Topological data analysis of continuum percolation with disks

Leo Speidel

Department of Statistics, University of Oxford, Oxford, UK and

Doctoral Training center in Systems Biology, University of Oxford, Oxford, UK

Heather A. Harrington and S. Jonathan Chapman

Mathematical Institute, University of Oxford, Oxford, UK

Mason A. Porter

Department of Mathematics, University of California Los Angeles, Los Angeles, CA, USA

(Dated: June 7, 2018)

Abstract

We study continuum percolation with disks, a variant of continuum percolation in two-dimensional Euclidean space, by applying tools from topological data analysis. We interpret each realization of continuum percolation with disks as a topological subspace of $[0, 1]^2$ and investigate its topological features across many realizations. We apply persistent homology to investigate topological changes as we vary the number and radius of disks. We observe evidence that the longest persisting invariant is born at or near the percolation transition.

I. INTRODUCTION

Percolation, and its many variants, is one of the most popular topics in statistical physics. It has been used in the study of a diverse variety of phenomena, ranging from biological and social contagions to connectivity of infrastructure [1–3]. Continuum percolation models can help provide a mechanistic understanding for the formation of structures embedded in space and are closely related to spatially-embedded random graphs [4]. Understanding the properties of spatially-embedded random graphs has long been of theoretical interest [4–8], and it also has numerous applications, such as for investigating properties of granular packings [9, 10] and sensor networks [11–13].

Studying spatial percolation processes can be helpful for investigations of spatially embedded systems, which is particularly important in light of the increased availability of rich, finely-grained data about such systems. For example, advances in imaging technology have led to increasingly accurate measurements of biophysical networks [14] (such as vascular networks, leaf-venation networks, bronchial trees in the lung, and neuronal networks), which — like many other networks and complex systems — are embedded in space [15]. Their properties and mechanisms of the formation of biophysical networks have been investigated as optimization problems [16], and they have also been studied using fluid dynamics [17] and in relation to random processes such as random sequential adsorption (RSA) [18]. Some of these situations, such as flocks of *Volvox barberi* [19], also form fascinating packing structures.

In many continuum percolation models (and in percolation problems more generally [1, 2]), there are known phase transitions (identified theoretically or computationally) as one varies the model parameters [4, 5]. For instance, a large percolating cluster may suddenly emerge as smaller clusters connect to each other. Percolation transitions are interesting, because they indicate a change in long-range spatial correlations. To detect such structural changes, one commonly studies changes in the sizes of connected clusters.

In the present paper, we take a different approach, as we focus instead on the *shape* of clusters. We do computations to investigate changes in the topological[1] properties of clusters [20–22]. The extent to which varying model parameters affects topological properties

[1] We use the word "topology" in its standard mathematical sense, rather than in the sense of examining connectivity structure in networks ("graph topology" or "network topology") that is often used to describe the structure of graphs and networks.

has been examined in some classical models of statistical physics, such as in models of interacting spins [23–25], and we do this in our focal problem as well. Specifically, we study one-dimensional topological invariants, which one can interpret as cycles or holes in the structure of a point cloud of data (including data from a network). We examine what we call a "continuum percolation with disks", a variant of continuum percolation, in which we place N disks of radius r independently and uniformly at random in two-dimensional (2D) Euclidian space. This percolation model is related to random geometric graphs (RGGs) [4, 6, 12, 26].

To identify topological invariants of continuum percolation with disks, we calculate its homology groups H_0 and H_1. The dimension of H_0 equals the number of connected components, and we note that the percolation and connectivity transitions are phase transitions related to H_0. The percolation transition occurs in the regime $0 < \lim_{N\to\infty} Nr^2 < \infty$, and the connectivity transition occurs in the regime $\lim_{N\to\infty} Nr^2 = \infty$ [4]. The latter regime is thus in the supercritical regime of the percolation transition. One can interpret the dimension of H_1 as the number of cycles, in the form of one-dimensional holes, in a structure. Previous studies have identified two topological phase transitions that are related to the first homology group H_1. They each correspond to a transition between zero and nonzero dimension of H_1 [7, 8]. The first phase transition occurs in the subcritical regime of the percolation transition (in which $\lim_{N\to\infty} Nr^2 = 0$), and the second phase transition occurs in the supercritical regime of the percolation transition.

It is partially understood how the first homology group changes in the critical regime of percolation transition (see Section IV A), but it is not well-understood how it changes at and near the percolation transition. We apply persistent homology (PH) [20, 22] and study topological changes close to the percolation transition. Based on our calculations, the longest persisting topological invariant appears to be born at the percolation transition. This suggests that there is a change in shape, in addition to a change in size, of clusters at the percolation transition.

The rest of our paper is organized as follows. First, we define continuum percolation more precisely and then define *continuum percolation with disks*. To help characterize continuum percolation with disks, we then discuss the sizes, number, and shapes of its clusters. We investigate continuum percolation with disks by computing PH and then conclude with a brief summary and discussion.

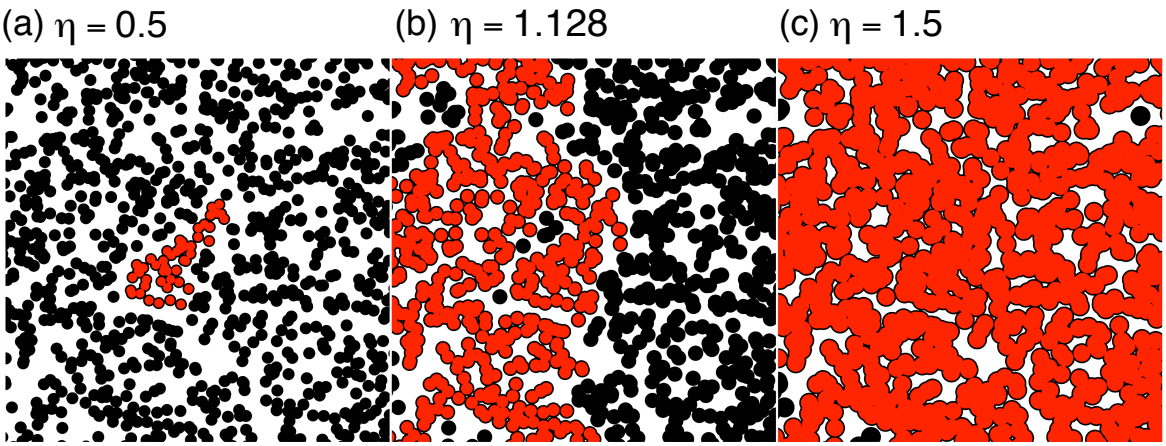


FIG. 1. One realization of continuum percolation with disks with filling factors of (a) $\eta(N,r) = 0.5$, (b) $\eta(N,r) = 1.128$, and (c) $\eta(N,r) = 1.5$, where $\eta(N,r) = N\pi r^2$ and there are $N = 1000$ disks of radius r. We show the largest connected cluster in red. It has been estimated that a percolation transition occurs at $\eta_c \approx 1.128$ [27].

II. CONTINUUM PERCOLATION

Consider a metric space (X, d). *Continuum-percolation* models are random processes in which subsets $D_i \subseteq X$ (where $i \in \{1, \ldots, N\}$) are chosen randomly with some probability distribution, resulting in a structure $\bigcup_{i=1}^{N} D_i$.

We consider the unit square $X = [0,1]^2$ equipped with the Euclidean metric. We sample points $\boldsymbol{x}_i \in X$ (with $i \in \{1, \ldots, N\}$) independently and uniformly at random and obtain disks $D_i(r) = \{\boldsymbol{y} \in X : d(\boldsymbol{y}, \boldsymbol{x}_i) < r\}$ of radius $r \geq 0$ that are centered at these points. The radius r is the same for all disks. The resulting structure $\mathcal{D}(r) = \bigcup_{i=1}^{N} D_i(r)$ is a union of potentially overlapping disks (see Fig. 1). We refer to this model as *continuum percolation with disks*.

A related model is a random geometric graph (RGG) [12, 28]. One can construct an RGG from a continuum-percolation model by defining the node set $V = \{1, \ldots, N\}$ and edge set $E = \{(i,j) : D_i \cap D_j \neq \emptyset\}$. One can induce an RGG from continuum percolation by placing nodes independently and uniformly at random in 2D Euclidean space and connecting pairs of nodes that lie within a specified distance $2r$ [6, 26].

Both continuum percolation and RGGs have been studied extensively over many years [5], with focus on identifying and characterizing phase transitions [6, 29] and relating continuum

percolation to percolation on lattices [30]. They have both been studied in a variety of metric spaces [4, 15, 26, 31]. One can also choose different probability distributions for placing points; consider higher-dimensional spaces; and consider "softer" potentials (in contrast to the Heaviside function above), in which nodes are connected to each other with a probability that decays with increasing distance between them [12].

III. SIZES AND NUMBER OF CLUSTERS

One can partition any realization of the structure $\mathcal{D}(r)$ into disjoint clusters. The distributions of the number of clusters and the number of disks in the largest cluster depend on the combination of the parameters N and r.

It is useful to study $\eta(N,r) = N\pi r^2$, which is called the *filling factor*. The size of the largest cluster undergoes a phase transition as one varies $\eta(N,r)$. For $\lim_{N\to\infty}\eta(N,r) < \eta_c$, the size of the largest cluster is, with probability 1, at most $O(\log N)$ of the number N of disks as $N \to \infty$ (see Theorem 10.3 in Ref. [4]). When $\lim_{N\to\infty}\eta(N,r)$ approaches the critical percolation transition value (i.e., the percolation "threshold") η_c from below, clusters merge until, at $\lim_{N\to\infty}\eta(N,r) = \eta_c$, there emerges a unique giant cluster (i.e., giant component) with size $O(N)$ (see Theorem 10.9 in Ref. [4]). Percolation transitions have been studied for many models and in numerous situations, including for lattice models and networks, continuum models with various geometric objects, and in different dimensions [2, 27, 32–34]. For continuum percolation with disks, the exact value η_c of the percolation threshold is unknown. It has been estimated numerically to be $\eta_c \approx 1.128$ [27].

In the supercritical regime $\lim_{N\to\infty}\eta(N,r) > \eta_c$, clusters keep merging as $\eta(N,r)$ increases. In a second ("connectivity") phase transition, which occurs in the limit $\lim_{N\to\infty}\eta(N,r) = \infty$, all clusters merge into a single cluster with probability 1. The critical value for this connectivity transition satisfies [4]

$$\lim_{N\to\infty} \frac{\eta_c^{\text{connect}}(N,r)}{\log N} = \frac{1}{4}\,. \tag{1}$$

IV. SHAPE OF CLUSTERS

We are interested in understanding the change in shape of connected clusters as we vary $\eta(N,r)$. To investigate this, we interpret $\mathcal{D}(r)$ as a topological subspace of (X,d) and study

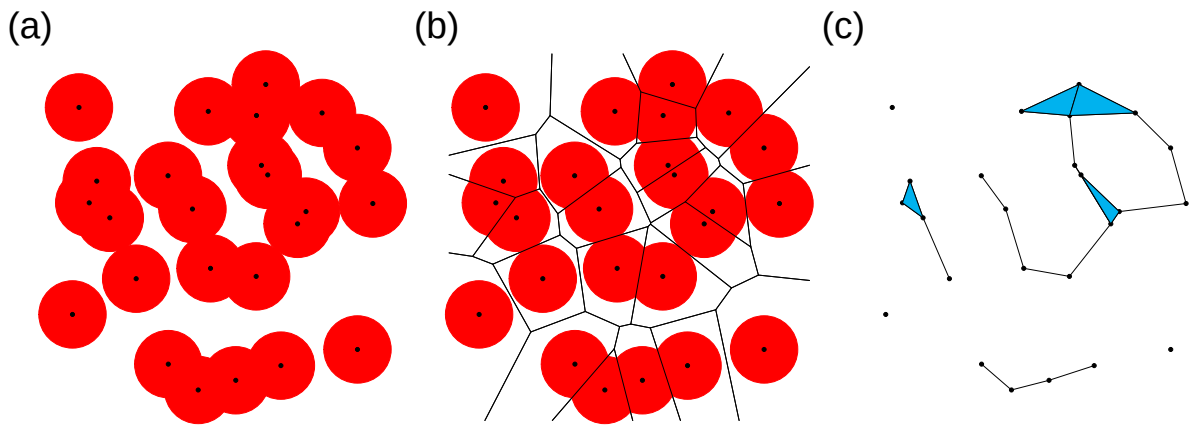


FIG. 2. (a) A realization of continuum percolation with disks with $N = 25$ and $r = 0.22$. (b) Voronoi tessellation constructed using the center of the disks in panel (a). (c) Alpha complex constructed from panel (b). We represent 0-simplices by nodes, 1-simplices by edges (which connect adjacent nodes to each other), and 2-simplicies by filled triangles. In this example, there are six clusters (so $\beta_0 = 6$) and one hole (so $\beta_1 = 1$).

its topological properties. Specifically, we study its *homology groups*, which correspond, intuitively, to the numbers of holes in different dimensions [35].

We can calculate the homology groups of $\mathcal{D}(r)$ by studying a different topological space with homology groups that are isomorphic to those of $\mathcal{D}(r)$. The topological space that we study is the *alpha complex* [36], which is a type of *simplicial complex*. (See e.g., Refs. [22, 35] for the definition of a simplicial complex.) To construct an alpha complex, we calculate the Voronoi tesselation from the centers of the disks. We define the intersection of a disk and its corresponding Voronoi cell by $V_i(r)$ (where $i \in \{1, \dots, N\}$). The alpha complex includes 0-simplices $\{V_i(r)\}$ (where $i \in \{1, \dots, N\}$) and k-simplices (with $k = 1$ and $k = 2$), given by $\{V_{i_0}(r), \dots, V_{i_k}(r)\}$, such that $V_{i_0}(r) \cap \dots \cap V_{i_k}(r) \neq \emptyset$. In our setting, the Nerve Theorem implies that $\mathcal{C}(r)$ and $\mathcal{D}(r)$ have the same homology (i.e., their homology groups are isomorphic) [20, 37]. In Fig. 4, we visualize the alpha complex for a small example of continuum percolation with disks.

The kth homology group $H_k(\mathcal{C}(r))$ describes the k-dimensional holes [22, 35]. Its dimension is the kth Betti number β_k. The 0th Betti number β_0 is equal to the number of clusters. As we discussed in Section III, β_0 has been studied extensively in percolation (though without explicitly invoking ideas from homology). Intuitively, the first Betti number β_1 counts

the number of holes. In this paper, we focus on β_1, which gives information about the "shapes" of the clusters.

A. Known phase transitions in the first Betti number

Theoretical work relating Betti numbers to continuum percolation indicates that the shapes of clusters change as one varies $\eta(N, r)$ [7, 8]. For instance, it has been shown, in the limit $\lim_{N\to\infty} \eta(N, r) = 0$, that there is a phase transition at which nontrivial 1D homology first appears. The critical value satisfies

$$\lim_{N\to\infty} \sqrt{N}\eta_{\mathrm{c}}^{\mathrm{exist}}(N, r) = c_{\mathrm{exist}}\,. \tag{2}$$

Below the critical value, $\beta_1 = 0$; above the critical value, $\beta_1 > 0$ with probability 1 (see, e.g., Theorem 4.12 in Ref. [7]).

Similar to the connectivity transition in Eq. (1), the first Betti number vanishes in the limit $\lim_{N\to\infty} \eta(N, r) = \infty$. The critical value for this phase transition satisfies (see, e.g., Theorem 4.12 in Ref. [7])

$$\lim_{N\to\infty} \frac{\eta_{\mathrm{c}}^{\mathrm{vanish}}(N, r)}{\log N} = c_{\mathrm{vanish}}\,. \tag{3}$$

In the critical regime, in which $\eta(N, r)$ approaches a finite, nonzero limit as $N \to \infty$, it is known that β_1 scales linearly in the number of disks:

$$\lim_{N\to\infty} \frac{E[\beta_1]}{N} = f(\eta)\,, \tag{4}$$

for some function $f(\eta) \in (0, \infty)$. (See, e.g., Theorem 4.6 in Ref. [7].) Other than Eq. (4), it is not understood what happens at or near the percolation transition. In Section V, we confirm Eq. (4) (see Fig. 4) and investigate the behavior of $f(\eta)$ near the percolation transition.

V. PERSISTENT HOMOLOGY AND RESULTS

We aim to apply persistent homology (PH) to characterize structural changes over a wide range of values of $\eta(N, r)$, especially near the percolation transition in the critical regime. PH is a method for topological data analysis (TDA) of point clouds that extracts qualitative features over multiple scales [20, 22, 38], which we examine by varying the filling factor

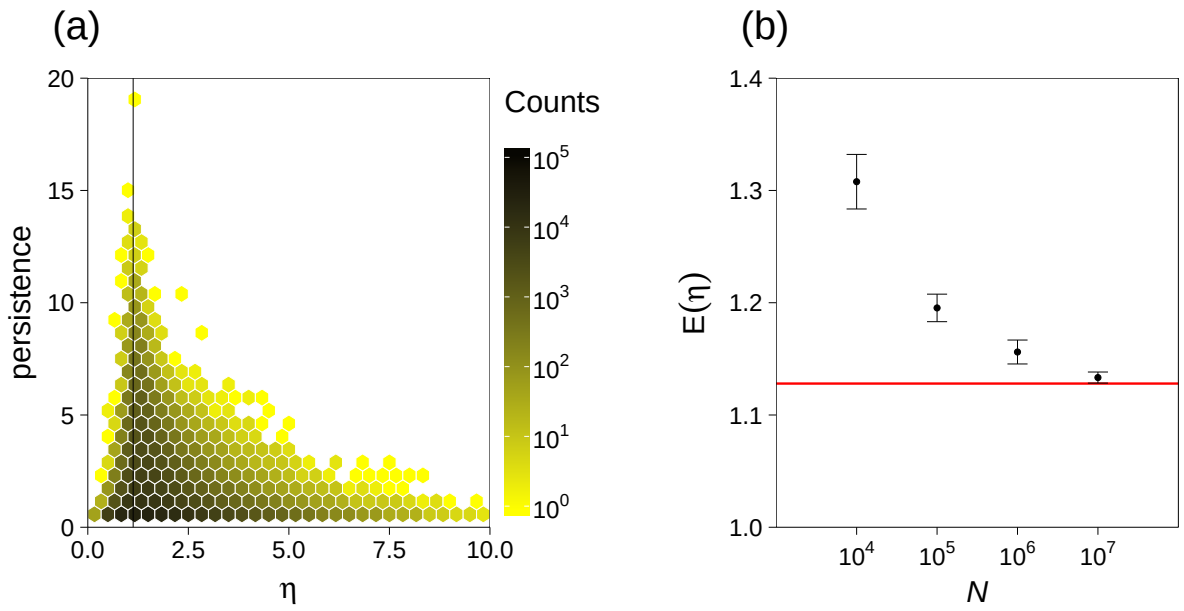


FIG. 3. (a) Birth–persistence diagram for 1D invariants of continuum percolation with disks with $N = 10^6$ disks, summarized using a hexagonal binning. The vertical black line shows $\eta_c = 1.128$. (b) Mean value of η of the longest persisting 1D invariant for different values of N. The error bars indicate the 95% confidence intervals of the mean. The horizontal line indicates the percolation threshold ($\eta_c \approx 1.128$) for $N \to \infty$.

$\eta(N, r)$. PH has been applied to investigate force networks in granular material [39, 40], to analyze brain functional networks [41], and for many other applications [22].

As we increase the value of r, we construct an alpha complex $\mathcal{C}(r)$. By increasing r, which causes $\eta(N, r)$ to change, we obtain a nested sequence (a so-called *filtration*) of simplicial complexes. Calculating PH of the filtration then allows us examine how the features evolve with the filling factor $\eta(N, r)$.

Based on the comparison of different implementations of PH by Otter et al. [22], we use the software package Dionysus [42]. (We test that our results are qualitatively the same when we use the software package (GUDHI) [43].) The output from the software consists of pairs of birth and death values. After rescaling, these give, respectively, the values of $\eta(N, r)$ at which a topological feature appears and disappears.

We visualize the output of PH using a birth–persistence diagram. In this diagram, we plot the persistence of an invariant, defined as the difference of death and birth values of $\eta(N, r)$, versus the value of $\eta(N, r)$ at which the invariant feature is born. In Fig. 3(a), we

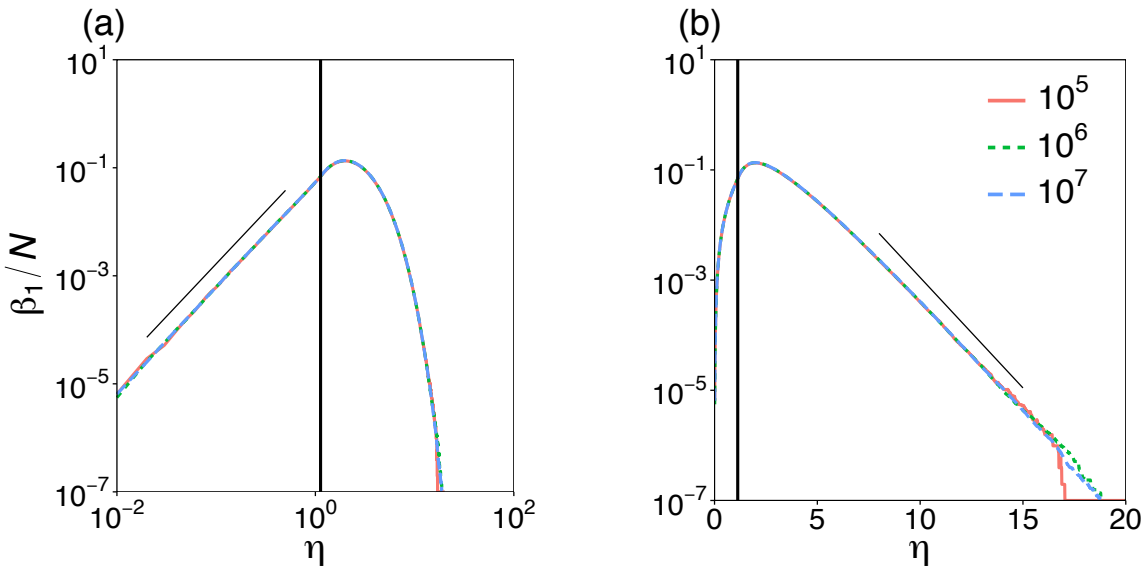


FIG. 4. Mean of β_1/N as a function of η for $N = 10^5$, $N = 10^6$, and $N = 10^7$ disks in continuum percolation with disks with (a) doubly-logarithmic axes and (b) a logarithmic vertical axis. For reference, we show the percolation threshold $\eta_c \approx 1.128$ using a vertical solid line. To guide the eye, we have drawn straight lines in both plots. The slopes of the straight lines are about 1.95 in panel (a) and about -0.92 in panel (b). We calculated these values using a least-squares fit.

show the birth–persistence diagram of one realization of continuum percolation with disks with $N = 10^6$ disks. We find that 95% of invariants persist for less than 2.11, and the mean $\eta(N, r)$ of birth is at approximately $\eta(N, r) = 1.6$ which is considerably larger than η_c. We observe an increasing trend in the persistence of the longest persisting invariants for $\eta(N, r) < \eta_c$ and a decreasing trend for $\eta(N, r) > \eta_c$. Close to $\eta(N, r) \approx \eta_c$, the longest persisting invariant is born. In Fig. 3(b), we see that the longest persisting invariant is born progressively closer to η_c for larger N. This suggests the intriguing possibility that, in the limit $N \to \infty$, the longest persisting invariant is born at η_c. It thereby suggests a possible link between PH in continuum percolation with disks and more conventional approaches to studying such percolation.

In Fig. 4, we show how β_1/N changes for $\eta(N, r) \in (0, \infty)$. We observe that the curves collapse onto a single curve for a wide range of $\eta(N, r)$, supporting the linear scaling of β_1 that is predicted by Eq. (4). In Fig. 4(a), we see that β_1/N increases in a manner that seems to resemble a power law until it peaks at a value of η that is larger than the percolation

threshold. We recall that in the subcritical regime of the percolation transition, the largest cluster is of order $\log N$ (Section III). This result therefore suggests a rapid increase in the number of holes per cluster in this regime. From Fig. 4(b), we observe that β_1/N subsequently appears to decay at least exponentially. From this, we can deduce that holes remaining in this regimes tend to be small and disappear quickly until eventually the entire unit square $[0,1]^2$ is covered by disks.

VI. CONCLUSIONS AND DISCUSSION

Continuum percolation with disks is a frequently-studied variant of continuum percolation, and many of its properties have been uncovered using a combination of theoretical and computational techniques [4–8]. In the present paper, we interpreted the union of potentially overlapping disks generated by continuum percolation with disks as a subspace of $[0,1]^2$ and examined its topological properties near the percolation transition. We computed persistent homology and identified structural changes that cannot be described solely by changes in the sizes of clusters.

We found evidence that the longest persisting invariants are born close to (though seemingly not exactly at) the percolation transition in the limit $N \to \infty$. We also provided numerical confirmation of a theoretical prediction that the first Betti number β_1 scales linearly with respect to the number N of disks, and we characterized β_1 as a function of $\eta(N,r)$.

There are several ways to extend our work. It will be interesting, for example, to apply PH to higher-dimensional versions of continuum percolation, and one can also ask how the shapes of objects in percolation models affect topological properties. The latter is reminiscent of important questions in the study of granular materials [40]. Applying PH to related percolation models, such as random sequential adsorption [18], that incorporate correlations between the placement of disks is also likely to be interesting. More generally illustrates that it is useful to examine not only the sizes of components in complex systems, but also their "shapes", a perspective that we expect will be insightful for many types of phase transitions.

VII. ACKNOWLEDGEMENTS

We thank Bernadette Stolz and Ulrike Tillmann for helpful comments. LS acknowledges the support provided by the Engineering and Physical Sciences Research Council (EPSRC) through grant number EP/G03706X/1. HAH gratefully acknowledges the EPSRC postdoctoral fellowship EP/KO4196/1 and a Royal Society University Research Fellowship.

---

[1] A. A. Saberi, Physics Reports **578**, 1 (2015).

[2] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks: A Tutorial* (Springer-Verlag (Vol. 4 in "Frontiers in Applied Dynamical Systems: Reviews and Tutorials"), Heidelberg, Germany, 2016).

[3] H. Kesten, Notices of the American Mathematical Society **53**, 572 (2006).

[4] M. D. Penrose, *Random Geometric Graphs*, Vol. 5 (Oxford University Press, 2003).

[5] R. Meester and R. Roy, *Continuum Percolation*, Vol. 119 (Cambridge University Press, 1996).

[6] E. N. Gilbert, Journal of the Society for Industrial and Applied Mathematics **9**, 533 (1961).

[7] M. Kahle and O. Bobrowski, AMS Contemporary Mathematics **620**, 201 (2014).

[8] M. Kahle, Discrete & Computational Geometry **45**, 553 (2011).

[9] D. S. Bassett, E. T. Owens, K. E. Daniels, and M. A. Porter, Phys. Rev. E **86**, 041306 (2012).

[10] J. Setford, "Models of granular networks in two and three dimensions," (2014), undergraduate Thesis, Department of Physics, University of Oxford (available at `http://www.math.ucla.edu/~mason/research/setford-final.pdf`).

[11] G. Németh and G. Vattay, Physical Review E **67**, 036110 (2003).

[12] C. P. Dettmann and O. Georgiou, Physical Review E **93**, 032313 (2016).

[13] C. Dettmann, O. Georgiou, and P. Pratt, arXiv:1803.04166 (2018).

[14] A. R. Kherlopian, T. Song, Q. Duan, M. A. Neimark, M. J. Po, J. K. Gohagan, and A. F. Laine, BMC Systems Biology **2**, 1 (2008).

[15] M. Barthlemy, Physics Reports **499**, 1 (2011).

[16] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, Physical Review Letters **104**, 048704 (2010).

[17] A. Pries, T. Secomb, and P. Gaehtgens, Cardiovascular Research **32**, 654 (1996).

[18] J. W. Evans, Reviews of Modern Physics **65**, 1281 (1993).

[19] R. N. Balasubramanian, arXiv:1803.03542 (2018).

[20] H. Edelsbrunner and J. L. Harer, *Computational Topology* (AMS, Providence, RI, USA, 2010).

[21] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational homology*, Vol. 157 (Springer Science & Business Media, New York, NY, USA, 2006).

[22] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, European Physical Journal — Data Science **6**, 17 (2017).

[23] M. Pettini, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Vol. 33 (Springer Science & Business Media, New York, NY, USA, 2007).

[24] I. Donato, M. Gori, M. Pettini, G. Petri, S. De Nigris, R. Franzosi, and F. Vaccarino, Physical Review E **93**, 052138 (2016).

[25] F. A. N. Santos, L. C. B. da Silva, and M. D. Coutinho-Filho, Journal of Statistical Mechanics: Theory and Experiment **2017**, 013202 (2017).

[26] J. Dall and M. Christensen, Physical Review E **66**, 016121 (2002).

[27] S. Mertens and C. Moore, Physical Review E **86**, 061109 (2012).

[28] M. D. Penrose, The Annals of Applied Probability **26**, 986 (2016).

[29] S. W. Haan and R. Zwanzig, Journal of Physics A: Mathematical and General **10**, 1547 (1977).

[30] J. McCarthy, Physical review letters **58**, 2242 (1987).

[31] M. D. Rintoul and S. Torquato, Journal of Physics A: Mathematical and General **30**, L585 (1997).

[32] G. R. Grimmett, *Percolation*, Grundlehren der mathematischen Wissenschaften, Vol. 321 (Springer, 1999).

[33] B. Bollobás and O. Riordan, *Percolation* (Cambridge University Press, 2006).

[34] M. E. J. Newman, *Networks: An Introduction* (Oxford university press, 2010).

[35] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, UK, 2002).

[36] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, IEEE Transactions on Information Theory **29**, 551 (1983).

[37] A. Björner, "Handbook of Combinatorics," (Elsevier Science B.V., 1995) Chap. Topological Methods, pp. 1819–1872.

[38] A. Zomorodian and G. Carlsson, Discrete & Computational Geometry **33**, 249 (2005).

[39] M. Kramár, A. Goullet, L. Kondic, and K. Mischaikow, Physical Review E **90**, 052203 (2014).

[40] L. Papadopoulos, M. A. Porter, K. E. Daniels, and D. S. Bassett, Journal of Complex Networks , advanced access (available at https://doi.org/10.1093/comnet/cny005) (2018).

[41] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Journal of the Royal Society Interface **11** (2014).

[42] D. Morozov, "Dionysus," Software available at http://www.mrzv.org/software/dionysus/.

[43] C. Maria, J.-D. Boissonnat, M. Glisse, and M. Yvinec, in *International Congress on Mathematical Software* (Springer, 2014) pp. 167–174.