On Edge Collapse of Random Simplicial Complexes

Jean-Daniel Boissonnat
Université Côte d'Azur, INRIA, Sophia Antipolis, France

Kunal Dutta
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Soumik Dutta
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Siddharth Pritam
Chennai Mathematical Institute, India

Abstract

We consider the edge collapse (introduced in [Boissonnat, Pritam. SoCG 2020]) process on the Erdős-Rényi random clique complex $X(n, c/\sqrt{n})$ on n vertices with edge probability $c/\sqrt{n}$ such that $c > \sqrt{\eta_2}$ where $\eta_2 = \inf\{\eta | x = e^{-\eta(1-x)^2}$ has a solution in $(0,1)\}$. For a given $c > \sqrt{\eta_2}$, we show that after t iterations of maximal edge collapsing phases, the remaining subcomplex, or *t-core*, has at most $(1+o(1))\binom{n}{2}p(1-(c^2/3)(1-(1-\gamma_t)^3))$ and at least $(1+o(1))\binom{n}{2}p(1-\gamma_{t+1}-c^2\gamma_t(1-\gamma_t)^2)$ edges asymptotically almost surely (a.a.s.), where $\{\gamma_t\}_{t\geq 0}$ is recursively determined by $\gamma_{t+1} = e^{-c^2(1-\gamma_t)^2}$ and $\gamma_0 = 0$. We also determine the upper and lower bound on the final core with explicit formulas. If $c < \sqrt{\eta_2}$ then we show that the final core contains $o(n\sqrt{n})$ edges. On the other hand, if, instead of c being a constant with respect to n, $c > \sqrt{2\log n}$ then the edge collapse process is no more effective in reducing the size of the complex.

Our proof is based on the notion of local weak convergence [Aldous, Steele. In Probability on discrete structures. Springer, 2004] together with two new components. Firstly, we identify the critical combinatorial structures that control the outcome of the edge collapse process. By controlling the expected number of these structures during the edge collapse process we establish a.a.s. bounds on the size of the core. We also give a new concentration inequality for typically Lipschitz functions on random graphs which improves on the bound of [Warnke. Combinatorics, Probability and Computing, 2016] and is, therefore, of independent interest. The proof of our lower bound is via the recursive technique of [Linial and Peled. A Journey Through Discrete Mathematics. 2017] to simulate cycles in infinite trees. These are the first theoretical results proved for edge collapses on random (or non-random) simplicial complexes.

2012 ACM Subject Classification Mathematics of computing; Theory of computation; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Computational Topology, Topological Data Analysis, Strong Collapse, Simple Collapse, Persistent homology, Random simplicial complexes

Digital Object Identifier 10.4230/LIPIcs.SoCG.2024.21

Related Version *Full Version*: https://www.researchgate.net/publication/379119064_On_Edge_Collapse_of_Random_Simplicial_Complexes

Supplementary Material
Software: https://github.com/soumikdt/Edge-Collapse-On-Random-Clique-Complex
archived at swh:1:dir:6e303142c2e9f3be976d75075fa445284a43999a

Funding J.-D. Boissonnat and S. Pritam received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007- 2013) / ERC Grant Agreement No. 339025 GUDHI (Algorithmic Foundations of Geometry Understanding in Higher Dimensions). K. Dutta and S. Dutta received funding from the Polish NCN SONATA Grant no. 2019/35/D/ST6/04525.

40th International Symposium on Computational Geometry (SoCG 2024).
Editors: Wolfgang Mulzer and Jeff M. Phillips; Article No. 21; pp. 21:1–21:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Simple collapse is a combinatorial process that simplifies a simplicial complex without changing its topology. It can be expressed as a series of elementary moves of removals of pairs of simplices σ and τ, such that σ is uniquely contained in τ. The notion of simple collapse was introduced by J.H.C Whitehead [29] to study homotopy types of cell complexes. Since then it has found usage in many different areas of topology, especially in computational topology. Recently new variants of simple collapses have been introduced, called strong collapses and more generally d-collapses [6, 8, 5]. In such collapses, one removes special vertices (more generally d-simplices) called dominated vertices (d-simplices) whose link is a *simplicial cone*. This can again be expressed as a series of simple collapses. They have been shown to be very powerful tools to solve many problems in computational topology. In particular, the recent studies [9, 8, 31] have shown that strong collapses and edge collapses (d-collapse for $d = 1$) can be used for efficient computation of one parameter as well as multi-parameter persistence homology (PH). Efficient computation of persistent homology is one of the central topics of research in topological data analysis (TDA).

The computation of persistent homology involves computing homology groups of a nested sequence of simplicial complexes called *filtrations*. This requires $O(n^3)$ time and $O(n^2)$ space, where n is the total number of simplices in the filtration. The general technique developed in [9, 7, 8, 11, 31] is to reduce a filtration to a smaller filtration using strong or edge collapse such that the persistent homology is preserved. In [9, 7, 8, 11, 31], it has been established through experiments that in practice the reduced filtrations are very small, and thereafter computation of persistent homology is extremely fast. The gain in efficiency is quite dramatic in the case of the *flag (or clique) complexes*, where the strong collapse and edge collapse can be computed using only the graph (1-skeleton) of the given complex [7, 11, 31]. One of the commonly used simplicial complexes in TDA is the Vietoris–Rips complex which is a flag complex.

As mentioned above the observed success of the procedure reported in [9, 7, 11, 31] is experimental and there is no theoretical guarantee over the size reduction. This is because in general the amount of reduction depends on the individual complex and its combinatorial structure. In fact, the reduction is dependent even on the order of the collapses and a different order can result in a different core, except in the case of the strong collapse. This makes it even harder to analyze when we reduce in a filtered simplicial complex. This motivates us to consider the case of random simplicial complexes and study the average reduction size by collapses.

Besides intrinsic mathematical interest, the study is motivated by the need for a theoretical explanation of the experimental success of [30, 7, 8]. These results also provide an average case analysis for edge collapse when no extra information about the data set is provided. With more structural information about the data set (which is often the case in practice), there is a greater reduction in size.

Even a constant fraction reduction in the number of edges translates to a much greater reduction in the size of the simplicial complex and hence in the PH computation time since the number of d-simplices is typically $O(m^{d-2})$ (m= # edges). Usual experimental gain is by a constant factor of 2-3 in magnitude. Secondly, in experiments, the dataset and the Vietoris-Rips complex (=clique complex) have additional geometric structure, which further helps the collapse process.

Edge collapse accelerates PH computation. Observed experimental and practical successes have led to its deployment in software packages like GUDHI. It performs several simple collapses in one step, so is more efficient than simple collapse. In flag complexes, it can be performed using only the underlying graph structure, avoiding the need to store the entire complex. Ours is an average-case analysis to explain the observed experimental reductions and provides a starting point for more specific theoretical bounds.

In this article, we study the problem of reduction size achieved by the edge collapses of a clique complex defined over an Erdős-Rényi random graph. This algorithm can be applied to real datasets to effectively reduce its size while preserving homology. More details can be found in [8].

Previous Work

The study of random simplicial complexes was initiated in the seminal paper of Linial and Meshulam [19]. Later Meshulam and Wallach [22] generalized the model of random complexes to obtain the *Linial-Meshulam* (LM) model of d-dimensional random complexes. Since then a large body of work from several authors has emerged on many different models of random simplicial complexes, studying various topological and geometric properties of such complexes [15]. Random clique complex was introduced in [14]. The study of simple collapses for random simplicial complexes has also been of interest to researchers and there have been numerous works in this direction. In the d-dimensional LM model, Kozlov [18] proved bounds on the threshold for vanishing of the d-th homology. Simple collapses on random complexes were first studied by Aronshtam, Linial, Łuczak, and Meshulam [4], who improved Kozlov's bound to get a tight bound on the threshold and also gave a bound on the threshold for collapsibility in the d-dimensional LM model. Later Aronshtam and Linial [2, 3] extended this line of work, obtaining first the threshold for the vanishing of the d-th homology in [2] and then the threshold for non-collapsibility of the d-dimensional LM complex [3]. In [20] Linial and Peled obtained precise asymptotic bounds on the size of the core of such complexes. Recently, Malen [21] has shown that the ER clique complex $X(n, p)$ is $(k+1)$-collapsible with high probability for $p = n^{-\alpha}$ when $\alpha > 1/(k+1)$. This result was improved in [16] by establishes asymptotic limits on Betti numbers of $X(n, p)$ with $p = \left(\frac{c}{n}\right)^{\frac{1}{k+1}}$ with local weak convergence of the X in different probability regimes. Their work generalizes to any random simplicial complex satisfying homogeneity and spatial independence properties. [24] discusses d-collapsibility of $X \sim X(n, \frac{c}{\sqrt[d]{n}})$ for $c < \frac{1}{2^{2d+1}d}$. They have shown that the pure d-dimensional part of the core consists of face disjoint copies of $(d+1)$-dimensional cross polytope boundary. Note that this $(d+1)$-dimensional cross polytope boundary is the minimal supporting structure for non-trivial dth homology for flag complex. In particular, setting $d = 2$ gives the structure of the core which is the concern of the current paper. But it should be noted that, in this case, this result holds for $c < \frac{1}{5 \cdot 2^5} = 0.00625$ while our results hold for any constant $c > \sqrt{\eta_2} \approx 1.57$.

Thus, to the best of our knowledge, work on collapses in random complexes has so far considered only simple collapses. We add that it is not too hard to show that the core from the strong collapse of $X \sim X(n, c/n)$ is essentially the 2-core of the corresponding Erdős-Rényi (ER) random graph. (The k-core of a graph G is the largest induced subgraph of G where each vertex has degree at least k. The size of the k-core for random graphs has been well studied [25, 13, 23].)

2 Our Contribution

We give bounds on the size of the t-*core* (i.e. the remaining complex after t phases of edge collapse as described in Algorithm 1) of a random simplicial complex in terms of the number of edges remaining. We show that for n-vertex Erdős-Rényi (ER) clique complexes with edge probability lying in a certain regime, the size of the t-core is a.a.s. a constant fraction of already existing edges, with the constant depending only on the edge probability, and bounded away from 1. Further, we also find a precise expression for the constant. Our first theorem is stated below.

For a simplicial complex K, we use $f_i(K)$ to denote the set of $i-$simplices of the complex. In a *pruning phase* run on K, a maximal subset of all the dominated (edge-collapsible, see Section 3) edges of K are collapsed by collapsing them one by one. Let $R_t(K)$ denote a complex obtained by running t pruning phases over K, i.e., t-core of K. The probability space of clique complexes on $G \sim G(n,p)$ will be referred to as the *Erdos Renyi Clique Complexes* and denoted as $X(n,p)$. Throughout this paper, we shall use the notation asymptotically almost surely (a.a.s.) for a series of events $(E_n)_{n\geq 1}$, when the probability of occurrence of E_n goes to 1 as $n \to \infty$. Define $\eta_2 = \inf\{\eta | x = e^{-\eta(1-x)^2} \text{has a solution in } (0,1)\}$. Note that η_2 is well-defined because $x = e^{-2.5(1-x)^2}$ has a solution in $(0,1)$. In particular, $\eta_2 \approx 2.45$.

▶ **Theorem 1.** *Let $X \sim X(n,p)$ such that $p = c/\sqrt{n}$ with $c > \sqrt{\eta_2}$. Let $\{\gamma_t\}_{t\geq 0}$ be defined recursively by $\gamma_{t+1} = e^{-c^2(1-\gamma_t)^2}$ and $\gamma_0 = 0$. Fix $t \leq \log n/(8 \log\log n) - 1$. Then with probability greater than $(1 - O(1/n^{11/2}))$*

$$(1 + o(1))\binom{n}{2}p(1 - \gamma_{t+1} - c^2\gamma_t(1 - \gamma_t)^2) \leq |f_1(R_t(X))|, \tag{1}$$

$$|f_1(R_t(X))| \leq (1 + o(1))\binom{n}{2}p(1 - (c^2/3)(1 - (1 - \gamma_t)^3)). \tag{2}$$

▶ Remark 2. For $X \sim X(n,p)$, let $P(X)$ denote the number of dominated edge-dominating vertex pairs. It can be shown that $\mathbb{E}[P(X)] = O(n^3p^3\left(1 - p^2(1-p)\right)^{(n-3)}) = O(n^3p^3 \exp(-np^2(1-p)))$. Now if $p \geq \sqrt{2\log(n)/n}$ then $\mathbb{E}[P(X)] = o_n(1)$. Thus, by Markov's inequality, with high probability, there is no dominated edge to start the collapsing procedure. On the other hand if $p < \sqrt{\eta_2/n}$ then it can be shown that the expected number of surviving edges after t pruning phases is $(O(1/2^t) + o(1))n\sqrt{n}$. Hence for $t = \theta(\log(\log(n)))$ the expected number of 2-simplices in the core becomes $o(n\sqrt{n})$. So we focus on the regime where $p = c/\sqrt{n}$ with $c > \sqrt{\eta_2}$.

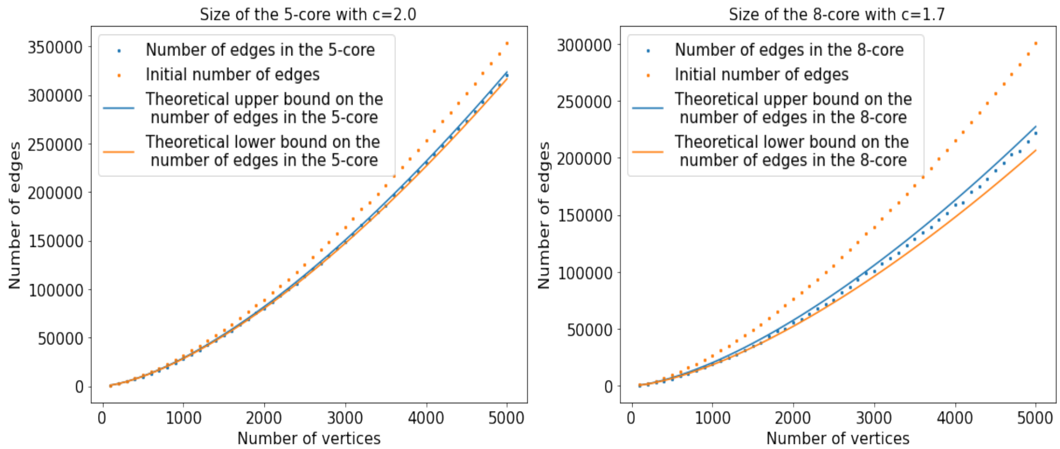
We ran experiments on ER clique complexes $X(n, c/\sqrt{n})$ and plotted the size of the t-core against the number of vertices $n \in [100, 5000]$ for different values of the parameters c and t. These experiments (Figure 1) clearly validate our theoretical results. In a different set of experiments (Fig. 2a) over ER clique complexes $X(n, c/\sqrt{n})$, we varied the constant $c \in [1.6, 2.5]$ keeping the number of vertices fixed to $n = 1500$. Python codes for these experiments can be found at https://github.com/soumikdt/Edge-Collapse-On-Random-Clique-Complex.

We also provide an upper and lower bound on the size of the final core, also referred to as the core, obtained by applying Algorithm 1 with t large enough so that no dominated edge is left at the end.

▶ **Theorem 3.** *Define $\eta_2 = \inf\{\eta | x = e^{-\eta(1-x)^2} \text{has a solution in } (0,1)\}$. Let $X \sim X(n,p)$ with $p = c/\sqrt{n}$ and $c > \sqrt{\eta_2}$. Let γ be the smallest positive solution of the equation $x = e^{-c^2(1-x)^2}$. Then a.a.s.*

$$|f_1(R_\infty(X))| \leq (1 + o(1))\binom{n}{2}(c/\sqrt{n})(1 - (c^2/3)(1 - (1 - \gamma)^3)), \tag{3}$$

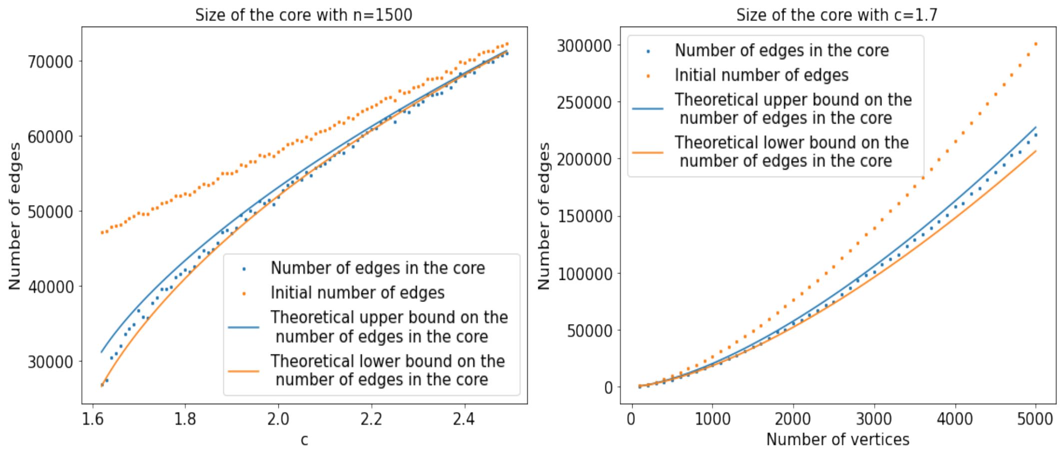
$$(1 + o(1))\binom{n}{2}p(1 - \gamma - c^2\gamma(1 - \gamma)^2) \leq |f_1(R_\infty(X))|. \tag{4}$$



(a) Size of the 5-core with $c = 2.0$.

(b) Size of the 8-core with $c = 1.7$.

Figure 1 Figure 1a shows the number of edges in the 5-core of $X \sim X(n, 2/\sqrt{n})$ along with the initial number of edges. The solid line indicates the theoretical upper and lower bounds on the expected number of edges. Figure 1b shows the same thing for the 8-core of $X \sim X(n, 1.7/\sqrt{n})$.



(a) Size of the core plotted against c.

(b) Upper and lower bound of the size of the core.

Figure 2 In Figure 2a, we plotted the number of edges in the core of $X \sim X(1500, c/\sqrt{1500})$ for $c \in [1.6, 2.5]$ along with the initial number of edges. The solid lines indicate the theoretical upper and lower bounds on the expected number of edges. Note that these bounds tend to coincide as c increases as shown in the figure. In Figure 2b, we ran the pruning procedure until there were no more dominated edges. The figure shows the number of edges in the final core of $X \sim X(n, 1.7/\sqrt{n})$ along with the initial number of edges. The solid line indicates the theoretical upper bound on the expected number of edges.

We also ran experiments (Figure 2b) on the size of the final core over ER clique complexes $X(n, c/\sqrt{n})$. We varied the number of vertices $n \in [100, 5000]$ keeping the constant fixed to $c = 1.7$.

Next, we prove the abstract version of the concentration inequality that we use in the subsequent sections. Fix n to be the number of vertices of a graph G, $p < 0.5$, and set $m = \frac{n(n-1)}{2}$. We order every pair of vertices G in an arbitrary manner. Let e_i denote the i-th such pair. We set $e_i = 1$ if the corresponding edge exists and $e_i = 0$ otherwise. For

$1 \leq i \leq m$, let $e_i \sim Bernoulli(p)$ be i.i.d. random variables. Clearly the Erdős-Rényi random graph $G(n,p)$ is equivalent to $X(n,p) = e_1 \times \cdots \times e_m$ as probability spaces. We shall use these notions interchangeably. Define $e_i(G)$ be the graph obtained after *flipping* the value e_i in G, i.e., if the potential edge e_i does not exist in G it exists in $e_i(G)$ and vice versa. Let $C_G(H)$ denote the number of copies of H in G as a subgraph.

▶ **Theorem 4.** *Assume notations as above. Let $\mathcal{G}$ be the set of labeled graphs on n vertices and H be an arbitrary but fixed graph on less than n vertices. Let $f : \mathcal{G} \to \mathbb{R}$ be a function that satisfy the following modified Lipschitz conditions:*

1. *If $C_G(H) = 0$ and $C_{e_i(G)}(H) = 0$ $\forall i \in \{1, \cdots, m\}$ then $|\Delta_{e_i}(f(G))| := |f(G) - f(e_i(G))| \leq C$ $\forall i$.*
2. *Otherwise $|\Delta_{e_i}(f(G))| \leq D$ $\forall i \in \{1, \cdots, m\}$.*

Now let $G \sim G(n,p)$ with $p \leq 1/2$ and $B := \mathbb{E}[C_G(H)]$. Then for all $T > B$

$$Pr\{|f - \mathbb{E}[f]| > s\} \leq 2 \exp\left(-\frac{s^2}{2(C + DT/p)(mpC + mDT + s/3)}\right) + B/T. \tag{5}$$

▶ Remark 5. [28] as well as several earlier papers [17, 27] have used stopped martingales to obtain concentration bounds on functions that are Lipschitz "on average". The idea is that at each step of the martingale process, we choose to continue if the function in the current point in the space, behaves "nicely" i.e. it is Lipschitz, or stop, if the function shows non-Lipschitz behavior. Thus the stopped martingale process works over a Lipschitz function and therefore provides a strong tail bound. However, one still needs to compute the probability that the stopped process differs from the original. To bound this, previous works have mainly used the union bound over the probability of stopping at each step.

Our approach differs from the previous approaches in that we choose to continue or stop the martingale based on the expected Lipschitz (or non-Lipschitz) behavior of the function at each step. We use a coupled martingale process which stops if and only if the original process stops, and control the probability of the martingale stopping, using the Optional Stopping Theorem. These ideas will be made precise in Section 6. This provides a sharper bound on the probability of ever stopping. To compare, in the same setting as Theorem 4, the inequality proved in [28] will have the following form

$$Pr\{|f - \mathbb{E}[f]| > s\} \leq 2 \exp\left(-\frac{s^2}{2(C + DT/p)(mpC + mDT + s/3)}\right) + mB/T. \tag{6}$$

In general, the m in the non-exponential part of the bound would be replaced by the length of the martingale. Thus when the length of the martingale is large the non-exponential part of the bound often makes the inequality ineffective. On the other hand, if one tries to improve the non-exponential part by increasing T then that increment in T kills the exponential part.

While our analysis of the edge collapse shares the general flow of the analysis of the simple collapsibility of LM complexes in e.g. [4, 2, 3, 20], there are several differences and difficulties. Firstly, our goal is to find the size of the t-core in a non-homogeneous complex (i.e., the maximal simplices of the complex do not have any fixed dimension, also known as non-pure complex), and for that, we need to define the local neighborhood of our graph appropriately. We only need edges that are susceptible to edge collapse in our local neighborhood rather than all possible edges. Secondly, as we are running the collapsing procedure in phases, it should be noted that all the edges that are dominated at the start of a phase do not get collapsed at that phase. Thus it is not enough to calculate the probability that some edge gets dominated in a certain phase to calculate the expected size of the t-core. We need to consider the higher-order structures in the complex to get proper estimates. Also, our analysis should accommodate the non-homogeneity of the complex. That is, maximal simplices in this model can have different sizes. Further, unlike in the LM model, the existence of a maximal simplex is not independent of the existence of all other possible maximal simplices. Finally, perhaps the most interesting difference of the random ER clique complex model in our context, is that the effect of collapsing an edge is not necessarily localized – a fact that requires a fair bit of innovation to handle (in several places).

Outline of the sections. Proof of Theorem 1 is done in two steps. First, Similar to [3], we show that a certain tree-like model of random simplicial complexes provides a good approximation of local neighborhoods, which is done in Section 4. This provides the bounds on the expected size of the t-core as shown in Section 5. Then we prove that the size of the t-core is concentrated around its expectation with probability in Section 7 using Theorem 4 which we discuss in Section 6. Theorem 3 is discussed in Section 8 and Section 9.

3 Preliminaries

In this section, we briefly introduce some topological and probabilistic notions. Readers can refer to [12] for a comprehensive introduction to topics related to topology and [10] for topics related to probability theory and random structures.

An **abstract simplicial complex** K is a collection of subsets of a non-empty finite set X such that the collection is closed under taking subsets of its elements. More formally, (i) $K \subset 2^X$ and (ii) $A \in K$ and $B \subset A$ implies $B \in K$.

An element of K of cardinality $d+1$ is called a ***d*-dimensional simplex** or ***d*-face** in short. The set of d-dimensional simplices of K will be denoted by $f_d(K)$. Elements of $f_0(K)$ and $f_1(K)$ are often referred to as vertices and edges respectively. If $\sigma, \tau \in K$ such that $\sigma \subset \tau$ then we say τ is a **coface** of σ. A simplex is called **maximal** if it is *not* a proper subset of any other simplex in K. A sub-collection L of K is called a **subcomplex** if it is a simplicial complex itself. A subcomplex K' of K is called a ***d*-skeleton** of K if it contains all the simplices of K of dimension at most d.

Dominated vertex and edge. A vertex $u \in K$ is dominated by another vertex $v \in K$ if all the maximal simplices of K that contain u also contain v. Similarly, an edge $e := \{u, w\} \in f_1(K)$ is dominated by a vertex $v \in f_0(K) \setminus \{u, w\}$ if all the maximal simplices of K that contain e also contain v.

Edge collapse. Given a complex K, the action of removing a dominated edge e and all its cofaces from K is called a **edge collapse** [8], denoted as $K \searrow\searrow^1 \{K \setminus e\}$. If after a series of edge collapses on a simplicial complex K produces another complex L then we write $K \searrow\searrow^1 L$. We further call a complex K **edge collapse minimal** if it does not have any dominated edge. A subcomplex L of K is called a **core** under edge collapse if $K \searrow\searrow^1 L$ and L is edge collapse minimal. A core of a complex K under edge collapse is not unique. Like simple collapses, edge collapses preserve the homotopy type of a simplicial complex.

For an abstract simplicial complex K and $\sigma \in K$, define link of σ as $Lk(\sigma) := \{\tau \in K \mid \tau \cap \sigma = \emptyset, \tau \cup \sigma \in K\}$.

Now we present the algorithms to find the t-core of simplicial complexes under edge collapse.

Algorithm 1 Compute t-core under edge collapse.

Data: Simplicial complex X
Result: $R_t(X)$, i.e., t-core of X under edge collapse
Calculate the set D of all the dominated edges of the current complex;
Shuffle elements of D uniformly randomly;
for e *in* D **do**
  If e is dominated by some vertex in the current complex collapse e and all its cofaces.
end
Repeat the above steps t times.

▶ Remark 6. Note that all the edges present in the set D in the above algorithm can *not* be removed at once because deleting one edge may change the dominated status of its neighboring edges as shown in Figure 3. This explains that all dominatd edges in a pruning phase can not be collapsed simultaneously. We need to use some fixed but arbitrary order of deletion. The resulting complex depends on this order. While random ordering is not necessary to perform this collapse, we use it in our analysis. Using a non-random ordering gives rise to a different process.

Clique complex. A **clique** or a **flag** complex of a graph G is a simplicial complex where each clique of size k in G makes a $k-1$-face of K. Note that if K is a flag complex and L is any subcomplex of K obtained by edge collapses then L is also a flag complex. In this paper, our main focus will be to analyze the effect of the Algorithm 1 on Erdős-Rényi Clique Complexes $X(n, p)$.

4 Random 2-tree and local 2-neighborhood structure

In this section, we shall first describe the process of building random 2-trees recursively which will be useful later. This construction is similar to the construction of Galton-Watson trees. It will be proved that for $e \in f_1(X)$, where $X \sim X(n, c/\sqrt{n})$ with $c > \sqrt{\eta_2}$, the local neighborhood of the edge e looks like a random 2-tree. More formally, the random simplicial complex exhibits a local weak convergence [1] to the Poisson 2-tree with parameter c^2.

▶ **Definition 7.** *A* 2*-tree is built recursively as follows: (i) Start with a single edge(root). (ii) Each* 1*-simplex added in the* $i-1$*st iteration shall have* m *children* 2*-simplices in the* i*th iteration where* $m \sim Poisson(c')$ *with* $c' \in \mathbb{R}_+$*. In other words, we draw a random variable with Poisson distribution with parameter* c' *for each* 1*-simplex added in the last iteration and add that many* 2*-simplices to it in the current iteration.*

Let $\mathcal{T}_n$ *denote the set of all possible rooted trees after* n*th iteration for* $n \geq 1$ *and* $\mathcal{T}_0$ *being the root itself. We also define the infinite Poisson tree* T_∞ *as the resulting tree when the number of iterations becomes unbounded. Define* $\mathcal{T} := \bigcup_{n \in \mathbb{N}} \mathcal{T}_n$*.*

Let γ_t be the probability that a tree $T \in \mathcal{T}_t$ is pruned to the root in *no more than* $t-1$ steps. Note that, in this process, we never prune the root itself even if its degree is 1. We call such a process *root collapsing*. Clearly, $\gamma_1 = e^{-c'}$. Set $\gamma_0 = 0$. Observe that the root survives the $t+1$st pruning phase if at least one of the children of the root survives the t-th pruning phase. In other words, to prune a rooted edge, all its children 2-*simplices* must be pruned, i.e. at least one edge from *each* child 2-simplex must be collapsed. This probability for one child 2-simplex is $(1-(1-\gamma_t)^2)$. Thus, we have the following recursive relation which is true in general:

$$\gamma_{t+1} = 1 - \sum_{k=0}^{\infty}(1-(1-(1-\gamma_t)^2)^k)\frac{c'^k e^{-c'}}{k!} = e^{-c'(1-\gamma_t)^2}.$$

Note that γ_t depends on c' and t. Sometimes we shall indicate this fact by writing $\gamma_t(c')$ instead of γ_t to avoid any confusion. Define $\gamma \equiv \gamma(c') := \lim_{t\to\infty}\gamma_t$. Define $\beta_t := 1-\gamma_{t+1}$. Thus, β_t is the probability that *at least* $t+1$ root collapsing steps are needed to isolate the root of a tree $T \in \mathcal{T}_t$.

▶ **Lemma 8.** *For $c > \eta_2$, we have $\gamma - \gamma_t \le \frac{e^{-c}}{1-f'(\gamma)}(f'(\gamma))^t$ and $f'(\gamma) < 1$.*

The last lemma provides an estimate of the rate of convergence of the sequence $\{\gamma_t\}_{t\ge 0}$. It is to be noted that, as $f'(\gamma) < 1$, the sequence converges to γ exponentially fast with respect to t. We shall leverage this fact while calculating the upper bound on the size of the final core.

Now we shall analyze the properties of the local neighborhood of an edge e in $X \sim X(n, c/\sqrt{n})$. Due to the probabilistic nature of the complex, the properties we consider are not absolute, rather, as we shall prove, they hold with high probability. As we are working on the edge collapse in non-homogeneous simplicial complexes, there are simplices neighboring e that do not have any effect on the edge collapsing procedure. Thus, for an effective analysis, we need to craft the definition of local neighborhoods carefully. Note that, for a flag complex K, an edge $e \in K$ is dominated by another vertex $v \in K$, if and only if all the vertices in K that have an edge with both vertices of e also have an edge with v. This suggests that we should only consider those vertices as neighbors that are adjacent to both the vertices of e. So we adopt the following definition of the neighborhood for our purpose. Note that $Cl(X)$ denotes the simplicial closure of X.

▶ **Definition 9.** *Let $e = \{u, v\} \in X$. Define $f_0(\mathcal{N}(e)) := \{w \in f_0(X) | e \cup \{w\} \in X\}$. Let $X[f_0(\mathcal{N}(e))]$ be the induced subgraph on the vertex set $f_0(\mathcal{N}(e))$. We define the neighborhood complex of e as $\mathcal{N}_1(e) := Cl(X[f_0(\mathcal{N}(e))])$ which is the clique complex build on $X[f_0(\mathcal{N}(e))]$. Also define $deg(e) := |\{w \in X | \{u, v, w\} \in f_2(X)\}|$.*

Next, we extend the previous definition to define a neighborhood of arbitrary depth. we start with a fixed edge and then we define the neighborhood recursively. At each recursive step, we add the neighborhood of already existing edges.

▶ **Definition 10.** *As before, fix $e = \{u, v\} \in X$. We recursively define the t-neighborhood of the edge e to be $\mathcal{N}_t(e) := Cl(X[\bigcup_{h\in f_1(\mathcal{N}_{t-1}(e))}\{u \in f_0(X) | h \cup \{u\} \in X\}])$ and $\mathcal{N}_0(e) := \{\{u\}, \{v\}, \{u, v\}\}$. We shall write $\mathcal{N}_t(e)$ as 2-neighborhood of e of depth t.*

First, we shall show that the degree of an edge, defined earlier, is bounded by $O(\log n)$ with high probability.

▶ **Lemma 11.** *Define event $E := \{deg(e) \le \log n \;\; \forall e \in f_1(X)\}$. Then $Pr\{E\} = 1 - o_n(1)$.*

We now show that the neighborhood is a 2-tree with high probability.

▶ **Lemma 12.** *Let $X \sim X(n, c/\sqrt{n})$ and $e \in f_1(X)$. Then, for $t \le \log n/(8\log\log n) - 1$, $Pr\{\{\mathcal{N}_t \in \mathcal{T}\} \cap E\} = 1 - O(n^{-1/4})$.*

The degree of an edge of this 2-tree comes from $Bin(n-2, c^2/n)$. For large n this distribution can be approximated by $Poisson(c^2)$. In short, the local neighborhood is a random 2-tree from the Poisson distribution with parameter c^2.

Now we shall briefly recall the notion of local weak convergence [1]. We shall modify the framework a little to suit our purpose. An edge-rooted (possibly infinite) simplicial complex (X, e) is a simplicial complex X with a distinguished edge $e \in f_1(X)$ called root of X. Let $\mathcal{X}_\star$ be the set of all such edge-rooted complexes. Under proper metric, $\mathcal{X}_\star$ becomes a complete and separable metric space that has its Borel σ-algebra (denoted by Σ). So one can define weak convergence of probability measures in this space.

Given a finite simplicial complex X one can then describe a probability measure on $\mathcal{X}_\star$ by choosing a root of X uniformly randomly from $f_1(X)$. Let us call this measure $X[U]$. For a sequence $(X_n)_{n\in\mathbb{N}}$ of finite simplicial complexes, we say that $X_n \to X_\infty$ locally weakly in $(\mathcal{X}_\star, \Sigma)$ if $X_n[U] \to X_\infty[P]$ where $P : f_1(X_\infty) \to \mathbb{R}_+ \cup \{0\}$ is some probability measure on the 1-simplices of X_∞. In other words, if we fix any vertex of X_n to be the root uniformly randomly and calculate the neighborhood with respect to that root then that neighborhood will converge (under the distance metric introduced in the last paragraph) to X_∞ where the root in X_∞ is chosen using probability measure P on the vertices of X_∞. Under the above framework, the previous lemma implies that $X(n, c/\sqrt{n}) \to T_\infty(c^2)$ locally weakly. Thus for any bounded continuous function f

$$\mathbb{E}[f(X(n, c/\sqrt{n}))] \to \mathbb{E}[f(T_\infty(c^2))].$$

5 Expected size of the t-core

As stated earlier edge collapses will be carried out in phases (also referred to as pruning phases). In each phase, a maximal set of the existing dominated edges shall be collapsed. Recall that, in a flag complex K, an edge $e = \{u, w\}$ is dominated by a vertex v if and only if all the common neighbors of u and w are also a neighbor of v.

As the complex locally looks like a random 2-tree, an edge survives the t-th collapsing phases if the neighborhood 2-tree is not collapsed to its root edge until the t-th phase. Due to the tree structure collapsing an edge in a 2-tree is effectively like removing a 2-simplex from the tree. t-core is achieved after t many pruning phases. In each phase, we mark all the dominated edges and randomly collapse a maximal set from the marked edges. Let us first calculate the lower bound on the size of the t-core.

Lower bound on the expected size of the t-core. To estimate the lower bound we shall calculate the total number of unmarked edges after t phases. Note that an edge $e \in f_1(X)$ unmarked after k pruning phases only if it had degree $deg(e) \geq 2$ after $k-1$ pruning phases. As probability of such an event is $(1 - \gamma_t - c^2\gamma_t(1 - \gamma_{t-1})^2)$, using Lemma 12, we get

$$\mathbb{E}[f_1(R_t(X))] \geq (1 - O(n^{-1/4}))\binom{n}{2}p(1 - \gamma_{t+1} - c^2\gamma_t(1 - \gamma_t)^2).$$

Upper bound on the expected size of the t-core. Calculating the upper bound is less straightforward because, in any particular phase, not all dominated edges get collapsed. For example, consider Figure 3.

Even if all the edges of an isolated triangle are dominated in any particular phase only one can be collapsed. Because after collapsing one edge rest of the edges are not dominated anymore. Thus not all the marked edges are collapsed. To get around this issue we observe the following lemma. We say a 2-simplex is marked if any of its edges are marked. By maximal 2-simplex we mean a 2-simplex which is also a maximal simplex.

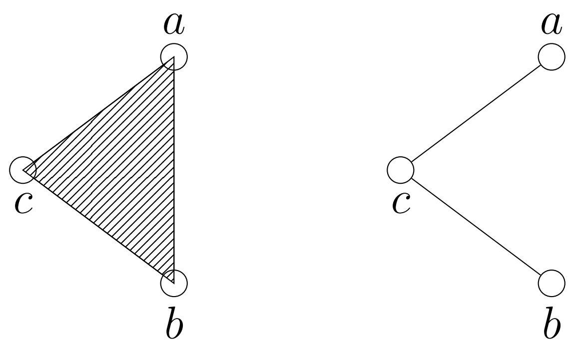


Figure 3 In the left complex each edge is dominated by the vertex opposite to it. In particular $\{a, b\}$ is dominated by c. So we collapse this edge to the figure on the right. But in the figure on the right, as we can check, the edges $\{c, b\}$ and $\{a, c\}$ are not dominated anymore.

▶ **Lemma 13.** *Any marked maximal 2-simplex σ gets collapsed in the same pruning phase it got marked.*

It can be shown that the expected number of non-maximal 2-simplices is $O(n)$. So the expected number of marked maximal 2-simplices is essentially the same as the expected number of marked 2-simplices. So, we can use marked 2-simplices to count collapsed 2-simplices. Let d_i denote the number of *collapsed* i-simplices during the t pruning phases. We shall focus on the expected values of d_1 and d_2. It can be shown that:

▶ **Lemma 14.** $\mathbb{E}[d_2] \geq (1 + o(1))(1 - (1 - \gamma_t)^3)\binom{n}{3}(c/\sqrt{n})^3$.

Now to calculate a lower bound on the expected number of collapsed edges, i.e. d_1, we need the following combinatorial observations.

▶ **Lemma 15.** *At any stage of the pruning phases, if collapsing one dominated edge removes k 2-simplices then at least $k - 1$ 3-simplices are also removed due to this.*

Proof. Let us assume that the edge $\{u, w\}$ is part of k 2-simplices, namely, $\{u, w, v\}$ and $\{u, w, u_i\}$ for $1 \leq i \leq k - 1$. Also assume, without loss of generality, that v dominates $\{u, w\}$. Then the edges $\{v, u_i\}$ exist for $1 \leq i \leq k - 1$. Thus the arrangement shown in Figure 4 is necessary for the event that collapsing one edge ($\{u, w\}$ in the figure) removes k many 2-simplices.

Therefore, removing $\{u, w\}$ kills the $\{u, v, w, u_i\}$ for $1 \leq i \leq k - 1$. So, at least $k - 1$ 3-simplices are removed. ◀

▶ **Corollary 16.** *let d_1, d_2 and d_3 as defined before. Then $d_2 \leq d_3 + d_1$.*

Proof. Let $d_{2,i}$ and $d_{3,i}$ denote the number of collapsed 2-simplices and 3-simplices respectively for ith edge collapse during the pruning phases. By Lemma 15, we have $d_{3,i} \geq d_{2,i} - 1$. Thus $d_3 = \sum_{i=1}^{d_1} d_{3,i} \geq \sum_{i=1}^{d_1} d_{2,i} - \sum_{i=1}^{d_1} 1 = d_2 - d_1$. ◀

Hence, $d_2 \leq d_1 + d_3 \leq d_1 + f_3(X)$. After taking expectation we get

$$\mathbb{E}[d_1] \geq \mathbb{E}[d_2] - \mathbb{E}[f_3(X)] \geq \mathbb{E}[d_2] - (1 + o(1))(c^6/24)n \geq (1 + o(1))\mathbb{E}[d_2].$$

So we get the desired upper bound. But

$$\mathbb{E}[f_1(R_t(X))] \leq \mathbb{E}[f_1(R_0(X))] - \mathbb{E}[d_1] \leq (1 + o(1))\binom{n}{2}(c/\sqrt{n})(1 - (c^2/3)(1 - (1 - \gamma_t)^3)).$$

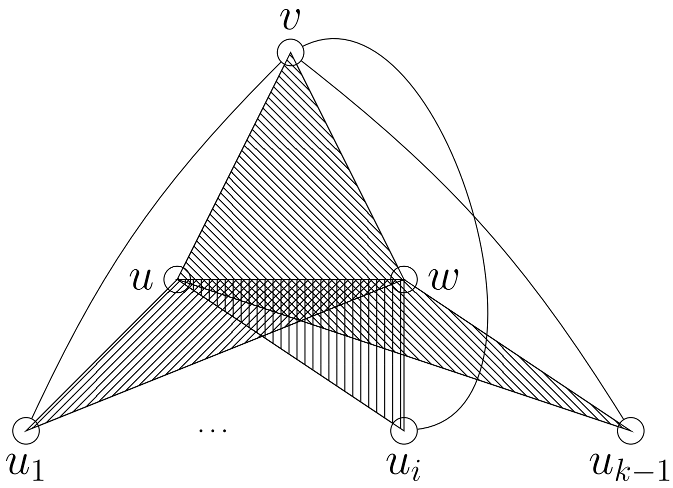


Figure 4 Observe that the edge $\{u, w\}$ is part of k 2-simplices, namely, $\{u, w, v\}$ and $\{u, w, u_i\}$ for $1 \leq i \leq k-1$. Also, $\{u, w\}$ is part of $k-1$ 3-simplices, namely, $\{v, u, w, u_i\}$ for $1 \leq i \leq k-1$.

Thus we get the following theorem:

▶ **Theorem 17.** *Define* $\eta_2 = \inf\{\eta | x = e^{-\eta(1-x)^2}$ *has a solution in* $(0,1)\}$. *Let* $X \sim X(n, c/\sqrt{n})$ *with* $c > \sqrt{\eta_2}$. *Let* $\mathbb{E}(|f_1(R_t(X))|)$ *denote the number of edges in* X *after* t *edge collapse phases and* γ_t *is as defined above. Then,*

$$(1+o(1))\binom{n}{2}p(1-\gamma_{t+1}-c^2\gamma_t(1-\gamma_t)^2) \leq \mathbb{E}(|f_1(R_t(X))|),$$

$$\mathbb{E}(|f_1(R_t(X))|) \leq (1+o(1))\binom{n}{2}p(1-(c^2/3)(1-(1-\gamma_t)^3)).$$

6 The concentration inequality

We use the martingale method to prove Theorem 4. In fact, we use two coupled martingales defined on the same filtration of σ-algebras such that one martingale stops if and only if the other martingale also stops. One of the martingales is to control the deviation of the concerned function f from its expected value. The purpose of the other martingale is to control the probability of stopping. The stopping time would be the moment when the conditional expected number of the subgraph H of G exceeds some certain threshold. This stopping probability can be estimated using the Optional Stopping Theorem.

7 Concentration of the size of the t-core

As we are working on the Erdős-Rényi clique complexes, the effect of edge collapse is not localized only in the neighborhoods of the collapsed edges throughout the pruning phases. Collapse of a dominated edge can affect a large region of the flag complex. But the occurrence of such events implies the existence of a specific class of subcomplexes, which we call *Critical Complexes*. It is easy to show that the probability of occurrence of these subcomplexes is vanishingly low in the original random complex. Thus the effect of edge collapse is *typically* bounded. This observation indicates that we can use Theorem 4 to prove the concentration of the size of the t-core. In this section, we shall prove the following concentration result using the theorem developed in the last section.

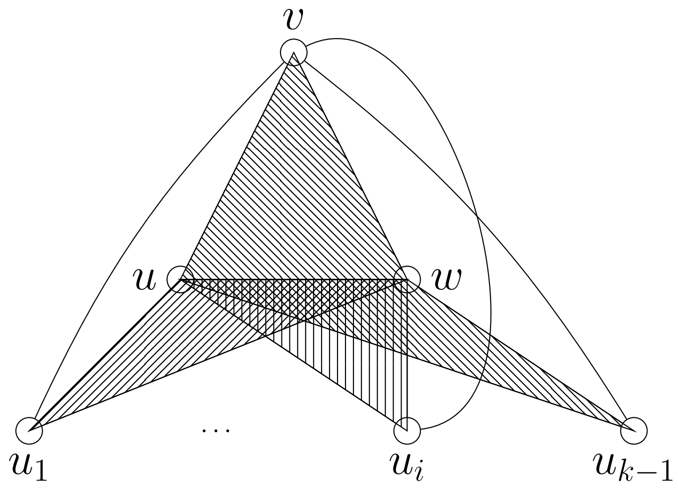


Figure 5 Critical Complex of order k.

▶ **Theorem 18.** *Let $X \sim X(n,p)$ with $p = c/\sqrt{n}$ for some constant c. Let $|f_1(R_t(X))|$ be the number of edges after t edge collapsing phases and $Y_0 = \mathbb{E}(|f_1(R_t(X))|)$ be its expected value. Then for any $s \geq 0$, we have,*

$$Pr\{||f_1(R_t(X))| - Y_0| \geq s \cdot n\} \leq 2\exp(-s^2\Theta(\sqrt{n})) + O(1/n^{5.5}).$$

We first prove some combinatorial results that will be useful to construct the 'bad' subgraph H of G, i.e., the 1-skeleton of a critical complex, so that we can use the concentration theorem from the last section.

▶ **Lemma 19.** *$Pr\{$Collapsing a dominated edge e gives birth to $2k$ newly generated dominated edges for $k \geq 2\} \leq O(\frac{1}{(\sqrt{n})^{k-4}})$*

Proof. We shall use the fact that for a general simplicial complex K: An edge $e \in K$ is dominated by another vertex $v \in K$, if and only if all the maximal simplices of K that contain e also contain v. Let $e, e' \in K$. Let v dominate $e := \{u, w\}$ and only after collapsing of e, e' gets dominated by v'. Assume that $v \neq v'$. Then there is a maximal simplex σ that contains both e and e' but $v' \notin \sigma$. First, assume that e and e_1 do not share any common vertex. So, $e_1 \subset \sigma \cap Lk(e)$. Collapsing e breaks the maximal simplex σ into two maximal simplices consisting of at least 3 vertices each, namely, $Cl((\sigma \cap Lk(e)) \cup \{u\})$ and $Cl((\sigma \cap Lk(e)) \cup \{w\})$, both containing e_1. Then $v' \in Lk(e) \cap \sigma \subset \sigma$ which is a contradiction. Thus e_1 must be of the form $\{u, u_1\}$ or $\{w, u_1\}$, i.e., $e_1 \subset e \cup \{u_1\}$ for some $u_1 \in K$. The case $v = v'$ is not possible as there can not be any maximal simplex σ_1 such that $e, e' \in \sigma_1$ but $v \notin \sigma_1$.

Now if $u_1 \neq v$ then the edge $\{u_1, v\}$ must exist. Thus the arrangement shown in Figure 5 is necessary for the case that collapsing e generates $2k - 2$ newly dominated edges. Including the case that $u_1 = v$ we get that Figure 5 is necessary for the event that collapsing e generates $2k$ newly dominated edges. We call the complex shown in Figure 5 a *Crtitical Complex of order k*.

It requires $k + 2$ vertices and $3k$ edges. Expected number is $O(\frac{1}{(\sqrt{n})^{k-4}})$. Hence the result follows from the Markov's inequality. ◀

Proof of Theorem 18. We shall use the notations introduced in Theorem 4. Fix H to be the 1-skeleton of the critical complex (as shown in Figure 5) of order $k = 20$. It has 22 vertices and 60 edges. So $B = \mathbb{E}[C_G(H)] = O(1/n^8)$. Flipping one edge in the initial complex will have a cascading effect on the consequent t pruning phases. But as H is forbidden, at the end of the t phases at most $\frac{39^{t+1}-1}{38} \leq 40^t$ edges would be flipped. On the other hand, if we do not put any constraint on the structure of the initial complex then at the end of the pruning phases all edges can be flipped. Thus we get $C = 40^t$ and $D = n^2$. Then by Theorem 4 we have

$$Pr\{||f_1(R_t(X))| - Y_0| \geq s \cdot n\} \leq 2\exp(-\frac{s^2 \cdot n^2}{2(40^t + n^{2.5}T/c)(n^{1.5} + n^4 T + sn/3)}) + O(1/n^8 T).$$

Setting $T = 1/n^{2.5}$ we get

$$Pr\{||f_1(R_t(X))| - Y_0| \geq s \cdot n\} \leq 2\exp(-s^2\Theta(\sqrt{n})) + O(1/n^{5.5}).$$ ◀

Proof of Theorem 1. Using Theorem 17 and Theorem 18 we get the desired result. ◀

8 Upper bound on the size of the core

In this section, we shall study the effect of a maximal number of edge-collapsing phases on random simplicial complexes. In other words, we keep running the pruning phases until there is no more dominated edge and observe the size of the core obtained in this way. Even though letting $t \to \infty$ in Equation 2 produces Equation 4 proving so is less straightforward simply because our initial assumption about tree-like neighborhood no longer is valid for arbitrary t. But Lemma 12 provides a bound on the largest t that still keeps our previous analysis valid. Fortunately, this bound is decent enough for our purpose.

Idea behind the proof of Theorem 3. First we observe that, for a fixed flag complex X, $|f_1(R_t(X))| \geq |f_1(R_{t+1}(X))|$. Thus $|f_1(R_t(X))| \geq |f_1(R_\infty(X))|$. Thus for any $t \geq 1$, $|f_1(R_t(X))|$ is an upper bound of the size of the final core. In particular, we shall take $t = \rho := \log n/16 \log\log n$. It can be shown that a.a.s. $(1+o(1))\binom{n}{2}(c/\sqrt{n})(1-(c^2/3)(1-(1-\gamma)^3)) \geq (|f_1(R_\rho(X))|$. Then the result follows immediately as $|f_1(R_\rho(X))| \geq |f_1(R_\infty(X))|$. ◀

9 Lower Bound on the size of the core

The proof for the lower bound on the size of the core follows a different approach, adapted from [20, 26]. The idea is, roughly speaking, to craft two properties, $\mathcal{P}$ and $\mathcal{A}$, defined on the space of rooted random 2-trees. The property $\mathcal{A}$ says that a 2-tree can not be pruned within a fixed finite number of steps. The property $\mathcal{P}$ implies that sufficiently lower descendant leaves of the root of the tree have property $\mathcal{A}$. Then one shows that property $\mathcal{A}$ implies property $\mathcal{P}$ with high probability. This recursive nature implies that 2-trees with property $\mathcal{A}$ can not be pruned in *any* finite number of steps. Thus calculating the expected number of trees with property $\mathcal{A}$ yields the following theorem.

▶ **Theorem 20.** *Let $X \sim X(n, c/\sqrt{n})$ with $c > \sqrt{\eta_2}$. Then a.a.s*

$$f_1(R_\infty(X)) \geq (1+o(1))(1-\gamma-c^2\gamma(1-\gamma)^2)c/\sqrt{n}\binom{n}{2}.$$

References

1 David Aldous and J Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

2 Lior Aronshtam and Nathan Linial. When does the top homology of a random simplicial complex vanish? *Random Struct. Algorithms*, 46(1):26–35, 2015. doi:10.1002/rsa.20495.

3 Lior Aronshtam and Nathan Linial. The threshold for d-collapsibility in random complexes. *Random Struct. Algorithms*, 48(2):260–269, 2016. doi:10.1002/rsa.20585.

4 Lior Aronshtam, Nathan Linial, Tomasz Luczak, and Roy Meshulam. Collapsibility and vanishing of top homology in random simplicial complexes. *Discret. Comput. Geom.*, 49(2):317–334, 2013. doi:10.1007/s00454-012-9483-8.

5 Dominique Attali, André Lieutier, and David Salinas. Vietoris-rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, 46(4):448–465, 2013.

6 J. A. Barmak and E. G. Minian. Strong homotopy types, nerves and collapses. *Discrete and Computational Geometry*, 47:301–328, 2012.

7 J-D. Boissonnat and S. Pritam. Computing persistent homology of flag complexes via strong collapses. *International Symposium on Computational Geometry (SoCG)*, 2019.

8 J-D. Boissonnat and S. Pritam. Edge collapse and persistence of flag complexes. *International Symposium on Computational Geometry (SoCG)*, 2020.

9 J-D. Boissonnat, S.Pritam, and D. Pareek. Strong Collapse for Persistence. In *26th Annual European Symposium on Algorithms (ESA 2018)*, volume 112, 2018.

10 Béla Bollobás. *Random graphs*. Number 73 in Cambridge studies in advanced mathematics. Cambridge University Press, 2 edition, 2001.

11 Marc Glisse and Siddharth Pritam. Swap, Shift and Trim to Edge Collapse a Filtration. In *38th International Symposium on Computational Geometry (SoCG 2022)*, volume 224, pages 44:1–44:15, 2022.

12 A. Hatcher. *Algebraic Topology.* Univ. Press Cambridge, 2001.

13 Svante Janson and Malwina J Luczak. A simple solution to the k-core problem. *Random Structures & Algorithms*, 30(1-2):50–62, 2007.

14 Matthew Kahle. Topology of random clique complexes. *Discrete mathematics*, 309(6):1658–1671, 2009.

15 Matthew Kahle. Random simplicial complexes, 2016. arXiv:1607.07069.

16 Shu Kanazawa. Law of large numbers for betti numbers of homogeneous and spatially independent random simplicial complexes. *Random Structures & Algorithms*, 60(1):68–105, 2022.

17 Jeong Han Kim and Van H Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

18 Dmitry N. Kozlov. The threshold function for vanishing of the top homology group of random d-complexes. *Proceedings of the American Mathematical Society*, 138(12):4517–4527, 2010. URL: http://www.jstor.org/stable/41059187.

19 Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Comb.*, 26(4):475–487, 2006. doi:10.1007/s00493-006-0027-9.

20 Nathan Linial and Yuval Peled. Random simplicial complexes: around the phase transition. *A Journey Through Discrete Mathematics*, pages 543–570, 2017.

21 Greg Malen. Collapsibility of random clique complexes. *Discrete Mathematics*, 346(3):113267, 2023. doi:10.1016/j.disc.2022.113267.

22 Roy Meshulam and N. Wallach. Homological connectivity of random k-dimensional complexes. *Random Struct. Algorithms*, 34(3):408–417, 2009. doi:10.1002/rsa.20238.

23 Michael Molloy. Cores in random hypergraphs and boolean formulas. *Random Structures & Algorithms*, 27(1):124–135, 2005.

24 Andrew Newman. One-sided sharp thresholds for homology of random flag complexes. *Journal of the London Mathematical Society*, 109(3):e12872, 2024.

25 Boris Pittel, Joel Spencer, and Nicholas Wormald. Sudden emergence of a giantk-core in a random graph. *Journal of Combinatorial Theory, Series B*, 67(1):111–151, 1996.

26 Oliver Riordan. The k-core and branching processes. *Combinatorics, Probability and Computing*, 17(1):111–136, 2008.

27 Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012.

28 Lutz Warnke. On the method of typical bounded differences. *Combinatorics, Probability and Computing*, 25(2):269–299, 2016. doi:10.1017/S0963548315000103.

29 J. H. C Whitehead. Simplicial spaces nuclei and m-groups. *Proc. London Math. Soc*, 45:243–327, 1939.

30 Adam C. Wilkerson, Harish Chintakunta, and Hamid Krim. Computing persistent features in big data: A distributed dimension reduction approach. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11–15, 2014. doi:10.1109/ICASSP.2014.6853548.

31 Siddharth Pritam Ángel Javier Alonso, Michael Kerber. Filtration-Domination in Bifiltered Graphs. In *SIAM Symposium on Algorithm Engineering and Experiments (ALENEX23)*, 2023.