Delta_n + 4s^4\Delta_n^2$ are given by:

$$-s^2\lambda_i(\Delta_n) + 4s^4\lambda_i(\Delta_n)^2, 1 \leq i \leq d.$$

The inequality $s \leq 1/2$ implies $4s^4 \leq s^2$. Moreover, each $\lambda_i(\Delta_n)$ is between 0 and 1, since $\|\Delta_n\| \leq 1$ and $\Delta_n \succeq 0$ (it is the conditional expectation of X_n^2). This implies that the above expression is at most:

$$-s^2\lambda_i(\Delta_n) + s^2\lambda_i(\Delta_n) = 0$$

for each i. Therefore, $-s^2\Delta_n + 4s^4\Delta_n^2 \preceq 0$ and (again using the monotonicity property (7.5)), $\mathbb{E}\left[e^{sX_n - 2s^2\Delta_n} \mid \mathcal{F}_{n-1}\right] \preceq I$ almost surely. $\square$

Proof: [of Theorem 1.2] One may assume that $M = 1$ (one can always rescale Z_n so that this is the case; the bound behaves accordingly). If $\lambda_{\max}(W_n) \leq \sigma^2$, $\sigma^2 I - W_n \succeq 0$ is positive semi-definite. Inequality (7.3) then implies that for all $s > 0$,

$$\lambda_{\max}(sX_n + 2s^2\sigma^2 I - 2s^2W_n) \geq \lambda_{\max}(sX_n) = s\lambda_{\max}(X_n).$$

Therefore,

$$\begin{aligned}\forall s > 0,\ \mathbb{P}\left(\lambda_{\max}(X_n) \geq t,\ \lambda_{\max}(W_n) \leq \sigma^2\right) &\leq \mathbb{P}\left(\lambda_{\max}(sX_n + 2s^2\sigma^2 I - 2s^2W_n) \geq st\right) \\ &\leq e^{-st}\mathbb{E}\left[\exp(\lambda_{\max}(sX_n + 2s^2\sigma^2 I - 2s^2W_n))\right].\end{aligned}$$

We now use the inequality "$e^{\lambda_{\max}(sZ)} \leq \mathrm{Tr}(e^{sZ})$", valid for any $s \geq 0$ and $Z \in \mathbb{C}^{d\times d}_{\mathrm{Herm}}$ (cf. (7.6)), together with the exponential inequality in Lemma 7.3, to deduce that for all $s \in [0, 1/2]$,

$$\begin{aligned} \mathbb{P}\left(\lambda_{\max}(sX_n + 2s^2\sigma^2 I - 2s^2 W_n) \geq st\right) &\leq e^{-st}\mathbb{E}\left[\mathrm{Tr}(\exp(sX_n + 2s^2\sigma^2 I - 2s^2 W_n))\right] \\ &\leq \mathrm{Tr}(\exp(2s^2\sigma^2 I))e^{-st} = d\, e^{2s^2\sigma^2 - st}. \end{aligned}$$

Set

$$s \equiv \frac{t}{4\sigma^2 + 2t}.$$

Notice that with this choice $s \leq 1/2$ always. Moreover,

$$2s^2\sigma^2 = \frac{t^2}{8\sigma^2(1 + t/2\sigma^2)^2} \leq \frac{t^2}{8\sigma^2(1 + t/2\sigma^2)} = \frac{st}{2}.$$

Hence:

$$\mathbb{P}\left(\lambda_{\max}(X_n) \geq t,\, \lambda_{\max}(W_n) \leq \sigma^2\right) \leq d\, e^{-\frac{t^2}{8\sigma^2 + 4t}},$$

as desired. □

Remark 7.1 *It is well-known in the scalar case that inequalities for martingales imply inequalities for independent sums. The same is true in the matrix setting. Let $X_1, \dots, X_n$ be mean-zero independent random matrices, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in $\mathbb{C}^{d\times d}_{\mathrm{Herm}}$ and such that there exists a $M > 0$ with $\|X_i\| \leq M$ almost surely for all $1 \leq i \leq m$. Letting $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$ ($i \in [n]$), one can see that:*

$$\{(Z_i \equiv \sum_{j=1}^{i} X_j, \mathcal{F}_i)\}_{i=0}^{n}$$

is a martingale satisfying the assumptions of the Theorem and that, moreover, W_n is deterministic in this case:

$$W_n \equiv \sum_{i=1}^{n} \mathbb{E}\left[(Z_i - Z_{i-1})^2 \mid \mathcal{F}_{i-1}\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right].$$

Thus one may take:

$$\sigma^2 = \lambda_{\max}\left(\sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right]\right)$$

in Theorem 1.2 and deduce the first half of the Corollary below. The other half comes from considering $-\sum_{i=1}^{n} X_i$.

Corollary 7.1 *Let $X_1, \dots, X_n$ be mean-zero independent random matrices, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in $\mathbb{C}^{d\times d}_{\mathrm{Herm}}$ and such that there exists a $M > 0$ with*

$\|X_i\| \le M$ *almost surely for all* $1 \le i \le m$. *Define:*

$$\sigma^2 \equiv \lambda_{\max}\left(\sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right]\right).$$

Then for all $t \ge 0$,

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} X_i\right) \ge t\right) \le d\, e^{-\frac{t^2}{8\sigma^2+4Mt}},$$

and

$$\mathbb{P}\left(\left\|\sum_{i=1}^{n} X_i\right\| \ge t\right) \le 2d\, e^{-\frac{t^2}{8\sigma^2+4Mt}}.$$

8 Final remarks

Sharpness of Theorem 1.2. One can show that Theorem 1.2 is close to sharp and that, in particular, the d factor in the bound is necessary for general martingale sequences. To see this, consider a sum Z_n of n independent, identically distributed $d \times d$ diagonal random matrices $X_1, \dots, X_n$ whose diagonal entries are independent, unbiased ± 1. The largest eigenvalue of Z_n is a maximum of d independent random sums, each with n terms of the kind ± 1 above. One can see that for large n and d and for $t \approx \sqrt{n \ln d}$,

$$\mathbb{P}\left(\lambda_{\max}(Z_n) \ge t\right) \ge de^{-(1+o(1))t^2/2n}$$

which is what Corollary 7.1 gives up to the constants in the exponent.

An interesting question is to understand the circumstances under which one can remove the d factor from the bound. For instance, can the sharper results of [38, 33] be reobtained via some variant of Theorem 1.2?

Other applications of Theorem 1.2. In a related paper (in preparation) we show how Theorem 1.2 can be used to show concentration of the matrices of random lifts of large graphs. A pleasing corollary of our result is this: consider a random k_1k_2-lift of a large graph G with minimum degree $\omega(\ln(k_1k_2n))$. The Laplacian of this lift is essentially indistinguishable from that of the (in principle very different) random graph obtained by performing a k_1-lift on G and then a k_2-lift on the resulting graph.

It would be interesting to see other applications of Theorem 1.2, especially in settings where the Christofides-Märkstrom bound is useless because its variance term is too large (cf. Remark 3.1).

The Laplacian of inhomogeneous random graphs. The results of the Section 6 can be extended to the Laplacian $\mathcal{L}_{n,p,\kappa}$ of $G_{n,p,\kappa}$. More precisely, add the following condition to Assumption 6.1: that there exists a $K_- > 0$ such that for all $x \in [0,1]$, $\kappa(x) \equiv \int_0^1 \kappa(x,y)\,dy \geq K_-$. Then there is a close correspondence between $\mathcal{L}_{n,p,\kappa}$ and the operator $S_\xi \equiv \mathrm{Id}_{L^2} - T_\xi$, where Id_{L^2} is the identity operator on $L^2([0,1])$ and T_ξ is the integral operator given by the symmetric, non-negative function:

$$\xi(\cdot,\cdot\cdot) = \frac{\kappa(\cdot,\cdot\cdot)}{\sqrt{\kappa(\cdot)\kappa(\cdot\cdot)}}.$$

That is, if $p \leq 1/K$ and $pnK_- \gg C\ln n$ for some C, we will have:

$$\|\mathcal{L}_{n,p,\kappa} - E_n S_\xi H_n\| = o\,(1) \text{ and } \|H_n \mathcal{L}_{n,p,\kappa} E_n - S_\xi\| = o\,(1)\,,$$

with consequences for the spectrum and eigenspaces of $\mathcal{L}_{n,p,\kappa}$. We omit the details.

Better bounds and extensions? We have mentioned the results on spectral gaps in references [33] and [29], on $G_{n,p}$ and random graphs with given expected degrees. These papers actually do much more than we described, as they show that, even is very sparse graphs, there is a large "core" set of vertices so that the matrices of the induced subgraph are well-behaved. It would be an interesting question to prove a similar result either for more general instances of bond percolation or inhomonegeous random graphs.

Cut convergence, eigenvalues and eigenvectors. It is not clear to the author what one can/cannot prove about eigenvectors and eigenvalues of sparse graphs while only assuming that they converge to a given κ in the cut norm. Ideally, one would wish to be able to prove that this suffices for the convergence of the given operators, at least under suitable assumptions, but it is not clear how one should proceed.

A Appendix: two perturbation results

The following functional-analytic perturbation results are needed in the main text. In what follows $\mathcal{H}$ is a real Hilbert space and $\|\cdot\|$ denotes both the Hilbert space norm and the induced norm on linear operators. Undefined notions and quoted results can be found in any textbook on Functional Analysis, eg. [55, 45].

Lemma A.1 *Suppose V, W are compact Hermitian linear operators on the Hilbert space $\mathcal{H}$ that satisfy $\|V - W\| \leq \epsilon$. Let* $\mathrm{spec}(V)$, $\mathrm{spec(W)}$ *denote the spectra of V and W (respectively). Let $S \subset \mathbb{R}$ be such that* $\inf_{s\in S}|s| > \epsilon$ *and let $m_V(S)$ be the sum of the multiplicities of all elements* *of* $\mathrm{spec}(V) \cap S$. *Then:*

$$m_V(S) \leq m_W(S^\epsilon)$$

where for $A \subset \mathbb{R}$, $A^\epsilon \equiv \{x \in \mathbb{R} : \exists a \in A, |x - a| \leq \epsilon\}$.

Proof: This is evident if both V and W have finite-dimensional rank. In this case one may restrict to the span of the two ranges, which is a finite-dimensional space isomorphic to some $\mathbb{R}^d$, and then apply (3.1). [Do notice that 0 might belong to the spectrum of the restriction of V or W to the finite-dimensional subspace, even though it does not belong to the original spectra. This, however, will not matter, due to the condition $\inf_{s\in S} |s| > \epsilon$.]

For the case of infinite-dimensional rank, V and W are the limit (in the operator norm) of operators of finite-dimensional rank. More specifically, recall from Section 2.2 that the spectral theorem for compact, self-adjoint operators states that V can be written as a sum:

$$V = \sum_{\alpha \in \mathrm{spec}(V)} \alpha P_\alpha$$

where the P_α are orthogonal projectors of orthogonal ranges, with finite rank if $\alpha \neq 0$. Moreover, for any $\delta > 0$, $\mathrm{spec}(V)\backslash(-\delta, \delta)$ is finite. Therefore, the finite-rank operator:

$$V_\delta = \sum_{\alpha \in \mathrm{spec}(V)\backslash(-\delta,\delta)} \alpha P_\alpha$$

satisfies $\|V_\delta - V\| \leq \delta$. One may similarly define W_δ with $\|W_\delta - W\| \leq \delta$ and it follows that $\|V_\delta - W_\delta\| \leq \epsilon + 2\delta$. Moreover, we have the simple fact:

$$\forall A \subset \mathbb{R}\backslash[-\delta, \delta],\ m_{V_\delta}(A) = m_V(A) \text{ and } m_{W_\delta}(A) = m_W(A). \tag{A.1}$$

Let $\delta > 0$ be small, so that $\inf_{s\in S} |s| > \epsilon + 3\delta$. The finite-dimensional result implies:

$$m_{V_\delta}(S) \leq m_{W_\delta}(S^{\epsilon+2\delta}).$$

Notice that $m_{V_\delta}(S) = m_V(S)$ because $S \subset \mathbb{R}\backslash[-\epsilon, \epsilon] \subset \mathbb{R}\backslash[-\delta, \delta]$ and therefore (A.1) applies. Moreover, $\forall x \in S^{\epsilon+2\delta}$,

$$|x| \geq \inf_{s\in S} |s| - \epsilon - 2\delta > \delta$$

by the choice of δ; therefore $S^{\epsilon+2\delta} \subset \mathbb{R}\backslash[-\delta, \delta]$ and we can apply (A.1) again to deduce that $m_{W_\delta}(S^{\epsilon+2\delta}) = m_W(S^{\epsilon+2\delta})$. These facts imply:

$$m_V(S) \leq m_W(S^{\epsilon+2\delta}).$$

It is an exercise to show that $m_W(S^{\epsilon+2\delta}) \to m_W(S^\epsilon)$ when $\delta \searrow 0$. This finishes the proof. $\square$

Lemma A.2 *Suppose V, W are compact Hermitian linear operators on the Hilbert space $\mathcal{H}$ that satisfy $\|V - W\| \le \epsilon$. Assume that $a < b$ and $\gamma > \epsilon$ be such that $a + \gamma < b - \gamma$ and V does not contain any eigenvalues in $(a-\gamma, a+\gamma) \cup (b-\gamma, b+\gamma)$. Define $\Pi_{a,b}(V)$ as the projector onto the span of the eigenvectors of V corresponding to $a \le \lambda_k(V) \le b$ and define $\Pi_{a,b}(W)$ similarly. Then:*

$$\|\Pi_{a,b}(V) - \Pi_{a,b}(W)\| \le \frac{(b-a+2\gamma)\,\epsilon}{\pi(\gamma^2 - \gamma\epsilon)}.$$

Proof: Suppose first that $\mathcal{H}$ is finite-dimensional, in which case one may assume that $\mathcal{H} = \mathbb{C}^d$ for some d and that V and W are matrices. In this case we use a standard technique involving contour integration in the complex plane and the resolvent of linear operators [41, Chapter 2].

Let $\mathcal{C}$ be the rectangular contour in the complex plane that passes through the points $a + \gamma\sqrt{-1}$, $a - \gamma\sqrt{-1}$, $b - \gamma\sqrt{-1}$, $b + \gamma\sqrt{-1}$ in counterclockwise order. The Cauchy formula implies that for all $\lambda \in \mathbb{R}\backslash\{a, b\}$,

$$\frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} \frac{dz}{z-\lambda} = \begin{cases} 1, & a < \lambda < b \\ 0, & \text{otherwise.} \end{cases}$$

Now consider the *resolvent*:

$$R_V(z) \equiv (zI - V)^{-1},\ z \in \mathbb{C}\backslash\{\lambda_i(V) \,:\, 0 \le i \le n-1\}.$$

The spectral theorem implies that:

$$R_V(z) = \sum_{k=0}^{d-1} \frac{\psi_{k,V}\psi_{k,V}^*}{z - \lambda_k(V)}.$$

where $\psi_{k,V}$ is the eigenvector of V corresponding to $\lambda_k(V)$. By assumption, V has no eigenvalues on $\mathcal{C}$, therefore:

$$\frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} R_V(z)\,dz = \sum_{k=0}^{d-1} \frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} \frac{\psi_{k,V}\psi_{k,V}^*}{z-\lambda_k(V)}\,dz = \sum_{k:\lambda_k(V)\in[a,b]} \psi_{k,V}\psi_{k,V}^* = \Pi_{a,b}(V).$$

Now define the resolvent $R_W(z) = (zI - W)^{-1}$. Recall that $|\lambda_i(V) - \lambda_i(W)| \le \epsilon < \gamma$ by (3.1) and that no eigenvalue of V lies in $(a-\gamma, a+\gamma) \cup (b-\gamma, b+\gamma)$ (by assumption). This implies that no eigenvalue of W can lie on a or b. Therefore, the same reasoning used above implies that:

$$\frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} R_W(z)\,dz = \Pi_{a,b}(W).$$

In particular,

$$\|\Pi_{a,b}(V) - \Pi_{a,b}(W)\| = \left\| \frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} (R_V(z) - R_W(z))\, dz \right\|.$$

It is not hard to show that:

$$\left\| \frac{1}{2\pi\sqrt{-1}} \int_{\mathcal{C}} (R_V(z) - R_W(z))\, dz \right\| \leq \frac{1}{2\pi} \int_{\mathcal{C}} \|R_V(z) - R_W(z)\|\, d|z|.$$

Since $\mathcal{C}$ has length $2(b-a) + 4\gamma$, we have:

$$\|\Pi_{a,b}(V) - \Pi_{a,b}(W)\| \leq \frac{(b-a+2\gamma)}{\pi} \max_{z\in\mathcal{C}} \|R_V(z) - R_W(z)\|. \tag{A.2}$$

We now bound the difference between the resolvents. Recall that for $T \in \mathbb{C}^{d\times d}$ with $\|T\| < 1$,

$$(I+T)^{-1} = \sum_{n\geq 0} T^n.$$

Suppose we can show that $\|(W-V)R_V(z)\| \leq \alpha < 1$ for $z \in \mathcal{C}$. Then:

$$\begin{aligned}
\|R_W(z) - R_V(z)\| &= \|((zI - V) - (W-V))^{-1} - R_V(z)\| \\
&= \|(zI-V)^{-1}\,(I - (W-V)(zI-V)^{-1})^{-1} - R_V(z)\| \\
&= \|R_V(z)\,\{(I - (W-V)R_V(z))^{-1} - I\}\| \\
&= \|\sum_{n\geq 1} R_V(z)[(W-V)\,R_V(z)]^n\| \\
&\leq \|R_V(z)\| \sum_{n\geq 1} \|(W-V)\,R_V(z)\|^n \\
&\leq \|R_V(z)\|\,\frac{\alpha}{1-\alpha}.
\end{aligned}$$

But in our case we have:

$$\|R_V(z)\| = \left\| \sum_{k=0}^{d-1} \frac{\psi_{k,V}\psi_{k,V}^*}{z - \lambda_k(V)} \right\| = \max_k |z - \lambda_k(V)|^{-1} \leq 1/\gamma$$

because all $\lambda_k(V)$ lie within distance $\geq \gamma$ from the contour $\mathcal{C}$ (this follows from the assumption that no $\lambda_k(V)$ is in $(a-\gamma, a+\gamma) \cup (b-\gamma, b+\gamma)$). Moreover, $\|W - V\| \leq \epsilon$ by assumption. Therefore, $\|(W-V)R_V(z)\| \leq \epsilon/\gamma < 1$ and, by the above,

$$\|R_W(z) - R_V(z)\| \leq \frac{\epsilon}{\gamma^2 - \gamma\epsilon}.$$

Together with (A.2), this finishes the proof for the finite-dimensional case.

We now consider the case of arbitrary $\mathcal{H}$. Recall the definitions of V_δ and W_δ from the previous proof. It is easy to deduce from the definition of V_δ that for any $v \in \mathbb{C}^d$,

$$\Pi_{a,b}(V)\, v = \lim_{\delta \searrow 0} \Pi_{a,b}(V_\delta)$$

and similarly

$$\Pi_{a,b}(W)\, v = \lim_{\delta \searrow 0} \Pi_{a,b}(W_\delta)\, v \text{ where } W_\delta \equiv \sum_{i:|\lambda_i| \geq \delta} \eta_i\, \psi_{i,W} \psi_{i,W}^*.$$

Since V_δ and W_δ have finite dimensional rank, one sees from the first part that for all small enough $\delta > 0$,

$$\|\, (\Pi_{a,b}(V_\delta) - \Pi_{a,b}(W_\delta))\; v\| \leq \|v\|\, \|\Pi_{a,b}(V_\delta) - \Pi_{a,b}(W_\delta)\| \leq \frac{(b - a + \gamma)\,(\epsilon + 2\delta)}{\pi(\gamma^2 - \gamma(\epsilon + 2\delta))}$$

since $\|V_\delta - W_\delta\| \leq \epsilon + 2\delta < \gamma$. Letting $\delta \searrow 0$ implies:

$$\|\, (\Pi_{a,b}(V) - \Pi_{a,b}(W))\; v\| \leq \|v\|\, \frac{(b - a + \gamma)\,\epsilon}{\pi(\gamma^2 - \gamma\epsilon)}$$

and since v is arbitrary this finishes the proof. $\square$

References

[1] Dimitris Achlioptas and Assaf Naor. “The two possible values of the chromatic number of a random graph.” Annals of Mathematics 162(3): 1333-1351 (2005).

[2] Rudolf Ahlswede and Andreas Winter. “Strong converse for identification via quantum channels.” IEEE Transactions on Information Theory 48(3): 569-579 (2002).

[3] Noga Alon, Itai Benjamini and Alan Stacey. “Percolation on Finite Graphs and Isoperimetric Inequalities.” Annals of Probability 32(3):1727-1745 (2004).

[4] Noga Alon and Nabil Kahale. “A Spectral Technique for Coloring Random 3-Colorable Graphs.” SIAM Journal of Computing 26: 1733-1748 (1997).

[5] Noga Alon and Michael Krivelevich. “The concentration of the chromatic number of random graphs.” Combinatorica 17(3):303-313 (1997).

[6] Noga Alon, Michael Krivelevich and Van Vu. “On the concentration of eigenvalues of random symmetric matrices.” Israel Journal of Mathematics 131(1):259-267 (2002).

[7] Noga Alon and Joel Spencer. *The Probabilistic Method* (2nd edition). Wiley Interscience (2004).

[8] Alon Amit and Nathan Linial. “Random graph coverings, I: General theory and graph connectivity.” Combinatorica 22:1-18 (2002).

[9] Béla Bollobás. *Random Graphs* (2nd edition). Cambridge University Press (2001).

[10] Béla Bollobás, Christopher Borgs, Jennifer Chayes and Oliver Riordan. “Percolation on dense graph sequences.” To appear in Annals of Probability.

[11] Béla Bollobás, Svante Janson and Oliver Riordan. “The phase transition in inhomogeneous random graphs.” Random Structures and Algorithms 31(1):3-122 (2007).

[12] Béla Bollobás, Svante Janson and Oliver Riordan. “The cut metric, random graphs, and branching processes.” Arxiv:0901.2091.

[13] Béla Bollobás and Oliver Riordan. “Metrics for sparse graphs.” Arxiv: 0708.1919.

[14] Béla Bollobás and Oliver Riordan. “Sparse graphs: metrics and random models.” Arxiv: 0812.2656.

[15] Christian Borgs, Jennifer T. Chayes, Remco van der Hofstad, Gordon Slade and Joel Spencer. “Random subgraphs of finite graphs: I. The scaling window under the triangle condition.” Random Structures and Algorithms 27: 137-184 (2005).

[16] Christian Borgs, Jennifer T. Chayes, Lazsló Lóvasz, Vera T. Sós and Katalin Vesztergombi. “Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing.” Advances in Mathematics 219(6): 1901-1851 (2008).

[17] Christian Borgs, Jennifer T. Chayes, Lazsló Lóvasz, Vera T. Sós and Katalin Vesztergombi. “Convergent sequences of dense graphs II: Convergent Sequences of Dense Graphs II. Multiway Cuts and Statistical Physics.” Preprint.

[18] Stéphane Boucheron, Olivier Bousquet and Gábor Lugosi. “Concentration inequalities.” In O. Bousquet, U.v. Luxburg, and G. Rtsch (editors), Advanced Lectures in Machine Learning, Springer, 208-240 (2004).

[19] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart. “Moment inequalities for functions of independent random variables.” Annals of Probability, 33:514-560 (2005).

[20] Herman Chernoff. “A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations.” Annals of Mathematical Statistics 23(4):493-507 (1952).

[21] Demetres Christofides and Klas Markström. "Expansion properties of random Cayley graphs and vertex transitive graphs via matrix martingales." Random Structures and Algorithms 32(1):88-100 (2007).

[22] Fan Chung Graham. *Spectral Graph Theory.* CBMS Series, American Mathematical Society (1997).

[23] Fan Chung, Ronald Graham and Richard Wilson. "Quasi-Random Graphs." Combinatorica 22(2):217-244 (2007).

[24] Fan Chung and Ronald Graham. "Sparse Quasi-Random Graphs." Combinatorica 22(2):217-244 (2007).

[25] Fan Chung and Ronald Graham. "Quasi-random graphs with given degree sequences." Random Structures and Algorithms 32(1):1-19 (2007).

[26] Fan Chung and Paul Horn. "The Spectral Gap of a Random Subgraph of a Graph." Internet Mathematics 4(2-3):225-244 (2007).

[27] Fan Chung, Paul Horn and Linyuan Liu. "The giant component in a random subgraph of a given graph." In *Proceedings of the Workshop on Algorithms and Models for the Web-Graph (WAW 2009).* Lecture Notes in Computer Science 5427. Springer Verlag (2009).

[28] Fan Chung, Linyuan Lu and Van H. Vu. "The Spectra of Random Graphs with Given Expected Degrees." Internet Mathematics 1(3):257-275 (2004)

[29] Amin Coja-Oghlan and Andre Lanka. "The spectral gap of random graphs with given expected degrees." Preprint.

[30] Amin Coja-Oghlan and Andre Lanka. "Partitioning Random Graphs with General Degree Distributions." Preprint.

[31] Igor Devetak. "The private classical capacity and quantum capacity of a quantum channel." IEEE Transactions on Information Theory 51(1): 44-55 (2005).

[32] Igor Devetak and Andreas Winter. "Relating Quantum Privacy and Quantum Coherence: An Operational Approach." Physical Review Letters 93: 080501 (2004).

[33] Uriel Feige and Eran Ofek. "Spectral Techniques Applied to Sparse Random Graphs." Random Structures and Algorithms 27(2): 251-275 (2005).

[34] David Freedman. "On tail probabilities for martingales." Annals of Probability 3:100-118 (1975).

[35] Joel Friedman. "Relative expanders or weakly relatively Ramanujan graphs." Duke Mathematical Journal 118(1):19-35 (2003).

[36] Joel Friedman. "A Proof of Alon's Second Eigenvalue Conjecture and Related Problems." Memoirs of the American Mathematical Society 195 (2008).

[37] Alan Frieze, Michael Krivelevich and Ryan Martin. "The emergence of a giant component in a random subgraph of pseudo-random graphs." Random Structures and Algorithms 24(1):42-50 (2003).

[38] Zoltan Füredi and Janos Komlos. "The eigenvalues of random symmetric matrices". Combinatorica, 1(3):233241 (1981).

[39] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." Journal of the American Statistical Association 58(301):13-30 (1963).

[40] Roger Horn and Charles Johnson. *Matrix Analysis.* Cambridge University Press (1990).

[41] Tosio Kato. *Perturbation Theory for Linear Operators* (reprint of the 1980 edition). Springer (1995).

[42] Michael Krivelevich and Benjamin Sudakov. "The largest eigenvalue of sparse random graphs." Combinatorics, Probability and Computing 12:61-72 (2003).

[43] Michael Krivelevich and Benjamin Sudakov. "Pseudo-random graphs." In *More Sets, Graphs and Numbers*, Bolyai Society Mathematical Studies 15: 199-262 (2006), Springer.

[44] Zeph Landau and Alexander Russell. "Random Cayley graphs are expanders: a simplified proof of the Alon-Roichman theorem." The Electronic Journal of Combinatorics, 11(1) (2004).

[45] Peter Lax. *Functional Analysis.* Wiley-Interscience, New York (2002).

[46] Michel Ledoux. *The concentration of measure phenomenon.* American Mathematical Society (2001).

[47] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces.* Springer (1991).

[48] Nathan Linial and Doron Puder. "Words Maps and Spectra of Random Graph Lifts." Math arXiv:0806.1993 (2008).

[49] Laszló Lóvasz and Balázs Szegedy. "Testing properties of graphs and functions." To appear in the Israel Journal of Mathematics.

[50] Françoise Lust-Piquard and Gilles Pisier. "Non commutative Khintchine and Paley inequalities." Arkiv för Matematik, 29(2): 241-260 (1991).

[51] Pascal Massart. "The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality." Annals of Probability 18(3):1269-1283 (1990).

[52] Asaf Nachmias. "Mean-field conditions for percolation on finite graphs." Geometric and Functional Analysis, to appear.

[53] Roberto Imbuzeiro Oliveira. "Concentration of random sums of matrices and a new proof of a bound by Rudelson". Submitted to Electronic Communications in Probability.

[54] Roberto Imbuzeiro Oliveira. "Spectral properties of random lifts of large graphs". In preparation.

[55] Frigyes Riesz and Béla Sz.-Nagy. *Functional Analysis.* Dover Publications (1990).

[56] Mark Rudelson. "Random vectors in the isotropic position." Journal of Functional Analysis, 164 (1): 60-72 (1999).

[57] Eli Shamir and Joel Spencer. "Sharp concentration of the chromatic number on random graphs $G_{n,p}$." Combinatorica 7(1): 121–129 (1987).

[58] C.J. Thompson. "Inequality with applications in statistical mechanics." Journal of Mathematical Physics, 6: 1812-1813 (1965).

[59] Avi Widgerson and David Xiao. "Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications." Theory of Computing 4: 53-76 (2008).

[60] Andreas Winter. "Extrinsic and Intrinsic Data in Quantum Measurements: Asymptotic Convex Decomposition of Positive Operator Valued Measures." Communications in Mathematical Physics 244(1):157-185 (2004).