

\section{Sparse Regime}

In the probabilistic analysis of the geometric complexes introduced in Section 3, we will focus on the \( p \)-th Betti numbers, for \( p \geq 1 \), stating two results originally proved in \(^1\). Indeed, \( \beta_0 \) behaves different from the other Betti numbers. However, it is also the most elementary of the concepts so that much more is known about this Betti number than about the others.

**Number of components.** The 0-th Betti number of a simplicial complex is the number of components, which is the same for the 1-skeleton. This implies

\[
\beta_0(R(X; r)) = \beta_0(C(X; r)) = \beta_0(G(X; r)).
\]

(6)

To be connected is a monotone property: if \( H \) is a subgraph of \( G \) with the same vertex set and \( H \) is connected, then \( G \) is connected. It therefore makes sense to defined the connectivity threshold:

\[
R_n = \min \{ r \mid G(X; r) \text{ is connected} \}.
\]

(8)

Note that \( R_n \) is the length of the longest edge in the minimum spanning tree of \( X \). For density functions with support that is convex with smooth boundary, the connectivity threshold behaves as for random graphs. Specifically, there are constants \( 0 < c < C < \infty \) such that

\[
\text{Prob}[G(X; c \left( \frac{\log n}{n} \right)^{1/2}) \text{ is connected}] \to 0,
\]

(9)

\[
\text{Prob}[G(X; C \left( \frac{\log n}{n} \right)^{1/2}) \text{ is connected}] \to 1,
\]

(10)

a.a.s. as \( n \to \infty \); see for example \(^2\). In the sparse regime, which is much below the connectivity threshold, the number of components grows linearly with \( n \), as we will see.

**Expected Betti numbers.** Other than for \( p = 0 \), the Betti numbers do not vary monotonically as we add simplices. We begin by stating the main results, which are bounds on the expected Betti numbers of the Vietoris-Rips and the Čech complex.


**Vietoris-Rips complex, sparse regime.** For \( 2 \leq d, 1 \leq p \), and \( r = o(n^{-1/4}) \), the expected \( p \)-th Betti number of \( R(X; r) \) satisfies

\[
\frac{\text{Exp}[\beta_p]}{n^{2p+2d(p+1)}} \to C_p
\]

(11)

as \( n \to \infty \), where the constant \( C_p \) depends only on \( p \) and on the probability density function.

For \( r = n^{-1/2} \) – which is of course not in the allowed range – the denominator is \( n \), while for smaller \( r \), the denominator is larger. We note that (11) also holds for \( d \leq p \). Indeed, the Vietoris-Rips complex can have non-trivial homology in dimensions beyond the dimension of the Euclidean space in which the points are sampled.

**Čech complex, sparse regime.** For \( d \geq 2, 1 \leq p \leq d - 1 \), and \( r = o(n^{-1/4}) \), the expected \( p \)-th Betti number of \( C(X; r) \) satisfies

\[
\frac{\text{Exp}[\beta_p]}{n^{2p+2d(p+1)}} \to D_p
\]

(12)

as \( n \to \infty \), where the constant \( D_p \) depends only on \( p \) and on the probability density function.

For \( r = n^{-1/2} \), the denominator is again \( n \), and for smaller \( r \), the denominator is larger but not as large as for Vietoris-Rips complexes. We note that the Čech complex is homotopy equivalent to the union of balls of radius \( r \) centered at the points in \( X \). This implies in particular that \( H_p = 0 \) for \( d \geq p \).

**Feasible subgraphs.** The proofs of both theorems heavily rely on a the expected number of induced subgraphs isomorphic to a given graph, a problem studied in the mentioned book by Penrose.

![Figure 8: The complete bipartite graph with seven vertices is not feasible in the plane.](image)

Recall that \( H \) is an induced subgraph of \( G \) if for every two vertices \( u \) and \( v \) of \( H \), the absence of the edge \( \{u, v\} \) in \( H \) implies its absence in \( G \). We call a connected graph
feasible if it arises as an induced subgraph of a geometric graph. For example, the complete bipartite graph with one vertex on one side and six vertices on the other side is not feasible in \( \mathbb{R}^2 \); see Figure 8. Letting \( 2r \) be the length of the longest of the six edges, it is not difficult to show that at least one of the pairs of the six vertices connected to the center is at distance at most \( 2r \). Write \( g_n(H) \) for the number of induced subgraphs of \( G(X_n; r_n) \) isomorphic to \( H \), and let \( j_n(H) \) for the number of components of \( G(X_n; r_n) \) isomorphic to \( H \). We have \( j_n(H) \leq g_n(H) \), but it turns out that \( j_n \) is not much smaller than \( g_n \).

**Expectations of Subgraph Counts.** Let \( H \) be a feasible connected graph with \( k \geq 2 \) vertices. Then

\[
\lim_{n \to \infty} \frac{\text{Exp}[g_n(H)]}{n^{k+2d(k-1)}} = \lim_{n \to \infty} \frac{\text{Exp}[j_n(H)]}{n^{k+2d(k-1)}} = C,
\]

in which \( C \) is a positive constant that depends only on \( H \) and the probability density function.

Note that the limit of the expectation for a feasible graph with \( k+1 \) vertices is \( \Theta(n^d) \) times that for a feasible graph with \( k \) vertices. In the sparse regime, \( n^d \) goes to zero, which implies that all subgraphs except the smallest ones can be neglected. For example, for \( k = 2 \), we see that the number of edges grows sublinearly with \( n \). This implies that in the sparse regime, the number of components grows linearly with \( n \). It follows that for \( 2 \leq d \) and \( r = o(n^{-\frac{d}{2}}) \), the 0-th Betti numbers of the Vietoris-Rips and the Čech complexes satisfy

\[
\frac{\text{Exp}[\beta_0(R(X_n; r))] \text{Exp}[\beta_0(C(X_n; r))]}{n} \to 1,
\]

as \( n \to \infty \).

**Smallest subcomplexes.** For the Vietoris-Rips complex, the smallest subcomplex that supports a \( p \)-cycle is the boundary complex of the \((p+1)\)-dimensional cross-polytope, which is the convex hull of the \( 2p \) points at which the unit \( p \)-sphere intersects the coordinate axes in \( \mathbb{R}^{p+1} \). The reason for this is the fact that the Vietoris-Rips complex is the clique complex of its edges for which smaller \( p \)-cycles are filled in and become boundaries. This fact is trivial for \( p = 0, 1 \).

**Lemma.** The smallest subcomplex of a clique complex supporting a non-bounding \( p \)-cycle has \( 2p + 2 \) vertices.

**Proof.** Let \( S_p \) be a subcomplex with the smallest number of vertices that supports a non-bounding \( p \)-cycle, and let \( u \) be a vertex of \( S_p \). The link of \( u \) in \( S_p \) supports a \((p-1)\)-cycle, and within the subcomplex defined by the vertices in the link, this cycle does not bound. Indeed, if it does then we can either remove \( u \) and get a smaller support for a \( p \)-cycle, or we get a contradiction to \( S_p \) being non-bounding. By induction hypothesis, we can now assume that the link of \( u \) has at least \( 2p \) vertices. If the link has \( 2p \) vertices, then we already have \( 2p+1 \) vertices in total, namely \( u \) and the vertices of its link. If these are all the vertices of \( S_p \), then we can repeat the argument for each of the other \( 2p \) vertices, and conclude that the \( 2p+1 \) vertices form a complete graph. But then the clique complex would contain all simplices spanned by subsets of these \( 2p+1 \) vertices, which contradicts that \( S_p \) does not bound. Hence, \( S_p \) has at least \( 2p + 2 \) vertices, as claimed.

![Figure 9: From left to right: 1-skeleta of the \((p+1)\)-dimensional cross-polytope realized as geometric graphs in the plane, for \( p = 1, 2, 3 \). The shading suggests four of the eight triangles of the octahedron and two of the sixteen tetrahedra in the boundary of the 4-dimensional cross-polytope.](image)

Setting \( k = 2p + 2 \) in the subgraph count, we get the bounds for the \( p \)-th Betti number of the Vietoris-Rips complex. It is perhaps surprising that the 1-skeleton of the \((p+1)\)-dimensional cross-polytope is feasible, even in \( \mathbb{R}^2 \). As illustrated in Figure 9, the regular \((2p+2)\)-gon provides an example.

For the Čech complex, the smallest subcomplex that supports a \( p \)-cycle is the boundary complex of a \((p+1)\)-simplex. It contains \( p + 2 \) vertices. However, in the Čech complex, any induced subcomplex is not determined by its edges. We therefore need an alternative derivation of the expectation limits given above, which can be found in [3]. Still, if we set \( k = 2p + 2 \) in the subgraph count, we get the correct bounds for the \( p \)-th Betti numbers of the Čech complex.

---