



# Maximal persistence in random clique complexes

Ayat Ababneh<sup>1</sup> · Matthew Kahle<sup>2</sup>

Received: 14 September 2022 / Revised: 31 May 2023 / Accepted: 23 June 2023 /  
Published online: 10 July 2023  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

## Abstract

We study the persistent homology of an Erdős–Rényi random clique complex filtration on  $n$  vertices. Here, each edge  $e$  appears independently at a uniform random time  $p_e \in [0, 1]$ , and the persistence of a cycle  $\sigma$  is defined as  $p_2(\sigma)/p_1(\sigma)$ , where  $p_1(\sigma)$  and  $p_2(\sigma)$  are the birth and death times of  $\sigma$ . We show that if  $k \geq 1$  is fixed, then with high probability the maximal persistence of a  $k$ -cycle is of order  $n^{1/(k+1)}$ .

**Keywords** Random graphs · Persistent homology

**Mathematics Subject Classification** 55N31 · 05C80

## 1 Introduction

Recently, the topology of random simplicial complexes has been an active area of study — see, for example, the surveys (Kahle 2017; Bobrowski and Kahle 2018). This study has had applications in topological data analysis, including in neuroscience (Giusti et al. 2015). One of the main methodologies of topological data analysis is *persistent homology*. We will assume that the reader is familiar with the notions of persistent homology and of a persistence diagram (Edelsbrunner et al. 2002). In topological inference, one sometimes considers points far from the diagonal in the persistence diagram to be representing “signal” and points near the diagonal as representing “noise”.

With this in mind, Bobrowski, Kahle, and Skraba studied maximally persistent cycles in random geometric complexes in Bobrowski et al. (2017). Both the Vietoris–Rips and Čech filtrations have an underlying parameter  $r$ . Persistence of a cycle is measured multiplicatively as  $r_2/r_1$  where  $r_1$  and  $r_2$  are the birth and death radius. We write  $f \asymp g$  if  $f$  and  $g$  grow at the same rate in the sense that there exist constants

---

✉ Matthew Kahle  
mkahle@math.osu.edu

<sup>1</sup> Department of Mathematics, The University of Jordan, Amman 11942, Jordan

<sup>2</sup> Department of Mathematics, The Ohio State University, Columbus, OH, USA

$c_1, c_2 > 0$  such that  $c_1 f(n) \leq g(n) \leq c_2 f(n)$  for all large enough  $n$ . They showed that with high probability the maximal persistence of a  $k$ -dimensional cycle in a random geometric complex in  $\mathbb{R}^d$  is  $\asymp (\log n / \log \log n)^{1/k}$ . Here the implied constants depend on  $d$  and  $k$ , but not on  $n$ .

Our main result is that for fixed  $k \geq 1$ , the maximal persistence of  $k$ -cycles in an Erdős–Rényi random clique complex filtration  $X(n, p)$  is of order  $n^{1/k(k+1)}$ . The definition of random clique complex is given in Sect. 2 and a precise statement of the theorem is given in Theorem 4.1. In a similar way, we measure the persistence of a cycle multiplicatively, but now as  $p_2/p_1$  where  $p_1$  and  $p_2$  are the birth and death edge probabilities, respectively.

The comparison between the Erdős–Rényi and random geometric settings may be more apparent if we renormalize so that the persistence of the associated filtrations can be measured on the same scale. One natural way to do this is to reconsider the earlier results for maximal persistence in random geometric complexes, using instead birth and death edge probability rather than radius.

The edge probability  $P$  in a random geometric complex is of order  $P \asymp r^d$ . So if

$$r_2/r_1 \asymp (\log n / \log \log n)^{1/k},$$

then

$$P_2/P_1 \asymp (\log n / \log \log n)^{d/k}.$$

As long as  $d$  is fixed,  $(\log n / \log \log n)^{d/k}$  is still much smaller than the maximal persistence of cycles in the random clique complex. However, this parameterization makes it clear that if  $d$  grows, we expect cycles to persist for longer. It is known that random geometric graphs in dimension growing quickly enough converge in total variation distance to Erdős–Rényi random graphs, and this connection has been further explored and quantified in a number of recent papers — see, for example (Bubeck et al. 2016; Brennan et al. 2020; Paquette and Werf 2021). From this point of view, our main result can be seen as a “curse of dimensionality” for topological inference—the ambient dimension gets bigger, noisy cycles persist for much longer.

## 2 Topology of random clique complexes

In this section, we review the definition of the random clique complex, and briefly survey the literature on topology of random clique complexes.

We use the notation  $[n] := \{1, 2, \dots, n\}$ . The following random graph is sometimes called the *Erdős–Rényi model*.

**Definition 2.1** For  $n \geq 1$  and  $p \in [0, 1]$ ,  $G(n, p)$  is the probability space of all graphs on vertex set  $[n]$  where every edge is included with probability  $p$ , jointly independently.

We use the notation  $G \sim G(n, p)$  to indicate that a graph  $G$  is chosen according to this distribution. We say that  $G$  has a given property *with high probability (w.h.p.)* if the probability that  $G$  has the property tends to one as  $n \rightarrow \infty$ .

The *clique complex* of a graph  $H$  is an abstract simplicial complex whose faces are sets of vertices in  $H$  which form cliques. We define the *random clique complex*  $X(n, p)$  to be the clique complex of the Erdős–Rényi random graph  $G(n, p)$ . We write  $X \sim X(n, p)$  to indicate that  $X$  is a random simplicial complex chosen according to this distribution.

In Kahle (2009), Kahle studied the topology of random clique complexes, and the following theorem identified the threshold for homology to appear as  $p = n^{-1/k}$ . We use the notation  $f \ll g$  to indicate  $\lim_{n \rightarrow \infty} f/g = 0$ .

**Theorem 2.2** *Let  $X \sim X(n, p)$  be the random clique complex, and assume  $k \geq 1$  is fixed.*

- (1) *If  $p \leq n^{-\alpha}$  with  $\alpha > \frac{1}{k}$ , then w.h.p.  $H_k(X) = 0$ . On the other hand,*
- (2) *If  $n^{-\frac{1}{k}} \ll p \ll n^{-\frac{1}{k+1}}$  then w.h.p.  $H_k(X) \neq 0$ .*

The following is the main result of a later paper, Kahle (2014), showing that the threshold for  $k$ th cohomology with coefficients in  $\mathbb{Q}$  to vanish is approximately  $p = n^{-1/(k+1)}$ .

**Theorem 2.3** *Let  $k \geq 1$  and  $\epsilon > 0$  be fixed, and  $X \sim X(n, p)$ . If*

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

*then w.h.p.  $H^k(X, \mathbb{Q}) = 0$ .*

Theorem 2.3 describes a sharp threshold for cohomology to vanish, in the same spirit as in Linial and Meshulam’s work (Linial and Meshulam 2006). By the universal coefficient theorem for cohomology, these results hold for homology vanishing as well.

The proof of Theorem 2.3 in Kahle (2014) depends on new results on spectral gaps of random graphs which appeared in Hoffman et al. (2021), together with Garland’s method, which is similar in spirit to combinatorial Hodge theory, relating spectra of Laplacians on  $k$  forms with  $k$ th cohomology. As such, the proof only works over a field of characteristic zero. Extending Theorem 2.3 to  $\mathbb{Z}$  coefficients remains one of the main open problems about the topology of random clique complexes, and is equivalent to the “bouquet-of-spheres conjecture.” See the discussion in Kahle (2014).

Since we depend here on an extension of Theorem 2.3 that we will prove in Sect. 4, our results also only hold with  $\mathbb{Q}$  coefficients, or over a field of characteristic zero.

Malen gave a topological strengthening of part (1) of Theorem 2.2 in Malen (2019).

**Theorem 2.4** (Malen, 2019). *Let  $k \geq 1$  be fixed and  $X \sim X(n, p)$ . If  $p \leq n^{-\alpha}$  with  $\alpha > \frac{1}{k}$ , then w.h.p.  $X$  collapses onto a subcomplex of dimension at most  $k - 1$ .*

This implies, in particular, that  $H_{k-1}(X)$  is torsion-free, so this represents an important step toward the “bouquet-of-spheres conjecture” described in Kahle (2009, 2014).

Newman recently refined Malen’s collapsing argument to give a probabilistic refinement (Newman 2021).

**Theorem 2.5** (Newman, 2021). Let  $k \geq 1$  be fixed and  $X \sim X(n, p)$ . If

$$p \ll n^{-1/k}$$

then w.h.p.  $X$  collapses onto a subcomplex of dimension at most  $k - 1$ .

In summary, earlier results show that there is one threshold where homology is born for the first time, when  $p \approx n^{-1/k}$ , and another where homology dies for the last time, when  $p \approx n^{-1/(k+1)}$ . Our main result is that there exist cycles that persist for nearly the entire interval of nontrivial homology.

### 3 The second moment method

We briefly review the second moment method, i.e. the use of Chebyshev's inequality, which is our main probabilistic tool. The *variance* of a random variable  $X$  is defined by

$$\sigma^2 := \text{Var}(X) := \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

The *covariance* of a pair of random variables  $X, Y$  is defined by

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \mathbb{E}(X_j).$$

**Theorem 3.1** (Chebyshev's Inequality). For any  $\lambda > 0$ ,

$$\mathbb{P}(|X - \mu| \geq \lambda) \leq \frac{\sigma^2}{\lambda^2}.$$

Where  $\mu$  is the expectation and  $\sigma^2$  is the variance.

If  $X$  can be written as a sum of indicator random variables  $X = \sum_i X_i$ , then the following is easy to derive and its proof appears, for example, in Chapter 4 of Alon and Spencer's book (Alon and Spencer 2016).

$$\begin{aligned} \text{Var}(X) &= \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &\leq \mathbb{E}(X) + \sum_{i \neq j} \text{Cov}(X_i, X_j). \end{aligned}$$

It follows from Theorem 3.1 that if  $\mathbb{E}(X) \rightarrow \infty$  and

$$\sum_{i \neq j} \text{Cov}(X_i, X_j) = o\left(\mathbb{E}(X)^2\right),$$

then  $X > 0$  w.h.p. In fact,  $X \sim \mathbb{E}(X)$  w.h.p., meaning that  $X/\mathbb{E}(X) \rightarrow 1$  in probability.

Finally, we note that if  $X_i, X_j$  are indicator random variables for events  $A_i, A_j$ , we have that

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(A_i \wedge A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Here  $A_i \wedge A_j$  denotes the event that both  $A_i$  and  $A_j$  occur.

## 4 Main result and proof

We consider the random graph  $G(n, p)$  as a stochastic process, as follows. Consider the random filtration of the complete graph  $K_n$  where each edge  $e$  appears at time  $p_e$ , chosen uniform randomly in the interval  $[0, 1]$ . Similarly, the random clique complex  $X(n, p)$  is a random filtration of the simplex on  $n$  vertices  $\Delta_n$ .

We assume the reader is familiar with persistence diagrams (Edelsbrunner et al. 2002; Cohen-Steiner et al. 2007). A point  $(x, y)$  in the persistence diagram for  $H_k$  with  $\mathbb{Q}$  coefficients and  $k \geq 1$  represents a  $k$ -dimensional cycle with birth time  $x$  and death time  $y$ . We measure the persistence of that cycle multiplicatively, as  $y/x$ . Define

$$M_k(n) = \max\{y/x\},$$

where the maximum is taken over all points in the persistence diagram for homology in degree  $k$ .

An equivalent definition is the following. Consider the natural inclusion map  $i : X(n, p_1) \hookrightarrow X(n, p_2)$ , where  $0 \leq p_1 \leq p_2 \leq 1$ . For every  $k \geq 1$ , there is an induced map on homology  $i_* : H_k(X(n, p_1)) \rightarrow H_k(X(n, p_2))$ . We define

$$M_k(n) := \max\{p_2/p_1 \mid i_* : H_k(X(n, p_1)) \rightarrow H_k(X(n, p_2)) \text{ is nontrivial}\},$$

where the maximum is taken as  $p_1$  and  $p_2$  range over all values with  $0 \leq p_1 \leq p_2 \leq 1$ . Our main result is the following.

**Theorem 4.1** *For fixed  $k \geq 1$  and  $\epsilon > 0$ ,*

$$n^{1/k(k+1)-\epsilon} \leq M_k(n) \leq n^{1/k(k+1)+\epsilon},$$

*with high probability.*

*Equivalently, if*

$$\tilde{M}_k(n) = \frac{\log M_k(n)}{\log n},$$

*then  $\tilde{M}_k(n)$  converges in probability to  $1/k(k+1)$ .*

Our results are actually slightly sharper than this. We show in the following that if

$$L_k(n) \ll n^{1/k(k+1)}$$

and

$$U_k(n) \gg n^{1/k(k+1)} (\log n)^{1/(k+1)},$$

then w.h.p.

$$L_k(n) \leq M_k(n) \leq U_k(n).$$

So our results are sharp, up to a small power of  $\log n$ .

**Proof of Theorem 4.1** First we prove an upper bound on  $M_k(n)$ . Suppose that  $p_1 \ll n^{-1/k}$ . By Theorem 2.5, w.h.p. we have that  $H_k(X(n, p_1)) = 0$ . Now let  $\epsilon > 0$ , and suppose that

$$p_2 \geq \left( \frac{(\frac{k}{2} + 1 + \epsilon) \log n}{n} \right)^{1/(k+1)}.$$

By Theorem 2.3, w.h.p.  $H_k(X(n, p_2), \mathbb{Q}) = 0$ . So for any cycle  $\sigma$  that is born after time  $p_1$  and before time  $p_2$ , the persistence of  $\sigma$  is at most  $p_2/p_1$ .

This seems like it might already imply an upper bound on  $M_k(n)$ , but unfortunately it is not quite enough. Theorem 2.5 does not state that w.h.p.  $H_k(X(n, p)) = 0$  for all  $p \in [0, p_1]$ . Similarly, Theorem 2.3 does not state that w.h.p.  $H_k(X(n, p)) = 0$  for all  $p \in [p_2, 1]$ . Although we believe such statements are almost certainly true, we have another way to get the desired upper bound on persistence.

Let  $I = \{0, 1, \dots, k-1\} \cup \{k+2, k+3, \dots, 2k(k+1)\}$ , and set

$$S = \left\{ \frac{i}{k(k+1)} \mid i \in I \right\}.$$

By repeatedly applying Theorems 2.5 and 2.3, w.h.p.  $H_k(X(n, p)) = 0$  whenever  $p = n^{-\alpha}$  with  $\alpha \in S$ . The point is that there are only a constant number of elements in  $S$ , since we are assuming throughout that  $k$  is fixed and  $n \rightarrow \infty$ .

For any  $k$ -cycles that are born and die after time  $p_2$ , the multiplicative persistence is at most  $n^{1/k(k+1)}$ , which is smaller than our desired upper bound. We can make the same argument for any  $k$ -cycles that are born and die before time  $p_1$ . It suffices to consider indices in  $I$  only up to  $2k(k+1)$  since w.h.p.  $G(n, p)$  has no edges when  $p \ll n^{-2}$ .

Set

$$f_k(n) = n^{1/k(k+1)} (\log n)^{1/(k+1)},$$

and let  $U_k(n)$  be any function such that  $U_k(n) \gg f_k(n)$ . We have showed that

$$M_k(n) \leq U_k(n).$$

Most of our work is in proving a lower bound for  $M_k(n)$ . We focus our attention on a particular type of nontrivial  $k$ -cycle, namely simplicial spheres which are combinatorially isomorphic to cross-polytope boundaries.

In the following, let  $Y$  and  $Z$  denote distinct subsets of  $2k + 2$  vertices. That is, we suppose that  $Y, Z \subseteq [n]$  with  $|Y| = |Z| = 2k + 2$ . A notation we can use for this is

$$Y, Z \in \binom{[n]}{2k+2}.$$

Suppose that  $Y = \{u_1, \dots, u_{k+1}\} \cup \{v_1, \dots, v_{k+1}\}$ , where  $u_1 < \dots < u_{k+1} < v_1 < \dots < v_{k+1}$ . Recall that every vertex is an element of  $[n]$ , so they come with a natural ordering. We use  $x \sim y$  and  $x \not\sim y$  to denote adjacency and non-adjacency of vertices  $x$  and  $y$ . For any choice of  $0 \leq p_1 \leq p_2 \leq 1$ , we say that  $Y$  is a  $(p_1, p_2)$  *special persistent cycle* in the random clique complex filtration if

- (1)  $u_i \sim u_j, v_i \sim v_j$ , and  $u_i \not\sim v_j$  for every  $i \neq j$  at time  $p_1$ ,
- (2)  $u_i \not\sim v_i$  for every  $i$  at time  $p_2$ , and
- (3)  $\{u_1, \dots, u_{k+1}\}$  have no common neighbors outside of vertex set  $Y$  at time  $p_2$ .

Condition (1) implies that  $Y$  spans a  $k$ -dimensional cycle at time  $p_1$ , namely a cycle that is combinatorially equivalent to the boundary of  $k + 1$ -dimensional cross-polytope. Conditions (2) and (3) together imply that  $Y$  is still not a boundary at time  $p_2$ . So then it is not only a nontrivial cycle at time  $p_1$ , but it persists at least until time  $p_2$ . Note that condition (2) already implies that  $\{u_1, \dots, u_{k+1}\}$  have no common neighbor within vertex set  $Y$ . So condition (3) implies that they have no common neighbor at all, and then  $\{u_1, \dots, u_{k+1}\}$  is a maximal  $k$ -dimensional face.

Let  $N_k = N_k(p_1, p_2)$  be the number of  $(p_1, p_2)$  special persistent cycles. We want to show that  $\mathbb{P}(N_k > 0) \rightarrow 1$ , which in turn will imply that  $M_k(n) \geq p_2/p_1$  with high probability. In the following, we will assume whenever necessary that

$$n^{-1/k} \ll p_1 \leq p_2 \ll n^{-1/(k+1)}.$$

In particular, we assume that  $np_1^k \rightarrow \infty$  and  $np_2^{k+1} \rightarrow 0$ .

Let  $A_Y$  be the event that the set of vertices in  $Y$  form a  $(p_1, p_2)$  special persistent cycle, and let  $I_Y$  be its indicator random variable for this event. Then we can write

$$N_k = \sum_{Y \in \binom{[n]}{2k+2}} I_Y,$$

where the sum is taken over all subsets  $Y \subseteq [n]$  of size  $|Y| = 2k + 2$ .

By edge independence, the probability of condition (1) is  $p_1^{2k(k+1)}$ , and the probability of condition (2) is  $(1 - p_2)^{k+1}$ , the probability of condition (3) is

$\left(1 - p_2^{k+1}\right)^{n-2k-2}$ . Moreover, these events are independent since they involve disjoint sets of edges. So we have

$$\mathbb{E}(I_Y) = \mathbb{P}(A_Y) = p_1^{2k(k+1)} (1 - p_2)^{k+1} \left(1 - p_2^{k+1}\right)^{n-2k-2}.$$

By linearity of expectation,

$$\begin{aligned} \mathbb{E}(N_k) &= \sum \mathbb{E}(I_Y) \\ &= \binom{n}{2k+2} p_1^{2k(k+1)} (1 - p_2)^{k+1} \left(1 - p_2^{k+1}\right)^{n-2k-2} \\ &= \frac{n^{2k+2}}{(2k+2)!} p_1^{2k(k+1)} (1 - o(1)), \end{aligned}$$

since  $np_2^{k+1} \rightarrow 0$ . Since we also assume that  $np_1^k \rightarrow \infty$ , we have  $\mathbb{E}(N_k) \rightarrow \infty$ . By Chebyshev's inequality, if we show that  $\text{Var}(N_k) = o(\mathbb{E}(N_k)^2)$ , then  $N_k > 0$  w.h.p.

We have the standard inequality

$$\text{Var}(N_k) \leq \mathbb{E}(N_k) + \sum_{Y \neq Z} \text{Cov}(I_Y, I_Z).$$

We recall that

$$\text{Cov}(I_Y, I_Z) = \mathbb{P}(A_Y \wedge A_Z) - \mathbb{P}(A_Y)\mathbb{P}(A_Z).$$

We always have

$$\mathbb{P}(A_Y)\mathbb{P}(A_Z) = p_1^{4k(k+1)} (1 - p_2)^{2k+2} \left(1 - p_2^{k+1}\right)^{2(n-2k-2)},$$

and we note the simpler estimate

$$\mathbb{P}(A_Y)\mathbb{P}(A_Z) = p_1^{4k(k+1)} (1 - o(1))$$

since  $k \geq 1$  is fixed and  $np_2^{k+1} \rightarrow 0$ .

Let  $m := |Y \cap Z|$ . In estimating  $\mathbb{P}(A_Y \wedge A_Z)$ , we consider cases depending on the value of  $m$ .

### Case I:

First, consider  $m = 0$ . It might be tempting to believe that if  $Y \cap Z = \emptyset$  then  $A_Y$  and  $A_Z$  are independent sets and the covariance is zero. Unfortunately, this is not the case. Conditions (1) and (2) for a  $(p_1, p_2)$  special persistent cycle only depend



on adjacency between vertices within the  $(2k + 2)$ -set, but condition (3) depends on connections with the rest of the graph and these are not independent.

Nevertheless, we still have in this case

$$\mathbb{P}(A_Y \wedge A_Z) = p_1^{4k(k+1)} (1 - o(1)),$$

as follows.

The term  $p_1^{4k(k+1)}$  is the probability of condition (1) holding for both vertex sets  $Y$  and  $Z$ . So this is also an upper bound on the probability of conditions (1), (2), and (3) holding for both vertex sets. For a lower bound on  $\mathbb{P}(A_Y \wedge A_Z)$ , we consider a slightly smaller event, slightly simpler but whose probability is of the same order of magnitude.

Let  $Y = \{u_1, \dots, u_{k+1}\} \cup \{v_1, \dots, v_{k+1}\}$ , where  $u_1 < \dots < u_{k+1} < v_1 < \dots < v_{k+1}$ , as before. Similarly, let  $Z = \{u'_1, \dots, u'_{k+1}\} \cup \{v'_1, \dots, v'_{k+1}\}$ , where  $u'_1 < \dots < u'_{k+1} < v'_1 < \dots < v'_{k+1}$ .

The event  $A_{YZ}^*$  is defined as follows.

- (1) We have  $u_i \sim u_j$ ,  $v_i \sim v_j$ , and  $u_i \sim v_j$ ,  $u'_i \sim u'_j$ ,  $v'_i \sim v'_j$ , and  $u'_i \sim v'_j$  for every  $i \neq j$  at time  $p_1$ . That is, condition (1) holds for both  $Y$  and  $Z$ . Some edges may be listed more than once if  $Y$  and  $Z$  overlap. This does not happen when  $m = 0$  but these are the cases we consider below.
- (2) Besides the edges that appear in the previous condition, no other edges occur between vertices in vertex set  $Y \cup Z$ , at time  $p_2$ . This happens with probability  $1 - O(p_2) = 1 - o(1)$ .
- (3) Neither  $\{u_1, \dots, u_{k+1}\}$  nor  $\{u'_1, \dots, u'_{k+1}\}$  has any mutual neighbors outside of vertex set  $Y \cup Z$ , at time  $p_2$ . The probability of this condition being satisfied can be bounded below by a union bound by  $1 - 2np_2^{k+1}$ , which is again  $1 - o(1)$  since  $np_2^{k+1} \rightarrow 0$ .

Putting it all together, we have that  $\mathbb{P}(A_{YZ}^*) \geq p_1^{4k(k+1)} (1 - o(1))$ . We note that  $A_{YZ}^*$  implies  $A_Y \wedge A_Z$ . Indeed, condition (1) is the same, condition (2) of  $A_{YZ}^*$  implies condition (2) of  $A_Y \wedge A_Z$ , and conditions (2) and (3) of  $A_{YZ}^*$  together imply condition (3) of  $A_Y \wedge A_Z$ .

Then

$$p_1^{4k(k+1)} \geq \mathbb{P}(A_Y \wedge A_Z) \geq \mathbb{P}(A_{YZ}^*) \geq p_1^{4k(k+1)} (1 - o(1)),$$

and

$$\mathbb{P}(A_Y \wedge A_Z) = p_1^{4k(k+1)} (1 - o(1)),$$

as desired.

So then

$$\mathbb{P}(A_Y \wedge A_Z) - \mathbb{P}(A_Y)\mathbb{P}(A_Z) = o\left(p_1^{4k(k+1)}\right).$$

Since the number of pairs  $Y, Z$  is bounded by  $n^{4k+4}$  we have that the total contribution to the variance,  $S_0$ , is bounded by

$$S_0 = o\left(n^{4k+4} p_1^{4k(k+1)}\right).$$

Comparing this to

$$\mathbb{E}(N_k)^2 = \binom{n}{2k+2}^2 p_1^{4k(k+1)} (1 - o(1))$$

we see that

$$S_0 = o\left(\mathbb{E}(N_k)^2\right).$$

### Case II:

An essentially identical calculation shows that when  $m = 1$ , we have

$$\mathbb{P}(A_Y \wedge A_Z) = p_1^{4k(k+1)} (1 - o(1)).$$

So in this case we have again

$$\mathbb{P}(A_Y \wedge A_Z) - \mathbb{P}(A_Y)\mathbb{P}(A_Z) = o\left(p_1^{4k(k+1)}\right).$$

Hence, the total contribution to the variance,  $S_1$ , is

$$S_1 = o\left(n^{4k+3} p_1^{4k(k+1)}\right)$$

and then,

$$S_1/\mathbb{E}(N_k)^2 = o\left(n^{-1}\right)$$

and in particular  $S_1 = o\left(\mathbb{E}(N_k)^2\right)$ .

### Case III:

When  $2 \leq m \leq 2k+1$ , we consider two sub-cases. The first subcase is that events  $A_Y$  and  $A_Z$  are not compatible in the sense that they cannot both occur due to the ways in which  $Y$  and  $Z$  overlap. This happens if for a certain pair of vertices  $u, v \in Y \cap Z$ ,  $u, v$  are required to be adjacent in one of  $Y, Z$  and non-adjacent in the other. In this subcase, we have

$$\mathbb{P}(A_Y \wedge A_Z) = 0,$$

so

$$\mathbb{P}(A_Y \wedge A_Z) - \mathbb{P}(A_Y)\mathbb{P}(A_Z) = -p^{4k(k+1)} (1 - o(1)) \leq 0.$$

The second subcase is that the events  $A_Y$  and  $A_Z$  are compatible, in the sense that they could possibly both happen. In this case, let  $j$  denote the number of pairs in  $Y \cap Z$  that are forced to be non-adjacent in  $A_Y \wedge A_Z$ . Then the same argument as in Case I shows that

$$\mathbb{P}(A_Y \wedge A_Z) \geq \mathbb{P}(A_{YZ}^*) = p_1^{4k(k+1) - \binom{m}{2} + j} (1 - o(1)).$$

So

$$\frac{\mathbb{P}(A_Y)\mathbb{P}(A_Z)}{\mathbb{P}(A_Y \wedge A_Z)} \leq p_1^{\binom{m}{2} - j} (1 - o(1))$$

Since  $p_1 \rightarrow 0$  as  $n \rightarrow \infty$ , we have

$$\frac{\mathbb{P}(A_Y)\mathbb{P}(A_Z)}{\mathbb{P}(A_Y \wedge A_Z)} \rightarrow 0.$$

So,

$$\mathbb{P}(A_Y \wedge A_Z) - \mathbb{P}(A_Y)\mathbb{P}(A_Z) = (1 - o(1)) \mathbb{P}(A_Y \wedge A_Z).$$

The total contribution  $S_m$  of a pair of events  $A_Y$  and  $A_Z$  with  $Y \cap Z = m$  to the variance is then bounded by

$$S_m \leq n^{4k+4-m} p_1^{4k(k+1) - \binom{m}{2} + j} (1 + o(1)).$$

Comparing this to

$$\mathbb{E}(N_k)^2 = \binom{n}{2k+2}^2 p_1^{4k(k+1)} (1 - o(1)).$$

we get

$$S_m / \mathbb{E}(N_k)^2 = O\left(n^{-m} p_1^{-\binom{m}{2} + j}\right).$$

We have

$$n^{-m} p_1^{-\binom{m}{2} + j} = \left(np_1^{\frac{m-1}{2}}\right)^{-m} p_1^j.$$

We are assuming that  $np_1^k \rightarrow \infty$ . Since  $m \leq 2k + 1$ , we have  $k \geq (m - 1)/2$ . Then

$$\left(np_1^{\frac{m-1}{2}}\right)^{-m} \rightarrow 0,$$

$p_1^j \rightarrow 0$ , and  $S_m = o(\mathbb{E}(N_k)^2)$ .

Summing the inequalities from the different cases, we conclude that

$$\sum_{Y \neq Z} \text{Cov}(I_Y, I_Z) = \sum_{m=0}^{2k+1} S_m = o(\mathbb{E}(N_k)^2),$$

since  $S_m = o(\mathbb{E}(N_k)^2)$  for each  $m$  and  $k$  is fixed.

We conclude that as long as

$$n^{-1/k} \ll p_1 \leq p_2 \ll n^{-1/(k+1)},$$

then  $N_k > 0$  with high probability.

It follows that if  $L_k(n) \ll n^{1/k(k+1)}$  then w.h.p.  $M_k(n) \geq L_k(n)$ , as desired.  $\square$

## 5 Future directions

Recall that we earlier defined

$$f_k(n) = n^{1/k(k+1)} (\log n)^{1/(k+1)}.$$

We believe that the  $M_k(n)$  is likely of order  $f_k(n)$ , in the following sense.

Let  $\omega(n)$  be any function that tends to infinity with  $n$ . We showed in the proof of Theorem 4.1 that

$$M_k(n) \leq f_k(n)\omega(n).$$

We believe that an analogous lower bound should hold.

**Conjecture 5.1** *Let  $M_k(n)$  denote the maximal persistence over all  $k$ -dimensional cycles in  $X(n, p)$ . Then*

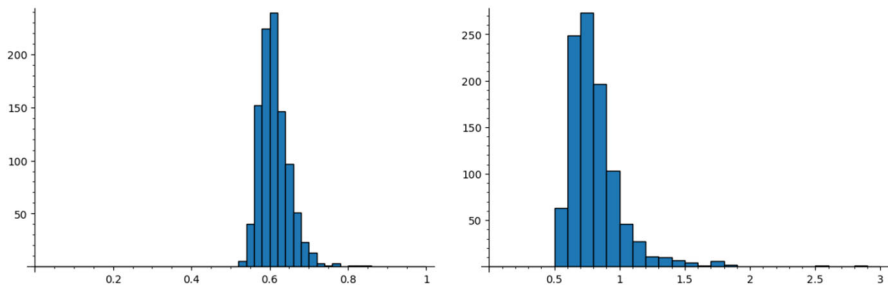
$$\frac{f_k(n)}{\omega(n)} \leq M_k(n)$$

*with high probability.*

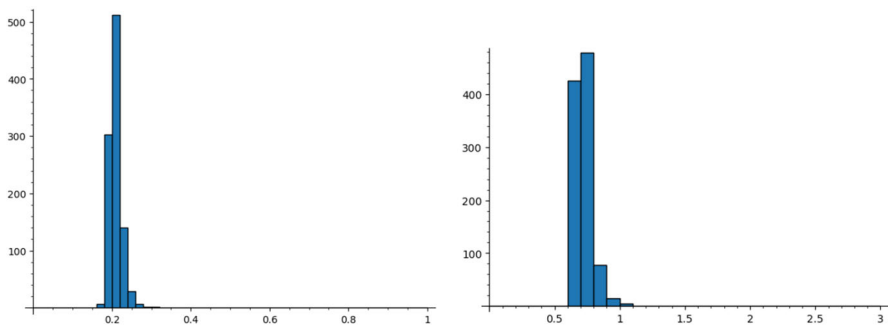
The following kind of limit theorem would provide precise answers to questions like, “Given a prior of this kind of distribution, what is the probability  $P(\lambda)$  that there exists a cycle of persistence greater than  $\lambda$ ?”

**Conjecture 5.2** *Let  $M_k(n)$  denote the maximal persistence over all  $k$ -dimensional cycles in  $X(n, p)$ . Then*

$$\frac{M_k(n)}{f_k(n)}$$



**Fig. 1** Maximally persistent 1-cycles. On the left, a histogram for  $\log M_1(n)/\log n$ . We prove that this converges to  $1/2$  as  $n \rightarrow \infty$ . On the right, a histogram for  $M_1(n)/f_1(n)$ . Both these figures are based on 1000 samples on  $n = 250$  vertices



**Fig. 2** Maximally persistent 2-cycles. On the left, a histogram for  $\log M_2(n)/\log n$ . We prove that this converges to  $1/6$  as  $n \rightarrow \infty$ . On the right, a histogram for  $M_2(n)/f_2(n)$ . Both these figures are based on 1000 samples on  $n = 150$  vertices

converges in law to a limiting distribution supported on an interval  $[\lambda_k, \infty)$  for some  $\lambda_k > 0$ .

See Figs. 1 and 2 for some numerical experiments illustrating these conjectures. These experiments were computed with the aid of Ulrich Bauer's software Ripser (Bauer 2021). We note that even though we have proved that asymptotically  $\log M_1(n)/\log n \rightarrow 1/2$  and  $\log M_2(n)/\log n \rightarrow 1/6$ , this is not apparent from our numerical experiments. So this is a hint that the rate of convergence may be slow.

It also seems natural to study more about the "rank invariant" of a random clique complex filtration. That is, given  $k \geq 1$ ,  $p_1$ , and  $p_2$ , how large do we expect the rank of the map  $i_* : H_k(X(n, p_1)) \rightarrow H_k(X(n, p_2))$  to be?

**Conjecture 5.3** Suppose that  $k \geq 1$  is fixed, and

$$n^{-1/k} \ll p_1 \leq p_2 \ll n^{-1/(k+1)}.$$

If  $i : X(n, p_1) \rightarrow X(n, p_2)$  is the inclusion map, and

$$i_* : H_k(X(n, p_1)) \rightarrow H_k(X(n, p_2))$$

is the induced map on homology, then

$$\text{rank}(i_*) = (1 - o(1)) \binom{n}{k+1} p_1^{\binom{k+1}{2}}.$$

In Kahle (2009), it is shown that

$$\dim H_k(X(n, p_1), \mathbb{Q}) = (1 - o(1)) \binom{n}{k+1} p_1^{\binom{k+1}{2}},$$

so this conjecture is that almost all of the homology persists for as long as possible.

Bobrowski and Skraba study limiting distributions for maximal persistence in their recent preprint (Bobrowski and Skraba 2022). They describe experimental evidence that there is a universal distribution for persistence over a wide class of models, including random Čech and Vietoris–Rips complexes. We do not know whether we should expect the random clique complex filtration studied here to be in the same conjectural universality class.

**Acknowledgements** We thank both anonymous referees for their corrections and helpful comments. MK also thanks Greg Malen and Andrew Newman for several helpful conversations.

**Funding** Both authors gratefully acknowledge the support of NSF-DMS #2005630.

## Declarations

**Conflict of interest** The authors have no competing interests to declare that are relevant to the content of this article.

## References

- Alon, N., Spencer, J.H.: The Probabilistic Method, 4th edn. Wiley, Hoboken (2016)
- Bauer, U.: Ripser: efficient computation of Vietoris–Rips persistence barcodes. *J. Appl. Comput. Topol.* **5**(3), 391–423 (2021)
- Bobrowski, O., Kahle, M.: Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.* **1**(3–4), 331–364 (2018)
- Bobrowski, O., Kahle, M., Skraba, P.: Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.* **27**(4), 2032–2060 (2017)
- Bobrowski, O., Skraba, P.: On the Universality of Random Persistence Diagrams. (submitted), [arXiv:2207.03926](https://arxiv.org/abs/2207.03926), (2022)
- Brennan, M., Bresler, G., Nagaraj, D.: Phase transitions for detecting latent geometry in random graphs. *Probab. Theory Related Fields* **178**(3–4), 1215–1289 (2020)
- Bubeck, S., Ding, J., Eldan, R., Rácz, M.Z.: Testing for high-dimensional geometry in random graphs. *Random Struct. Algorithms* **49**(3), 503–532 (2016)
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete Comput. Geom.* **37**(1), 103–120 (2007)
- Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. *Discrete Comput. Geom.* **28**(4), 511–533 (2002)
- Giusti, C., Pastalkova, E., Curto, C., Itskov, V.: Clique topology reveals intrinsic geometric structure in neural correlations. *Proc. Natl. Acad. Sci.* **112**(44), 13455–13460 (2015)
- Hoffman, C., Kahle, M., Paquette, E.: Spectral gaps of random graphs and applications. *Int. Math. Res. Not. IMRN* **2021**(11), 8353–8404 (2021)

- Kahle, M.: Topology of random clique complexes. *Discrete Math.* **309**(6), 1658–1671 (2009)
- Kahle, M.: Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. Math.* **179**(3), 1085–1107 (2014)
- Kahle, M.: Random simplicial complexes. In *Handbook of Discrete and Computational Geometry*, pp. 581–603. Chapman and Hall/CRC, (2017)
- Linial, N., Meshulam, R.: Homological connectivity of random 2-complexes. *Combinatorica* **26**(4), 475–487 (2006)
- Malen, G.: Collapsibility of random clique complexes. [arXiv:1903.05055](https://arxiv.org/abs/1903.05055), (2019)
- Newman, A.: One-sided sharp thresholds for homology of random flag complexes. [arXiv:2108.04299](https://arxiv.org/abs/2108.04299), (2021)
- Paquette, E., Werf, A.V.: Random geometric graphs and the spherical Wishart matrix. [arXiv:2110.10785](https://arxiv.org/abs/2110.10785), (2021)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.