Computation of Persistent Betti numbers using Cubical and Permutahedral Complexes

Fintan Hegarty
Supervisor: Prof. Graham Ellis

National University of Ireland, Galway

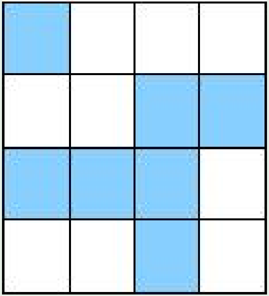
January 26th, 2011

science foundation ireland
fondúireacht eolaíochta éireann

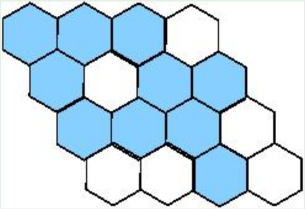
Example

A representation X of the pure cubical complex K and a representation Y of the pure permutahedral complex L where

$$K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad L = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$



X



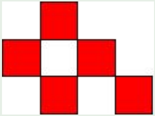
Y

The *betti numbers* $\beta_n(X)$ of a space X can be thought of as the number of holes with n-dimensional boundary in X, with $\beta_0(X)$ being the number of connected path components in X.

Suppose we are given $T \subset X$ which is a random sample of points from X - can we calculate the betti numbers of X from T?

Example

T_0 is a random sample of 710 points from some space Y.



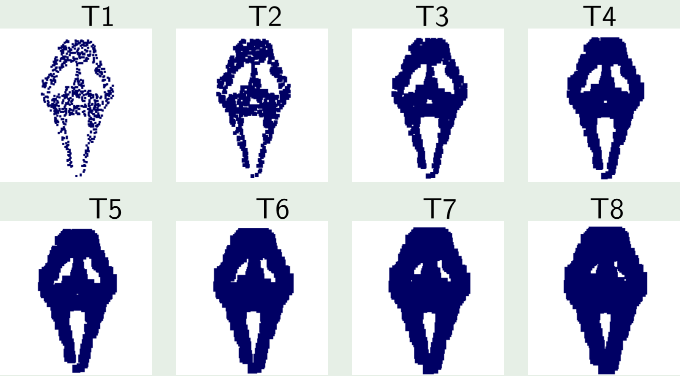
T_0.

$\beta_1(T_0) = 1$.

Thickening a complex can sometimes provide more information.

Example

$T_1 \dots T_8$ are a series of thickenings of T_0.



The betti numbers at any given stage do not give a clear indication of the betti numbers of the original space.

Example

T_5 and T_6 both have $B_1 = 3$, but one hole is created and another destroyed when T_5 is thickened, so just because $B_1(T_5) = B_1(T_6)$ it does not mean that we can assume that they provide an accurate representation of $B_1(X)$.

We need to analyse which holes *persist* throughout a number of thickenings.

Given a sequence of subcomplexes $M_1 \subset M_2 \subset \ldots$ we can represent the persistence of holes through the sequence as a matrix where the i^{th} entry in the j^{th} row is the number of holes which exist in the i^{th} complex *and* the j^{th} complex.

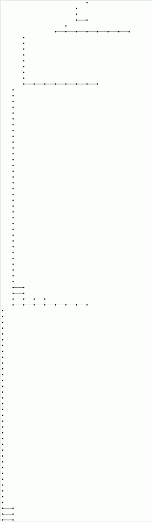
Example

The peristence matrix for $T_1 \ldots T_{14}$.

$$\begin{pmatrix}
37 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 41 & 4 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 13 & 3 & 3 & 2 & 2 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 3 & 3 & 2 & 2 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 2 & 2 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 3 & 2 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 4 & 3 & 3 & 2 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 4 & 2 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 2 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Following the work of Carlsson we represent the matrix as a *bar code* with horizontal edges and vertices arranged in columns.

Example



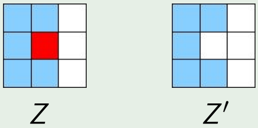
We can see that there are 3 holes which persist for a relatively large number of thickenings, and 1 which persists for a shorter but non-negligible number of thickenings.

The calculations required to find the (persistent) betti numbers involve matrix operations which can be computationally expensive if the matrices are large. By reducing the size of the space we can greatly reduce the size of the matrices to be dealt with. One of the matrices used for calculating $B_1(T_6)$, for example, is of size 15165×29902. This can be reduced to a matrix of size 27×40 if we reduce the space we are dealing with to a smaller space which has the same betti numbers.

A neighbourhood is *positive* if its central cell can be removed without changing any of the betti numbers of the neighbourhood.

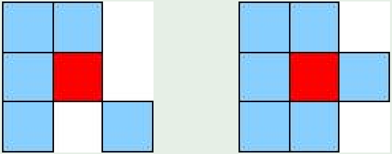
Example

The space Z below is positive, as $\beta_0(Z) = \beta_0(Z') = 1$ and $\beta_1(Z) = \beta_1(Z') = 0$. Consequently, we can remove the central cell without affecting the betti numbers of the space they are in.



Example

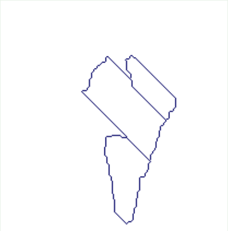
The two neighbourhoods below are not positive, as removing the red cell in either would change either β_0 or β_1 respectively.



There are only 2^8 possible combinatorial types of neighbourhood in R^2_C so we can compute and store a list $\mathbb{L}^2_C$ indicating whether each neighbourhood is positive or not. We *contract* a complex by checking all its cells against $\mathbb{L}^2_C$ and removing those whose neighbourhoods are positive.

Example

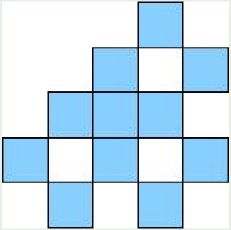
Here, $CT_6 = \texttt{Contracted}(T_6)$ has the same betti numbers, but only 445 cells, as opposed to 12843.



It is possible, by recursively finding bounding spaces and contracting them, to find a smaller yet homotopy equivalent retract of a space X, which we refer to as a *zig-zag retraction* $ZZ(X)$.

Example

Here is $ZZ(CT_6)$ which consists of only 11 cells, as opposed to T_6's original 12843:



Analogous methods exist in higher dimensions, though $\mathbb{L}_C^4$ the list indicating positivity of neighbourhoods in the 4-dimensional cubical case, with 2^{80} entries, is too large to store.

In the permutahedral case, however, the 4-dimensional list of positive neighbourhoods $\mathbb{L}_P^4$ is of length 2^{30}, and so is possible to compute and store. This enables us to efficiently contract, and hence calculate the betti numbers etc of 4-dimensional data. The pure permutahedral complex and pure cubical complexes will not necessarily have the same betti numbers everywhere...

Example

$\beta_1(T_6) = 3$ in the pure cubical case, but $\beta_1(PT_6) = 2$, where PT_6 is the result of 6 thickenings of T_0 in the pure permutahedral case.

T_6

PT_6

Example

...but the persistent betti numbers and bar codes yielded should be similar.