

Spectral Gaps of Random Graphs and Applications

Christopher Hoffman^{1,*}, Matthew Kahle², and Elliot Paquette²

¹Department of Mathematics, University of Washington, Seattle, Washington 98195 USA and ²Department of Mathematics, The Ohio State University, Columbus, Ohio 43210 USA

**Correspondence to be sent to: e-mail: hoffman@math.washington.edu*

We study the spectral gap of the Erdős–Rényi random graph through the connectivity threshold. In particular, we show that for any fixed $\delta > 0$ if

$$p \geq \frac{(1/2 + \delta) \log n}{n},$$

then the normalized graph Laplacian of an Erdős–Rényi graph has all of its nonzero eigenvalues tightly concentrated around 1. This is a strong expander property. We estimate both the decay rate of the spectral gap to 1 and the failure probability, up to a constant factor. We also show that the $1/2$ in the above is optimal, and that if $p = \frac{c \log n}{n}$ for $c < 1/2$, then there are eigenvalues of the Laplacian restricted to the giant component that are separated from 1. We then describe several applications of our spectral gap results to stochastic topology and geometric group theory. These all depend on Garland's method [24], a kind of spectral geometry for simplicial complexes. The following can all be considered to be higher-dimensional expander properties. First, we exhibit a sharp threshold for the fundamental group of the Bernoulli random 2-complex to have Kazhdan's property (T). We also obtain slightly more information and can describe the large-scale structure of the group just before the (T) threshold. In this regime, the random fundamental group is with high probability the free product of a (T) group with a free group, where the free group has one generator for every

Received July 18, 2017; Revised February 28, 2019; Accepted March 15, 2019
Communicated by Prof. Assaf Naor

isolated edge. The (T) group plays a role analogous to that of a “giant component” in percolation theory. Next we give a new, short, self-contained proof of the Linial–Meshulam–Wallach theorem [35, 39], identifying the cohomology-vanishing threshold of Bernoulli random d -complexes. Since we use spectral techniques, it only holds for \mathbb{Q} or \mathbb{R} coefficients rather than finite field coefficients, as in [35] and [39]. However, it is sharp from a probabilistic point of view, providing for example, hitting time type results and limiting Poisson distributions inside the critical window. It is also a new method of proof, circumventing the combinatorial complications of cocycle counting. Similarly, results in an earlier preprint version of this article were already applied in [33] to obtain sharp cohomology-vanishing thresholds in every dimension for the random flag complex model.

1 Introduction

Studying the spectral properties of random matrices has played a central role in probability theory ever since Wigner’s paper establishing the semicircular law for symmetric matrices with independent entries of equal variance [44]. The theory of these matrices is rich and welldeveloped, and its techniques and theorems provide great insight into the adjacency matrices of random graphs.

In this paper we study the normalized Laplacian matrix of a Bernoulli (also Erdős–Rényi) random graph $G(n, p)$, which has n vertices and whose every edge is included independently with probability $p = p(n)$. For a connected graph G , the normalized Laplacian has smallest eigenvalue $\lambda_1 = 0$, and the remainder of its eigenvalues $\{\lambda_i\}_{i=2}^n$ lie in the interval $0 < \lambda_i \leq 2$. The spectral gap, λ_2 , is the principal quantity of interest in many applications, and it has received much attention in the literature [8–10, 12].

Our focus is on typical behavior of random graphs for large values of n . So, we will use the terminology *with high probability* (w.h.p.) as a qualifier for a statement holds with probability tending to 1 as n tends to infinity. We will also use the expression *with overwhelming probability*, meaning the statement holds with failure probability smaller than $O(n^{-C})$ for all $C > 0$.

We will make use of the Landau notations $O, o, \omega, \Omega, \Theta$ in the asymptotic sense, so that $f = O(g)$ means f/g is eventually bounded above as $n \rightarrow \infty$ and $f = o(g)$ means f/g tends to 0 as $n \rightarrow \infty$. Also, $f = \omega(g)$ means $g = o(f)$ and $f = \Omega(g)$ means $g = O(f)$. Finally, we will use $f = \Theta(g)$ to mean $f = O(g)$ and $f = \Omega(g)$.

We will also make use of the notion of *thresholds*. A function $f = f(n)$ is said to be a threshold for a property \mathcal{P} if $p = \omega(f)$ implies $G \in \mathcal{P}$ w.h.p. and $p = o(f)$ implies $G \notin \mathcal{P}$ w.h.p. Such a threshold is only defined up to n -independent scalar multiples. If there is a function $g = o(f)$ so that $p \geq f + g$ implies $G \in \mathcal{P}$ w.h.p. and $p \leq f - g$ implies $G \notin \mathcal{P}$ w.h.p. the threshold is *sharp*. If no such g exists, the threshold is *coarse*.

A fundamental result of random graph theory is that every nontrivial monotone property has a threshold [21], which need not be sharp. For example, the appearance of triangles in $G(n, p)$ has the threshold $1/n$, which is coarse. On the other hand, the Erdős–Rényi theorem shows that $\log n/n$ is the sharp threshold for connectivity of the graph. Similarly, we will need that $\frac{1}{2} \log n/n$ is the sharp threshold for the graph to consist only of one *giant component* \tilde{G} and isolated vertices, which is an easy extension of the Erdős–Rényi theorem. We will use \tilde{G} to denote the largest connected component of $G(n, p)$, which is well-defined w.h.p. for $p = \omega(1/n)$ (see [31] for a detailed discussion or Lemma 5.8).

For the Erdős–Rényi graph, as we shall show, the eigenvalues $\{\lambda_i\}_{i=2}^n$ tend to cluster around 1, and hence we define $\lambda(G) = \max_{i \neq 1} |1 - \lambda_i|$. The quantity $1 - \lambda(G)$ is sometimes referred to as the *absolute gap*. The methods in the previous papers are successful in establishing the correct order for $\lambda(G)$ of $C(np)^{-1/2}$ when the density of edges is sufficiently large, but they do not extend to p very near the connectivity threshold $\log n/n$.

Our main result on spectral gaps are contained in the following two theorems.

Theorem 1.1. Fix $\delta > 0$ and let $p \geq (\frac{1}{2} + \delta) \log n/n$. Let $d = p(n-1)$ denote the expected degree of a vertex. For every fixed $\epsilon > 0$, there is a constant $C = C(\delta, \epsilon)$, so that

$$\lambda(\tilde{G}) < \frac{C}{\sqrt{d}}$$

with probability at least $1 - Cn \exp(-(2 - \epsilon)d) - C \exp(-d^{1/4} \log n)$.

This result improves on a number of previous results. These earlier results are discussed in more detail in Section 2. In brief, the state of the art is due to Coja–Oghlan [12] who obtains gap $1 - O(d^{-1/2})$ for $p \geq C \log n/n$, where $C > 0$ is a sufficiently large constant. We are able to extend this to $C = 1$, and appropriately modifying the statement for the giant component, we extend this to $C = \frac{1}{2}$.

We note that Theorem 1.1 is vacuous for $p \leq \frac{1}{2} \log n/n$. Indeed, the next result shows that for smaller values of p , the gap is no longer $1 - o(1)$.

Theorem 1.2. For p satisfying $p = \omega(\sqrt{\log n}/n)$ and $p \leq \frac{1}{2} \log n/n$,

$$\lambda(\tilde{G}) \geq \frac{1}{2},$$

w.h.p.

For $p = O(\sqrt{\log n}/n)$, Fountoulakis and Reed [19] show that the mixing time is large, and hence provide a lower bound for $\lambda(\tilde{G})$ in this regime. So $G(n, p)$ has $\lambda(\tilde{G})$ bounded away from 0, but at $\frac{1}{2} \log n/n$ there is a phase transition, and at this point $\lambda(\tilde{G}) = o(1)$. We in fact prove a slightly stronger result than Theorem 1.2 in Section 5 (c.f. Lemma 4.2).

We also consider an Erdős–Rényi process version (see Section 6 for definitions) of the spectral gap theorem. In particular, we show that if random edges are added one at a time, at the moment of connectivity the random graph already has spectral gap $1 - o(1)$. More precisely, we have the following.

Theorem 1.3. Let τ_c be the connection time for the Erdős–Rényi graph process $G(n, m)$. Then there is a constant C so that w.h.p.

$$\lambda(G(n, \tau_c)) \leq C/\sqrt{\log n}.$$

This theorem follows immediately from Theorem 6.1 in the 2D case: that theorem shows that the largest component of the Erdős–Rényi graph process has gap $\lambda(\tilde{G}) \leq C/\sqrt{\log n}$ for all time after $(\frac{1}{4} + \delta)n \log n$ edges have been added, w.h.p. Hence, at the connection time τ_c , which occurs when about $(\frac{1}{2})n \log n$ edges have been added, $\lambda(G) = \lambda(\tilde{G}) \leq C/\sqrt{\log n}$.

1.1 Applications to stochastic topology

As we will see, Theorem 1.1 is useful in the study of random topological spaces and random groups. We now provide several examples where this theorem yields sharp results. All of these new results depend on the combination of the spectral gap theorem with “Garland’s method” and its refinements by Ballman and Świątkowski [3], and by Żuk [45, 46].

- **Kazhdan’s property (T).** Linial and Meshulam [35] introduce an analogous measure $Y_2(n, p)$ to the binomial random graph for random 2D simplicial complexes. This is the probability distribution on all simplicial complexes with vertex set

$[n] = \{1, 2, \dots, n\}$, with complete 1-skeleton (i.e., with all possible $\binom{n}{2}$ edges), and such that each of the $\binom{n}{3}$ possible 2D faces are included independently with probability p . We use the notation $Y \sim Y_2(n, p)$ to indicate a complex drawn from this distribution. We will call an edge *isolated* if no triangle contains it.

We prove here a structure theorem for the random fundamental group, for a certain range of p .

Theorem 1.4. Suppose $\delta > 0$ is fixed,

$$p \geq \frac{(1 + \delta) \log n}{n},$$

and $Y \sim Y_2(n, p)$. Then w.h.p. $\pi_1(Y)$ is isomorphic to the free product of a (T) group G , and a free group F , where the free group F has one generator for every isolated edge in Y .

As a corollary, we also show that the threshold for $\pi_1(Y)$ to have property (T) agrees precisely with the homology-vanishing threshold found by Linial and Meshulam [35]. For the proof, along with further details and explanation, see Section 3.2.

It might be that $\pi_1(Y)$ is a free product of a (T) group and a free group for smaller p . The most interesting conjecture about the structure of $\pi_1(Y)$ might be the birth of a giant (T) free factor at $p = c/n$ for some constant $c \approx 2.7538$. This is the same point as the homological phase transition studied by Linial and Peled [36].

• **Random d -dimensional simplicial complexes.** Meshulam and Wallach further generalize the 2D model to random d -dimensional complexes $Y_d(n, p)$ [39]. Their main result is that $p = d \log n/n$ is a sharp threshold for vanishing of cohomology $H^{d-1}(Y, \mathbf{k})$ where \mathbf{k} is a finite field or a field of characteristic 0. The proof requires delicate cocycle counting arguments.

The new spectral gap results give a new proof of the Meshulam–Wallach theorem, in the case that k is a field of characteristic 0. The Meshulam–Wallach theorem is stronger topologically, since homology vanishing over a finite field implies vanishing over \mathbb{Q} . However, our new proof is very short (given the spectral gap theorem), and the result is slightly sharper probabilistically. For example, we obtain hitting time results in an accompanying stochastic growth process (see Corollaries 3.3 and 3.9 for representative examples of “hitting time results”), and also we recover a simple proof of the Poisson distribution of Betti numbers in the critical window (Corollary 3.4).

Gundert and Wagner show that the Laplacian on $(d - 1)$ -forms in a random d -complex has a large spectral gap for $p \geq C_d \log n/n$ for some sufficiently large C_d [28]. Combining their argument with the results in this paper would yield a hitting time result, and in particular this shows that the gap for these higher Laplacians is already large for $p \geq d \log n/n$.

Parzanchevski, Rosenthal, and Tessler [41] combine Gundert and Wagner’s argument with earlier work of Pach [40] to show that for $p \geq C_d \log n/n$, w.h.p. Y has the “geometric overlap” property (a sequence of d -dimensional simplicial complexes S_n with F_n d -dimensional faces has the *geometric overlap* property if there exists a constant $\lambda > 0$ so that for every geometric map: $S_n \rightarrow \mathbb{R}^d$ (i.e., affine linear on each face), there exists a point $p \in \mathbb{R}^d$ that lies in the image of at least λF_n d -faces. See, for example, recent work by Gromov and collaborators in [26], [27], and [20]). It also seems possible to use the new spectral gap results to sharpen this result, and show that in the process version of the random complex, random d -complexes already have the geometric overlap property as soon as they are pure d -dimensional.

As far as we can tell, these suggested sharpenings of the main theorems in [28] and [14] are not written down anywhere, and we do not further elaborate on them in this note. It seems that these sharper results only depend on substituting our Theorem 1.1 for earlier results on spectral gap $G(n, p)$.

- **Triangular random groups.** Antoniuk *et al.* [1] study the phase transitions that occur in the triangular model of random groups. Similarly, by using our spectral gap results, their results can be strengthened, for example, to show a hitting time result.
- **Random flag complexes.** Let $X(n, p)$ denote the random clique complex, that is, the maximal simplicial complex, with respect to inclusion of faces, whose 1-skeleton is given by an Erdős–Rényi graph $G(n, p)$.

Combining the spectral gap theorem from an earlier version of this paper with Garland’s method, similar cohomology vanishing results were recently obtained for $X(n, p)$ by the 2nd author in [33]. Combining with several earlier results [32], as a corollary this shows that for every $d \geq 3$, there is a wide range of p for which $X(n, p)$ is rationally homotopy equivalent to a bouquet of d -dimensional spheres (a simplicial complex is *rationally homotopy equivalent* to a bouquet of d spheres if it is simply connected and all of its nontrivial reduced, rational, homology is in degree d).

- **Random right-angled Coxeter groups.** Group cohomology of random right-angled Coxeter groups were studied in [15]. Applying the same techniques as in the random flag complex paper [33], it is shown that for a certain measure and range of parameter, random right-angled Coxeter groups are rational duality groups w.h.p. This is actually

a special case of a more general statement that shows that the same holds for random graph products of finite groups.

1.2 Organization

Section 2 contains the background about the spectrum of the normalized Laplacian of Erdős–Rényi random graphs. Section 3 does the same for our applications of our spectral results to random topology. In Section 4 we show how to transfer adjacency matrix estimates to the normalized Laplacian under some assumptions on the structure of the graph. In Section 5 we show that an Erdős–Rényi graph satisfies these structural conditions w.h.p. In Section 6 we show that the Linial–Meshulam process has large gap in a local spectral sense. In Section 7 we show how to apply the Ballman–Świątkowski criterion to prove the structure theorem for rational cohomology, and in Section 8 we show how to apply Żuk’s criterion to prove the structure theorem for the fundamental group. In Section 9 we apply the Kahn–Szemerédi machinery to show that the adjacency matrix of the Erdős–Rényi graph has a gap of the correct order for any p with $p = \Omega(\log n/n)$. Finally, we include one appendix that proves the precise versions of the tail bounds for binomial variables that we use.

2 Background: Spectra of Random Graphs

There are multiple common notions of spectra of a graph. The most elementary definition is given by the eigenvalues of the adjacency matrix A . The subjects of our main theorems are the eigenvalues of the normalized Laplacian L (see (1) for a precise definition). When the graph is regular, these two notions of spectra are just shifted rescalings of one another.

Appropriately, when the graph is nearly regular, as is the case for $G(n, p)$ with $p = \omega(\log n/n)$, these two spectra behave in nearly the same way. Coarse statements about the spectral gap of $G(n, p)$ in this regime can largely be considered a statement about either spectra, and indeed, the primary method for estimating the gap of L in the setting of Erdős–Rényi graphs is by comparison with A .

We will now give a precise definition of the normalized Laplacian. A good general introduction to the properties of the normalized Laplacian is available in [10]. Let π_+ be the projection map onto the vertices with positive degree, let T be the diagonal matrix of degrees, and let A be the adjacency matrix. The normalized Laplacian is defined as

$$L = \pi_+ - T^{-1/2}AT^{-1/2}, \quad (1)$$

where $T^{-1/2}$ is taken to be 0 in coordinates where the degree is 0. Note that some authors use an alternate definition of normalized Laplacian, with π_+ replaced by Id. We let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$ be the eigenvalues of L .

The principal nontrivial property we will employ about L is that the dimension of the kernel is equal to the number of components of G . An immediate consequence is that for a graph with multiple nontrivial components, $\lambda_2 = 0$. In particular, when $np - \log n \rightarrow -\infty$ the normalized Laplacian has no spectral gap w.h.p. That said, it still makes sense to consider the spectral gap of L restricted to the giant component.

Techniques for estimating eigenvalues

As A has independent and identically distributed (i.i.d.) entries above the diagonal, many off-the-shelf techniques can be applied to it directly. In particular, the original *trace method* bound of Füredi and Komlós [23] can be extended to show that when $p = \omega(\log^6 n/n)$, the 2nd largest eigenvalue of the adjacency matrix of an Erdős–Rényi graph is of smaller order than the largest eigenvalue. Improvements and corrections to this argument brought the bound to $p = \omega(\log^4 n/n)$, [43] and later to as low as $p \gg \log^2 n/n$ [8]. Newer methods have been pursued in [5, 6, 34].

The alternative method of Kahn and Szemerédi [22], first developed for bounding the spectral gap of d -regular graphs, has been adapted quite successfully for estimating the spectral gap in the $p = \Theta(\log n/n)$ regime by Feige and Ofek [18]. In particular, they show that there are constants $c > 0$ and $K > 0$ so that for $p > c \log n/n$, all but the first eigenvalue are at most $K\sqrt{np}$.

One contribution of this paper is a sharpening of this estimate (see Proposition 5.2). Indeed, we show that for *any* $c > 0$, there is a $K > 0$ so that for $p > c \log n/n$, all but the first eigenvalue are at most $K\sqrt{np}$. Conversely, it is easily checked that for $p = o(\log n/n)$, there are many eigenvalues greater in magnitude than \sqrt{np} , coming from the existence of high-degree stars in the graph. Thus, in a sense, we sharpen the Kahn–Szemerédi analysis of the *full* adjacency matrix of $G(n, p)$ to its natural endpoint.

However, our main contribution in this paper is a technique for exactly characterizing *when and why* the extremal eigenvalues of the normalized Laplacian stop tracking the extremal eigenvalues of the adjacency matrix. Throughout the $p = \Theta(\log n/n)$, the extremal eigenvalues of the adjacency matrix do not undergo a phase transition (see Proposition 5.2).

In contrast, for the Laplacian, there is a transition at $p = \log n/n$, before which point the graph has isolated vertices. Each isolated vertex contributes a

0-eigenvalue to the spectra of the Laplacian, but as a consequence of Theorem 1.1, the remaining eigenvalues will be $1 + O(1/\sqrt{np})$ as anticipated. There is a 2nd transition at $p = \frac{1}{2} \log n/n$ below which there are quadruplets of vertices in the giant component on which the induced graph is a path. These quadruplets each contribute an eigenvalue near to $\frac{1}{2}$, but the remainder of the spectra will again be $1 + O(1/\sqrt{np})$. Continuing in this way, we conjecture that there are a whole family of transitions at $\frac{1}{k} \log n/n$ for any natural number k , where the spectral gap of the giant component is asymptotically the spectral gap of a path on k vertices.

2.1 Comparing spectra and the gap theorem proof approach

While it is relatively straightforward to transfer estimates on the gap of A to the gap of L in the $p = \omega(\log n/n)$ regime, Coja-Oghlan [12] sharpens this analysis to show that there are $c > 0$ and $K > 0$ so that for $p \geq c \log n/n$, all but the smallest eigenvalue of L are at most K/\sqrt{np} in modulus w.h.p.

There are some similarities between our approach and the method of Coja-Oghlan [12]. His analysis rests on applying the Kahn–Szemerédi machinery to the adjacency matrix of a sufficiently regular subgraph of $G(n, p)$ and then arguing this core of the graph determines the eigenvalues of the Laplacian of the whole graph. We make a finer analysis of the structure of $G(n, p)$ in the $p = \Theta(\log n/n)$ regime in order to show that in fact the spectra of the adjacency matrix and the spectra of the normalized Laplacian only fail to be comparable when small sparse subgraphs appear.

To bound $\max_{i>1} |1 - \lambda_i|$ it suffices instead to bound the spectrum of what is essentially $I - L$. Given the graph G with vertices $\{1, 2, \dots, n\}$ we define the matrix

$$M_{u,v} = \begin{cases} \frac{1}{\sqrt{\deg(u)}\sqrt{\deg(v)}} & \text{if } u \text{ is adjacent to } v, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if all degrees are positive we have

$$M = T^{-1/2}AT^{-1/2},$$

and it is easily checked that for any vertex set W of a connected component of V , $T^{1/2}\mathbf{1}_W$ is an eigenvector with eigenvalue one.

Set $S = \{x | x^t \mathbf{1} = 0\}$. The standard Kahn–Szemerédi machinery applied to the adjacency matrix shows that

$$|x^t A y| \leq C\sqrt{d}\|x\|\|y\|,$$

where $d = np$, for all $x \in S$ and all $y \in \mathbb{R}^n$, provided $p = \Omega(\log n/n)$.

When $p > (1 + \epsilon)\log n/n$, the comparison is relatively straightforward, by virtue of the fact that w.h.p. all the degrees in the graph are larger than d/M for some sufficiently large M . In particular, this means that $\|T^{-1/2}\| \leq \sqrt{M}/\sqrt{d}$. One must additionally show that $T^{-1/2}\mathbf{1}$ is nearly parallel to $\mathbf{1}$, that is, $T^{-1/2}$ nearly maps the space S to itself. In sum, these two facts show that for $x \in S$, $T^{-1/2}x$ is still nearly in S and has norm $\|T^{-1/2}x\| \leq \sqrt{M}\|x\|/\sqrt{d}$. Thus,

$$|x^t M x| = |(T^{-1/2}x)^t A (T^{-1/2}x)| \approx C\sqrt{d}\|T^{-1/2}x\|^2 \leq CM\|x\|^2/\sqrt{d},$$

giving the desired result.

Likewise, when $p > \frac{\log n + (\log n)^{1/2+\delta} \log \log n}{n}$, the minimal degree of the graph is still at least $d^{1/2+\delta}$ w.h.p. In this case, the $T^{-1/2}$ still nearly maps S to S , but now $\|T^{-1/2}x\| \leq d^{-1/4-\delta/2}$. This allows one to show that

$$\max_{i>1} |1 - \lambda_i| < d^{-\delta},$$

which is essentially the approach taken by an earlier version of this paper.

To get theorems that hold all the way down to below $p = \log n/n$, where the minimum degree drops to 0 an additional argument is needed. This is because it is no longer the case that $\|T^{-1/2}\| = O(1/\sqrt{d})$. The key structure theorem that allows the comparison to go through is an analysis of the graph structure surrounding low-degree vertices. Precisely, we show that near the connectivity threshold, there are no edges between low-degree vertices, and low-degree vertices do not even have shared neighbors (see Proposition 5.3). Thus, they are only connected through the large, high-degree core. This is enough to ensure that the desired spectral properties persist all the way down to around $p \sim 1/2 \log n/n$.

On the other hand, below $p \sim 1/2 \log n/n$, low-degree vertices in the giant component begin to connect w.h.p. Indeed, it is possible to show that there are even two degree 2 vertices that connect to each other and the high-degree core. This is enough to ensure that λ_2 of the giant component is at most a little above $\frac{1}{2}$ and λ_n is at least $\frac{3}{2}$.

2.2 Further discussion

For p satisfying $np - \log n \rightarrow \infty$, we have provided a bound on $\lambda(G)$ that is sharp up to a constant multiplicative factor. For the adjacency matrix in many regimes, much more is known about the behavior of the second largest eigenvalue.

Recall that a Wigner matrix is a symmetric matrix with independent, centered, variance 1 entries above the diagonal. From Wigner's celebrated semicircle law, it can be inferred that the largest eigenvalue of such a matrix is around $2\sqrt{n}$. In fact a much stronger result is known for a large class of Wigner matrices, for which it is seen that

$$n^{1/6}(\lambda_1 - 2\sqrt{n}) \Rightarrow X,$$

where X follows the Gaussian orthogonal ensemble (GOE) Tracy–Widom law. When the entry distributions are Bernoulli(p)—that is, when this is the adjacency matrix of an Erdős–Rényi graph—it was recently shown by Knowles, Erdős, Yau, and Yin [17] that for $p \gg n^{-1/3}$, the analogous results hold for the second largest eigenvalue. One of the limits of comparing the spectra of the adjacency matrix and the Laplacian matrix is that such a fine statement about the spectra does not easily transfer. It is appealing to speculate that at $p \sim \log n/n$, the smallest nonzero eigenvalue of the normalized Laplacian is exactly $1 - (2 - o(1))\sqrt{np}$, consistent with what would be predicted by the semicircle law of the adjacency matrix.

The spectral gap of the normalized Laplacian is strongly related to other probabilistic quantities of the graph, in particular to properties of simple random walk (see [10] for more details) and to the Cheeger constant. Direct analysis of these quantities is also possible, which then implicitly give bounds on the spectral gap. Benjamini *et al.* [7] take a combinatorial approach and study the Cheeger constant (also called isoperimetric constant, or conductance) throughout the evolution of the random graph process. Likewise Fountoulakis and Reed study the mixing time of simple random walk on the giant component through the conductance [19] in the strictly supercritical regime $\frac{1+\epsilon}{n} < p < \frac{\sqrt{\log n}}{n}$. Ding *et al.* [16] studied probabilistic aspects of the graph including the mixing time of simple random walk on the giant component as the graph emerges from the critical window. All these works show that the giant component can be partitioned into a well-connected expanding core together with small (logarithmic size) graphs attached to the core. We also employ a version of this decomposition to analyze the spectral properties of the graph.

3 Random Topology

In [35], Linial and Meshulam introduce an analogous measure $Y_2(n, p)$ to the binomial random graph for random 2D simplicial complexes. This is a probability distribution over all simplicial complexes with vertex set $[n] = \{1, 2, \dots, n\}$ with complete 1-skeleton (i.e., with all possible $\binom{n}{2}$ edges). Each of the $\binom{n}{3}$ possible 2D faces are included independently with probability p . We use the notation $Y \sim Y_2(n, p)$ to indicate a complex drawn from this distribution. Meshulam and Wallach [39] extend this definition to a d -dimensional complex $Y_d(n, p)$, formed by taking the complete $(d - 1)$ -skeleton of the n -vertex simplex, and including d -dimensional faces independently with probability p .

The distributions can be made into stochastic growth processes in a natural way. Let $Y_2(n, m)$ be the random 2-complex that has the uniform distribution over all simplicial complexes with n vertices, $\binom{n}{2}$ edges, and exactly m 2D faces. In the random complex process $\{Y_2(n, m)\}$, faces are added one at a time, uniformly randomly from all faces that have not already been chosen. In the same way, we can define the process $\{Y_d(n, m)\}$ by including d -faces one at a time.

We also define a time-changed version of this process $Y_t^d(n)$, more suitable to working with the binomial complex. Instead of including the faces one at a time, create independent $\text{Exp}(1)$ clocks for every d -face. When one of the clocks rings, include the corresponding face. If we let $p(t) = 1 - e^{-t}$, then $Y_t^d(n)$ has the distribution $Y_d(n, p(t))$.

3.1 Cohomology vanishing

The foundational work on the Linial–Meshulam complexes is a cohomological analogue of the Erdős–Rényi connectivity theorem.

Linial–Meshulam–Wallach theorem. Let \mathbf{k} be any finite field, $d \geq 2$ fixed, $f(n) \rightarrow \infty$ be any slowly growing function, and $Y \sim Y_d(n, p)$. If

$$p \geq \frac{d \log n + f(n)}{n},$$

then w.h.p. $H^{d-1}(Y, \mathbf{k}) = 0$, and if

$$p \leq \frac{d \log n - f(n)}{n},$$

then w.h.p. $H^{d-1}(Y, \mathbf{k}) \neq 0$.

For the case that $d = 2$ and $\mathbf{k} = \mathbb{Z}_2$, this is due to [35], while for the version stated, this is due to [39]. By the universal coefficient theorem, these results imply

the corresponding theorem for the cohomology with \mathbb{Q} coefficients. For \mathbb{Z} coefficients, it is shown by the authors in [30] that for $p \geq 80d \log n/n$, $H^{d-1}(Y, \mathbb{Z}) = 0$ by other techniques. For $d = 2$, work of [37] establishes $2 \log n/n$ as the sharp threshold for vanishing \mathbb{Z} homology.

The threshold $p \sim d \log n/n$ is also the threshold for the existence of isolated $(d-1)$ -faces in the complex, that is, faces that are not included in any d -face. Indeed, the presence of isolated faces is precisely the reason that the cohomology is nonzero below this threshold. In fact, a finer statement can be made about the number of isolated $(d-1)$ -faces.

Lemma 3.1. Let I denote the number of isolated $(d-1)$ faces in $Y_d(n, p)$. Suppose that for fixed c ,

$$p = \frac{d \log n + c + o(1)}{n}.$$

Then I converges in law to $\text{Poisson}(e^{-c}/d!)$.

The proof of this lemma is standard and can be proved in the same manner as the Poisson convergence of the number of isolated vertices in $G(n, p)$. See Proposition 4.13 of [42].

Using spectral techniques, we give a new proof of the Linial–Meshulam–Wallach theorem, although only with \mathbb{Q} or \mathbb{R} coefficients. However, for \mathbb{Q} coefficients, we also sharpen the theorem by proving a process version. More strikingly, this theorem shows that long before the last isolated $(d-1)$ -faces disappear, the only obstruction to vanishing cohomology are those isolated $(d-1)$ -faces. Its proof follows almost immediately from spectral arguments and Garland’s method (see Section 7).

Theorem 3.2. Consider the random complex process $\{Y_t^d(n)\}$. Let I_t denote the number of isolated $(d-1)$ -faces in the complex at time t . Fix any $\delta > 0$ and define t_0 so $p(t_0) = (d-1+\delta) \log n/n$. Then w.h.p. for all time $t \geq t_0$,

$$H_{d-1}(Y_t^d(n), \mathbb{Q}) \cong \mathbb{Q}^{I_t}.$$

As w.h.p. $I_{t_0} > 0$ we immediately get the following hitting time corollary.

Corollary 3.3. Consider the random complex process $\{Y_d(n, m)\}$. Let

$$M_1 = \min\{m \mid Y_d(n, m) \text{ has no isolated } (d-1) \text{ - dimensional faces}\},$$

and let

$$M_2 = \min\{m \mid H^{d-1}(Y_d(n, m), \mathbb{Q}) = 0\}.$$

Then w.h.p. $M_1 = M_2$.

Further, it is standard to show at this point that the Betti numbers are asymptotically Poisson.

Corollary 3.4. Suppose that for fixed c ,

$$p = \frac{d \log n + c + o(1)}{n}.$$

Then $b_{d-1}(Y_d(n, p))$ converges in law to $\text{Poisson}(e^{-c}/d!)$.

Note that this follows immediately from Lemma 3.1 and Theorem 3.2.

3.2 The fundamental group

For the 2D complex, a fair bit is known about the fundamental group $\pi_1(Y)$. Babson and the 1st two authors find the threshold for the fundamental group to be trivial [2].

Theorem 3.5 (Babson–Hoffman–Kahle). If $p = n^{-\alpha}$, where $\alpha < 1/2$ then w.h.p. $\pi_1(Y)$ is a nontrivial word hyperbolic group. If $p \geq n^{-1/2} \log(n)$ then $\pi_1(Y)$ is trivial.

Cohen *et al.* [11] show that if $p = o(1/n)$, then w.h.p. $\pi_1(Y)$ is free. Finally, Costa and Farber describe the cohomological dimension $\text{cd } \pi_1(Y)$ in various regimes [13, 41].

Theorem 3.6 (Costa–Farber). Let $Y \sim Y_2(n, p)$, and set $p = n^{-\alpha}$.

- (1) If $\alpha > 1$ then w.h.p. $\text{cd } \pi_1(Y) = 1$;
- (2) if $1 > \alpha > 3/5$ then w.h.p. $\text{cd } \pi_1(Y) = 2$; and
- (3) if $3/5 > \alpha > 1/2$ then w.h.p. $\text{cd } \pi_1(Y) = \infty$.

For the 2D complex, we combine the new spectral results with Garland’s method to show a threshold theorem for $\pi_1(Y)$ to have property (T). A group G is said to have property (T) if every unitary action of G on a Hilbert space that has almost invariant vectors also has a nonzero invariant vector. The 1st explicit examples of expanders, due to Margulis, were constructed using Cayley graphs on quotients of (T) groups such

as $SL(3, \mathbb{Z})$ [38]. Conversely, expansion properties of some graphs associated with the generating set of a group can imply property (T) (see [46]).

Property (T) has found use in many different areas of mathematics. For example, groups with property (T) lead to good mixing properties in ergodic theory—a process that mixes slowly must leave some subsets almost invariant. In particular, if a group Γ has property (T), then every ergodic Γ system is also strongly ergodic [25]. See the monograph [4] for a comprehensive overview of property (T).

We recall for convenience the statement of Theorem 1.4.

Theorem. Suppose $\delta > 0$ is fixed,

$$p \geq \frac{(1 + \delta) \log n}{n},$$

and $Y \sim Y_2(n, p)$. Then w.h.p. $\pi_1(Y)$ is isomorphic to the free product of a (T) group G , and a free group F , where the free group F has one generator for every isolated edge in Y .

Theorem 1.4 might be viewed as a group-theoretic analogue of the fact that for $p \geq (1/2 + \delta) \log n/n$, the random graph $G \sim G(n, p)$ is w.h.p. a giant component, which is an expander, and isolated vertices.

We anticipate that the true threshold for $\pi_1(Y)$ being the free product of a free group and a nontrivial (T) group is much lower, and that it occurs in the range $p = \Theta(1/n)$. The significance of the threshold $\log n/n$ is that this is the threshold at which the free group is generated by isolated edges.

For example, if $p = \delta \log n/n$ with $0 < \delta < 1$ fixed, then w.h.p. there exists a triangle abc in $Y_2(n, p)$ such that edges ab and ac are not contained in any other triangle. In other words, the edge bc is a connected component in the link of vertex a . In this case, the edge ab and triangle abc can be collapsed by an elementary collapse. This is a homotopy equivalence—after the collapse, the edge ac is a generator of a free \mathbb{Z} factor in $\pi(Y)$, but before the collapse there is no isolated edge generating this element of the group.

On the other hand, this is also the point our argument in Section 8 ceases to apply. To apply Zuk's criterion in Section 8, we first delete all isolated edges and the resulting complex has connected vertex links with good expansion properties. In the case above, the vertex link a is not connected, even after such deletions.

We have the following corollary of Theorem 1.4, which shows that the threshold for property (T) is the same as the Linial–Meshulam theorem for vanishing of $\mathbb{Z}/2$ -homology.

Corollary 3.7. Let $\omega \rightarrow \infty$ as $n \rightarrow \infty$, and $Y \sim Y_2(n, p)$. If

$$p \geq \frac{2 \log n + \omega}{n}$$

then $\mathbb{P}[\pi_1(Y) \text{ has property (T)}] \rightarrow 1$.

We also describe a process version of this structure theorem that holds below the connectivity threshold.

Theorem 3.8. Consider the random complex process $\{Y_t^d(n)\}$. Let \tilde{F}_t be a free group with the number of generators equal to the number of isolated edges in the complex $Y_t^d(n)$. Fix any $\delta > 0$ and define t_0 so $p(t_0) = (1 + \delta) \log n/n$. Then w.h.p. for all $t \geq t_0$,

$$\pi_1(Y_2(n, p(t))) \cong G_t * \tilde{F}_t$$

where G_t has property (T).

Note that Theorem 1.4 follows immediately from this. As the number of isolated edges at time t_0 is positive w.h.p, we get the following hitting time corollary.

Corollary 3.9. Consider the random complex process $\{Y_2(n, m)\}$. Let

$$M_1 = \min\{m \mid Y_2(n, m) \text{ has no isolated edges}\},$$

and let

$$M_2 = \min\{m \mid \pi_1(Y_2(n, m)) \text{ is (T)}\}.$$

Then w.h.p. $M_1 = M_2$.

Remark 3.10. We can additionally give an explicit Kazhdan pair for the (T) group. Setting S to be the canonical generating set based at vertex 1, that is, all loops cycles of the form $1 \rightarrow x \rightarrow y \rightarrow 1$ for distinct vertices x and y , then $(S, \sqrt{2}(1 - o(1)))$ is a Kazhdan pair (see Remark 5.5.3 of [4]).

4 Spectral Estimates

In this section we give some conditions on an arbitrary graph G on n vertices that facilitate a large spectral gap. Fix positive constants C_1, C_2, C_3 and M . In this section d can be any function of n with $d = d(n) \geq 1$, and this is always satisfied by $d = (n-1)p$, the convention taken in other sections.

Recall that T is the diagonal matrix of degrees. Let W denote the set of vertices x for which $\deg x > 0$ and I be the number of isolated vertices in the graph. For any set of vertices S , let $\mathbf{1}_S$ denote the vector that is one in every coordinate corresponding to S and 0 elsewhere. Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of the normalized Laplacian $L[G]$, so that $\lambda_1 = \lambda_2 = \dots = \lambda_{I+1} = 0$. We also define a set of vertices of small degree. Let

$$\mathfrak{N}_M = \{v \in V : \deg(v) \leq d/M\}. \quad (2)$$

We now define four conditions that will ensure a spectral gap.

- (1) **Bounded degree condition (b.d.c.)** Every vertex has degree at most $C_1 d$.
- (2) **Adjacency matrix**

$$\sup_{\substack{\|x\|=1, x^t \mathbf{1}=0 \\ \|y\|=1}} |x^t A y| \leq C_2 \sqrt{d}.$$

- (3) **Fuzz** There are no edges between vertices of \mathfrak{N}_M , $|\mathfrak{N}_M| \leq \frac{n}{2}$, and

$$\max_{u \in \mathfrak{N}_M^c} e(u, \mathfrak{N}_M) \leq 1,$$

where $e(U, V)$ denotes the number of edges between sets of vertices U and V .

- (4) **Parallel eigenspaces**

$$\sup_{\substack{\|x\|=1, \\ x^t T^{1/2} \mathbf{1}_W=0}} |x^t T^{-1/2} \mathbf{1}_{\mathfrak{N}_M^c}| \leq C_3 \frac{\sqrt{n}}{d}.$$

The final condition states that a vector x that is orthogonal to the kernel of L will not have such a large component in the direction of the principal eigenvector of $T^{-1/2}A$. The vector $\mathbf{1}_{\mathfrak{N}_M^c}$ can be considered as a good approximation to this principal right eigenvector. Otherwise said, the 0-eigenspace of L and the principal right eigenspace of $T^{-1/2}A$ are nearly parallel.

With these definitions we can now state our main result on spectral gaps.

Lemma 4.1. Let G be a graph on n vertices and let C_1, C_2, C_3 and M be constants. If G satisfies the four conditions above then there is a constant $C = C(C_1, C_2, C_3, M)$ so that

$$\max_{i>I+1} |1 - \lambda_i| < \frac{C}{\sqrt{d}}.$$

Proof. Let W be the set of vertices x for which $\deg x > 0$. By the spectral theorem, L admits a basis of orthogonal eigenvectors. Let v be a normalized eigenvector of L corresponding to an eigenvalue λ_i with $i > I + 1$. Setting l_1, l_2, \dots, l_I to be the isolated vertices, a basis for the kernel of L is given by $\{T^{1/2}\mathbf{1}, \delta_{l_1}, \delta_{l_2}, \dots, \delta_{l_I}\}$, where δ_a is 1 in the a^{th} coordinate and 0 elsewhere. As v is orthogonal to all of these, it is orthogonal to $T^{1/2}\mathbf{1}_W$. Hence,

$$|1 - \lambda_i| = \left| v^t T^{-1/2} A T^{-1/2} v \right| \leq \sup_{\substack{\|x\|=1, \\ x^t T^{1/2} \mathbf{1}_W = 0}} \left| x^t T^{-1/2} A T^{-1/2} x \right|.$$

As this holds for all such $i > I + 1$, it suffices to bound the right-hand side.

Orthogonally decompose $T^{-1/2}x = u + v$, where u is supported on vertices in \aleph_M^c and v is supported on vertices in \aleph_M . Further decompose $u = u_0 + u_1$ by letting u_1 be the projection of u along $\mathbf{1}_{\aleph_M^c}$. Expanding the quadratic form, we may write

$$\left| x^t T^{-1/2} A T^{-1/2} x \right| \leq 2|u_0^t A u| + |u_1^t A u_1| + |v^t A v| + 2|v^t A u|. \quad (3)$$

Each of these terms will be seen to have the right order bound, completing the proof.

As $u_0 \perp \mathbf{1}_{\aleph_M^c}$ and is supported only on \aleph_M^c , we have that $u_0 \perp \mathbf{1}$. By the definitions of \aleph_M and x , we have that

$$\|u_0\|^2 \leq \|u\|^2 = \sum_{i \in \aleph_M^c} \frac{|x_i|^2}{\deg i} \leq \frac{M}{d}.$$

Hence, by the adjacency matrix condition and the above equation we have that

$$|u_0^t A u| \leq C_2 \sqrt{d} \|u_0\| \|u\| \leq \frac{C_2 M}{\sqrt{d}}. \quad (4)$$

As u_1 is the projection of u along $\mathbf{1}_{\aleph_M^c}$, we have

$$u_1 = (u^t \mathbf{1}_{\aleph_M^c}) \frac{\mathbf{1}_{\aleph_M^c}}{|\aleph_M^c|} = (x^t T^{-1/2} \mathbf{1}_{\aleph_M^c}) \frac{\mathbf{1}_{\aleph_M^c}}{|\aleph_M^c|}.$$

Because $|\mathbb{N}_M^c| \geq \frac{n}{2}$, the parallel eigenspaces condition implies that we have $\|u_1\| \leq \frac{\sqrt{2}C_3}{d}$. The norm of A is at most the maximum degree of the graph, and by the b.d.c. this is at most $C_1 d$. Hence, we get that

$$|u_1^t A u_1| \leq \frac{2C_1 C_3^2}{d} \quad (5)$$

For the 3rd term, we note that by the \mathbb{N}_M condition there are no edges between vertices of \mathbb{N}_M , and hence

$$v^t A v = 0. \quad (6)$$

Finally, we may expand $v^t A u$ as

$$v^t A u = \sum_{i \in \mathbb{N}_M} \frac{x_i}{\sqrt{\deg i}} \sum_{\substack{j \in \mathbb{N}_M^c \\ j \sim i}} u_j.$$

By Cauchy–Schwarz, this is bounded by

$$|v^t A u|^2 \leq \sum_{i \in \mathbb{N}_M} \frac{1}{\deg i} \left(\sum_{\substack{j \in \mathbb{N}_M^c \\ j \sim i}} u_j \right)^2 \leq \sum_{i \in \mathbb{N}_M} \sum_{\substack{j \in \mathbb{N}_M^c \\ j \sim i}} (u_j)^2.$$

Now each $j \in \mathbb{N}_M^c$ has at most one neighbor in \mathbb{N}_M , and hence we have

$$|v^t A u| \leq \|u\| = \frac{\sqrt{M}}{\sqrt{d}}. \quad (7)$$

Plugging (4), (5), (6), and (7) into (3) completes the proof. \blacksquare

In the remainder of this section we prove a condition on a graph that will imply an upper bound on the spectral gap. This lemma shows that our previous argument breaks down when the set \mathbb{N}_M fails to be isolated.

Lemma 4.2. Suppose that H is a connected graph and that there are vertices u, v, w, x for which the induced graph on u, v, w, x is a path with endpoints u and x . Suppose further that $\deg v = \deg w = 2$ and $\deg u, \deg x \geq m$. Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{|H|}$ be the eigenvalues of the normalized Laplacian $L[H]$, then

$$\lambda_{|H|} \geq \frac{3}{2}$$

and

$$\lambda_2 \leq \frac{1}{2} + O(1/\sqrt{m}).$$

Proof. For each case, we construct an appropriate approximate eigenvector. For the first, consider the vector f with $f(v) = 1$, $f(w) = -1$ and $f(y) = 0$ for all other y . This vector is orthogonal to $T^{1/2}\mathbf{1}$, the 1st eigenvector of L . Now $T^{-1/2}f$ is just $f/\sqrt{2}$ while $f^t A f = -2$. Thus,

$$\frac{f^t T^{-1/2} A T^{-1/2} f}{\|f\|^2} = -\frac{1}{2},$$

and so $\lambda_{|S|} \geq 1 - \frac{-1}{2} = \frac{3}{2}$.

For the lower bound let f be given by $f(v) = f(w) = 1/\sqrt{2}$ while $f(x) = -1/\sqrt{\deg x}$ and $f(u) = -1/\sqrt{\deg u}$. Then we have $f \perp T^{1/2}\mathbf{1}$. By direct computation,

$$f^t T^{-1/2} A T^{-1/2} f = \frac{1}{2} - \frac{1}{\deg x} - \frac{1}{\deg u},$$

while

$$\|f\|^2 \leq 1 + \frac{1}{\deg x} + \frac{1}{\deg u}.$$

Thus, combining everything, we have that

$$\lambda_2 \leq 1 - \frac{\frac{1}{2} - \frac{2}{m}}{\sqrt{1 + \frac{2}{m}}} = \frac{1}{2} + O(1/\sqrt{m}).$$

■

5 Probability Bounds

In this section we show various estimates on $G(n, p)$, which when combined with the deterministic lemmas on spectral gaps from the previous section, will complete the proofs of Theorems 1.1 and 1.2. In this section we again use that $d = (n - 1)p$ is the expected degree of a vertex.

Lemma 5.1. For each $\delta > 0$ and $m \geq 0$, there is a constant $C = C(\delta, m)$ so that the following conditions hold with probability at least $1 - C \exp(-md)$ and $1 - C \exp(-md^{1/4} \log n)$, respectively, in $G(n, p)$ with $p \geq \delta \log n/n$.

- (1) **B.d.c** Every vertex has degree at most Cd .
- (2) **Discrepancy** For every pair of vertex sets A and B , letting $e(A, B)$ denote the number of edges between the sets and $\mu(A, B) = \frac{|A||B|d}{n}$, one of

- (a) $\frac{e(A,B)}{\mu(A,B)} \leq C$
 - (b) $e(A,B) \log \frac{e(A,B)}{\mu(A,B)} \leq C(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}$
 - (c) $|A| \leq d^{1/4}/100, |B| \leq d^{1/4}/100$
- (c) occurs.

Both of these bounds are consequences of tail bounds of binomial variables, and they are relatively standard in the literature (see, e.g., [12, 18, 22]). This one differs in that we look for more control over the order of decay of the failure probability.

Proposition 5.2. For each $\delta > 0$ and $m \geq 0$, there is a constant $C = C(\delta, m)$ sufficiently large so that if $p \geq \delta \log n/n$ then

$$\sup_{\substack{\|x\|=1, x^t \mathbf{1}=0 \\ \|y\|=1}} |x^t A y| \leq C\sqrt{d}$$

with probability at least $1 - C \exp(-md^{1/4} \log n) - C \exp(-md)$.

This follows from the standard Kahn–Szemerédi argument, and it is essentially proven in both Feige and Ofek [18] and the original Friedman, Kahn, and Szemerédi paper [22]. This version has a sharper estimate on the failure probability than [18], which in turn follows from Lemma 5.1. We will delay the proof of both this and the previous lemma to Section 9.

Additionally, the b.d.c. is needed to make estimates about low-degree vertices. Recall the definition of \aleph_M from (2). We show that this set is both small and structurally very simple for sufficiently large M .

Proposition 5.3. For each $\delta > 0$ and each $\epsilon > 0$, there is an $M = M(\delta, \epsilon) > 1$ such that for $p \geq (\frac{1}{2} + \delta) \log n/n$, $G(n, p)$ satisfies the following:

- (1) $|\aleph_M| < n/(100d)$;
- (2) \aleph_M is an independent set; and
- (3) $\max_{u \in \aleph_M^c} e(u \aleph_M) \leq 1$

with probability at least $1 - Cn \exp(-(2 - \epsilon)d) - C \exp(-cn)$ for some absolute constant $c > 0$.

Proof. (i) We start by estimating the size of \aleph_M , which we do by a simple union bound. Namely by symmetry we have

$$\Pr[|\aleph_M| \geq k] \leq \binom{n}{k} \Pr[\deg u_i \leq d/M, 1 \leq i \leq k].$$

Let S be the set of vertices u_{k+1}, \dots, u_n , then we have

$$\Pr[\deg u_i \leq d/M, 1 \leq i \leq k] \leq \Pr[e(u_i S) \leq d/M, 1 \leq i \leq k],$$

which are now independent $\text{Binom}(n-k, p)$ variables. Applying Lemma A.1, we get

$$\log \Pr[|\mathfrak{N}_M| \geq k] \leq k \left[\left(1 + \log \frac{n}{k}\right) - (d - kp) + \frac{d}{M} (1 + \log(M)) \right].$$

Setting $k = \lfloor n/(100d) \rfloor$, we may make M sufficiently large that

$$\left(1 + \log \frac{n}{k}\right) - (d - kp) + \frac{d}{M} (1 + \log(M)) \leq -\frac{d}{2}$$

for all $n \geq n_0(\delta)$. Hence, we have that $|\mathfrak{N}_M| < n/(100d)$ with probability at least $1 - O \exp(-cn)$ for some absolute constant $c > 0$.

(ii) We begin by bounding the probability that there is an edge between any two vertices of \mathfrak{N}_M . Note that we may assume that $d < n/100$, lest $\mathfrak{N}_M = \emptyset$ by the previous bound.

From the union bound and symmetry, we have that

$$\Pr[\mathfrak{N}_M \text{ is not an independent set}] \leq n^2 \Pr[v \in S, w \in S, v \leftrightarrow w].$$

Thus, it suffices to compute this probability, which we do by conditioning $\deg v = d_1$ and $\deg w = d_2$. Note that the law of the neighborhood N of $\{v, w\}$ under this conditioning is *not* uniform over all such neighborhoods. For a possible neighborhood H of $\{v, w\}$, let $E(H)$ denote the number of edges in this neighborhood. Then we have that

$$\Pr[N = H | \deg v = d_1, \deg w = d_2] = \frac{1}{Z} \left(\frac{p}{1-p} \right)^{E(H)},$$

for a suitable normalization constant Z .

Thus, we have that

$$\begin{aligned} \Pr[v \leftrightarrow w | \deg v = d_1, \deg w = d_2] &\leq \frac{\Pr[v \leftrightarrow w | \deg v = d_1, \deg w = d_2]}{\Pr[v \leftrightarrow w | \deg v = d_1, \deg w = d_2]} \\ &= \frac{1-p}{p} \frac{\binom{n-2}{d_1-1} \binom{n-2}{d_2-1}}{\binom{n-2}{d_1} \binom{n-2}{d_2}}. \end{aligned}$$

As we consider only d_1 and d_2 that are less than d/M , and as $d < n/100$, we may bound this as Cd/n for some absolute constant C . It remains to estimate the probability that both v and w are in \aleph_M . Hence, we have

$$\Pr[\deg v \leq d/M, \deg w \leq d/M] \leq \Pr[X \leq d/M]^2,$$

where $X \sim \text{Binom}(n-2, p)$. Applying Lemma A.1, we have that

$$\Pr[\deg v \leq d/M, \deg w \leq d/M] \leq \exp\left[-2d + \frac{2d}{M}(1 + \log M + O(1))\right]. \quad (8)$$

Thus, by adjusting M to be sufficiently large, we have

$$\Pr[\aleph_M \text{ is not an independent set}] = O(nd \exp(-(2 - \epsilon/2)d)) = O(n \exp(-(2 - \epsilon)d)).$$

(iii) This follows in much the same way as the proof of (ii). Here though, we require that the degrees of \aleph_M^c are not too large. By Lemma 5.1, these degrees can be bounded by some Cd with probability at least $1 - O(\exp(-2d))$, and so it suffices to assume it. From the union bound and symmetry, we have that

$$\begin{aligned} \Pr[\exists u \in \aleph_M^c : e(u\aleph_M) \geq 2 \cap \mathbf{b.d.c.}] \\ \leq n^3 \Pr[u \in \aleph_M^c, v \in \aleph_M, w \in \aleph_M, u \leftrightarrow v, u \leftrightarrow w \cap \mathbf{b.d.c.}]. \end{aligned}$$

Again we condition on the degrees $\deg u = d_1, \deg v = d_2$, and $\deg w = d_3$, and bound

$$\begin{aligned} \Pr[u \leftrightarrow v, u \leftrightarrow w | \deg u = d_1, \deg v = d_2, \deg w = d_3] \\ \leq \frac{\Pr[u \leftrightarrow v, u \leftrightarrow w | \deg u = d_1, \deg v = d_2, \deg w = d_3]}{\Pr[u \leftrightarrow v, u \leftrightarrow w, v \leftrightarrow w | \deg u = d_1, \deg v = d_2, \deg w = d_3]} \\ = \left(\frac{1-p}{p}\right)^2 \frac{\binom{n-3}{d_1-2} \binom{n-3}{d_2-1} \binom{n-3}{d_3-1} + \frac{p}{1-p} \binom{n-3}{d_1-2} \binom{n-3}{d_2-2} \binom{n-3}{d_3-2}}{\binom{n-3}{d_1} \binom{n-3}{d_2} \binom{n-3}{d_3}}. \end{aligned}$$

As before, we have d_1 and d_2 are less than d/M . As we also require the $\mathbf{b.d.c.}$ to hold, we may take $d_1 \leq c_1 d$ and as $d < n/100$, we may bound this as $Cc_1 d^2/n^2$ for some absolute constant C .

From (8), we have that

$$\Pr[v \in \aleph_M, w \in \aleph_M] = O(\exp(-(2 - \epsilon/2)d)),$$

and so we conclude that

$$\Pr \left[\max_{u \in \mathcal{N}_M^c} e(u \mathcal{N}_M) > 1 \right] = O(n \exp(-(2 - \epsilon)d)).$$

■

Our next lemma shows that the variance of the degree distribution is not too much larger than its expectation.

Lemma 5.4. For each fixed $\delta > 0$ and $m \geq 0$, there is a constant $C = C(\delta, m)$ sufficiently large so that if $p \geq \delta \log n/n$ then

$$\sum_{v \in V} (\deg v - d)^2 \leq Cnd$$

with probability at least $1 - C \exp(-md)$.

Proof. Note that this sum is the square Euclidean norm of the vector $(A - dI)\mathbf{1}$. Further, it is possible to write the norm as

$$\|(A - dI)\mathbf{1}\| = \sup_{\|x\|=1} |x^t(A - dI)\mathbf{1}|.$$

For any fixed vector x , we orthogonally decompose it as $x = v + c\mathbf{1}$, where $|c| \leq 1/\sqrt{n}$. We have that $v^t(A - dI)\mathbf{1} = v^tA\mathbf{1}$, and so by Proposition 5.2, for any m there is a constant C so that

$$\sup_{\substack{\|v\|=1 \\ v^t\mathbf{1}=0}} |v^tA\mathbf{1}| \leq C\sqrt{nd}$$

with probability at least $1 - O(\exp(-md))$. It remains to bound $\mathbf{1}^t(A - dI)\mathbf{1}$, which is

$$\mathbf{1}^t(A - dI)\mathbf{1} = \left(\sum_{v \in V} \deg v \right) - nd.$$

Note that $\sum_{v \in V} \deg v \sim 2 \text{Binom}(\binom{n}{2}, p)$, and so by standard Chernoff bounds, we have that

$$\Pr \left[|\mathbf{1}^t(A - dI)\mathbf{1}| \geq t \right] \leq C \exp\left(-\frac{t^2}{Cnd}\right)$$

for some absolute constant C and all $t \leq nd$. By taking $t = mn\sqrt{d}$, we have that $|\mathbf{1}^t(A - dI)\mathbf{1}| \leq mn\sqrt{d}$ with probability at least $1 - O(\exp(-mn))$ for sufficiently

large n . Recalling that $|c| \leq 1/\sqrt{n}$, we have that

$$|c\mathbf{1}^t(A - dI)\mathbf{1}| = O(\sqrt{nd}),$$

which completes the proof. \blacksquare

Using the previous lemma, we show that $T^{-1/2}$ tends to map the orthogonal complement of the 1st eigenvector of M to the approximate orthogonal complement of the 1st eigenvector of A .

Lemma 5.5. Let W be the set of vertices x for which $\deg x > 0$, and let \aleph_M be as in Proposition 5.3. For each $\delta > 0$ and $m \geq 0$, there is a constant $C = C(\delta, m)$ sufficiently large so that if $p \geq \delta \log n/n$ then

$$\sup_{\substack{\|x\|=1, \\ x^t T^{1/2} \mathbf{1}_W = 0}} |x^t T^{-1/2} \mathbf{1}_{\aleph_M^c}| \leq C \frac{\sqrt{n}}{d}$$

with probability at least $1 - C \exp(-md)$.

Proof. As we have that $|\aleph_M| < n/(100d)$ by Proposition 5.3, it follows that

$$|x^t T^{1/2} \mathbf{1}_{\aleph_M}| \leq \|T^{1/2} \mathbf{1}_{\aleph_M}\| \leq \sqrt{d|\aleph_M|} = O(\sqrt{n}).$$

Further, we have that $x^t T^{1/2} \mathbf{1}_{\aleph_M} = -x^t T^{1/2} \mathbf{1}_{\aleph_M^c}$, and hence it suffices to show that

$$\sup_{\substack{\|x\|=1, \\ x^t T^{1/2} \mathbf{1} = 0}} |x^t (T^{-1/2} - T^{1/2}/d) \mathbf{1}_{\aleph_M^c}| \leq C \frac{\sqrt{n}}{d}.$$

Taking norms,

$$\|x^t (T^{-1/2} - T^{1/2}/d) \mathbf{1}_{\aleph_M^c}\| \leq \|(T^{-1/2} - T^{1/2}/d) \mathbf{1}_{\aleph_M^c}\|.$$

Squaring this norm, we get

$$\|(T^{-1/2} - T^{1/2}/d) \mathbf{1}_{\aleph_M^c}\|^2 = \sum_{v \in \aleph_M^c} \left(\frac{1}{\sqrt{\deg v}} - \frac{\sqrt{\deg v}}{d} \right)^2 \leq \frac{M}{d^3} \sum_{v \in \aleph_M^c} (\deg v - d)^2.$$

Lemma 5.4 completes the proof. \blacksquare

Proof of Theorem 1.1 We show that we satisfy the conditions in Lemma 4.1. In Lemma 5.1, we show the b.d.c. In Proposition 5.2, we show the adjacency matrix condition. In Proposition 5.3, we show the fuzz condition. Finally, in Lemma 5.5, we show the parallel eigenspaces condition. Summing the failure probabilities, the failure probability in Theorem 1.1 is the sum of $Cn \exp(-(2 - \epsilon)d)$ from Proposition 5.3 and $C \exp(-md^{1/4} \log n)$ from Proposition 5.2, with all other errors much smaller in magnitude. Without the condition that $p \geq (\frac{1}{2} + \delta) \log n/n$, for some $\delta > 0$, the failure probability in Proposition 5.3 is not in control. ■

We wish to now show the lower bounds for $\lambda(\tilde{G})$. We will use Lemma 4.2, and this requires that we show the following.

Proposition 5.6. If $p = \omega(\sqrt{\log n}/n)$ and $p \leq \frac{1}{2} \log n/n$ then w.h.p., there are four distinct vertices a, b, c, d in the giant component for which the degrees of a and d are at least $np/2$, the degrees of b and c are 2, and the induced subgraph on (a, b, c, d) is a path.

We first show by the 2nd moment method that such four-tuples (a, b, c, d) exist in the graph w.h.p. We then show that w.h.p., the small components have maximal degree $o(np)$, and hence these four tuples must have been part of the giant component.

Lemma 5.7. Suppose that $p = \omega(1/n)$ and that $p \leq \frac{1}{2} \log n/n$. Then, w.h.p., there are four-tuples (a, b, c, d) for which the degrees of a and d are at least $np/2$, the degrees of b and c are 2, and the induced subgraph on (a, b, c, d) is a path.

Proof. Define the pair of events

$$A(a, b, c, d) = \{a \leftrightarrow b \leftrightarrow c \leftrightarrow d, \deg b = \deg c = 2\} \text{ and}$$

$$B(a, b, c, d) = A(a, b, c, d) \cap \{\deg a \geq np/2, \deg d \geq np/2\}.$$

Set S to be the number of occurrences of B , that is,

$$S = \sum_{a,b,c,d} 1[B(a, b, c, d)],$$

with the sum over ordered four tuples of distinct vertices (a, b, c, d) . We need to show that $S > 0$ w.h.p.

The probability of A can be explicitly calculated as

$$\Pr[A(a, b, c, d)] = p^3(1-p)^{2(n-3)}.$$

Meanwhile, conditional on $A(a, b, c, d)$, the probability of $B(a, b, c, d)$ is exactly the probability of having two specific vertices of degree at least $np/2 - 1$ in $G(n-2, p)$. Set $Q = \Pr[X \geq np/2]$, where $X \sim \text{Binom}(n, p)$. Note that as $np \rightarrow \infty$, we have that $Q = 1 - o(1)$.

Furthermore, as $np \rightarrow \infty$ we have that

$$\Pr[B(a, b, c, d)|A(a, b, c, d)] = Q^2(1 - o(1)),$$

simply by conditioning on the edge between a and d . By summing over all possible tuples, it follows that $\mathbb{E}S = \Theta(nQ^2(np)^3e^{-2np}) = \omega(1)$.

For the variance of S , we need to compute probabilities of the pairs $B((a_i)_{i=1}^4) \cap B((b_i)_{i=1}^4)$. Note that if $a_2 = b_2$ then the only way both can happen is if $a_i = b_i$ for all $i \in [4]$. Analogous conclusions hold if $a_2 = b_3$ or if $a_3 \in \{b_2, b_3\}$. Thus, the only nontrivial way for the events $B((a_i)_{i=1}^4)$ and $B((b_i)_{i=1}^4)$ to intersect is if

- (1) all a_i and b_i are distinct;
- (2) $a_1 = b_1$ and the rest are distinct;
- (3) $a_1 = b_1, a_4 = b_4$, and the rest are distinct; or
- (4) $a_i = b_i$ for all i .

Note that there's no need to consider $a_1 = b_4$, as the event $B((b_i)_{i=1}^4)$ is preserved under reversing the a_i . Likewise, there's no need to consider $a_4 = b_4$, as one can reverse both a_i and b_i . Set T_i to be the pairs of tuples satisfying each of the four cases.

If the pair is in T_1 , then

$$\Pr[B((a_i)_{i=1}^4) \cap B((b_i)_{i=1}^4) | A((a_i)_{i=1}^4) \cap A((b_i)_{i=1}^4)] = Q^4(1 - o(1))$$

as once more; this is the statement that four vertices in $G(n-4, p)$ have degree at least $(np/2 - 1)$. We also have that

$$\Pr[A((a_i)_{i=1}^4) \cap A((b_i)_{i=1}^4)] = p^6(1-p)^{4n-16},$$

so that

$$\Pr[B((a_i)_{i=1}^4) \cap B((b_i)_{i=1}^4)] = \Pr[B((a_i)_{i=1}^4)]^2 (1 - o(1)).$$

Thus, the contribution of the pairs in T_1 to the variance of S is $o((\mathbb{E}S)^2)$.

For terms from T_2 , the same reasoning as above shows that

$$\Pr\left[B((a_i)_{i=1}^4) \cap B((b_i)_{i=1}^4)\right] = Q^3 p^6 (1-p)^{4n} (1-o(1)).$$

For such pairs, however, we have that $|T_2| = \Theta(n^7)$, and hence the contribution to the variance of S is $o((\mathbb{E}S)^2)$. In the same way, the contributions of T_3 and T_4 are smaller still. As each is individually of order $o((\mathbb{E}S)^2)$, we have that $S > 0$ w.h.p. \blacksquare

Lemma 5.8. Suppose that $p = \omega(1/n)$, then for any $\epsilon > 0$, the number of vertices not in the giant component is at most $ne^{-(1-\epsilon)np}$ w.h.p.

Proof. Set R to be the number of vertices not in the largest component of $G(n, p)$. If W is the set of these vertices, then W satisfies $e(W, W^c) = 0$. Therefore, if there is no collection W of at least r vertices such that $e(W, W^c) = 0$, then $R < r$.

The expected number $\mathbb{E}N_r$ of such collections W is given by

$$\mathbb{E}N_r = (1-p)^{r(n-r)} \binom{n}{r}.$$

Set $r_0 = ne^{-(1-\epsilon)np}$. We will show that $\sum_{r=r_0}^{n/2} \mathbb{E}N_r \rightarrow 0$, which implies the lemma.

Subdivide the sum into two pieces S_1 and S_2 , given by $S_1 = \sum_{r_0}^{\lfloor \epsilon n/4 \rfloor} \mathbb{E}N_r$ and $S_2 = \sum_{\lfloor \epsilon n/4 \rfloor}^{n/2} \mathbb{E}N_r$. For $\lfloor \epsilon n/4 \rfloor \leq r \leq n/2$,

$$\mathbb{E}N_r = (1-p)^{r(n-r)} \binom{n}{r} \leq e^{-c_\epsilon n^2 p 2^n},$$

for some $c_\epsilon > 0$, which decays exponentially in n as $np \rightarrow \infty$. Hence, $S_2 \rightarrow 0$.

As for S_1 , we claim that for any $\alpha > 0$ there is an $n \geq n_0(\alpha, \epsilon)$ sufficiently large so that for all $r_0 < r < \epsilon n/4$, $\mathbb{E}N_{r+1} \leq \alpha \mathbb{E}N_r$ for all $n \geq n_0(\alpha, \epsilon)$. Estimating for these r ,

$$\begin{aligned} \frac{\mathbb{E}N_{r+1}}{\mathbb{E}N_r} &= (1-p)^{n-2r+1} \frac{n-r-1}{r+1} \\ &\leq \frac{ne^{-np+2rp}}{r} \\ &\leq \frac{ne^{-(1-\epsilon/2)np}}{r} \\ &\leq e^{-\epsilon np/2}. \end{aligned}$$

Hence, as $np \rightarrow \infty$, this is eventually less than any positive α .

As S_1 is dominated by a geometric series $S_1 = O(\mathbb{E}N_{r_0})$. For this leading term, we get that

$$\mathbb{E}N_{r_0} \leq e^{-pr_0(n-r_0)} \left(\frac{en}{r_0}\right)^{r_0} \leq \exp\left(-\epsilon n^2 p e^{-(1-\epsilon)np} (1 - o(1))\right) \rightarrow 0,$$

completing the proof. \blacksquare

Lemma 5.9. If $p = \omega(\sqrt{\log n}/n)$, then w.h.p., the maximum degree of the vertices not in the giant component is at most $np/100$.

Proof. Set R to be the number of vertices not in the giant component. By Lemma 5.8, we have that $R \leq ne^{-np/2}$ w.h.p. Suppose that W is a fixed collection of vertices of size r . Conditional on there being no edges between W and W^c , the law of the induced graph on W is simply that of $G(r, p)$.

Let $X \sim \text{Binom}(r-1, p)$. Then by Lemma A.2 there are absolute constants $c > 0$ and $M > 0$ so that

$$\Pr[X > np/100] \leq \exp(-cnp \log(n/r))$$

provided $r < n/M$. Setting E_W to be the event that W and W^c are not connected,

$$\Pr\left[\max_{w \in W} \deg w > np/100 \mid E_W\right] \leq r \exp(-cnp \log(n/r)).$$

Let Y be the max degree of all vertices not in the largest component. As the previous bound holds for all W in consideration, we get that

$$\Pr[Y > np/100 \mid R = r] \leq r \exp(-cnp \log(n/r)).$$

This bound is monotone increasing in r , and so we get that

$$\Pr[Y > np/100 \mid R \leq ne^{-np/2}] \leq n \exp(-c(np)^2(1 - o(1)))$$

for some absolute constant c . Thus, by the assumption on np , the desired claim holds. \blacksquare

Proof of Theorem 1.2 and Proposition 5.6 For Proposition 5.6, the previous three Lemmas 5.7, 5.8, and 5.9 show the desired claim that w.h.p. there are tuples (a, b, c, d)

of vertices in the giant component for which $\deg a$ and $\deg d$ are at least $np/2$, vertices b and c have degree 2, and the induced graph on these vertices is a path.

Letting H be the giant component of the graph, then there is a constant C so that the eigenvalues of the Laplacian of H satisfy

$$\lambda_{|H|} \geq \frac{3}{2}$$

and

$$\lambda_2 \leq \frac{1}{2} + C/\sqrt{np}$$

by Lemma 4.2. ■

6 Gap Process Theorem

In this section we prove a general process-version theorem for the spectral gap below the connectivity threshold. We recall the definition of $Y_t^d(n)$, the continuous time Linial–Meshulam process. Let F_k denote the collection of all possible k -faces on n vertices, and let $F_k(S)$ for simplicial complex S be all k -faces of S . Let $\{T_\sigma, \sigma \in F_k\}$ be an i.i.d. family of $\text{Exp}(1)$ variables. Define $\{Y_t^d(n), t \geq 0\}$ to be the continuous time Markov process where $Y_t^d(n)[0][k]$ is the complete $(k-1)$ -skeleton of the n -simplex and its k -faces are given by

$$F_k(Y_t^d(n)) = \{\sigma \in F_k : T_\sigma \leq t\}.$$

Thus, $Y_t^d(n)$ is the complex whose k -faces have been born up to time t , and $Y_t^d(n)[\infty][k]$ is the complete k -skeleton of the n -simplex. For $k=1$, this recovers the standard continuous time Erdős–Rényi process. For fixed t , $Y_t^d(n)$ is the Bernoulli complex $Y_k(n, p(t))$ with $p(t) = 1 - e^{-t}$. Let $d(t) = (n-1)p(t)$. Fix $\delta \in (0, \frac{1}{2})$ and define t_0 by the relation that

$$p(t_0) = \begin{cases} (\frac{1}{2} + \delta) \log n/n & k=1, \\ (k-1 + \delta) \log n/n & k>1. \end{cases}$$

For any $(k-2)$ -dimensional face f of a k -dimensional simplicial complex S , we identify its *link* with a graph, denoted $\text{lk}(f)$. We will only consider links of $(k-2)$ -dimensional faces. This graph $\text{lk}(f)$ has vertex set given by all $(k-1)$ -dimensional faces containing f . Two of these edges e and g are connected if and only if $e \cup g$, which is a k -dimensional face, is contained in S .

For example, when $k = 1$ and S is a graph, the only $(k - 2)$ -dimensional face is the empty set. Its link has vertex set given by all 0-dimensional faces (all vertices), and vertices are connected if and only if they are contained in an edge. Hence, in this case $\text{lk}(\emptyset)$ can be identified with the original graph S .

In $Y_t^d(n)$, which has a complete $(k - 1)$ -skeleton, each link is distributed as a $G(n - k + 1, p(t))$. These links $\{\text{lk}(f)\}$, where f ranges over all $(k - 2)$ -dimensional faces, are not independent, and in fact analysis rests in some ways on exploiting their exact dependency structure.

Recall that we refer to a $(k - 1)$ -dimensional face f as *isolated* if and only if it is not contained in any k -dimensional face. Note that a face f is isolated if and only if it is an isolated vertex in $\text{lk}(g)$ for all $(k - 2)$ -dimensional $g \subset f$.

Theorem 6.1. Let $\tilde{Y}_t^k(n)$ denote the process derived from $Y_t^d(n)$ by removing every isolated $(k - 1)$ -face. There is a constant $C = C(k, \delta)$ so that w.h.p. the normalized Laplacian of $\text{lk}(f)$ of every dimension- $(k - 2)$ face f of $\tilde{Y}_t^k(n)$ has

$$\max_{i>1} |1 - \lambda_i| < \frac{C}{\sqrt{d(t)}}$$

for all $t \geq t_0$.

An equivalent formulation is that each $\text{lk}(f)$ for codimension-2 $f \in Y_t^d(n)$ consists of isolated vertices and a giant component whose gap is $1 - C/\sqrt{d(t)}$ for all time $t \geq t_0$. In the higher-dimensional setting, the proof is more complicated than simply studying each link individually and taking the union bound. The key is to study the “fuzz” globally. To this end, for each $\text{lk}(f)$ and for any $M \geq 1$, let

$$\aleph_M^f(t) = \{w \in \mathcal{V}(\text{lk}(f)) : \deg_{\text{lk}(f)}(w) \leq d(t_0)/M\}. \quad (9)$$

Note that this makes each $\aleph_M[f][t]$ monotone decreasing in t .

Lemma 6.2. There is an $M = M(k, \delta)$ and an $\epsilon = \epsilon(k, \delta)$ so that

$$\sum_{f \in F_{k-2}} |\aleph_M[f][t_0]|^2 \leq n^{1-\epsilon}$$

with overwhelming probability.

Proof. For $k = 1$, there is only one link to consider, and so it suffices to show that $\left| \mathbb{N}_M^\emptyset[t_0] \right| \leq n^{1/2-\epsilon}$. For $k > 1$, we proceed by showing that for any ϵ there is an M so that both

- (1) $\max_{f \in F_{k-2}} \left| \mathbb{N}_M^f(t_0) \right| \leq n^\epsilon$ and
- (2) $\sum_{f \in F_{k-2}} \left| \mathbb{N}_M^f(t_0) \right| \leq n^{1-2\epsilon}$

hold with overwhelming probability.

The 1st condition follows from an identical argument to the 1st part of Proposition 5.3; the $k = 1$ case follows from an identical argument, and we just sketch the $k > 1$ case. As before, for any $1 > \eta > 0$, there is an $M(\delta, \eta)$ sufficiently large so that for a fixed set of vertices $w_1, w_2, \dots, w_{\lceil n^\epsilon \rceil}$,

$$\Pr \left[\deg_{\text{lk}(f)}(w_i) \leq d(t_0)/M, \forall 1 \leq i \leq \lceil n^\epsilon \rceil \right] = O(\exp(-n^\epsilon d(t_0)(1 - \eta))).$$

This overwhelms the $O(\exp((1 - \epsilon)n^\epsilon \log n))$ possible choices of vertices as

$$d(t_0)/\log n > (1 + \delta)(1 + o(1))$$

and η may be chosen sufficiently small. As there are only $O(n^{k-1})$ many links to consider, this may be taken to hold for all links simultaneously with overwhelming probability.

We now turn to the 2nd condition. For a fixed $(k - 1)$ -dimensional face f , let X_f denote the number of k -faces in $Y_t^d(n)[t_0][k]$ containing f . If f is a vertex in a $(k - 2)$ -dimensional face of $Y_t^d(n)[t_0][k]$, then X_f is the degree of that vertex in $\text{lk}(f)$. Hence,

$$\frac{1}{k} \sum_{f \in F_{k-2}} \left| \mathbb{N}_M^f(t_0) \right| = \sum_{f \in F_{k-1}} \mathbf{1}[X_f \leq d(t_0)/M].$$

Thus, by adjusting ϵ , it suffices to show the claim for the right-hand side. Call a collection S of $(k - 1)$ -faces *balanced* if

$$\max_{w \in F_{k-2}} |\{\sigma \in S : w \subset \sigma\}| \leq n^\epsilon.$$

Observe that we have shown that with overwhelming probability the set

$$S = \left\{ f \in F_{k-1} : X_f \leq d(t_0)/M \right\}$$

is balanced with overwhelming probability.

By symmetry we have

$$\begin{aligned} \Pr \left[\exists f_1, f_2, \dots, f_r \in F_{k-1} : X_{f_i} \leq d(t_0)/M, 1 \leq i \leq r, \{f_i\} \text{ balanced} \right] \\ \leq \binom{n}{r} \Pr \left[X_{f_i} \leq d(t_0)/M, 1 \leq i \leq r, \{f_i\} \text{ balanced} \right]. \end{aligned}$$

Let X denote the number of k -faces that contain some f_i . If every $X_{f_i} \leq d(t_0)/M$, it follows that $X \leq rd(t_0)/M$. Each f_i is contained in $n - k$ possible k -faces, but it may be possible that some f_i and f_j are both contained in a single k -face. If this occurs, however, it must be that $|f_i \cap f_j| = k - 1$. In other words, each contains a common $(k - 2)$ -face. Furthermore, there is at most one k -face that contains both f_i and f_j .

A fixed face f_j contains k distinct $(k - 2)$ -faces q_1, q_2, \dots, q_k . As $\{f_i\}$ is balanced, each q_l is contained in at most n^ϵ distinct f_i . Thus, there are at most $n^\epsilon k$ many k -faces that contain f_j and some other f_i , and this implies there are at least $r(n - k - n^\epsilon k)$ distinct possible k -faces that contain some f_i . It follows that X stochastically dominates a Binom $(\lceil r(n - k - n^\epsilon k) \rceil, p(t_0))$ variable. Applying Lemma A.1, we get

$$\begin{aligned} \Pr \left[X_{f_i} \leq d(t_0)/M, 1 \leq i \leq r, \{f_i\} \text{ balanced} \right] &\leq \Pr \left[X \leq rd(t_0)/M \right] \\ &\leq \exp \left(-r(n - k - n^\epsilon k)p(t_0) + \frac{rd(t_0)}{M} \left(1 + \log \frac{M(1+r(n-k-n^\epsilon k)p(t_0))}{rd(t_0)} \right) \right). \end{aligned}$$

Thus, we get

$$\begin{aligned} \log \Pr \left[\exists f_1, f_2, \dots, f_r : X_{f_i} \leq d(t_0)/M, 1 \leq i \leq r \right] \\ \leq r \left[(k \log n - \log r) - d(t_0) + \frac{d(t_0)}{M} (1 + \log(M)) \right] (1 + o(1)). \end{aligned}$$

Since $d(t_0) \geq (k - 1 + \delta) \log n - o(1)$, we can set $r = \lceil n^{1-\delta/2} \rceil$ and make M sufficiently large that

$$(k \log n - \log r) - d(t_0) + \frac{d(t_0)}{M} (1 + \log(M)) \rightarrow -\infty.$$

Taking $\epsilon = \delta/4$, we have shown the desired claim. ■

With global control on the number of exceptional vertices, the proof now reduces to essentially a union bound over all later times and links.

Lemma 6.3. There is a constant $C = C(k)$ so that w.h.p., every $\text{lk}(f)$ where $f \in Y_t^d(n)$ has dimension $(k - 2)$ satisfies the following:

- (1) **B.d.c** Every vertex has degree at most $Cd(t)$.
- (2) **Adjacency matrix** The adjacency matrix of the link satisfies

$$\sup_{\substack{\|x\|=1, x^t \mathbf{1}=0 \\ \|y\|=1}} |x^t A y| \leq C\sqrt{d(t)}.$$

- (3) **Parallel eigenspaces** Setting $\aleph_M = \aleph_M[f][t]$ and T to be the diagonal matrix of degrees of the link

$$\sup_{\substack{\|x\|=1, \\ x^t T^{1/2} \mathbf{1}_w=0}} |x^t T^{-1/2} \mathbf{1}_{\aleph_M^c}| \leq C \frac{\sqrt{n}}{d(t)}$$

for all $t \geq t_0$.

Proof. Let I be the interval $[t_1, t_2]$, where $t_0 \leq t_1 \leq t_2$. The probability that there are two faces that appear in this interval can be bounded by

$$\Pr[\exists \sigma_1, \sigma_2 : T_{\sigma_1} \in I \text{ and } T_{\sigma_2} \in I] \leq \binom{n}{k}^2 (p(t_2) - p(t_1))^2.$$

Let r be the smallest integer so that $p(t_0) + rn^{-2k-1} \geq 1$. Set $p_i = p(t_0) + in^{-2k-1}$ for all $0 \leq i < r$, and set $p_r = 1$. Let t_i be such that $p(t_i) = p_i$, and set $t_r = \infty$. Note that for $t \in [t_i, t_{i+1})$, $Y_t^d(n) \neq Y_{t_i}^k(n)$ and $Y_t^d(n) \neq Y_{t_{i+1}}^k(n)$ imply there must be two faces σ_1 and σ_2 for which $T_{\sigma_1}, T_{\sigma_2} \in [t_i, t_{i+1})$. Hence,

$$\begin{aligned} \Pr[\exists t \geq t_0 : Y_t^d(n) \neq Y_{t_i}^k(n) \forall 0 \leq i \leq r] &\leq \sum_{i=0}^{r-1} \Pr[\exists \sigma_1, \sigma_2 : T_{\sigma_1}, T_{\sigma_2} \in I] \\ &\leq \sum_{i=0}^{r-1} n^{-2k-2} \leq n^{-2}. \end{aligned}$$

By applying Lemma 5.1, Proposition 5.2, and Lemma 5.5 with m sufficiently large, we may thus assure that there is a constant sufficiently large that these properties occur for all links of all $Y_{t_i}^k(n)$, for $0 \leq i \leq r - 1$. ■

Lemma 6.4. There is an $M = M(k, \delta)$ and a constant $C = C(M, k)$ so that with t_1 satisfying $p(t_1) = C \log n/n$, all $\aleph_M^f(t) = \emptyset$ for $t \geq t_1$ w.h.p. Further, for all $t_1 \geq t \geq t_0$ every $\text{lk}(f)$ of $Y_t^d(n)$ satisfies the following:

- (1) $|\aleph_M^f| \leq \frac{n}{2}$;
- (2) \aleph_M^f is an independent set; and
- (3) $\max_{u \in \aleph_M^c} e(u, \aleph_M^f) \leq 1$

with $\aleph_M^f = \aleph_M^f(t)$.

Proof. There is an M_1 so that this holds for $Y_{t_0}^k(n)$ by Proposition 5.3 and by taking the union bound over all links. Likewise, there is an M_2 so that the conclusions of Lemma 6.2 holds. Take M to be the maximum of these, and note that from monotonicity, the conclusions of both the proposition and lemma hold. As $\aleph_M^f(t)$ is monotone in t also, we have that

$$|\aleph_M^f(t)| \leq |\aleph_M^f(t_0)| \leq n/2$$

is satisfied for all n sufficiently large.

From a union bound and Lemma A.1, we may choose $C = C(M, k)$ sufficiently large so that with probability going to 1,

$$\aleph_M^f(t_1) = \emptyset$$

for all $f \in F_{k-2}$.

Let τ_i be the times at which the i^{th} face is added to $Y_t^d(n)$ after time t_0 , and let $\tau_0 = t_0$. Likewise, let Δ_i denote the i^{th} face, and let $\mathcal{F}(\tau_i) = \sigma(Y_{\tau_i}^k(n))$. Let N denote the largest i so that $\tau_i \leq C \log n/n$. From Chernoff bounds, there are at most $100C(\log n)n^k$ many k -dimensional faces in $Y_{\tau_1}^k(n)$ with overwhelming probability, and hence $N \leq 100(\log n)n^k$ with overwhelming probability.

We begin by bounding the probability that a newly added face creates an edge between two vertices of $\aleph_M^f(t)$ for some $f \in F_{k-2}$.

$$\begin{aligned} \Pr \left[\exists u, v \in \aleph_M^f(\tau_i) : u, v \in \Delta_{i+1} \mid \mathcal{F}(\tau_i) \right] &\leq \frac{|\aleph_M^f(\tau_i)|^2}{|F_k| - |Y_{\tau_1}^k(n)|} \\ &\leq \frac{|\aleph_M^f(\tau_0)|^2}{|F_k| - |Y_{\tau_1}^k(n)|}. \end{aligned} \tag{10}$$

Let $E_{i,f}$ denote the event that

- (1) the number of k -dimensional faces $|Y_{t_1}^k(n)| \leq 100Cn^k \log n$,
- (2) $\sum_{f \in F_{k-2}} |\mathbb{N}_M^f(t_0)|^2 \leq n^{1-\epsilon}$,
- (3) there exists u and v in $\mathbb{N}_M^f(\tau_i)$ so that $u \in \Delta_{i+1}$ and $v \in \Delta_{i+1}$.

By conditioning, we have that

$$\begin{aligned} \Pr \left[\bigcup_{i,f} E_{i,f} \right] &\leq \mathbb{E} \sum_{i=0}^N \sum_{f \in F_{k-2}} \frac{|\mathbb{N}_M^f(\tau_0)|^2 \mathbf{1}\{E_{i,f}\}}{|F_k| - |Y_{t_1}^k(n)|} \\ &\leq \mathbb{E} \sum_{i=0}^N \frac{\sum_{f \in F_{k-2}} |\mathbb{N}_M^f(\tau_0)|^2 \mathbf{1}\{E_{i,f}\}}{|F_k| - 100Cn^k \log n} \\ &\leq \mathbb{E} \sum_{i=0}^N \frac{n^{1-\epsilon} \mathbf{1}\{|Y_{t_1}^k(n)| \leq 100Cn^k \log n\}}{|F_k| - 100Cn^k \log n} \\ &\leq \frac{(100Cn^k \log n)n^{1-\epsilon}}{|F_k| - 100Cn^k \log n} = O(n^{-\epsilon} \log n). \end{aligned}$$

Thus, w.h.p., no face added between t_0 and t_1 creates an edge between two elements of any $\mathbb{N}_M^f(t)$.

We now turn to bounding the probability that a newly added face connects an element of $\mathbb{N}_M^f(t)$ to a neighbor of $\mathbb{N}_M^f(t)$. Let $\mathcal{N}_M^f(t)$ be the set of neighbors of $\mathbb{N}_M^f(t)$, and let $D(t)$ be an upper bound for the degree of a vertex of any link of $Y_t^d(n)$. Note that $|\mathcal{N}_M^f(t)| \leq D(t)|\mathbb{N}_M^f(t)|$. Then

$$\Pr \left[\exists u \in \mathbb{N}_M^f(\tau_i), v \in \mathcal{N}_M^f(t) : u, v \in \Delta_{i+1} \mid \mathcal{F}(\tau_i) \right] \leq \frac{D(\tau_i)|\mathbb{N}_M^f(\tau_i)|^2}{|F_k| - |Y_{t_i}^k(n)|}.$$

W.h.p., there is a constant K so that all the degrees can be bounded by $K \log n$ for all $t \leq t_1$. This failure probability is at most a logarithmic factor more than the failure probability in (10). Hence, the same proof shows that w.h.p., no added face increases

$$\max_{u \in \mathcal{V}(\text{lk}(f)) \setminus \mathbb{N}_M^f(t)} e(u, \mathbb{N}_M^f(t)).$$

■

Proof of Theorem 6.1 We are essentially ready to apply Lemma 4.1. The only concern is that in (9), the set $\mathbb{N}_M^f(t)$ is defined in terms of $d(t_0)$ and not $d(t)$. However, as noted in Lemma 6.4, all these sets disappear once $p(t_1) = C \log n/n$, at which point $d(t)$ has only

risen by a factor of $K = \frac{p(t_1)}{p(t_0)}$. Thus,

$$Q^f(t) = \{w \in \mathcal{V}(\text{lk}(f)) : \deg_{\mathbb{S}(\text{lk}(f))}(w) \leq d(t)/KM\} \subseteq \mathbb{N}_M^f(t),$$

for all $t \leq t_1$, and by monotonicity, all the desired properties of $\mathbb{N}_M^f(t)$ transfer to $Q^f(t)$. Thus, Lemmas 6.3 and 6.4 show all the needed properties of Lemma 4.1 hold, completing the proof. ■

7 Cohomology Structure Theorem

The structure theorem for cohomology relies on the following theorem of Ballman–Świątkowski [3]. A simplicial complex Δ is called *pure k -dimensional* if it is k -dimensional and every face is contained in a k -dimensional one.

Ballman–Świątkowski criterion. If Δ is a finite, pure k -dimensional simplicial complex, so that for every $(k - 2)$ -dimensional face σ , the normalized Laplacian $L = L[\text{lk}(\sigma)]$ satisfies $\lambda_2 > 1 - \frac{1}{k}$ then $H^{k-1}(\Delta, \mathbb{Q}) = 0$.

Proof of Theorem 3.2 Recall that we define t_0 so that $p(t_0) = (k - 1 + \delta) \log n/n$. Let \tilde{Y}_t denote the simplicial complex $Y_t^d(n)$ with all its isolated $(k - 1)$ -faces deleted. By Theorem 6.1, w.h.p. for all $t \geq t_0$, all links of \tilde{Y}_t have $\lambda_2(L) = 1 - o(1)$.

We need to check that \tilde{Y}_t is pure k -dimensional, that is, that every face is contained in some k -dimensional face. Note that this can only fail if there is some $(k - 2)$ -dimensional face of $Y_t^d(n)$ that is not contained in any k -dimensional face. As this is a monotone property, it suffices to check that $Y_t^d(n)[t_0][k]$ has no such $(k - 2)$ -faces.

Put I to be the number of isolated $(k - 2)$ -faces in $Y_t^d(n)[t_0][k]$. Then

$$\mathbb{E}I = \binom{n}{k-1} (1 - p(t_0))^{n^2/2} (1 - o(1)),$$

which decays exponentially in n . Hence, \tilde{Y}_t is pure k -dimensional w.h.p. for all $t \geq t_0$, and so Theorem 7 applies. It follows that $H^{k-1}(\tilde{Y}_t, \mathbb{Q}) = 0$, and it remains to compare $H^{k-1}(\tilde{Y}_t, \mathbb{Q})$ and $H^{k-1}(Y_t^d(n), \mathbb{Q})$.

For what remains, fix $t \geq t_0$. It will follow from induction that each additional $(k - 1)$ -face we glue to \tilde{Y} increases the dimension of the $(k - 1)$ -cohomology by 1. Let Z be the complex formed by including one of the isolated $(k - 1)$ -faces of Y back into \tilde{Y} . Let B be a neighborhood of the included $(k - 1)$ -face that is homotopic to a single

$(k - 1)$ -simplex. Then the Mayer–Vietoris sequence (see Chapter 3 of [29]) for the $(k - 1)$ -dimensional cohomology is

$$\dots \rightarrow H^{k-1}(Z, \mathbb{Q}) \rightarrow H^{k-1}(\tilde{Y}, \mathbb{Q}) \oplus H^{k-1}(B, \mathbb{Q}) \rightarrow H^{k-1}(\tilde{Y} \cap B, \mathbb{Q}) \rightarrow H^k(Z, \mathbb{Q}).$$

As $\tilde{Y} \cap B$ is homotopic to a $(k - 2)$ -dimensional sphere, $H^{k-1}(\tilde{Y} \cap B, \mathbb{Q}) = 0$. Also, $H^{k-1}(B, \mathbb{Q}) = \mathbb{Q}$, and so this sequence becomes

$$0 \rightarrow H^{k-1}(Z, \mathbb{Q}) \rightarrow H^{k-1}(\tilde{Y}, \mathbb{Q}) \oplus \mathbb{Q} \rightarrow 0,$$

or otherwise stated, $H^{k-1}(Z, \mathbb{Q}) \cong H^{k-1}(\tilde{Y}, \mathbb{Q}) \oplus \mathbb{Q}$. Each additional isolated $(k - 1)$ -faces increases the dimension by one by the very same argument, which completes the proof. ■

8 Property (T)

The proof here is nearly identical to the proof of the cohomology vanishing structure theorem. To establish our results concerning property (T) of random fundamental groups, we will use the following theorem of Żuk.

Żuk’s criterion. If X is a pure 2D locally finite simplicial complex so that for every vertex v , the vertex link $\text{lk}(v)$ is connected and the normalized Laplacian $L = L[\text{lk}(v)]$ satisfies $\lambda_2(L) > 1/2$, then $\pi_1(X)$ has property (T).

Proof. of Theorem 1.4 Recall that we define t_0 so that $p(t_0) = (1 + \delta) \log n/n$. Let \tilde{Y}_t denote the simplicial complex $Y_t^d(n)[t][2]$ with all its isolated edges deleted. By Theorem 6.1, w.h.p. for all $t \geq t_0$, all links of \tilde{Y}_t have $\lambda_2(L) = 1 - o(1)$. Then by Żuk’s criterion, $\pi_1(\tilde{Y}_t)$ has property (T) for all $t \geq t_0$.

Fix $t \geq t_0$. It only remains to compare the fundamental groups $\pi_1(\tilde{Y})$ and $\pi_1(Y)$. But attaching a 1-cell to a connected CW complex W adds a free \mathbb{Z} -factor to the fundamental group $\pi_1(W)$, by the Seifert–van Kampen theorem (see Theorem 1.20 of [29]). So we only need to check that deleting all the isolated edges in Y does not result in a disconnected complex \tilde{Y} .

Removing less than $n - 1$ edges from the complete graph K_n can not disconnect it; indeed, to separate a component of order k from the rest of the graph requires removing at least $k(n - k)$ edges, which is minimized when $k = 1$. Thus, we need only check that the number of isolated edges is fewer than $n - 1$. From monotonicity, it suffices to show that at time t_0 the number of isolated edges is w.h.p. $o(n)$.

By linearity of expectation, the expected number of edges deleted $\mathbb{E}[D]$ is given by

$$\begin{aligned}\mathbb{E}[D] &= \binom{n}{2} (1 - p(t_0))^{n-2} \\ &\leq \frac{1}{2} n^2 \exp(-p(t_0)(n-2)) \\ &\leq O(n^{1-c})\end{aligned}$$

for some constant $c > 0$. By the 2nd moment method, for example, D is tightly concentrated around its mean, so w.h.p. \tilde{Y} is connected. The claim follows. ■

Corollary 3.7 quickly follows.

Proof of Corollary 3.7 Let I denote the number of isolated edges. The expected number of isolated edges $\mathbb{E}[I]$ is

$$\mathbb{E}[i] = \binom{n}{2} (1 - p)^{n-2} \leq n^2 e^{-np}.$$

Taking $p = (2 \log n + f(n))/n$, where $f(n) \rightarrow \infty$, this is seen to go to 0, completing the proof. ■

9 Kahn–Szemerédi argument

In this section we give the proof of Proposition 5.2 and Lemma 5.1, which are minor modifications of the standard Kahn–Szemerédi argument.

We begin with a proof of the regularity conditions.

Proof of Lemma 5.1 For any vertex v , $\deg(v)$ is a binomial random variable with mean $d > \delta \log(n)$. By Lemma A.2, $\mathbb{P}(\deg(v) > c_0 d) \leq \exp\left(-\frac{dc_0 \log c_0}{3}\right)$ provided $c_0 > 4$. Thus, taking the union bound over all vertices, we get that

$$\Pr[\mathbf{b.d.c. fails}] \leq \exp(d(\frac{1}{\delta} - \frac{c_0 \log c_0}{3})).$$

By taking c_0 sufficiently large, we may take

$$\frac{1}{\delta} - \frac{c_0 \log c_0}{3} \leq -m,$$

completing the proof of the 1st claim.

We will now turn to showing the discrepancy property, for which we need to show there are constants $c_i = c_i(\delta, m)$ so that at least one of

- (1) $\frac{e(A,B)}{\mu(A,B)} \leq c_1$
- (2) $e(A, B) \log \frac{e(A,B)}{\mu(A,B)} \leq c_2(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}$
- (3) $|A| \vee |B| \leq d^{1/4}/100$.

Note that these properties are monotone in c_i , and so we are free to increase the constants as needed throughout the proof.

Let D be the event that the discrepancy condition fails and let $D(A, B)$ be the event that the discrepancy condition fails for sets A and B . Then by the union bound

$$\begin{aligned} \mathbb{P}(D) &\leq \mathbb{P}(\exists A, B \text{ with } |A| \wedge |B| \geq n/e : D(A, B) \text{ occurs}) \\ &\quad + \mathbb{P}(\exists A, B \text{ with } |A| \vee |B| \geq n/e \geq |A| \wedge |B| : D(A, B) \text{ occurs}) \\ &\quad + \sum_{A, B: |A| \vee |B| < n/e} \mathbb{P}(D(A, B)). \end{aligned}$$

Taking $c_1 > e^2$, then when $|A| \wedge |B| \geq \frac{n}{e}$,

$$e(A, B) > c_1 \mu(A, B) > c_1 (n/e)^2 d/n > nd.$$

Thus, there are at least nd edges in the graph. The distribution of the number of edges is binomial with mean $n(n-1)p/2 = nd/2$, and so the probability of this is going to zero exponentially in nd , that is,

$$\mathbb{P}(\exists A, B \text{ with } |A| \wedge |B| \geq n/e : D(A, B) \text{ occurs}) = O(\exp(-cnd)) \quad (11)$$

for some absolute constant $c > 0$.

If $|A| \vee |B| \geq \frac{n}{e} > |A| \wedge |B|$, and if the b.d.c. holds, then $e(A, B) \leq (|A| \vee |B|)c_0 d$ and

$$\frac{e(A, B)}{\mu(A, B, n)} \leq \frac{c_0 n d (|A| \vee |B|)}{|A| |B| d} = \frac{c_0 n}{|A| \wedge |B|} \leq c_0 e.$$

Thus, taking $c_1 > c_0 e$, we have that

$$\begin{aligned} \mathbb{P}(\exists A, B \text{ with } |A| \vee |B| \geq n/e \geq |A| \wedge |B| : D(A, B) \text{ occurs}) &\leq \mathbb{P}(\mathbf{b.d.c. fails}) \\ &= O(\exp(-md)). \end{aligned} \quad (12)$$

Now we need to deal with the case that both A and B are less than $\frac{n}{e}$, but at least one is greater than $d^{1/4}/100$. Take $c_2 > 18 + 1200m$. For emphasis, we will write $\mu(A, B, n) = \mu(A, B) = \frac{|A||B|d}{n}$. Choose $r = r(A, B, n) = c_1 \vee r_1$, where r_1 is the solution to

$$\mu(A, B, n)r_1 \log(r_1) = c_2(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}.$$

For any A, B and n we must have either

- $e(A, B) \leq r\mu(A, B, n)$ and $r = c_1$
- $e(A, B) \leq r\mu(A, B, n)$ and $r = r_1$ or
- $e(A, B) > r\mu(A, B, n)$.

Thus, if $D(A, B)$ occurs then at least one of the following three events occur:

- $D_1 = D_1(A, B) = \left\{ \begin{array}{l} e(A, B) \leq r\mu(A, B, n), r = c_1 \text{ and} \\ e(A, B) > c_1\mu(|A|, |B|, n) \end{array} \right\}$
- $D_2 = D_2(A, B) = \left\{ \begin{array}{l} e(A, B) \leq r\mu(A, B, n), r = r_1 \text{ and} \\ e(A, B) \log \frac{e(A, B)}{\mu(A, B, n)} > c_2(|A| \vee |B|) \log \frac{n}{|A| \vee |B|} \end{array} \right\}$
- $D_3 = D_3(A, B) = \{e(A, B) > r\mu(A, B, n)\}$.

For D_1 the conditions are mutually exclusive as $e(A, B)$ can not be simultaneously greater than and less than or equal to $c_1\mu(A, B, n)$. Thus, $D_1(A, B)$ is empty. For D_2 we get similar contradiction after a little work.

$$e(A, B) \log \frac{e(A, B)}{\mu(A, B, n)} > c_2(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}$$

$$e(A, B) \log \frac{e(A, B)}{\mu(A, B, n)} > \mu(A, B, n)r_1 \log r_1$$

$$\frac{e(A, B)}{\mu(A, B, n)} \log \frac{e(A, B)}{\mu(A, B, n)} > r_1 \log r_1$$

$$\frac{e(A, B)}{\mu(A, B, n)} > r_1$$

$$e(A, B) > r_1\mu(A, B, n)$$

$$e(A, B) > r\mu(A, B, n).$$

This is a contradiction so $D_2(A, B)$ is also empty.

Now we bound $\mathbb{P}(D_3(A, B))$. As $e(A, B)$ is binomial with mean at most $\mu(A, B, n)$, Lemma A.2 implies

$$\mathbb{P}(D_3(A, B)) \leq \exp\left(-\frac{\mu(|A|, |B|, n)r \log r}{3}\right)$$

for any $r \geq 4$.

For all A, B we have $D \subset D_1 \cup D_2 \cup D_3$ and $\mathbb{P}(D_1(A, B)) = \mathbb{P}(D_2(A, B)) = 0$. Combining this with (11) and (12) we get

$$\begin{aligned}
\mathbb{P}(D) &\leq \mathbb{P}(\exists A, B : D(A, B) \text{ occurs}) \\
&\leq \mathbb{P}(\exists A, B : |A|, |B| < n/e \text{ and } D(A, B) \text{ occurs}) + O(\exp(-md)) \\
&\leq \mathbb{P}(\exists A, B : |A|, |B| < n/e \text{ and } D_3(A, B) \text{ occurs}) + O(\exp(-md)) \\
&\leq \sum_{|A|, |B|} \mathbb{P}(D_3(A, B)) + O(\exp(-md)) \\
&\leq \sum_{a, b} \sum_{|A|=a, |B|=b} \exp\left(-\frac{\mu r \log r}{3}\right) + O(\exp(-md)) \\
&\leq \sum_{a, b} \binom{n}{a} \binom{n}{b} \exp\left(-\frac{\mu(a, b, n) r \log r}{3}\right) + O(\exp(-md)),
\end{aligned}$$

where the sums are over all pairs (a, b) with $d^{1/4}/100 \leq a \vee b \leq n/e$. To evaluate the last term we get

$$\begin{aligned}
\frac{\mu r \log r}{3} &\geq (6 + 400m) \left((|A| \vee |B|) \log \frac{n}{|A| \vee |B|} \right) \\
&> (2 + 2 + 2 + (400m)) \left((|A| \vee |B|) \log \frac{n}{|A| \vee |B|} \right) \\
&> 2|A| \left(\log \frac{n}{|A|} \right) + 2|B| \left(\log \frac{n}{|B|} \right) + 2 \log n + 4md^{1/4} \log \frac{100n}{d^{1/4}} \\
&> |A|(1 + \log \frac{n}{|A|}) + |B|(1 + \log \frac{n}{|B|}) + 2 \log n + 3md^{1/4} \log n.
\end{aligned}$$

The 1st line is due to the definitions of r and c_2 . In the 3rd line we use the monotonicity of $x \log \frac{n}{x}$ on $[1, n/e]$ by substituting in $|A|, |B|, 1$ and $d^{1/4}/100$ for x . In the 4th line we use that $|A| \vee |B| \leq \frac{n}{e}$ so $\log \frac{n}{|A|}, \log \frac{n}{|B|} > 1$

Exponentiating we get

$$\exp\left[\frac{\mu r \log r}{3}\right] \geq \left(\frac{en}{|A|}\right)^n \left(\frac{en}{|B|}\right)^n n^2 \exp(3md^{1/4}(\log n)).$$

It follows that

$$\begin{aligned}
\binom{n}{a} \binom{n}{b} \exp\left(-\frac{\mu(a, b, n) r \log r}{3}\right) &\leq \binom{n}{a} \binom{n}{b} \left(\frac{en}{a}\right)^{-n} \left(\frac{en}{b}\right)^{-n} n^{-2} \exp(-3md^{1/4} \log n) \\
&\leq n^{-2} \exp(-3md^{1/4} \log n).
\end{aligned}$$

Putting this together we get

$$\begin{aligned} \mathbb{P}(D) &\leq \sum_{d/100 \leq a \vee b \leq n/e} \binom{n}{a} \binom{n}{b} \exp\left(-\frac{\mu(a,b,n)r \log r}{3}\right) + O(\exp(-md)) \\ &\leq n^2 n^{-2} \exp(-3md^{1/4} \log n) + O(\exp(-md)). \end{aligned}$$

Thus, the lemma is satisfied. ■

We finally give a quick sketch of how Proposition 5.2 follows from Lemma 5.1. This is nearly the same as Theorem 2.5 of [18], and so we will cite heavily.

Proof of Proposition 5.2 We recall that we wish to bound

$$\sup_{\substack{\|x\|=1, x^t \mathbf{1}=0 \\ \|y\|=1}} |x^t A y| \leq C\sqrt{d}.$$

For this we will relax the supremum to a finite, discrete space. Define

$$\mathcal{U} = \left\{ \frac{z}{2\sqrt{n}} : z \in \mathbb{Z}^n, \|z\|^2 \leq 4n \right\} \text{ and } \mathcal{T} = \{z \in \mathcal{U} : z \perp \mathbf{1}\}.$$

As \mathcal{U} is $\frac{1}{2}$ -net of the sphere, and $S = \{x : \|x\| = 1, x^t \mathbf{1} = 0\}$ is in the convex hull of \mathcal{T} (by Lemma 2.3 of [18]), we have that

$$\sup_{\substack{\|x\|=1, x^t \mathbf{1}=0 \\ \|y\|=1}} |x^t A y| \leq 4 \sup_{\substack{x \in \mathcal{T} \\ y \in \mathcal{U}}} |x^t A y|.$$

Further, we have that $|\mathcal{T}| \leq |\mathcal{U}| \leq C^n$ for some absolute constant C .

For a fixed pair of vectors $(x, y) \in \mathcal{T} \times \mathcal{U}$, define the **light couples** $\mathcal{L} = \mathcal{L}(x, y)$ to be all those ordered pairs $(u, v) \in \{1, 2, \dots, n\}^2$ so that $|x_u y_v| \leq \frac{\sqrt{d}}{n}$, and let the **heavy couples** $\mathcal{H} = \mathcal{H}(x, y)$ be all those pairs that are not light. We will use the notation

$$\text{light}(x, y) = \sum_{(u,v) \in \mathcal{L}} x_u A_{uv} y_v,$$

and the notation

$$\text{heavy}(x, y) = \sum_{(u,v) \in \mathcal{H}} x_u A_{uv} y_v.$$

For the light couples, we recall Bernstein's inequality, which says that for independent, centered random variables $\{X_i\}_1^N$ such that $|X_i| \leq M$ almost surely for all $1 \leq i \leq N$ and all $t \geq 0$,

$$\Pr \left[\sum_{i=1}^N X_i > t \right] \leq \exp \left(\frac{-t^2}{2 \sum_{i=1}^N \mathbb{E} X_i^2 + \frac{2}{3} M t} \right).$$

To realize $\text{light}(x, y)$ as a sum of independent variables, we need to account for the symmetry in A . Let N be the number of undirected edges $\{u, v\}$ so that either (u, v) or (v, u) appear in \mathcal{L} . Enumerate these edges and define for i with $1 \leq i \leq N$ corresponding to $\{u, v\}$,

$$X_i = (A_{uv} - p)x_u y_v \mathbf{1}[(u, v) \in \mathcal{L}] + (A_{uv} - p)x_v y_u \mathbf{1}[(v, u) \in \mathcal{L}].$$

For our purposes, it will be enough to use the bound

$$\sum_{i=1}^N \mathbb{E} X_i^2 \leq \sum_{i=1}^N 2p \left\{ (x_u y_v)^2 + (x_v y_u)^2 \right\} \leq 2p \sum_{(u,v)} x_u^2 y_v^2 \leq 2p,$$

where we have used the normalization of the vectors. In summary, by Bernstein's inequality,

$$\Pr [|\text{light}(x, y) - \mathbb{E} \text{light}(x, y)| > t] \leq \exp \left(\frac{-nt^2}{4d + \frac{2}{3} \sqrt{d} t} \right).$$

To control the expectation, note that on account of $x \in \mathcal{T}$,

$$\mathbb{E} \text{light}(x, y) + \mathbb{E} \text{heavy}(x, y) = 0.$$

However,

$$|\mathbb{E} \text{heavy}(x, y)| \leq \sum_{(u,v) \in \mathcal{H}} p |x_u y_v| \leq \sum_{(u,v) \in \mathcal{H}} \frac{np}{\sqrt{d}} |x_u y_v|^2 \leq \sqrt{d}.$$

As \mathcal{T} is only of cardinality $e^{O(N)}$, for each m there is a constant $C = C(m)$ so that

$$\Pr \left[\sup_{(x,y) \in \mathcal{T} \times \mathcal{U}} |\text{light}(x, y)| > C\sqrt{d} \right] \leq C e^{-mn}.$$

To control the heavy couples, we use the discrepancy property (c.f. Corollary 2.11 of [18] or Section 2.3 of [22]). The proof is nearly identical to either of those two claims, although it is not exactly either one, on account of the slightly altered definition of discrepancy. ■

Lemma 9.1. Suppose c_1, c_2, C_1 are constants greater than 1 and $d > 0$. There is a constant $C > 0$ depending only on c_1, c_2, C_1 so that for any graph with the property that all degrees are bounded by $C_1 d$ and for all subsets A and B of vertices

- (1) $\frac{e(A,B)}{\mu(A,B)} \leq c_1$
- (2) $e(A, B) \log \frac{e(A,B)}{\mu(A,B)} \leq c_2(|A| \vee |B|) \log \frac{n}{|A| \vee |B|}$
- (3) $|A| \vee |B| \leq d^{1/4}/100$

then for all $x, y \in \mathcal{U}$

$$\sum_{\{u,v\} \in \mathcal{H}} |x_u A_{u,v} y_v| \leq C\sqrt{d}.$$

By Lemma 5.1, all these conditions hold with the desired probability, and hence the proof of Proposition 5.2 is complete.

Proof of Lemma 9.1 We will partition the summands into blocks where each term x_u or y_v has approximately the same magnitude. Let $\gamma_i = 2^i$, $n^* = \lceil \log_2 \sqrt{n} \rceil$ and put

$$\begin{aligned} A_i &= \left\{ u \mid \frac{\gamma_{i-1}}{\sqrt{n}} \leq |x_u| < \frac{\gamma_i}{\sqrt{n}} \right\}, & 0 \leq i \leq n^*. \\ B_i &= \left\{ u \mid \frac{\gamma_{i-1}}{\sqrt{n}} \leq |y_u| < \frac{\gamma_i}{\sqrt{n}} \right\}, & 0 \leq i \leq n^*. \end{aligned}$$

Let $\hat{\mathcal{H}}$ denote those pairs (i, j) so that $\gamma_i \gamma_j \geq \sqrt{d}$. The contribution of the absolute sum can, in these terms, be bounded by

$$\sum_{(u,v) \in \mathcal{H}} |x_u A_{u,v} y_v| \leq \sum_{(i,j) \in \hat{\mathcal{H}}} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j).$$

In what follows, we will bound the contribution of the summands where $|A_i| \geq |B_j|$. By symmetry, the contribution of the other summands will have the same bound. The heavy couples will now be partitioned into 6 classes $\{\hat{\mathcal{H}}_i\}_{i=1}^6$ where their contribution is bounded in a different way. Let $\hat{\mathcal{H}}_i \subseteq \hat{\mathcal{H}}$ be those pairs (i, j) that satisfy the i^{th} property from the following list but none of the prior properties:

- (1) $|A_i| < d^{1/4}/100$.
- (2) $\frac{e(A_i, B_j)}{\mu(A_i, B_j)} \leq c_1 \frac{\gamma_i \gamma_j}{\sqrt{d}}$.
- (3) $\gamma_j > \frac{1}{4} \sqrt{d} \gamma_i$.
- (4) $\log \frac{e(A_i, B_j)}{\mu(A_i, B_j)} > \frac{1}{2} \log \frac{n}{|A_i|}$.
- (5) $\frac{n}{|A_i|} > \gamma_i^4$.
- (6) $\frac{n}{|A_i|} \leq \gamma_i^4$.

■

Bounding the contribution of $\hat{\mathcal{H}}_1$

For these terms, we have that $e(A_i, B_j) \leq |A_i||B_j| \leq \frac{\sqrt{d}}{10000}$. Hence,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_1} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq \sum_{i,j=0}^{n^*} \frac{\gamma_i \gamma_j}{n} \frac{\sqrt{d}}{10000} \leq \frac{16\sqrt{d}}{10000},$$

where in the last line we have used that $\sum_{i=0}^{n^*} 2^i \leq 4\sqrt{n}$.

Bounding the contribution of $\hat{\mathcal{H}}_2$

Applying the bound directly to the sum, we have that

$$\sum_{(i,j) \in \hat{\mathcal{H}}_2} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq c_1 \sum_{(i,j) \in \hat{\mathcal{H}}_2} \frac{\gamma_i^2 \gamma_j^2}{n\sqrt{d}} \mu(A_i, B_j) = c_1 \sqrt{d} \sum_{(i,j) \in \hat{\mathcal{H}}_2} \frac{\gamma_i^2 \gamma_j^2}{n} \frac{|A_i||B_j|}{n}.$$

Further,

$$\sum_{i=0}^{n^*} \frac{\gamma_i^2 |A_i|}{n} \leq 4 \sum_{u=1}^n |x_u|^2 \leq 4,$$

and the same bound holds for the sum over $|B_j|$. Hence,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_2} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq c_1 \sqrt{d} \sum_{i,j=0}^{n^*} \frac{\gamma_i^2 \gamma_j^2}{n} \frac{|A_i||B_j|}{n} = 16c_1 \sqrt{d}.$$

Bounding the contribution of $\hat{\mathcal{H}}_3$.

By the bound on the degrees, we have that $e(A_i, B_j) \leq C_1 |B_j| d$. Hence,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_3} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq C_1 d \sum_{(i,j) \in \hat{\mathcal{H}}_3} \frac{\gamma_i \gamma_j}{n} |B_j|.$$

Since $\gamma_i < 4\gamma_j/\sqrt{d}$, upon summing over all possible i , we get that for fixed j ,

$$\sum_{i:(i,j) \in \hat{\mathcal{H}}_3} \gamma_i \leq \frac{8\gamma_j}{\sqrt{d}}.$$

Therefore,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_3} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq c_1 \sqrt{d} \sum_{j=0}^{n^*} \frac{8\gamma_j^2}{n} |B_j| \leq 32c_1 \sqrt{d}.$$

Bounding the contribution of $\hat{\mathcal{H}}_4$.

As we are not in $\hat{\mathcal{H}}_1$ or $\hat{\mathcal{H}}_2$, it must be that $(i, j) \in \hat{\mathcal{H}}_4$ satisfy the 2nd discrepancy condition, that is,

$$\frac{1}{2} e(A_i, B_j) \log \frac{n}{|A_i|} \leq e(A_i, B_j) \log \frac{e(A_i, B_j)}{\mu(A_i, B_j)} \leq c_2 |A_i| \log \frac{n}{|A_i|}.$$

Hence, applying this bound and summing over all j so that $\gamma_j \leq \frac{1}{4} \sqrt{d} \gamma_i$,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_4} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq c_2 \sqrt{d} \sum_{i=0}^{n^*} \gamma_i^2 \frac{|A_i|}{n} \leq 4c_2 \sqrt{d}.$$

Bounding the contribution of $\hat{\mathcal{H}}_5$.

For $(i, j) \in \hat{\mathcal{H}}_5$ we have

$$e(A_i, B_j) \leq \mu(A_i, B_j) \left(\frac{n}{|A_i|} \right)^{1/2} = d |B_j| \left(\frac{n}{|A_i|} \right)^{-1/2} \leq d |B_j| \gamma_i^{-2}.$$

Hence,

$$\sum_{(i,j) \in \hat{\mathcal{H}}_5} \frac{\gamma_i \gamma_j}{n} e(A_i, B_j) \leq \sum_{(i,j) \in \hat{\mathcal{H}}_5} \frac{d \gamma_j^2 |B_j|}{n \gamma_i \gamma_j} \leq \frac{2}{\sqrt{d}} \sum_{j=0}^{n^*} \frac{d \gamma_j^2 |B_j|}{n} \leq 8\sqrt{d},$$

where we have used in the penultimate bound that the sum over i is dominated by the series

$$\sum_{i: \sqrt{d} \leq \gamma_j \gamma_i} \frac{1}{\gamma_i} \leq \frac{2\gamma_j}{\sqrt{d}}.$$

Bounding the contribution of $\hat{\mathcal{H}}_6$.

For $(i, j) \in \hat{\mathcal{H}}_6$, we have that

$$e(A_i, B_j) \log \frac{c_1 \gamma_i \gamma_j}{\sqrt{d}} \leq e(A_i, B_j) \log \frac{e(A_i, B_j)}{\mu(A_i, B_j)} \leq c_2 |A_i| \log \frac{n}{|A_i|} \leq 4c_2 |A_i| \log \gamma_i.$$

This brings us to the bound

$$\sum_{(i,j) \in \hat{\mathcal{H}}_6} \frac{\gamma_i \gamma_j}{n} e^{(A_i, B_j)} \leq 4c_2 \cdot \sum_{(i,j) \in \hat{\mathcal{H}}_6} \frac{\gamma_i |A_i| \log \gamma_i}{n} \frac{\gamma_j}{\log(c_1 \gamma_i \gamma_j) - \log \sqrt{d}}.$$

The sum in j only runs over those terms such that $4\gamma_j \leq \sqrt{d}\gamma_i$ and such that $\gamma_j \gamma_i \geq \sqrt{d}$. For j such that $\gamma_j \leq \gamma_i \sqrt{d}/(1 + \log(\gamma_i))$ we bound the sum over j by

$$\sum_j \frac{\gamma_j}{\log(c_1 \gamma_i \gamma_j) - \log \sqrt{d}} \leq \sum_j \frac{\gamma_j}{\log c_1} \leq \frac{2\gamma_i \sqrt{d}}{(\log c_1)(1 + \log \gamma_i)}.$$

For larger j , we bound the sum by

$$\sum_j \frac{\gamma_j}{\log(c_1 \gamma_i \gamma_j) - \log \sqrt{d}} \leq \sum_j \frac{\gamma_j}{\log c_1 \gamma_i^2 - \log(1 + \log \gamma_i)} \leq \frac{\gamma_i \sqrt{d}}{2(\log c_1)(\log \gamma_i)},$$

having applied the inequality $\log(1 + x) \leq x$. Hence, we conclude that

$$\sum_{(i,j) \in \hat{\mathcal{H}}_6} \frac{\gamma_i \gamma_j}{n} e^{(A_i, B_j)} \leq \frac{10c_2 \sqrt{d}}{\log c_1} \cdot \sum_{(i,j) \in \hat{\mathcal{H}}_6} \frac{\gamma_i^2 |A_i|}{n} \leq \frac{40c_2 \sqrt{d}}{\log c_1}.$$

A Estimates of Binomial Random Variables

Lemma A.1. Let X be a binomial random variable with mean μ . Then for any $t \leq \mu$,

$$\mathbb{P}[X \leq t] \leq \exp[-\mu + t(1 + \log \frac{\mu}{t})].$$

Proof of Lemma A.1 The proof follows from a standard estimate on the Laplace transform combined with Markov's inequality. For any $\lambda \in \mathbb{R}$, the Laplace transform of $X \sim \text{Binomial}(n, p)$ can be bounded by

$$\begin{aligned} \mathbb{E}e^{\lambda X} &= (pe^\lambda + (1-p))^n \\ &= (1 + p(e^\lambda - 1))^n \\ &\leq \exp[\mu(e^\lambda - 1)]. \end{aligned}$$

Provided that $\lambda < 0$, the tail bound now can be bounded by Markov's inequality by

$$\begin{aligned}\mathbb{P}[X \leq t] &= \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda t}\right] \\ &\leq \left[\mathbb{E}e^{\lambda X}\right] e^{-\lambda t} \\ &\leq \exp\left[\mu(e^\lambda - 1) - \lambda t\right].\end{aligned}$$

Assuming that $t < \mu$, this bound holds with $\lambda = \log(t/\mu)$, which upon evaluation gives

$$\mathbb{P}[X \leq t] \leq \exp\left[\mu(e^{\log(t/\mu)} - 1) - \log(t/\mu)t\right] = \exp\left[-\mu + t(1 + \log \frac{\mu}{t})\right]. \quad \blacksquare$$

Lemma A.2. Let X be a binomial random variable with mean μ . Then for any $t > 4$,

$$\mathbb{P}[X \geq t\mu] \leq \exp\left[-\frac{t\mu \log(t)}{3}\right].$$

Proof of Lemma A.2 The proof here is identical in approach to the proof of Lemma A.1. As there, it is possible to bound the Laplace transform of X as

$$\mathbb{E}e^{\lambda X} \leq \exp\left[\mu(e^\lambda - 1)\right],$$

for any real λ . For $\lambda > 0$, the tail bound follows from Markov's inequality by

$$\begin{aligned}\mathbb{P}[X \geq t\mu] &= \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda t\mu}\right] \\ &\leq \left[\mathbb{E}e^{\lambda X}\right] e^{-\lambda t\mu} \\ &\leq \exp\left[\mu(e^\lambda - 1) - \lambda t\mu\right].\end{aligned}$$

For $t > 1$, it is possible to take $\lambda = \log t$. This gives the bound on the tail probability

$$\mathbb{P}[X \geq t\mu] \leq \exp\left[\mu(t - 1 - t \log t)\right].$$

To complete the proof, it remains to show that $t - 1 \leq \frac{2}{3}t \log t$ when $t \geq 4$. The function $\frac{t}{t-1} \log t$ is monotonically increasing for $t > 1$, and thus it suffices to show that $\frac{4}{3} \log 4 \geq \frac{3}{2}$, or equivalently that $\log 4 \geq \frac{9}{8}$. This follows from $\log 4 = \int_1^4 \frac{1}{x} dx$ and bounding the integral from below by a right Riemann sum. \blacksquare

Funding

The work was supported by American Mathematical Society (AMS) Centennial Fellowship [to C.H.]; National Science Foundation (NSF) [DMS-0806024 to C.H., DMS-1352386 to M.K., DMS-0847661, DMS-0806024, and DMS-1304057 to E.P.]; National Security Agency (NSA) [H98230-10-1-0227 to M.K.]; Defense Advanced Research Projects Agency (DARPA) [N66001-12-1-4226 to M.K.]; and Alfred P. Sloan Research Fellowship [to M.K.].

References

- [1] Antoniuk, S., T. Łuczak, and J. Świątkowski. "Random triangular groups at density $1/3$." *Compos. Math.* 151, no. 1 (2015): 167–78.
- [2] Babson, E., C. Hoffman, and M. Kahle. "The fundamental group of random 2-complexes." *J. Amer. Math. Soc.* 24, no. 1 (2011): 1–28.
- [3] Ballmann, W. and J. Świątkowski. "On L^2 -cohomology and property (T) for automorphism groups of polyhedral cell complexes." *Geom. Funct. Anal.* 7, no. 4 (1997): 615–45.
- [4] Bekka, B., P. de la Harpe, and A. Valette. *Kazhdan's Property (T)*. New Mathematical Monographs 11. Cambridge: Cambridge University Press, 2008.
- [5] Benaych-Georges, F., C. Bordenave, and A. Knowles. "Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs." (2017): preprint [arXiv:1704.02953](https://arxiv.org/abs/1704.02953).
- [6] Benaych-Georges, F., C. Bordenave, and A. Knowles. "Spectral radii of sparse random matrices". (2017): preprint [arXiv:1704.02945](https://arxiv.org/abs/1704.02945).
- [7] Benjamini, I., S. Haber, M. Krivelevich, and E. Lubetzky. "The isoperimetric constant of the random graph process." *Random Structures Algorithms* 32, no. 1 (2008): 101–14.
- [8] Chung, F., L. Lu, and V. Vu. "The spectra of random graphs with given expected degrees." *Internet Math.* 1, no. 3 (2004): 257–75.
- [9] Chung, F. and M. Radcliffe. "On the spectra of general random graphs." *Electron. J. Combin.* 18, no. 1 (2011): 215–14.
- [10] Chung, F. R. K. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics 92. Published for the Conference Board of the Mathematical Sciences. Washington, DC, 1997.
- [11] Cohen, D., A. Costa, M. Farber, and T. Kappeler. "Topology of random 2-complexes." *Discrete Comput. Geom.* 47, no. 1 (2012): 117–49.
- [12] Coja-Oghlan, A. "On the Laplacian eigenvalues of $G_{n,p}$." *Combin. Probab. Comput.* 16, no. 6 (2007): 923–46.
- [13] Costa, A. E. and M. Farber. "The asphericity of random 2-dimensional complexes." *Random Structures Algorithms* 46, no. 2 (2015): 261–73.
- [14] Costa, A. E. and M. Farber. "Geometry and topology of random 2-complexes." *Israel J. Math.* 209, no. 2 (2015): 883–927.
- [15] Davis, M. W. and M. Kahle. "Random graph products of finite groups are rational duality groups." *J. Topol.* 7, no. 2 (2014): 589–606.
- [16] Ding, J., J. H. Kim, E. Lubetzky, and Y. Peres. "Anatomy of a young giant component in the random graph." *Random Structures Algorithms* 39, no. 2 (2011): 139–78.
- [17] Erdős, L., A. Knowles, H.-T. Yau, and J. Yin. "Spectral statistics of Erdős–Rényi graphs II: eigenvalue spacing and the extreme eigenvalues." *Comm. Math. Phys.* 314, no. 3 (2012): 587–640.
- [18] Feige, U. and E. Ofek. "Spectral techniques applied to sparse random graphs." *Random Structures Algorithms* 27, no. 2 (2005): 251–75.
- [19] Fountoulakis, N. and B. A. Reed. "The evolution of the mixing rate of a simple random walk on the giant component of a random graph." *Random Structures Algorithms* 33, no. 1 (2008): 68–86.

- [20] Fox, J., M. Gromov, V. Lafforgue, A. Naor, and J. Pach. "Overlap properties of geometric expanders." *J. Reine Angew. Math.* 671 (2012): 49–83.
- [21] Friedgut, E. and G. Kalai. "Every monotone graph property has a sharp threshold." *Proc. Amer. Math. Soc.* 124, no. 10 (1996): 2993–3002.
- [22] Friedman, J., J. Kahn, and E. Szemerédi. "On the second eigenvalue of random regular graphs." In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing, STOC '89*, 587–98. New York: Association for Computing Machinery, 1989.
- [23] Füredi, Z. and J. Komlós. "The eigenvalues of random symmetric matrices." *Combinatorica* 1, no. 3 (1981): 233–41.
- [24] Garland, H. " p -adic curvature and the cohomology of discrete subgroups of p -adic groups." *Ann. of Math.* 97, no. 2 (1973): 375–423.
- [25] Glasner, E. *Ergodic Theory Via Joinings*. Mathematical Surveys and Monographs 101. Providence, RI: American Mathematical Society, 2003.
- [26] Gromov, M. "Singularities, expanders and topology of maps. I. Homology versus volume in the spaces of cycles." *Geom. Funct. Anal.* 19, no. 3 (2009): 743–841.
- [27] Gromov, M. "Singularities, expanders and topology of maps. Part 2: from combinatorics to topology via algebraic isoperimetry." *Geom. Funct. Anal.* 20, no. 2 (2010): 416–526.
- [28] Gundert, A. and U. Wagner. "On Laplacians of random complexes." In *Computational Geometry (SCG'12)*, 151–60. New York: Association for Computing Machinery, 2012.
- [29] Hatcher, A. *Algebraic Topology*. Cambridge, New York: Cambridge University Press, 2002. Autre(s) tirage(s): 2003, 2004, 2005, 2006.
- [30] Hoffman, C., M. Kahle, and E. Paquette. "The threshold for integer homology in random d -complexes." *Discrete Comput. Geom.* 57, no. 4 (2017): 810–23.
- [31] Janson, S., D. E. Knuth, T. Łuczak, and B. Pittel. "The birth of the giant component." *Random Structures Algorithms* 4, no. 3 (1993): 233–358.
- [32] Kahle, M. "Topology of random clique complexes." *Discrete Math.* 309, no. 6 (2009): 1658–71.
- [33] Kahle, M. "Sharp vanishing thresholds for cohomology of random flag complexes." *Ann. of Math. (2)* 179, no. 3 (2014): 1085–107.
- [34] Le, C. M., E. Levina, and R. Vershynin. "Concentration and regularization of random graphs." *Random Structures Algorithms* 51, no. 3 (2017): 538–61.
- [35] Linial, N. and R. Meshulam. "Homological connectivity of random 2-complexes." *Combinatorica* 26, no. 4 (2006): 475–87.
- [36] Linial, N. and Y. Peled. "On the phase transition in random simplicial complexes." *Ann. of Math. (2)* 184, no. 3 (2016): 745–73.
- [37] Łuczak, T. and Y. Peled. "Integral homology of random simplicial complexes." *Discrete Comput. Geom.* 59, no. 1 (2018): 131–42.
- [38] Margulis, G. A. "Explicit constructions of expanders." *Problemy Peredachi Informatsii* 9, no. 4 (1973): 71–80.
- [39] Meshulam, R. and N. Wallach. "Homological connectivity of random k -dimensional complexes." *Random Structures Algorithms* 34, no. 3 (2009): 408–17.

- [40] Pach, J. "A Tverberg-type result on multicolored simplices." *Comput. Geom.* 10, no. 2 (1998): 71–6.
- [41] Parzanchevski, O., R. Rosenthal, and R. J. Tessler. "Isoperimetric inequalities in simplicial complexes." *Combinatorica* 36, no. 2 (2016): 195–227.
- [42] Ross, N. "Fundamentals of Stein's method." *Probab. Surv.* 8 (2011): 210–93.
- [43] Vu, V. H. "Spectral norm of random matrices." *Combinatorica* 27, no. 6 (2007): 721–36.
- [44] Wigner, E. P. "On the distribution of the roots of certain symmetric matrices." *Ann. of Math.* (2) 67 (1958): 325–7.
- [45] Žuk, A. "La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres." *C. R. Acad. Sci. Paris Sér. I Math.* 323, no. 5 (1996): 453–8.
- [46] Žuk, A. "Property (T) and Kazhdan constants for discrete groups." *Geom. Funct. Anal.* 13, no. 3 (2003): 643–70.