TDA and Persistent Homology: a new method for analysing temporal graphs

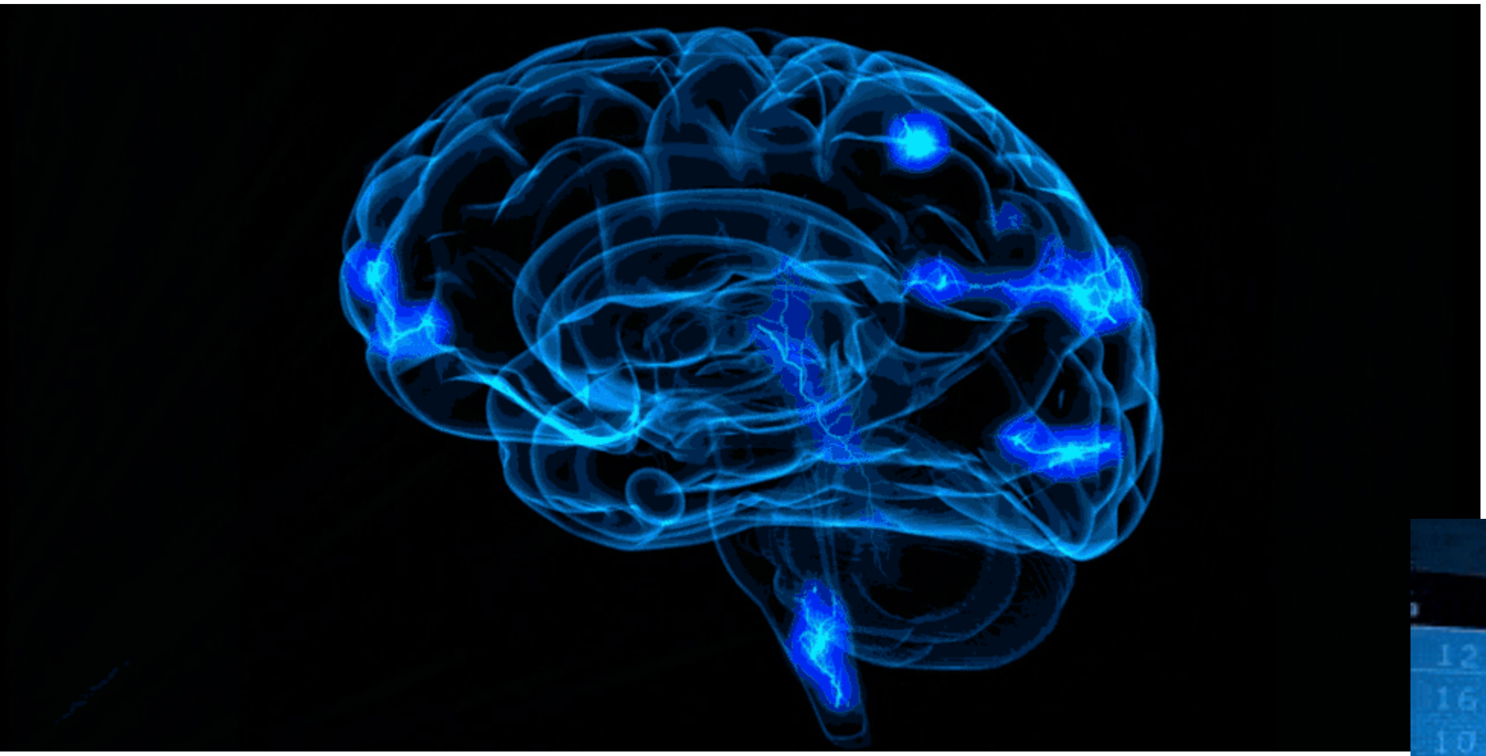
Marco Piangerelli - Emanuela Merelli
marco.piangerelli@unicam.it

Algorithmic Aspects on Temporal Graphs II
ICALP2019 - Patras
08/07/2019

Outline

- Complex Systems
- From Complex System to temporal graphs
- Why Topological Data Analysis?
- Topology, Filtration & Homology
- Persistent Entropy
- Results

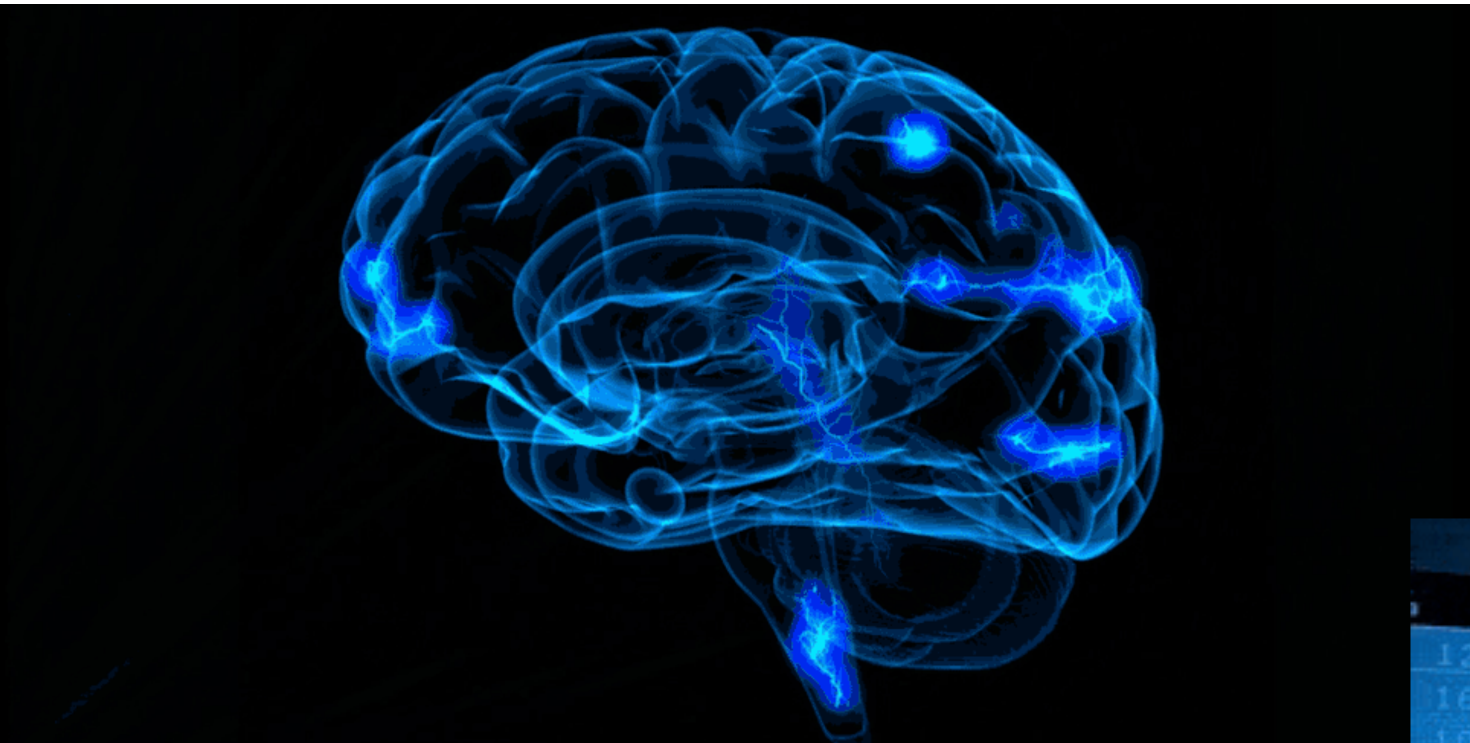
Complex Systems



The Human Brain

The Stock Market

Complex Systems

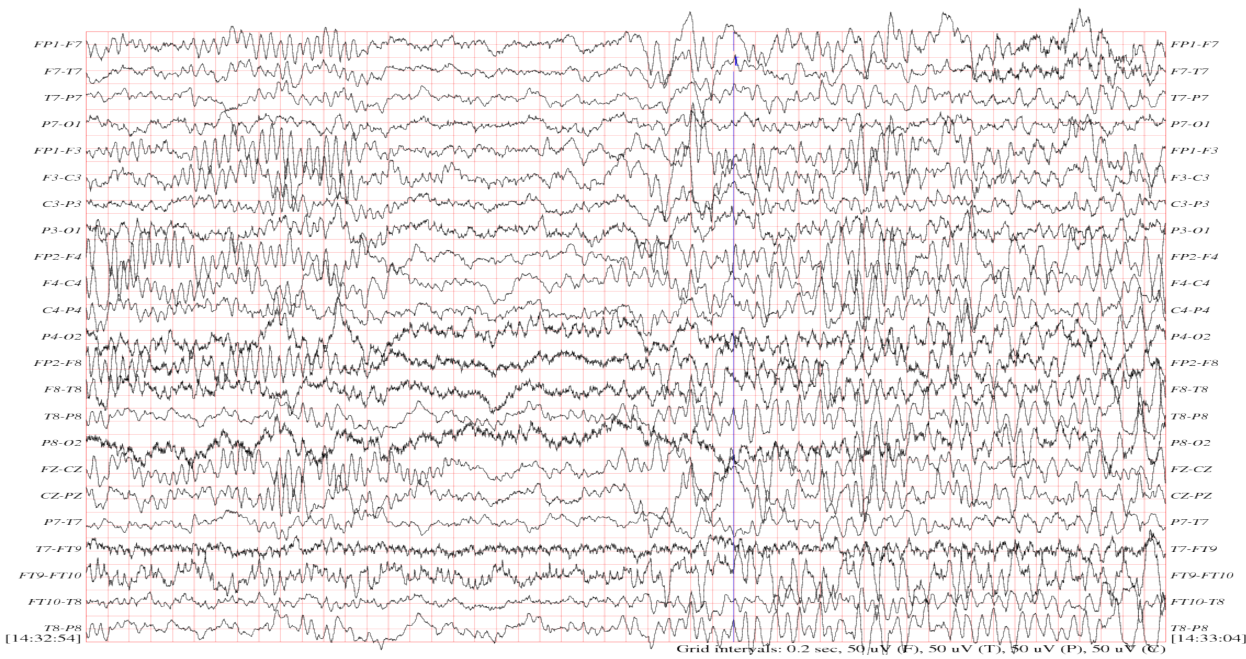


The Human Brain

The Stock Market

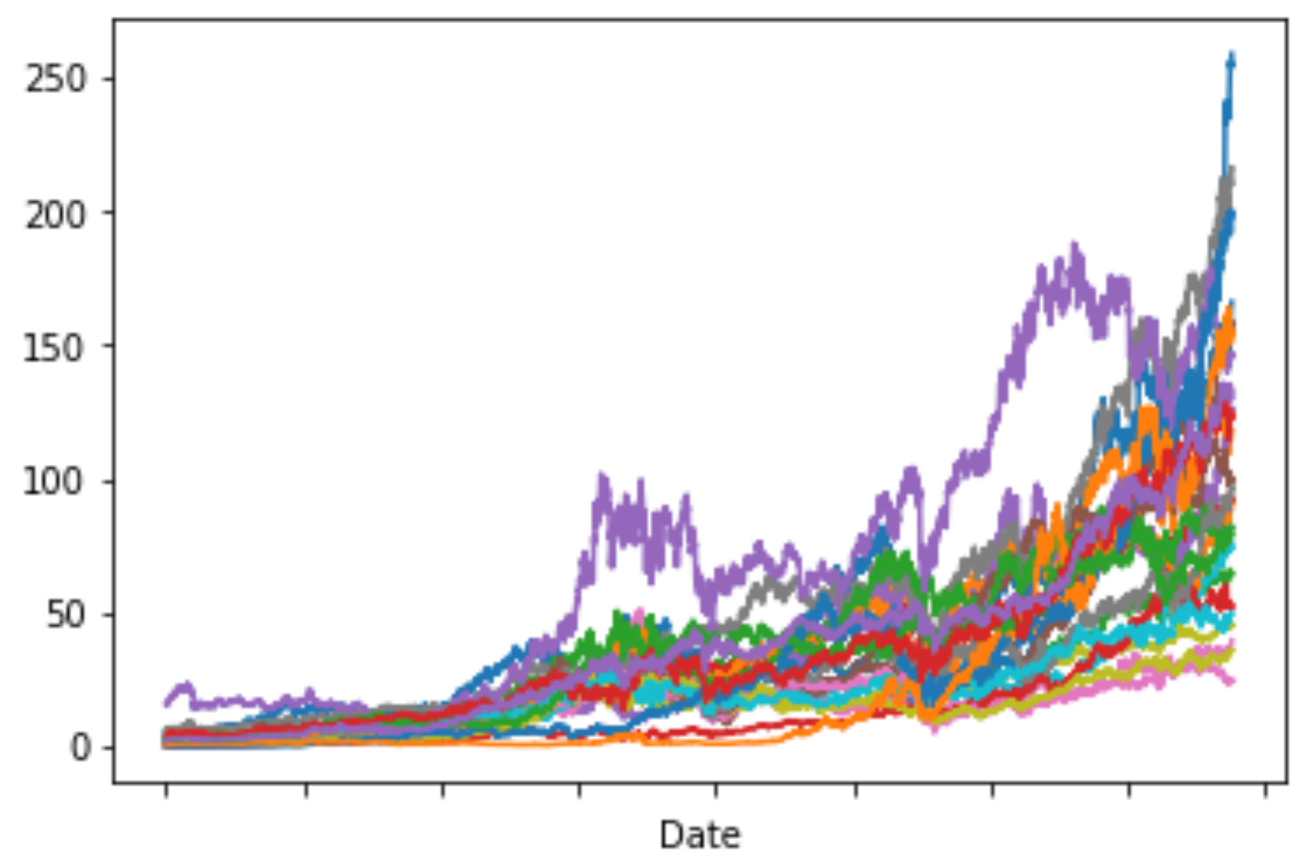
Extracting Emerging GLOBAL behaviors

Complex Systems

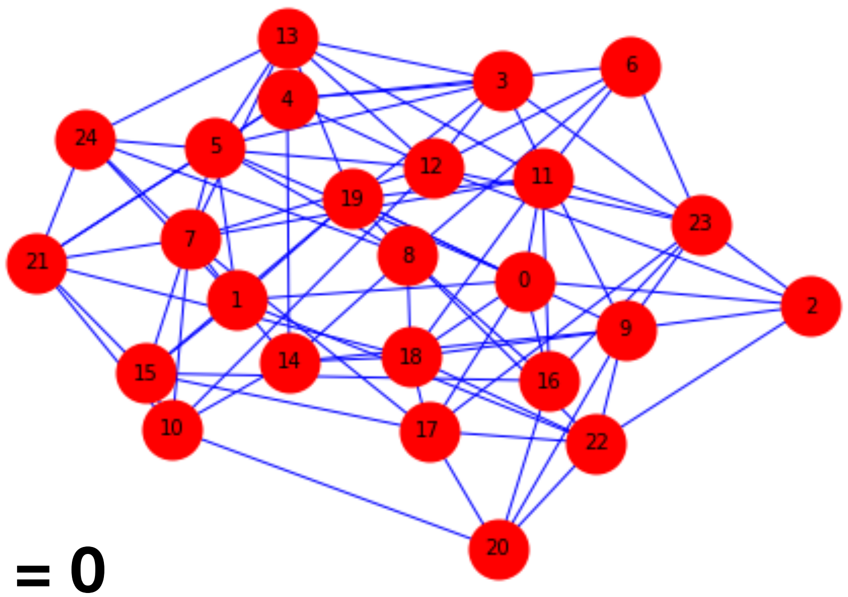


The Human Brain
(Epileptic Seizures (1h))

The Stock Market
(Dow Jones (1980-2017))

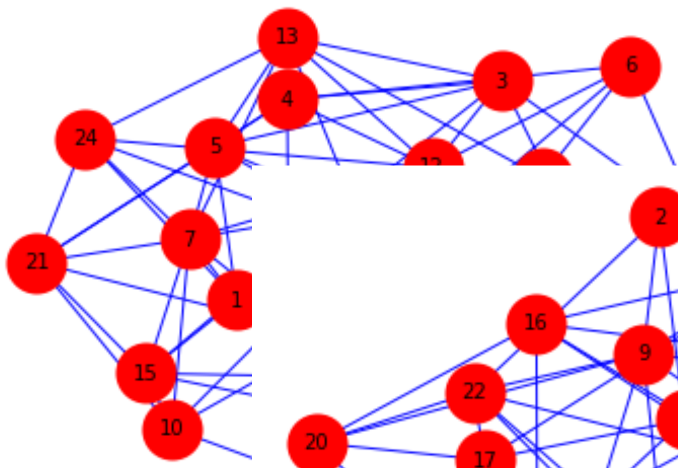


Temporal Graphs

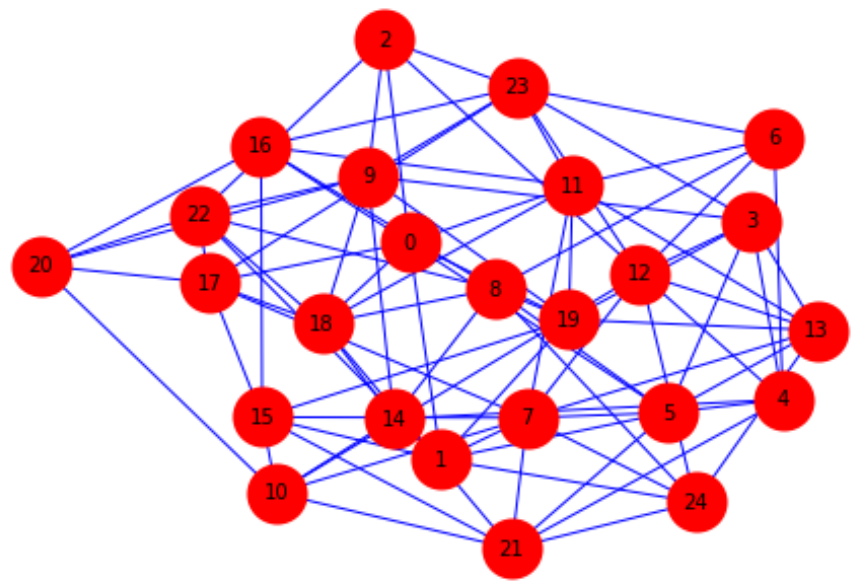


t = 0

Temporal Graphs



t = 0

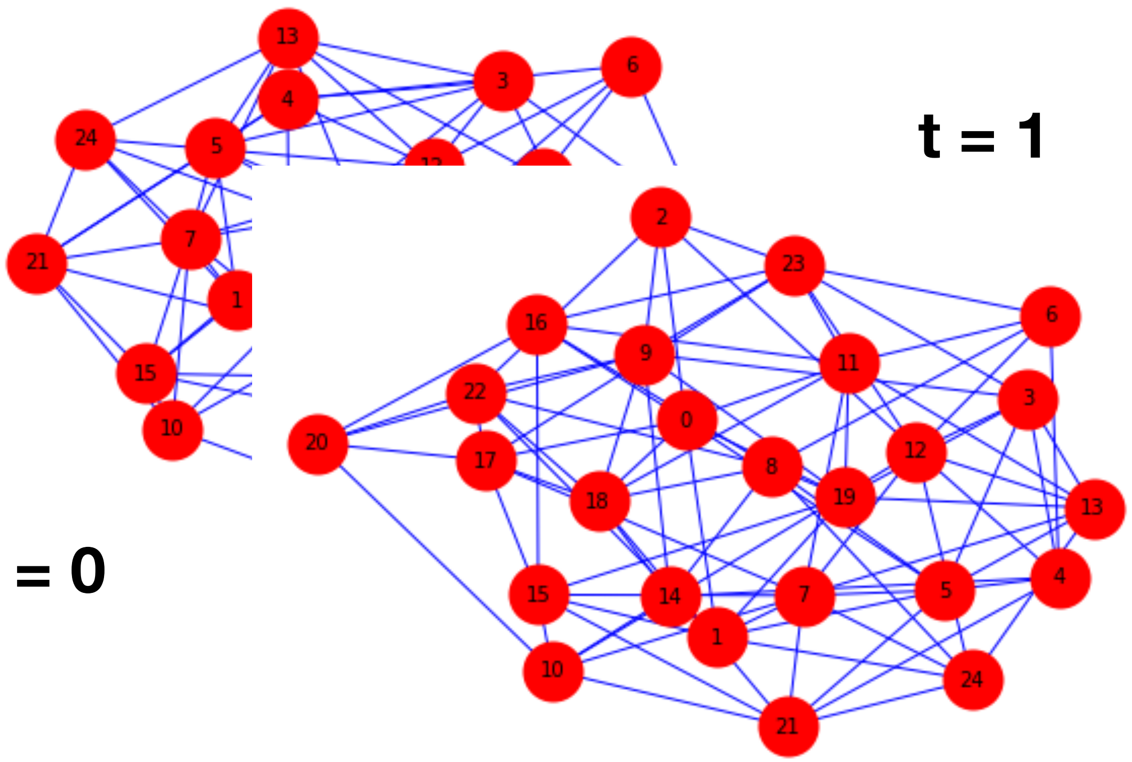


t = 1

...

Temporal Graphs

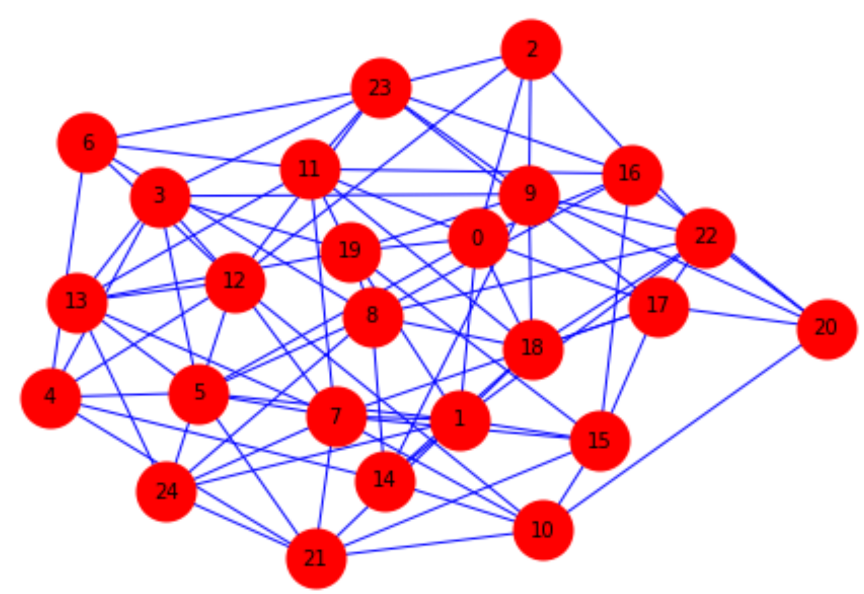
t = 1



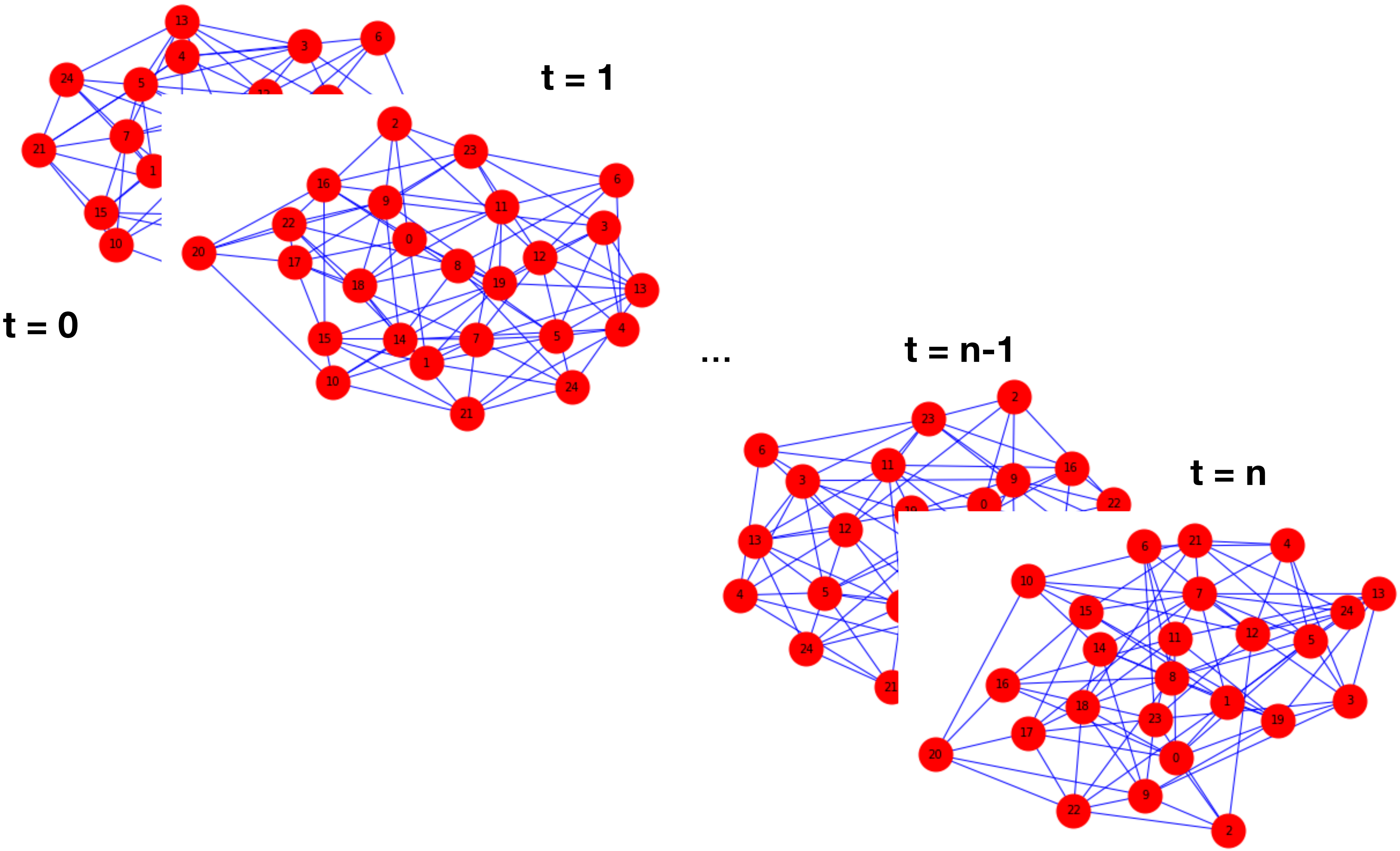
t = 0

...

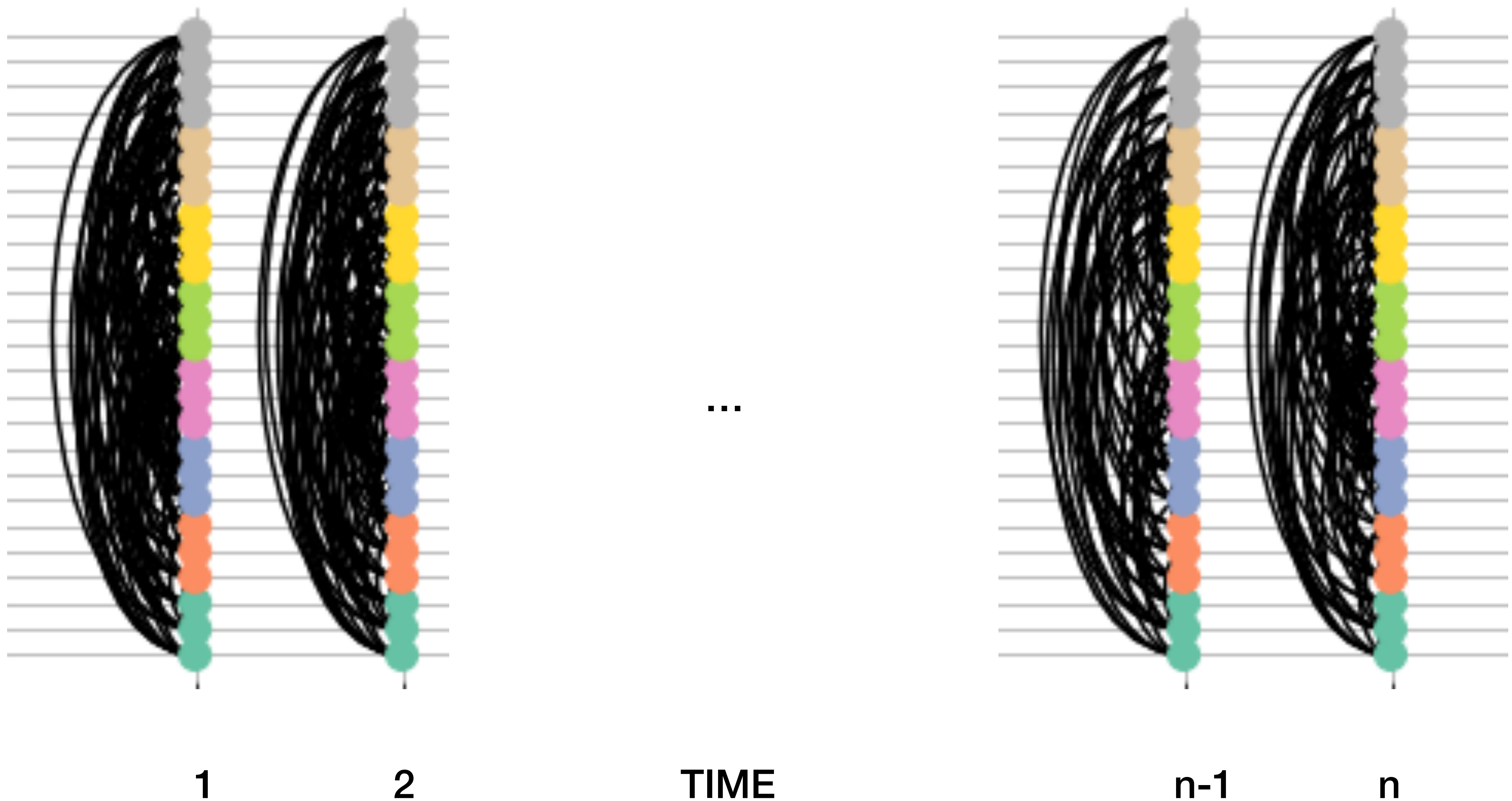
t = n-1



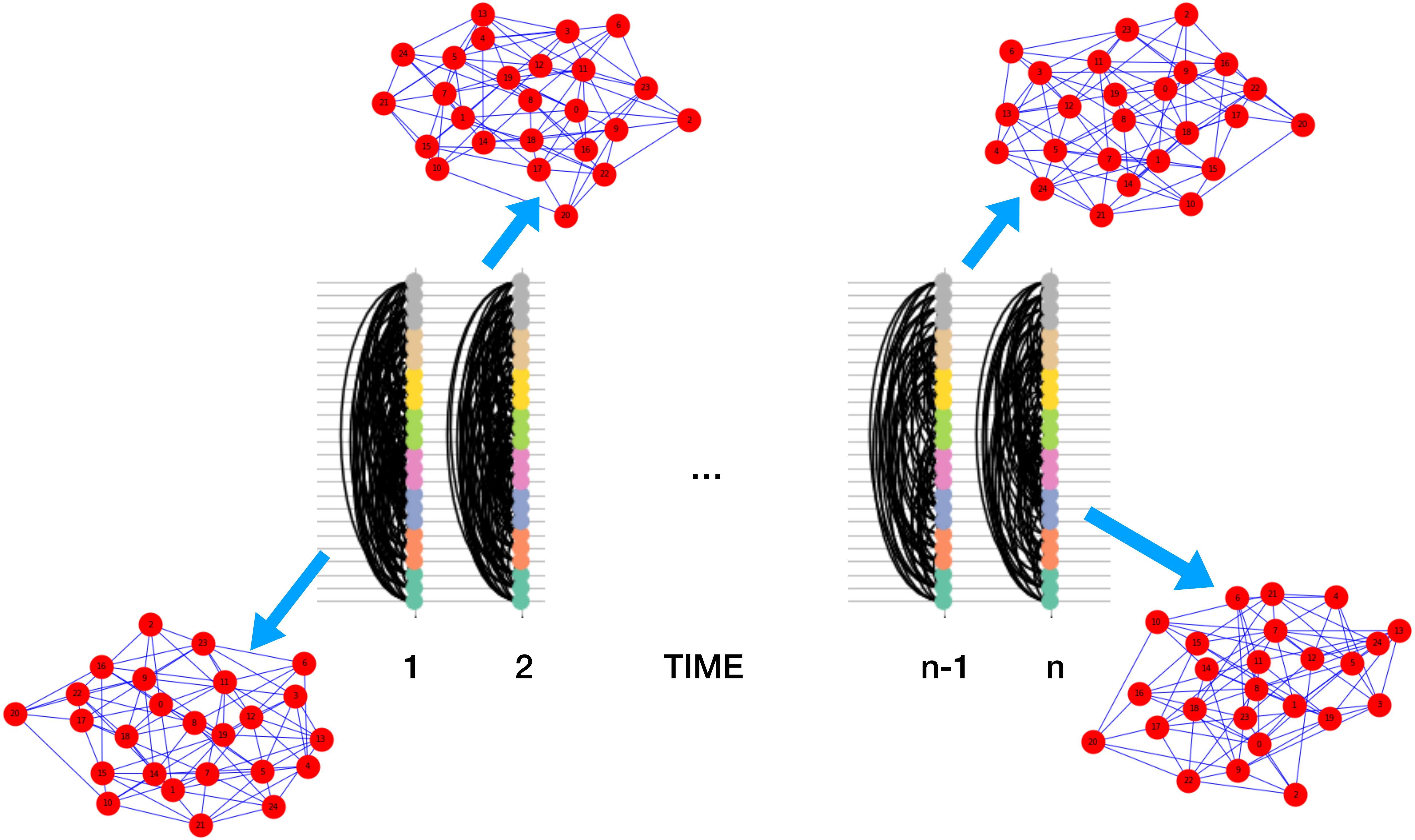
Temporal Graphs



Temporal Graphs



Temporal Graphs

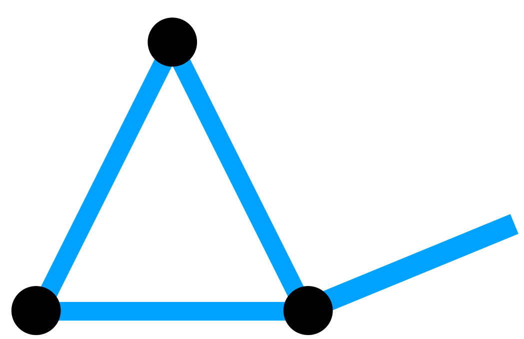


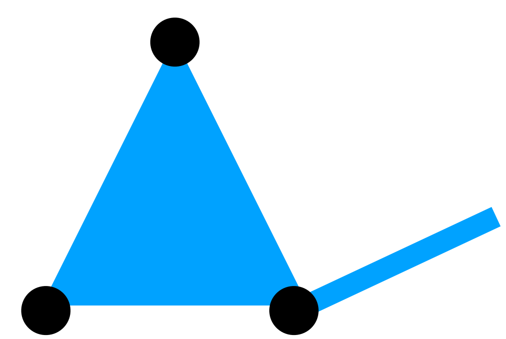
Why topological data analysis?

Data $\longrightarrow$ (Global) Information $\longrightarrow$ Knowledge

Why topological data analysis?

Data $\longrightarrow$ (Global) Information $\longrightarrow$ Knowledge



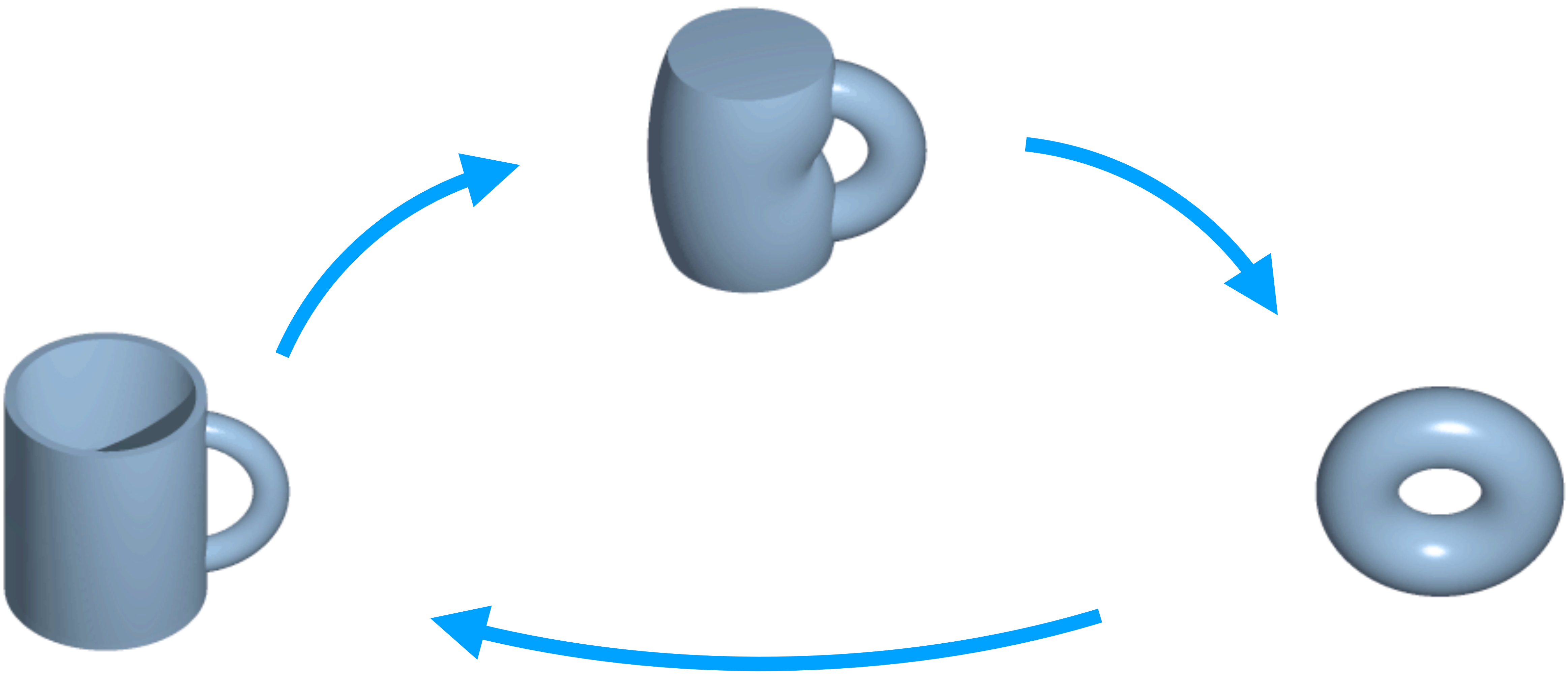


What is topology?

In mathematics, **topology** (from the Greek τόπος, *place*, and λόγος, *study*) is concerned with the properties of space that are preserved under continuous deformations:

- Allowed: Stretching, Twisting, Bending
- Forbidden: Cutting, Gluing

What is topology?



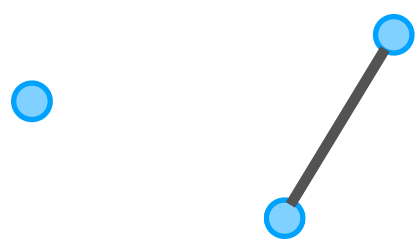
Topological Data Analysis (TDA)

A simplicial complex is a discrete topological space, obtained from the union of simplicies

0-simplex

Topological Data Analysis (TDA)

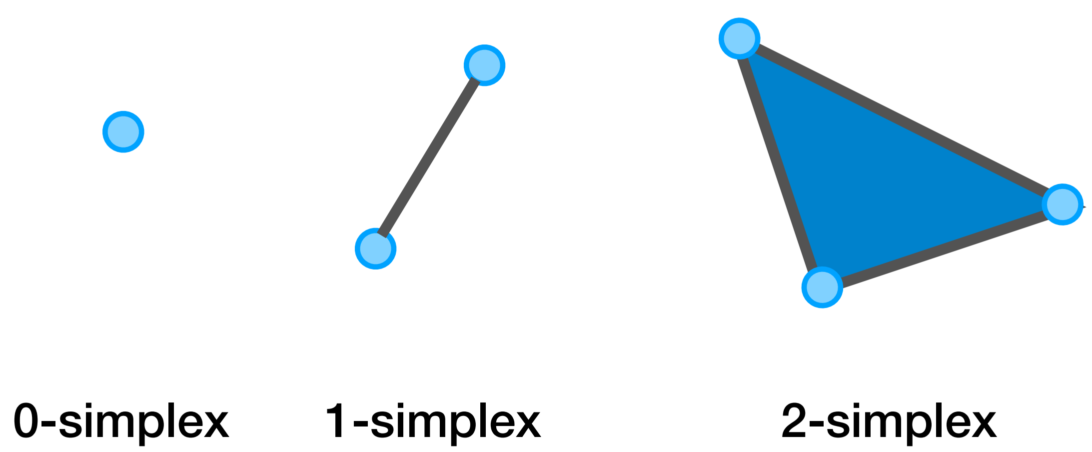
A simplicial complex is a discrete topological space, obtained from the union of simplicies



0-simplex 1-simplex

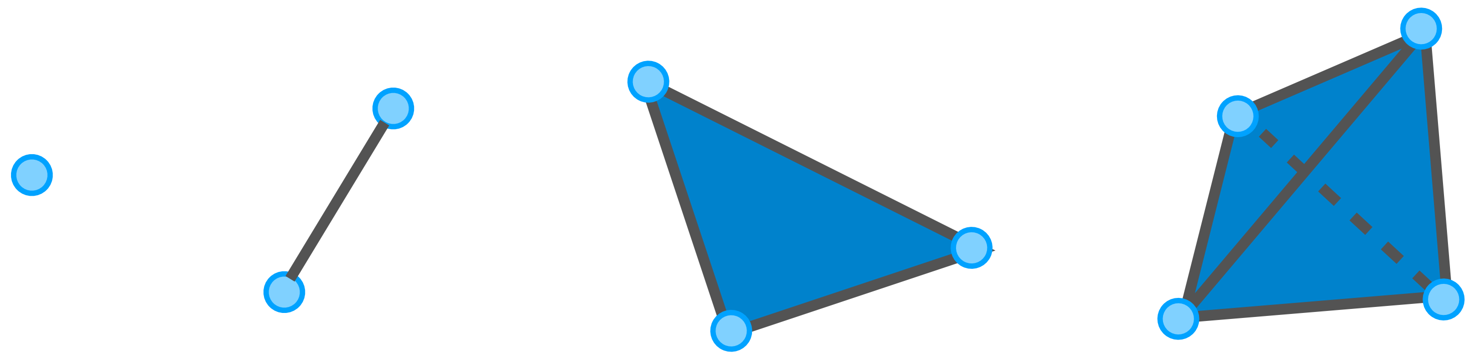
Topological Data Analysis (TDA)

A simplicial complex is a discrete topological space, obtained from the union of simplicies



Topological Data Analysis (TDA)

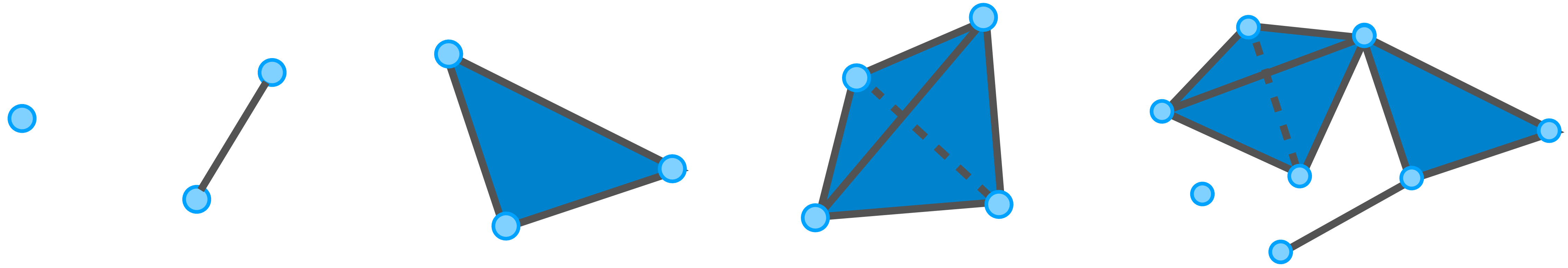
A simplicial complex is a discrete topological space, obtained from the union of simplicies



0-simplex 1-simplex 2-simplex 3-simplex

Topological Data Analysis (TDA)

A simplicial complex is a discrete topological space, obtained from the union of simplicies



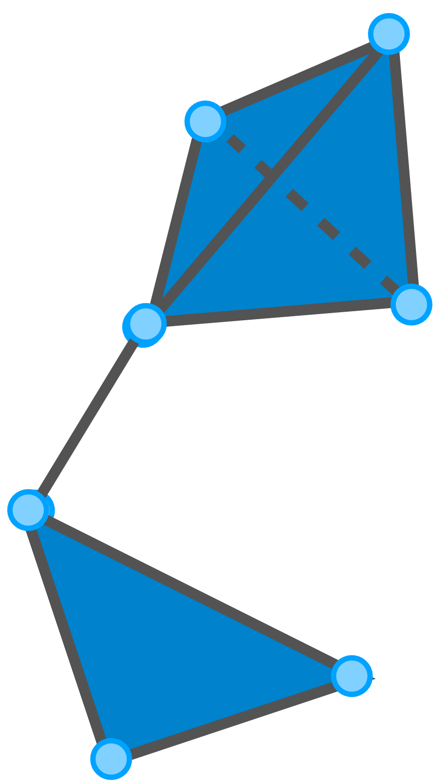
0-simplex 1-simplex 2-simplex 3-simplex Simplicial Complex

Topological Data Analysis (TDA)

Homology allows to compute the number of n-dimesional holes

Topological Data Analysis (TDA)

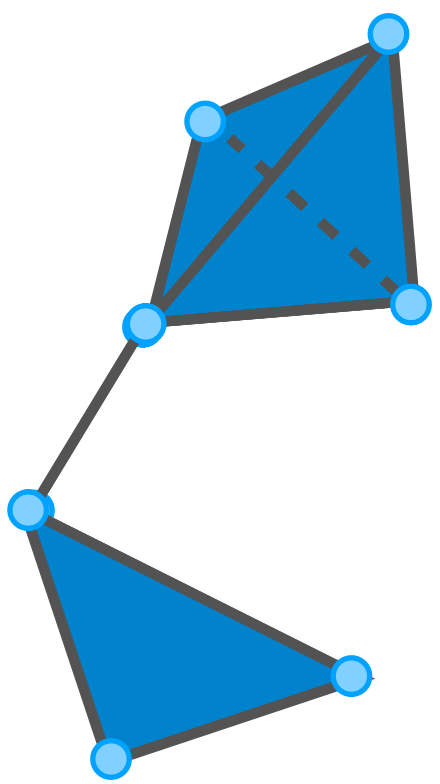
Homology allows to compute the number of n-dimesional holes



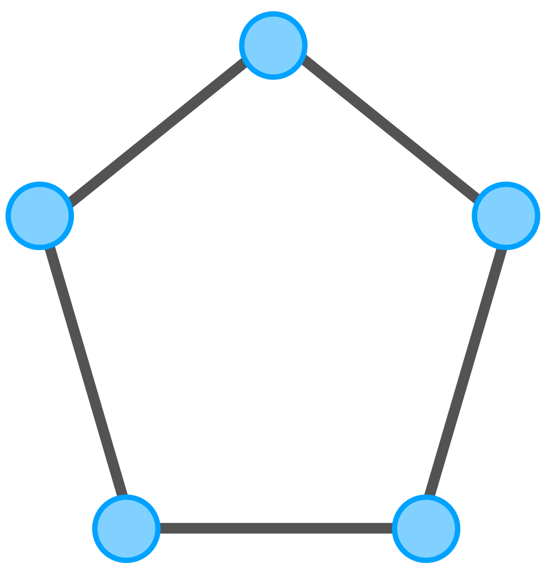
A connected component is a 0-dimensional hole

Topological Data Analysis (TDA)

Homology allows to compute the number of n-dimesional holes



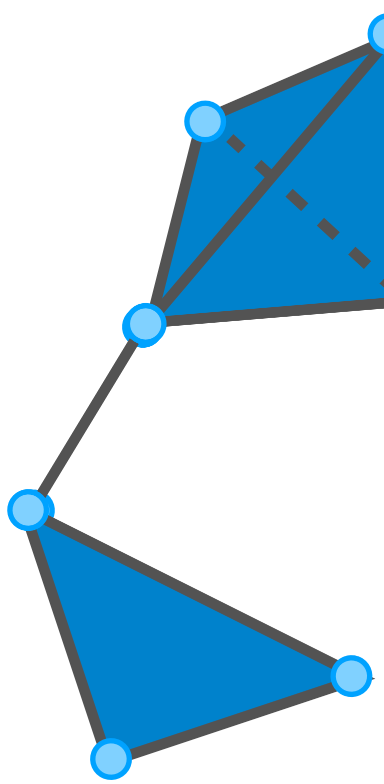
A connected component is a 0-dimensional hole



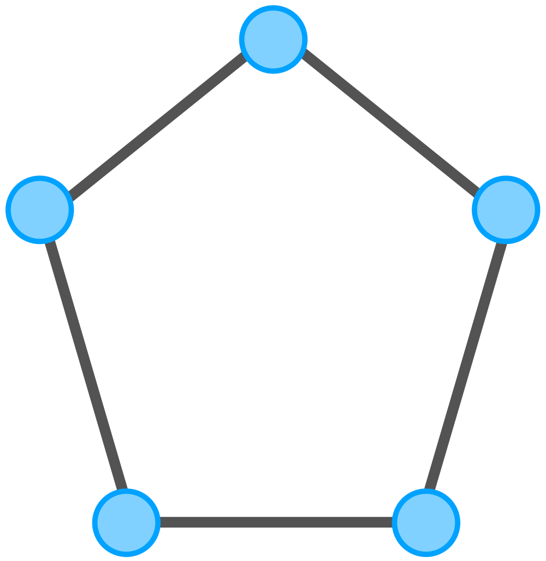
A loop of more than 3 vertices is a 1-dimensional hole

Topological Data Analysis (TDA)

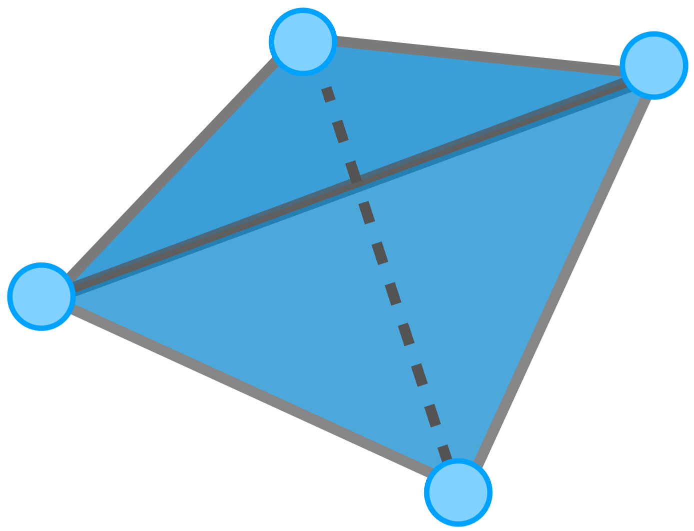
Homology allows to compute the number of n-dimesional holes



A connected component is a 0-dimensional hole



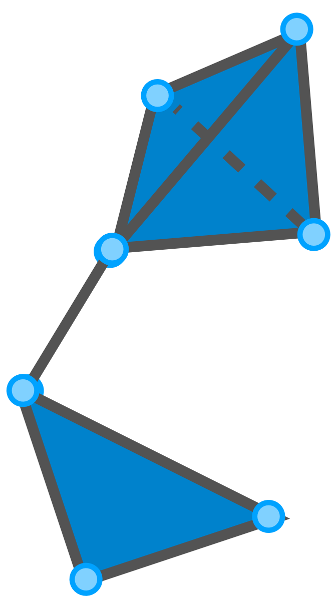
A loop of more than 3 vertices is a 1-dimensional hole



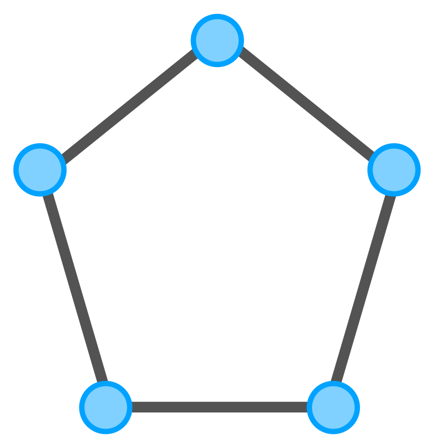
An empty solid is a cavity, or a tunnel, and it is a 2-dimensional hole

Topological Data Analysis (TDA)

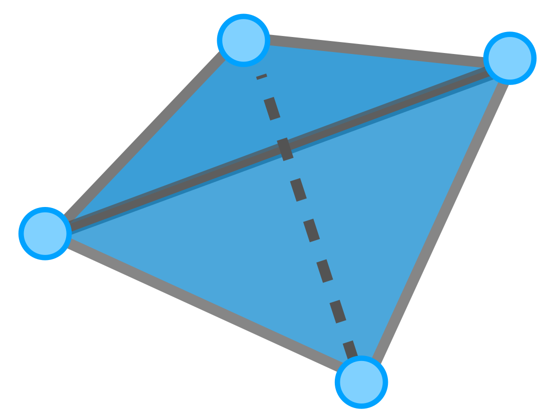
Homology allows to compute the number of n-dimesional holes



A connected component is a 0-dimensional hole



A loop of more than 3 vertices is a 1-dimensional hole



An empty solid is a cavity, or a tunnel, and it is a 2-dimensional hole

?

3-dimensional hole

Topological Data Analysis (TDA)

- We want to recover the space of origin of our data
- We want to obtain some quantity for characterizing the space
- Those quantities are the topological invariants
- Many topological invariants exist:

A. Euler Characteristics

B. Betti Numbers (β_0, β_1 , …)

C. Torsion Coefficients

D. …

Persistent Homology

$$H_k = \frac{ker\,\partial_k(C_k)}{Img\,\partial_{k+1}(C_k)} = \frac{Z_n}{B_n}$$

$$rank(H_k) := \beta_k$$

Persistent Homology

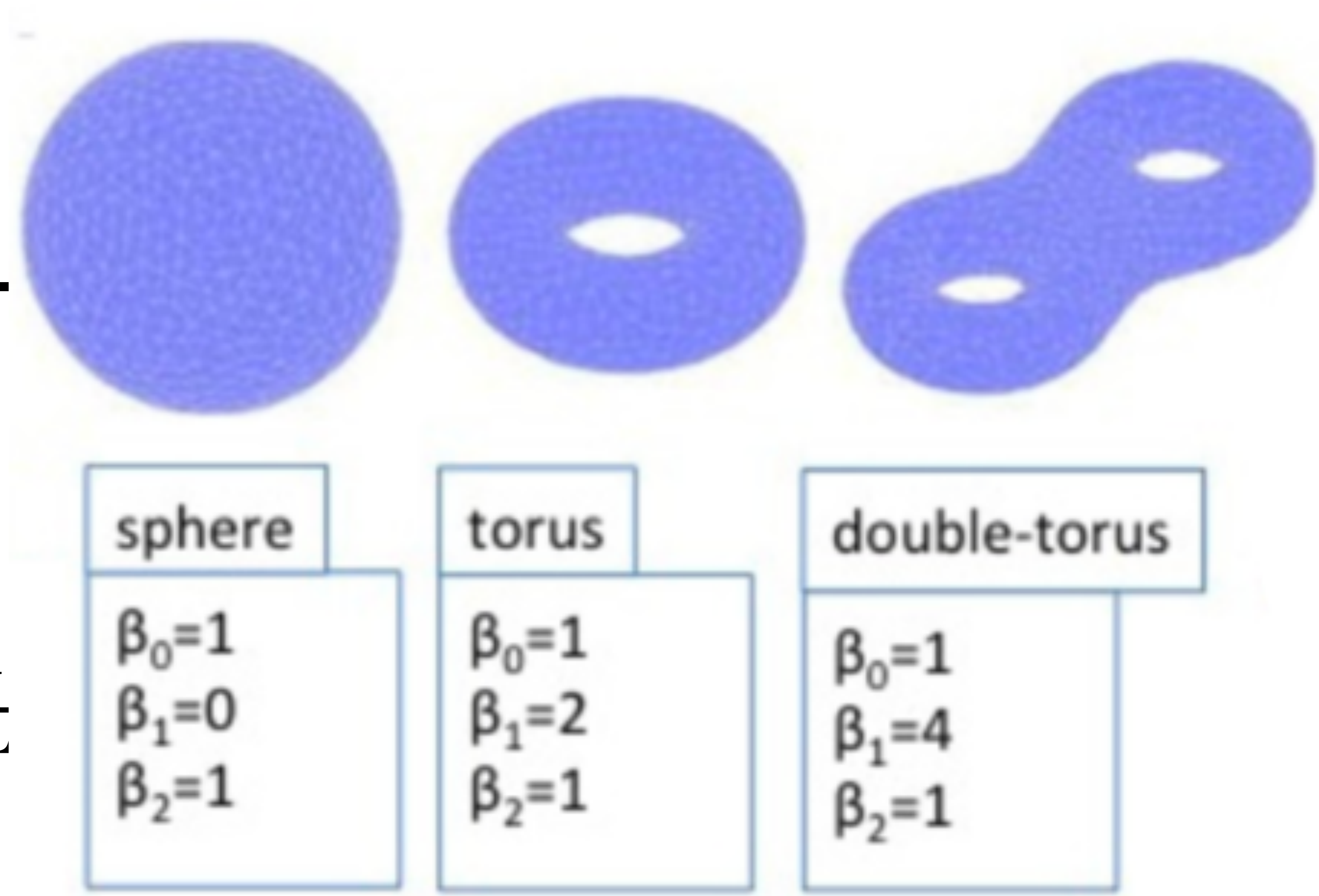
$$H_k = \frac{ker\,\partial_k(C_k)}{Img\,\partial_{k+1}(C_k)} = \frac{Z_n}{B_n}$$

$\longrightarrow$ Linear Algebra

$$rank(H_k) := \beta_k$$

Persistent Homology

$H_k =$ -



lgebra

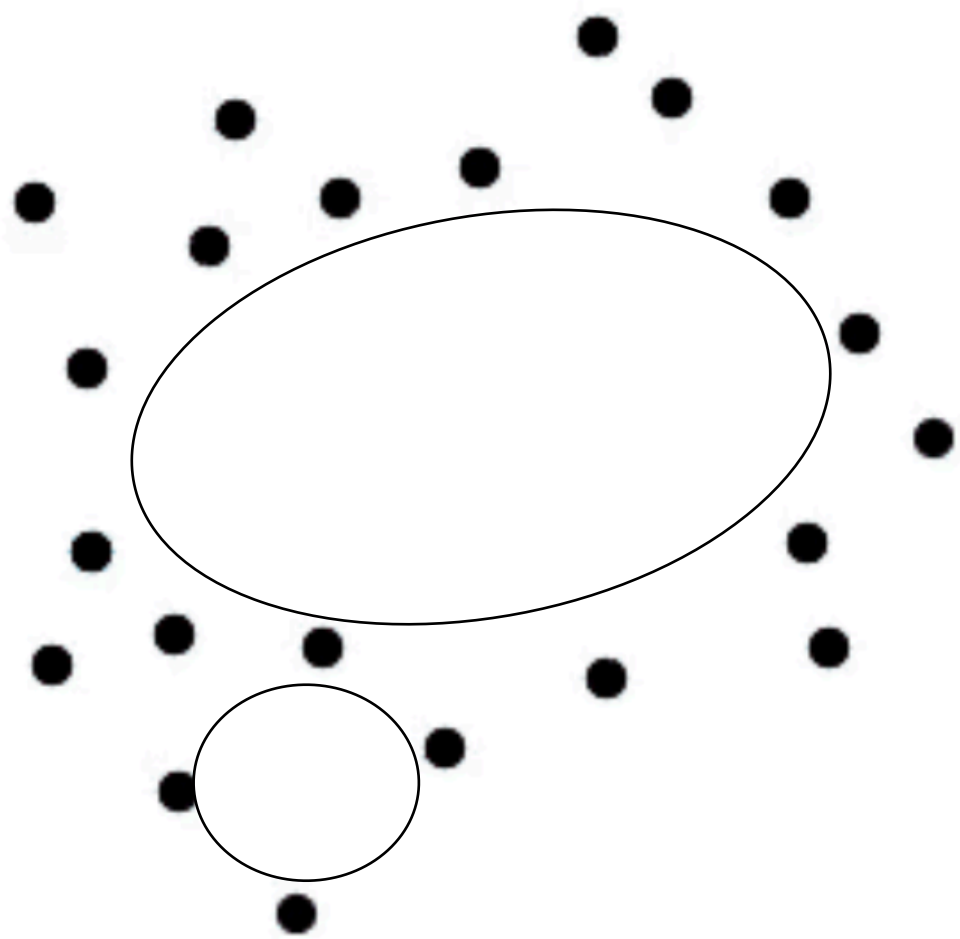
H

Filtration

- Cech & Vietoris Rips Filtration
- Clique Weighted Rank Filtration

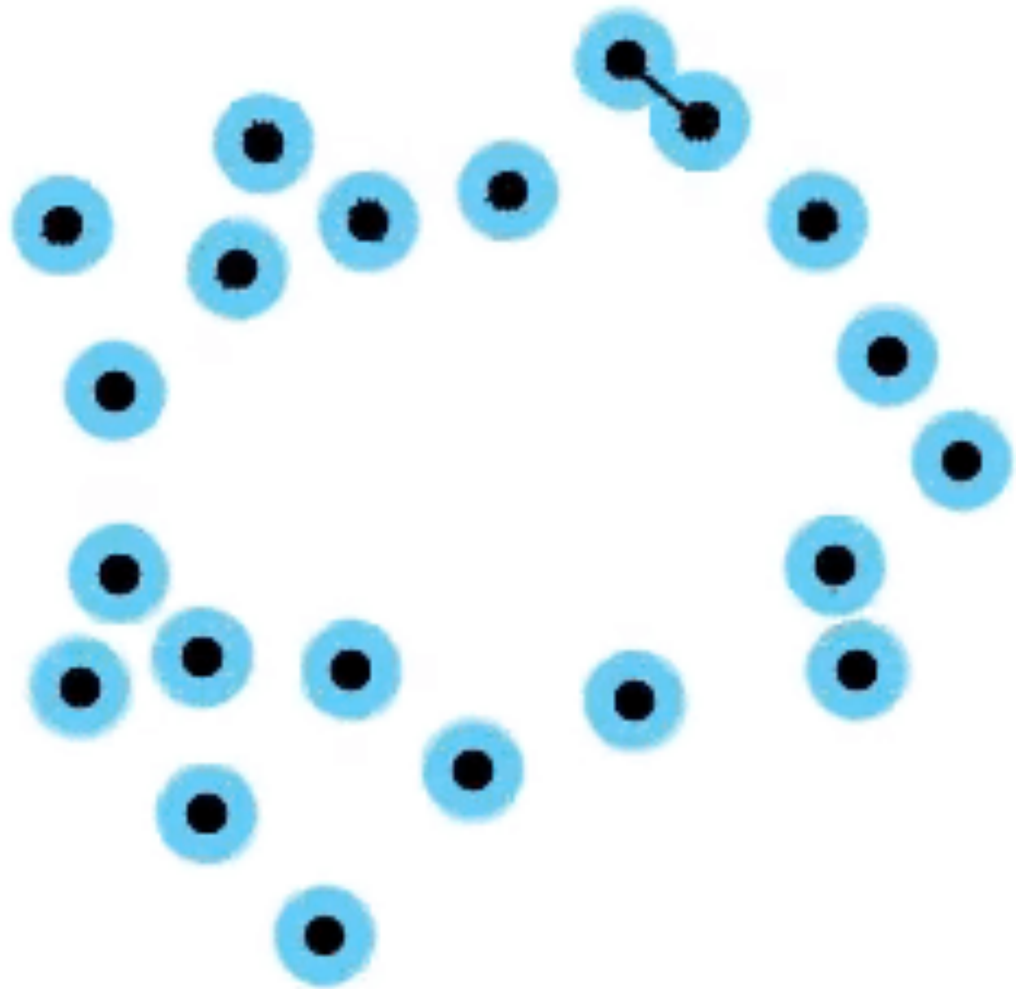
Vietoris Rips Filtration

Point Cloud



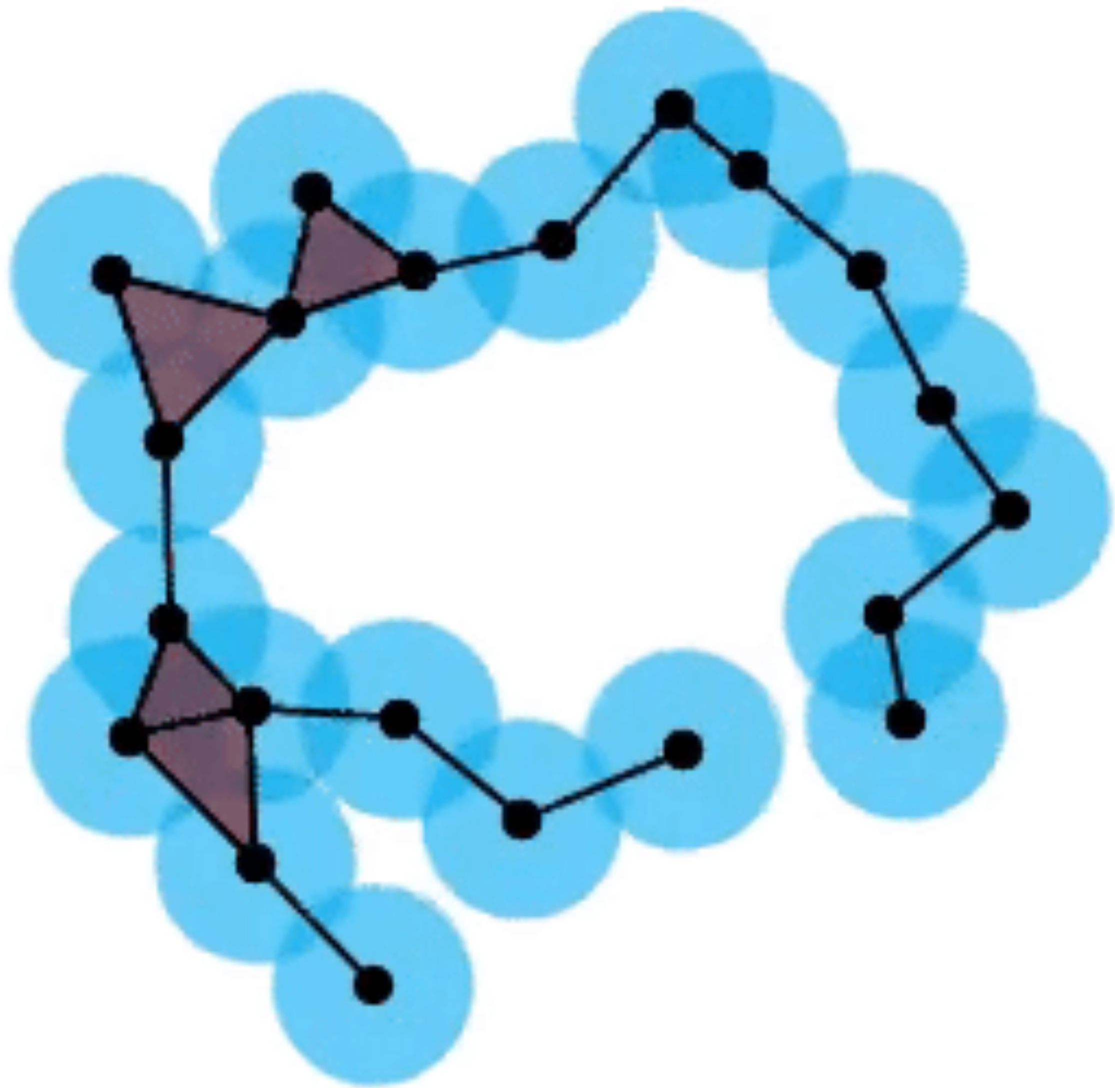
Vietoris Rips Filtration

Point Cloud



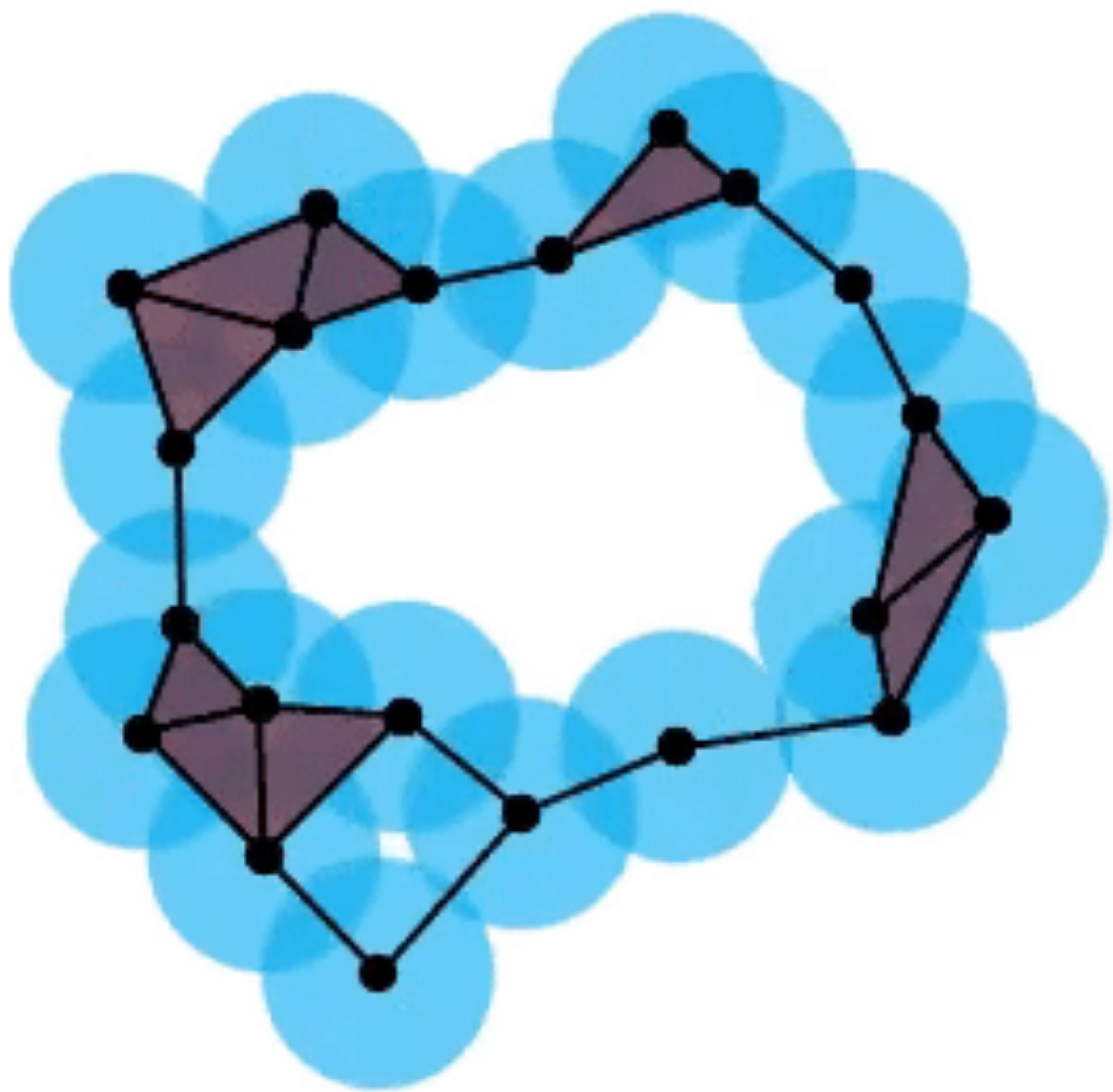
Vietoris Rips Filtration

Point Cloud



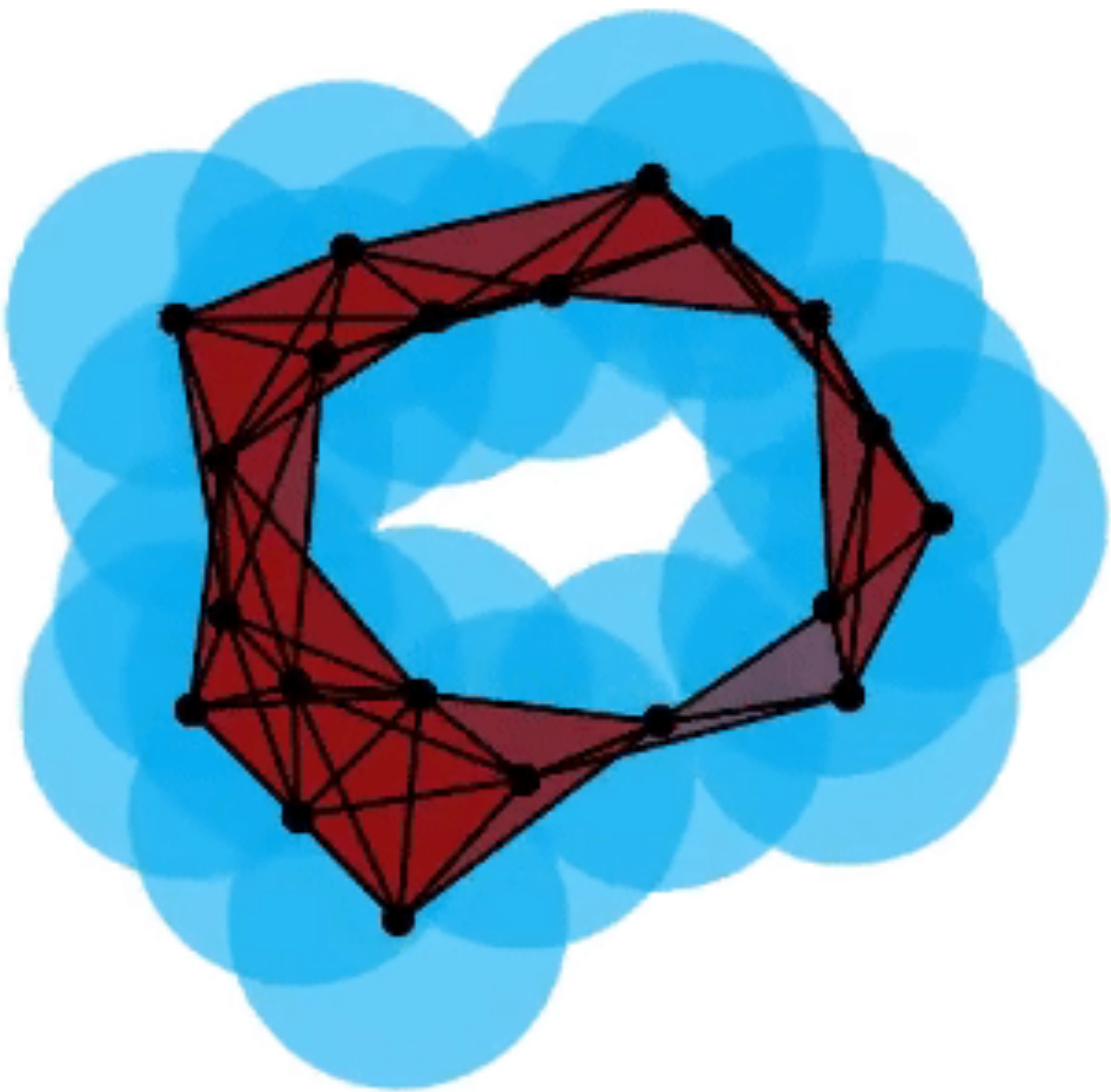
Vietoris Rips Filtration

Point Cloud

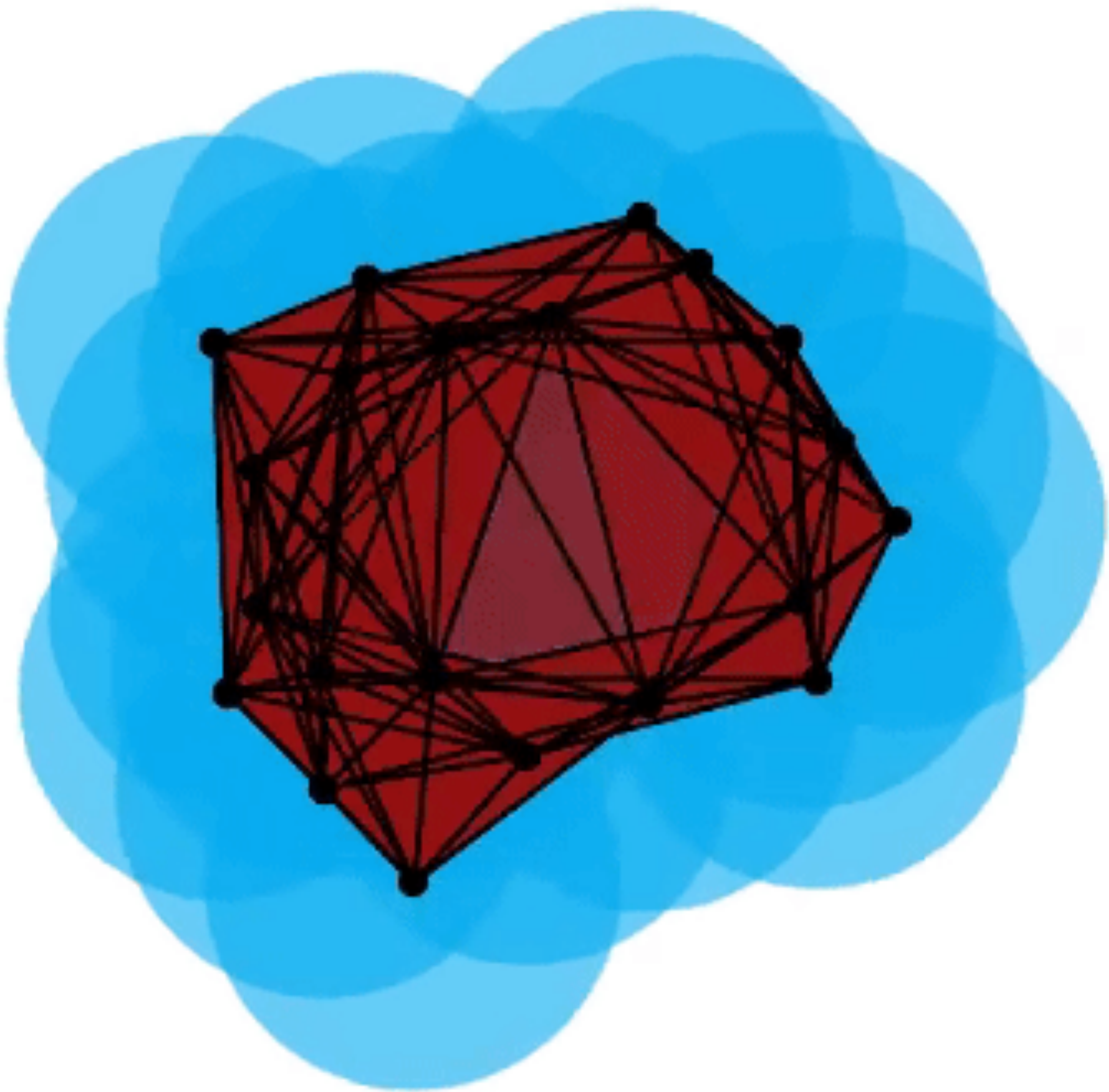


Vietoris Rips Filtration

Point Cloud



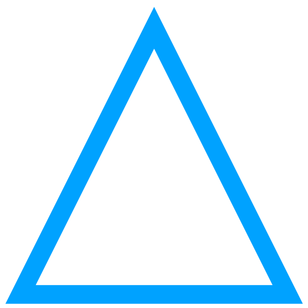
Vietoris Rips Filtration



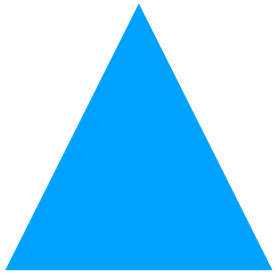
Point Cloud

Clique Weighted Rank Filtration

- A k-Clique is equivalent to a (k-1)-simplex



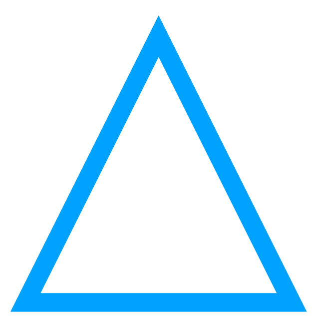
3 - clique



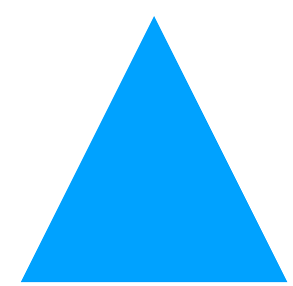
2-simplex

Clique Weighted Rank Filtration

- A k-Clique is equivalent to a (k-1)-simplex



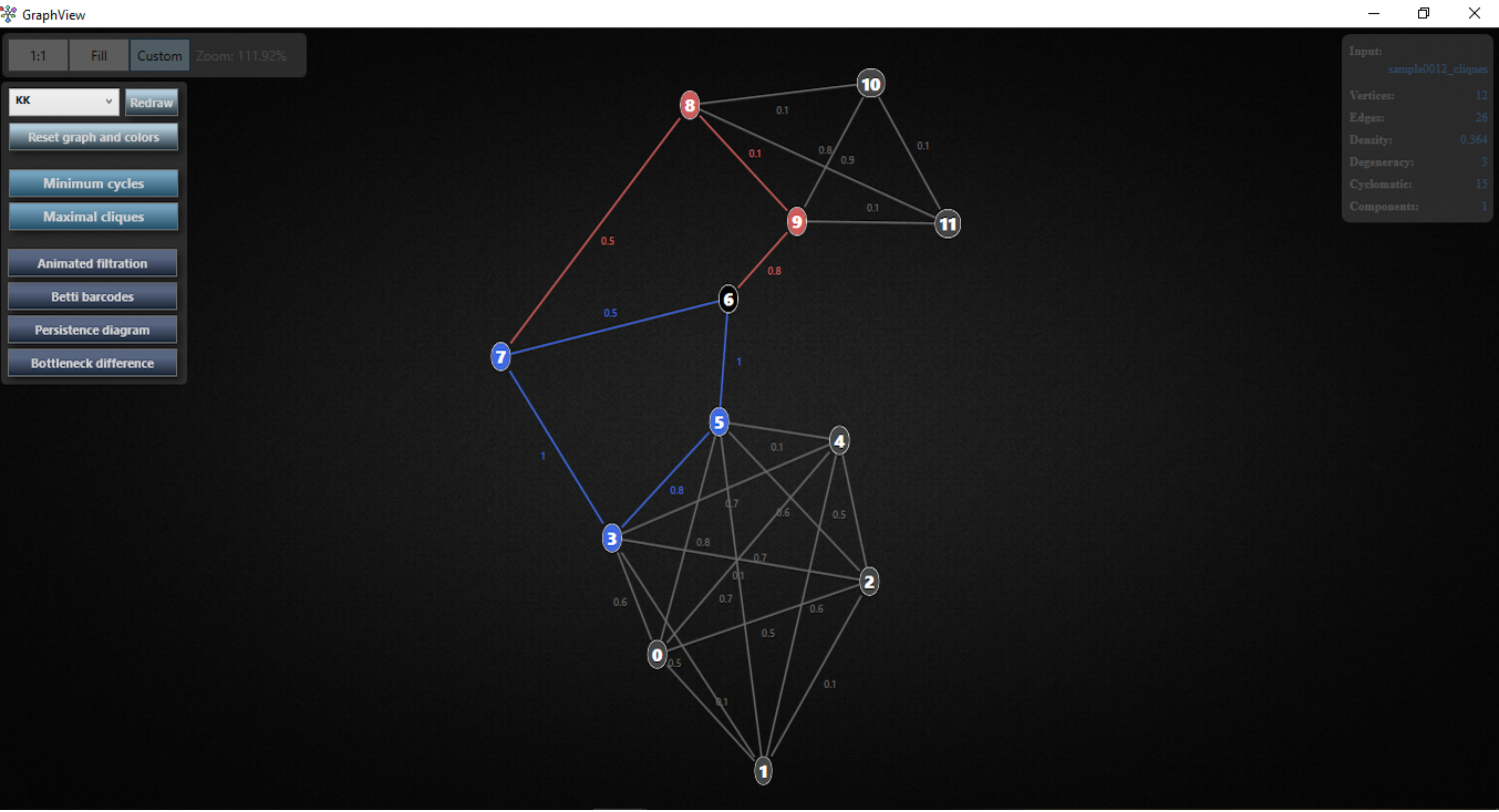
3 - clique



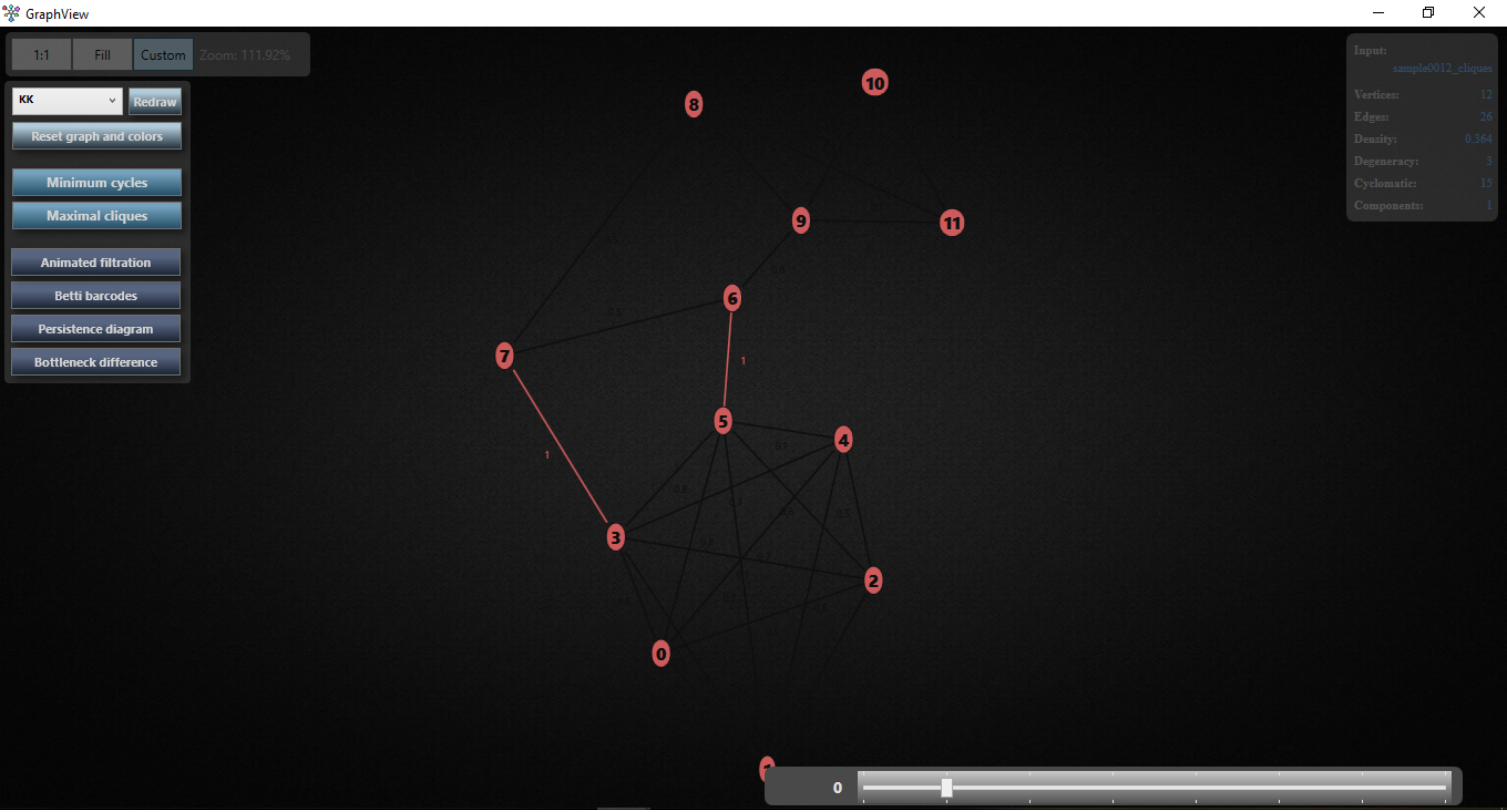
2-simplex

Bron-Kerbosch ($O(3^{n/3})$)

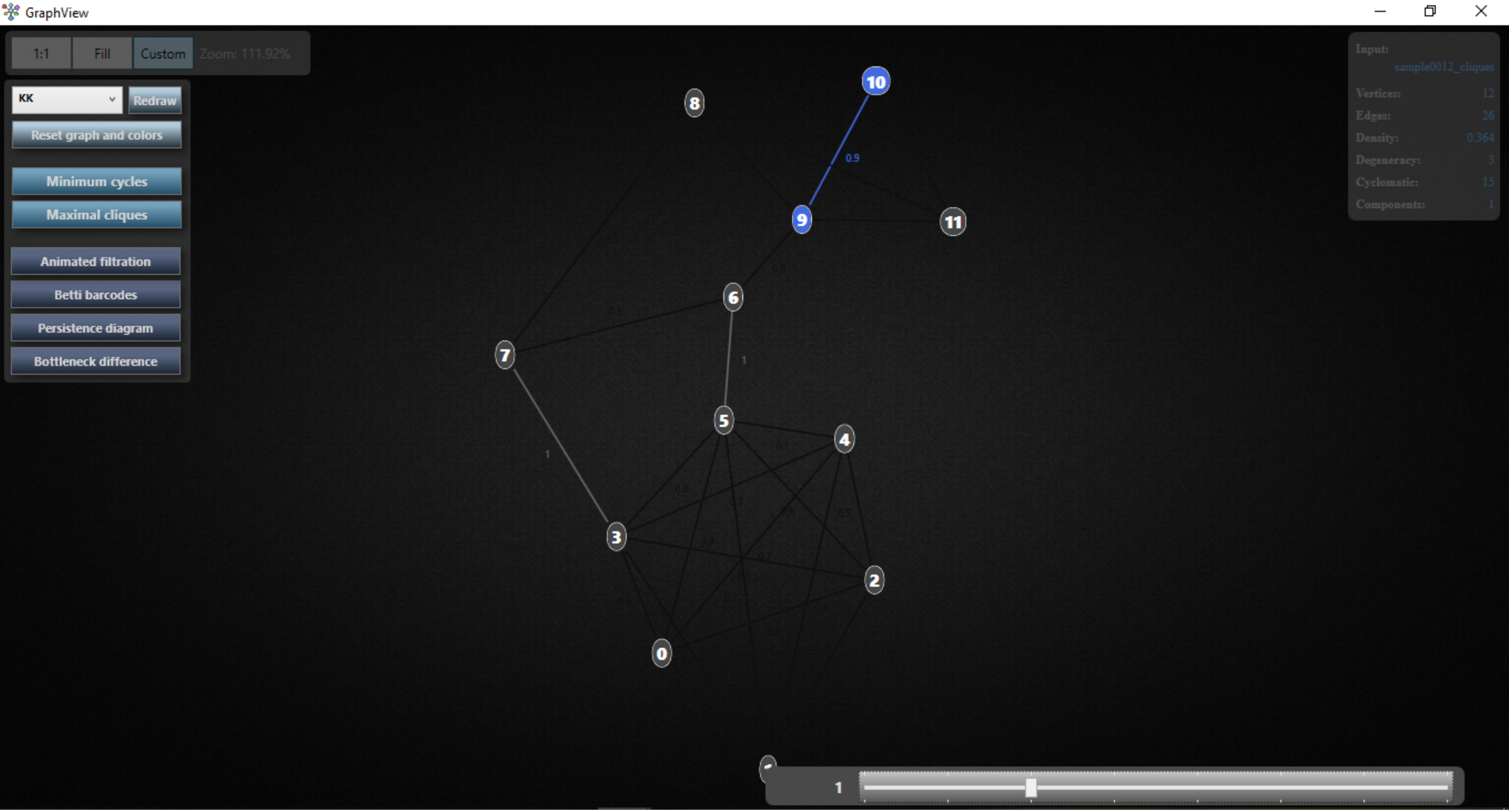
Clique Weighted Rank Filtration



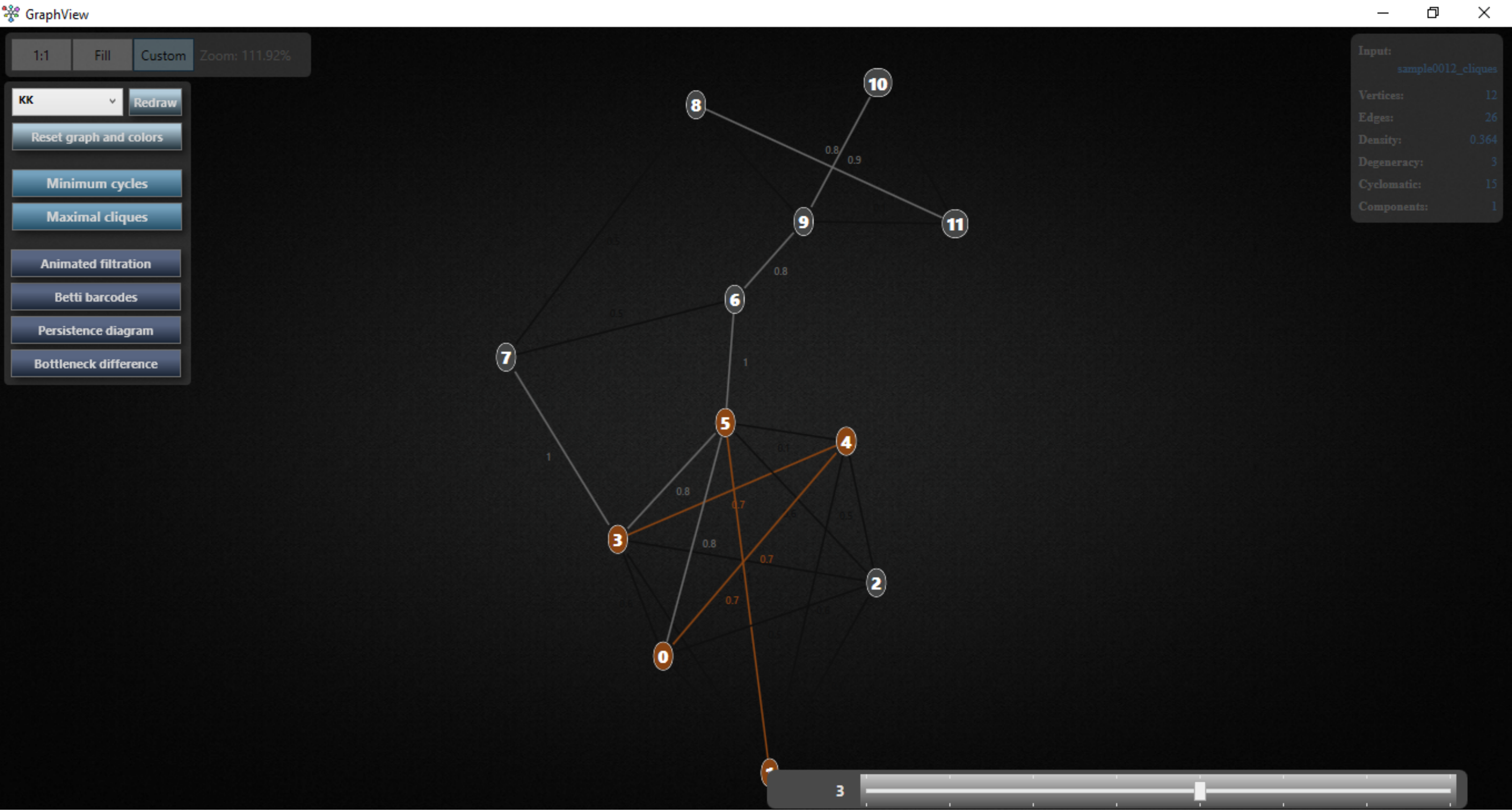
Clique Weighted Rank Filtration



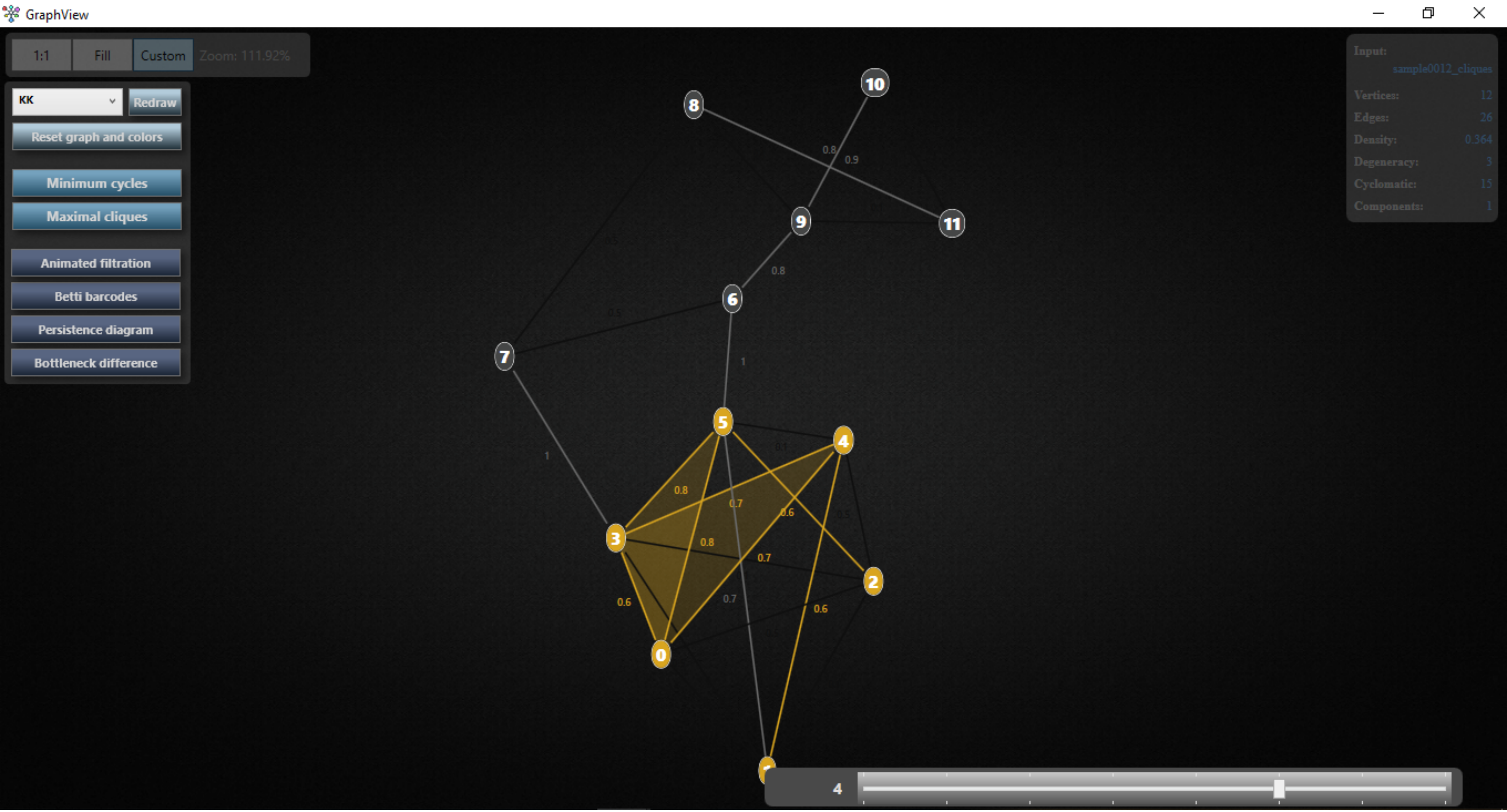
Clique Weighted Rank Filtration



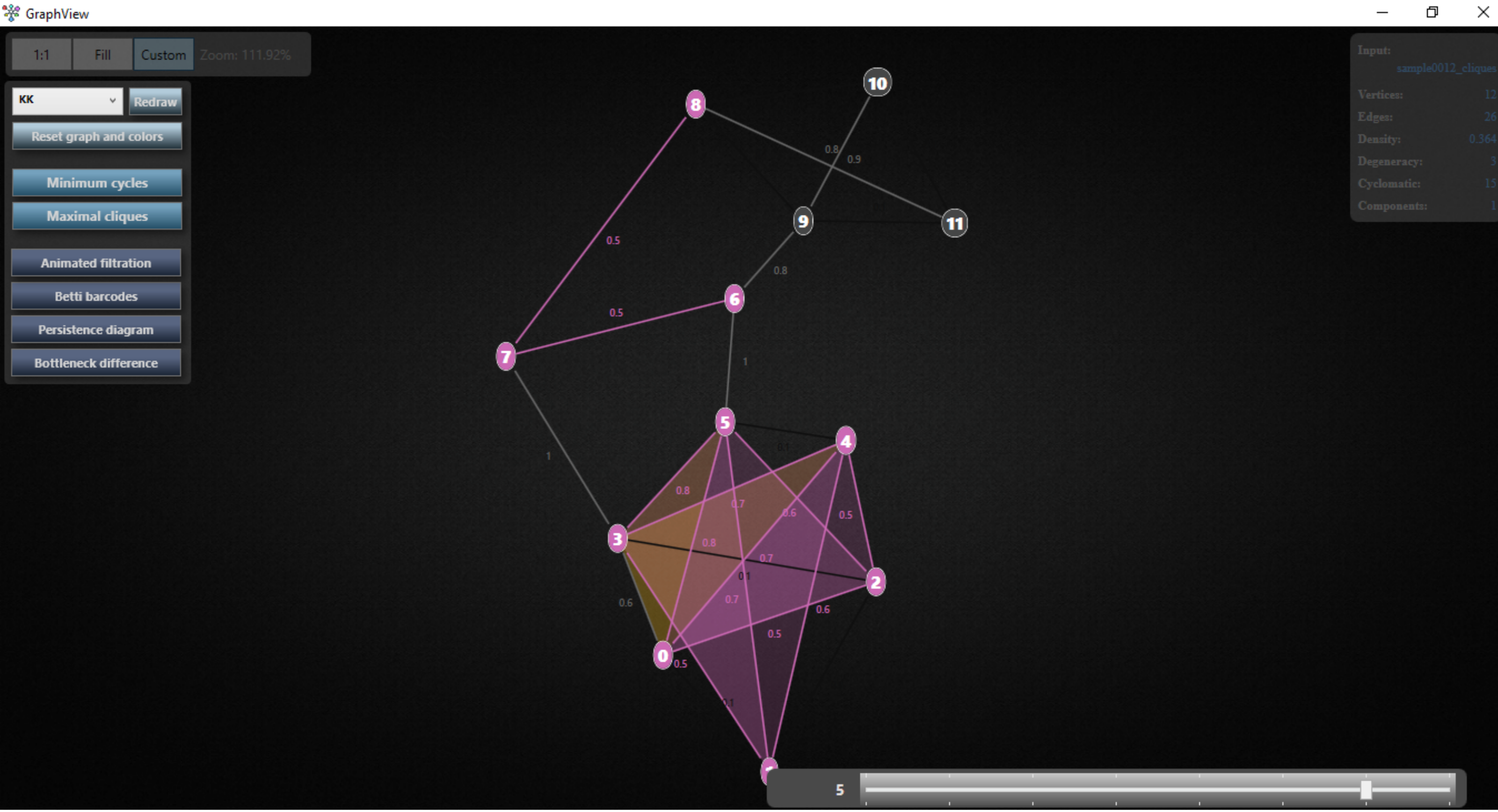
Clique Weighted Rank Filtration



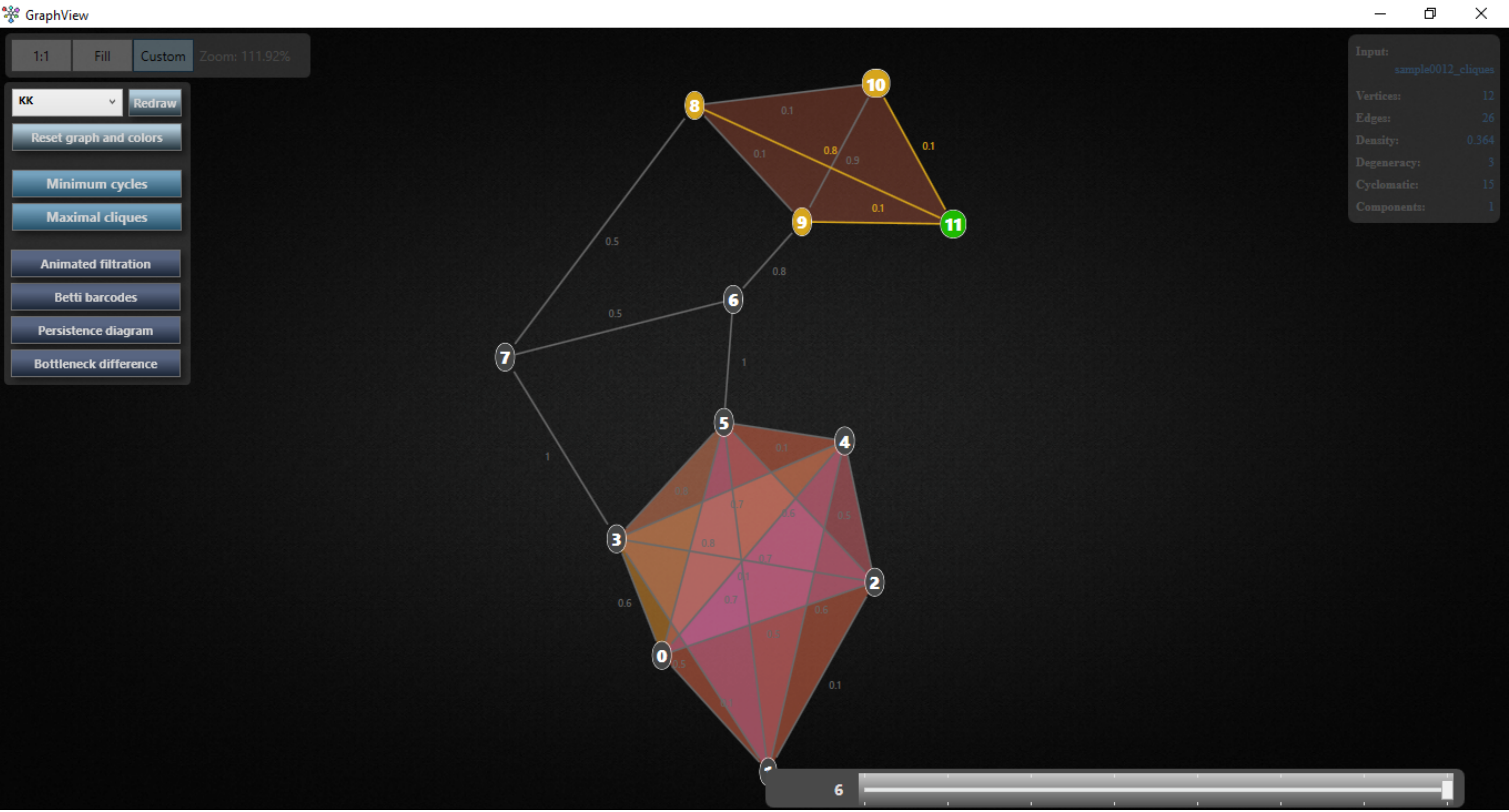
Clique Weighted Rank Filtration



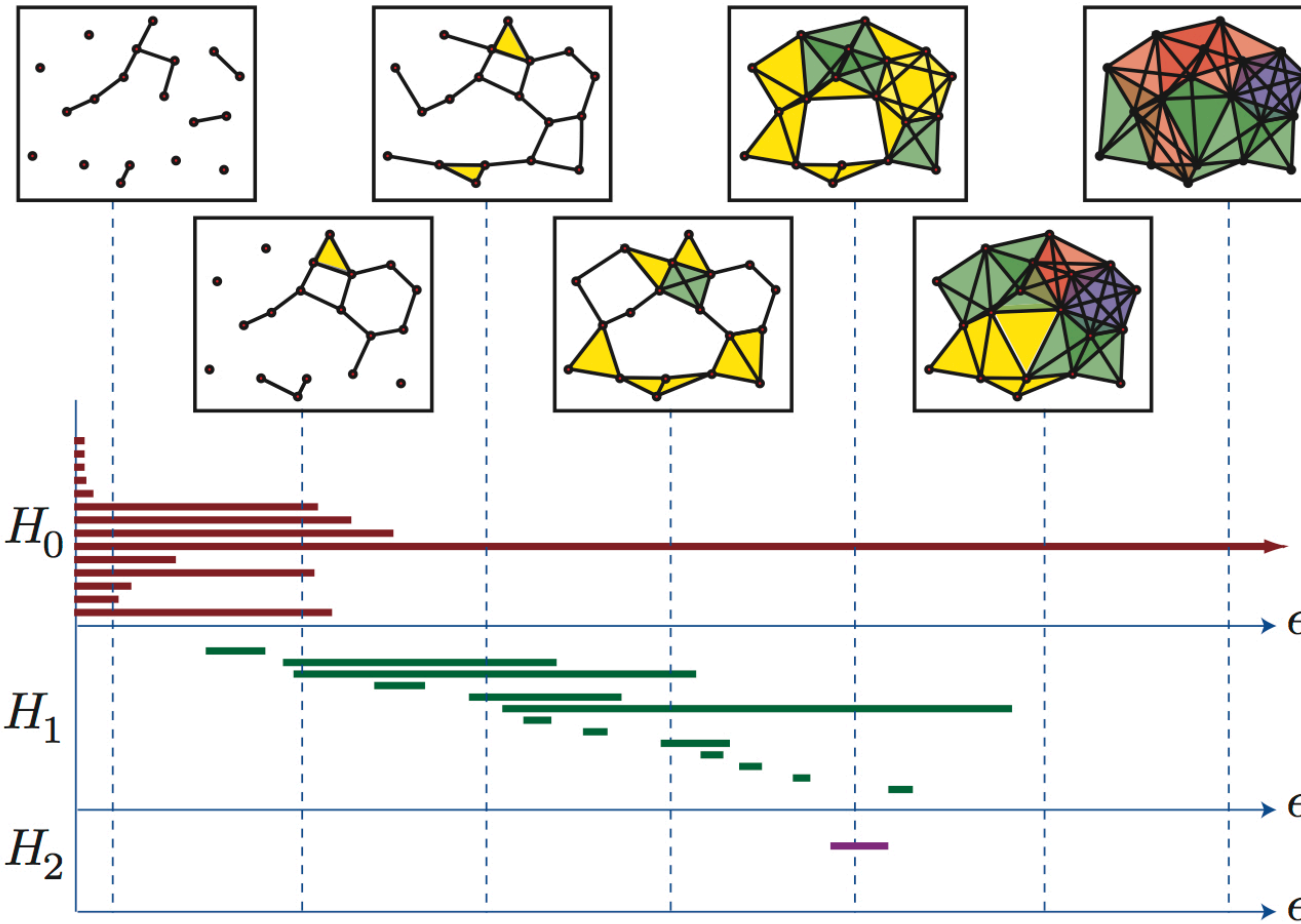
Clique Weighted Rank Filtration



Clique Weighted Rank Filtration



Barcodes & Diagrams



Ghrist, 2008,BARCODES: THE PERSISTENT TOPOLOGY OF DATA

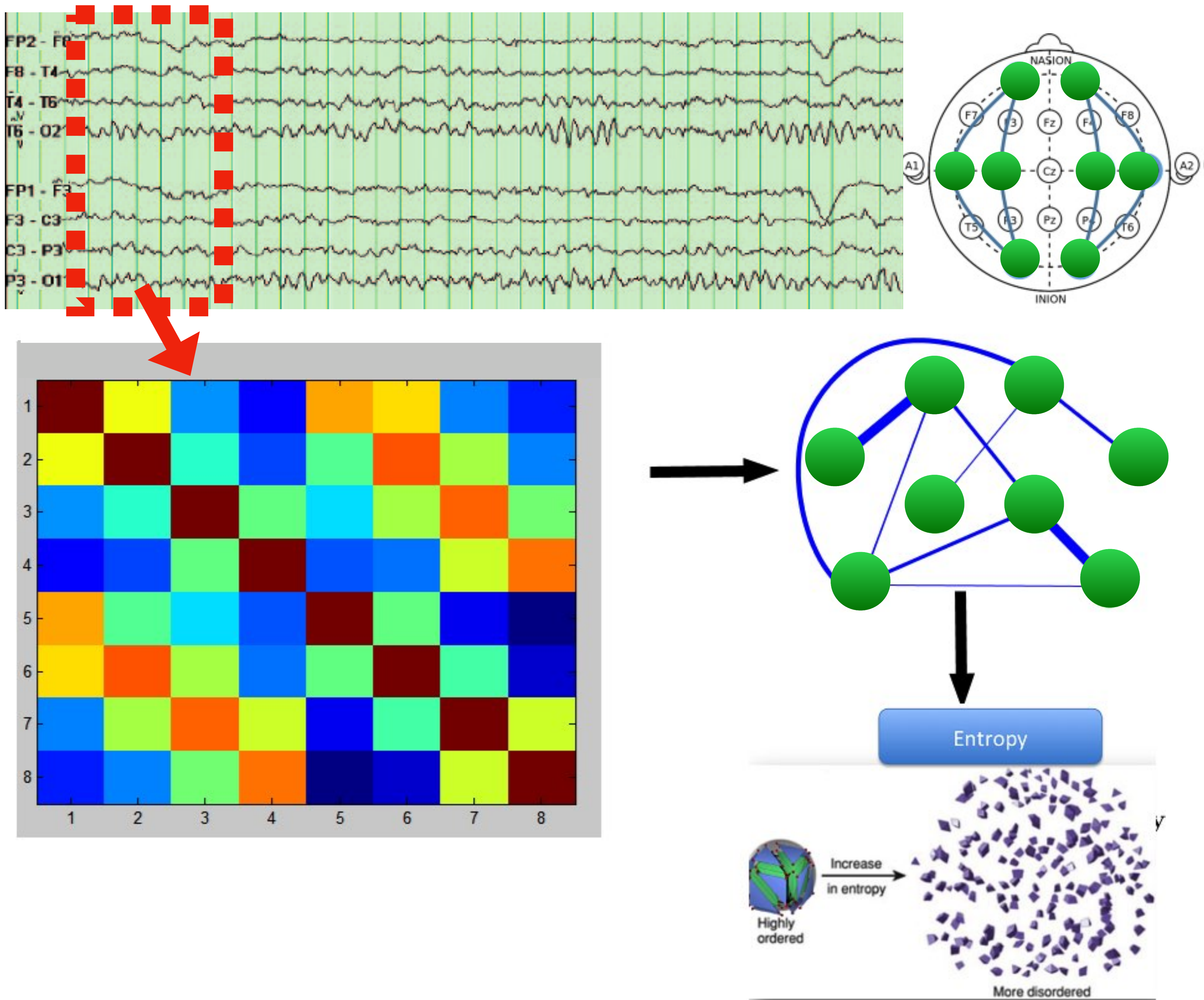
Persistent Entropy

$$PE_{H_k} = -\sum_{i}^{n=N_k} \frac{l_i}{L_{tot}} log \frac{l_i}{L_{tot}}$$

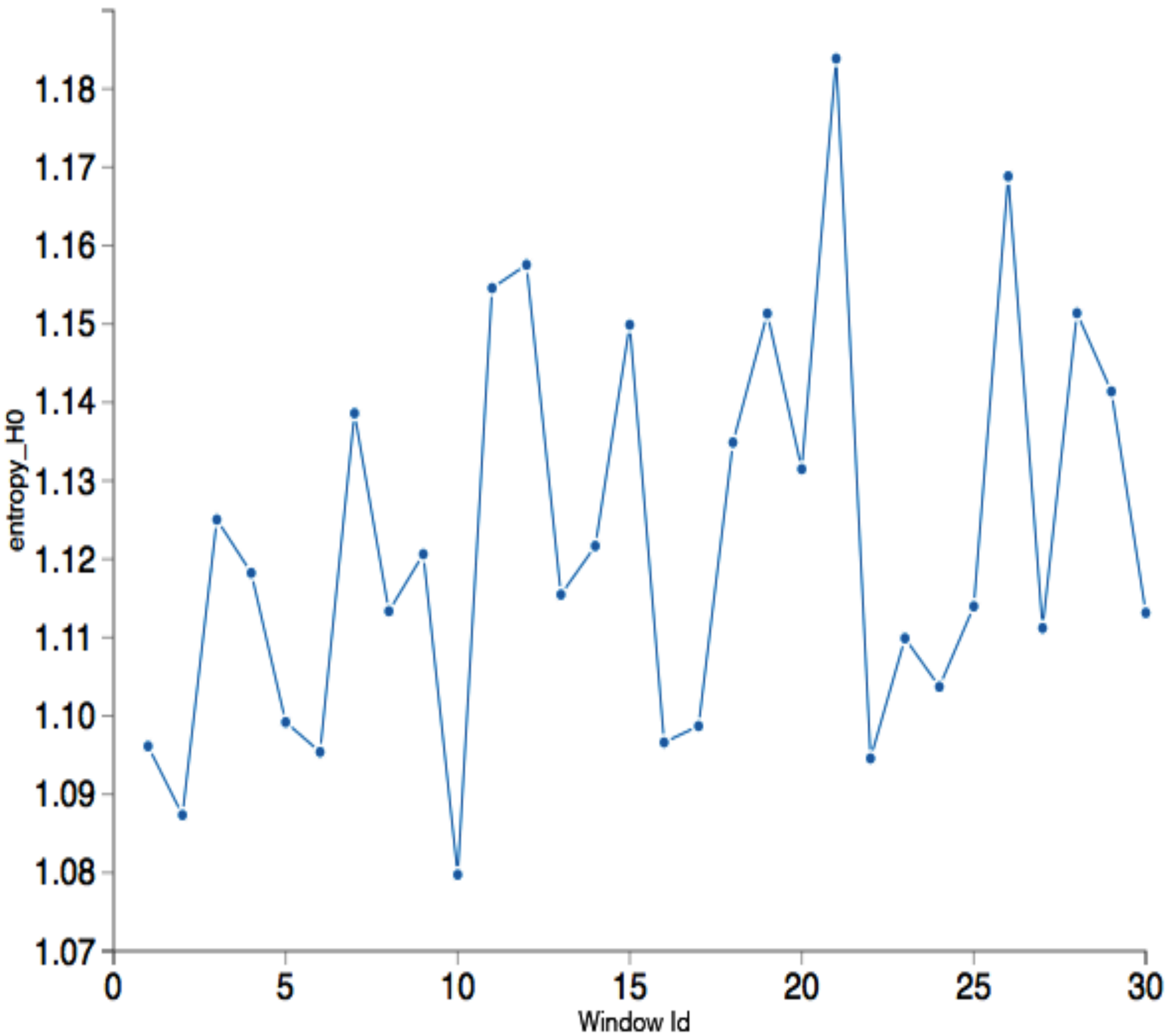
$$l_i = [death_i - birth_i]; L_{Tot} = \sum_i l_i$$

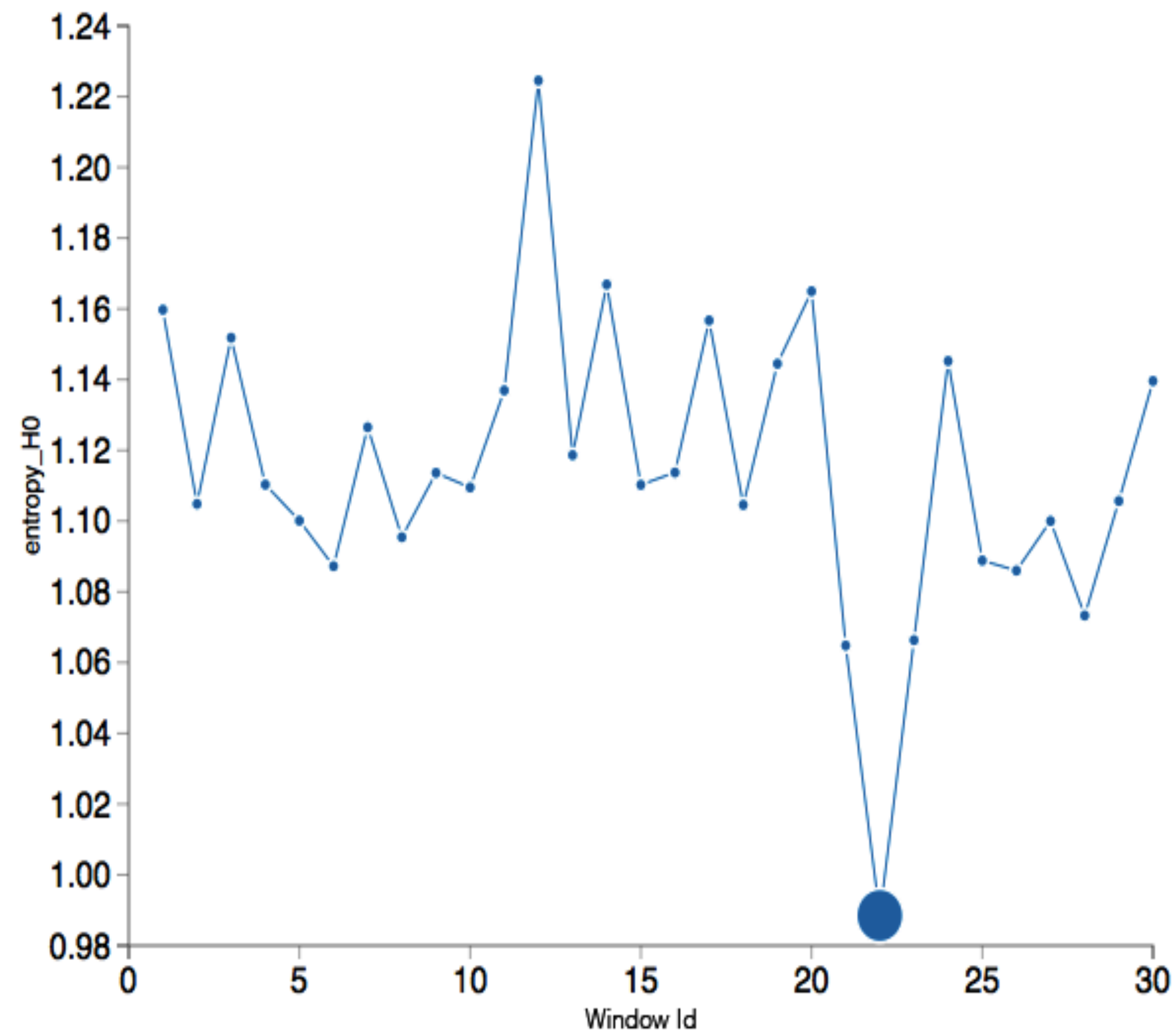
$$PE_{Tot} = \sum_k PE_{H_k}$$

Results I


FP2 - F8
F8 - T4
T4 - T6
T6 - O2
FP1 - F3
F3 - C3
C3 - P3
P3 - O1
NASION
INION
A1
A2
1
2
3
4
5
6
7
8
Entropy
Increase
in entropy
Highly
ordered
More disordered

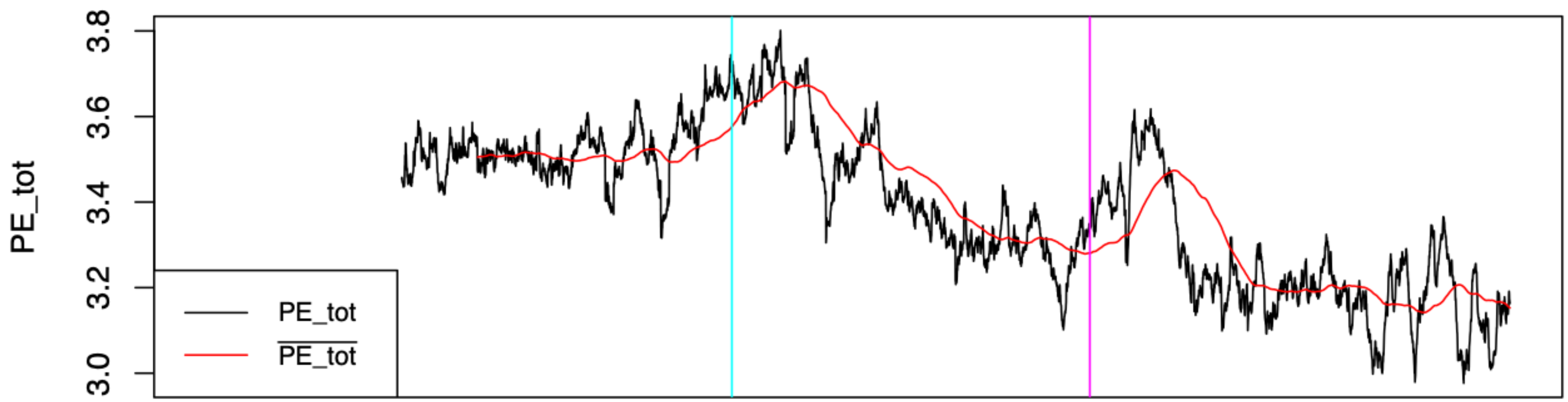
Results I



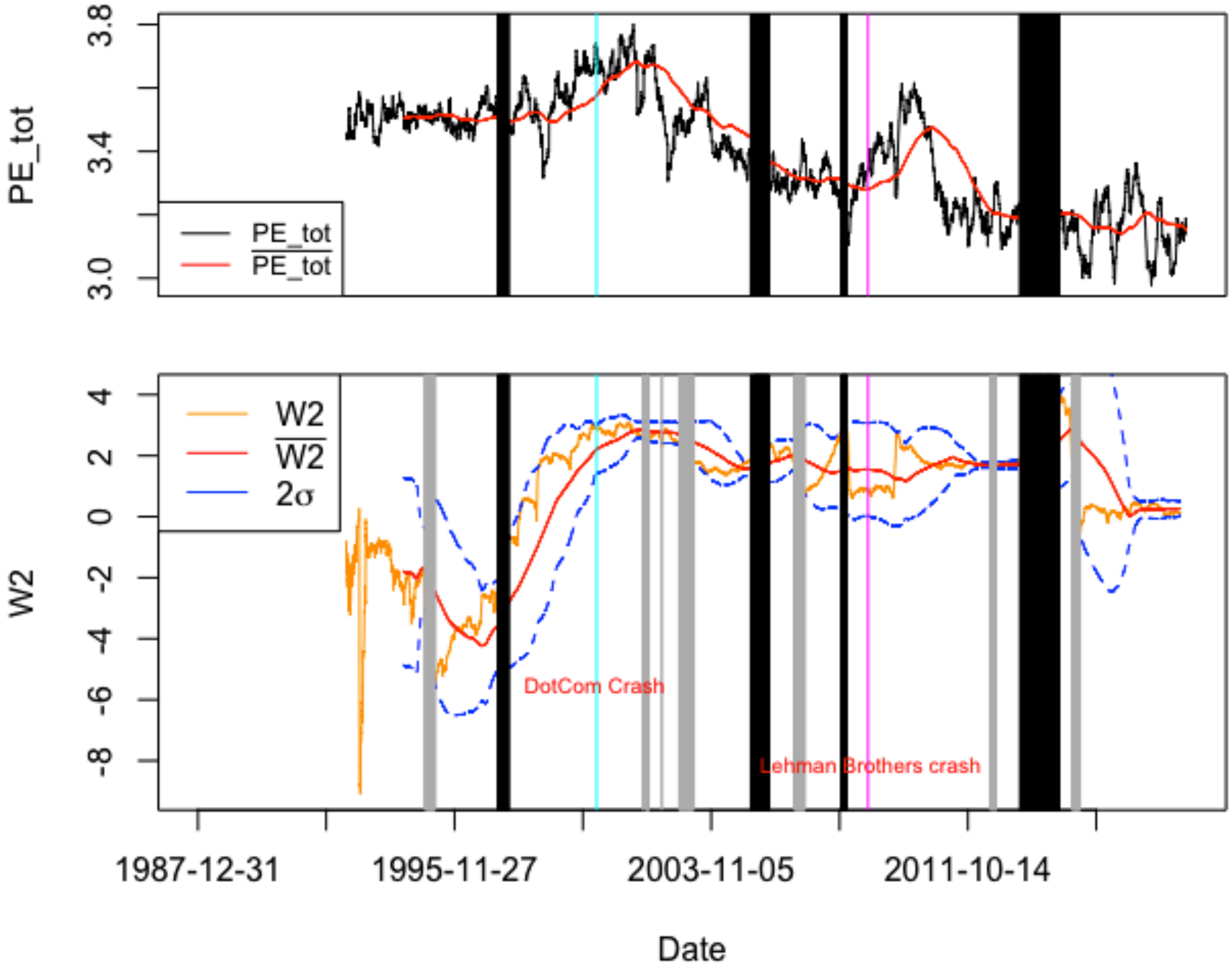


Merelli, Rucco, Piangerelli, & Toller, D. (2015). A topological approach for multivariate time series characterization: the epilepsy case study.

Results II



Results II



Piangerelli, Tesei, Merelli. (2019). A Persistent Entropy Automaton for the Dow Jones Stock Market. FSEN 2019

Take home message

- TDA is a new paradigm for data analysis
- TDA allows to go behind the graph representation
- TDA is versatile but computationally expensive
- TDA sliding window-based, naturally, tracks, the evolution in time of the global behavior (Persistent Entropy)

Thank you!