Extremal Betti Numbers and Persistence in Flag Complexes

Lies Beers
Vrije Universiteit Amsterdam, The Netherlands

Magnus Bakke Botnan
Vrije Universiteit Amsterdam, The Netherlands

Abstract

We investigate several problems concerning extremal Betti numbers and persistence in filtrations of flag complexes. For graphs on n vertices, we show that $\beta_k(X(G))$ is maximal when $G = \mathcal{T}_{n,k+1}$, the Turán graph on $k+1$ partition classes, where $X(G)$ denotes the flag complex of G. Building on this, we construct an edgewise (one edge at a time) filtration $\mathcal{G} = G_1 \subseteq \cdots \subseteq \mathcal{T}_{n,k+1}$ for which $\beta_k(X(G_i))$ is maximal for all graphs on n vertices and i edges. Moreover, the persistence barcode $\mathcal{B}_k(X(G))$ achieves a maximal number of intervals, and total persistence, among all edgewise filtrations with $|E(\mathcal{T}_{n,k+1})|$ edges.

For $k = 1$, we consider edgewise filtrations of the complete graph K_n. We show that the maximal number of intervals in the persistence barcode is obtained precisely when $G_{\lceil n/2\rceil \cdot \lfloor n/2 \rfloor} = \mathcal{T}_{n,2}$. Among such filtrations, we characterize those achieving maximal total persistence. We further show that no filtration can optimize $\beta_1(X(G_i))$ for all i, and conjecture that our filtrations maximize the total persistence over all edgewise filtrations of K_n.

2012 ACM Subject Classification Mathematics of computing → Algebraic topology

Keywords and phrases Topological data analysis, Extremal graph theory

Digital Object Identifier 10.4230/LIPIcs.SoCG.2025.14

Related Version *Full Version*: `https://arxiv.org/abs/2502.21294` [3]

Acknowledgements We thank Ulrich Bauer for suggesting the problem of maximizing the number of intervals in the barcode of a Vietoris–Rips filtration.

1 Introduction

A central theme in topological data analysis (TDA) is the computation of homological invariants from simplicial complexes. These complexes are often constructed from point cloud data, with the *Vietoris–Rips* complex being one of the most widely used constructions. For a finite set P in a metric space (M, d), the Vietoris–Rips complex at scale r, denoted $\mathrm{VR}_r(P)$, includes all simplices $\sigma = \{p_0, \dots, p_n\}$ such that $d(p_i, p_j) \leq 2r$ for all $0 \leq i, j \leq n$. This complex is a *flag complex*, meaning it is the largest simplicial complex with a given underlying graph (the 1-skeleton). In particular, the edges in the connectivity graph of P at scale r fully determine the higher-dimensional simplices in $\mathrm{VR}_r(P)$. Varying the scale parameter r induces a filtration of simplicial complexes:

$$\mathrm{VR}(P)_{r_0} \hookrightarrow \mathrm{VR}(P)_{r_1} \hookrightarrow \cdots \hookrightarrow \mathrm{VR}(P)_{r_m},$$

where $r_0 < r_1 < \cdots < r_m$. Applying k-dimensional homology over a field $\mathbf{k}$ to this filtration yields a persistence barcode in degree k, denoted $\mathcal{B}_k(\mathrm{VR}(P))$. This barcode, consisting of intervals $[a, b)$, encodes the birth and death of topological features across scales and serves as a powerful tool for extracting topological information from P. When P is a sufficiently dense sampling of an underlying space X, the number of "long bars" in $\mathcal{B}_k(\mathrm{VR}(P))$ determines the k-th Betti number $\beta_k(X)$, which describes the k-dimensional topological features of X [7]. For a comprehensive introduction to topological data analysis, we refer the reader to [8].

© Lies Beers and Magnus Bakke Botnan;
licensed under Creative Commons License CC-BY 4.0

41st International Symposium on Computational Geometry (SoCG 2025).
Editors: Oswin Aichholzer and Haitao Wang; Article No. 14; pp. 14:1–14:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Given the central role of such filtrations in TDA, this paper addresses a fundamental question: how many topological features can arise in a data set of n points? Specifically, we investigate bounds on quantities such as the maximal value of $\beta_k(\mathrm{VR}_r(P))$, the maximal number of intervals in $\mathcal{B}_k(\mathrm{VR}(P))$, the length of the longest interval, and the sum of the lengths of the intervals (the total persistence).

▶ Remark 1. In this paper, we focus on edgewise filtrations of graphs, i.e., sequences of graphs $\mathcal{G} = G_1 \subseteq G_2 \subseteq \cdots \subseteq G_m = \{G_i\}_{i=1}^m$, on n vertices, where G_{i+1} and G_i differ by precisely one edge. Notably, every filtration of flag complexes can be realized as the Vietoris–Rips filtration of an appropriate metric on the vertices of G_m; see the full version of this paper [3, A]. Thus, our results are directly applicable to data analysis using persistent homology.

1.1 Overview and contributions

In Section 2, we introduce the relevant background material and compute the Betti numbers of flag complexes on Turán graphs.

In Section 3, we examine extremal Betti numbers and establish tight upper bounds on $\beta_k(F)$, where F is a flag complex on n vertices (Theorem 10). This upper bound is achieved by the Turán graph $\mathcal{T}_{n,k+1}$.

In Section 4, we study filtrations of flag complexes on n vertices with at most e edges, where e is the number of edges in $\mathcal{T}_{n,k+1}$. We prove that our filtration maximizes β_k at all filtration steps (Theorem 16).

In Section 5, we analyze the longest possible interval in the barcode of a filtration of flag complexes on n vertices (Corollary 21).

Additionally, in Section 6, we focus on degree 1 homology and show that any filtration of flag complexes on n vertices will achieve the maximal number of intervals in its degree 1 barcode if and only if the filtration contains $\mathcal{T}_{n,2}$ (Corollary 25).

In Section 7, we explore flag filtrations in homology degree 1 that achieve both a maximal number of intervals and maximal total persistence. Equivalently, we maximize the total persistence of all filtrations of graphs containing $G_{\lceil n/2\rceil \cdot \lfloor n/2 \rfloor} = \mathcal{T}_{n,2}$. Given the importance of Turán graphs in extremal graph theory, we believe that identifying extremal filtrations of Turán graphs is interesting in its own right. Our result relies on an elaborate combinatorial analysis that precisely classifies the extremal filtrations (Theorem 31).

Finally, the paper concludes with a discussion in which we outline several important conjectures for future work.

1.2 Related work

The study of extremal values of $\mathbb{Z}$-linear functions on the f- and β-vectors of simplicial complexes with n vertices has a rich history. Classical questions include determining the extremal Euler characteristic and the maximal sum of Betti numbers. These problems were addressed for general simplicial complexes in [4] and later extended to arbitrary $\mathbb{Z}$-linear functions in [11]. Specifically, [11, Theorem 3.4] shows that for flag complexes, extremal values are always realized by $\mathcal{T}_{n,k}$ for some k. Consequently, Theorem 10 follows directly from [11, Theorem 3.4].

Our proof of Theorem 10 adapts the approach of [1, Theorem 1.1], which gives an alternative proof for the extremal total Betti number of flag complexes. The key observation in our proof of Theorem 10 plays a central role in our results on extremal interval lengths (Section 5).

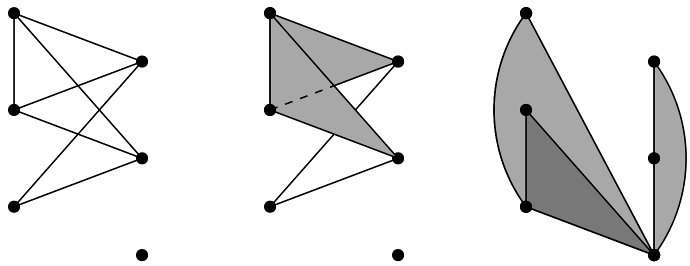


Figure 1 A graph (left), its flag complex (middle) and independence complex (right).

Another related work is [10], which examines asymptotic bounds for β_k in Vietoris–Rips complexes of point samples in $\mathbb{R}^d$. Similar questions have been explored for Čech complexes; see, e.g., [9].

Our work diverges from earlier work by considering *filtered* flag complexes. This approach is closely tied to the problem of finding extremal values for complexes with precisely n vertices and e edges; we return to this in Section 8. Importantly, and in contrast to classical work, graphs achieving extremal values need not have $\mathcal{T}_{n,k}$ as a subgraph.

2 Background

Graphs and Simplicial Complexes. Let $G = (V, E)$ be a graph. We write $\{v, w\}$ for the edge connecting the vertices v and w. For a vertex $v \in V$, the *neighborhood* is $N_G(v) := \{w \colon \{v, w\} \in E\}$, and the *closed neighborhood* is $N_G[v] := N_G(v) \cup \{v\}$. The *degree* of v is $\deg(v) := \deg_G(v) := |N_G(v)|$. The complete graph on n vertices is denoted K_n.

The *complement* of G, denoted $\overline{G}$, is the graph on the same vertex set as G, where two distinct vertices of $\overline{G}$ are adjacent if and only if they are non-adjacent in G. The *join* of disjoint graphs G_1 and G_2 is the graph $G_1 \vee G_2$ with vertex set $V(G_1) \sqcup V(G_2)$ and edge set $E(G_1) \cup E(G_2) \cup \{\{v_1, v_2\} \colon v_1 \in V(G_1), v_2 \in V(G_2)\}$. The *complete bipartite graph* K_{n_1,n_2} is defined as the join $G_1 \vee G_2$, where each G_i is an empty graph on n_i vertices. The sets $V(G_1)$ and $V(G_2)$ are called the *partition classes* of K_{n_1,n_2}.

For a vertex v in a simplicial complex K, the *link* and *(closed) star* of v are the simplicial complexes given respectively by

$$\mathrm{lk}_K(v) = \{\tau \in K : v \notin \tau, \{v\} \cup \tau \in K\} \qquad \mathrm{st}_K(v) = \{\tau \in K : \{v\} \cup \tau \in K\}.$$

We write $K - v$ for the simplicial complex with simplices $\tau \in K$ for which $v \notin K$.

For a graph G, we let $X(G)$ denote the simplicial complex with m-simplices given by the $(m + 1)$-cliques in G. A simplicial complex K is called a *flag complex* if $K = X(G)$ for some graph G. The *independence complex* of G is the simplicial complex $\mathrm{Ind}(G) = X(\overline{G})$. Examples of flag and independence complexes are given in Figure 1.

Homology. For a simplicial complex K, we let $\beta_k(K)$ denote the dimension of the reduced homology group $\tilde{H}_k(K; \mathbf{k})$, and let $C_k(K)$ denote the vector space of k-chains. We use coefficients in a fixed, but arbitrary, field $\mathbf{k}$; our optimal constructions are torsion-free and thus our results do not depend on the choice of coefficient field. For that reason, we simply write $\tilde{H}_k(K)$. We also employ the notation $\beta_k^{\mathrm{FL}}(G) = \beta_k(X(G))$ for a graph G.

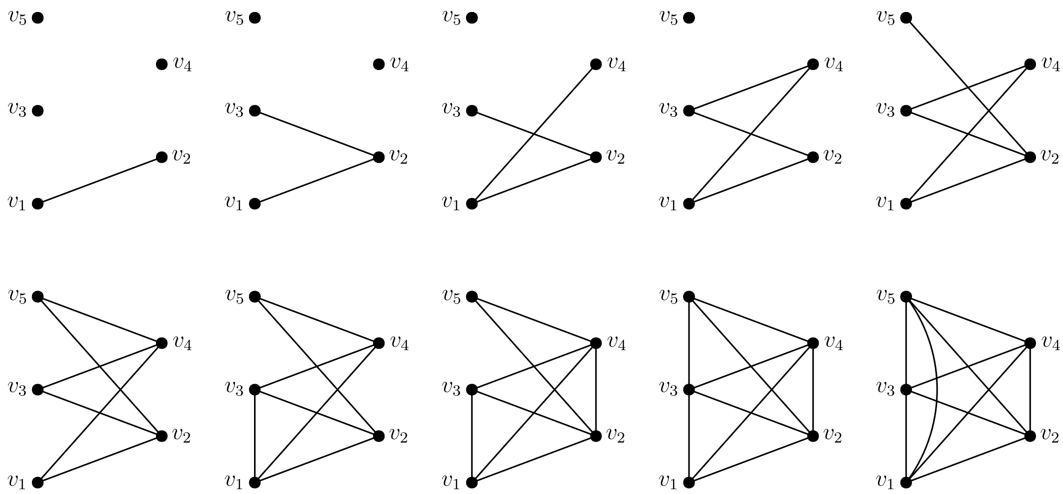


Figure 2 An edgewise filtration of K_5.

Persistent Homology. A filtration $\mathcal{K}$ of simplicial complexes is a collection of simplicial complexes $\{K_i\}_{i=1}^m$ such that $K_i \subseteq K_j$ for $i \leq j$. Applying $\tilde{H}_k(-;\mathbf{k})$ to a filtration yields a sequence of vector spaces and linear maps $\tilde{H}_k(\mathcal{K}) : \tilde{H}_k(K_1) \to \cdots \to \tilde{H}_k(K_m)$ called a *persistence module*. Provided all the vector spaces are finite-dimensional, $\tilde{H}_k(\mathcal{K})$ is uniquely described by a collection of intervals in $\{1, \ldots, m\}$, called the *degree k barcode of* $\mathcal{K}$. We shall denote this barcode by $\mathcal{B}_k(\mathcal{K})$. The *total persistence* of $\mathcal{K}$ is given by

$$T\beta_k(\mathcal{K}) = \sum_{[a,b)\in\mathcal{B}_k(\mathcal{K})} (b - a) = \sum_{i=1}^{m} \beta_k(K_i).$$

For a more thorough introduction to persistent homology, (generalized) persistence modules, and examples, see, e.g., [8, Chapter 3] or [5].

For a filtration $\mathcal{G}$ of graphs (1-dimensional simplicial complexes), we get a filtration $X(\mathcal{G})$ of simplicial complexes by taking the flag complex at every index. If $G_{i+1} - G_i$ is a single edge for all i, then we say that the filtration $\mathcal{G}$ is *edgewise*. We shall employ the notation $\mathcal{B}_k^{\mathrm{FL}}(G) = \mathcal{B}_k(X(\mathcal{G}))$.

▶ **Example 2.** An edgewise filtration of K_5 can be found in Figure 2.

2.1 Elementary homological properties

The following two lemmas are well-known and important in combinatorial topology. For completeness, their proofs can be found in the full version [3, B.1].

▶ **Lemma 3.** *Let K be a simplicial complex and $v \in V(K)$ a vertex. Then, for all $k \geq 1$,*

$$\beta_k(K) \leq \beta_k(K - v) + \beta_{k-1}(\mathrm{lk}_K(v)).$$

▶ **Lemma 4.** *For any two graphs G and H, and $k \geq -1$,*

$$\beta_k(\mathrm{Ind}(G \sqcup H)) = \sum_{i,j\geq -1; i+j=k-1} \beta_i(\mathrm{Ind}(G))\beta_j(\mathrm{Ind}(H)).$$

The following observation is essential for our work in Section 7.

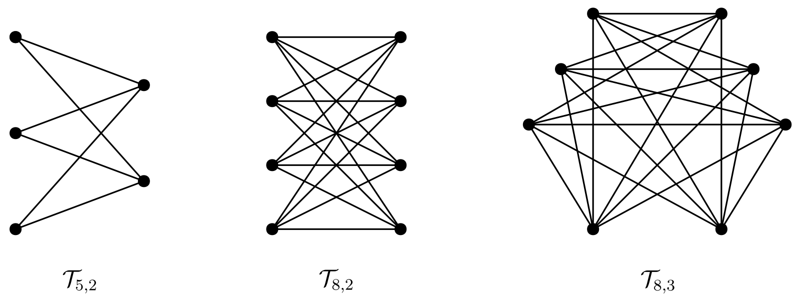


Figure 3 From left to right: the Turán graphs $\mathcal{T}_{5,2}$, $\mathcal{T}_{8,2}$ and $\mathcal{T}_{8,3}$.

▶ **Proposition 5.** *Let $V(G) = V_1 \sqcup V_2$ be the vertex set of a graph G containing all edges of the form $\{v, w\}$ where $v \in V_1$ and $w \in V_2$. Let $d_i \geq 1$ denote the number of connected components of G restricted to V_i. Then, $\beta_1^{\mathrm{FL}}(G) = \beta_0^{\mathrm{FL}}(G_1)\beta_0^{\mathrm{FL}}(G_2) = (d_1 - 1)(d_2 - 1)$.*

Proof. Let $\overline{G}$ denote the complement graph of G and observe that $\overline{G} = \overline{G}_1 \sqcup \overline{G}_2$ where G_1 and G_2 are the full subgraphs of G with vertices V_1 and V_2, respectively. From Lemma 4,

$$\beta_1^{\mathrm{FL}}(G) = \beta_1(\mathrm{Ind}(\overline{G})) = \beta_0(\mathrm{Ind}(\overline{G}_1))\beta_0(\mathrm{Ind}(\overline{G}_2)) = \beta_0(X(G_1))\beta_0(X(G_2)).$$ ◀

2.2 Turán graphs

▶ **Definition 6.** *Let $n \geq 0$ and $k \geq 1$, and let n' be the smallest positive integer such that $n' \equiv n \bmod k$. Then, the (n, k) graph $\mathcal{T}_{n,k}$ is the complement graph of the graph*

$$\overline{\mathcal{T}_{n,k}} = \bigsqcup_{i=1}^{k} K_{n_i}, \qquad n_i = \begin{cases} \lceil n/k \rceil & \text{if } 1 \leq i \leq n', \\ \lfloor n/k \rfloor & \text{otherwise.} \end{cases}$$

▶ **Example 7.** See Figure 3 for some examples of Turán graphs.

The proof of the following can be found in the full version [3, B.1].

▶ **Proposition 8.** *For all integers $n \geq 1$ and $k \geq 1$, we let n' be the smallest positive integer such that $n' \equiv n \bmod k$. We have*

$$\beta_i^{\mathrm{FL}}(\mathcal{T}_{n,k}) = \begin{cases} (\lceil n/k \rceil - 1)^{n'} \cdot (\lfloor n/k \rfloor - 1)^{k-n'} & \text{if } i = k-1, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, if n is a multiple of k, then $\beta_{k-1} = (n/k - 1)^k$.

3 The maximum value of $\beta_k(X(G))$

The following is well-known; see the full version [3, B.2] for a proof.

▶ **Lemma 9.** *Let S be a positive integer, and $x_i \geq 1$ be integers satisfying $\sum_{i=1}^{n} x_i = S$. Then, $\prod_{i=1}^{n}(x_i - 1) \leq \prod_{i=1}^{n}(y_i - 1)$ where*

$$y_i = \begin{cases} \lceil S/n \rceil & \text{if } 1 \leq i \leq S \bmod n \\ \lfloor S/n \rfloor & \text{otherwise.} \end{cases} \tag{1}$$

As said in Section 1.2, the following proof is a modification of the proof of [1, Theorem 1.1].

▶ **Theorem 10.** *Let G be a graph on n vertices. Then, for all $k \geq 0$,*

$$\beta_k^{\mathrm{FL}}(G) \leq (\lceil n/(k+1) \rceil - 1)^{(n \bmod k+1)} \cdot (\lfloor n/(k+1) \rfloor - 1)^{(k+1)-(n \bmod k+1)}.$$

Proof. We shall work inductively on n. The result is trivially true for $n = 1$, so assume that it holds for all $n' < n$.

Note that $X(G) = \mathrm{Ind}(\overline{G})$. Let d denote the minimal degree over all $v \in V(\overline{G})$. Assume that v is a vertex with $\deg(v) = d$, and let $\{v_1, \ldots, v_d\}$ denote the neighbours of v in $\overline{G}$. Moreover, let $\overline{G}_i = \overline{G} - \{v_1, \ldots, v_i\}$ for $i = 1, \ldots, d$ and $\overline{G}_0 = \overline{G}$. Applying Lemma 3,

$$\begin{aligned}
\beta_k(\mathrm{Ind}(\overline{G})) &\leq \beta_k(\mathrm{Ind}(\overline{G}_1)) + \beta_{k-1}(\mathrm{Ind}(\overline{G} - N_{\overline{G}}[v_1]) \\
&\leq \beta_k(\mathrm{Ind}(\overline{G}_2)) + \beta_{k-1}(\mathrm{Ind}(\overline{G}_1 - N_{\overline{G}_1}[v_2])) + \beta_{k-1}(\mathrm{Ind}(\overline{G} - N_{\overline{G}}[v_1]))) \\
&\;\;\vdots \\
&\leq \beta_k(\mathrm{Ind}(\overline{G}_d)) + \sum_{i=0}^{d-1} \beta_{k-1}(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}]))
\end{aligned}$$

Here $\overline{G}_d$ has an isolated vertex and thus $\mathrm{Ind}(\overline{G}_d)$ is a cone, and therefore $\beta_k(\mathrm{Ind}(\overline{G}_d)) = 0$. Since every vertex of $\overline{G}$ has degree at least d, it follows that $\overline{G}_i - N_{\overline{G}_i}(v_{i+1})$ contains at most $n - (d+1) = n - d - 1$ vertices. By the induction hypothesis it follows that

$$\beta_k(\mathrm{Ind}(\overline{G})) \leq d \cdot (x_1 - 1) \cdots (x_k - 1) = ((d+1) - 1) \cdot (x_1 - 1) \cdots (x_k - 1)$$

for integers $x_i \geq 1$ for which $(d+1) + \sum_{i=1}^{k} x_i = d + 1 + n - d - 1 = n$. Hence, by Lemma 9,

$$\beta_k(\mathrm{Ind}(\overline{G})) \leq (\lceil n/(k+1) \rceil - 1)^{(n \bmod k+1)} \cdot (\lfloor n/(k+1) \rfloor - 1)^{(k+1)-(n \bmod k+1)}. \qquad ◀$$

Let $\mathbf{G}(n) = \{G\colon \text{ graph on } n \text{ vertices}\}$. Combining Theorem 10 and Proposition 8, we have

▶ **Corollary 11.** *Of all graphs on n vertices, $\mathcal{T}_{n,k+1}$ maximizes the k-th Betti number. I.e.,*

$$\max_{G \in \mathbf{G}(n)} \beta_k^{\mathrm{FL}}(G) = \beta_k^{\mathrm{FL}}(\mathcal{T}_{n,k+1}).$$

In particular, if n is a multiple of k, then $\max_{G \in \mathbf{G}(n)} \beta_k^{\mathrm{FL}}(G) = (n/k - 1)^k$.

4 The maximum value of $\beta_k(X(G))$ for $|E(G)| \leq |E(\mathcal{T}_{n,k+1})|$

In this section, we fix $n, k \geq 1$, and the notation

$$e_{n,k+1} \coloneqq |E(\mathcal{T}_{n,k+1})| \text{ and } \Delta^{k+1}_{m-1,m} \coloneqq e^{k+1}_{m} - e^{k+1}_{m-1}. \tag{2}$$

The following result follows from combining Lemma 3 and Corollary 11, and will be integral in proving the main result of this section.

▶ **Corollary 12.** *Let v be a vertex of degree $d \geq 1$ in a graph G. Then,*

$$\beta_k^{\mathrm{FL}}(G) \leq \beta_k^{\mathrm{FL}}(G - v) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}).$$

Proof. If $K = X(G)$, then $K - v = X(G - v)$, and Lemma 3 implies that

$$\beta_k^{\mathrm{FL}}(G) \leq \beta_k^{\mathrm{FL}}(G - v) + \beta_{k-1}(\mathrm{lk}_{X(G)}(v)).$$

Now observe that $\mathrm{lk}_{X(G)}(v) = X(N_G[v])$, and since $\mathrm{lk}_{X(G)}(v)$ has d vertices, it follows from Corollary 11 that $\beta_{k-1}(\mathrm{lk}_K(v)) \leq \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k})$. ◀

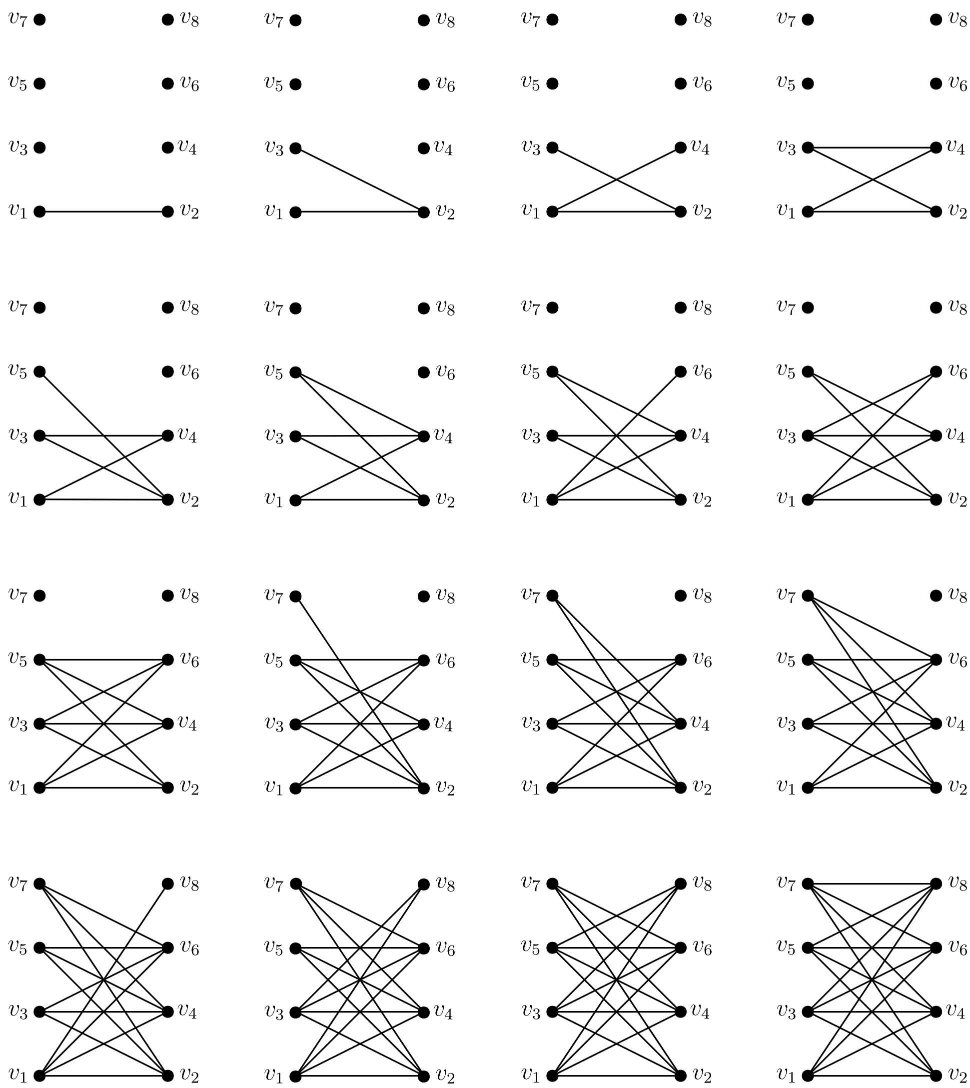


Figure 4 The filtration $\mathcal{H}$ of $\mathcal{T}_{8,2}$ from Definition 13.

We now define an edgewise filtration $\mathcal{H}^{n,k+1} = \mathcal{H} = \{H_i\}_{i=1}^{e_{n,k+1}} = \{H_i^{n,k+1}\}_{i=1}^{e_{n,k+1}}$ of the Turán graph $\mathcal{T}_{n,k+1}$. In this section, we shall show, for $m = 1, \dots, |\mathcal{T}_{n,k+1}|$, that $\beta_k^{\mathrm{FL}}(H_m)$ is maximal over all graphs with m edges, i.e., that $\mathcal{H}$ is *fiberwise optimal.*

▶ **Definition 13.** *Let V denote the vertices of $\mathcal{T}_{n,k+1}$ and label the elements of V such that v_i and v_{i+k+1} are in the same partition. Writing each edge as $e_{i,j} = \{v_i, v_j\}$ for $i < j$, we order the edges by the following co-lexicographic order: $e_{i,j} < e_{k,l}$ if $\max\{i,j\} < \max\{k,l\}$, or if $i = k$ and $j < l$, or if $j = l$ and $i < k$. Following this order, let $H_i = H_i^{n,k+1}$ denote the subgraph of $\mathcal{T}_{n,k+1}$ with the first i edges.*

Note that, since $\mathcal{T}_{n,k+1}$ contains no $k+1$-simplices, $\beta_k(H_i)$ is increasing as a function of i.

▶ **Example 14.** See Figure 2 for an edgewise filtration $\mathcal{G} = \{G_i\}_{i=1}^{6}$ of K_5, where $G_i \cong H_i^{5,2}$ for $i = 1, \dots, 6$. Furthermore, in Figure 4 one can find the filtration $\mathcal{H}^{8,2}$ of $\mathcal{T}_{8,2}$.

$$\begin{tikzcd}
\boxed{\mathcal{T}_{m-2,k}} \arrow[r, "+d'"] \arrow[rr, bend right, "+\Delta_{m-2,m-1}"'] & \boxed{G-\{v\}} \arrow[r, "+\widehat{e}"] \arrow[rr, bend right, "+d"'] & \boxed{\mathcal{T}_{m-1,k}} \arrow[r, "+d-\widehat{e}"] & \boxed{G} & \boxed{\mathcal{T}_{m,k}}
\end{tikzcd}$$

Figure 5 An illustration of Case 2 from the proof of Theorem 16.

The next lemma shows that, once a vertex has been connected to the growing component, and until the next vertex gets connected, the change in β_k^{FL} from adding a single edge increases with the number of edges already added. Its proof, similar to that of Corollary 12, can be found in the full version [3, B.3].

▶ **Lemma 15.** *For e edges, let m be maximal such that $\mathcal{T}_{m,k+1}$ is a subgraph of H_e. Then,*

$$\beta_k^{\mathrm{FL}}(H_e) = \beta_k^{\mathrm{FL}}(\mathcal{T}_{m,k+1}) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{e-e_m^{k+1},k}).$$

In particular, for $e_m^{k+1} \leq e_1 < e_2 < e_{m+1}^{k+1}$, we have that

$$\beta_k^{\mathrm{FL}}(H_{e_1+1}) - \beta_k^{\mathrm{FL}}(H_{e_1}) \leq \beta_k^{\mathrm{FL}}(H_{e_2+1}) - \beta_k^{\mathrm{FL}}(H_{e_2}).$$

▶ **Theorem 16.** *Let $e \leq e_{n,k+1}$, let H_e be as above, and let G be any other graph on n vertices and e edges. Then, $\beta_k^{\mathrm{FL}}(G) \leq \beta_k^{\mathrm{FL}}(H_e)$.*

Proof. We prove this inductively on the number of edges e. The statement is clearly true for $e = 1$, so let's assume that the result holds for all $e' < e$.

Let m be the number of vertices in H_e with positive degree, and let n' denote the number of vertices in G with positive degree. If $n' < m$, then by Theorem 10,

$$\beta_k^{\mathrm{FL}}(G) \leq \beta_k^{\mathrm{FL}}(\mathcal{T}_{m-1,k+1}) \leq \beta_k^{\mathrm{FL}}(H_e).$$

Hence, we may assume that $n' \geq m$. In particular, the average degree of the positive-degree-vertices in G is no larger than the average degree of positive-degree-vertices in H_e.

Choose a vertex v in G with minimal positive degree d, and observe that

$$d \leq e_m^{k+1} - e_{m-1}^{k+1} = \Delta_{m-1,m},$$

since $\mathcal{T}_{m-1,k+1} \subsetneq H_e \subseteq \mathcal{T}_{m,k+1}$. In fact, if $d = \Delta_{m-1,m}$, then we must have that the average degree in H_e is at least d. Importantly, this happens if and only if $H_e = \mathcal{T}_{m,k+1}$ and m is a multiple of $k+1$.

By the induction assumption,

$$\begin{aligned}\beta_k^{\mathrm{FL}}(G-v) &\leq \beta_k^{\mathrm{FL}}(H_{e-\deg(v)}), \text{ and thus, by Corollary 12,} \\ \beta_k^{\mathrm{FL}}(G) &\leq \beta_k^{\mathrm{FL}}(H_{e-\deg(v)}) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}).\end{aligned}$$

If $d = \Delta_{m-1,m}$, then this becomes, by Lemma 15,

$$\beta_k^{\mathrm{FL}}(G) \leq \beta_k^{\mathrm{FL}}(\mathcal{T}_{m-1,k+1})) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}) = \beta_k^{\mathrm{FL}}(\mathcal{T}_{m,k+1}) = \beta_k^{\mathrm{FL}}(H_e).$$

We may therefore assume that $d \leq \Delta_{m-1,m} - 1 \leq \Delta_{m-2,m-1}$.

Let $\hat{e}$ denote the number of edges that is added to H_{e-d} before a new vertex gets positive degree in the filtration $\mathcal{H}$. Let d' be the degree of the last vertex in H_{e-d} that obtained positive degree in $\mathcal{H}$. We consider two cases.

- **Case 1:** $\mathcal{T}_{m-1,k} \subseteq H_{e-d}$. Then,

$$\begin{aligned}\beta_k^{\mathrm{FL}}(G) &\leq \beta_k^{\mathrm{FL}}(H_{e-d}) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}) \\ &= \beta_k^{\mathrm{FL}}(H_{e-d}) + \beta_k^{\mathrm{FL}}(H_{e_m^{k+1}+d}) - \beta_k^{\mathrm{FL}}(H_{e_m^{k+1}}) \\ &\leq \beta_k^{\mathrm{FL}}(H_{e-d}) + \beta_k^{\mathrm{FL}}(H_e) - \beta_k^{\mathrm{FL}}(H_{e-d}) \\ &= \beta_k^{\mathrm{FL}}(H_e).\end{aligned}$$

- **Case 2:** $\mathcal{T}_{m-2,k} \subseteq H_{e-d} \subset \mathcal{T}_{m-1,k}$. Note that $d' + \hat{e} = \Delta_{m-2,m-1} \geq d$ and $d - \hat{e} \geq 0$ (see Figure 5). In particular,

$$\begin{aligned}\beta_k^{\mathrm{FL}}(G) &\leq \beta_k^{\mathrm{FL}}(H_{e-d}) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}) \\ &= \beta_k^{\mathrm{FL}}(H_{e-d}) + \big(\beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d,k}) - \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d-\hat{e},k})\big) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d-\hat{e},k}) \\ &\leq \beta_k^{\mathrm{FL}}(H_{e-d}) + \big(\beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d'+\hat{e},k}) - \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d',k})\big) + \beta_{k-1}^{\mathrm{FL}}(\mathcal{T}_{d-\hat{e},k}) \\ &= \beta_k^{\mathrm{FL}}(H_e).\end{aligned}$$

In both cases, the inequality follows from Corollary 12 and Lemma 15. ◀

5 Tight bounds on the vanishing of homology and extremal interval lengths

For a graph G with n vertices, it is guaranteed that there exists a vertex of degree $n-1$ when the average degree exceeds $n-2$. Specifically, if the number of edges satisfies $m > \frac{n(n-2)}{2}$, then $X(G)$ must be a cone, implying that $\beta_k^{\mathrm{FL}}(G) = 0$ for all k. In practical scenarios, however, the primary interest is with $\beta_k^{\mathrm{FL}}(G)$ for small k. In this section, we provide tight bounds for the vanishing of $\beta_k^{\mathrm{FL}}(G)$ for a fixed k.

▶ **Lemma 17.** *Let G be a graph with minimum degree u. Let v be a vertex of degree u and let $N_G(v) = \{v_1, \ldots, v_u\}$. Let $d_i \coloneqq \deg(v_i)$. If G has n vertices and m edges, then*

$$\big|V\big(G - N_G[v_i]\big)\big| = n - d_i - 1 \text{ and } \big|E\big(G - N_G[v_i]\big)\big| \leq m - \left(d_i + \left\lceil \frac{(u-1)d_i}{2} \right\rceil\right).$$

Proof. Let $\widehat{G} \coloneqq G - N_G[v_i]$. The fact that $\big|V\big(\widehat{G}\big)\big| = n - d_i - 1$ is trivial. For showing the inequality, note that, by removing $N_G[v_i]$ from G, we remove the d_i edges containing v_i, and all edges containing the vertices of $N_G(v_i)$. Those vertices all have degree at least u, which means that they have at least $u-1$ neighbors apart from v_i. Because it might be the case that $N_G\big(N_G[v_i]\big) = N_G[v_i]$, it follows that

$$\big|E\big(\widehat{G}\big)\big| \leq m - \left(d_i + \left\lceil \frac{(u-1)d_i}{2} \right\rceil\right). \qquad ◀$$

▶ **Theorem 18.** *Let G be a graph with n vertices and $m > \binom{n-1}{2} + k$ edges. Then $\beta_k^{\mathrm{FL}}(G) = 0$.*

Proof. Let us first consider the case $k = 0$. This case is immediate from the fact that the maximal number of edges in a graph on n vertices with an isolated vertex is $\binom{n-1}{2}$.

Working inductively on k, assume that result holds for all $k' < k$, and that $m > \binom{n-1}{2} + k$. Let $\overline{G}$ be the complement graph of G, and note that the number of edges $\overline{m} = |E(\overline{G})|$ satisfies

$$\overline{m} = \binom{n}{2} - m < \binom{n}{2} - \binom{n-1}{2} - k = n - 1 - k.$$

From the proof of Theorem 10, we have that

$$\beta_k^{\mathrm{FL}}(G) = \beta_k(\mathrm{Ind}(\overline{G})) \leq \sum_{i=0}^{d-1} \beta_{k-1}(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}])),$$

where $\overline{G}_i = \overline{G} - \{v_1, \ldots, v_i\}$, and $\{v_1, \ldots, v_u\}$ are the neighbors of a vertex $v \in V(\overline{G})$ with minimal degree u. We shall show that all the terms in the sum are zero.

If we let $u' \geq u - i$ denote the minimal degree of a vertex in $V(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}])$,

$$\begin{aligned}
&\left|E\left(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}]\right)\right| \\
&\overset{(i)}{\leq} |E(\overline{G}_i)| - \left(\deg_{\overline{G}_i}(v_{i+1}) + \left\lceil \frac{(u'-1)\deg_{\overline{G}_i}(v_{i+1})}{2} \right\rceil\right) \\
&\overset{(ii)}{\leq} |E(\overline{G}_i)| - \left(u - i + \left\lceil \frac{(u-i-1)(u-i)}{2} \right\rceil\right) = |E(\overline{G}_i)| - \sum_{j=1}^{u-i} j \\
&\overset{(iii)}{\leq} |E(\overline{G})| - \sum_{j=u-i+1}^{u} j - \sum_{j=1}^{u-i} j = |E(\overline{G})| - \sum_{j=1}^{u} j = \overline{m} - \left(\frac{u(u+1)}{2}\right).
\end{aligned}$$

Here, (i) follows from Lemma 17, (ii) from $\deg_{\overline{G}_i}(v_i) \geq u'$, and (iii) from the fact that $\overline{G}_i$ is obtained by removing i vertices from $\overline{G}$ with degree at least u in $\overline{G}$. Now write

$$n' = |V(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}]))| \leq n - u - 1,$$

and observe that

$$|E(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}]))| \geq \binom{n'}{2} - \overline{m} + u(u+1)/2.$$

It follows that,

$$\begin{aligned}
&|E(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}]))| - \binom{n'-1}{2} \\
&\geq \binom{n'}{2} - \overline{m} + u(u+1)/2 - \binom{n'-1}{2} \\
&= n' - 1 - \overline{m} + u(u+1)/2 > n' - 1 - (n-1-k) + u(u+1)/2 \geq k + u(u+1)/2 \geq k.
\end{aligned}$$

Hence, $\beta_{k-1}(\mathrm{Ind}(\overline{G}_i - N_{\overline{G}_i}[v_{i+1}])) = 0$ by the induction hypothesis. ◀

▶ Remark 19. Observe that the difference between the bound for a cone and the bound given in Theorem 18 is $n/2 - k - 1$. While this difference is linear in the number of vertices, it can result in a significant reduction in the number of higher-dimensional simplices; a reduction which has the potential to speed up current implementations of persistent homology for flag complexes, e.g., Ripser [2].

In the following example, we show that that bounds in the previous theorem are tight.

▶ **Example 20.** Let $n = (k+1)p$, and let $K_{1,p-1}$ be a star graph on p vertices and $p-1$ edges, i.e., a central vertex connected to all other vertices. Observe that $\beta_0(\mathrm{Ind}(K_{1,p-1})) = 1$ as one vertex is completely disconnected in the complement graph. If, $H = \bigsqcup_{i=1}^{k+1} K_{1,p-1}$, then it follows from repeated application of Lemma 4 that $\beta_k(\mathrm{Ind}(H)) \geq \prod_{i=1}^{k+1} \beta_0(\mathrm{Ind}(K_{1,p-1})) = 1$. The number of edges in $\mathrm{Ind}(H)$ is

$$\binom{n}{2} - (k+1)(p-1) = \binom{n}{2} - n + (k+1) = \binom{n}{2} - (n-1) + k = \binom{n-1}{2} + k.$$

▶ **Corollary 21.** *Let $\mathcal{G}$ be an edgewise filtration on n vertices. Then for any $[a,b) \in \mathcal{B}_k^{\mathrm{FL}}(\mathcal{G})$, we have*

$$2k(k+1) \leq a < b \leq \binom{n-1}{2} + k.$$

Proof. The bound on a follows from Theorem 16, and the bound on b from Theorem 18. ◀

It is straightforward to define a filtration $\mathcal{F}$ of the complement graph of H from Example 20 such that $\mathcal{B}_k^{\mathrm{FL}}(\mathcal{F})$ contains the interval $[a,b)$ from Corollary 21.

6 The maximum value of $|\beta_1^{\mathrm{FL}}(\mathcal{G})|$

▶ **Proposition 22.** *Let $\mathcal{G} = \{G_i\}_{i=1}^m$ be an edgewise filtration. Then, there exists a triangle-free subgraph H of G_m such that $\beta_1^{\mathrm{FL}}(H) \geq |\mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})|$.*

Proof. Choose a representative cycle $c_{[a,b]}$ for each non-empty interval $[a,b) \in \mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})$, e.g., by running the standard algorithm for persistent homology. For a cycle c, let $e(c)$ denote the set of edges on which the cycle has a non-zero coefficient, and let H be the subgraph of G_m with edges given by

$$\bigcup_{[a,b)\in\mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})} e(c_{[a,b)]}).$$

Note that the cycles $\{c_{[a,b)} : [a,b) \in \mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})\}$ are linearly independent (as elements of $C_1(G)$) by virtue of being representatives for non-trivial intervals.

If $X(H)$ contains a 2-simplex $\tau = \{v_1, v_2, v_3\}$, then any cycle $c_{[a,b)}$ supported on the edge $\{v_2, v_3\}$ is homologous in $\tilde{H}_1(X(H))$ to the cycle $c'_{[a,b)} = c_{[a,b)} - \partial_2(\tau)$. Hence, we can remove $\{v_2, v_3\}$ from H, without reducing the 1st Betti number. Doing this for all triangles, we end up with a triangle-free graph $\hat{H}$ containing at least $|\mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})|$ linearly independent 1-cycles. In particular, $\beta_1^{\mathrm{FL}}(\hat{H}) \geq |\mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})|$. ◀

▶ Remark 23. This argument does not apply to homology in higher degrees. For instance, the removal of an edge can result in the removal of many 2-simplices, some of which are faces of 3-simplices, and others that are not.

The previous result actually gives a short proof of Corollary 11 for the case $k = 1$. We shall include this proof here, as it ensures uniqueness of the extremal complex.

▶ **Proposition 24.** *For any graph on n vertices, we have $\beta_1^{\mathrm{FL}}(G) \leq \beta_1^{\mathrm{FL}}(\mathcal{T}_{n,2})$ with equality if and only if $G = \mathcal{T}_{n,2}$.*

Proof. Define any edgewise filtration $\mathcal{G} = \{G_i\}_{i=1}^m$ for which $G_m = G$. Then, by Proposition 22, there exists a triangle-free graph H such that $\beta_1^{\mathrm{FL}}(H) \geq \mathcal{B}_1^{\mathrm{FL}}(\mathcal{G}) \geq \beta_1^{\mathrm{FL}}(G)$. Hence, it suffices to maximize $\beta_1^{\mathrm{FL}}(G)$ over triangle-free graphs. By the Euler-Poincaré formula,

$$\beta_1^{\mathrm{FL}}(G) = -\beta_0^{\mathrm{FL}}(G) - 1 - |V(G)| + |E(G)|.$$

In conclusion, we are seeking a triangle-free graph with a maximal number of edges, but this is well-known to be uniquely the graph $\mathcal{T}_{n,2}$ by Turán's theorem [6, Theorem 11.17]. ◀

Combined we arrive at the following.

▶ **Corollary 25.** *If $\mathcal{G}$ is a filtration of K_n, then $|\mathcal{B}_1^{\mathrm{FL}}(\mathcal{G})| \leq \beta_1^{\mathrm{FL}}(\mathcal{T}_{n,2})$ with equality if and only if $G_{\lceil n/2\rceil\cdot\lfloor n/2\rfloor} = \mathcal{T}_{n,2}$.*

7 Extremal filtrations for degree 1 persistent homology

In this section, we consider edgewise filtrations $\mathcal{G} = G_1 \subseteq \cdots \subseteq G_{\binom{n}{2}}$ of the complete graph K_n that maximize the total persistence and the number of bars for degree 1 homology. We shall answer the following question:

For a fixed number of vertices n and for $k = 1$, which edgewise filtrations of the complete graph K_n with a maximal number of bars achieve maximal total persistence?

By Corollary 25, this translates to finding

$$\max\left\{ \sum_{i=1}^{n(n-1)/2} \beta_1^{\mathrm{FL}}(G_i) \colon \mathcal{G} = \left\{G_i \colon 1 \leq i \leq \binom{n}{2}, G_{|E(\mathcal{T}_{n,2})|} = \mathcal{T}_{n,2} \text{ and } G_{n(n-1)/2} = K_n \right\} \right\}.$$

First, we give some preliminaries and establish some notation.

- All graphs $G = (V, E)$ have n vertices. The join of two graphs G_1 and G_2, notation $G_1 \vee G_2$, is as in Section 2.
- For a filtration $\mathcal{G} = \{G_i\}_{i=1}^m$, we write $T\beta_1^{\mathrm{FL}}(\mathcal{G}) = \sum_{i=1}^m \beta_1^{\mathrm{FL}}(G_i)$ for its total persistence of degree 1.
- Denote $e_n \coloneqq |E(\mathcal{T}_{n,2})|$. If $|E(G)| \geq e_n$, then G has $\mathcal{T}_{n,2}$ as a subgraph, with partition classes V_1 and V_2 of sizes $|V_1| = \lceil n/2 \rceil$ and $|V_2| = \lfloor n/2 \rfloor$.
- Let d_i be the number of connected components of $G|_{V_i}$. Then, by Proposition 5,

$$\beta_1^{\mathrm{FL}}(G) = (d_1 - 1)(d_2 - 1). \tag{3}$$

We also define what it means for two graphs and two filtrations to be (fiberwise) isomorphic.

▶ **Definition 26.** *Two graphs G_1 and G_2 with vertex sets $V(G_i)$ and edge sets $E(G_i)$ are* isomorphic, *notation $G_1 \cong G_2$, if there is a bijection $\varphi \colon V(G_1) \to V(G_2)$, such that $\{u_1, u_2\} \in E(G_1)$ if and only if $\{\varphi(u_1), \varphi(u_2)\} \in E(G_2)$. Furthermore, two filtrations $\mathcal{G}^1 = \{G_i^1\}_{i=1}^m$ and $\mathcal{G}^2 = \{G_i^2\}_{i=1}^m$ are* fiberwise isomorphic *if $G_i^1 \cong G_i^2$ for all $i = 1, \ldots, m$.*

First, by Corollary 25, any filtration with a maximal number of bars must contain the Turán graph $\mathcal{T}_{n,2}$. Moreover, Theorem 16 establishes that the filtration $\mathcal{H} = \{H_i\}_{i=1}^{e_n}$ from Definition 13 is fiberwise optimal, reducing our problem to maximizing the total persistence of an edgewise filtration of the complete graph, beginning with the Turán graph plus one edge. This is a bit more involved, because we cannot find a fiberwise optimal filtration of the complete graph, as is shown in the following example.

▶ **Example 27.** If $|E| = e_n + 2$, then one gets optimal degree 1 homology by adding one edge to V_1 and one edge to V_2 (left graph in Figure 6). However, if $|E| = e_n + 3$, then optimal degree 1 homology is obtained by adding all three edges to V_1 to form a K_3 (second-to-left graph in Figure 6). It is clear that a filtration of K_n cannot contain both graphs.

We can prove the following about the structure of the graphs in an optimal filtration of K_n.

▶ **Lemma 28.** *Let $\mathcal{G}$ be an edgewise filtration $\{G_i\}_{i=1}^{\binom{n}{2}}$ of K_n with maximal total persistence that includes $\mathcal{T}_{n,2}$. Then, for $t = 1, 2$ and $i = e_n + 1, \ldots, \binom{n-1}{2} + 1$, one of the subgraphs $G_i|_{V_t}$ consists of isolated vertices and a single component K_m, and the other subgraph $G_i|_{V_t}$ consists of isolated vertices, and a single component with l vertices, such that K_{l-1} is a subgraph of this component.*

▶ **Example 29.** Cf. Figure 6 for four examples of graphs with the properties from Lemma 28.

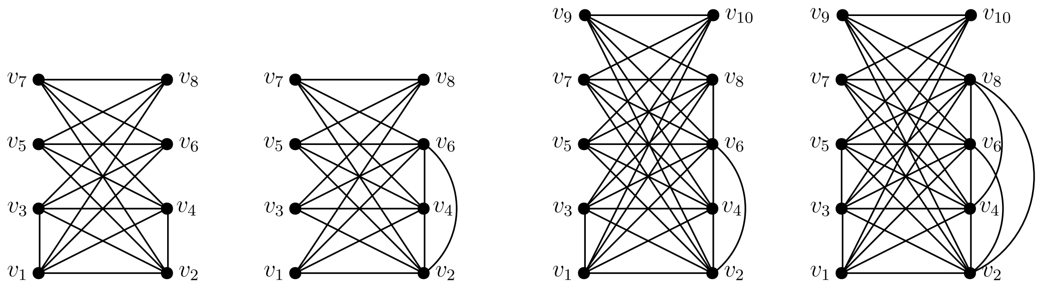


Figure 6 The graphs from Example 27 and 29.

Proof. Suppose that for some i, $E(G_i) \setminus E(G_{i-1}) = \{w_1, w_2\}$, where w_1 and w_2 are isolated vertices in $G_{j-1}|_{V_1}$ (without loss of generality). Also assume that $G_{j-1}|_{V_1}$ contains a component C of size $|C| > 1$. To maximize total persistence, it is more optimal to connect w_1 to C, since this allows us to add $|C| - 1$ edges of the form $\{w_1, c\}$ (where $c \in C$) without increasing the number of connected components (cf. (3)). Because these additional edges do not decrease the degree 1 homology, we should add them as early in the filtration as possible to maximize total persistence. Hence, by this same argument, $G_{j-1}|_{V_2}$ should be isomorphic to K_m and some isolated vertices.

Furthermore, note that we added the restriction $j \leq \binom{n-1}{2} + 1$ because Theorem 18 ensures that beyond this point, $\beta_1^{\mathrm{FL}}(G_j) = 0$. This means that

$$G_{\binom{n-1}{2}+1} = (K_1 \sqcup K_{\lceil n/2 \rceil - 1}) \vee (K_1 \sqcup K_{\lfloor n/2 \rfloor - 1}), \tag{4}$$

but if $j > \binom{n-1}{2} + 1$, then G_j could be any graph containing $G_{\binom{n-1}{2}+1}$ as a subgraph. ◀

Because of this, we restrict our focus to edgewise filtrations of the graph $G_{\binom{n-1}{2}+1}$ from (4).

Moreover, by Lemma 28, we only have to consider filtrations of the described form. We can represent them by a sequence of tuples of the form

$$(K_1, K_1) = (K_{l_1}, K_{r_1}), (K_{l_2}, K_{r_2}), \ldots, (K_{l_c}, K_{r_c}) = (K_{\lceil n/2 \rceil - 1}, K_{\lfloor n/2 \rfloor - 1}), \tag{5}$$

such that $l_i \geq l_{i-1}$ and $r_i \geq r_{i-1}$. The representation from (5) corresponds to the filtration $\mathcal{G} = \{G_i\}_{i=1}^{m}$, where, for $i \geq 1$,

$$G_{e_n + |E(K_{l_i})| + |E(K_{r_i})|} \cong \left(K_{l_i} \sqcup \bigsqcup_{j=l_i+1}^{\lceil n/2 \rceil} K_1\right) \bigvee \left(K_{r_i} \sqcup \bigsqcup_{j=r_i+1}^{\lfloor n/2 \rfloor} K_1\right) \text{ and}$$

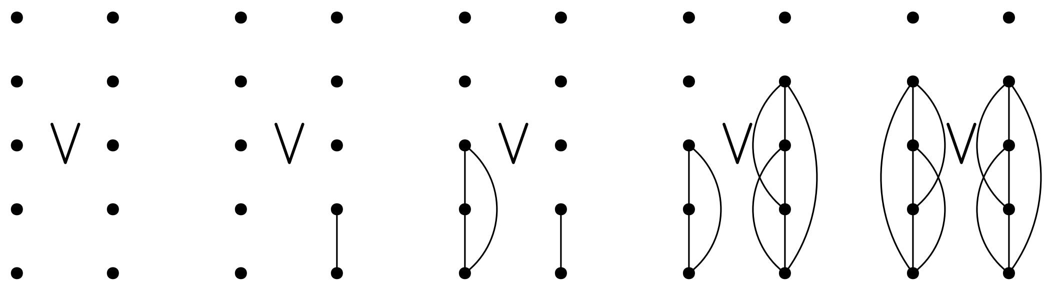
$$G_{e_n + |E(K_{l_{i+1}})| + |E(K_{r_i})|} \cong \left(K_{l_{i+1}} \sqcup \bigsqcup_{j=l_i+1}^{\lceil n/2 \rceil} K_1\right) \bigvee \left(K_{r_i} \sqcup \bigsqcup_{j=r_i+1}^{\lfloor n/2 \rfloor} K_1\right).$$

In other words, the representation from (5) means that, starting from the Turán graph $\mathcal{T}_{n,2}$:

1. we begin adding edges by forming a K_{l_2} in V_1,
2. then add edges to form a K_{r_2} in V_2,
3. increase the sizes of the complete subgraphs in V_1 and V_2 by constructing a K_{l_3} (containing the smaller K_{l_2}) in V_1, a K_{r_3} in V_2, and so on.

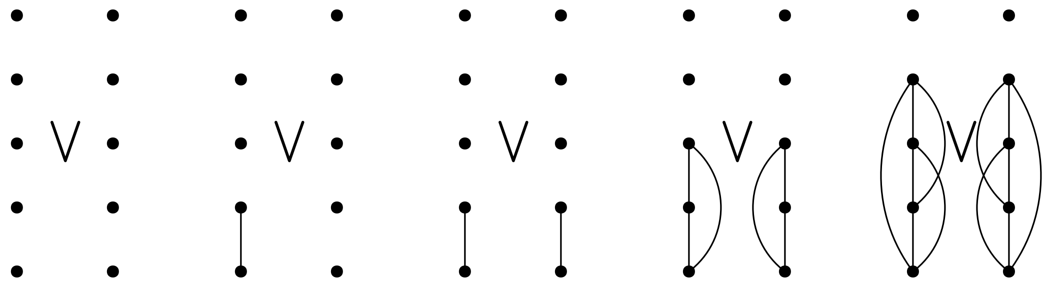
Note that this representation is not unique ((K_3, K_1) is equivalent to $(K_2, K_1), (K_3, K_1)$).

▶ **Example 30.** In Figure 7, one finds a filtration and a corresponding representation.



$G_{25} = \mathcal{T}_{10,2}$ (K_1, K_1) | G_{26} (K_1, K_2) | G_{29} (K_3, K_2) | G_{34} (K_3, K_4) | G_{37} (K_4, K_4)

Figure 7 A filtration and its representation. Here, $\bigvee$ denotes the join of the left and right graphs.



$G'_{25} = \mathcal{T}_{10,2}$ (K_1, K_1) | G'_{26} (K_2, K_1) | G'_{27} (K_2, K_2) | G'_{31} (K_3, K_3) | G'_{37} (K_4, K_4)

Figure 8 A filtration with the same total persistence as the one from Figure 7.

Surprisingly, the optimal strategy is to first construct two large cliques in V_1 and V_2 of sizes approximately $\frac{3}{4}|V_1|$ and $\frac{3}{4}|V_2|$, respectively, and then, alternating between V_1 and V_2, increase the sizes of the cliques one by one. We prove this in the main result of this section.

▶ **Theorem 31.** *Let $n \geq 4$ and let $j_n := \lfloor (3n-2)/8 \rfloor$ and $k_n := \lfloor (3n+7)/8 \rfloor$. Up to fiberwise isomorphism, the edgewise filtration $\mathcal{G}_{n,\max} = \{G_i\}_{i=1}^{\binom{n-1}{2}+1}$ of $G_{\binom{n-1}{2}+1}$ with maximal total persistence and a maximum number of bars is given by $G_i \cong H_i$ (Definition 13) for $i = 1, \ldots, e_n$. After $G_{e_n} = \mathcal{T}_{n,2}$, for $n \not\equiv 0 \mod 8$, the filtration is unique up to fiberwise isomorphism and represented by a sequence of tuples (cf.* (5)*) as follows*

$$\begin{cases} (K_1, K_1), (K_{j_n}, K_{j_n}), (K_{j_n+1}, K_{j_n+1}), \ldots, (K_{n/2-1}, K_{n/2-1}), & \text{if } n \equiv 2,4,6 \mod 8, \\ (K_1, K_1), (K_{k_n}, K_{k_n-1}), (K_{k_n+1}, K_{k_n}), \ldots, (K_{(n-1)/2}, K_{(n-3)/2}), & \text{if } n \equiv 1 \mod 2. \end{cases}$$

If $n \equiv 0 \mod 8$, then there are, up to fiberwise isomorphism, two optimal filtrations, represented by

$$(K_1, K_1), (K_{j_n}, K_{j_n}), (K_{j_n+1}, K_{j_n+1}), \ldots, (K_{n/2-1}, K_{n/2-1}) \text{ and}$$
$$(K_1, K_1), (K_{j_n+1}, K_{j_n+1}), (K_{j_n+2}, K_{j_n+2}), \ldots, (K_{n/2-1}, K_{n/2-1}).$$

▶ Remark 32. We prove the even case. The proof of the odd case is more or less the same (although the values are slightly different). However, we need to prove that, for any representation as in (5), $l_i \geq r_i$ for $i = 1, \ldots, c$. In the even case, we may assume this.

See Figure 8 for an example of and the full version [3, B.4] for a proof of this assumption.

Proof. It follows from Corollary 25 that, for a filtration $\mathcal{G}$ to have a maximum number of bars, we must have $G_{e_n} = \mathcal{T}_{n,2}$. For a filtration $\mathcal{G}$ to also have maximal total persistence, it immediately follows by Theorem 16 that $G_i \cong H_i$ for $i = 1, \ldots, e_n$.

Any filtration $\mathcal{G}$ that we consider in this proof will therefore be such that $G_i \cong H_i$ for $i = 1, \ldots, e_n$. We are interested in the remainder of the filtration, which we can represent by a sequence of tuples as in (5). Within the proof, we will further reduce it by writing

$$\big(l_1, r_1\big), \big(l_2, r_2\big), \ldots, \big(l_c, r_c\big) \text{ instead of } \big(K_{l_1}, K_{r_1}\big), \big(K_{l_2}, K_{r_2}\big), \ldots, \big(K_{l_c}, K_{r_c}\big). \tag{6}$$

Furthermore, given a filtration $\mathcal{G}$ represented by $\big(l_1, r_1\big), \big(l_2, r_2\big), \ldots, \big(l_c, r_c\big)$, we define its *alternation depth* $d_{\mathcal{G}}$ as follows:

- We let $d_{\mathcal{G}} \coloneqq 1$ if $\big(l_i, r_i\big) = \big(l_{i-1}+1, r_{i-1}+1\big)$ for all $i = 2, \ldots, c$.
- Otherwise, let $d_{\mathcal{G}} \geq 2$ be such that $\big(l_i, r_i\big) = \big(l_{i-1}+1, r_{i-1}+1\big)$ for all $i = d_{\mathcal{G}}+1, \ldots, c$, but $\big(l_{d_{\mathcal{G}}}, r_{d_{\mathcal{G}}}\big) \neq \big(l_{d_{\mathcal{G}}-1}+1, r_{d_{\mathcal{G}}-1}+1\big)$.

The idea of the proof is to start with a filtration $\mathcal{G} = \big\{G_i\big\}_{i=1}^{\binom{n-1}{2}+1}$, with a representation as in (6), and show that we can change the filtration in steps, to obtain in every step a filtration $\mathcal{G}'$ such that $T\beta_1^{\mathrm{FL}}\big(\mathcal{G}'\big) \geq T\beta_1^{\mathrm{FL}}\big(\mathcal{G}\big)$, and end up with an optimal solution.

Now, writing $n = 2p$, let $\mathcal{G}$ be a filtration, represented by

$$\big(1,1\big) = \big(l_1, r_1\big), \big(l_2, r_2\big), \ldots, \big(l_c, r_c\big) = \big(n/2-1, n/2-1\big) = \big(p-1, p-1\big).$$

We shall show that

1. If the alternation depth $d_{\mathcal{G}} > j_n$, then we can change $\mathcal{G}$ to a filtration $\mathcal{G}'$ such that $d_{\mathcal{G}'} < d_{\mathcal{G}}$ and $T\beta_1^{\mathrm{FL}}\big(\mathcal{G}'\big) \geq T\beta_1^{\mathrm{FL}}\big(\mathcal{G}\big)$.
2. If $l_2 < j_n$, then we can change $\mathcal{G}$ to a filtration $\mathcal{G}'$ represented by $\big(l'_1, r'_1\big), \ldots, \big(l'_{c'}, r'_{c'}\big)$ with $l'_2 = l_2 + 1$, such that $c' = c-1$ or $c' = c$ (depending on $\mathcal{G}$) and $T\beta_1^{\mathrm{FL}}\big(\mathcal{G}'\big) \geq T\beta_1^{\mathrm{FL}}\big(\mathcal{G}\big)$.

For the first part, let $\mathcal{G}$ be a filtration with alternation depth $d_{\mathcal{G}} > j_n$. Its representation, substituting n by $2p$, is of the form

$$\big(1,1\big), \ldots, \big(t,u\big), \big(v,w\big), \big(d_{\mathcal{G}}, d_{\mathcal{G}}\big), \big(d_{\mathcal{G}}+1, d_{\mathcal{G}}+1\big), \ldots, \big(p-1, p-1\big) \text{ with } \big(v,w\big) \neq \big(d_{\mathcal{G}}-1, d_{\mathcal{G}}-1\big).$$

Since we assumed that $v \geq w$, we must have $w < d_{\mathcal{G}} - 1$. We transform $\mathcal{G}$ into a filtration $\mathcal{G}'$ which is the same as $\mathcal{G}$, except that we add an extra alternation step. We consider two cases.

- Case 1: $v < d_{\mathcal{G}} - 1$. In this case, $\mathcal{G}'$ is represented by
$$\big(1,1\big), \ldots \big(t,u\big), \big(v,w\big), \big(d_{\mathcal{G}}-1, d_{\mathcal{G}}-1\big), \big(d_{\mathcal{G}}, d_{\mathcal{G}}\big), \big(d_{\mathcal{G}}+1, d_{\mathcal{G}}+1\big), \ldots, \big(p-1, p-1\big).$$
- Case 2: $v = d_{\mathcal{G}} - 1$. In this case, $\mathcal{G}'$ is represented by
$$\big(1,1\big), \ldots \big(t,u\big), \big(v = d_{\mathcal{G}}-1, d_{\mathcal{G}}-1\big), \big(d_{\mathcal{G}}, d_{\mathcal{G}}\big), \big(d_{\mathcal{G}}+1, d_{\mathcal{G}}+1\big), \ldots, \big(p-1, p-1\big).$$

In both cases, we see that $d_{\mathcal{G}'} < d_{\mathcal{G}}$. Moreover, as we show in more detail in [3, B.4],

$$T\beta_1^{\mathrm{FL}}\big(\mathcal{G}'\big) - T\beta_1^{\mathrm{FL}}\big(\mathcal{G}\big) = \frac{1}{6}\big(d_{\mathcal{G}} - 1 - w\big)\big(d_{\mathcal{G}} - w\big)\big(4d_{\mathcal{G}} + 2w - 3p - 2\big) \geq 0. \tag{7}$$

To see that this difference is non-negative, first recall that $w < d_{\mathcal{G}} - 1$, so the first two terms are strictly positive. Furthermore, since $d_{\mathcal{G}} > \lfloor 3p/4 - 1/4 \rfloor$, we have $d_{\mathcal{G}} \geq \lceil 3p/4 \rceil \geq 3p/4$, so

$$4d_{\mathcal{G}} + 2w - 3p - 2 \geq 2w - 2 \geq 0,$$

since $w \geq 1$. Note that $T\beta_1^{\mathrm{FL}}\big(\mathcal{G}'\big) - T\beta_1^{\mathrm{FL}}\big(\mathcal{G}\big) = 0$ if and only if $w = 1$ and $d_{\mathcal{G}} = 3p/4$.

We now show the second part. We let $\mathcal{G}$ be a filtration with representation

$(1,1),(j,k),(l,m),\dots,(p-1,p-1)$ with $j \leq 3p/4-1$.

We transform $\mathcal{G}$ into a filtration $\mathcal{G}'$ such that its start is different from $\mathcal{G}$'s, but from the tuple (l,m) onwards, $\mathcal{G}$ and $\mathcal{G}'$ are the same. We again consider two cases.

- Case 1: $l > j+1$. In this case, $\mathcal{G}'$ is represented by $(1,1),(j+1,k),(l,m),\dots,(p-1,p-1)$.
- Case 2: $l = j+1$. In this case, $\mathcal{G}'$ is represented by $(1,1),(j+1=l,m),\dots,(p-1,p-1)$.

In both cases, as we show in more detail in the full version [3, B.4], we have

$$T\beta_1^{\mathrm{FL}}(\mathcal{G}')-T\beta_1^{\mathrm{FL}}(\mathcal{G}) = (k-1)\big(j(p-j-1)-\frac{k}{6}(3p-2k-2)\big) \overset{(i)}{\geq} 2\big((j-k+1)(j-k)\big) \overset{(ii)}{\geq} 0, \quad (8)$$

where we use in (i) that $j \leq 3p/4-1$, and in (ii) that $j \geq k$. If $n \equiv 2,4,6 \mod 8$, then $j \leq 3p/4-1$ implies that $j \leq \lfloor 3p/4-1 \rfloor = \lfloor 3p/4-5/4 \rfloor = j_n - 1$. If $n \equiv 0 \mod 8$, then $j \leq 3p/4-1 = j_n$. Only in this case, and if $j = k = 3p/4-1$, then the inequality is sharp.

Hence, if we start with an arbitrary filtration $\mathcal{G}$, then we can transform it step by step into a more optimal filtration $\mathcal{G}^1$ with $d_{\mathcal{G}^1} \leq j_n$, by the first part. Then, by the second part, we can transform $\mathcal{G}^1$ step by step into a more optimal filtration $\mathcal{G}^2$ that starts with $(1,1),(j_n,j_n)$, with alternation depth $d_{\mathcal{G}^2} = j_n$. Hence, $\mathcal{G}^2$ is the optimal filtration as described in the statement for $n \equiv 2,4,6 \mod 8$.

In the case that $n \equiv 0 \mod 8$, we saw in the proof of case 1 and case 2, that the filtration $\mathcal{G}^3$ starting with $(1,1),\big(j_n+1,j_n+1\big)$ and having alternation depth $d_{\mathcal{G}^3} = j_n+1$ satisfies $T\beta_1^{\mathrm{FL}}(\mathcal{G}^3) = T\beta_1^{\mathrm{FL}}(\mathcal{G}^2)$, hence both $\mathcal{G}^2$ and $\mathcal{G}^3$ are optimal. ◀

8 Discussion

In this paper, we have answered several key questions in the context of topological data analysis. However, many questions remain open, and we propose the following conjectures.

▶ **Conjecture 33.** *If $\mathcal{G}$ is a filtration on n vertices, then the number of intervals in the barcode of $\mathcal{G}$ in homology degree k satisfies $|\mathcal{B}_k^{\mathrm{FL}}(\mathcal{G})| \leq \beta_k^{\mathrm{FL}}(\mathcal{T}_{n,k+1})$.*

▶ **Conjecture 34.** *The extremal filtrations described in Section 7 achieve the maximal total persistence over any edgewise filtration of flag complexes in homology degree 1.*

Proving the latter conjecture likely requires new ideas, as no filtration $\mathcal{G}$ can maximize $\beta_1^{\mathrm{FL}}(G_i)$ for all i. In fact, extremal values need not be realized by graphs containing Turán graphs as spanning subgraphs, as illustrated by the following example.

▶ **Example 35.** Let $n = 8$ and $m = 16+5$. Suppose G contains the Turán graph $\mathcal{T}_{8,2}$. By Proposition 5, we seek to add 5 edges to the Turán graph to maximize the product $(d_1-1)(d_2-1)$. It is easy to see that this product can be at most 1. However, by partitioning the vertices as $V = V_1 \cup V_2$, with $|V_1| = 3$ and $|V_2| = 5$, and adding all edges between V_1 and V_2, we need to add 6 more edges. Adding these as a K_4 in the larger partition gives $d_1 = 2$ and $d_2 = 1$; see Figure 9. The value 2 is extremal; see Figure 10.

Finding extremal values of $\mathbb{Z}$-linear functions defined on simplicial complexes with *precisely* n vertices and m edges presents a challenging and interesting direction for future research.

▶ **Conjecture 36.** *Let $\mathbf{G}(n,m)$ denote the collection of graphs on n vertices and m edges. If $G \in \mathbf{G}(n,m)$ and $\beta_1^{\mathrm{FL}}(G) = \max_{H\in\mathbf{G}(n,m)} \beta_1^{\mathrm{FL}}(H)$, then G contains a complete bipartite spanning subgraph.*

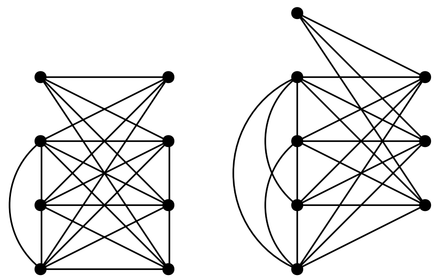


Figure 9 Two graphs G_l (left) and G_r (right) such that $\beta_1^{\mathrm{FL}}(G_l) = 1$ and $\beta_1^{\mathrm{FL}}(G_r) = 2$.

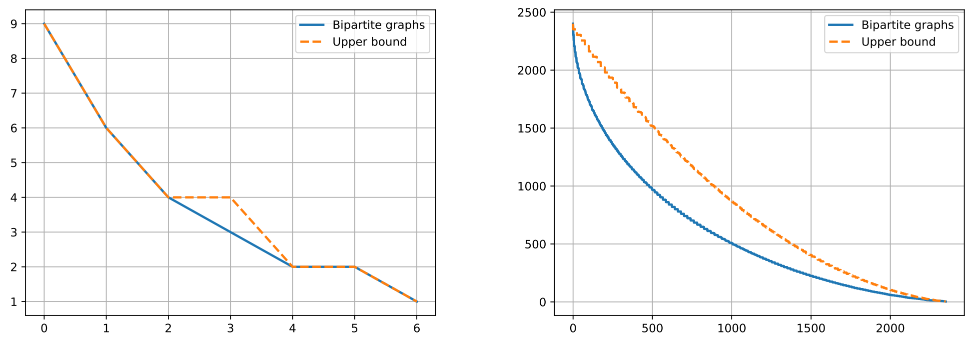


Figure 10 The vertical axis is β_1^{FL}, and the value k along the x-axis represents $(n/2)^2 + k$ edges. Here n is the number of vertices in the underlying graph; $n = 8$ (left) and $n = 100$ (right). The solid curves give the optimal value of $\beta_1^{\mathrm{FL}}(G)$ for any graph G containing a complete bipartite spanning subgraph, and the dashed line is an upper bound derived from the proof of Theorem 10 in conjuction with Lemma 17 (details omitted).

References

1 Michał Adamaszek. Extremal problems related to Betti numbers of flag complexes. *Discrete Applied Mathematics*, 173:8–15, 2014. doi:10.1016/J.DAM.2014.04.006.

2 Ulrich Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *Journal of Applied and Computational Topology*, 5(3):391–423, 2021. doi:10.1007/S41468-021-00071-5.

3 Lies Beers and Magnus Bakke Botnan. Extremal Betti Numbers and Persistence in Flag Complexes, 2025. arXiv:2502.21294.

4 Anders Björner and Gil Kalai. An extended Euler-Poincaré theorem, 1988.

5 Magnus Botnan and Michael Lesnick. An introduction to multiparameter persistence. In *Representations of Algebras and Related Structures*, pages 77–150. EMS Press, 2023.

6 Gary Chartrand and Ping Zhang. *A first course in graph theory*. Courier Corporation, 2013.

7 Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 232–241, 2008.

8 Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.

9 Herbert Edelsbrunner and János Pach. Maximum Betti numbers of Čech complexes. In *40th International Symposium on Computational Geometry (SoCG 2024)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.SoCG.2024.53.

10 Michael Goff. Extremal Betti numbers of Vietoris–Rips complexes. *Discrete and Computational Geometry*, 46(1):132, 2011. doi:10.1007/s00454-010-9274-z.

11 Dmitry N Kozlov. Convex hulls of f-and β-vectors. *Discrete & Computational Geometry*, 18:421–431, 1997. doi:10.1007/PL00009326.