Matthew Kahle
Research statement

INTRODUCTION

I am broadly interested in interactions of probability and statistical physics with topology, geometry, and combinatorics. These interactions arise naturally in a variety of settings.

The field of topological statistics has been very active in recent years [9]. It is hoped that applying topological methodology to point cloud data should be helpful in understanding its qualitative structure. My recent article [18] gives detailed results about the expected topological properties of simplicial complexes built on Poisson point processes in Euclidean space, including central limit theorems and concentration of measure results for homology. This work helps establish a solid probabilistic foundation for the field of topological statistics.

Phase transitions are well-studied in probability, and some of my research is concerned with phase transitions for topological and geometric properties of various types of random spaces. A classical result in this area is the theorem of Erdős and Rényi characterizing the threshold for connectivity of the random graph [13]. Part of my Ph.D. thesis generalizes the Erdős-Rényi theorem to higher dimensions and to non-monotone settings [19].

Hard discs in a box is a well-studied model in statistical physics, where a phase transition from "liquid" to "solid" is observed empirically once the discs take up a certain fraction of the area of the box, but where little is known mathematically. Persi Diaconis pointed out that the set Config$(n; r)$ of all configurations of n nonoverlapping discs of radius r in the unit square can be viewed as a topological space as well as a probability space, and that little seems know about the topology of these spaces [12]. I recently found thresholds for ergodicity of the Metropolis Markov chain, and also gave examples to show that even simple topological properties like path-connectivity are not necessarily monotone in the radius of the discs [17].

A general motivation for this kind of work comes from the probabilistic method, which has provided existence proofs in combinatorics when constructive methods are difficult or impossible to come by [1]. Random finitely presented groups have also provided important examples, and in some cases conjectural counterexamples, in geometric group theory [15, 23]. Random cell complexes will likely provide interesting and extremal examples of topological spaces [20].

1. RANDOM SIMPLICIAL COMPLEXES

In several articles I have studied topological properties of random simplicial complexes, developing several different techniques for doing so. There are various motivations for this, discussed in the context of each article.

1.1. **Random geometric complexes.** In some of my recent work, I studied simplicial complexes built on random geometric graphs [18], a problem suggested by Persi Diaconis [11]. Random geometric graphs have vertices on n points sampled i.i.d. in $\mathbb{R}^d$ from a probability distribution with a bounded measurable density function. Then edges connect pairs of points if they are within distance r. Usually $r = r(n)$ and one considers properties asymptotically as $n \to \infty$. In this article I studied Vietoris-Rips and Čech complexes built on these random geometric graphs.

This work generalizes some of Mathew Penrose's earlier work on component counts in random geometric graphs [24] to higher dimensions and introduces new ideas to geometric probability. The main results are as follows.

(1) In the sparse range, $r = o(n^{-1/d})$, I compute the expectations and variances of the Betti numbers, establish Central Limit Theorems, and use martingales to prove a concentration of measure result. The central limit theorem for Betti numbers states that

$$\frac{\beta_k - E[\beta_k]}{\text{Var}[\beta_k]} \to \mathcal{N}(0,1)$$

in distribution as $n \to \infty$, where $\mathcal{N}(0,1)$ is the standard normal distribution.
The concentration of measure result gives that

$$\Pr(\beta_k - E[\beta_k]) \geq a) \leq 2\exp(-ca^2/n),$$

where c is a constant depending on k and the underlying density.

(2) In the critical range, $r = \theta(n^{-1/d})$, and here I showed that $E[\beta_k]$ and $\text{Var}[\beta_k]$ grow linearly with n.

(3) In the dense range, $r = \omega(n^{-1/d})$, I introduce Morse-theoretic arguments to bound the expectation of the Betti numbers. From these estimates it follows that the expectations of the Betti numbers grow sub-linearly in the dense range, and that the threshold for contractibility of the Čech and Rips complexes is approximately the same as for connectivity of the random geometric graph.

This main motivation for this work is that it helps provide a probabilistic "null model" for topological statistics. It is also perhaps interesting as a flexible model of random simplicial complex; indeed, one may choose the underlying density function on $\mathbb{R}^d$, so the parameter space is essentially infinite-dimensional.

1.2. **The fundamental group of random 2-complexes.** Nati Linial and Roy Meshulam [21] defined the random 2-complex $Y(n,p)$ to be the complete graph with each of the $\binom{n}{3}$ possible triangular 2-cells included with probability p independently. They found the sharp threshold for vanishing of the cohomology $H^1(Y(n,p),\mathbb{Z}/2)$ to be $p = 2\log n/n$ [21], a cohomological analogue of the Erdős-Rényi Theorem characterizing the threshold for connectivity of the random graph [13].

In joint work with Eric Babson and Chris Hoffman [3], we found the threshold for vanishing of $\pi_1(Y(n,p))$ to be quite different, approximately $p = n^{-1/2}$. The argument is somewhat delicate, combining methods of geometric group theory and combinatorial homotopy theory. The main idea was to show that in the nonvanishing range, π_1 is a hyperbolic group [14], and this required the use of Gromov's local-to-global method, that one only need check hyperbolicity on balls of bounded radius [23].

This work seems to be interesting for bringing together several different areas of math and also for providing a new model of random groups in geometric group theory.

1.3. **The random clique complex.** The random clique complex $X(n,p)$ is the clique complex of the random graph $G(n,p)$, or the Erdős-Rényi graph with all possible simplicial faces of dimension ≥ 2 added. Clique complexes are also sometimes called *flag complexes* in the literature. As part of my thesis, I studied the topology of $X(n,p)$ when $p = O(n^{-\epsilon})$ [19], where these complexes are a.a.s. d-dimensional for some constant d depending on ϵ. The main results can be summed up as follows.

(1) A d-dimensional random clique complex a.a.s. has nontrivial homology in dimension $k = \lfloor d/2 \rfloor$. I computed the expected dimension of homology $E[\beta_k]$ and the variance. More recently I proved a central limit theorem for β_k when $k = \lfloor d/2 \rfloor$ (preprint in preparation).

(2) Homology is a.a.s. vanishing below dimension $\lfloor d/4 \rfloor$ and above dimension $\lfloor d/2 \rfloor$. This gives a d-dimensional generalization of the Erdős-Rényi theorem, but in a non-monotone setting. The results suggest strongly that $X(n,p)$ a.a.s. has only one nontrivial reduced homology group $H_k(X(n,p),\mathbb{Z}/2)$, at $k = \lfloor d/2 \rfloor$. This would be

quite interesting, for a variety of reasons. In particular it would imply that a kind of Poincare duality holds for almost all d-dimensional flag complexes.

(3) I used discrete Morse theory to show that $E[\beta_k]$ is small when $k \neq \lfloor d/2 \rfloor$, supporting the conjecture that there is only one nontrivial homology group. This method was further refined in the article on random geometric complexes, where Morse theory is used to show that certain homology groups vanish altogether.

A common type of theorem in topological combinatorics is that a certain poset or simplicial complex is homotopy equivalent to a wedge of spheres of the same dimension [5]. My work on random clique complexes suggests that this kind of behavior may actually be generic in a measure theoretic sense. Figure 1 illustrates several of the theorems from the paper, and seemingly supports the conjecture that almost all d-dimensional flag complexes have only one nontrivial homology group.

2. PROBABILITY AND GRAPH MAPS

Graph homomorphisms have been studied in recent years through interactions with both algebraic topology [22, 4] and statistical physics [8]. In both settings, lower bounds on chromatic number $\chi(H)$ (i.e. obstructions of maps from H to the right) are found by probing H with graph homomorphisms from the left. Some of my work tests the strength of each of these types of bounds for various kinds of random graphs. This helps illustrate qualitatively which types of graphs will be most amenable to these kinds of bounds, and also leads to new quantitative conjectures relating the topological and statistical physics points of view.

2.1. The neighborhood complex of a random graph.

In 1978, Lovász used algebraic topology to put lower bounds on the chromatic number of graphs [22]. In particular he showed that the connectivity of the neighborhood complex is a lower bound on chromatic number, conn$[\mathcal{N}(G)] \leq \chi(G)$. Lovász's bound was tight for Kneser graphs, settling a long standing conjecture. In the article [16], I computed the connectivity of the neighborhood complex (within a constant factor) of an Erdős-Rényi graph $G(n,p)$, and compared it with known values for expected chromatic number. In both sparse and dense regimes, conn$[\mathcal{N}(G)]$ is much smaller than $\chi(G)$ for a random graph G.

To prove this meant understanding of which homology groups of the neighborhood complex are vanishing and non-vanishing. For example, the expected dimension of the random neighborhood complex $\mathcal{N}[G(n,1/2)]$ is

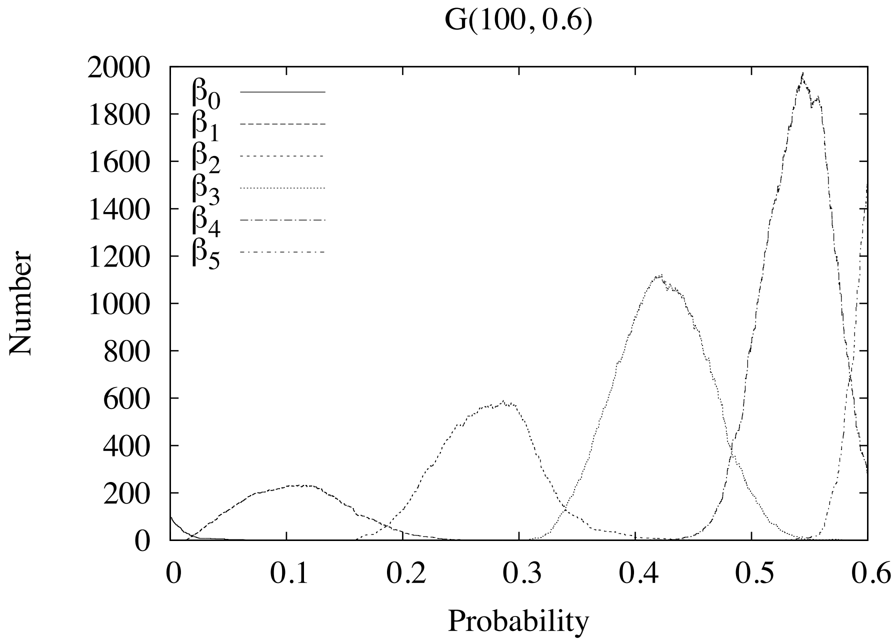

FIGURE 1. A random clique complex on $n = 100$ vertices and with 0–3000 edges. The number of random edges is the horizontal axis, and the Betti numbers β_0–β_5 are graphed as six different curves superimposed. Image courtesy of Afra Zomorodian.

approximately $n/2$, and I showed that homology $H_k(\mathcal{N}, \mathbb{Z})$ a.a.s. vanishes for $k < \log_2 n$ and $k > 4\log_2 n$, and is non-vanishing when $(4/3)\log_2 n < k < 2\log_2 n$.

The results also suggest which types of graphs are likely to be most amenable to topological lower bounds on chromatic numbers. In particular, Erdős-Rényi graphs do not have enough complete bipartite subgraphs to keep the neighborhood complex from collapsing to a complex of small dimension, so we might consider graphs with many bipartite subgraphs or large odd girth. This is the part of the motivation for studying random graphs on spheres, discussed in the next subsection.

2.2. **The chromatic number of random graphs on spheres.** (Preprint in preparation.) We define a random graph geometrically on the d-dimensional unit sphere S^d by taking n points uniformly and independently for vertices, and connecting two vertices by an edge if the distance between them is at least $2 - \epsilon$ where $\epsilon = \epsilon(n)$ tends to zero as $n \to \infty$. If ϵ goes to zero slowly enough, i.e. if $\epsilon = \Omega(\log n/n)$, then a Morse-theoretic argument gives that the neighborhood complex is a.a.s. $(d-1)$-connected.

One can check that these graphs are $(d+2)$-colorable, so by Lovász's bound, the chromatic number of these graphs a.a.s. $= d+2$. This gives new infinite families of graphs for which topological bounds on chromatic number are tight.

2.3. **Warmth and mobility of random graphs.**

(Preprint in preparation.) In combinatorial statistical physics a graph G can be thought of as a space, and another graph H as physical laws or constraints. Then a graph homomorphism $\sigma : G \to H$ can be interpreted as one possible state of the space. Motivated by this point of view, Brightwell and Winkler introduced new graph parameters [8]. The *warmth* $w(G)$ is slightly technical to define, but it essentially measures the amount of freedom in mapping regular trees into G. One of Brightwell and Winkler's results is to show that $w(G) \leq \chi(G)$ for every graph G.

In joint work with Sukhada Fadnavis, a graduate student at Stanford, we investigated warmth for Erdős-Rényi random graphs $G(n,p)$, and found that warmth, although not a monotone graph property is nevertheless statistically monotone with p and for most values of $p = O(n^{-\epsilon})$ is concentrated on a single value. This allows us to establish a conjecture of Lovász that $m(G) \leq \chi(G)$, for almost all graphs.

These results also lead to a new conjecture relating the statistical physics and topological parameters in graph homomorphisms. In particular, is it true that $w(G) \leq \text{conn}[\mathcal{N}(G)]$, where $\text{conn}[\mathcal{N}(G)]$ denotes the connectivity of the neighborhood complex?

3. CONFIGURATION SPACES OF HARD DISCS

Hard discs in a box is a well-studied model for matter in statistical physics. Phase transitions have been observed empirically through computer simulations, but so far there does not seem to be a good mathematical explanation of this. Persi Diaconis recently pointed out [12] that Config$(n; r)$ = the configuration space of all possible positions of n nonoverlapping discs of radius r in a unit square is naturally endowed with a topology as well as probability measure, and that "very, very little is known about the topology of these spaces." Besides their interest from statistical physics, these spaces provide a generalization of classical configuration spaces, interesting to understand topologically for their own sake.

3.1. **Thresholds for ergodicity and classical homotopy type.**

A well studied Markov chain on these configuration spaces is the Metropolis algorithm. Using a disc packing construction, I showed that this Markov chain is not ergodic in the range of parameters commonly studied, as is apparently commonly believed [17]. I showed that if $r = \Omega(1/n)$ then the configuration space has many Markov disconnected states. This result is also best possible, up to a constant: a result of Diaconis, Mechel, and Lebeau gives that if $r = O(1/n)$, the Metropolis Markov chain is ergodic.

More recently, in joint work with Yuliy Barishnikov and Peter Bubenik, we showed that for some $r = O(1/n)$, not only is the configuration space path connected, but that it is homotopy equivalent to the classical configuration space of n labeled points in the plane, well studied in algebraic topology. In ongoing work, we have upper and lower bounds on the topological complexity of Config$(n; r)$ as measured by the Lusternik–Schnirelmann category.

3.2. **A computational approach.** (Preprint in preparation.) With Gunnar Carlsson and Jackson Gorham, an undergraduate researcher at Stanford, we studied these configuration spaces computationally. In particular, combining a Morse theoretic approach with more traditional Monte Carlo methods and the "nudged elastic band" method to find low energy paths, we were able to recover the complete dendrogram for persistent path components of Config$(n; r)$ (with respect to r) when when the number n of discs is 5. Even here the story is complicated, and we were surprised to find out that the number of path components is not monotone in the radius, nor is path-connectedness a monotone property.

This work contributes to computational topology by giving an example of how Morse theory can in principle handle hierarchical clustering, even in a non-monotone setting where agglomerative and divisive clustering methods can not tell the whole story. Non-monotonicity of the number of path components arises quite naturally in this setting.

4. FUTURE DIRECTIONS

The following are a few more open problems and continuing projects.

(1) Persistent homology on random point processes. Although the work on random geometric complexes [18] gives detailed information about the homology of complexes on random point processes, for applications to topological data analysis one would like information about persistent homology. In ongoing work with Dmitriy Morozov, we have bounds on the expected persistent homology of random geometric complexes by using stability of persistent homology, together with geometric probability [10].

(2) Giant components and structure theorems in random homology. In the Erdős-Rényi $G(n, p)$ setting, much more is know than the threshold for connectivity and the distribution on the number of components. In particular, percolation on the random graph is extremely well studied in the range near $p = 1/n$ [6], and there are many things known about the size of the components and the structure of the giant and other components as it emerges. There should be higher dimensional analogues of these statements as well, using normed cohomology to measure the size of d-dimensional homology classes, as in earlier work of Babson and Benjamani [2].

Bollobás and Riordan recently posted a preprint on clique percolation where they exhibit higher dimensional analogues of percolation for the random clique complex described earlier [7]. Analogous statements are likely to be true for other kinds of random complexes, and one might be able to frame a fairly general theory of higher-dimensional percolation in topological terms.

(3) Torsion in random homology. After several papers on various kinds of random simplicial complexes have been written, none of them have addressed the torsion in $\mathbb{Z}$-homology when it is non-vanishing. It would seem that more subtle techniques are needed. The question of the threshold for $\mathbb{Z}$-homology in the Linial-Meshulam 2-complexes is of particular interest for geometric group theoretic reasons.

(4) Statistical physics, topology, and graph maps. The papers on probability and graph maps have started to shed light on the possible connections between the statistical physics and topological points of view on graph homomorphisms, and have led to new conjectures. In ongoing work, I am investigating whether the warmth of G (in the sense of Brightwell–Winkler [8]) and the connectivity of the neighborhood complex (as defined by Lovász [22]) are related by the inequality

$$w(G) \leq \text{conn}[\mathcal{N}(G)].$$

(5) Combinatorial rigidity and long-range action. One can consider a point chosen uniformly in $\text{Config}(n; r)$ together with a connectivity threshold radius $r' > r$ to be a kind of random geometric graph. The fact that the underlying point process is not Poisson makes properties such as subgraph counts much more subtle to analyze than for

Poisson random geometric graphs. An ongoing project is to understand very basic local features of these graphs, such as expectation of vertex degree. Even a coarse understanding of the combinatorial structure of these graphs, combined with the right rigidity theory, might give a heuristic explanation for long-range action.

REFERENCES

[1] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[2] Eric Babson and Itai Benjamini. Cut sets and normed cohomology with applications to percolation. *Proc. Amer. Math. Soc.*, 127(2):589–597, 1999.

[3] Eric Babson, Chris Hoffman, and Matthew Kahle. The fundamental group of random 2-complexes. Preprint, arXiv:0711.2704 (submitted), 2008.

[4] Eric Babson and Dmitry N. Kozlov. Proof of the Lovász conjecture. *Ann. of Math. (2)*, 165(3):965–1007, 2007.

[5] A. Björner. Topological methods. In *Handbook of combinatorics, Vol. 1, 2*, pages 1819–1872. Elsevier, Amsterdam, 1995.

[6] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[7] Bela Bollobás and Oliver Riodan. Clique percolation. Preprint, arXiv:0804.0867, 2008.

[8] Graham R. Brightwell and Peter Winkler. Graph homomorphisms and long range action. In *Graphs, morphisms and statistical physics*, volume 63 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 29–47. Amer. Math. Soc., Providence, RI, 2004.

[9] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[10] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

[11] Persi Diaconis. Application of topology. Quicktime video, September 2006. (Talk given at MSRI Workshop on "Application of topology in science and engineering," Quicktime video available on MSRI webpage).

[12] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009.

[13] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[14] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.

[15] M. Gromov. Random walk in random groups. *Geom. Funct. Anal.*, 13(1):73–146, 2003.

[16] Matthew Kahle. The neighborhood complex of a random graph. *J. Combin. Theory Ser. A*, 114(2):380–387, 2007.

[17] Matthew Kahle. An application of disc packing to statistical mechanics. Preprint, arXiv:0908.1830 (submitted), 2009.

[18] Matthew Kahle. Random geometric complexes. Preprint, arXiv:0910.1649 (submitted), 2009.

[19] Matthew Kahle. Topology of random clique complexes. *Discrete Math.*, 309(6):1658–1671, 2009.

[20] Nathan Linial. Going up in dimension: Probabilistic and combinatorial aspects of simplicial complexes, August 2009. (Talk given at Random Structures and Algorithms, .pdf file available online).

[21] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

[22] László Lovász. Kneser's conjecture, chromatic number, and homotopy. *J. Combin. Theory Ser. A*, 25(3):319–324, 1978.

[23] Yann Ollivier. *A January 2005 invitation to random groups*, volume 10 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2005.

[24] Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.