Topology of random simplicial complexes: a survey

Matthew Kahle

1. Introduction

This expository article is based on a lecture from the Stanford Symposium on Algebraic Topology: Application and New Directions, held in honor of Gunnar Carlsson, Ralph Cohen, and Ib Madsen.

> "I predict a new subject of statistical topology. Rather than count the number of holes, Betti numbers, etc., one will be more interested in the distribution of such objects on noncompact manifolds as one goes out to infinity." — Isadore Singer, 2004 [48]

1.1. Earlier work. Random triangulated surfaces were studied by Pippenger and Schleich [46]. Their model is randomly gluing together n oriented triangles, uniformly over all such gluings, and they compute the expected genus $\mathbb{E}[g_n]$ of the resulting oriented surface as $n \to \infty$. Such random surfaces arise in 2-dimensional quantum gravity and as world-sheets in string theory.

Dunfield and Thurston considered this same model around the same time, and they pointed out that in general one can not make a random 3-manifold by gluing together tetrahedra in an analogous way, as the probability that a gluing results in a manifold tends to 0 as the number of tetrahedra $n \to \infty$. They introduced a new model for random 3-manifolds M where one takes a random walk on the mapping class group, resulting in a random gluing of a two handlebodies [17]. They were able to compute the probabilities that the resulting manifolds have finite covers of particular kinds.

The configuration space for a random planar linkage, where the number of links tends to infinity, was a different kind of random manifold introduced by Farber and Kappeler [20]. They gave a formula the mean of the Betti numbers, and showed exponential concentration around the mean.

Random functions on manifolds were studied by Adler and Taylor, and they discovered the "Gaussian kinematic" formula for the expected Euler characteristic of the sub-level sets. See for example, Chapter 12 of their book [2]. Giving a formula for the expectation of the individual Betti numbers seems to be an open problem. For a survey of this area see [1].

We will will focus on random simplicial complexes.

Random 2-dimensional simplicial complexes were first studied by Linial and Meshulam [**41**], and the k-dimensional version by Meshulam and Wallach [**43**]. One way to think about these kinds of models is as higher-dimensional analogues of random graphs — for example the main results of [**41**] and [**43**] are cohomological analogues of the Erdős–Rényi theorem which characterizes the threshold for connectivity of a random graph. These kinds of theorems describe topological phase transitions where cohomology passes from nontrivial to trivial within a very short window of parameter.

Since the influential paper [**41**], random complexes and their topological properties have continued to be explored by several teams of researchers [**43, 7, 39, 4, 5, 35**]. Random clique complexes provide another way of generalizing random graphs to higher dimensions [**31, 33**], which puts a measure on a wide range of topologies.

1.2. Motivation.
1.2.1. *Randomness models the natural world.*

(1) The field of topological data analysis has received a lot of attention over the past several years — see for example the survey articles [**14, 22**]. In order to quantify statistical significance of topological features, it will be necessary to have firmer probabilistic foundations. In other words, random topology may be viewed as the null hypothesis for statistical topology.

(2) Certain situations in physics seem to be well modeled by probabilistic topology. For one famous example, John Wheeler suggested in the 1960's that inside a black hole, one would need a topological and geometric theory to account for relativity, and that near the presumed singularity one would also require a quantum theory, necessarily stochastic. He reasoned that inside a black hole the topology of space-time itself may best be understood as a probability distribution over shapes, rather than any particular fixed shape.

(3) One might also want to understand why certain mathematical phenomena are so ubiquitous. For example, a folklore theorem is that "almost all groups are hyperbolic". This turns out to be true under a variety of different measures — see Ollivier's survey [**44**].

It is known that many simplicial complexes and posets found in the wild are homotopy equivalent to wedges of spheres of the same dimension, and many researchers in topological combinatorics have wondered if there is a deeper reason why [**21**]. Is it possible that asymptotically almost all complexes are relatively simple, topologically speaking? One the goal of this article is to describe a yes to this question, at least according to certain natural measures, and in terms of homology with rational coefficients.

(4) Random topology might also model certain number-theoretic objects.

The following natural class of simplicial complexes were introduced by Björner [**10**]: Δ_n has primes less than n as its vertices, and its faces correspond to square-free numbers i with $1 \le i \le n$. He pointed out that the Euler characteristic $\chi(\Delta_n)$ coincides with the Mertens function

$$M(n) = \sum_{k=1}^{n} \mu(k),$$

where $\mu(k)$ is the Möbius function. The Riemann hypothesis is well known to be equivalent to the statement that $M(n)$ satisfies

$$|M(n)| = O(n^{1/2+\epsilon})$$

for every fixed $\epsilon > 0$, suggesting that studying the topology of Δ_n might be quite interesting.

Björner proved these complexes are all homotopy equivalent to wedges of spheres (not necessarily of the same dimension), hence homology $H_*(\Delta_n)$ is torsion-free. He also provided some estimates for Betti numbers, showing that

$$\beta_k \approx \frac{n}{2\log n} \frac{(\log\log n)^k}{k!}$$

for k fixed, and also that

$$\sum_{k \geq 0} \beta_k\left(\Delta_n\right) = \frac{2n}{\pi^2} + O\left(n^\theta\right)$$

for all $\theta > 17/54$.

Unfortunately, this does not seem to get us any closer to proving the Riemann hypothesis, but Björner's work suggests further study. The primes are not random, but they are "pseudorandom" and for many purposes behave as if they were a random subset of the integers with density predicted by the prime number theorem — the Green–Tao theorem is a celebrated example [23].

1.2.2. *The probabilistic method provides existence proofs.* This is a complementary point of view. Random objects often have desirable properties, and in some cases it is difficult to construct explicit examples. This has been one of the most influential ideas in discrete mathematics of the past several decades — for a broad overview of the subject, see Alon and Spencer's book [3].

(1) In Ramsey theory and extremal graph theory, the probabilistic method has proved to be an extremely powerful tool. Almost all graphs are known to have strong Ramsey properties (i.e. no large cliques or independent sets), but after several decades of research no one knows how to give large explicit examples nearly as strong. This paradoxical situation is sometimes referred to as the problem of finding hay in a haystack.

(2) Since early work of Pinsker [45], and even earlier work of Barzdin and Kolmogorov [9], it has been known that "almost all graphs are expanders." (For a survey of expander graphs and their applications, see [30].) Some of the work surveyed in this article may be viewed as higher-dimensional analogues of this paradigm. One of our goals is to describe expander-like qualities of random simplicial complexes.

(3) The probabilistic method has found applications in other areas of mathematics as well. Gromov asked, "What does a random group look like? As we shall see the answer is most satisfactory: nothing like we have ever seen before," [24], and then later fulfilled his own prediction by proving the existence of a finitely generated group Γ whose Cayley graph admits no uniform embedding in the Hilbert space [25]. His argument is a probabilistic one.
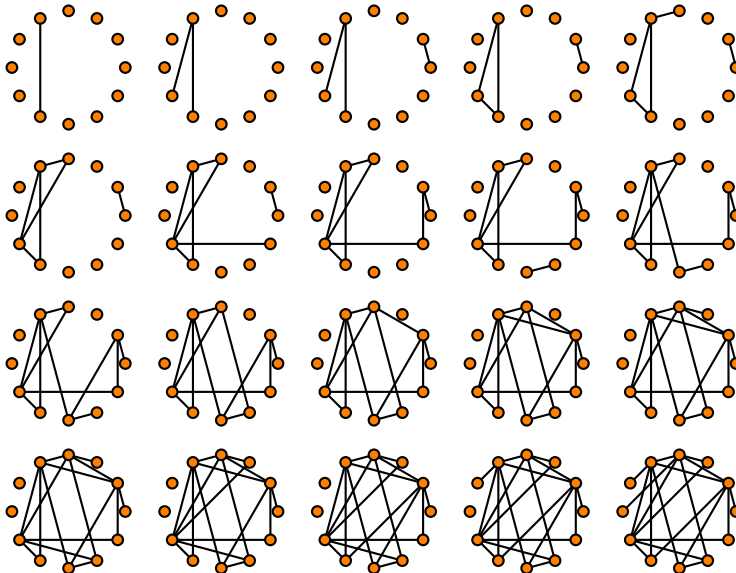
FIGURE 1. The beginning of a random graph process on $n = 12$ vertices.

2. Random graphs

The random graph $G(n,p)$ has vertex set $[n] = \{1,2,\ldots,n\}$ and each edge appears independently with probability p, independently. The closely related random graph $G(n,m)$ is selected uniformly among all $\binom{\binom{n}{2}}{m}$ graphs on n vertices with m edges.

One can also think about random graphs as part of a stochastic process. In the andom graph process

$$\{G(n,m)\}_{m=1}^{\binom{n}{2}},$$

for example, the mth edge is selected uniformly randomly from the remaining $n - \binom{m}{2} - 1$ edges. See Figure 1 for the beginning of a random graph process.

In random graph theory, one is usually interested in the asymptotic behavior of such graphs as $n \to \infty$, and $p = p(n)$. A celebrated theorem about random graphs, but also perhaps the most the topology of random graphs is the following [19].

THEOREM 2.1 (Erdős–Rényi, 1959). *Let $\epsilon > 0$ be fixed. Then as $n \to \infty$,*

$$\mathbb{P}[G(n,p) \text{ is connected}] \to \begin{cases} 1 & : p \geq \frac{(1+\epsilon)\log n}{n} \\ \\ 0 & : p \leq \frac{(1-\epsilon)\log n}{n} \end{cases}$$

The Erdős–Rényi theorem is actually slightly sharper than this, as follows. Let

$$\tilde{\beta}_0(G) = \# \text{ of connected components of } G - 1,$$

i.e. to a topologist, the reduced 0th Betti number of G.

THEOREM 2.2 (Erdős–Rényi, 1959). *Let $G = G(n,p)$ where*

$$p = \frac{\log n + c}{n},$$

and $c \in \mathbb{R}$ is fixed. Then as $n \to \infty$, $\tilde{\beta}_0(G)$ is asymptotically Poisson distributed with mean e^{-c}. In particular,

$$\mathbb{P}[G \text{ is connected}] \to e^{-e^{-c}}$$

COROLLARY 2.3. *Let $\omega \to \infty$ arbitrarily slowly as $n \to \infty$. Then*

$$\mathbb{P}[G(n,p) \text{ is connected}] \to \begin{cases} 1 & : p \geq \frac{\log n + \omega}{n} \\ 0 & : p \leq \frac{\log n - \omega}{n} \end{cases}$$

We sometimes write "with high probability (w.h.p.)" for an event if the probability of the event tends to 1 as the number of vertices $n \to \infty$.

Why is $p = \log n / n$ the right answer here? To answer this, we set $p = (\log n + c)/n$ and ask how many isolated vertices we expect to see. The probability that a vertex is isolated is $(1 - p)^{n-1}$, by independence. Then by linearity of expectation, the expected number of isolated vertices X_0 is

$$\mathbb{E}[X_0] = n(1 - p)^{n-1}.$$

It follows easily that $\mathbb{E}[X_0] \to e^{-c}$ as $n \to \infty$, and with only a little more work one can show that X_0 is asymptotically Poisson distributed with mean e^{-c}. See, for example, Chapter 8 of [3] to learn how to prove limit theorems like this using the method of moments.

To finish the proof of Theorem 2.2, one also needs a structure theorem, namely that for p is in this range, $G(n, p)$ w.h.p. consists of only two kinds of connected components: a unique "giant component", and isolated vertices. Given this structure, the graph is connected if and only if there are no isolated vertices. See Chapter 7 of [11] for a complete proof. Moreover, we can understand the limiting distribution for the total number of connected components by understanding the limiting distribution for the number of isolated vertices.

Corollary 2.3 shows that $p = \log n / n$ is a *sharp threshold* for connectivity of $G(n, p)$, meaning that the phase transition from probability 0 to probability 1 happens within a very narrow window. More precisely, a function f is said to be a sharp threshold for a graph property $\mathcal{P}$ if there exists a function $g = o(f)$ such that

$$\mathbb{P}[G(n,p) \in \mathcal{P}] \to \begin{cases} 1 & : p \geq f + g \\ 0 & : p \leq f - g \end{cases}$$

The message of Corollary 2.3 is that the threshold function for "G is connected" is the same as the threshold function for "G has no isolated vertices." The following result of Bollobás and Thomasson takes this idea all the way to its logical conclusion [12].

THEOREM 2.4. *For a random graph process $\{G(n, m)\}_{m=1}^{\binom{n}{2}}$, with high probability*

$$\min\{M : G(n, M) \text{ has no isolated vertices}\} = \min\{M : G(n, M) \text{ is connected}\}.$$

We leave it to the reader to convince themselves that Theorem 2.4 is even sharper than Corollary 2.3.
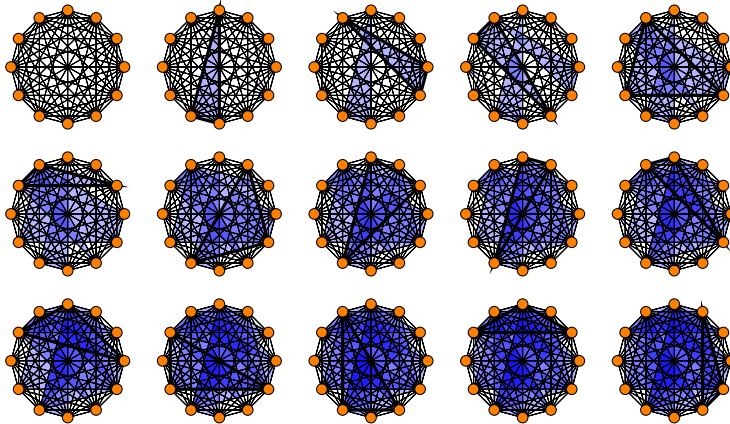
FIGURE 2. The beginning of a random 2-complex process on $n = 12$ vertices.

It is worth noting that there is another important topological phase transition for $G = G(n, p)$, namely where cycles first appear, or to a topologist, where $H_1(G)$ first becomes nontrivial. See [**47**] for a proof of the following.

THEOREM 2.5. *Let $p = c/n$ where $c > 0$ is constant.*

$$\mathbb{P}[G(n,p) \text{ has no cycles}] \to \begin{cases} 1 & : c \geq 1 \\ \frac{\sqrt{1-c}}{\exp(c/2 + c^2/4)} & : c < 1 \end{cases}$$

In contrast to the connectivity threshold, the threshold described in Theorem 2.5 is not sharp in the sense described above, or one might say more precisely that is sharp on one side, but not on the other.

3. Random 2-complexes

Linial and Meshulam initiated the topological study of random 2-dimensional simplicial complexes $Y(n, p)$ in [**41**]. This model random simplicial complex is defined to have vertex set $[n]$, edge set $\binom{[n]}{2}$ (i.e. the underlying graph is a complete graph), and each of the $\binom{n}{3}$ possible triangle faces is included with the same probability $p = p(n)$, independently. A random 2-complex complex process is illustrated in Figure 2.

The main result of [**41**] is a perfect cohomological analogue of Theorem 2.1.

THEOREM 3.1 (Linial–Meshulam, 2006). *Let $\epsilon > 0$ be fixed and $Y = Y(n, p)$. Then as $n \to \infty$,*

$$\mathbb{P}[H^1(Y, \mathbb{Z}/2) = 0] \to \begin{cases} 1 & : p \geq (2 + \epsilon) \log n / n \\ 0 & : p \leq (2 - \epsilon) \log n / n \end{cases}$$

Although Theorem 3.1 is analogous to Theorem 2.1, the proof is much harder. Perhaps this is not surprising — cohomology is after all cocycles modulo coboundaries, and in degree 0 there are not too many coboundaries. The combinatorics are considerably more complicated in degree 1.

A few comments are in order.

(1) The Linial–Meshulam theorem is sharper than this, analogous to Corollary 2.3 but we sometimes trade the strongest result for a simpler statement.

(2) The threshold $p = 2 \log n / n$ is exactly what is required in order to ensure that there are no isolated edges. The analogue of the "stopping time" Theorem 2.4 for the random 2-complex process was recently established in [**35**].

(3) It was shown by Meshulam and Wallach [**43**] that the same result holds with $(\mathbb{Z}/\ell)$-coefficients for every fixed ℓ. The threshold for $\mathbb{Z}$-coefficients is still unknown. It might seem that the $\mathbb{Z}$-threshold would follow from this, but the problem is that there could be ℓ-torsion growing with n. We return to speculation about random torsion in Section 4.

(4) As mentioned in the original paper, the argument is cohomological but then universal coefficients for homology and cohomology give the corresponding result for homology.

The threshold for simple connectivity has been shown to be much larger [**7**].

THEOREM 3.2 (Babson et al., 2011). *Let $\epsilon > 0$ be fixed and $Y = Y(n, p)$. Then as $n \to \infty$,*

$$\mathbb{P}[\pi_1(Y) = 0] \to \begin{cases} 1 & : p \geq \frac{n^{\epsilon}}{\sqrt{n}} \\ \\ 0 & : p \leq \frac{n^{-\epsilon}}{\sqrt{n}} \end{cases}$$

On the way to showing the fundamental group is nontrivial when $p \leq \frac{n^{-\epsilon}}{\sqrt{n}}$, one first shows that it is word hyperbolic.

The study of fundamental groups of random 2-complexes is continued in [**29**]. A group G is said to have Kazhdan's property (T) if the trivial representation is an isolated point in the unitary dual of G equipped with the Fell topology. More intuitively, Property (T) is an "expander-like" property of groups, and first explicit examples of expanders, due to Margulis, were constructed from Cayley graphs of quotients of (T) groups such as $SL_3(\mathbb{Z})$.

The following theorem shows that the threshold for $\pi_1(Y)$ to have Kazhdan's Property (T) is the same as for $H^1(Y, \mathbb{Z}/2)$ to vanish [**29**].

THEOREM 3.3 (Hoffman et al., 2012). *Let $\epsilon > 0$ be fixed and $Y = Y(n, p)$. Then as $n \to \infty$,*

$$\mathbb{P}[\pi_1(Y) \text{ is } (T)] \to \begin{cases} 1 & : p \geq (2 + \epsilon) \log n / n \\ \\ 0 & : p \leq (2 - \epsilon) \log n / n \end{cases}$$

The proof that $Y(n, p)$ is (T) when $p \geq (2 + \epsilon) \log n / n$ utilizes the following theorem of Żuk [**51**].

THEOREM 3.4 (Żuk). *If X is a pure 2-dimensional locally-finite simplicial complex so that for every vertex v, the vertex link $lk_v(X)$ is connected and the normalized graph Laplacian $L = L(lk_v(X))$ has smallest positive eigenvalue $\lambda_2(L) > 1/2$, then $\pi_1(X)$ has property (T).*

The link of a vertex in the random 2-complex $Y(n, p)$ has the same probability distribution as a random graph $G(n - 1, p)$. So Żuk's theorem reduces the proof of Theorem 3.3 to a question about Laplacians of random graphs. However, new

results about such Laplacians are still required in order to prove Theorem 3.3. Establishing the following comprises most of the work in [**29**].

THEOREM 3.5. *(Hoffman et al., 2012) Fix $k \geq 0$ and $\epsilon > 0$. Let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of the normalized Laplacian of the random graph $G(n, p)$. There is a constant $C = C(k)$ so that when*

$$p \geq \frac{(k+1)\log n + C\sqrt{\log n}\log\log n}{n}$$

is satisfied, then

$$\lambda_2 > 1 - \epsilon,$$

with probability at least $1 - o(n^{-k})$.

The proof of Theorem 3.3 only requires Theorem 3.5 with $k = 1$ and $\epsilon = 1/2$. The proof is almost immediate once Theorem 3.5 is established: There are n vertex links, and for each one, the probability that its spectral gap is smaller than $1/2$ is $o(n^{-1})$. So the probability that there is at least one vertex link with spectral gap too small is $o(1)$ by a union bound — the probability that at least one bad event occurs is never more than the sum of the probabilities of the individual bad events.

The $k = 0$ case of Theorem 3.5 is also of particular interest, as this gets very close to the connectivity threshold for $G(n, p)$. It would be interesting to see just how close one can get. Consider a random graph process, for example, adding one edge at a time: is the graph already an expander at the moment of connectivity? We discuss applications of Theorem 3.5 with other values of k in Section 4.

Both Theorem 3.1 and Theorem 3.3 have the following corollary.

COROLLARY 3.6. *Let $\epsilon > 0$ be fixed and $Y = Y(n, p)$. Then as $n \to \infty$,*

$$\mathbb{P}[H_1(Y, \mathbb{Q}) = 0] \to \begin{cases} 1 & : p \geq (2 + \epsilon)\log n/n \\ \\ 0 & : p \leq (2 - \epsilon)\log n/n \end{cases}$$

This follows from Theorem 3.1 by universal coefficients for homology and co-homology, and follows from Theorem 3.3, since a finitely presented group with Property (T) has finite abelianization.

Homology vanishing for any choice of coefficients, simple connectivity, and Property (T) are all monotone properties for random 2-complexes, meaning that if at some point in the random complex process, you have one of these properties, the property continues to hold from that point on. This is in contrast to what we will see in Section 4, where each homology group passes through two distinct phase transitions, vanishing-to-nonvanishing and nonvanishing-to-vanishing.

4. Random flag complexes

The *flag complex* $X(H)$ of a graph H is the maximal simplicial complex compatible with H as its 1-skeleton; in other words, the i-dimensional faces of $X(H)$ correspond to the cliques of order $i + 1$ in H. (Such complexes have apparently arisen independently several times, and $X(H)$ is also sometimes called the *clique complex* or the *Vietoris–Rips complex* of H.)

We define the random flag complex $X(n, p)$ to be the flag complex of the random graph $G(n, p)$. Every simplicial complex is homeomorphic to a flag complex, e.g. by taking the barycentric subdivision. So $X(n, p)$ puts a measure on a wide variety of topologies as $n \to \infty$.
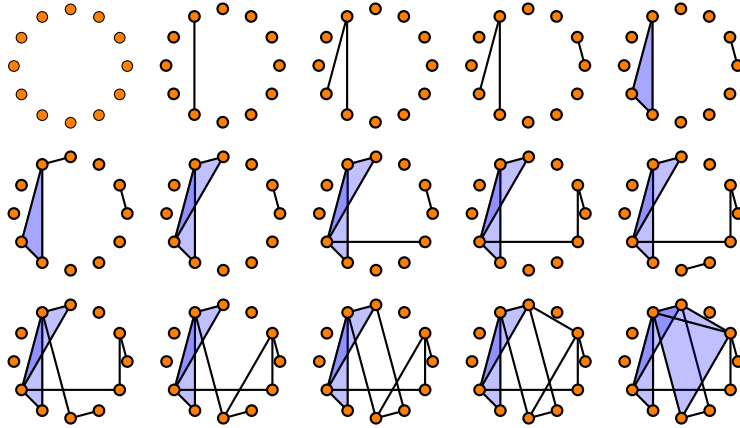
FIGURE 3. A random flag complex process. Vanishing of H_k for $k \geq 1$ is no longer a monotone property.
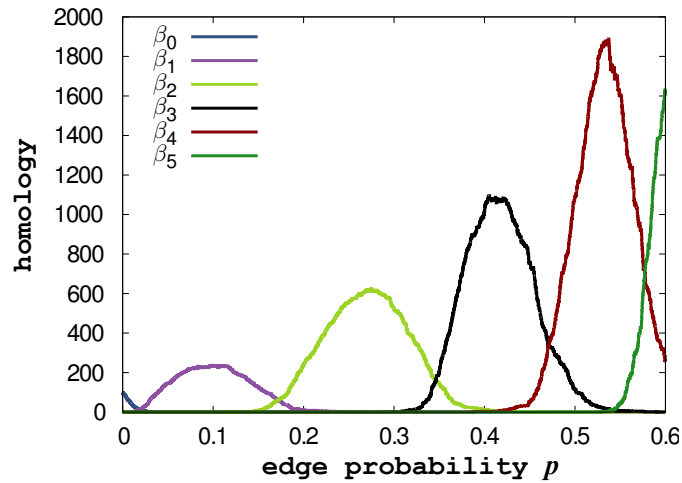

FIGURE 4. The Betti numbers for a random flag complex process on $n = 100$ vertices. *(Computation and image courtesy of Afra Zomorodian.)*

One can also consider a random flag complex process, which is the same probability space as the random graph process, edges being added one at a time. See Figure 3. The Betti numbers of an instance of such a process on $n = 100$ vertices and roughly 3000 steps are illustrated in Figure 4.

We see immediately that homology no longer behaves in a monotone way with respect to the underlying parameter. Instead homology passes through two phase transitions, vanishing-nonvanishing and nonvanishing-vanishing, and that in-between the dimension of homology is roughly unimodal.

4.1. Vanishing homology. First of all, we check that, as suggested by Figure 4, there is a range of p outside of which $H_k = 0$ with high probability [**31**].

THEOREM 4.1. *Let $\epsilon > 0$ be fixed and $X = X(n, p)$.*

(1) *If*

$$p \leq \frac{n^{-\epsilon}}{n^{1/k}}$$

then

$$\mathbb{P}[H_k(X, \mathbb{Z}) = 0] \to 1$$

as $n \to \infty$.

(2) *Also, if*

$$p \geq \frac{n^{\epsilon}}{n^{1/(2k+1)}}$$

then

$$\mathbb{P}[H_k(X, \mathbb{Z}) = 0] \to 1$$

as $n \to \infty$.

Note that this theorem actually holds for homology with integer coefficients.

The proof of (1) is essentially local and geometric, showing first that homology is supported on cycles of small support (bounded in size as $n \to \infty$), and then showing that every such cycle is a boundary.

The key observation for (2) is that link of a vertex in a random flag complex is a random flag complex with shifted parameter. Indeed, even intersecting the links of several vertex links results in another random flag complex. Then the Nerve Lemma allows one to bootstrap local information about connectivity of a large number of random graphs into global information about cohomology vanishing. This argument shows something stronger topologically: that if

$$p \geq \frac{n^{\epsilon}}{n^{1/(2k+1)}}$$

then w.h.p. X is k-connected, i.e. $\pi_i(X) = 0$ for $i \leq k$. Recent work of Babson shows that this exponent is tight when $k = 1$ [**6**].

In Section 4.4 we will see that the exponent in (2) can be improved, with a spectral gap argument, if one relaxes to cohomology with rational coefficients.

4.2. Nonvanishing homology and cohomology. It is also known that for every $k \geq 0$ there is a range of $p = p_k(n)$ for which $H_k(X(n, p)) \neq 0$ with high probability. Here are three ideas for how one might try to prove this.

(1) **Linear algebra.** Let f_i denote the number of i-dimensional faces of X. Then if $f_k > f_{k-1} + f_{k+1}$, we already have that $H_k \neq 0$ for dimensional reasons, i.e.

$$\beta_k \geq -f_{k-1} + f_k - f_{k+1}.$$

(2) **Homological argument: sphere.** Try to find a subcomplex Y homeomorphic to a sphere S^k, and a simplicial map $f : X \to Y$ such that $f \mid_Y = \mathrm{id}$. Then the homology of Y is naturally a summand of the homology of X, and in particular $H_k(X) \neq 0$.

(3) **Cohomological argument: isolated face.** If σ is a k-dimensional face not contained in any $(k+1)$-dimensional face, then the characteristic function of σ represents a cocycle. If one can somehow show that this function is not also a coboundary, then one has a nontrivial class.

All three of these approaches work, and in roughly the same range of param-eter. They also work equally well for with any choice of coefficients. Any of these approaches yields the following, for example.

THEOREM 4.2. *Let $\epsilon > 0$ be fixed and $X = X(n, p)$ If*

$$\frac{n^\epsilon}{n^{1/k}} \leq p \leq \frac{n^{-\epsilon}}{n^{1/(k+1)}}$$

then

$$\mathbb{P}[H_k = 0] \to 0$$

as $n \to \infty$.

The first two approaches are discussed in [**31**], and the third approach in [**33**]. In Section 4.4 we will see that the third approach has a slight edge on the other two approaches at the upper end of the nonvanishing window. In this case, a much sharper estimate may be obtained. So the exponent $1/k$ in Theorems 4.1 and 4.2 is sharp. The exponent $1/(k+1)$ in Theorem 4.2 is also sharp, as we will see in Section 4.4.

4.3. Limit theorems. Much more can be shown in the nonvanishing regime. Not only do we know that $H^k \neq 0$ w.h.p., but we can also understand the expec-tation of β^k and its limiting distribution.

The following asymptotic formula for the expectation follows from the linear algebra approach described above.

THEOREM 4.3. *Let $\epsilon > 0$ be fixed and $X = X(n, p)$ If*

$$\frac{n^\epsilon}{n^{1/k}} \leq p \leq \frac{n^{-\epsilon}}{n^{1/(k+1)}}$$

then

$$\frac{\mathbb{E}[\beta^k]}{\binom{n}{k+1}p^{\binom{k+1}{2}}} \to \mathcal{N}(0, 1)$$

as $n \to \infty$.

Here $\mathbb{E}[\beta^k]$ denotes the expectation of β^k. A similar formula can be given for the asymptotic variance $\text{Var}[\beta^k]$.

The following central limit theorem characterizes the limiting distribution [**34**].

THEOREM 4.4 (Meckes et al.). *Let $\epsilon > 0$ be fixed and $X = X(n, p)$ If*

$$\frac{n^\epsilon}{n^{1/k}} \leq p \leq \frac{n^{-\epsilon}}{n^{1/(k+1)}}$$

then

$$\frac{\beta^k - \mathbb{E}[\beta^k]}{\sqrt{\text{Var}[\beta^k]}} \to \mathcal{N}(0, 1)$$

as $n \to \infty$.

Here $\mathcal{N}(0, 1)$ is the standard normal distribution with mean 0 and variance 1, and the convergence is in distribution.

4.4. Sharp thresholds for rational cohomology. The following gives a sharp (upper) threshold for rational cohomology [**33**] of random flag complexes. It may be seen as a generalization of the Erdős-Renyi theorem, which corresponds to the $k = 0$ case.

THEOREM 4.5. *Let $k \geq 1$ and $\epsilon > 0$ be fixed, and $X = X(n, p)$.*

(1) *If*

$$p \geq \left(\frac{(k/2 + 1 + \epsilon) \log n}{n} \right)^{1/(k+1)},$$

then

$$\mathbb{P}[H^k(X, \mathbb{Q}) = 0] \to 1,$$

(2) *and if*

$$n^{-1/k+\epsilon} \leq p \leq \left(\frac{(k/2 + 1 - \epsilon) \log n}{n} \right)^{1/(k+1)},$$

then

$$\mathbb{P}[H^k(X, \mathbb{Q}) = 0] \to 0,$$

as $n \to \infty$.

The main tools used to prove this are Theorem 3.5 above which gives a concentration result for the spectral gap, together with the following Theorem 4.6.

THEOREM 4.6 (Garland, Ballman–Świątkowski). *Let Δ be a pure D-dimensional finite simplicial complex such that for every $(D - 2)$-dimensional face σ, the link $lk_\Delta(\sigma)$ is connected and has spectral gap is at least $\lambda_2[lk_\Delta(\sigma)] > 1 - 1/D$. Then $H^{D-1}(\Delta, \mathbb{Q}) = 0$.*

Theorem 4.6 is a special case of Theorem 2.5 of Ballman–Świątkowski [**8**], which in turn based on earlier work of Garland. For a deeper discussion of Garland's method, see A. Borel's account in Séminaire Bourbaki [**13**]. It is worth noting that Kazhdan had already proved other cases of Serre's conjecture in 1967 [**37**], and that this is the paper in which he introduced Property (T).

As a corollary, many random flag complexes have nontrivial rational homology only in middle degree.

COROLLARY 4.7. *Let $d \geq 1$ and $\epsilon > 0$ be fixed. If*

$$\frac{n^\epsilon}{n^{2/d}} \leq p \leq \frac{n^{-\epsilon}}{n^{2/(d+1)}},$$

then w.h.p. $X(n, p)$ is d-dimensional, and

$$\widetilde{H}_i(X(n, p), \mathbb{Q}) = 0 \ unless \ i = \lfloor d/2 \rfloor.$$

The same argument shows that if $p = O(n^{-\epsilon})$ for any fixed $\epsilon > 0$ then w.h.p. $X(n, p)$ has at most two nontrivial rational homology groups. This might be reminiscent of the concentration of chromatic number of a graph $\chi(G(n, p))$ on at most two values if say $p = O(n^{-1/2})$.

There is a slightly subtle point to be made here. In contrast to the earlier k-dimensional examples $Y_k(n, p)$, now the dimension d of the complex is a random variable. But by choosing p in the indicated regime, the complex is d-dimensional w.h.p.

Since the link of face in a random flag complex has the same distribution as another random flag complex (with shifted parameter), with a little more work one can show the following.

COROLLARY 4.8. *Let $d \geq 1$ and $\epsilon > 0$ be fixed. If*

$$\frac{n^\epsilon}{n^{2/d}} \leq p \leq \frac{n^{-\epsilon}}{n^{2/(d+1)}},$$

then w.h.p. $X(n, p)$ is d-dimensional, and the $\lfloor d/2 \rfloor$-skeleton of $X(n, p)$ is Cohen–Macaulay over $\mathbb{Q}$.

It is conceivable that Theorem 4.5 could be sharpened to the following.

CONJECTURE 4.9. *If*

$$p = \left(\frac{(k/2 + 1) \log n + (k/2) \log \log n + c}{n} \right)^{1/(k+1)},$$

where $c \in \mathbb{R}$ is constant, then the dimension of kth cohomology β^k approaches a Poisson distribution with mean

$$\mu = \frac{(k/2 + 1)^{k/2}}{(k + 1)!} e^{-c}.$$

In particular,

$$\mathbb{P}[H^k(X, \mathbb{Q}) = 0] \to \exp\left[-\frac{(k/2 + 1)^{k/2}}{(k + 1)!} e^{-c} \right],$$

as $n \to \infty$.

By earlier results, this conjecture is equivalent to showing that in this regime, cohomology is generated by characteristic functions on isolated k-faces.

4.5. Torsion. The question of torsion in random homology is still fairly mysterious, but for certain models torsion will be quite large. It may be surprising to learn, for example, that there exists a 2-dimensional $\mathbb{Q}$-acyclic simplicial complex S on 31 vertices with $H_1(S, \mathbb{Z})$ cyclic of order

$$|H_1(S, \mathbb{Z})| = 736712186612810774591.$$

The complex is relatively easy to define. The vertices are the elements of the cyclic group $\mathbb{Z}/31$, the 1-skeleton is a complete graph, and a set of three vertices $\{x, y, z\}$ span a 2-dimensional face if and only if

$$x + y + z \equiv 1, 2, \text{or } 9 \pmod{31}.$$

This type of "sum complex" was introduced and proved to be $\mathbb{Q}$-acyclic by Linial, Meshulam, and Rosenthal [40], and I found this example with a calculation in Sage [50].

Work of Kalai [36] showed that for $\mathbb{Q}$-acyclic complexes, the expected size of the torsion in homology is enormous. For example, for a random 2-dimensional $\mathbb{Q}$-acyclic complex S_n on n vertices,

$$\mathbb{E}\left[|H_1(S, \mathbb{Z})|\right] \geq e^{cn^2}$$

for some constant $c > 0$.

Compare this to Kowalski's note on torsion in homology of Dunfield–Thurston random 3-manifolds [**38**], where he shows that even though these manifolds M_n have $H_1(M_n, \mathbb{Q}) = 0$, with probability going to 1,

$$\mathbb{E}\left[|H_1(M_n, \mathbb{Z})|\right] \geq e^{\alpha n},$$

for some $\alpha > 0$.

For random 2-complexes or random flag complexes, my guess is that there may be a window when $H_i(X, \mathbf{k})$ vanishes for every fixed field $\mathbf{k}$ but such that $H_i(X, \mathbb{Z})$ is nonvanishing, but that this window is fairly small. Once you have a random finite abelian group of finite order, it should only take a few more random relations thrown in before the group becomes trivial. So in particular I would guess the following.

CONJECTURE 4.10. *Let $d \geq 6$ and $\epsilon > 0$ be fixed. If*

$$\frac{n^\epsilon}{n^{2/d}} \leq p \leq \frac{n^{-\epsilon}}{n^{2/(d+1)}},$$

then w.h.p. $X(n, p)$ is homotopy equivalent to a wedge of $\lfloor d/2 \rfloor$-spheres.

It is shown in [**31**] that for $p \geq n^{-1/3+\epsilon}$, $X(n, p)$ is simply connected w.h.p. So by uniqueness of Moore spaces, Conjecture 4.10 is equivalent to showing that for this range of p, $H_*(X(n, p))$ is torsion-free.

There may well be a window of nontrivial integer homology. Perhaps the prettiest thing we could hope for would be that for some model, Cohen–Lenstra heuristics hold. The Cohen–Lenstra heuristics propose roughly that the probability of a finite abelian p-group is inversely proportional to the size of its automorphism group. Could this hold for integer homology of a random $\mathbb{Q}$-acyclic complex, according to an appropriate determinantal measure, for example?

Lyons has carefully thought about these measures and described a way to generalize uniform spanning trees to higher dimensions, even on infinite cell complexes via ℓ^2-cohomology [**42**]. In closely related enumerative work, Duval, Klivans, and Martin extended Kalai's result, generalizing the Matrix Tree Theorem to higher dimensions [**18**].

5. Comments

Now that several different models of random simplicial complex have been studied, we start to see a few common themes emerging.

5.1. There are at least two different kinds of topological phase transitions. The "upper" phase transition where homology or cohomology passes from nonvanishing to vanishing seems easier to understand cohomologically. Examples of this kind of phase transition include the Erdős–Rényi theorem, the Linial–Meshulam theorem, and Theorem 4.5 above. These thresholds tend to be sharp, happening in a very narrow window.

The "lower" phase transition where homology or cohomology passes from vanishing to nonvanishing seems easier to understand homologically. Theorem 2.5 characterizes the first appearance of cycles in $G(n, p)$. The higher-dimensional analogue of this phase transition in random complexes is apparently much more subtle, but interesting recent work by Aronshtam, Linial, et al. studying this phase transition appears in [**4**] and [**5**]. These thresholds tend to be sharp on one side, not on the other.

The content of a certain number of theorems about random graphs is that non-sharp thresholds for monotone graph properties come from local properties, such as "contains a triangle", whereas sharp thresholds come from global properties, such as connected. Certifying that a graph contains a triangle only checking 3 relations of edge or non-edge. But certifying that a vertex is isolated involves checking $n-1$ other relations. Since isolated faces generate cohomology near the upper phase transition, by the same argument this is a global property and hence we should guess the upper threshold is sharp.

5.2. Homology and cohomology try to be as small as possible. With this motto we mean something more geometric, namely that near the lower threshold. homology tends to be supported on small classes. For example, homology of random geometric complexes in the sparse regime has a basis of vertex-minimal spheres [**32**]. This happens for both Čech and Vietoris–Rips complexes, even though the minimal spheres are combinatorially different in the two cases. This already accounts for the difference in the formulas for expectation of the Betti numbers in the sparse regime.

Near the upper threshold, cohomology is generated by small classes, namely characteristic functions on isolated k-faces.

The existence of these two phase transitions already implies something about what a theory of random persistent homology will have to look like. Perhaps a deeper understanding of these transitions will enable us to understand better how individual random homology classes are likely to begin and end.

This kind of paradigm also allows one to prove Poisson and normal approximation for Betti numbers. Since the homology is generated by small cycles, these cycles don't intersect very often, so they are "mostly independent". Then one can use the method of moments or Stein's method, for example, to prove something about limiting distributions [**34, 35**].

5.3. Nature abhors homology. (With apologies to Aristotle.) Homology is, after all said to measure the number of "holes" in a topological space, and holes are made out of vacuum. More seriously, it is often the case that unless there is a good reason random homology is forced to be there, then it is likely to vanish or be "small."

Suppose you have some measure on "pairs of random linear maps $f : A \to B$ and $g : B \to C$ satisfying $g \circ f = 0$". What can we say about the resulting distribution on H_B? Sometimes you can guess the answer from random linear algebra.

For example, one might guess that if

$$\dim A \ll \dim B \ll \dim C$$

then there is a good chance that the map $g : B \to C$ is injective, in which case $H_b = 0$. Or else perhaps $\ker g$ is merely small, but then this still bounds the size of homology. Similarly, if

$$\dim A \gg \dim B \gg \dim C,$$

then one might expect $f : A \to B$ is probably surjective (or nearly so) and so one expects that H_B is small. In fact only place the one place where we actually expect to see large homology, if dimension where the only consideration, would be if

$$\dim A \ll \dim B \gg \dim C.$$

So you might guess that $\dim H_B \approx \max\{0, -\dim A - + \dim B - \dim C\}$.

The random flag complex $X(n, p)$ is an example of where this kind of argument works surprisingly well.

5.4. Random simplicial complexes are expanders. One of the takeaway messages is that random simplicial complexes have expander-like properties. Higher-dimensional analogues of expanders have attracted a lot of attention — see for example the discussion in [16], and also Gromov's recent work on "geometric overlap" properties of expanders [26, 27]. As expander graphs have had so many applications in mathematics and theoretical computer science [30], one expects that higher-dimensional expanders will eventually find applications as well.

Expander graphs are well understood to be impossible to embed in Euclidean space with small metric distortion, thanks to work of Bourgain and many others. A recent "volume distortion" analogue of this was recently proved for random simplicial complexes by Dotterrer [15].

We would like to better understand the higher-dimensional analogues of the Cheeger–Buser inequalities relating spectral gap and the "bottleneck" expansion constant. See also Jerrum–Sinclair. For recent work on higher-dimensional analogues of Cheeger-Buser, see Gunder–Wagner [28], and Steenbergen–Klivans–Mukherjee [49].

5.5. A multi-parameter model. A model that deserves more attention is the multi-parameter random simplicial complex $\Delta(n; p_1, p_2, \dots)$. Here there are n vertices, the probability of an edge is $p_1 = p_1(n)$, and the complex is built inductively by dimension in so that the probability of every k-dimensional simplex, conditioned that its entire $(k-1)$-dimensional boundary is already in place, is $p_k = p_k(n)$, independently.

Several of the random simplicial complexes discussed here are special cases of this model, the random graph

$$G(n, p) = \Delta(n; p_1, 0, 0, \dots),$$

random 2-complex

$$Y(n, p) = \Delta(n; 1, p_2, 0, 0, \dots),$$

and random flag complex

$$X(n, p) = \Delta(n; p_1, 1, 1, \dots).$$

References

1. Robert J. Adler, Omer Bobrowski, Matthew S. Borman, Eliran Subag, and Shmuel Weinberger, *Persistent homology for random fields and complexes*, Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown, Inst. Math. Stat. Collect., vol. 6, Inst. Math. Statist., Beachwood, OH, 2010, pp. 124–143. MR 2798515 (2012j:60133)
2. Robert J. Adler and Jonathan E. Taylor, *Random fields and geometry*, Springer Monographs in Mathematics, Springer, New York, 2007. MR 2319516 (2008m:60090)
3. Noga Alon and Joel H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., Hoboken, NJ, 2008, With an appendix on the life and work of Paul Erdős. MR 2437651 (2009j:60004)
4. L. Aronshtam and N. Linial, *When does the top homology of a random simplicial complex vanish?*, arXiv:1203.3312, 2012.
5. L. Aronshtam, N. Linial, T. Luczak, and R. Meshulam, *Collapsibility and vanishing of top homology in random simplicial complexes*, arXiv:1010.1400, 2010.
6. Eric Babson, *Fundamental groups of random clique complexes*, arXiv:1207.5028, 2012.
7. Eric Babson, Christopher Hoffman, and Matthew Kahle, *The fundamental group of random 2-complexes*, J. Amer. Math. Soc. **24** (2011), no. 1, 1–28.

8. W. Ballmann and J. Świątkowski, *On L^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes*, Geom. Funct. Anal. **7** (1997), no. 4, 615–645. MR 1465598 (98m:20043)

9. Ya. M. Barzdin and A. N. Kolmogorov, *Selected works of A. N. Kolmogorov. Vol. III*, Mathematics and its Applications (Soviet Series), vol. 27, pp. 194–202, Kluwer Academic Publishers Group, Dordrecht, 1993, English translation. MR 1228446 (94c:01040)

10. Anders Björner, *A cell complex in number theory*, Adv. in Appl. Math. **46** (2011), no. 1-4, 71–85. MR 2794014 (2012f:05327)

11. Béla Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. MR MR1864966 (2002j:05132)

12. Béla Bollobás and Andrew Thomason, *Random graphs of small order*, Random graphs '83 (Poznań, 1983), North-Holland Math. Stud., vol. 118, North-Holland, Amsterdam, 1985, pp. 47–97. MR 860586 (87k:05137)

13. A. Borel, *Cohomologie de certains groupes discretes et lapiacien p-adique [d'après H. Garland]*, Séminaire Bourbaki vol. 1973/74 Exposés 436–452 (1975), 12–35.

14. Gunnar Carlsson, *Topology and data*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 2, 255–308. MR MR2476414

15. Dominic Dotterrer, *Higher dimensional distortion of random complexes*, arXiv:1210.6951, 2012.

16. Dominic Dotterrer and Matthew Kahle, *Coboundary expanders*, (2012), arXiv:1012.5316, to appear in Journal of Topology and Analysis.

17. Nathan M. Dunfield and William P. Thurston, *Finite covers of random 3-manifolds*, Invent. Math. **166** (2006), no. 3, 457–521. MR 2257389 (2007f:57039)

18. Art M. Duval, Caroline J. Klivans, and Jeremy L. Martin, *Simplicial matrix-tree theorems*, Trans. Amer. Math. Soc. **361** (2009), no. 11, 6073–6114. MR 2529925 (2011a:05385)

19. P. Erdős and A. Rényi, *On random graphs. I*, Publ. Math. Debrecen **6** (1959), 290–297. MR 0120167 (22 #10924)

20. Michael Farber and Thomas Kappeler, *Betti numbers of random manifolds*, Homology, Homotopy Appl. **10** (2008), no. 1, 205–222. MR 2386047 (2009d:55026)

21. Robin Forman, *A user's guide to discrete Morse theory*, Sém. Lothar. Combin. **48** (2002), Art. B48c, 35. MR 1939695 (2003j:57040)

22. Robert Ghrist, *Barcodes: the persistent topology of data*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 61–75. MR 2358377 (2008i:55007)

23. Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547. MR 2415379 (2009e:11181)

24. M. Gromov, *Spaces and questions*, Geom. Funct. Anal. (2000), no. Special Volume, Part I, 118–161, GAFA 2000 (Tel Aviv, 1999). MR 1826251 (2002e:53056)

25. _____, *Random walk in random groups*, Geom. Funct. Anal. **13** (2003), no. 1, 73–146. MR 1978492 (2004j:20088a)

26. _____, *Singularities, expanders and topology of maps. Part 1. Homology versus volume in the spaces of cycles*, Geom. Funct. Anal. **19** (2009), no. 3, 743–841. MR 2563769 (2012a:58062)

27. _____, *Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry*, Geometric And Functional Analysis **20** (2010), no. 2, 416–526.

28. A. Gundert and U. Wagner, *On laplacians of random complexes*, Proceedings of the 2012 symposuim on Computational Geometry, ACM, 2012, pp. 151–160.

29. Christopher Hoffman, Matthew Kahle, and Elliott Paquette, *A sharp threshold for Kazhdan's Property (T)*, arXiv:1201.0425, submitted, 2012.

30. Shlomo Hoory, Nathan Linial, and Avi Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561 (electronic). MR 2247919 (2007h:68055)

31. Matthew Kahle, *Topology of random clique complexes*, Discrete Math. **309** (2009), no. 6, 1658–1671. MR MR2510573

32. _____, *Random geometric complexes*, Discrete Comput. Geom. **45** (2011), no. 3, 553–573. MR 2770552 (2012a:52041)

33. _____, *Sharp vanishing thresholds for cohomology of random flag complexes*, submitted, arXiv:1207.0149, 2012.

34. Matthew Kahle and Elizabeth Meckes, *Limit theorems for Betti numbers of random simplicial complexes*, arXiv:1009.4130, to appear in Homology, Homotopy and Applications, 2012.

35. Matthew Kahle and Boris Pittel, *Inside the critical window for cohomology of random k-complexes*, submitted, arXiv:1301.1324, 2013.

36. G. Kalai, *Enumeration of Q-acyclic simplicial complexes*, Israel Journal of Mathematics **45** (1983), no. 4, 337–351.

37. D. A. Každan, *On the connection of the dual space of a group with the structure of its closed subgroups*, Funkcional. Anal. i Priložen. **1** (1967), 71–74. MR 0209390 (35 #288)

38. E. Kowalski, *On the complexity of Dunfield–Thurston random 3-manifolds*, available at http://www.ufr-mi.u-bordeaux.fr/kowalski/notes-unpublished.html.

39. D.N. Kozlov, *The threshold function for vanishing of the top homology group of random d-complexes*, Proc. Amer. Math. Soc **138** (2010), no. 12, 4517–4527.

40. N. Linial, R. Meshulam, and M. Rosenthal, *Sum complexes — a new family of hypertrees*, Discrete & Computational Geometry **44** (2010), no. 3, 622–636.

41. Nathan Linial and Roy Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica **26** (2006), no. 4, 475–487. MR MR2260850 (2007i:55004)

42. Russell Lyons, *Random complexes and l^2-Betti numbers*, J. Topol. Anal. **1** (2009), no. 2, 153–175. MR 2541759 (2010k:05130)

43. R. Meshulam and N. Wallach, *Homological connectivity of random k-dimensional complexes*, Random Structures Algorithms **34** (2009), no. 3, 408–417. MR 2504405 (2010g:60015)

44. Yann Ollivier, *A January 2005 invitation to random groups*, Ensaios Matemáticos [Mathematical Surveys], vol. 10, Sociedade Brasileira de Matemática, Rio de Janeiro, 2005. MR 2205306 (2007e:20088)

45. M. Pinsker, *On the complexity of a concentrator*, 7th annual teletraffic conference, 1973, pp. 1–4.

46. Nicholas Pippenger and Kristin Schleich, *Topological characteristics of random triangulated surfaces*, Random Structures Algorithms **28** (2006), no. 3, 247–288. MR MR2213112 (2007d:52019)

47. B. Pittel, *A random graph with a subcritical number of edges*, Trans. Amer. Math. Soc. **309** (1988), no. 1, 51–75. MR 957061 (89i:05247)

48. M. Raussen and C. Skau, *Interview with Michael Atiyah and Isadore Singer*, Notices of the AMS **52** (2005), no. 2.

49. John Steenbergen, Caroline Klivans, and Sayan Mukherjee, *A cheeger-type inequality on simplicial complexes*, arXiv:1209.5091, 2012.

50. W. A. Stein et al., *Sage Mathematics Software (Version 5.5)*, The Sage Development Team, 2012, http://www.sagemath.org.

51. A. Żuk, *Property (T) and Kazhdan constants for discrete groups*, Geom. Funct. Anal. **13** (2003), no. 3, 643–670. MR 1995802 (2004m:20079)

The Ohio State University, Department of Mathematics
E-mail address: mkahle@math.osu.edu