Topological Data Analysis Applied to Idiotypic Network

Matteo Rucco[1]

F. Castiglione[2], E. Merelli [1]

matteo.rucco@unicam.it

[1]University of Camerino - [2]Institute for Computing Applications (IAC) "M. Picone"

February 25, 2015

OUTLINE

Theoretical introduction
    Topological data analysis
    Clique Weight Rank Persistent Homology (CWRPH)

Experiment: Topological Data Analysis of Idiotypic Network
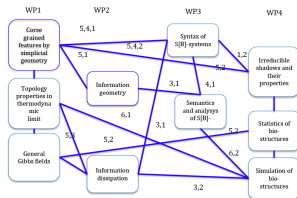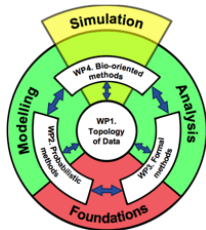   Idiotypic network
    Simulation
    Persistent Entropy
    Topological-holes analysis: communities detection

Conclusions

TOPDRIM - TOPOLOGY DRIVEN METHODS FOR COMPLEX SYSTEMS

The goal of this project is to provide methods driven by the **topology of data** for describing the dynamics of multi-level complex systems

http://www.topdrim.eu

TOPOLOGICAL DATA ANALYSIS FOR STUDYING COMPLEX SYSTEMS

Topological Data Analysis (TDA) is a subarea of computational topology that develops topological techniques for robust analysis of **scientific data: point cloud and complex networks**.
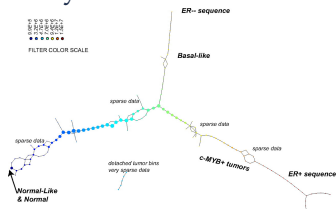
- In Petri et al. (2014), topological methods have been applied for understanding the effect of psilocybin.

- In Reidys et al. (2011), topological methods have been used for predicting RNA pseudoknots

SIMPLICIAL COMPLEX
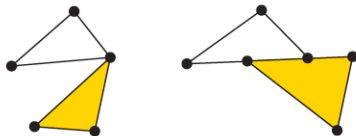
A topological space is generated by simplicies (a convex hull of k+1 vertices):

A nested collection of simplices is a simplicial complex

More formally: *a simplicial complex K on a finite set $V = \{v_1, v_2, \ldots, v_n\}$ of vertices is a non-empty subset of the power set of V, so that the simplicial complex K is closed under the formation of subsets*

Left: a simplicial complex. Right: a collection of simplices but that is not a simplicial complex.

# HOW TO BUILT A SIMPLICIAL COMPLEX

Vietoris-Rips, Witness, Neighborhood, Clique-complexes
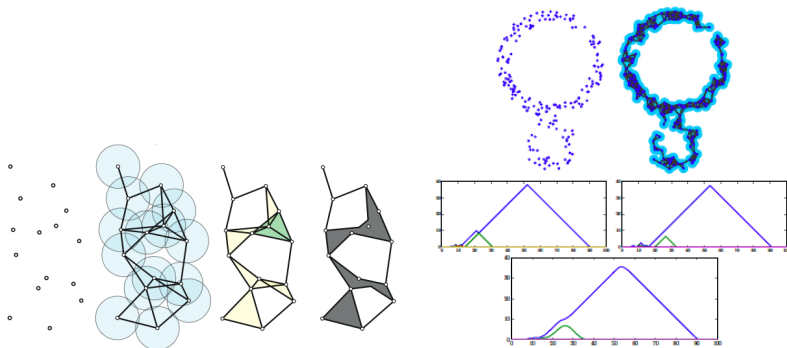
Left: example of Vietoris-Rips. Right: example of landscape analysis

PERSISTENT HOMOLOGY - FILTRATION

The basic aim of persistent homology is to measure the lifetime of certain topological properties of a simplicial complex (s.c.) when simplices are added to the complex or removed from it
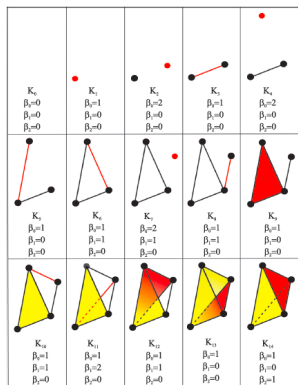
At each stage of filtration a simplices is added (red). The intuitive notion of Betti numbers as the number of k-dimensional holes is clear in this representation.

VISUALIZATION OF PERSISTENCE HOMOLOGY: BETTI BARCODES

Persistent homology represents an algebraic invariant that detects the birth and death of each topological features. It is advantageous to encode the persistent homology in the form of a parameterized version of Betti numbers.
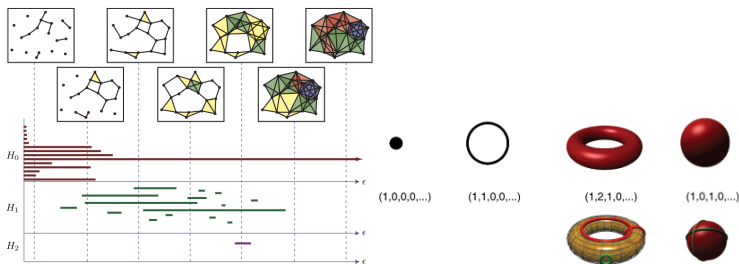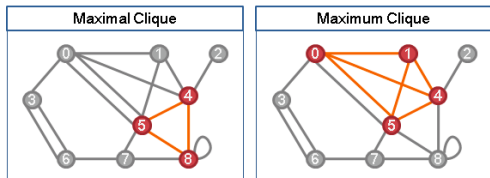
CLIQUE

A **clique** of a graph G is a complete subgraph of G ($d(G) = 1$), and the clique of largest possible size is referred to as a **maximum clique**.

A **maximal clique** is a clique that cannot be extended by including one more adjacent vertex, meaning it is not a subset of a larger clique. Maximum cliques are therefore maximal cliqued (but not necessarily vice versa).
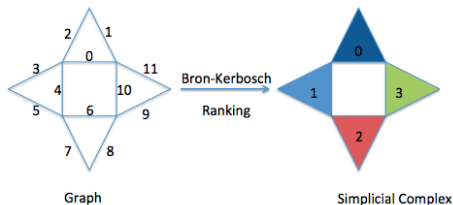
CLIQUE WEIGHT RANK PERSISTENT HOMOLOGY (CWRPH)

A recent development in Toplogical Data Analysis (TDA) providing a new approach to the study of weighted networks that allows to recover complete and accurate long-range information from noisy redundant network, by building on persistent homology.

Petri, Giovanni, et al. "Topological strata of weighted complex networks." PloS one 8.6 (2013): e66506.

CWRPH IN A NUTSHELL - JHOLES

Rucco, Matteo, et al. "jHoles: A tool for understanding biological complex networks via clique weight rank persistent homology." Electronic Notes in Theoretical Computer Science 306 (2014): 5-18.

TOPOLOGY DRIVEN MODELLING: THE IS METAPHOR

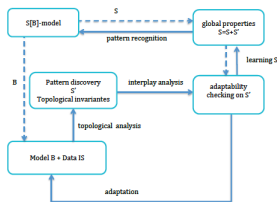This work has been inspired by...

Natural Computing
June 2014,
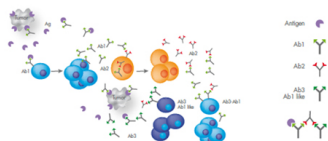
Open Access

Date: 24 Jun 2014

Topology driven modeling: the IS metaphor

Emanuela Merelli, Marco Pettini, Mario Rasetti

The S[B] model mimics the interplay capabilities of the immune system to identify, classify and learn new patterns.

IMMUNE SYSTEM: THE IDIOTYPIC NETWORK

Idiotypic network is formed by a chain of antibodies Abs. Two Abs are connected if and only if they show **Affinity**

An increase in the first antibody, AB1, recognized by AB2, would lead to an increase of the latter. The same would be seen regarding AB2, AB3, and so forth, always showing a tendency to restore the equilibrium of the system: **immune memory**

EXPERIMENTAL SETUP

Idiotypic Network has been simulated by the agent model based simulator *C-ImmSim*. Two simulations have been executed, in the first simulation the antigen has been injected only once while in the second simulation the injection has been executed twice and repeated with three configurations: $[Ag_2] = [Ag_1]$, $[Ag_2] > [Ag_1]$, $[Ag_2] < [Ag_1]$

```
PARAMETER FILE: DEFAULT
----------- Random Seed ------------------------------------
<12345> RS                /* random seed */
<0> Generating logfile  {0,1}
----------- Dimension determining params -------------------
<2> topology     0=cube, 1=sphere, 2=ellipsoid, 3=readfromfile
<12> NBITSTR             [even number]
<11> minmatch           [NBITSTR/2, NBITSTR-1]
<10> #mL                /* microL, integer number */
<480>
----------- Number of time steps--------------------------
<2190> num_step / a year is 1095 */
```

Graph pre-processing has been computed with *Matlab*, while TDA has been executed by **jHoles**.

FROM AFFINITY TO COEXISTENCE

The *Idiotypic Network* is represented as a weighted graph where the weight function is the pairwise Hamming distance between antibodies: the *Affinity* - $Aff(Ab_i, Ab_j)$ matrix.

Insted of $Aff(Ab_i, Ab_j)$ it is possible to use the so-called *coexistence function* $C_{Ab_{i,j}}(t)$:

$$C_{Ab_{i,j}}(t) = \frac{Aff(Ab_i(t), Ab_j(t)) * [Ab_i(t)] * [Ab_j(t)]}{\sum_{k=1}^{n}[Ab_k(t)]} \qquad (1)$$

For lower values of *Affinity* the concentration must be more significant, the match between antibodies is less probable.

PERSISTENT ENTROPY

Instead of computing an entropy for graph I defined an entropy Shannon-like based on *persistent homology*.
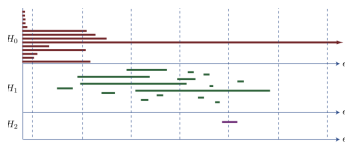
Def Persistent entropy: given a filtered topological space equipped with an ascending filtration algorithm, the set of filtration value F and the corresponding persistence bar-code $B = [a_j ; b_j] : j \in J$. A persistent line in a bar-code is conventionally represented as $[a_j ; \infty)$ here it is substitute with $[a_j ; m)$ where m=(max$\{F\} + 1$).
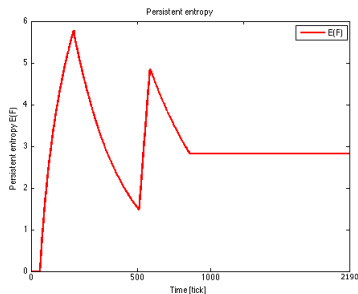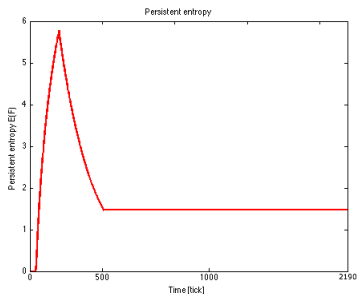
$$E(F) = -\sum_{j \in J} p_j log(p_j) \qquad (2)$$

Where $p_j = l_j/L$, $l_j = b_j - a_j$, and $L = \sum_{j \in J} l_j$

REMARKS ON PERSISTENT ENTROPY

From an information-theoretic view point, we have to interpret the number of significant intervals as the coding length. Entropy measures how different bars of the barcodes are in length. A barcode with uniform lengths has small entropy.

AVERAGE PERSISTENT ENTROPY VERSUS TIME [TICK]

Left: IS stimulated once. Right: IS stimulated twice.
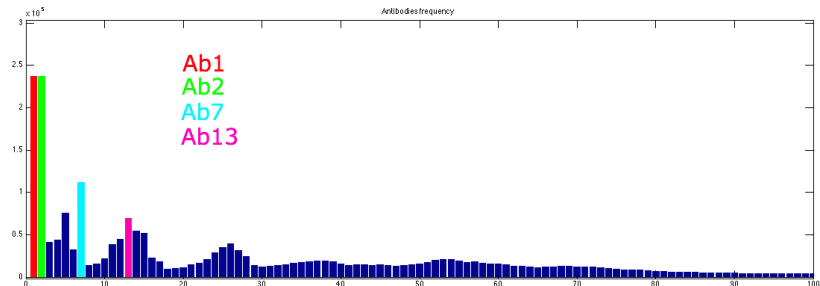A peak indicates the *immune activation*, the plateau indicates the *immune memory*.

Statistics over 100 trials.

TOPOLOGICAL-HOLES ANALYSIS: COMMUNITIES DETECTION

Topological-holes are formed by *generators*:
0,1,2,...,n-simplices.
From the loop frequency analysis we found a loop present in both the *immune responses* and in the *immune memory*: *[1, 2] + [2, 7] + [7, 13] + [1, 13].*

CONCLUDING REMARKS

TDA efficiently recognizes the persistent holes and their generators: the antibodies and how Ab_s are connected. Its computational complexity in the worst case is $O(n^2)$

Persistent entropy allows to study the roles of the persistent antibodies highglighting that they have the central role both in the functional and connectivity stability of the networks.

REFERENCES

▶ Merelli, Emanuela, and Mario Rasetti. "The immune system as a metaphor for topology driven patterns formation in complex systems." Artificial Immune Systems. Springer Berlin Heidelberg, 2012. 289-291.

▶ Petri, G., et al. "Homological scaffolds of brain functional networks." Journal of The Royal Society Interface 11.101 (2014): 20140873.

▶ Felice, Domenico, Stefano Mancini, and Marco Pettini. "Quantifying networks complexity from information geometry viewpoint." Journal of Mathematical Physics 55.4 (2014): 043505.

▶ Felice, Domenico, et al. "A geometric entropy measuring networks complexity." arXiv preprint arXiv:1410.5459 (2014).

ACKNOWLEDGEMENTS