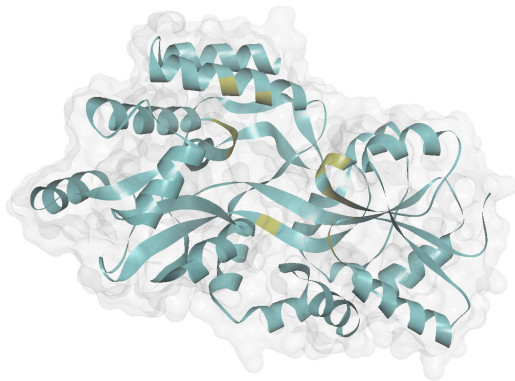
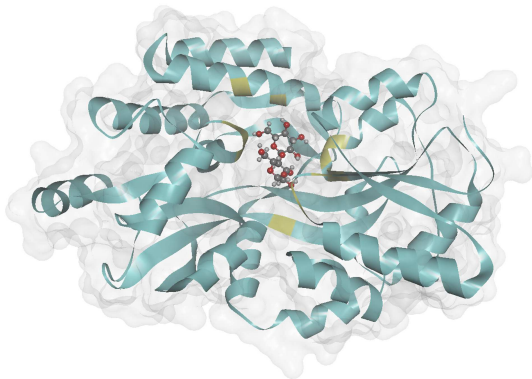


Maltose Binding Protein, two conformations



V. Kovacev-Nikolic, P. Bubenik, D. Nikolic, and G. Heo. Using persistent homology and dynamical distances to analyze protein binding. *Statistical Applications in Genetics and Molecular Biology*, **15** (2016) no. 1, 19–38.

Maltose Binding Protein, two conformations



V. Kovacev-Nikolic, P. Bubenik, D. Nikolic, and G. Heo. Using persistent homology and dynamical distances to analyze protein binding. *Statistical Applications in Genetics and Molecular Biology*, **15** (2016) no. 1, 19–38.

Maltose Binding Protein Data

The Data

Fourteen MBP structures from the Protein Data Bank.

- 7 closed conformations
- 7 open conformations

X-ray crystallography: coordinates of atoms.

Represent each amino acid residue by its $C\alpha$ atom.

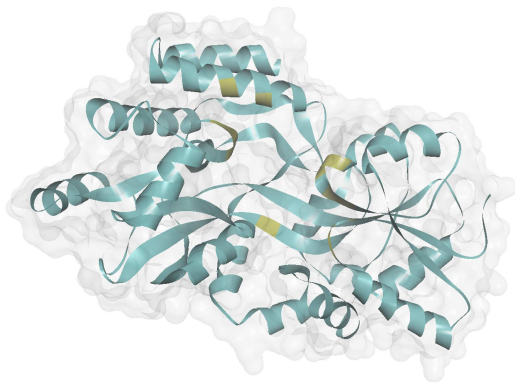
Have 14 sets of 370 points in \mathbb{R}^3 .

The Goal

Can we use topological data analysis to distinguish the open and closed conformations?

Topological Analysis of our Protein data

We have spatial coordinates of 370 amino acid residues.

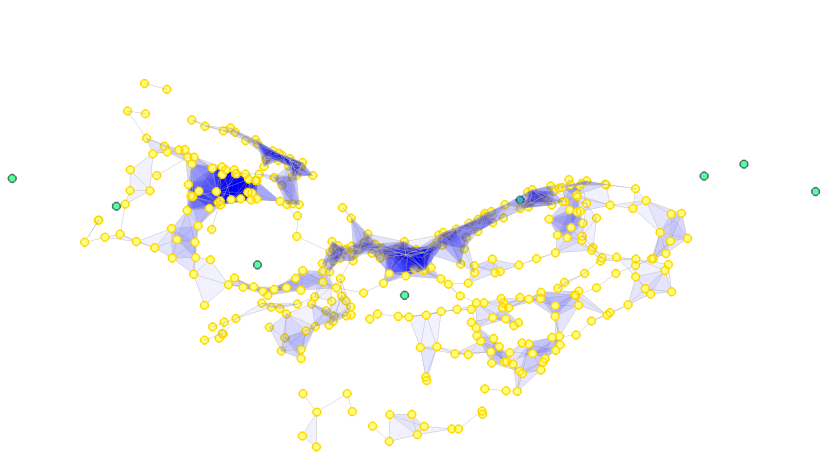


- Construct elastic network model
- Calculate correlations
- Use these to get distances

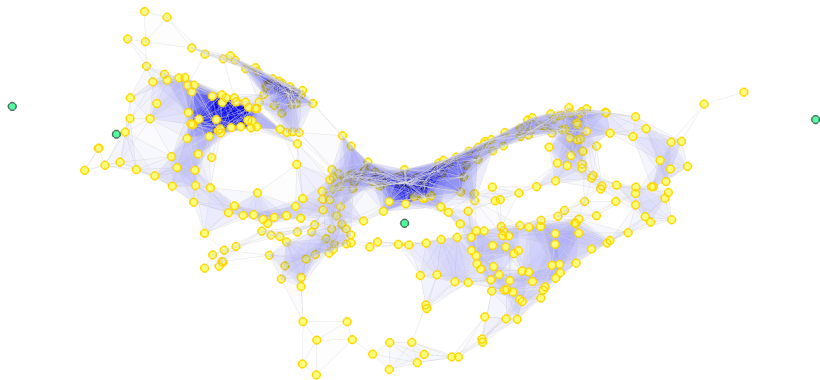
MBP Vietoris-Rips complex



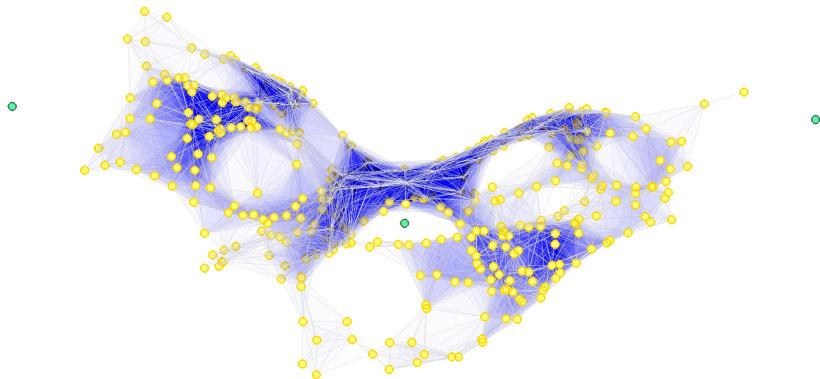
MBP Vietoris-Rips complex



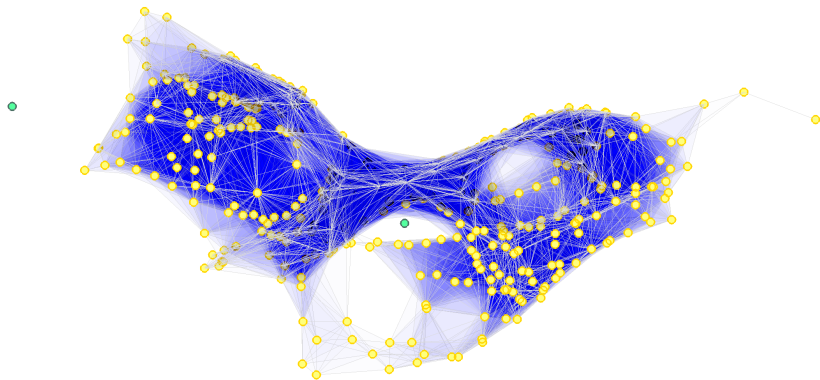
MBP Vietoris-Rips complex



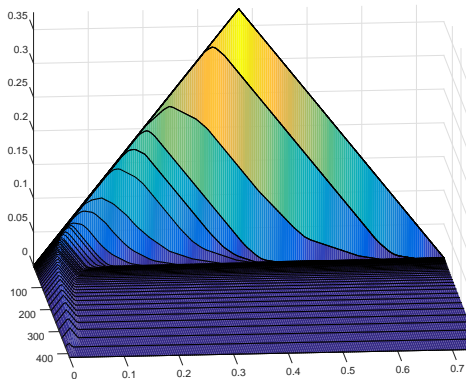
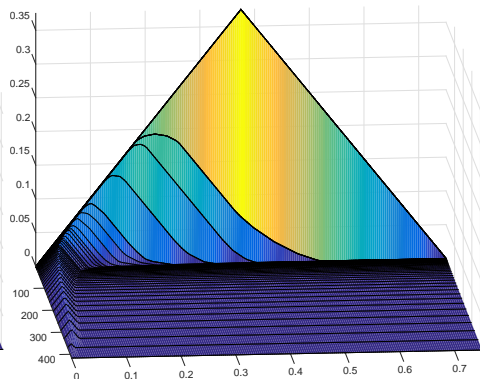
MBP Vietoris-Rips complex



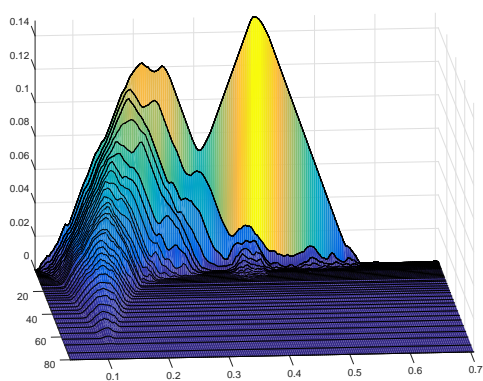
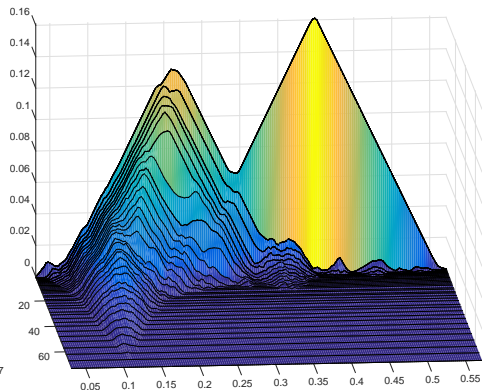
MBP Vietoris-Rips complex



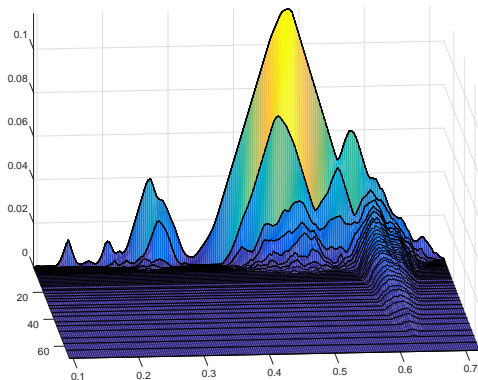
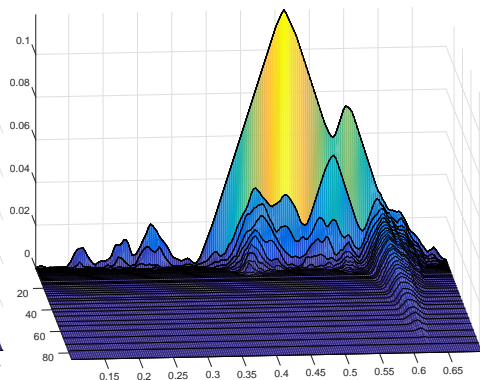
MBP average persistence landscapes

 H_0 closed H_0 open

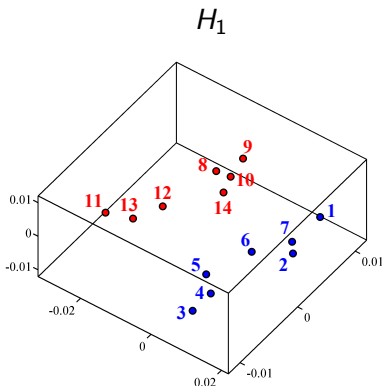
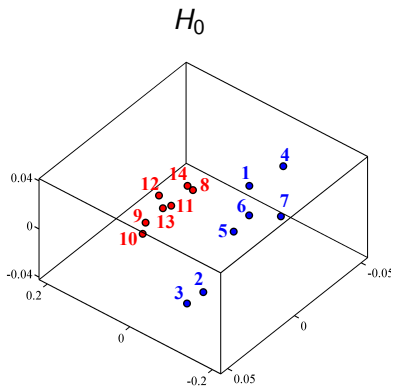
MBP average persistence landscapes

 H_1 closed H_1 open

MBP average persistence landscapes

 H_2 closed H_2 open

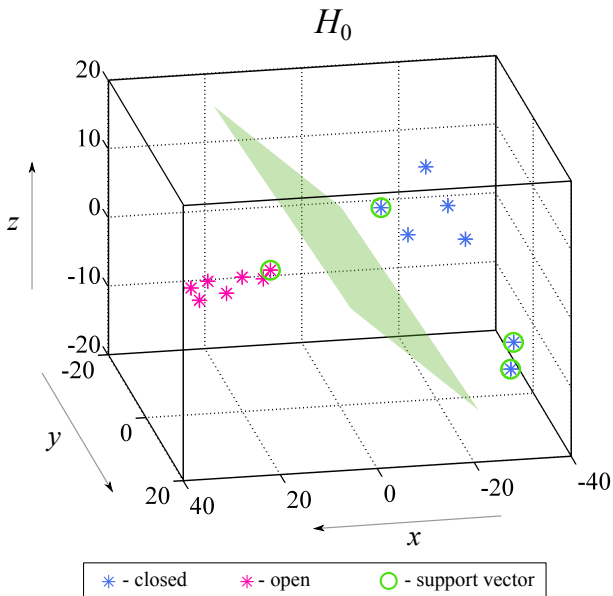
Classification of protein conformations



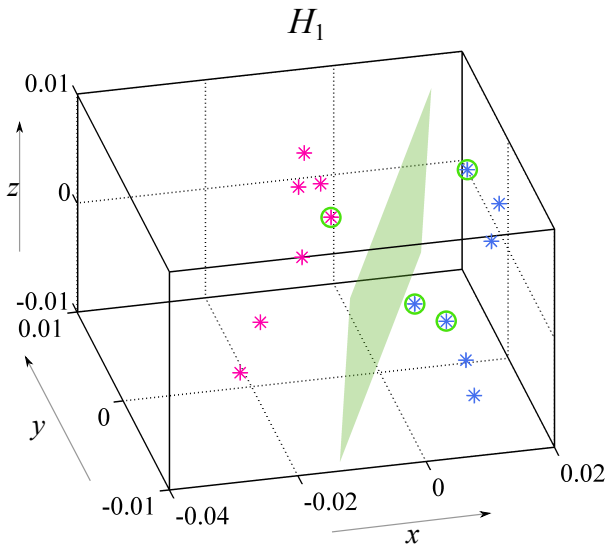
Projection of the L^2 distance matrix to \mathbb{R}^3 using Isomap.

Find the best plane that separates the two sets of points.
Method called Support Vector Machine (SVM).

Classification of protein conformations



Classification of protein conformations

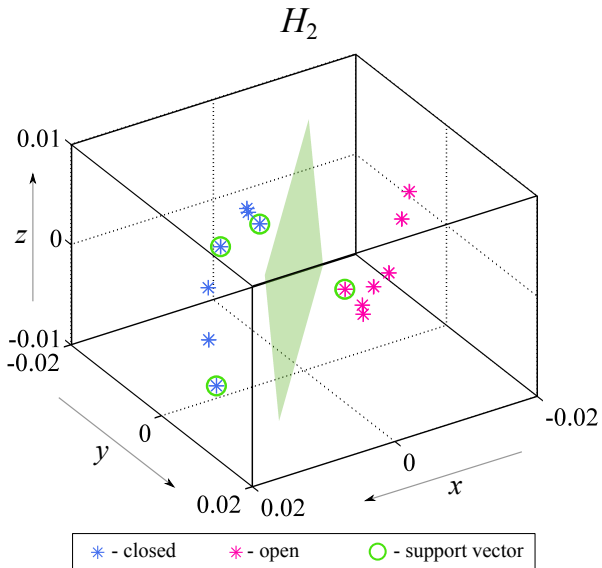


* - closed

* - open

○ - support vector

Classification of protein conformations



Classification of protein conformations

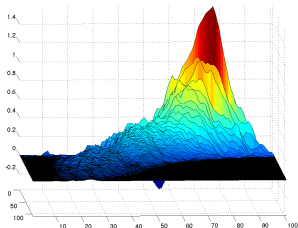
Support vector classification with 5-fold cross validation:

	true	
pred	closed	open
closed	7	0
open	0	7

Exploratory Data Analysis

If we find significant topological features and differences, what does it tell us?

1 Brain arteries: Female - Male difference

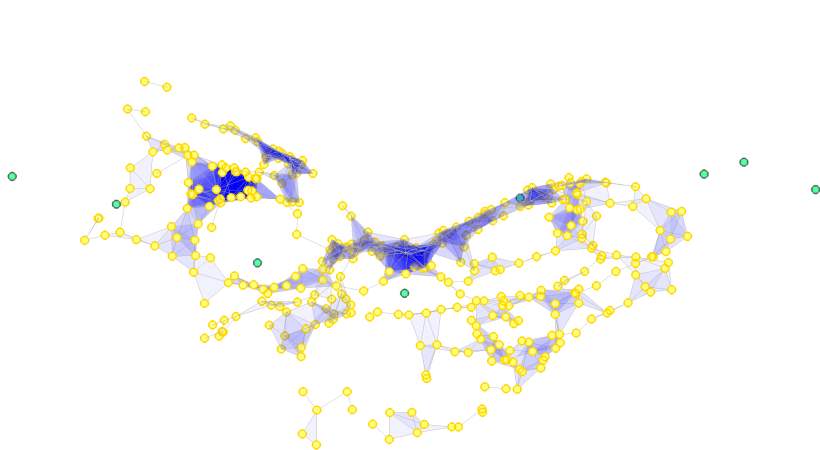


2 Protein data: the most persistent cycle

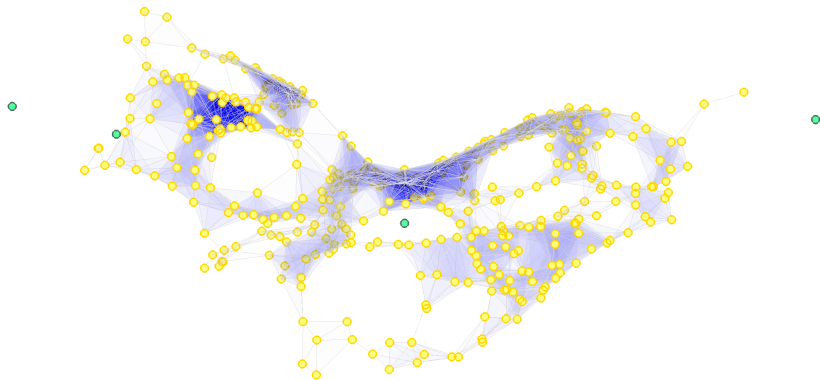
MBP Vietoris-Rips complex



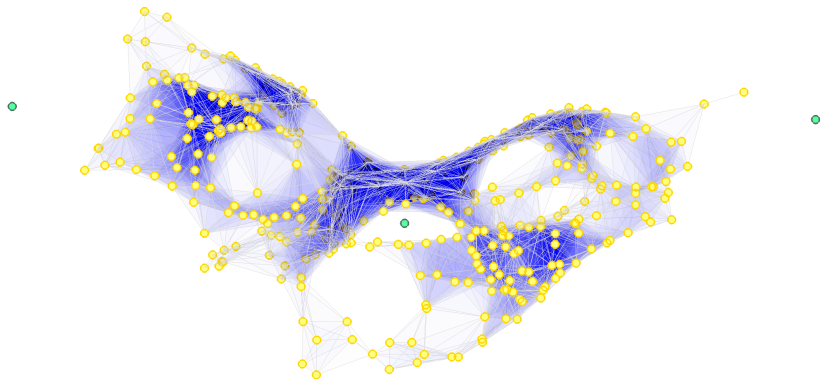
MBP Vietoris-Rips complex



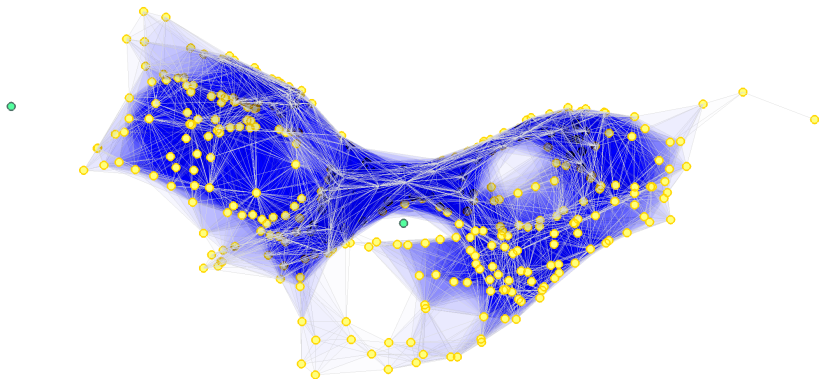
MBP Vietoris-Rips complex



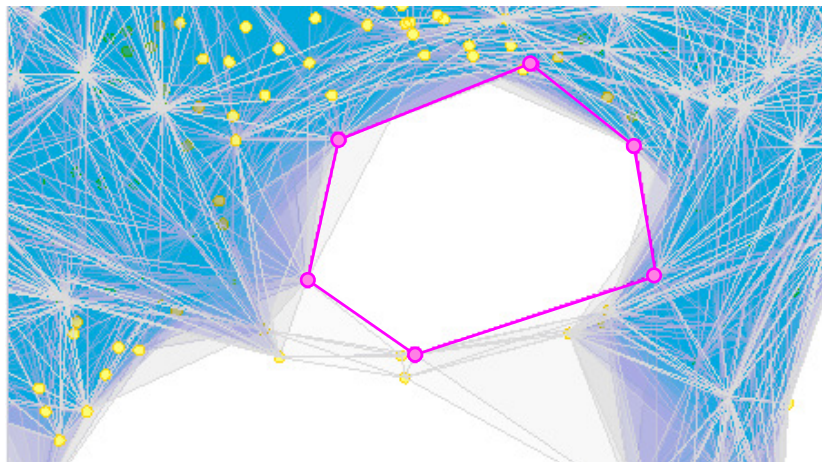
MBP Vietoris-Rips complex



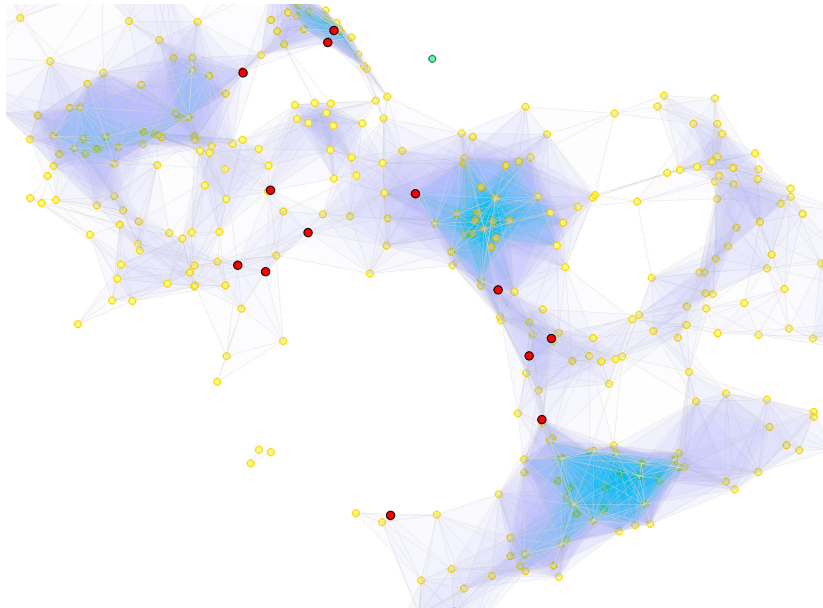
MBP Vietoris-Rips complex



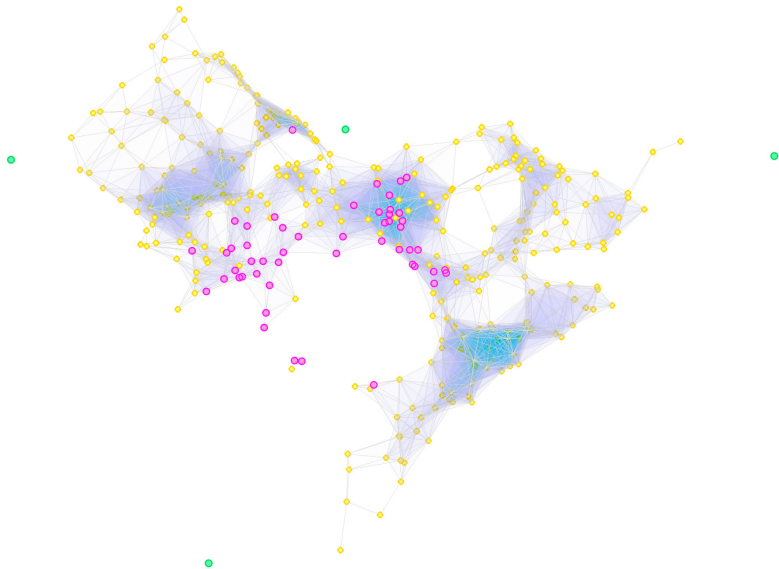
The most persistent cycle



Active sites and the most persistent cycle



Allosteric pathways and the most persistent cycle



Topological Data Analysis Summary

