Adv. Appl. Probab. **57,** 940–968 (2025)
*doi:*10.1017/apr.2024.66

THE ASYMPTOTICS OF THE EXPECTED BETTI NUMBERS OF PREFERENTIAL ATTACHMENT CLIQUE COMPLEXES

CHUNYIN SIU,* ** *Cornell University and Stanford University*
GENNADY SAMORODNITSKY,* ***
CHRISTINA LEE YU,* **** AND
RONGYI HE,* ***** *Cornell University*

Abstract

The preferential attachment model is a natural and popular random graph model for a growing network that contains very well-connected 'hubs'. We study the higher-order connectivity of such a network by investigating the topological properties of its clique complex. We concentrate on the Betti numbers, a sequence of topological invariants of the complex related to the numbers of holes (equivalently, repeated connections) of different dimensions. We prove that the expected Betti numbers grow sublinearly fast, with the trivial exceptions of those at dimensions 0 and 1. Our result also shows that preferential attachment graphs undergo infinitely many phase transitions within the parameter regime where the limiting degree distribution has an infinite variance. Regarding higher-order connectivity, our result shows that preferential attachment favors higher-order connectivity. We illustrate our theoretical results with simulations.

Keywords: Topological data analysis; complex network; phase transition

2010 Mathematics Subject Classification: Primary 05C82, 60C05, 05E45
Secondary 55U10, 55N31, 62R40

1. Introduction

1.1. Preferential attachment graphs

In [2], the preferential attachment model was proposed to explain the emergence of well-connected 'hub' nodes and of a power-law degree sequence in growing networks. Since then, many variants have been proposed.

Received 17 August 2023; accepted 25 October 2024.
* Postal address: The Center for Applied Mathematics, 657 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA.
** Postal address: Brain Dynamics Lab, 1520 Page Mill Road, Palo Alto, CA 94304, USA. Email address: siuc@stanford.edu
*** Postal address: The School of Operations Research and Information Engineering, 206 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA. Email address: gs18@cornell.edu
**** Postal address: The School of Operations Research and Information Engineering, 206 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA. Email address: cleeyu@cornell.edu
***** Postal address: The School of Operations Research and Information Engineering, 206 Rhodes Hall, Cornell University, Ithaca, NY 14853, USA. Email address: rh463@cornell.edu

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

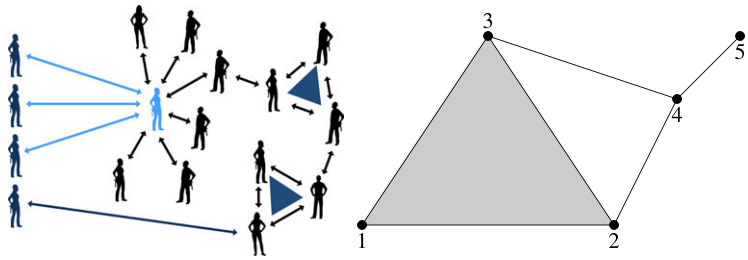

FIGURE 1. Left: An illustration of the preferential attachment mechanism (cf. Equation (1) and Definition 2.1) and the clique-building mechanism (cf. Definitions 2.3 and 1.5). When new nodes (drawn as people) in the left column are added to the network, they are more likely to attach to already popular nodes (which have high degrees), like the light blue person in the figure. Fully connected subsets of nodes form triangles, tetrahedra, or their higher-dimensional analogues in the clique complex. Note that in order to have triangles, each new node must connect to at least two nodes, but we draw only one connection for each new node to keep the illustration simple. Right: An illustration of a simplicial complex X whose simplices are $\{1, 2, 3\}$, $\{2, 4\}$, $\{3, 4\}$, $\{4, 5\}$ and their nonempty subsets. Its homology groups are as follows: $H_0(X) \cong H_1(X) \cong \mathbb{Z}$ and $H_q(X) \cong 0$ for $q \notin \{0, 1\}$. The generator of $H_1(X)$ can be represented by the cycle $[2, 3] + [3, 4] - [2, 4]$.

These graphs are typically built inductively. At each discrete time step, a new node is added to the graph and it is randomly connected to m existing nodes, where m is a positive integer. The key common feature of preferential attachment graphs is that

$$\mathbb{P}_{\text{current}}(\text{a new node connects to a node } v) \propto f(\text{degree of } v) \tag{1}$$

for some increasing function f, where $\mathbb{P}_{\text{current}}$ denotes the conditional probability given the current graph.

When $f(k) = k + \delta$ with $\delta > -m$, such variants of preferential attachment graphs are called *affine*. We precisely define the affine variant we consider in this paper in Definition 2.1, where we spell out the initial graph and the dependency between the m edges. An illustration of the preferential attachment mechanism is shown in the left panel of Figure 1.

The constant δ quantifies the strength of preferential attachment. The smaller (or more negative) it is, the stronger the preferential attachment effect, i.e. the more likely new nodes attach to nodes with large degrees. This gives rise to giant 'hubs' with large degrees.

This phenomenon manifests quantitatively as a heavier tail in the degree sequence. For instance, for the variant affine preferential attachment graphs in Section 8.2 of [43], Theorem 8.3 therein states that the (random) proportion $p_T(k)$ of nodes with degree k in the graph with T nodes is approximately $c_{m,\delta} k^{-(3+\delta/m)}$ for some constant $c_{m,\delta} > 0$ for large k and T. More precisely, let $p_k = (2 + \delta/m)\frac{\Gamma(k+\delta)\Gamma(m+2+\delta+\delta/m)}{\Gamma(m+\delta)\Gamma(k+3+\delta+\delta/m)}$. Then $p_k = c_{m,\delta} k^{-(3+\delta/m)}(1 + O(1/k))$. The theorem states that, for some constant $C_{m,\delta} > 0$,

$$\lim_{T\to\infty} \mathbb{P}\left(\max_k |P_T(k) - p_k| \geq C_{m,\delta}\sqrt{\frac{\log T}{T}}\right) = 0.$$

See also [9, 33], and Lemma 5.9 of [44] for similar results for other variants.

In particular, the graph undergoes a *phase transition* when δ becomes negative, as the limiting variance of the degree distribution becomes infinite for $\delta \in (-m, 0)$.

Besides their degree distributions, there has been a lot of interest in the higher-order connectivity of preferential attachment graphs. Their *clustering coefficients* (number of pairwise

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

connected triples divided by a suitable normalization constant) have been well studied, and they often vanish asymptotically unless the graphs are specifically engineered to behave otherwise (see, for instance, [8, 15, 25, 36, 37]). In [20], the growth rates, often sublinear, of the expected counts of *small motifs* (subgraphs isomorphic to a given finite graph) in preferential attachment graphs were determined. In [5, 12], some forms of preferential attachment *simplicial complexes* were considered, and power laws for some forms of higher-dimensional degrees were found.

For details about preferential attachment graphs, we refer the reader to [43, 44].

1.2. Higher-order connectivity and algebraic topology

A simplicial complex may be seen as a hypergraph where each subset of a hyperedge is also a hyperedge of the hypergraph. This closure condition enables us to investigate the higher-order connectivity of this object using tools from algebraic topology.

Formally, we have the following definition.

Definition 1.1. (Simplicial complex [34, Section 3].) *A finite simplicial complex X is a collection of nonempty subsets of a finite set that is closed under inclusion; i.e., for nonempty subsets σ, τ of X, if $\tau \subseteq \sigma$ and $\sigma \in X$, then $\tau \in X$.*

Elements of this collection are called simplices. Subsets of a simplex are called faces of the simplex. The dimension of a simplex is one less than the number of elements. Simplices of dimensions 0, 1, 2, and 3 are called vertices, edges, triangles, and tetrahedra, respectively.

Remark 1.2. Simplicial complexes as defined above are called abstract simplicial complexes in [34], to distinguish them from their geometric realizations, which are simply called 'simplicial complexes' there. In this paper, all simplicial complexes are abstract and we do not concern ourselves with their geometric realizations.

We think of two- and three-dimensional simplices as triangles and tetrahedra. A geometric illustration is shown in the right panel of Figure 1.

The higher-order connectivity of a simplicial complex can be measured by its *Betti numbers*, which generalize component counts and cycle counts to the counts of repeated connection (or equivalently, higher-dimensional holes). They have proven to be useful statistics in recent *topological-data-analytic* applications [1, 10, 11]. For instance, it has been observed that the way holes emerge in (biological) neural networks helps distinguish different stimuli to the brains of different animals [39]. We present the formal definition below.

Definition 1.3. (Homology group, Betti number, cycle, and boundary [34, Section 5].) *Let X be a finite simplicial complex. Impose a total ordering on the vertices.*

Denote by $[x_0, \ldots, x_q]$ the q-dimensional simplex $\sigma = \{x_0, \ldots, x_q\}$ with $x_0 < \ldots < x_q$. For each nonnegative integer q, let $C_q(X)$ be the free abelian group generated by q-dimensional simplices of X, and let $\partial_q : C_q \to C_{q-1}$ be the homomorphism defined by

$$\partial_q[x_0, \ldots, x_q] = \sum_{0 \le i \le q} (-1)^i [x_0, \ldots, \hat{x}_i, \ldots, x_q],$$

where the hat means removal (e.g. $[x_0, \hat{x}_1, x_2] = [x_0, x_2]$). The homology group $H_q(X)$ of X at dimension q is the quotient group

$$H_q(X) = \ker \partial_q / \operatorname{im} \partial_{q+1},$$

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

and its rank

$$\beta_q(X) = \mathrm{rk}\, H_q(X)$$

is called the Betti number of X at dimension q. Elements of $H_q(X)$, $\ker \partial_q$, $\mathrm{im}\, \partial_{q+1}$, *and* $C_q(X)$ *are called, respectively, homology classes, cycles, boundaries, and chains of dimension q.*

Remark 1.4. The homology groups, and hence the Betti numbers, are independent of the choice of ordering of the vertices. The quotient in the definition of $H_q(X)$ is well-defined: one can verify that $\partial_q \partial_{q+1} = 0$, and hence $\mathrm{im}\, \partial_{q+1} \subseteq \ker \partial_q$.

The Betti number at dimension 0 is simply the number of connected components. The Betti number at a higher dimension is often interpreted as the number of holes of that dimension. Equivalently, it is also the number of repeated connections, since a hole can be decomposed into two sides that 'connect' or 'fill up' their intersection, e.g. a loop can be formed by two arcs with the same endpoints, and a sphere can be formed by two hemispheres that share the same equator.

To utilize these concepts from algebraic topology for the study of random graphs, we consider the *clique complex* of a graph, whose simplices are fully-connected cliques.

Definition 1.5. (Clique complex.) *Let G be an undirected finite graph without loops and repeated edges, and let V be its set of vertices. The clique complex associated with G is*

$$\{\sigma \subseteq V : \text{the induced subgraph of } V \text{ on } \sigma \text{ is a nonempty complete graph}\}.$$

A finite simplicial complex K is said to be a clique complex if it is the clique complex of its underlying graph, i.e. the graph with the same vertex set and edge set as K.

1.3. Literature review on random simplicial complexes

The literature on *random simplicial complexes*, especially on their Betti numbers, has been growing rapidly in recent years. The Betti numbers of Erdös–Rényi clique complexes were found to be supported at a critical dimension with high probability in [27, 29]. This result can be seen as a generalization of the classical result on the phase transition of Erdös–Rényi graphs [16]. For readers interested in percolation theory, the emergence of *giant cycles*, homological analogues of giant components, on the torus (a lattice with periodic boundary condition) was established in [14].

A clique complex of a graph may be regarded as a (radius-1) *Rips complex* with respect to the graph metric. Rips complexes are geometric clique complexes where nearby points are connected. They have gained substantial attention because of their applications in topological data analysis [6]. Limit theorems for the Betti numbers of random Rips complexes constructed from independent and identically distributed points were established in [28, 31]. They are generalizations of connectivity results for random geometric graphs in [38].

For a more comprehensive overview of the literature on general random simplicial complexes, we refer the reader to [7, 30].

Regarding the preferential attachment mechanism, to the best of our knowledge, there have been no prior analytical results on the Betti numbers of preferential attachment simplicial complexes, with the notable exception of those of [35], which considers a two-dimensional model and determines the asymptotics of the Betti number at dimension 1. After the public release of our preprint, a central limit theorem was established for Betti numbers of a model of scale-free graphs in the parameter regime where the degree distribution has a finite limiting variance [24].

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

TABLE 1. Asymptotic notation

Notation	Definition
$f(n) = O(g(n))$	$f(n) \le Cg(n)$ for large enough n for some positive constant factor C.
$f(n) = \Omega(g(n))$	$f(n) \ge cg(n)$ for large enough n for some positive constant factor c.
$f(n) = \Theta(g(n))$	$cg(n) \le f(n) \le Cg(n)$ for large enough n for some positive constant factors C and c.
$f(n) = o(g(n))$	$\lim_{n\to\infty} f(n)/g(n) = 0$

1.4. Main result: the topology of preferential attachment clique complexes

Our main contribution is to determine the orders of magnitude of the expected Betti numbers of preferential attachment clique complexes at *all* dimensions.

We now set up the notation to state our main results.

We adopt the asymptotic notation in Table 1, where f and g are assumed to be nonnegative functions defined on $\mathbb{N}$.

Let

$$x(\delta, m) = 1 - \frac{1}{2 + \delta/m} \in (0, 1), \tag{2}$$

which is often denoted by χ in the literature on preferential attachment graphs, e.g. in [20]. We deviate from this convention because χ often denotes the Euler characteristic in algebraic topology.

When δ decreases (or becomes more negative), x decreases, and hence x decreases with the strength of the preferential attachment. The quantity x controls the rate at which the probability that a late node is attached to an early node converges to 0. This is made precise at Corollary 3.7.

Our main theorem is as follows.

Theorem 1.6. *Let $X(T, \delta, m)$ be the preferential attachment clique complex, which is precisely defined in Definition 2.3. Let $q \ge 2$ and suppose $m \ge 2q$. Then*

$$\mathbb{E}[\beta_q(X(T, \delta, m)] = \begin{cases} \Theta(T^{1-2qx(\delta,m)}) & \text{if } 1 - 2qx(\delta, m) > 0, \\ \Theta(\log T) & \text{if } 1 - 2qx(\delta, m) = 0, \\ O(1) & \text{otherwise,} \end{cases}$$

where $\mathbb{E}$ denotes expectation, and the big-O and big-Θ constants (see Table 1) depend only on q, δ, m.

One could keep track of the big-Θ constants in the proof. For instance, when $m = 7$, $\delta = -5$, $q = 2$, the big-Θ constants can be chosen to be $C = 2.16 \times 10^{14}$ and $c = 1.18 \times 10^{-34}$. Finding the optimal constants, however, is beyond the scope of this work.

The topological behavior at dimension 1 is different from the behavior at higher dimensions. The proof technique is different as well.

Proposition 1.7. *Let $X(T, \delta, m)$ be the preferential attachment clique complex. Then $\mathbb{E}[\beta_1(X(T, \delta, m))] = (m - 1)T + o(T)$.*

Next, we address the trivial cases left out by the results above.

Proposition 1.8. *Let $X(T, \delta, m)$ be the preferential attachment clique complex. Then the following hold:*

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

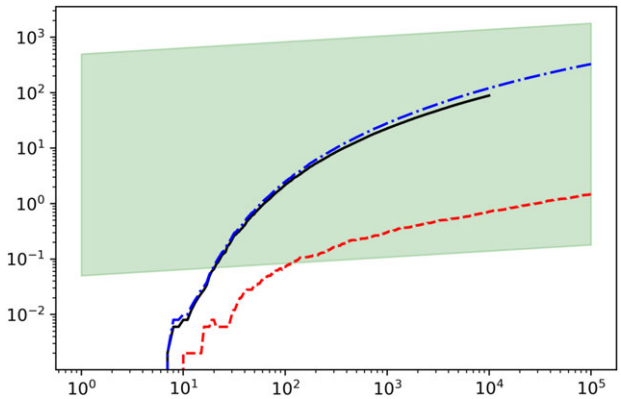

FIGURE 2. The log–log plot of the evolution of the mean Betti number at dimension 2 for 500 (synthetic) preferential attachment clique complexes. The horizontal axis is the number of nodes in log scale; the black curve corresponds to the mean Betti number, also in log scale. The dotted curves correspond to the mean upper and lower bounds in our argument (specifically in Proposition 4.1). The slope of the shaded region is the asymptotic growth rate of the expected Betti number. The position and the width of the shaded region are chosen post hoc manually, because the theoretical constants are too conservative.

- *For* $q = 0$, $\beta_0(X(T, \delta, m)) \equiv 1$.
- *For* $q \geq 2$ *and* $m < 2q$, $\beta_q(X(T, \delta, m)) \equiv 0$.

1.5. Simulation

In Figure 2, we illustrate the sublinear growth of the average Betti number at dimension 2 through a simulation. Our theorem and our choice of parameters dictate that the curve for the evolution of the expected Betti number is eventually contained in a band with slope 2/9, which is the slope of the shaded region. The evolution of our estimate of the expected Betti number is plotted as the solid curve, and it is plausible that it remains inside the shaded region when extrapolated indefinitely. We discuss the simulation in greater detail in Section 8. Codes for simulating Betti numbers of preferential attachment clique complexes are available at the GitHub repository `carolinerongyi/Preferential_Attachment_Clique_Complex`.

1.6. Discussion

Theorem 1.6 suggests that preferential attachment graphs undergo infinitely many phase transitions in the regime $\delta < 0$, where the limiting variance of the degree distribution is infinite. Indeed, for each dimension larger than 1, when the strength of preferential attachment exceeds a critical threshold, the expected Betti number at that dimension ceases to be bounded, and diverges to infinity as the number of nodes increases. The critical thresholds for the lower dimensions are illustrated in Figure 3.

Recall that Betti numbers can be interpreted as the number of higher-order repeated connections. Our theorem then implies that with a stronger strength of preferential attachment, there are more repeated connections of higher order (or equivalently, more holes of higher dimensions) in expectation. In other words, preferential attachment favors higher-order connectivity.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

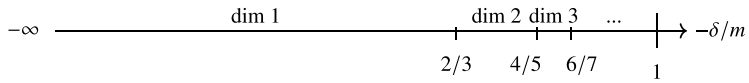

FIGURE 3. The top dimensions with unbounded expected Betti numbers for different values of $-\delta/m \in (-\infty, 1)$ for m not too small (recall that $-\delta/m$ increases with the strength of preferential attachment effect; see Theorem 1.6 for the precise condition on m). The critical thresholds for dimensions 2, 3, and 4 respectively are 2/3, 4/5, and 6/7.

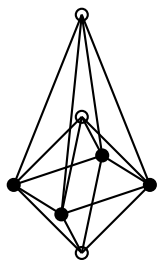

FIGURE 4. The graph Γ_3. All nodes marked by solid circles precede all nodes marked by hollow circles.

Our results suggest the following similarities between preferential attachment clique complexes, Erdös–Rényi clique complexes [31], and random Rips complexes [28, 47]:

- The major contributions to the Betti numbers are due to holes (homology classes) represented by small cycles (recall from Definition 1.3 that homology classes are equivalence classes of cycles).
- The expected Betti numbers have a dominating dimension, which is 1 in this case.

On the other hand, our complexes have two distinctive features:

- Unlike Erdös–Rényi clique complexes, there is a range of dimensions with positive expected Betti numbers, which grows with the number of nodes.
- Compelled by the preferential attachment mechanism, holes (homology classes) are predominantly represented by highly interconnected cycles.

An instance of a clique complex with such interconnected cycles at dimension 2 is illustrated in Figure 4. We discuss such cycles more in Section 4.

1.7. Paper outline

The rest of the paper is organized as follows. We present the precise definition of preferential attachment graphs and their clique complexes in Section 2. There, we also discuss some details of our set-up. In Section 3, we review the background materials for our proofs. Technical topological materials are deferred to Appendix B.

We begin proving the main theorem, Theorem 1.6, in Section 4 by presenting a proof synopsis and a decomposition result. We complete the proof of Theorem 1.6 in Section 5. In Section 6, we prove the edge cases of Propositions 1.7 and 1.8. We present and discuss simulation results in Section 8, and we discuss future directions in Section 9. We collect technical background materials in Appendices A and B.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

2. Set-up

Throughout this paper, we adopt the following notation for degrees in an undirected graph Γ without self-loops and with totally ordered nodes. Let v be a node in Γ.

- The degree of v is denoted by $d_\Gamma(v)$.
- The pre-degree (resp. post-degree) of a node v in a graph Γ, denoted by ${d^{\text{pre}}}_{\Gamma(v)}$ (resp. ${d^{\text{post}}}_\Gamma(v)$), is the number of nodes preceding (resp. preceded by) v that are connected to v. For example, if Γ has exactly two nodes 1, 2, and exactly one edge, then ${d^{\text{pre}}}_\Gamma(1) = 0$.
- Γ's sequence of degrees (resp. pre- or post-degrees) is the sequence whose kth term is the degree (resp. pre- or post-degree) of the kth node of Γ. This is not to be confused with the *degree sequence* (resp. pre- or post-degree sequence), whose kth term is the number of nodes whose degree (resp. pre- or post-degree) is k.

We now choose our variant of the preferential attachment model. We adopt the affine model in [20], since we will rely heavily on the subgraph count results therein.

Definition 2.1. (Affine preferential attachment graphs [20, Definition 1].) *Let T, m be positive integers with $T \geq 2$ and let $\delta \in (-m, \infty)$. The preferential attachment graph $G(T, \delta, m)$ is the random graph, with no self-loops but possibly with repeated edges, that is constructed inductively as follows:*

- *The graph $G(2, \delta, m)$ is the deterministic graph with two nodes, indexed by 1 and 2, and m edges between the two nodes.*
- *The graph $G(T, \delta, m)$ is constructed by adding a node, indexed by T, to $G(T-1, \delta, m)$ and adding m edges between node T and m sequentially and randomly chosen nodes in $G(T-1, m, \delta)$ from the following conditional distribution:*

$$\begin{aligned}&\mathbb{P}(\text{the } i\text{th edge of node } T \text{ connecting it with node } v | G(T-1, \delta, m, i-1))\\ &= \frac{1}{C(T, \delta, m, i)}(d_{G(T-1,\delta,m,i-1)}(v) + \delta)\end{aligned}$$

for $1 \leq i \leq m$, where $G(T-1, \delta, m, i-1)$ is the graph after adding $i-1$ edges between T and nodes in $G(T-1, \delta, m)$, and the normalization constant is $C(T, \delta, m, i) = 2(T-2)m + i - 1 + (T-1)\delta$.

Remark 2.2.

- In [20], the graph is said to be directed, with edges always pointing from later nodes to earlier nodes. However, we simply treat the graph as an undirected graph, because the edges' orientations can be inferred from the vertex labels.
- Regarding equivalent models in the literature, to the best of our knowledge, the case for $m = 1$ but with general δ was first considered in [33]. The definition above is equivalent to the model $PA_T^{(e)}(m, \delta)$ in Definition 4.3.1 of [19] and $PA_T^{m,\delta}(d)$ defined by (1.3.65) of [44], modulo typos in the latter [45]. For lists of other variants of the preferential attachment model, see [37] and Chapter 4.3 of [19].

We define the preferential attachment clique complex $X(T, \delta, m)$ as follows.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

Definition 2.3. (Preferential attachment clique complex.) *Let $G(T, \delta, m)$ be the preferential attachment graph, and let $G_{\text{simple}}(T, \delta, m)$ be the graph obtained by replacing all repeated edges in $G(T, \delta, m)$ with simple edges. The preferential attachment clique complex $X(T, \delta, m)$ is the clique complex of $G_{\text{simple}}(T, \delta, m)$ in the sense of Definition 1.5.*

We conclude this subsection by commenting on a few innocuous yet technical choices in our set-up.

First, regarding the treatment of multiple edges, we simply replace them by simple edges, because it is easier to define clique complexes for simple graphs. A notion of clique complex for multigraphs is defined in [3]. Our argument gives the same bound for this notion of clique complex upon slight modifications.

Second, to readers interested in *persistent homology*, we remark that our method gives estimates of the expected *persistent Betti numbers* as well for the filtration of node arrival (the complex at time t consists of the first t nodes), because our proof also shows that the death of homology classes occurs much less frequently than their birth. However, generalizing this to persistence diagrams is difficult because the number of points in a box in a persistence diagram is the difference of persistent Betti numbers, which have the same order of magnitude in this case. To keep the exposition simple, we do not further discuss persistence in the present paper.

Finally, regarding the choice of coefficients for the homology groups, we prove our homological results with coefficients in $\mathbb{Z}$. The same argument works for arbitrary field coefficients. Homological computations of our numerical simulations are done with coefficients in $\mathbb{Z}/2\mathbb{Z}$.

3. Preliminaries

In this section we recall results in the literature that are relevant for our proofs. We defer technical topological definitions and theorems to Appendix B.

First, we introduce a few simplicial complexes and subcomplexes in Section 3.1. In particular, as in [29], *spheres* and *links* will play a crucial role in our proofs.

Then we state two technical results. In Section 3.2, we state the subgraph count results in [20]. In Section 3.3, we generalize a result on minimal cycles due to [18, 27] to the setting of *relative homology* using an *exact sequence* argument. (Relative homology and exact sequence are defined in Appendix B.) Subgraph count results and the characterization of minimal cycles are typically crucial in the estimation of Betti numbers in the literature random simplicial complexes.

3.1. Instances of simplicial complexes

In this subsection, we define a few simplicial complexes and subcomplexes that will appear in our proofs. A subcomplex is a collection of simplices in a simplicial complex that itself forms a simplicial complex.

For an integer $q \geq 1$, the octahedral q-sphere S^q is the clique complex of the graph whose vertices are $1, \ldots, 2(q+1)$, and where i and j are connected by an edge if and only if $i - j \not\equiv 0 \bmod (q+1)$. It can be visualized in $\mathbb{R}^{q+1}$ as the ℓ^1 unit sphere by mapping vertices $1, \ldots, q+1$ to $e_1, \ldots, e_{q+1}$ and vertices $q+2, \ldots, 2(q+2)$ to $-e_1, \ldots, -e_{q+1}$, where the e_i are the standard basis of $\mathbb{R}^{q+1}$.

For an integer $q \geq 1$, the octahedral $(q+1)$-ball D^{q+1} is the clique complex formed by adding a vertex to S^q that is connected to all vertices of S^q by edges. It can be visualized in $\mathbb{R}^{q+1}$ as the closed ℓ^1 unit ball by mapping the new vertex to the origin.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

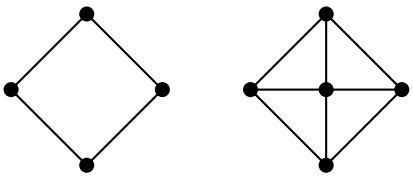

FIGURE 5. Illustrations of the underlying graphs of the clique complexes S^1 (left) and D^2 (right). The clique complex D^2 has four triangles, whereas S^1 has none.

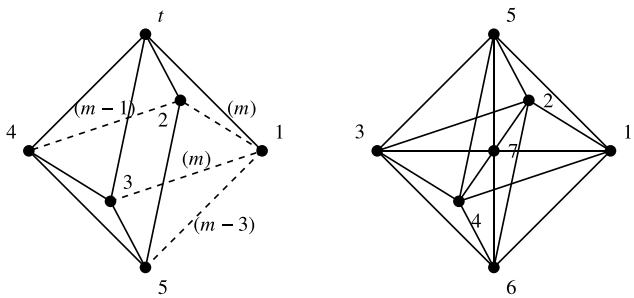

FIGURE 6. Illustrations of the underlying graphs of the clique complexes S^2 (left) and D^3 (right). The labels and the different line styles for the left illustration are for $\Gamma^{(t)}$ in the proof of Lemma 5.1, and those for the right illustration are for Example 4.2. Labels without parentheses denote node indices in $G(T, \delta, m)$, and labels in parentheses denote edge multiplicity of the dashed edges in $G(T, \delta, m)$.

For instance, S^1 and D^2, illustrated in Figure 5, are the unfilled and filled squares. The clique complexes S^2 and D^3 are illustrated in Figure 6.

The homology groups of the spheres S^n with $n \geq 1$ are as follows:

$$H_q(S^n) = \begin{cases} \mathbb{Z} & \text{if } q \in \{0, n\}, \\ 0 & \text{otherwise.} \end{cases}$$

Next we define some general subcomplexes. Let X be a simplicial complex.

The *star* of a vertex v in X, denoted by $\mathrm{St}_X(v)$, is the subcomplex of X consisting of all simplices containing v (and the faces of these simplices). (Our notion of star is called the *closed* star in [34, p. 11].)

The *link* of a vertex v in X, denoted by $\mathrm{Lk}_X(v)$, is the subcomplex of $\mathrm{St}_X(v)$ consisting of all simplices that do *not* contain v (cf. [34, p. 11]).

For instance, the link of the central vertex in D^2 is S^1.

The star of a vertex is an example of a *cone* complex. The cone CK of a simplicial complex K can be defined as follows: CK has all the vertices of K, as well as an extra vertex, denoted by v. For each simplex σ of K, let $v * \sigma = \{v\} \cup \sigma$. Then

$$CK = K \cup \{v * \sigma : \sigma \in \{\emptyset\} \cup K\}$$

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

(cf. Sections 8 and 62 of [34]). It can be readily verified that the star of a vertex is the cone of the link of the vertex (cf. the proof of Lemma 35.4 in [34]). The homology groups of cones are trivial.

Proposition 3.1. (Theorem 8.2 of [34].) *Cones are acyclic, i.e. all their homology groups are 0, except that H_0 is $\mathbb{Z}$. In particular, stars are acyclic.*

3.2. Preferential attachment graphs

Recall that preferential attachment graphs are defined in Definition 2.1, and we need certain subgraph count results, namely Theorem 3.4 and Proposition 3.6 below. We develop a few definitions to simplify their statements.

The preferential attachment graph $G(T, \delta, m)$ is a random subgraph of the underlying *attachment graph* $U(T, m)$, which we define as follows.

Definition 3.2. (Attachment graph.) *The (T, m)-attachment graph U(T, m) is the multigraph such that*

- *the vertex set is* $\{1, \ldots, T\}$*, and*
- *there are m edges between any pair of distinct nodes.*

The following definition will simplify the expression for the estimate in the theorem.

Definition 3.3 (Vertex power.) *Let* Γ *be a subgraph of U(T, m), and let v be a vertex in* Γ*. The power* $p_\Gamma(v; \delta, m)$ *of v is*

$$\begin{aligned} p_\Gamma(v; \delta, m) &= -\left[{d^{\text{pre}}}_\Gamma(v) + ({d^{\text{post}}}_\Gamma(v) - {d^{\text{pre}}}_\Gamma(v))(1 - x(\delta, m))\right] \\ &= -\left[(1 - x(\delta, m)){d^{\text{post}}}_\Gamma(v) + x(\delta, m){d^{\text{pre}}}_\Gamma(v)\right], \end{aligned}$$

where $x(\delta, m)$ *is defined in Equation (2).*

This expression is called the power because it appears in the exponents of the asymptotics of subgraph counts in the following two results.

Theorem 3.4. (Theorem 1 of [20].) *Consider the preferential attachment graph* $G(T, \delta, m)$*. Let* Γ *be a subgraph of U(T, m) with vertex set* $V_\Gamma = \{v_1 < \ldots < v_{|V_\Gamma|}\}$*, and with pre-degrees bounded above by m and degrees bounded below by 1. Then the expected count of subgraphs in* $G(T, \delta, m)$ *that are isomorphic to* Γ *is*

$$\Theta(T^A(\log T)^{r-1}),$$

where A is the maximum value of the sequence $a_0, \ldots, a_{|V_\Gamma|}$ *defined by*

$$a_k = |V_\Gamma| - k + \sum_{l>k} p_\Gamma(v_l; \delta, m), \tag{3}$$

and r is the number of maximizers. The big-Θ *constants are independent of T but do depend on m,* δ*, and the sequences of pre- and post-degrees of* Γ*.*

Remark 3.5. The original theorem is stated in terms of $\tau = 3 + \delta/m$ rather than δ, m, or $x(\delta, m)$; see the end of Section 2.1 of [20].

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

Theorem 3.4 can be proven from the following lemma, which we will need for the lower bound on the expected Betti numbers.

Proposition 3.6. (Lemma 1 of [20].) *Let Γ be a subgraph of $U(T, m)$ with vertex set $V_\Gamma \subseteq \{1, \ldots, T\}$ and with pre-degrees bounded above by m and degrees bounded below by 1. Then the probability that $G(T, \delta, m)$ contains Γ is*

$$\Theta\left(\prod_{v \in V_\Gamma} v^{p_\Gamma(v;\delta,m)}\right),$$

where the big-Θ constants depend only on δ, m, and the post-degree sequence of Γ.

We discuss the intuition for and formulation of these results further in Appendix A.

This proposition gives an interpretation of the quantity $x(\delta, m)$ as the order of magnitude of the probability that a late node attaches to an early node. More precisely, we have the following corollary. (See also Exercise 8.13 and Lemma 8.17 of [43] for an analogous theorem about a slightly different preferential attachment model.)

Corollary 3.7 (Proposition 3.2 of [41].) *Let $v \leq T$ be positive integers. Denote by $P(T \to v)$ the probability that node T is attached to node v in the preferential attachment graph via at least one edge. Then*

$$\mathbb{P}(T \to v) = \Theta\left(\frac{1}{v^{1-x(\delta,m)}T^{x(\delta,m)}}\right),$$

where the big-Θ constants depend only on m and δ. In particular, if v is treated as a constant, then

$$\mathbb{P}(T \to v) = \Theta\left(\frac{1}{T^{x(\delta,m)}}\right),$$

where the big-Θ constants depend only on m, δ and v.

Proof. Denote by $\mathbb{P}(T \xrightarrow{i} v)$ the the probability that T is attached to v via the ith edge of T. Then Proposition 3.6 implies that

$$\mathbb{P}(T \xrightarrow{i} v) = \Theta\left(\frac{1}{v^{1-x(\delta,m)}T^{x(\delta,m)}}\right).$$

The result then follows from the fact that $\mathbb{P}(T \xrightarrow{1} v) \leq \mathbb{P}(T \to v) \leq \sum_{1 \leq i \leq m} \mathbb{P}(T \xrightarrow{i} v)$. □

We will not use this corollary in this paper other than to give an interpretation of $x(\delta, m)$.

3.3. Minimal clique cycles

In this subsection, we state a result, namely Proposition 3.9, about minimal cycles in clique complexes, which will be needed to characterize the Γ_k's in the introduction as the dominating cycles. It is a slight generalization of Lemmas 5.2 and 5.3 in [27] and Lemma 2.1.4 in [18] to the context of relative homology (defined in Appendix B.2). Its proof, which is based on an exact sequence argument, is deferred to Section 7.

Definition 3.8 (Clique-minimal complex.) *Let X be a clique complex and A a subcomplex. For $q \geq 0$, X is said to be (A, q)-clique-minimal if for every clique subcomplex Y of X that contains A,*

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

$$\beta_q(Y, A) > 0 \text{ if and only if } X = Y.$$

If A is the empty set, we abbreviate (A, q)-clique-minimality as q-clique-minimality.

The expression $\beta_q(Y, A)$ is the Betti number of the pair (Y, A), which is defined in Appendix B.2.

For example, for $q > 0$, the q-dimensional octahedral sphere is q-clique-minimal. The q-dimensional octahedral sphere with an extra edge (i.e. the simplicial complex formed by adding an extra node to the q-dimensional octahedral sphere and an extra edge connecting the new node to one of the old nodes) is not q-clique-minimal, because the complex still has a positive Betti number at dimension q upon the removal of the extra edge.

Proposition 3.9. *Let $q \geq 0$, and let A be an induced subcomplex of a clique complex X (i.e. A contains a simplex σ of X whenever A contains all vertices of σ). Suppose X is (A, q)-clique-minimal. Then the following hold:*

- *If A is just a vertex, then X has at least $2q + 2$ vertices; otherwise, X has at least one more vertex than A does.*
- *We have* $\deg v \geq 2q$ *for every vertex v in X not in A, where deg denotes the degree of a vertex in the underlying graph of X.*

Remark 3.10. Note that a clique-minimal complex is not necessarily minimal (as defined by dropping all instances of 'clique' in the definition above). For example, let A be a triangulated annulus whose inner boundary has three edges. Consider the simplicial complex T formed by gluing together two identical copies of A along the boundaries. Let T' be the simplicial complex formed by gluing a triangle to the inner boundary of A in T. Then T is 2-minimal but not 2-clique-minimal (because T is not a clique complex). On the other hand, T' is not 2-minimal, but it is 2-clique-minimal, if the triangulation of A is nice enough.

4. Proof synopsis and a decomposition result

In this section, we begin our proof of Theorem 1.6.

4.1. Proof synopsis

Recall that in Section 1.6, homology classes in preferential attachment clique complexes are said to be predominantly represented by small interconnected cycles, like those in the clique complex in Figure 4. Now, we describe such cycles at dimension 2 in greater detail. The case for higher dimensions is similar.

For each positive integer k, let Γ_k be the graph consisting of four vertices forming a square, with no diagonals, and k other vertices that are each connected to all corners of the square. (The graph Γ_k is unique up to the permutation of nodes in the square. Permuting the other nodes gives the same graph.) The graph in Figure 4 is Γ_3.

Note that Γ_k contains k distinct (but not disjoint) copies of Γ_1. Since one can show that

$$\beta_2(\text{Clique}(\Gamma_k)) = k - 1,$$

where $\text{Clique}(\Gamma_k)$ denotes the clique complex of Γ_k, the Betti number can be approximated by the number of copies of Γ_1 in the graph.

When we approximate the Betti number of the whole preferential attachment clique complex $X(T, \delta, m)$ with the number of copies of Γ_1 in it, the error term consists of a few parts.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

First, other subcomplexes, such as (the clique complexes of) subgraphs with vertices attached to pentagons rather than squares, may also add to the Betti number. However, such subcomplexes are more complicated, and hence fewer of them arise from the preferential attachment mechanism.

On the other hand, copies of Γ_k in $X(T, \delta, m)$ may be boundaries (recall from Definition 1.3 that boundaries do not contribute to the Betti numbers). However, boundaries of higher-dimensional chains are, again, more complicated, and fewer of them arise. For technical reasons, we will analyze two types of boundaries separately.

In the next subsection, we make our approximation precise by establishing a decomposition result, Proposition 4.1.

4.2. A decomposition result

Let X be a clique complex with vertices $\{1, \ldots, T\}$. We make the following definitions:

- Let $X^{(t)}$ be the subcomplex of X such that $X^{(t)}$ consists of all simplices of X whose vertices are all in $\{1, \ldots, t\}$.
- Let $L^{(t)}$ be the link of t in $X^{(t)}$, and let $f^{(t)} : L^{(t)} \to X^{(t-1)}$ be the inclusion map ($L^{(t)} \subseteq X^{(t-1)}$ because t itself does not lie in its link; recall that 'link' and 'inclusion' are respectively defined in Section 3.1 and Appendix B.1).
- For a subcomplex S of X, an integer $q \geq 2$, nodes s, t of X such that all indices of nodes in S precede s, and $s < t$, let $\mathcal{S}(S, q, s, t)$ be the event where
  - S is isomorphic to S^{q-1}, and
  - $S \subseteq L^{(s)} \cap L^{(t)}$.

Recall that S^{q-1} is the octahedral sphere in Section 3.1. Note that when $\mathcal{S}(S, q, s, t)$ happens, X contains a q-dimensional sphere with S as the equator and s and t as the poles.

We need the following terms for our estimate (all of them depend on the dimension q, but we drop the dependence from the notation since we will not change q):

$$\begin{aligned}
\ell^{(t)}(S, s) &= \mathbf{1}[\mathcal{S}(S, q, s, t)], \\
b_{IK}^{(t)}(S, s) &= \mathbf{1}[\mathcal{S}(S, q, s, t)]\mathbf{1}[\beta_q(L^{(t)}, S) > 0], \\
b_{KL}^{(t)} &= \beta_q(L^{(t)}), \\
u^{(t)} &= \beta_{q-1}(L^{(t)}),
\end{aligned}$$

where $\mathbf{1}[\Lambda]$ denotes the indicator of the event Λ, and $\beta_q(L^{(t)}, S)$ is the relative Betti number of the pair $(L^{(t)}, S)$ (defined in Appendix B.2).

The letters u and ℓ stand for 'upper bound' and 'lower bound', and b stands for 'boundary'. The subscripts IK and KL denote the two types of boundaries mentioned in the introduction, and we will explain them below. Note that despite their notational difference, $u^{(t)}$ and $b_{KL}^{(t)}$ are just Betti numbers of $L^{(t)}$ at different dimensions. They will be estimated in the same way in Lemma 5.3.

We now state our decomposition result, which we prove at the end of this section.

Proposition 4.1. *Let X be a clique complex with vertices labeled by positive integers. Let $q \geq 2$, let S be a subcomplex of X, and let s be a node whose label is larger than all node labels in S.*

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

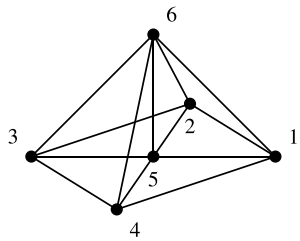

FIGURE 7. Illustration for the underlying graph of $X^{(6)}$ in Example 4.3.

Then

$$\sum_{s<t\leq T} (\ell^{(t)}(S,s) - b_{IK}^{(t)}(S,s)) - \sum_{t\leq T} b_{KL}^{(t)} \leq \beta_q(X) \leq \sum_{t\leq T} u^{(t)}.$$

Returning to the discussion in the previous subsection regarding the number of copies of Γ_k as an approximation of the Betti number at dimension 2, one may check that, for each $t \leq T$, $u^{(t)} = 1$ on the event where $L^{(t)}$ is a square, in which case the square along with vertex t gives a copy of Γ_1. We will show that the count for squares is the dominating term for $u^{(t)}$ in expectation at the end of the proof of Lemma 5.3. Similarly, for a fixed S and s and each $s < t \leq T$, $\ell^{(t)}(S,s)$ counts a subset of copies of Γ_1 formed by S and t in the complex. As mentioned in the previous subsection, the Betti number of Clique(Γ_k) at dimension 2 is one less than the number of copies of Γ_1. Here we capture this difference by not counting the copy of Γ_1 formed by S and s.

KL stands for 'kill', as a cycle is killed in the homology group when a boundary is formed. We give an example where $b_{KL}^{(t)} = 1$.

Example 4.2 (Kill.) *Let $q = 2$ and $t = 7$. Let $X^{(6)} \cong S^2$, and let $X^{(7)}$ be the clique complex with vertex 7 connected to 1, 2, 3, 4, 5, 6. The underlying graph of $X^{(7)}$ is shown in the right panel of Figure 6. Then $X^{(7)} \cong D^3$. The Betti number at dimension 2 drops by 1 when the new simplices are added.*

IK stands for 'instant kill', as the defining event happens when a new cycle is killed as soon as it forms. We give an example where $b_{IK}^{(t)}(S,s) = 1$.

Example 4.3. (Instant kill.) *Let $q = 2$, $s = 5$, $t = 6$. Let $X^{(5)} \cong D^2$ with vertices 1,2,3,4 on the boundary and vertex 5 in the center. Let $S = X^{(4)}$. Let $X^{(6)}$, which is illustrated in Figure 7, be the clique complex with vertex 6 connected to 1, 2, 3, 4, 5. The addition of new simplices creates a 2-cycle, namely a signed sum of the triangles of the two copies of (the clique complex of) Γ_1 spanned by 1, 2, 3, 4, 5 and 1, 2, 3, 4, 6. However, this does not add to the Betti number, since the new cycle is also the boundary of a signed sum of all tetrahedra in $X^{(6)}$. One may check that indeed $b_{IK}^{(6)}(X^{(4)}, 5) = 1$, as the exact sequence of the pair $(L^{(6)}, X^{(4)})$ shows $H_2(L^{(6)}, X^{(4)}) \cong H_1(X^{(4)}) \cong \mathbb{Z}$.*

We now prove Proposition 4.1, which is a direct corollary of the first and third bullet points of the following lemma.

Lemma 4.4. *Let $f_{q-1}^{(t)}$ be the homomorphism between homology groups at dimension $q-1$ that is induced by the map $f^{(t)}$ (cf. Appendix B.1). Then the following statements are true under the assumptions of Proposition 4.1:*

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

- *We have*

$$\sum_{t\le T}\left[\operatorname{rk}\ker f_{q-1}^{(t)} - \beta_q(L^{(t)})\right] \le \beta_q(X) \le \sum_{t\le T}\beta_{q-1}(L^{(t)}).$$

- *Suppose* $\mathcal{S}(S, q, s, t)$ *happens. Let* j *: Plugging in the definitions* $S \to L^{(t)}$ *be the inclusion map. If* $\operatorname{rk}\ker j_{q-1} = 0$*, then* $\operatorname{rk}\ker f_{q-1}^{(t)} \ge 1$.
- *Suppose* $\mathcal{S}(S, q, s, t)$ *happens. If* $\beta_q(L^{(t)}, S) = 0$*, then* $\operatorname{rk}\ker j_{q-1} = 0$*, where* j *still denotes the inclusion map of* S *in* $L^{(t)}$*, and hence* $\operatorname{rk}\ker f_{q-1}^{(t)} \ge 1$.

We remark that we will not use the second bullet point until our discussion on our simulation results in Section 8. Here it is merely a stepping stone towards the third bullet point.

Proof of Proposition 4.1. Plugging in the definitions of $u^{(t)}$ and $b_{KL}^{(t)}(S, s)$ into the first bullet point of the lemma gives

$$\sum_{t\le T}\operatorname{rk}\ker f_{q-1}^{(t)} - \sum_{t\le T} b_{KL}^{(t)} \le \beta_q(X) \le \sum_{t\le T} u^{(t)}.$$

It remains to show that the first sum $\sum_{t\le T}\operatorname{rk}\ker f_{q-1}^{(t)}$ is at least

$$\sum_{s<t\le T}(\ell^{(t)}(S, s) - b_{IK}^{(t)}(S, s)) = \sum_{s<t\le T}\mathbf{1}[\mathcal{S}(S, q, s, t)]\mathbf{1}[\beta_q(L^{(t)}, S) = 0]. \tag{4}$$

Since each term in $\sum_{t\le T}\operatorname{rk}\ker f_{q-1}^{(t)}$ is nonnegative, keeping only terms where

- $s < t$,
- $\mathcal{S}(S, q, s, t)$ happens, and
- $\beta_q(L^{(t)}, S) = 0$

shows that $\sum_{t\le T}\operatorname{rk}\ker f_{q-1}^{(t)}$ is at least

$$\sum_{s<t\le T}\mathbf{1}[\mathcal{S}(S, q, s, t)]\mathbf{1}[\beta_q(L^{(t)}, S) = 0]\operatorname{rk}\ker f_{q-1}^{(t)},$$

which, by the third bullet point of the lemma, is at least the right-hand side of Equation (4). □

Proof of Lemma 4.4. For the first bullet point, let C be the star of t in $X^{(t)}$. Then by Proposition 3.1, C is acyclic. Consider the Mayer–Vietoris sequence (Theorem B.3) for the decomposition $X^{(t)} = X^{(t-1)} \cup C$:

$$H_q(L^{(t)}) \xrightarrow{f_q^{(t)}} H_q(X^{(t-1)}) \to H_q(X^{(t)}) \to H_{q-1}(L^{(t)}) \xrightarrow{f_{q-1}^{(t)}} H_{q-1}(X^{(t-1)}),$$

where the trivial summands $H_q(C) \cong H_{q-1}(C) \cong 0$ are suppressed (we assume $q \ge 2$, and hence $q > q-1 > 0$). Hence we have the finite-length exact sequence

$$0 \to \operatorname{im} f_q^{(t)} \to H_q(X^{(t-1)}) \to H_q(X^{(t)}) \to \ker f_{q-1}^{(t)} \to 0.$$

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

Since the alternating sum of ranks vanishes (Proposition B.2), we have

$$\operatorname{rk} H_q(X^{(t)}) - \operatorname{rk} H_q(X^{(t-1)}) = \operatorname{rk}\ker f_{q-1}^{(t)} - \operatorname{rk}\operatorname{im} f_q^{(t)},$$

where $\operatorname{rk}\ker f_{q-1}^{(t)} \le \operatorname{rk} H_{q-1}(L^{(t)}) = \beta_{q-1}(L^{(t)})$ because the latter group contains the former, and $\operatorname{rk}\operatorname{im} f_q^{(t)} \le \beta_q(L^{(t)})$ by the rank–nullity theorem. The first bullet point then follows by summing over t.

For the second bullet point, inclusions induce the following commutative diagram (commutativity follows from Lemma B.1):

$$\begin{tikzcd} H_{q-1}(S) \arrow[r, "j_{q-1}"] \arrow[d, "0"] & H_{q-1}(L^{(t)}) \arrow[d, "f_{q-1}^{(t)}"] \\ H_{q-1}(\mathrm{St}_{X^{(t-1)}}(s)) \arrow[r] & H_{q-1}(X^{(t-1)}). \end{tikzcd}$$

The left vertical map is 0 because $\mathrm{St}_{X^{(t-1)}}(s)$ is acyclic (cf. Proposition B.2). Commutativity then implies $\operatorname{im} j_{q-1}$ lies in $\ker f_{q-1}^{(t)}$, and hence

$$\operatorname{rk}\ker f_{q-1}^{(t)} \ge \operatorname{rk}\operatorname{im} j_{q-1} = 1 - \operatorname{rk}\ker j_{q-1} = 1,$$

where the first equality follows from the rank–nullity theorem (recall $\beta_{q-1}(S) = 1$), and the second one holds by assumption. The second bullet point then follows.

The first part of the last bullet point follows from the long exact sequence for $(L^{(t)}, S, \emptyset)$ (Theorem B.4):

$$\operatorname{rk}\ker j_{q-1} = \operatorname{rk}\operatorname{im}(H_q(L^{(t)}, S) \to H_{q-1}(S)) \le \operatorname{rk} H_q(L^{(t)}, S) = 0.$$

The second part follows from the second bullet point. □

5. Proof of Theorem 1.6

It remains to estimate the expectations of all the terms in Proposition 4.1 for the preferential attachment complex for some choice of S and s. Throughout our proof, we fix $S = X^{(2q)}$ and $s = 2q+1$.

A direct application of Proposition 3.6 gives a lower bound on $\sum \mathbb{E}[\ell^{(t)}(X^{(2q)}, 2q+1)]$.

Lemma 5.1. *Consider the preferential attachment complex $X = X(T, \delta, m)$. Let $q \ge 0$ and suppose $m \ge 2q$. Then*

$$\sum_{t \le T} \mathbb{E}[\ell^{(t)}(X^{(2q)}, 2q+1)] = \begin{cases} \Omega(T^{1-2q\chi(\delta,m)}) & \text{if } 1 - 2q\chi(\delta, m) > 0, \\ \Omega(\log T) & \text{if } 1 - 2q\chi(\delta, m) = 0, \end{cases}$$

where the big-Ω constants depend only on q, δ, m.

Proof. For each $t > 2q+1$, let $\Gamma^{(t)}$ be a subgraph of $U(t, m)$ (possibly with repeated edges) with the following properties:

- The vertices are $1, \ldots, 2q+1$ and t.
- Each of t and $2q+1$ is connected to each of $1, \ldots, 2q$, and t is not connected to $2q+1$.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

- All edges incident on t are simple.
- The pre-degree of every vertex other than 1 and t is m.
- Removing t and edges incident on t and replacing repeated edges with simple edges gives the underlying graph of D^q.

For example, for $q = 2$, $\Gamma^{(t)}$ may be the graph illustrated in the left panel of Figure 6, which by definition has the following edges:

- m edges between 1 and 2, and m edges between 1 and 3;
- $m - 1$ edges between 4 and 2, and one edge between 4 and 3;
- $m - 3$ edges between 5 and 1, and one edge between 5 and each of 2, 3, 4; and
- one edge between t and each of 1, 2, 3, 4.

The $\Gamma^{(t)}$ can be chosen to be isomorphic to each other. Then $\mathcal{S}(X^{(2q)}, q, 2q+1, t)$ holds for $X(T, \delta, m)$ whenever $X(T, \delta, m)$ contains $\Gamma^{(t)}$.

To simplify notation, let $p(v) = p_{\Gamma^{(t)}}(v;\delta, m)$ for every vertex v of $\Gamma^{(t)}$. Note that $p(t) = -2qx(\delta, m)$. By Proposition 3.6, we have

$$\mathbb{P}[\mathcal{S}(X^{(2q)}, q, 2q+1, t)] = \Omega(t^{p(t)} \prod_{k \le 2q+1} k^{p(k)}) = \Omega(t^{p(t)}) = \Omega(t^{-2qx(\delta,m)}).$$

Summing over t and applying the integral test gives the desired result. □

Next we use Theorem 3.4 and Proposition 3.9 to estimate $\sum \mathbb{E}[u^{(t)}]$, $\sum \mathbb{E}[b_{KL}^{(t)}]$, and $\sum \mathbb{E}[b_{IK}^{(t)}(X^{(2q)}, 2q+1)]$, but before that, we need an auxiliary lemma to simplify the application of Theorem 3.4.

Lemma 5.2. *Let Γ be a subgraph of $U(T, m)$ with vertex set $V_\Gamma = \{v_1 < \ldots < v_{|V_H|}\}$. Then the sequence (a_k) in Theorem 3.4 satisfies*

$$a_k - a_{k-1} \begin{cases} = d_\Gamma(v_k)x(\delta, m) - 1 & \text{if } d^{\text{post}}{}_\Gamma(v_k) = 0, \\ \ge (d_\Gamma(v_k) - 2)x(\delta, m) & \text{if } d^{\text{post}}{}_\Gamma(v_k) > 0. \end{cases}$$

Proof. This can be verified directly. □

The next lemma gives a matching upper bound on $\sum \mathbb{E}[u^{(t)}]$ and an upper bound on $\sum \mathbb{E}[b_{KL}^{(t)}]$ with a smaller order of magnitude.

Lemma 5.3. *Consider the preferential attachment complex $X = X(T, \delta, m)$. Let $q \ge 1$ and suppose $m \ge 2(q+1)$. Then*

$$\sum_{t \le T} \mathbb{E}[\beta_q(L^{(t)})] = \begin{cases} O(T^{1-2(q+1)x(\delta,m)}) & \text{if } 1 - 2(q+1)x(\delta, m) > 0, \\ O(\log T) & \text{if } 1 - 2(q+1)x(\delta, m) = 0, \\ O(1) & \text{otherwise,} \end{cases}$$

where the big-O constants depend only on q, δ, m.

Proof. Since $L^{(t)}$ has at most m vertices, it has at most $\binom{m}{q+1}$ simplices of dimension q, where $\binom{m}{q+1}$ denotes the binomial coefficient. By the weak Morse inequality (Theorem 1.7 of [17]),

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

$$\beta_q(L^{(t)}) \leq \binom{m}{q+1} \mathbf{1}[\beta_q(L^{(t)}) > 0], \tag{5}$$

and hence $\sum \beta_q(L^{(t)})$ is at most $\binom{m}{q+1}$ times the number of values of t such that $\beta_q(L^{(t)}) > 0$. We will construct a distinct graph $\Gamma^{(t)}$ for each such t, and bound the expected count of such graphs. (These $\Gamma^{(t)}$ are different from those in Lemma 5.1.)

Whenever $\beta_q(L^{(t)}) > 0$, $L^{(t)}$ contains a q-clique-minimal subcomplex (note that $L^{(t)}$ is also clique; see Lemma 7.1), which by Proposition 3.9 has at least $2q+2$ vertices, whose degrees are at least $2q$. Since these vertices are all connected to node t in $G(T, \delta, m)$, this gives rise to a subgraph $\Gamma^{(t)}$ in $G(T, \delta, m)$ with the following properties:

- $\Gamma^{(t)}$ has at least $2q+3$ vertices, and at most $m+1$ vertices,
- the vertices of $\Gamma^{(t)}$ all have degrees at least $2q+1$, and
- the last vertex of $\Gamma^{(t)}$ (which is t) is connected to all other vertices.

Note that for $t \neq s$, $\Gamma^{(t)}$ and $\Gamma^{(s)}$ are distinct subgraphs in $G(T, \delta, m)$ because their last nodes are different.

Therefore, Equation (5) implies that $\sum \mathbb{E}[\beta_q(L^{(t)})]$ is at most $\binom{m}{q+1}$ times the expected number of subgraphs of $G(T, \delta, m)$ satisfying the properties above.

We use Theorem 3.4 to give an upper bound on the expected count of all such subgraphs. Since such subgraphs have at most $m+1$ vertices, there are only finitely many isomorphism classes of such graphs. Fix an isomorphism class and pick a representative Γ in the class.

We claim the sequence (a_k) in Equation (3) attains its maximum at $|V_\Gamma| - 1$ or $|V_\Gamma|$. To establish this claim, it suffices to show $a_0 < a_1 < \ldots < a_{|V_\Gamma|-1}$. Let $0 < k < |V_\Gamma|$. Then v_k is not the last vertex of Γ. Since its degree is at least $2q+1 \geq 3$, and since it is connected to the last vertex, Lemma 5.2 implies that

$$a_k - a_{k-1} \geq (3-2)x(\delta, m) > 0.$$

The claim then follows.

By definition, $a_{|V_\Gamma|} = 0$. Hence, by Lemma 5.2 again,

$$a_{|V_\Gamma|-1} = -(a_{|V_\Gamma|} - a_{|V_\Gamma|-1}) = 1 - d_\Gamma(v_{\text{last}})x(\delta, m),$$

where v_{last} is the last node of Γ. Therefore, the expected count of Γ is

$$O(T^{A_\Gamma} \log^{r_\Gamma} T),$$

where

$$A_\Gamma = \max\left(0, 1 - d_\Gamma(v_{\text{last}})x(\delta, m)\right),$$
$$r_\Gamma = \mathbf{1}[1 - d_\Gamma(v_{\text{last}})x(\delta, m) = 0].$$

The sum of counts for all isomorphism classes is dominated by the classes of the Γ's with the minimum $d_\Gamma(v_{\text{last}})$. Our criteria for Γ require $d_\Gamma(v_{\text{last}}) \geq 2q+2$. The result then follows. Note that the minimal $d_\Gamma(v_{\text{last}})$ is attained by D^{q+1}, which is the only minimizer such that the corresponding $\beta_q(L^{(t)})$ is positive, and this justifies our discussion of Γ_1's in the introduction. □

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

A similar argument using relative homology gives an upper bound on $\sum_{t\le T}\mathbb{E}[b_{IK}^{(t)}(X^{(2q)},2q+1)]$ with a smaller order of magnitude.

Lemma 5.4. *Suppose $q\ge 2$ and $m\ge 2q$. Let S be a (possibly random) subcomplex S of $X(T,\delta,m)$, and let s be a (possibly random) node in $X(T,\delta,m)$ that is (almost surely) a later node than all nodes in S. Then*

$$\mathbb{E}[\sum_{s<t\le T} b_{IK}^{(t)}(S,s)] = \begin{cases} O(T^{1-(2q+1)x(\delta,m)}) & \text{if } 1-(2q+1)x(\delta,m)>0,\\ O(\log T) & \text{if } 1-(2q+1)x(\delta,m)=0,\\ O(1) & \text{otherwise,}\end{cases}$$

where the big-O constants depend only on q, δ, m.

Proof. Again, on the event $\beta_q(L^{(t)},S)>0$, $L^{(t)}$ contains an (S,q)-clique-minimal subcomplex, which by Proposition 3.9 has at least $2q+1$ vertices, $2q$ of them (from S) with degree at least $(2q-2)$, and the rest with degree at least $2q$.

Since these vertices are all connected to node t in $G(T,\delta,m)$, this gives rise to a subgraph $\Gamma^{(t)}$ in $G(T,\delta,m)$ with the following properties:

- $\Gamma^{(t)}$ has at least $2q+2$ vertices, and at most $m+1$ vertices,
- $2q$ vertices of Γ have degrees at least $2q-1$, and the rest have degrees at least $2q+1$, and
- the last vertex of $\Gamma^{(t)}$ (which is t) is connected to all other vertices.

Then $\mathbb{E}[\sum_{s<t\le T} b_{IK}^{(t)}(S,s)]$ is at most the expected number of such subgraphs.

Appealing to Theorem 3.4 again, for each such Γ, the maximum A_Γ of the sequence (a_k) is attained by one of the last two terms. The result then follows. □

The proof of Theorem 1.6 is now completed by plugging the estimates in Lemmas 5.1, 5.3, and 5.4 into Proposition 4.1.

6. Proofs of Propositions 1.7 and 1.8

We first handle the trivial cases.

Proof of Proposition 1.8. The claim for $q=0$ is trivial, because preferential attachment graphs are connected by construction. The claim for $m<2q$ follows from the fact that there are not enough edges to form q-dimensional holes. This can be seen by applying Proposition 3.9 to the last node in a hypothetical q-minimal subcomplex. □

Finally, we prove Proposition 1.7 using the Morse inequality.

Proof. Let $|\bar V|,|\bar E|,|\bar F|$ be the expected numbers of vertices, edges, and triangles in $X(T,\delta,m)$. The strong Morse inequality (Theorem 1.8 of [17]) implies that

$$|\bar E|-|\bar V|-|\bar F|\le \mathbb{E}\beta_1(X(T,\delta,m))\le \mathbb{E}\beta_0(X(T,\delta,m))+|\bar E|-|\bar V|.$$

Obviously, $|\bar V|=T$ and $\beta_0(X(T,\delta,m))=1$. Theorem 3.4 implies that the expected numbers of triangles and of bi-angles (the two-node graph with two distinct edges from one node to the other) are both $o(T)$, and hence

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

$$m(T-1) - 2o(T) \leq |\bar{E}| \leq m(T-1),$$
$$|\bar{F}| = o(T).$$

The result then follows. □

7. Proof of Proposition 3.9

Our proof follows the argument of Lemmas 5.2 and 5.3 in [27]. We generalize these lemmas to the setting of relative homology in Lemma 7.2. To ensure clique-minimality conditions are met in our argument, we need Lemma 7.1 to ensure that certain subcomplexes of a clique complex are clique complexes. Proposition 3.9 is a corollary of Lemma 7.2.

For every simplicial complex X and every vertex v of X, we denote by $X - v$ the simplicial complex that consists precisely of simplices that do not contain v. Note that $\mathrm{Lk}_X v = \mathrm{St}_X v \cap (X - v)$ (recall that Lk and St are the link and star in Section 3.1).

Lemma 7.1. *If X is a clique complex, then $\mathrm{Lk}_X v$ and $X - v$ are clique complexes for every vertex v in X.*

Proof. The claim for $X - v$ is trivial. For $\mathrm{Lk}_X v$, let $w_0, \ldots, w_q$ be distinct and pairwise connected vertices in $\mathrm{Lk}_X v$. It suffices to show that $\sigma = \{w_0, \ldots, w_q\}$ lies in the link. Since σ does not contain v, it lies in $X - v$; hence it suffices to show that σ lies in $\mathrm{St}_X v$.

Since each w_i lies in the link, it is a vertex of a simplex in $\mathrm{St}_X v$. This simplex contains v by definition. Therefore, $\{w_i, v\}$ is an edge in this simplex, and hence in X.

Since this is true for all w_i, the clique complex X contains the simplex $v * \sigma_v = \{w_0, \ldots, w_q\} \cup \{v\}$. By definition $v * \sigma$ lies in $\mathrm{St}_X v$, and hence σ does too. The result then follows. □

Next, we generalize Lemmas 5.2 and 5.3 of [27] to the setting of relative homology. Only the first part of the following lemma is novel. The second claim is Lemma 2.1.4 of [18] and Lemma 5.3 of [27] phrased differently. We reproduce the proofs in those papers using our terminology.

Lemma 7.2. *Let X be a clique complex, and let A be a (not necessarily clique) subcomplex of X. Suppose X is (A, q)-clique-minimal. Then the following statements are true:*

- *For every $q > 0$ and every vertex v in X but not in A,*

$$\beta_{q-1}(\mathrm{Lk}_X(v), B) > 0$$

  whenever B is empty or it is an acyclic subcomplex of $\mathrm{Lk}_X(v) \cap A$.

- *If $q \geq 0$ and A consists of one single vertex, then X has at least $2q + 2$ vertices.*

Proof. For the first claim, fix a vertex v in X but not in A. We have the following commutative diagram:

$$\begin{tikzcd}
H_q(\mathrm{Lk}_X v, B) \arrow[r] \arrow[d] & H_q(\mathrm{St}_X v, B) \arrow[r, "0"] \arrow[d] & H_q(\mathrm{St}_X v, \mathrm{Lk}_X v) \arrow[r, hook, "\psi"] \arrow[d, "\varphi;EX"] & H_{q-1}(\mathrm{Lk}_X v, B) \arrow[d] \\
H_q(X - v, A) \arrow[r, "\mathrm{rk}\,0"] & H_q(X, A) \arrow[r, "\eta"] & H_q(X, X - v) \arrow[r] & H_{q-1}(X - v, A),
\end{tikzcd}$$

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

where the two rows are long exact sequences of triples (Theorem B.4), and the vertical maps are induced by inclusion (defined in Appendix B.1). We would like to show that the top-right group has positive rank.

We first check commutativity. The far-right square commutes by the naturality of the long exact sequence of triples, and the other squares commute because all maps are induced by inclusions (cf. Lemma B.1).

We explain the annotations in the diagram. The map φ, marked by 'EX', is an isomorphism by the excision theorem (Theorem B.5 with A and B being $\mathrm{St}_X v$ and $X - v$). The map marked by '0' is zero because Lemma B.7 implies $H_q(\mathrm{St}_X v, B) \cong 0$ (recall from Proposition 3.1 that $\mathrm{St}_X v$ is acyclic). The map marked by 'rk 0' is rank-0 by the clique-minimality of X.

Exactness implies ψ is injective. We also have

$$\beta_q(X, X - v) \geq \mathrm{rk\ im}\ \eta = \beta_q(X, A) - \mathrm{rk\ ker}\ \eta = \beta_q(X, A) > 0, \tag{5}$$

where the two equalities hold by the rank–nullity theorem and exactness, and the last inequality holds by assumption.

Since the top-right group contains $\psi\varphi^{-1}H_q(X, X - v)$, it must have a positive rank. The first claim then follows.

For the second claim, the case for $q = 0$ is trivial. For $q > 0$, suppose for the sake of contradiction that X has strictly fewer than $2q + 2$ vertices.

We first consider the main case when there is a vertex v connected to the vertex in A. The first claim implies $\beta_{q-1}(\mathrm{Lk}_X(v), A) > 0$. Therefore, $\mathrm{Lk}_X(v)$ has an $(A, q - 1)$-clique-minimal subcomplex, and hence by induction the link has at least $2q$ nodes. Since we have assumed X has strictly fewer than $2q + 2$ vertices, all nodes other than v are in the link. Since X is a clique complex, this means X is $\mathrm{St}_X(v)$, and hence is acyclic (Proposition 3.1), in contradiction to the assumption of (A, q)-clique-minimality.

We now consider the case when A is not connected to any other vertices. Since $\beta_q(X, A) > 0$, $X - A$ has at least one edge, say with endpoints v, w in $X - A$. It can be directly verified that $X - A$ is q-minimal, and hence $(\{w\}, q)$-minimal. The main case above implies $X - A$ has at least $2q + 2$ vertices, and hence so does X. □

Proof of Proposition 3.9. The first part of the first claim is just the second claim of Lemma 7.2. The second part of the first claim is trivial, because if X has the same vertex set as A, then $X = A$ (because A is an induced subcomplex), and hence $H_q(X, A) = 0$.

The second claim is trivial for $q = 0$. It is also straightforward for $q = 1$: the minimality of X implies that it has no degree-0 or degree-1 vertices outside of A, because their removal does not change $H_q(X, A)$. For $q > 1$, since the removal of isolated vertices not in A does not change the (relative) Betti number at dimension q, clique-minimality implies $\deg v \geq 1$, and hence $\mathrm{Lk}_X(v)$ is nonempty. The first claim in Lemma 7.2 implies $\beta_{q-1}(\mathrm{Lk}_X(v), \emptyset) > 0$, and hence $\beta_{q-1}(\mathrm{Lk}_X(v), \{w\}) > 0$ whenever $w \in \mathrm{Lk}_X(v)$. The second claim of Lemma 7.2 then implies $\mathrm{Lk}_X(v)$ contains a subcomplex with at least $2q$ vertices. This means that v is connected to at least $2q$ vertices in X. □

8. Numerical simulation

We discuss the simulation we mentioned in the introduction in greater detail. Recall that the right panel of Figure 2 illustrates the evolution of the mean Betti numbers. Below, we explain the set-up of the simulation, and in the last paragraph of this section we discuss the results shown in Figure 2.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

Numerical computations related to topology and graph theory are done with Ripser [4, 42] and igraph [13], respectively. Other numerical computations are done with NumPy [22] and SciPy [46]. Codes are compiled with Numba [32]. Plots are generated with Matplotlib [26].

We generate 500 preferential attachment clique complexes with $T = 10^4$ nodes and with parameters $m = 7$ and $\delta = -5$. We compute the sample mean of their Betti numbers (with coefficients in $\mathbb{Z}/2\mathbb{Z}$) at dimension $q = 2$. The black curve corresponds to the evolution of mean Betti numbers. We remark that the median of means gives a similar estimate of the expectation.

We also compute the sample mean of the upper bound $\sum u^{(t)}$ in prop:decomposition and the sample mean of a lower bound $\sum_{s<t\leq T} (\ell^{(t)}(S, s) - \hat{b}^{(t)}_{IK}(S, s)) - \sum_{t\leq T} b^{(t)}_{KL}$, where S, s, and $\hat{b}^{(t)}_{IK}(S, s)$ will be defined below. The evolutions of the means of these bounds are plotted in dotted lines. We compute these quantities for graphs with $T = 10^5$ nodes, because their computation is cheaper than that of Betti numbers.

We now define S, s and $\hat{b}_{IK}(S, s)$:

- S is the first induced subcomplex of $X^{(20)}$ (in an arbitrary but deterministically consistent ordering) that is isomorphic to S^{q-1}, if one exists; otherwise, it is the first subcomplex of $X^{(20)}$;
- s is the first node whose label is larger than all of those in S such that $S \subseteq \mathrm{Lk}_X(s)$, if it exists; otherwise it is node 21;
- $\hat{b}^{(t)}_{IK}(S, s) = \mathbf{1}[\mathcal{S}(S, q, s, t)]\mathbf{1}[\mathrm{rk}\ker j_{q-1} = 1]$, where $j : S \to L^{(t)}$ denotes the inclusion map.

We make the following remarks on these definitions:

- The 'otherwise' statements in the above definitions are unimportant, because in those cases, $\ell^{(t)}(S, s) = \hat{b}^{(t)}_{IK}(S, s) = 0$.
- We do not fix $S = X^{(2q)}$ and $s = 2q + 1$ as in our proofs, because $\mathcal{S}(X^{(2q)}, q, 2q + 1, t)$ happens so rarely that $\mathbb{E}[\sum \ell^{(t)}]$ is too small to be numerically estimated.
- We change the definition of $\hat{b}^{(t)}_{IK}$ because the computation of relative Betti numbers is numerically inconvenient. We numerically compute $\mathrm{rk}\ker j_{q-1}$ by computing the persistence diagram for the inclusion $S^{(t)} \subseteq L^{(t)}$ with Ripser. By the first two bullet points of Lemma 4.4, the new expression does give a lower bound.

Finally, we draw a band that contains all curves. The slope of the band is determined by Theorem 1.6. While the discussion following the theorem suggests some values for the y-intercepts of the band, the corresponding band trivially covers the entire plot. Therefore we manually choose other values for the y-intercepts.

It is apparent from the plot that the convergence is slow. In particular, at $T = 10^5$, the mean upper bound still grows at a rate faster than the asymptotic rate. However, it is obvious that the curve is concave, and hence has a decreasing slope. We also note that the mean upper bound is a good approximation of the mean Betti numbers.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

9. Future directions

We have established analytically the asymptotics of the expected Betti numbers of affine preferential attachment clique complexes and illustrated them numerically. A number of open questions remain.

It would be desirable to have sharper estimates of the expected Betti numbers, and finer descriptions of the distributions of the Betti numbers.

Other topological properties of preferential attachment graphs are also of interest. To understand the robustness of the complex, it would be helpful to understand the evolution of Betti numbers as nodes are removed. One may also consider the Betti numbers of the Rips complexes of the graph with respect to the graph metric. Beyond Betti numbers, one may also consider the homotopy type of the random simplicial complexes. Since holes are filled by later nodes, it is possible that all holes are filled if nodes are added ad infinitum and the number of edges added each time grows slowly. In particular, in a private conversation, Weinberger conjectured that the resultant complex is *contractible* (X is said to be contractible if there exists an $x_0 \in X$ and a map $f : X \times [0, 1] \to X$ such that $f(\cdot, 0)$ is the identity on X, $f(\cdot, 1) \equiv x_0$, and $f(x_0, \cdot) \equiv x_0$; cf. [34, p. 108]).

Other scale-free simplicial complexes are also of interest. It is not even clear whether our result is universal across different formulations of the preferential attachment models (e.g. when m edges are added simultaneously rather than sequentially at each stage). The configuration is another popular scale-free model. Different techniques are likely necessary, as there is no natural ordering of the vertices in the configuration model. In [24], the topology of the age-dependent random connected model is investigated. It remains open whether the limiting distribution of Betti numbers is a heavy-tailed stable distribution when the degree distribution has infinite variance.

Appendix A. More on subgraphs in preferential attachment graphs

In this appendix, we give the intuition behind Theorem 3.4, and we discuss the formulation of Proposition 3.6, which is very different from its original form.

For the intuition for Theorem 3.4, the main idea is that one may count the isomorphic copies of a graph Γ in $G(T, \delta, m)$ with node labels (or equivalently, the arrival times of the nodes) having specific orders of magnitude separately. For example, one may count the number of triangles such that the first node is approximately $T^0 = 1$, the second node is approximately $T^{1/2}$, and the third node is approximately $T^1 = T$. The order of magnitude of the total count is then the maximum of all orders of magnitude for the counts with specified magnitudes of node labels, and this gives rise to the optimization problem. It turns out the optimization problem is linear in the logarithm of the vertex labels (0, 1/2, and 1 in our example), and hence at the maximum they are either $T^0 = 1$ or $T^1 = T$. To determine the maximum, it suffices to identify the first node label whose order of magnitude is $T^1 = T$. This is why the maximum in Theorem 3.4 ranges from 0 to the number of nodes in Γ. We refer the reader to [20] for further discussion on the theorem.

We drastically paraphrased Proposition 3.6 both to facilitate its application and to avoid a potential confusion about the edge orientations. In Lemma 1 of [20], an ℓ-edge subgraph Γ of $U(T, m)$ is denoted by three vectors $\mathbf{u} = (u_1, \ldots, u_\ell)$, $\mathbf{v} = (v_1, \ldots, v_\ell)$, and $\mathbf{j} = (j_1, \ldots, j_\ell)$, where the kth edge of Γ is said to be the j_kth edge (among the m edges starting from u_k) *from* node u_k *to* node v_k. In the rest of the paper, edges point from later nodes to earlier nodes. However, based on the steps in the proof, as well as the application of the lemma, one can see

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

that nodes in **u** precede their counterparts in **v**. See, for instance, Equations (17) and (23) in the proof in [20] and the application of the lemma at Equation (27) therein.

Now, we claim that

$$\prod_{v\in V_\Gamma} v^{p_\Gamma(v;\delta,m)} = \prod_{1\le k\le \ell} u_k^{x-1} v_k^{-x},$$

where the right-hand side is the expression used in [20] (we have changed the indexing variable there from l to k, and we have changed $\chi = \frac{m+\delta}{2m+\delta}$ there to x, which is indeed the same as χ by Equation (2)). To see this, consider the right-hand side. For each node v in Γ, it appears as an entry in **u** and in **v** for ${d^{\text{post}}}_\Gamma(v)$ times and ${d^{\text{pre}}}_\Gamma(v)$ times, respectively. Collecting factors shows that the exponent for v is $(x-1){d^{\text{post}}}_\Gamma(v) - x{d^{\text{pre}}}_\Gamma(v) = p_\Gamma(v;\delta,m)$.

Appendix B. Homology theory

In this section, we develop the theory of simplicial homology by defining all relevant topological terminology and stating all relevant topological facts. We mainly follow the exposition in [34] to minimize point-set topological technicalities. We refer the reader to [21] for an elementary introduction, and to Chapter 2 of [23] for a thorough exposition.

B.1. Simplicial maps, inclusion, and induced homomorphisms (cf. [34, Section 2])

A **simplicial map** $f : X \to Y$ between two simplicial complexes is a function between the vertex sets of the two complexes such that $\{f(v) : v \in \sigma\}$ is a simplex in Y for every simplex $\sigma \in X$.

If X is a subcomplex of Y, the **inclusion map** $j : X \to Y$ is the simplicial map defined by $j(v) = v$.

For every integer q, every simplicial map $f : X \to Y$ induces a homomorphism $f_\# : C_q(X) \to C_q(Y)$ defined as follows.

For each simplex $[v_0, \ldots, v_q]$, if $f(v_0), \ldots, f(v_q)$ are distinct, we denote by π the permutation such that $f(v_{\pi(0)}), \ldots, f(v_{\pi(q)})$ is increasing with respect to the ordering of vertices in Y. (Obviously π depends on $[v_0, \ldots, v_q]$.) We define

$$f_\#[v_0, \ldots, v_q] = \begin{cases} \operatorname{sgn}(\pi)[f(v_{\pi(0)}), \ldots, f(v_{\pi(q)})] & \text{if } f(v_0), \ldots, f(v_q) \text{ are distinct,} \\ 0 & \text{otherwise,} \end{cases}$$

where $\operatorname{sgn}(\pi) = \pm 1$ is the sign of the permutation π.

This map in turn **induces** a **homomorphism** $f_q : H_q(X) \to H_q(Y)$ defined by

$$f_q(z + \operatorname{im} \partial^X_{q+1}) = f_\#(z) + \operatorname{im} \partial^Y_{q+1},$$

for every q-cycle z of X, where the two superscripted im ∂_{q+1}'s are the boundary groups of X and Y respectively.

Induced homomorphisms are **functorial**, in the sense that

- $(\mathrm{id}_X)_q = \mathrm{id}_{H_q(X)}$ for every simplicial complex X, where id denotes the identity map or homomorphism, and
- $(gf)_q = g_q f_q$ for every integer q and every simplicial maps $f : X \to Y$ and $g : Y \to Z$.

(Cf. Theorem 12.2 of [34].)

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

B.2. Pairs and relative homology (cf. [34, Section 9])

A 2-tuple (X, A) of simplicial complexes is said to be a **pair** if A is a subcomplex of X, and a 3-tuple (Y, X, A) is said to be a **triple** if (Y, X) and (X, A) are two pairs. A simplicial complex X can be identified with the pair $(X, \emptyset)$, where $\emptyset$ denotes the empty set.

For a pair (X, A), the qth relative chain group is the quotient group $C_q(X)/C_q(A)$. Elements of relative chain groups are called relative chains. The boundary homomorphism $\partial_q : C_q(X) \to C_{q-1}(X)$ induces the relative boundary homomorphism $\partial_q : C_q(X, A) \to C_{q-1}(X, A)$, defined by $[v_0, \ldots, v_q] + C_q(A) \mapsto \partial[v_0, \ldots, v_q] + C_{q-1}(A)$. The qth **relative homology group** $H_q(X, A)$ is the quotient group $\ker \partial_q / \operatorname{im} \partial_{q+1}$, where the boundary homomorphisms here are relative boundary homomorphisms. Its rank is called the **(relative) Betti number** of the pair and is denoted by $\beta_q(X, A)$. The elements of the kernels and the images of relative boundary maps are called relative cycles and relative boundaries, respectively.

The interpretation of relative homology may be found in Proposition 2.22 of [23].

The definition of simplicial maps and induced homomorphisms extends to pairs, triples, and relative homology. We refer the reader to Section 12 of [34] for the relevant details. In particular, functoriality still holds, and we have the following lemma.

Lemma B.1. *If $(A, A') \subseteq (B \cap C, B' \cap C') \subseteq (D, D')$ (entrywise inclusion), then the following diagram commutes for every q:*

$$\begin{array}{ccc} H_q(A, A') & \longrightarrow & H_q(B, B') \\ \downarrow & & \downarrow \\ H_q(C, C') & \longrightarrow & H_q(D, D'), \end{array}$$

where all maps are induced by inclusion.

B.3. Exact sequences

An **exact sequence** is a sequence of abelian groups (A_n) with homomorphisms $\varphi_n : A_n \to A_{n+1}$ between adjacent abelian groups, such that $\ker \varphi_n = \operatorname{im} \varphi_{n-1}$. The sequence is said to be **finite-length** if it starts and ends with the trivial group 0. We used the following fact in our proofs.

Proposition B.2. (Alternating sum of ranks of an exact sequence [40, Exercise 3.16].) *Let $0 \to A_0 \to \ldots \to A_n \to 0$ be a finite-length exact sequence of finitely-generated abelian groups. Then*

$$\sum_{0 \le k \le n} (-1)^k \operatorname{rk} A_k = 0.$$

The case $n = 3$ gives the the rank–nullity theorem, which states that $\operatorname{rk} A = \operatorname{rk} \ker f + \operatorname{rk} \operatorname{im} f$ for every homomorphism $f : A \to B$ between finitely-generated abelian groups.

B.4. Properties of homology groups

Finally, we state three key homological facts that are used in our proofs.

Theorem B.3. (Theorem 25.1 of [34].) *Let X and Y be subcomplexes of a simplicial complex Z. Then there exist maps such that the following sequence is exact:*

$$\ldots \to H_q(X \cap Y) \to H_q(X) \oplus H_q(Y) \to H_q(X \cup Y) \to H_{q-1}(X \cap Y) \to \ldots \to H_0(X \cup Y) \to 0.$$

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

Theorem B.4. *(Long exact sequence for triples; Exercise 24.1 and Lemma 24.4 of [34].) For every triple (Y, X, A), there exists a homomorphism* $\partial_q : H_q(Y, X) \to H_{q-1}(X, A)$ *for each q such that the following sequence is exact:*

$$\ldots \to H_q(X, A) \to H_q(Y, A) \to H_q(Y, X) \xrightarrow{\partial_q} H_{q-1}(X, A) \to H_{q-1}(Y, A) \to \ldots \to H_0(Y, X) \xrightarrow{0} 0.$$

All unmarked maps are induced by inclusions. Further, the map ∂_q *is natural, in the following sense: for every simplicial map* $f : (Y, X, A) \to (Y', X', A')$ *and every q,* f_q *and* f_{q-1} *commute with the boundary maps* $\partial_q^{(Y,X,A)}$ *and* $\partial_q^{(Y',X',A')}$ *in the long exact sequences of (Y, X, A) and (Y', X', A'), i.e. the diagram*

$$\begin{array}{ccc} H_q(Y, X) & \xrightarrow{\partial_q^{(Y,X,A)}} & H_{q-1}(X, A) \\ \downarrow f_q & & \downarrow f_{q-1} \\ H_q(Y', X') & \xrightarrow{\partial_q^{(Y',X',A')}} & H_{q-1}(X', A') \end{array}$$

commutes, where the two vertical maps are restrictions of f to the respective domains.

Theorem B.5. *(Excision theorem for complexes; Theorem 9.1 of [34] and Corollary 2.24 of [23].) Let A and B be subcomplexes of a simplicial complex X. If* $X = A \cup B$*, then the inclusion* $j : (A, A \cap B) \to (X, B)$ *induces an isomorphism* $j_q : H_q(A, A \cap B) \to H_q(X, B)$ *for every q.*

Remark B.6. The above form of the excision theorem is a special case of Corollary 2.24 of [23], which is phrased in terms of the more general CW complexes. Theorem 9.1 of [34] implies our version if we put $K = X$, $K_0 = B$, and $U = X - A$, and hence $L = A$ and $L_0 = A \cap B$. We check that U is indeed an open set contained in $|K_0| = B$. Since $K = A \cup B$, $U = X - A \subseteq B$. Thus U is open because all subcomplexes are closed (cf. Lemma 2.2 of [34]).

We conclude with an elementary application of the long exact sequence for triples (Theorem B.4).

Lemma B.7. *Let* $q \geq 1$ *and let A be a subcomplex of X. If each of X and A is either empty or acyclic, then* $H_q(X, A) \cong 0$.

Proof. Consider the following segment of the long exact sequence for $(X, A, \emptyset)$:

$$H_q(X) \to H_q(X, A) \xrightarrow{\varphi} H_{q-1}(A) \xrightarrow{\psi} H_{q-1}(X).$$

The assumption on X implies that the first group is 0, and hence exactness implies that the second homomorphism $\varphi : H_q(X, A) \to H_{q-1}(A)$ is injective. It therefore suffices to show that $\operatorname{im} \varphi \cong 0$.

If $q \geq 2$ or A is empty, then $H_{q-1}(A)$ is also 0, and hence $\operatorname{im} \varphi$ must also be 0.

If $q = 1$ and A is nonempty, then the last homomorphism $\psi : H_{q-1}(A) \to H_{q-1}(X)$ in the segment above is the identity on $\mathbb{Z}$. (To see this, let $a \in A$. Then $H_0(A)$ and $H_0(X)$ are generated by the homology classes of a from the respective simplicial complexes.) Exactness then implies $\operatorname{im} \varphi = \ker \psi \cong 0$. □

Acknowledgements

The authors would like to thank Shmuel Weinberger, Jason Manning, Takashi Owada, Gesine Reinert, Andrew Thomas, Eduardo Paluzo and Benjamin Thompson for insightful

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

discussion. The authors also thank the anonymous editor and reviewers for their valuable comments and suggestions. The authors also thank Avhan Misra for editorial assistance.

Funding information

This research was partially supported by the AFOSR grant FA9550-22-1-0091 and by the Cornell University Center for Advanced Computing, which receives funding from Cornell University, the National Science Foundation, and members of its Partner Program.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] AKTAS, M. E., AKBAS, E. AND FATMAOUI, A. E. (2019). Persistence homology of networks: methods and applications. *Appl. Network Sci.* **4**, article no. 61.
[2] ALBERT, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957.
[3] AYZENBERG, A. AND RUKHOVICH, A. (2020). *Clique complexes of multigraphs, edge inflations, and tournaplexes*. Preprint. Available at https://arxiv.org/abs/2012.07600.
[4] BAUER, U. (2021). Ripser: efficient computation of Vietoris–Rips persistence barcodes. *J. Appl. Comput. Topol.* **5**, 391–423.
[5] BIANCONI, G. AND RAHMEDE, C. (2016). Network geometry with flavor: from complexity to quantum geometry. *Phys. Rev. E* **93**, article no. 032315.
[6] BOBROWSKI, O. AND KAHLE, M. (2018). Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.* **1**, 331–364.
[7] BOBROWSKI, O. AND KRIOUKOV, D. (2022). Random simplicial complexes: models and phenomena. In *Higher-Order Systems*, Springer, Cham, pp. 59–96.
[8] BOLLOBÁS, B. AND RIORDAN, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH, Weinheim, pp. 1–34.
[9] BOLLOBÁS, B., RIORDAN, O., SPENCER, J. AND TUSNÁDY, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18**, 279–290.
[10] CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* **46**, 255–308.
[11] CHAZAL, F. AND MICHEL, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers Artificial Intellig.* **4**, article no. 667963.
[12] COURTNEY, O. T. AND BIANCONI, G. (2017). Weighted growing simplicial complexes. *Phys. Rev. E* **95**, article no. 062301.
[13] CSARDI, G. AND NEPUSZ, T. (2006). The igraph software package for complex network research. *InterJournal* **1695**, 1–9.
[14] DUNCAN, P., KAHLE, M. AND SCHWEINHART, B. (2023). *Homological percolation on a torus: plaquettes and permutohedra*. Preprint. Available at https://arxiv.org/abs/2011.11903.
[15] EGGEMANN, N. AND NOBLE, S. (2011). The clustering coefficient of a scale-free random graph. *Discrete Appl. Math.* **159**, 953–965.
[16] ERDÖS, P. AND RÉNYI, A. (1959). On random graphs I. *Publ. Math. Debrecen* **6**, 290–297.
[17] FORMAN, R. (2002). A user's guide to discrete Morse theory. *Sém. Lotharingien Combinatoire* **48**, article no. B48c.
[18] GAL, S. R. (2005). Real root conjecture fails for five- and higher-dimensional spheres. *Discrete Comput. Geom.* **34**, 269–284.
[19] GARAVAGLIA, A. (2019). Preferential attachment models for dynamic networks. Doctoral Thesis, Technische Universiteit Eindhoven.
[20] GARAVAGLIA, A. AND STEGEHUIS, C. (2019). Subgraphs in preferential attachment models. *Adv. Appl. Prob.* **51**, 898–926.
[21] GIBLIN, P. (2010). *Graphs, Surfaces and Homology*, 3rd edn. Cambridge University Press.
[22] HARRIS, C. R. *et al.* (2020). Array programming with NumPy. *Nature* **585**, 357–362.
[23] HATCHER, A. (2002). *Algebraic Topology*. Cambridge University Press.
[24] HIRSCH, C. AND JUHASZ, P. (2023). *On the topology of higher-order age-dependent random connection models*. Preprint. Available at https://arxiv.org/abs/2309.11407.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press

[25] HOLME, P. AND KIM, B. J. (2002). Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65**, article no. 026107.
[26] HUNTER, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95.
[27] KAHLE, M. (2009). Topology of random clique complexes. *Discrete Math.* **309**, 1658–1671.
[28] KAHLE, M. (2011). Random geometric complexes. *Discrete Comput. Geom.* **45**, 553–573.
[29] KAHLE, M. (2014a). Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. Math.* **179**, 1085–1107.
[30] KAHLE, M. (2014b). Topology of random simplicial complexes: a survey. In *Algebraic Topology: Applications and New Directions*, American Mathematical Society, Providence, RI, pp. 201–222.
[31] KAHLE, M. AND MECKES, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology Homotopy Appl.* **15**, 343–374.
[32] LAM, S. K., PITROU, A, and SEIBERT, S. (2015). Numba: a LLVM-based Python JIT compiler. In *LLVM '15: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, Association for Computing Machinery, New York, article no. 7, 6 pp.
[33] MÓRI, T. (2002). On random trees. *Studia Sci. Math. Hung.* **39**, 143–155.
[34] MUNKRES, J. R. (1984). *Elements of Algebraic Topology*. Benjamin/Cummings, Menlo Park, CA.
[35] OH, S. M., LEE, Y., LEE, J. AND KAHNG, B. (2021). Emergence of Betti numbers in growing simplicial complexes: analytical solutions. *J. Statist. Mech.* 2021, article no. 083218.
[36] OSTROUMOVA, L., RYABCHENKO, A. AND SAMOSVAT, E. (2013). Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient. In *Algorithms and Models for the Web Graph*, eds A. BONATO, M. MITZENMACHER, and P. Praat, Springer, Cham, pp. 185–202.
[37] OSTROUMOVA PROKHORENKOVA, L. (2017). General results on preferential attachment and clustering coefficient. *Optimization Lett.* **11**, 279–298.
[38] PENROSE, M. (2003). *Random Geometric Graphs*. Oxford University Press.
[39] REIMANN, M. W. *et al.* (2017). Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers Comput. Neurosci.* **11**, article no. 48.
[40] ROTMAN, J. (2008). *An Introduction to Homological Algebra*. Springer, New York.
[41] SIU, C. (2024). *The topological behavior of preferential attachment graphs*. Preprint. Available at https://arxiv.org/abs/2406.17619.
[42] TRALIE, C., SAUL, N. AND BAR-ON, R. (2018). Ripser.py: a lean persistent homology library for python. *J. Open Source Software* **3**, article no. 925.
[43] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*, Vol. 1. Cambridge University Press.
[44] VAN DER HOFSTAD, R. (2024a). *Random Graphs and Complex Networks*, Vol. 2. Cambridge University Press.
[45] VAN DER HOFSTAD, R. (2024b). Corrigenda *Random Graphs and Complex Networks* Volume Two. Available at https://www.win.tue.nl/rhofstad/CorrigendaNotesRGCNII.pdf.
[46] VIRTANEN, P. *et al.* (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Meth.* 17, 261–272.
[47] YOGESHWARAN, D. AND ADLER, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann. Appl. Prob.* **25**, 3338–3380.

https://doi.org/10.1017/apr.2024.66 Published online by Cambridge University Press