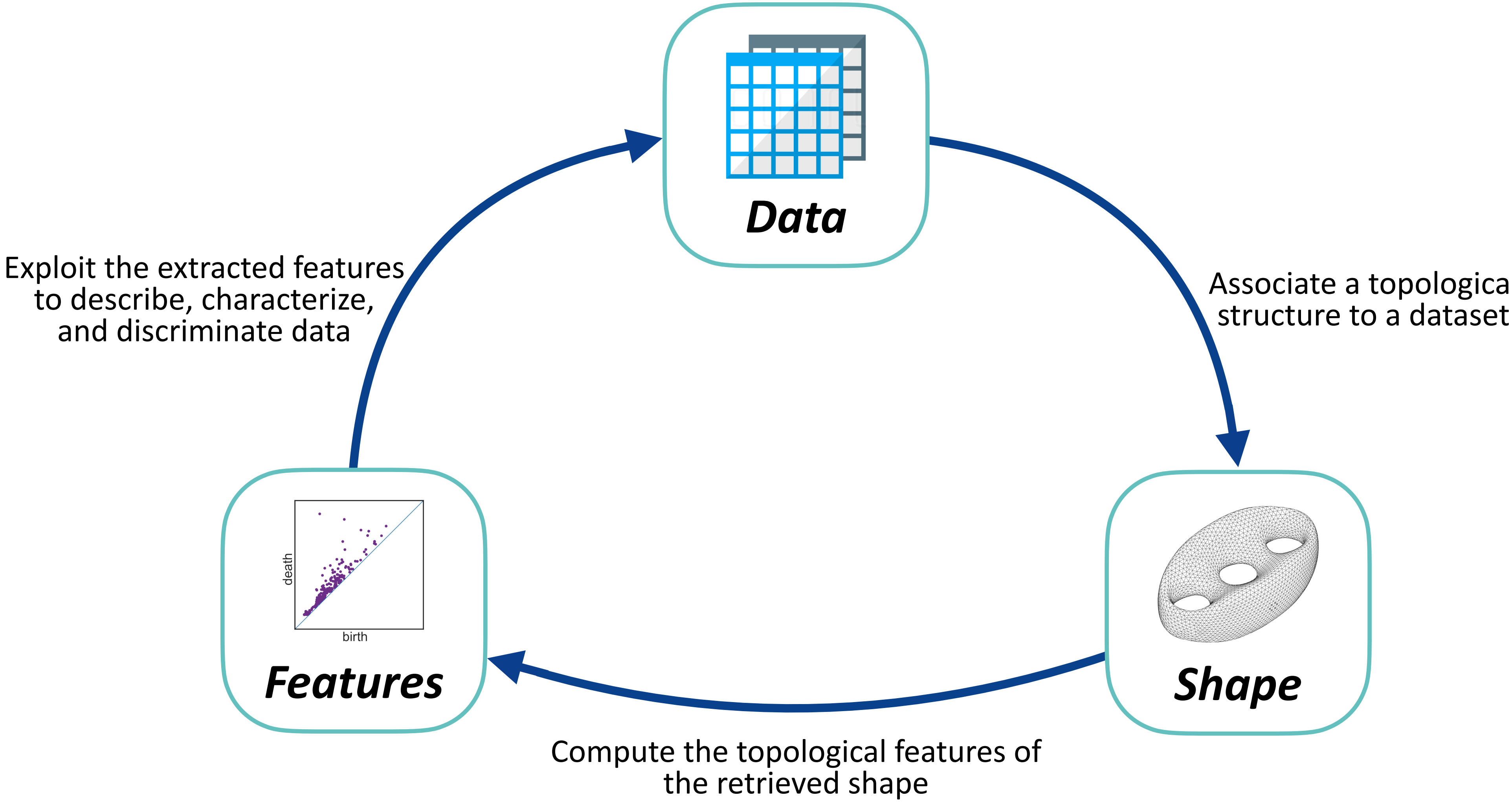
Topological Data Analysis

Kernels & Networks

Ulderico Fugacci

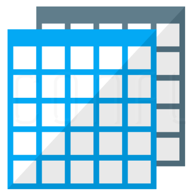
CNR - IMATI

Topological Data Analysis

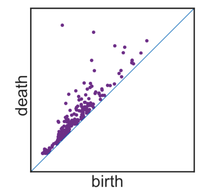


Kernels & Networks

Topological Data Analysis allows for assigning to (almost) ***any dataset*** a collection of features representing a ***topological summary*** of the input data



Data



Features

Goal:

Today, we address two main questions:

- ***Is this information immediately suitable for statistics and machine learning?***
- ***Do homology cycles provide a meaningful and descriptive information?***

Kernels & Networks

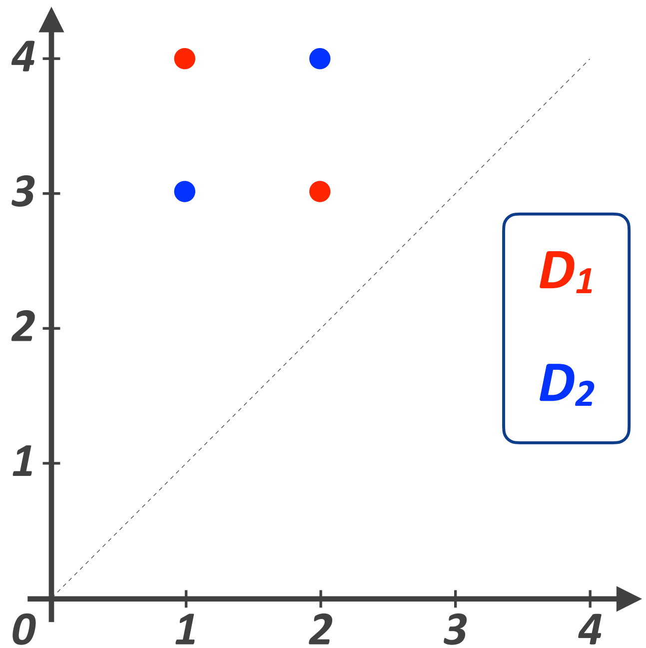
- *Kernels for Persistent Homology*
- *Persistence and Complex Networks*

Kernels & Networks

- ***Kernels*** ***for Persistent Homology***
- *Persistence and Complex Networks*

Kernels for Persistent Homology

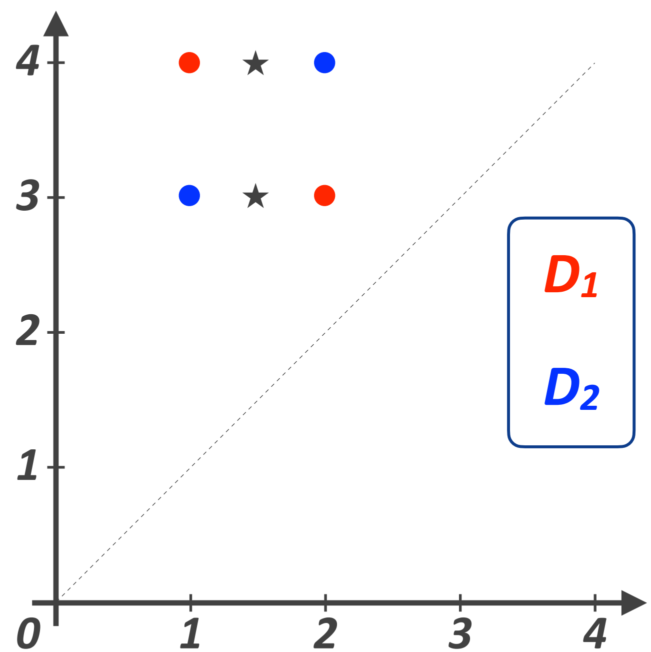
A Naive Example:



Mean *of persistence diagrams is* ***not unique***

Kernels for Persistent Homology

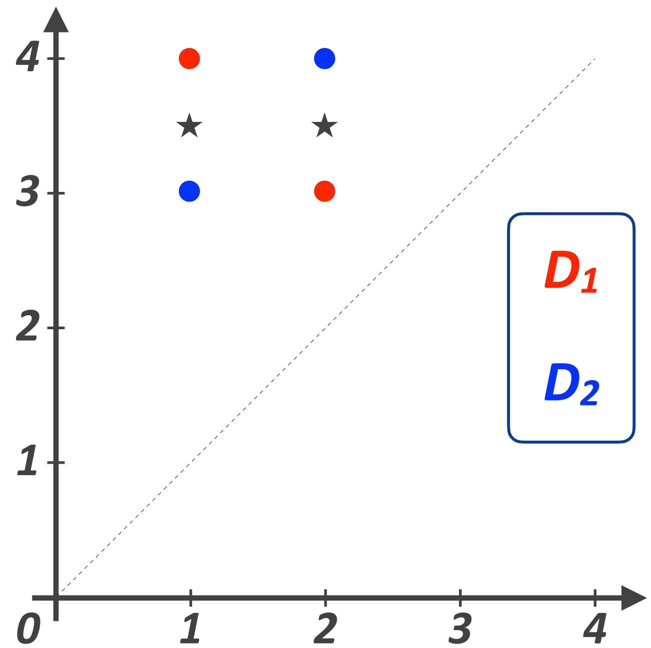
A Naive Example:



Mean *of persistence diagrams is* ***not unique***

Kernels for Persistent Homology

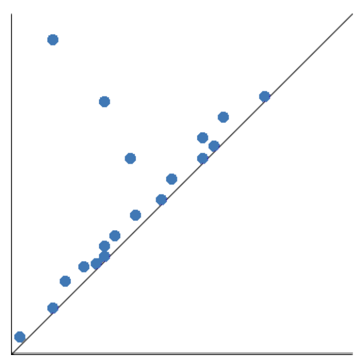
A Naive Example:

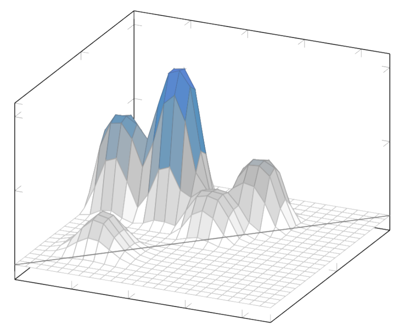


Mean** of persistence diagrams is **not unique

Kernels for Persistent Homology

Adopted Strategy:





Represent persistence diagrams as elements of a ***Hilbert space***

Kernels for Persistent Homology

Definitions:

*A **Hilbert space H** is*

*a **real or complex vector space** endowed with an **inner product***

$\langle\, .\, ,\, .\, \rangle : H \times H \longrightarrow \mathbb{R}$ such that, with respect to the distance induced by $\langle\, .\, ,\, .\, \rangle$,

*H is a **complete metric space***

*Recall that, a metric space H is called **complete** if*

every Cauchy sequence in H converges in H

Kernels for Persistent Homology

Example:

The space ***L^2*** of ***square-integrable functions*** on $\mathbb{R}^2$ is a Hilbert space

- $$\|f\|_{L^2} := \left(\int_{\mathbb{R}^2} |f|^2 d\mu\right)^{\frac{1}{2}} < +\infty$$
- $$\langle f, g\rangle_{L^2} := \int_{\mathbb{R}^2} f \cdot g \; d\mu$$

Kernels for Persistent Homology

Kernel Trick:

$$
\begin{tikzcd}
\Omega \times \Omega \arrow[rr, "k"] \arrow[d, "\phi \times \phi"'] & & \mathbb{R} \\
H \times H \arrow[urr, "\langle \cdot , \cdot \rangle"'] & &
\end{tikzcd}
$$

Definition:

*A **kernel k** for an input space Ω is a map $k\colon \Omega \times \Omega \longrightarrow \mathbb{R}$ such that there exist a **Hilbert space H** and a **feature map** $\phi\colon \Omega \longrightarrow H$ for which*

$$k(X, Y) = \langle \phi(X) , \phi(Y) \rangle$$

Kernels for Persistent Homology

Pseudo-Distance:

A kernel k: Ω × Ω ⟶ ℝ implicitly induces on Ω

a ***pseudo-distance*** ***d_k: Ω × Ω*** ⟶ ℝ defined, for each X, Y ∈ Ω, as

$$d_k(X,Y) := \|\phi(X) - \phi(Y)\|_H = \Big(k(X,X) + k(Y,Y) - 2k(X,Y)\Big)^{1/2}$$

Stability:

A kernel k: Ω × Ω ⟶ ℝ is ***stable*** *w.r.t a distance d in Ω*

if there is a constant C > 0 such that, for all X, Y ∈ Ω,

$$d_k(X,Y) \leq C \cdot d(X,Y)$$

Kernels for Persistent Homology

Our Goal:

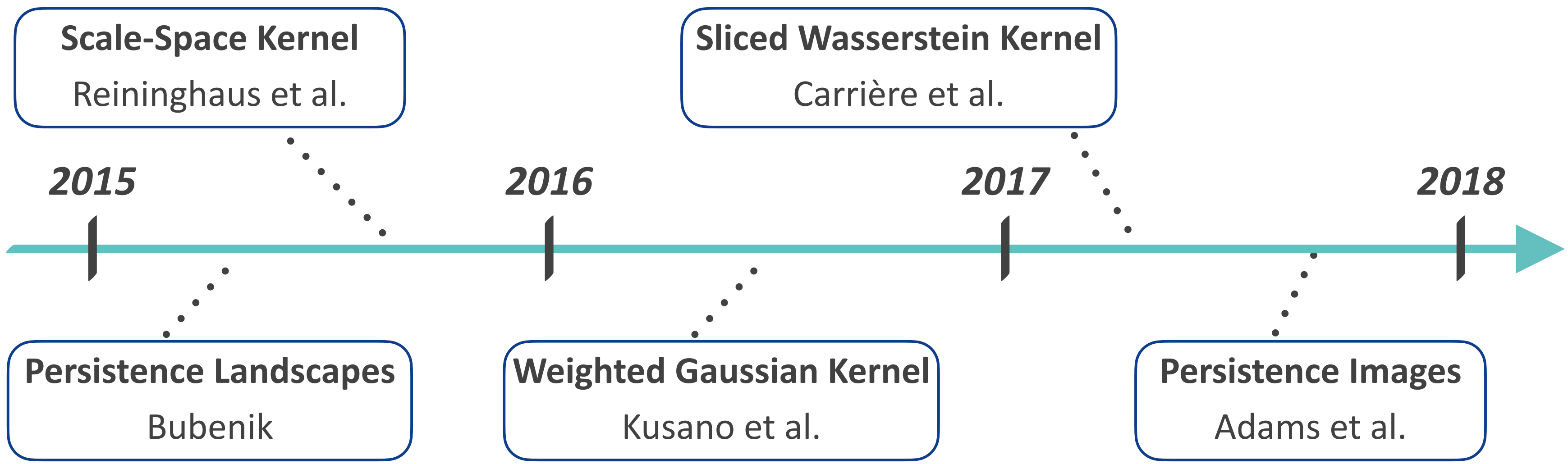
Defining a ***kernel*** for the set Ω of finite ***persistence diagrams***:

- ***Stable***
- ***Easy to be computed***
- Possible endowed with an ***explicit feature map*** ϕ***: Ω*** $\longrightarrow$ ***H***

The idea of a kernel for persistence diagrams has

- *Originally* ***born in the '90s*** *(see [Donatini et al. 1998; Ferri et al. 1998])*
- *Spread in the literature and* ***widely adopted in applications*** ***just recently***

Kernels for Persistent Homology

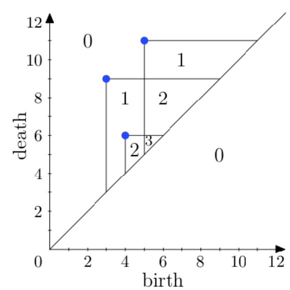


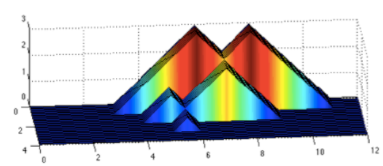
Kernels for Persistent Homology

2015 **2016** **2017** **2018**

Persistence Landscapes

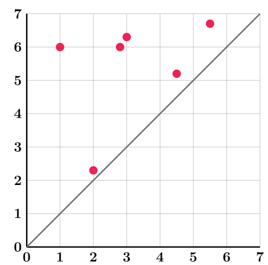
Bubenik

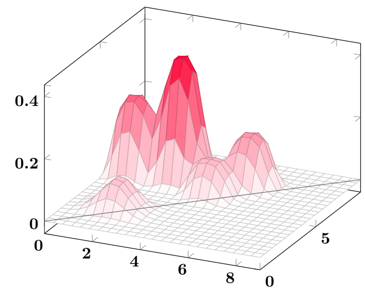




Kernels for Persistent Homology

Scale-Space Kernel
Reininghaus et al.





2015 ***2016*** ***2017*** ***2018***

Persistence Landscapes
Bubenik

Kernels for Persistent Homology

Scale-Space Kernel

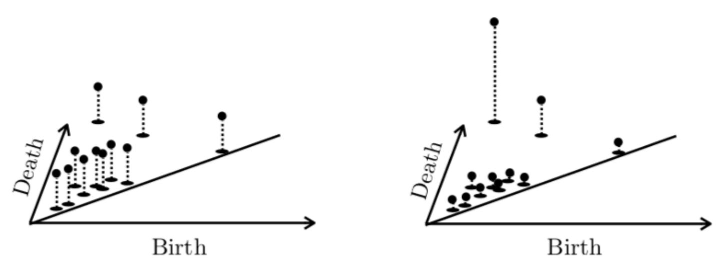
Reininghaus et al.

2015 *2016* *2017* *2018*

Persisten

Weighted Gaussian Kernel

Kusano et al.



Kernels for Persistent Homology

$$k_{SW}(D,E) := \exp\left(-\frac{d_{SW}(D,E)}{2\sigma^2}\right)$$

Sliced Wasserstein Kernel

Carrière et al.

Scale

Reini

2015

2016

2017

2018

Persistence Landscapes

Bubenik

Weighted Gaussian Kernel

Kusano et al.

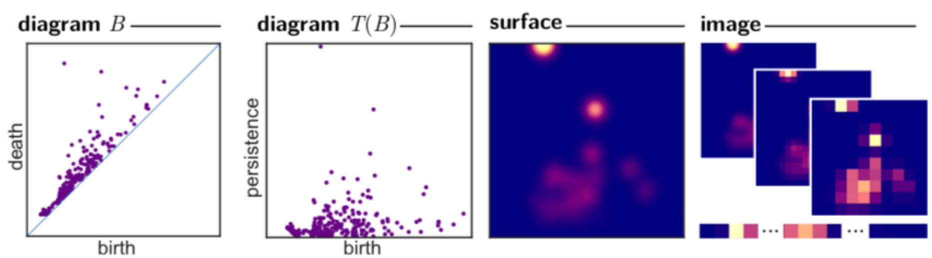
Kernels for Persistent Homology

Scale-Space Kernel
Reininghaus et al.

Sliced Wasserstein Kernel
Carrière et al.

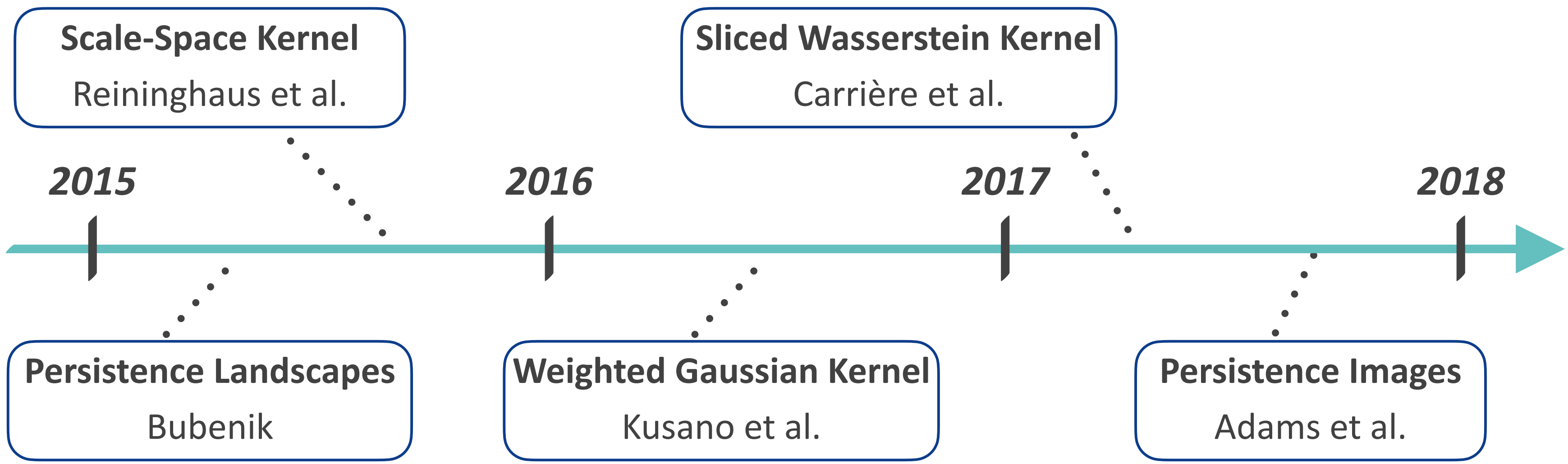
2015 2016 2017 2018

Persi



Persistence Images
Adams et al.

Kernels for Persistent Homology



Kernels for Persistent Homology

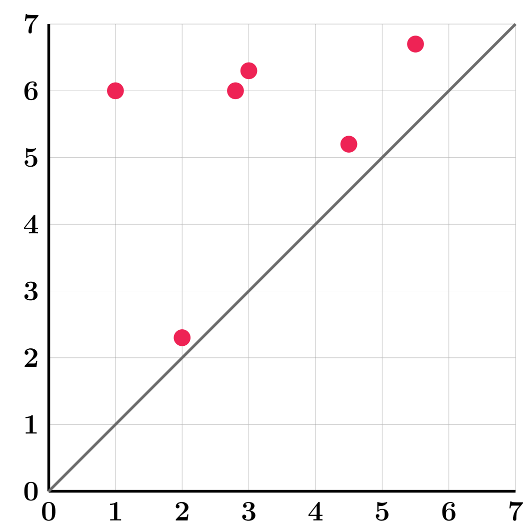
Scale-Space Kernel:

Given a finite persistence diagram D, we consider the solution

$\phi : \Delta^{+} \times \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}$ of the following ***heat diffusion problem***:

- having a Dirichlet ***boundary condition*** on the diagonal
- setting as an ***initial condition*** a sum of Dirac deltas

$$\begin{cases} \Delta_p \phi = \partial_\sigma \phi & \text{in } \Delta^{+} \times \mathbb{R}_{>0} \\ \phi = 0 & \text{on } \partial\Delta^{+} \times \mathbb{R}_{\geq 0} \\ \phi = \sum_{q \in D} \boldsymbol{\delta}_q & \text{on } \Delta^{+} \times \{0\} \end{cases}$$



Kernels for Persistent Homology

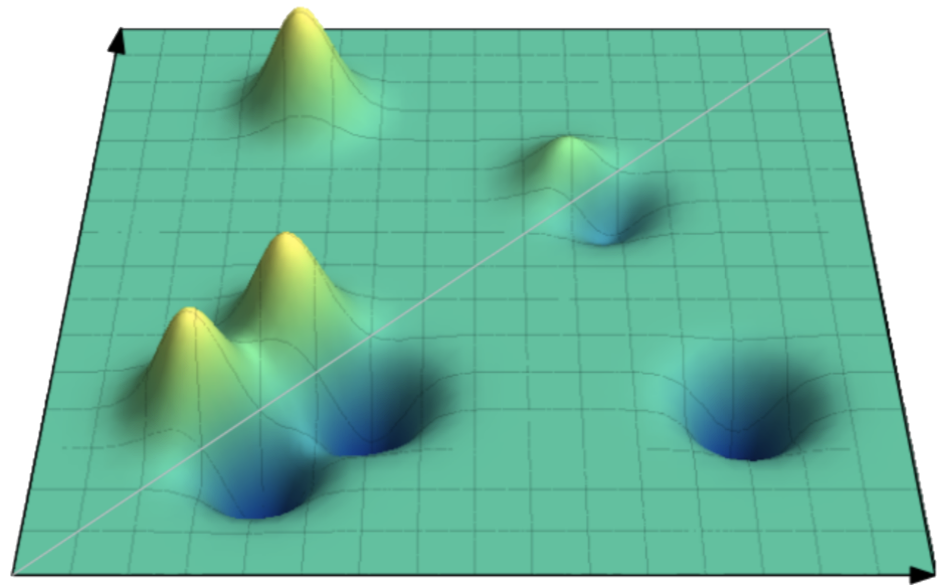
Scale-Space Kernel:

A *solution* is found by:

- ***extending*** Δ^+ to $\mathbb{R}^2$
- ***replacing*** the initial condition with

$$\phi = \Sigma_{q \in D} \boldsymbol{\delta}_q - \boldsymbol{\delta}_{q'} \qquad \text{on } \mathbb{R}^2 \times \{0\}$$

where, if $q=(a,b)$, then $q'=(b,a)$



Solution:

$$\phi_\sigma(p) = \frac{1}{4\pi\sigma} \sum_{q \in D} \Big(\exp\big(-\frac{\|p-q\|^2}{4\sigma}\big) - \exp\big(-\frac{\|p-q'\|^2}{4\sigma}\big)\Big)$$

Kernels for Persistent Homology

Scale-Space Kernel:

Stability Theorem:

Given two finite persistence diagrams D, E, we have that

$$\|\phi_\sigma(D) - \phi_\sigma(E)\|_{L^2} \leq \frac{1}{2\sqrt{\pi}\sigma} d_{W,1}(D,E)$$

where, for $r \geq 1$, the ***r-Wasserstein distance*** *is defined as*

$$d_{W,r}(D,E) := \left(\inf_\gamma \sum_{p \in D} \|p - \gamma(p)\|_\infty^r \right)^{1/r}$$

with γ running over all bijections from D to E

Kernels for Persistent Homology

Definitions:

A kernel k for the set Ω of finite persistence diagrams is called:

- ***additive*** if, for all D, E, F ∈ Ω, $k(D \cup E, F) = k(D, F) + k(E, F)$
- ***trivial*** if, for all D, E ∈ Ω, $k(D, E) = 0$

Theorem:

*Any **non-trivial additive** kernel k for the set Ω*

*is **not stable** with respect to $d_{W,r}$ for any $1 < r \leq \infty$*

(Notice that $d_{W,\infty} = d_B$)

Kernels for Persistent Homology

Sliced Wasserstein Kernel:

A standard way to construct a kernel is to exponentiate the negative of an Euclidean distance

$$k_\sigma(X,Y) := \exp\left(-\frac{\|X-Y\|^2}{2\sigma^2}\right)$$

Theorem:

$$k_\sigma(X,Y) := \exp\left(-\frac{f(X,Y)}{2\sigma^2}\right)$$

defines a ***valid kernel*** *for all $\sigma > 0$* ***if and only if*** *f is* ***conditionally negative definite*** *function*

I.e., for any $n \in \mathbb{N}$, for any $X_1, \dots, X_n \in \Omega$, and for any $a_1, \dots, a_n \in \mathbb{R}$ such that $\Sigma_i a_i = 0$,

one has $\Sigma_{i,j} a_i a_j f(X_i, X_j) \leq 0$

Kernels for Persistent Homology

Sliced Wasserstein Kernel:

Issue:

*None of the already introduced distances (and neither their squares) between persistence diagrams **is conditionally negative definite***

Solution:

*The **Sliced Wasserstein distance** $\mathbf{d}_{sw}$ is specifically designed to be **conditionally negative definite***

*Based on it, one can define the **Slice Wasserstein kernel** $\mathbf{k}_{sw}$ as*

$$k_{SW}(D,E) := \exp\left(-\frac{d_{SW}(D,E)}{2\sigma^2}\right)$$

Kernels & Networks

- *Kernels for Persistent Homology*
- ***Persistence and Complex Networks***

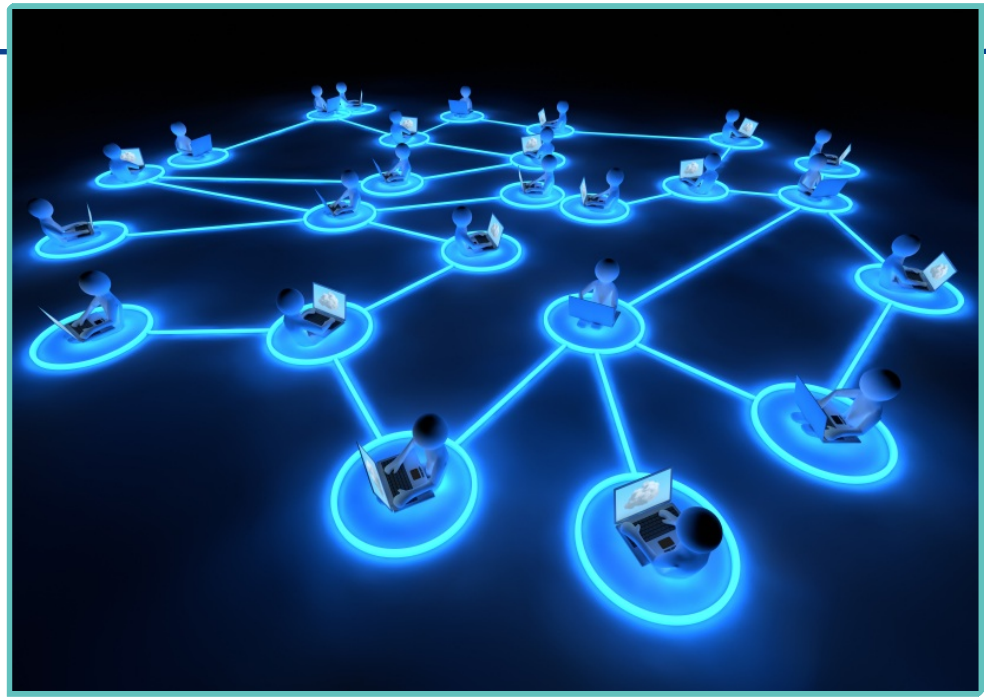
Persistence and Complex Networks

Networks:

*A **network** is a complex system consisting of **individuals or entities connected by specific ties** such as friendship, common interest, and shared knowledge*

E.g.

- ***Social*** Networks
- ***Sensor*** Networks
- ***Biological*** Networks
- ***Collaborative*** Networks
- …



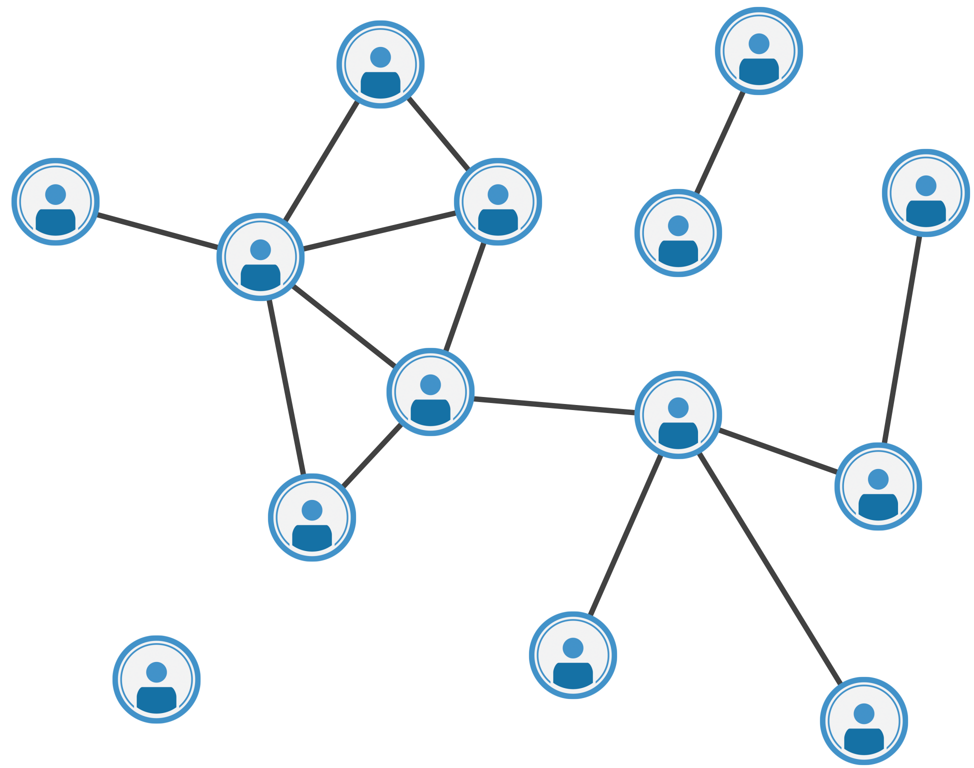
Persistence and Complex Networks

Representation:

A network can be represented by a ***graph*** ***G = (V, E)*** such that:

Nodes of V ⟺ *Entities or individuals*

Arcs of E ⟺ *Ties between entities*



Persistence and Complex Networks

Representation:

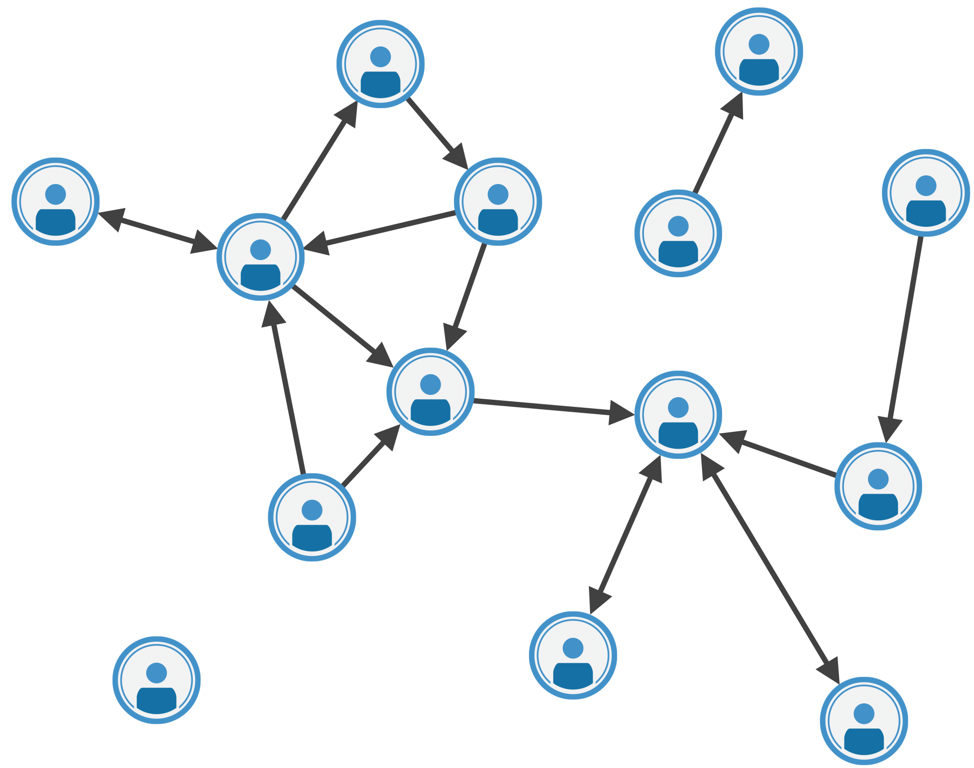
A network can be represented by a *graph* ***G = (V, E)*** such that:

Nodes of V ⟺ *Entities or individuals*

Arcs of E ⟺ *Ties between entities*

Arcs can be:

- ***Directed***
- ***Weighted***



Persistence and Complex Networks

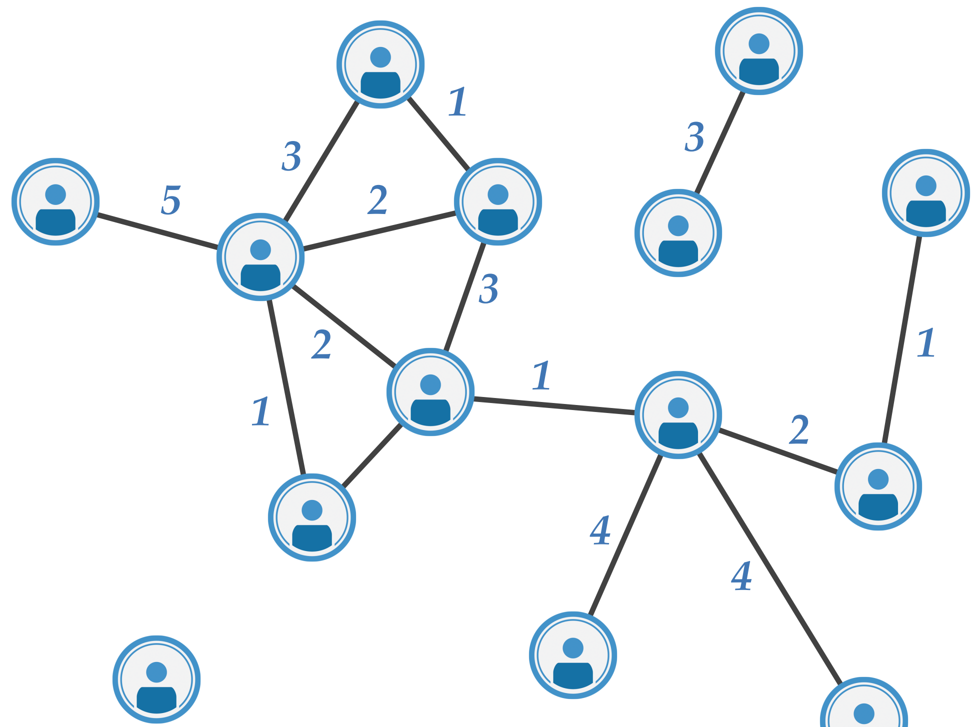
Representation:

A network can be represented by a ***graph*** ***G = (V, E)*** such that:

Nodes of V	⟺	*Entities or individuals*
Arcs of E		*Ties between entities*

Arcs can be:

- ***Directed***
- ***Weighted***



Persistence and Complex Networks

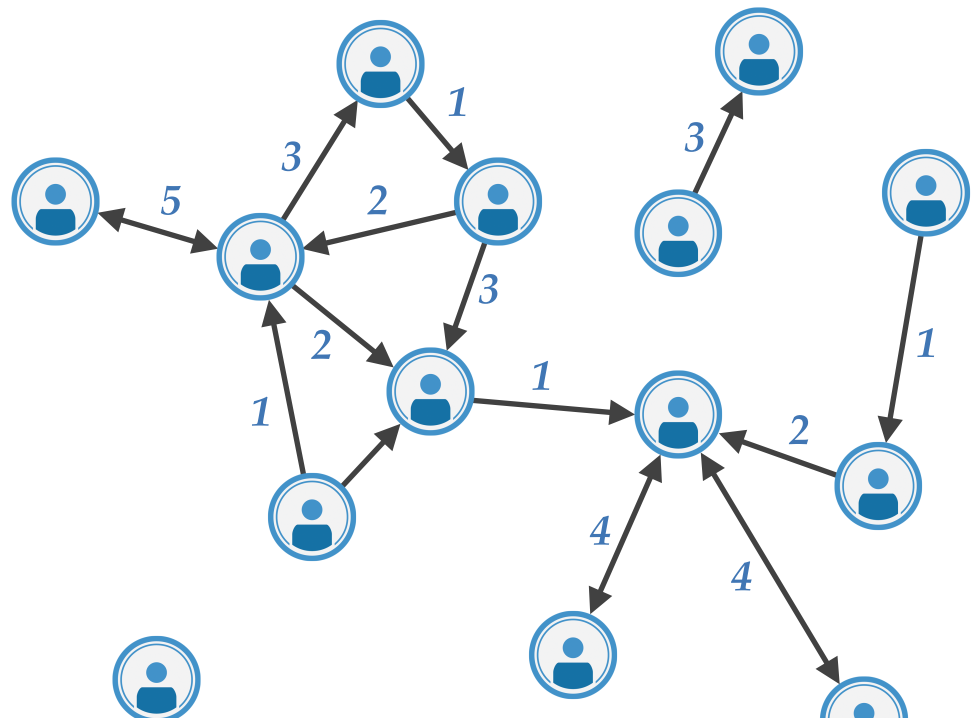
Representation:

A network can be represented by a *graph* ***G = (V, E)*** such that:

Nodes of V	⟺	*Entities or individuals*
Arcs of E		*Ties between entities*

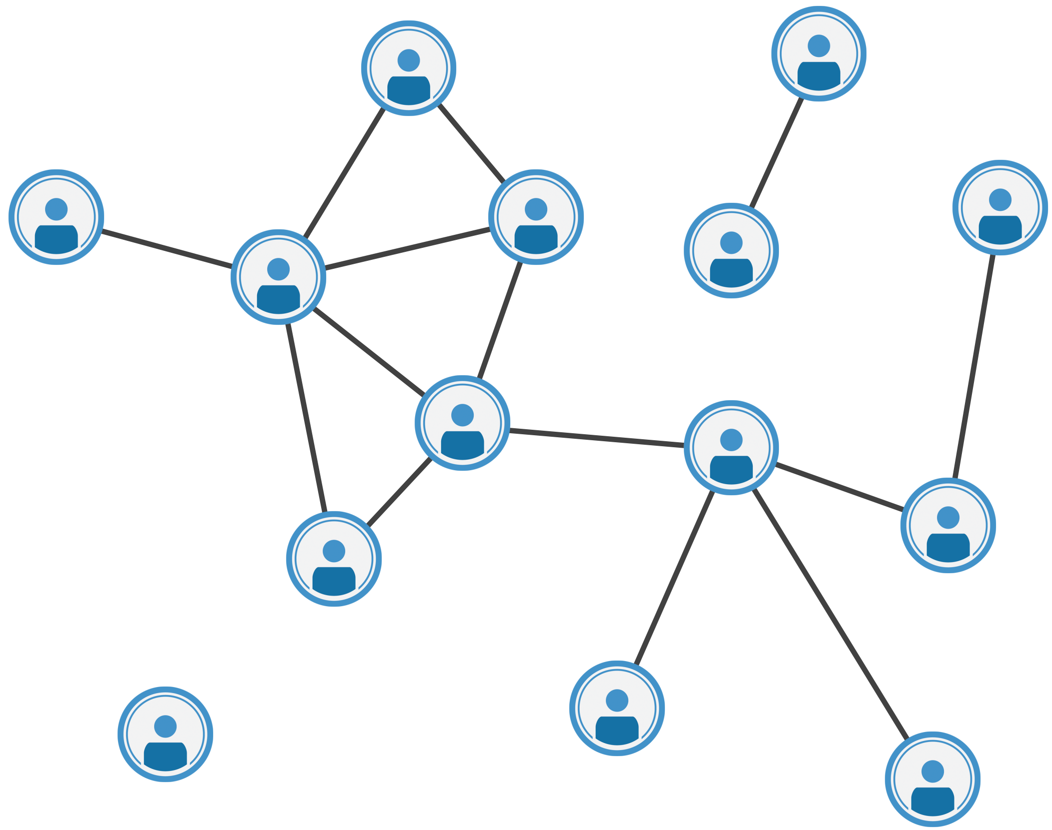
Arcs can be:

- ***Directed***
- ***Weighted***



Persistence and Complex Networks

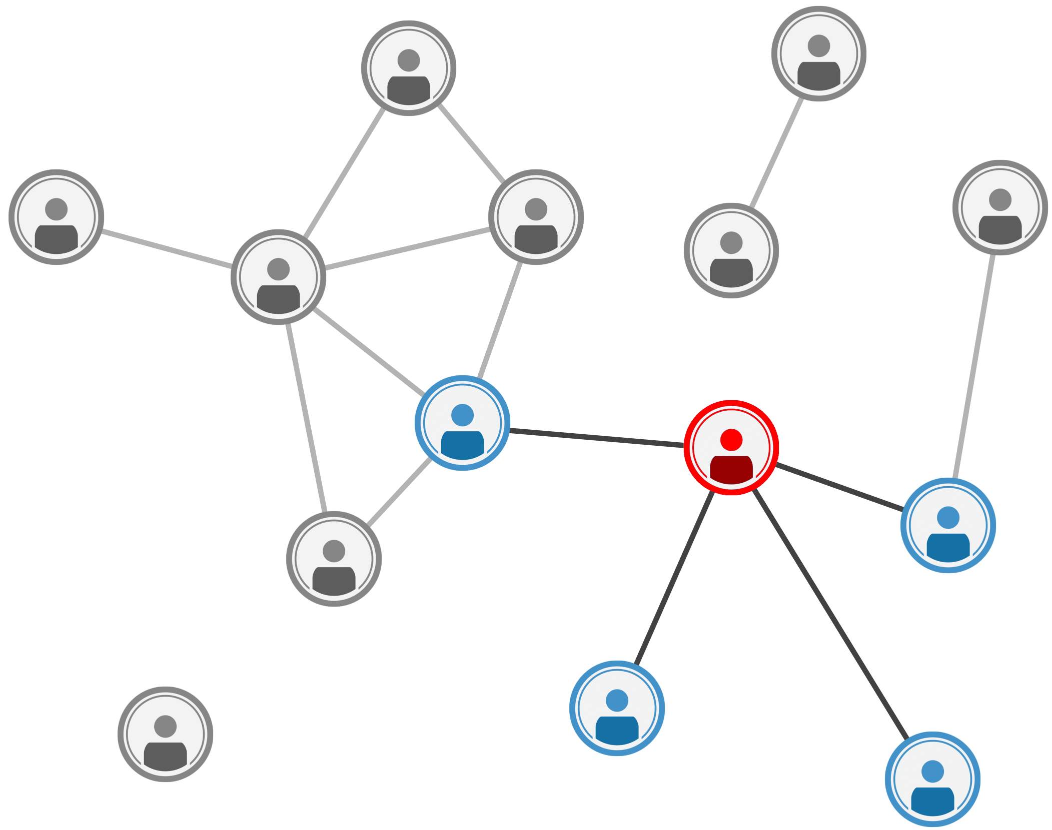
A Two-Level Analysis:



Persistence and Complex Networks

A Two-Level Analysis:

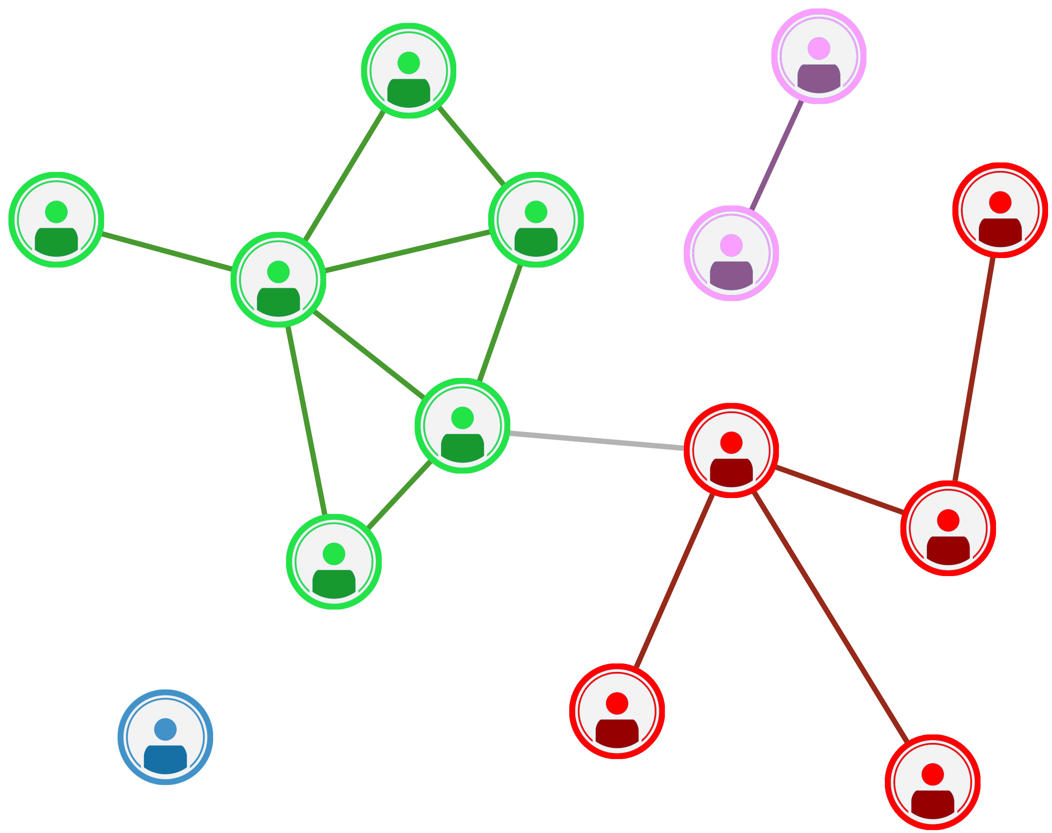
- ***Egocentric***
- *Sociocentric*



Persistence and Complex Networks

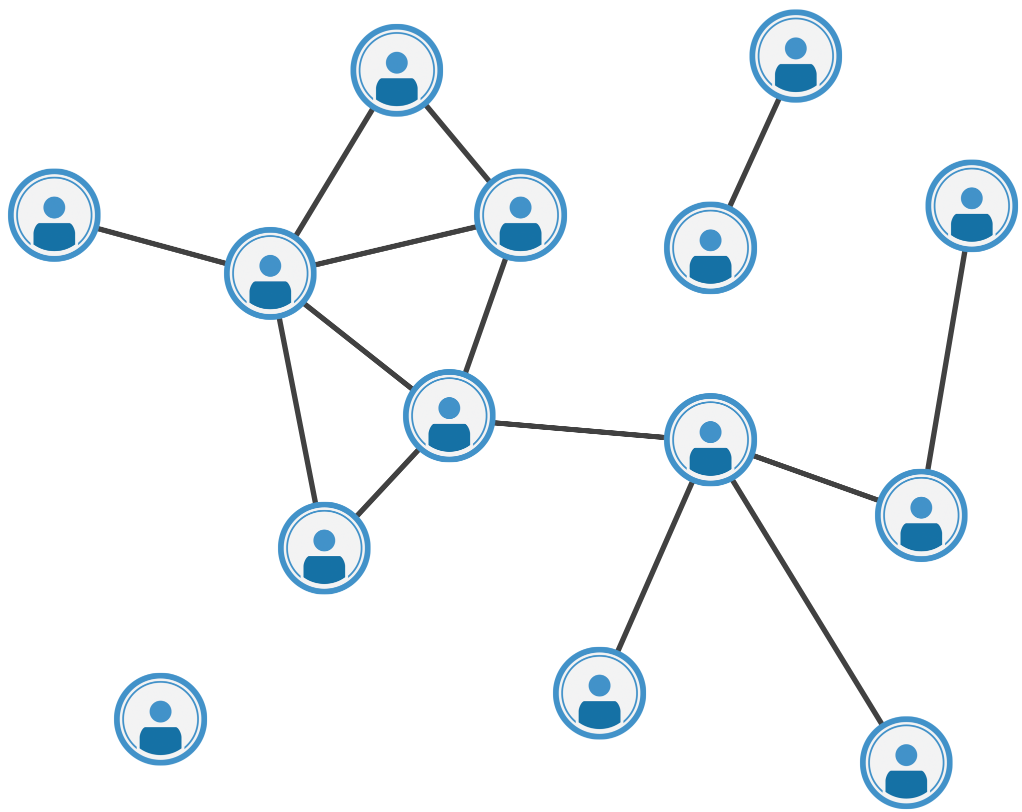
A Two-Level Analysis:

- *Egocentric*
- ***Sociocentric***



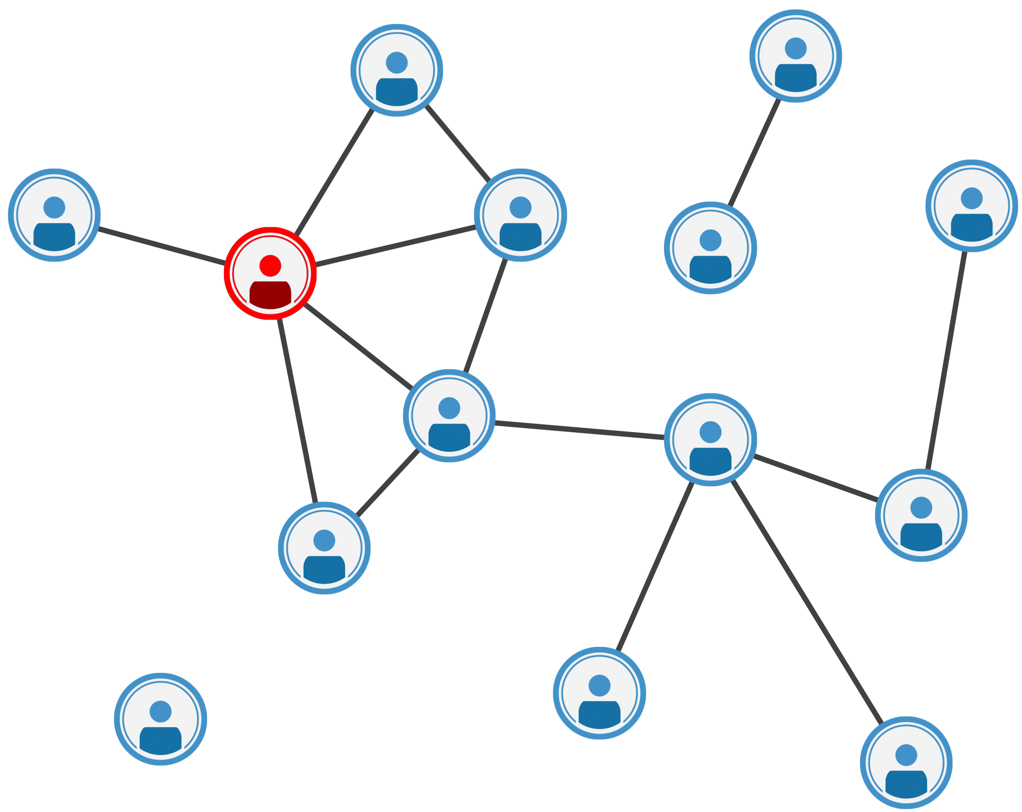
Persistence and Complex Networks

Who is the most important individual?



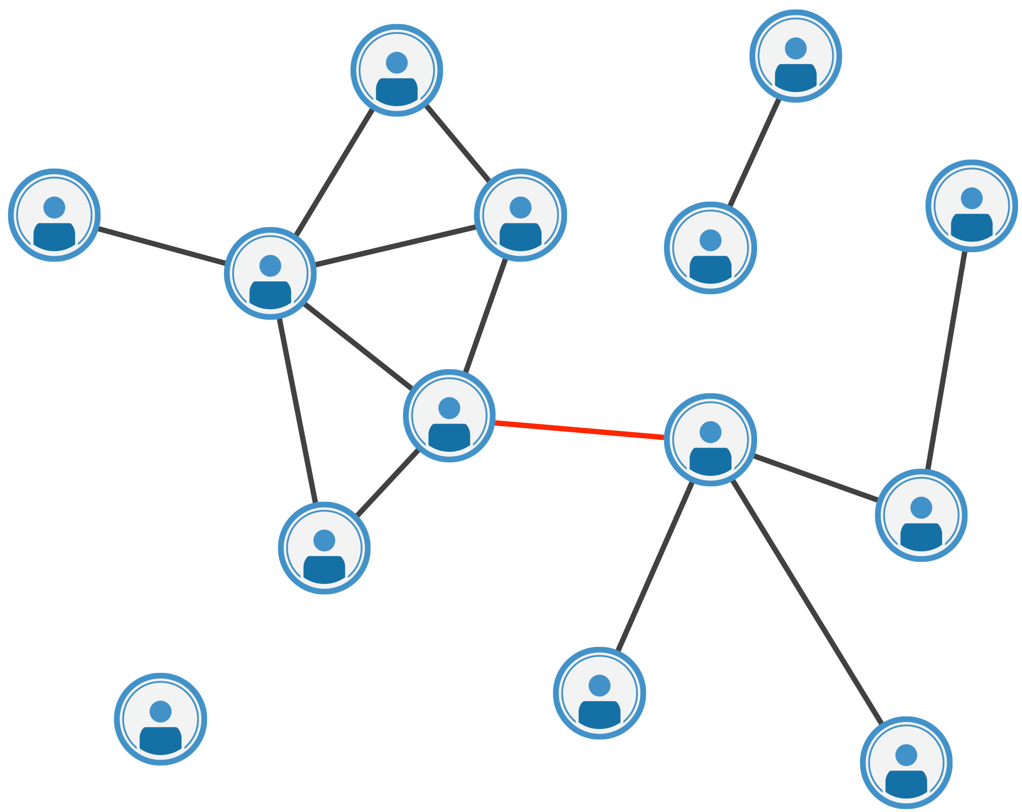
Persistence and Complex Networks

Who is the most important individual?



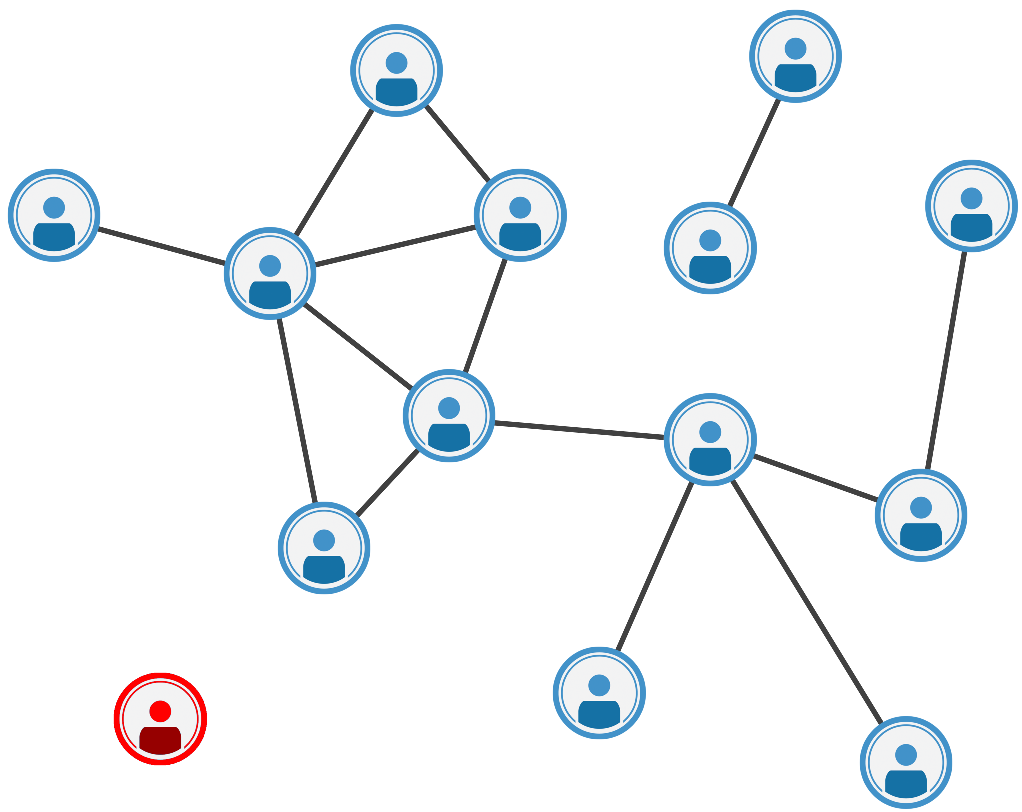
Persistence and Complex Networks

Who is the most important individual?



Persistence and Complex Networks

Who is the most important individual?



Persistence and Complex Networks

Centrality Measures:

Different criteria to underline ***different roles***:

- *Key players*
- *Brokers*
- *Bridges*
- *Isolated*
- *...*

Definition:

*A **centrality measure** is a function $F : V \longrightarrow \mathbb{R}$ assigning to each node a **"centrality"** value:*

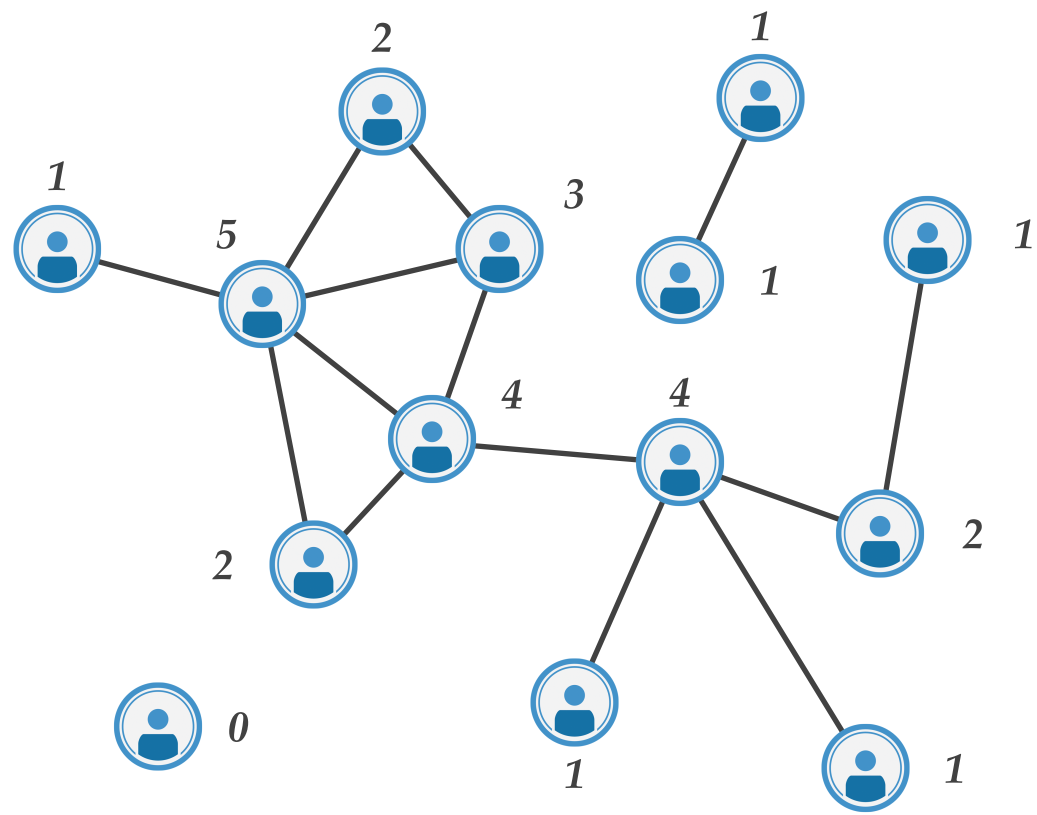
- ***Degree** centrality*
- ***Betweenness** centrality*
- ***Closeness** centrality*
- ***Eigenvector** centrality*
- ***Erdös** distance*

Persistence and Complex Networks

Degree Centrality:

Given a node v of G = (V, E),

$$D(v) := \#\{u \in V \,|\, (u,v) \in E\}$$

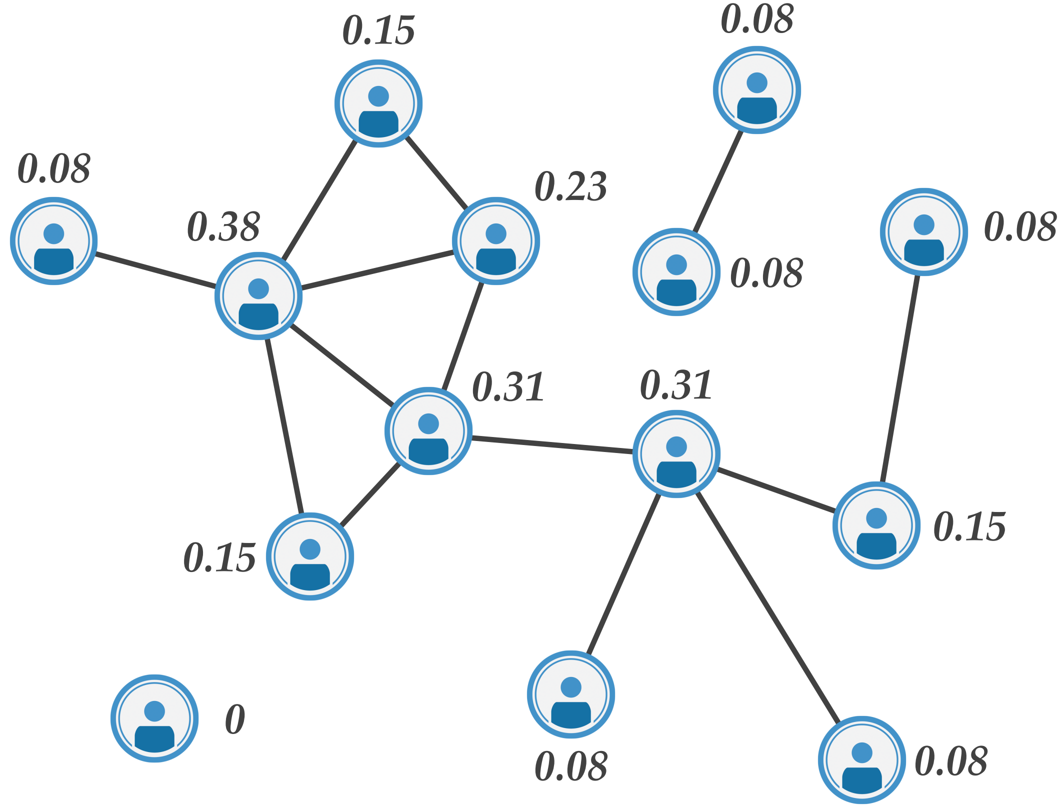


Persistence and Complex Networks

Degree Centrality:

Given a node v of G = (V, E),

$$D(v) := \frac{\#\{u \in V \mid (u,v) \in E\}}{\#V - 1}$$



Persistence and Complex Networks

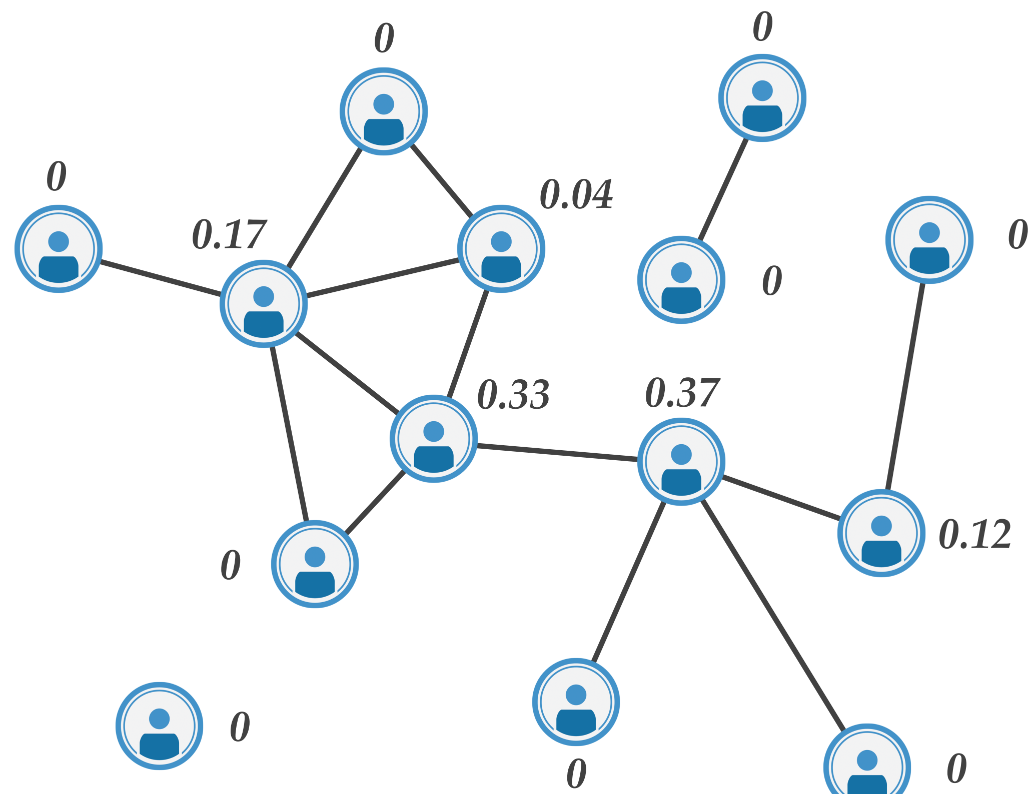
Betweenness Centrality:

Given a node v of G = (V, E),

$$B(v) := \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st} is the number of ***shortest paths from s to t***
- $\sigma_{st}(v)$ is the number of those paths ***passing through v***

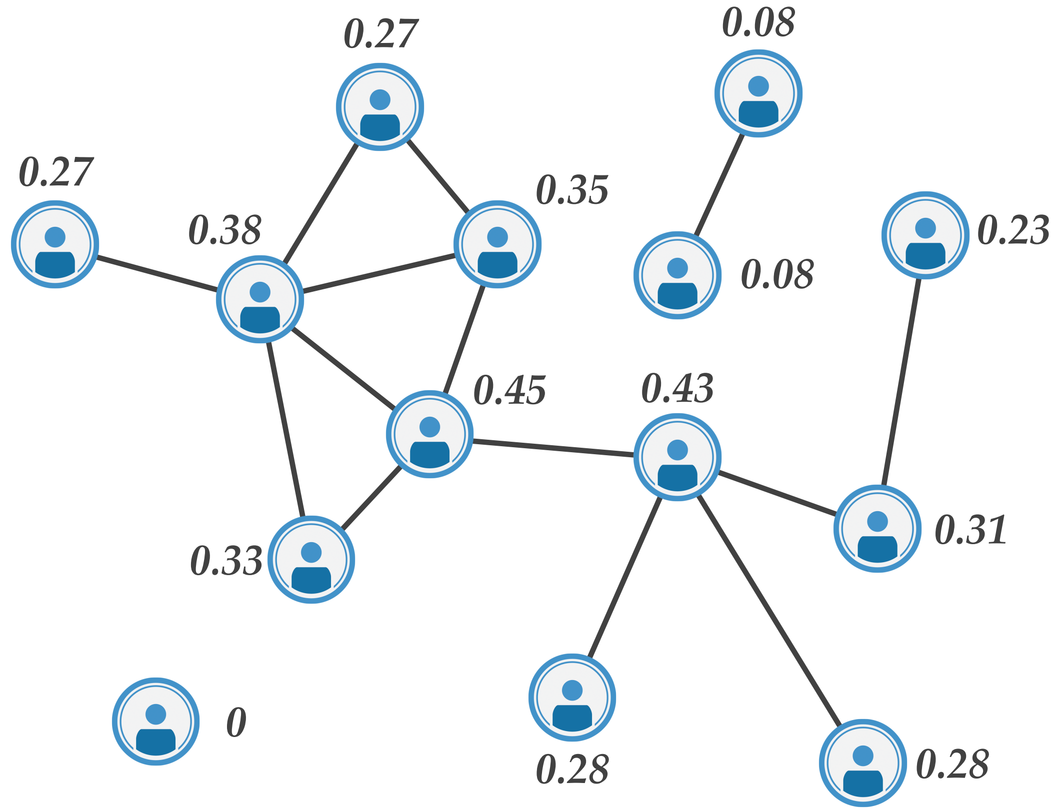


Persistence and Complex Networks

Closeness Centrality:

Given a node v of G = (V, E),

$$C(v) := \frac{\#V - 1}{\sum_{u \in V} d(u, v)}$$



Persistence and Complex Networks

Eigenvector Centrality:

Given a node v of G = (V, E),

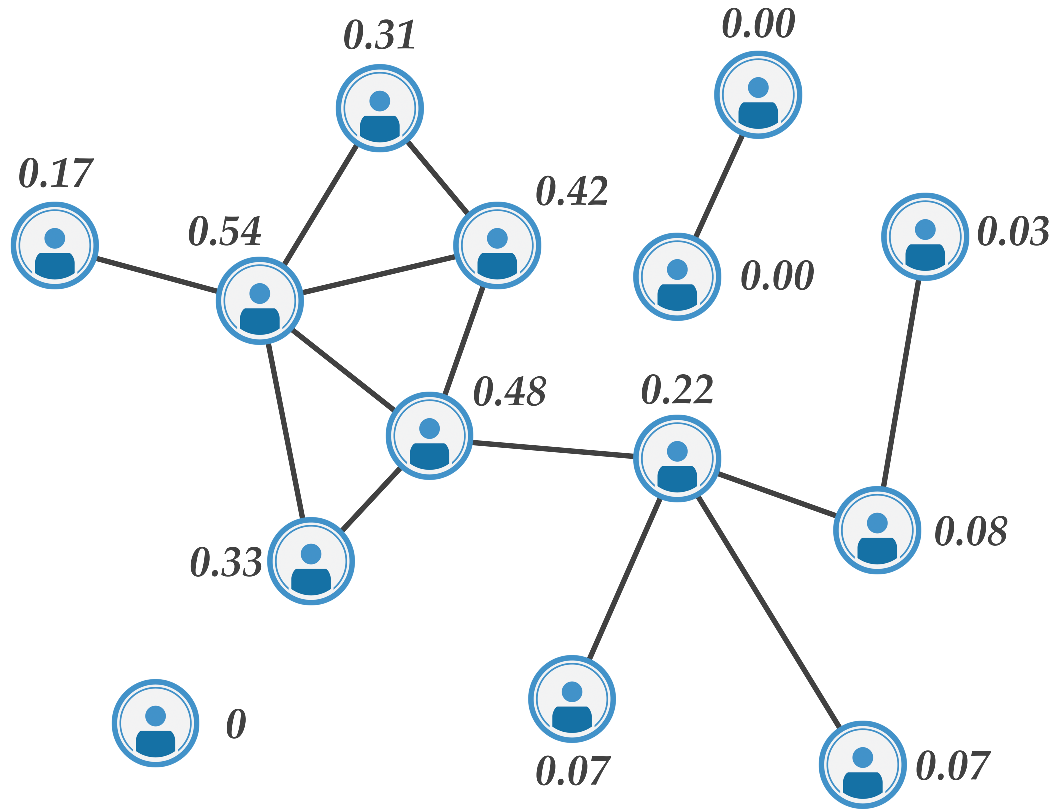
$$\boxed{x_v := \frac{1}{\lambda} \sum_{u \in V} A_{uv}\, x_u}$$

where λ is constant and

$$A_{uv} := \begin{cases} 1 & if (u,v) \in E \\ 0 & otherwise \end{cases}$$

I.e. the v^{th} entry of the eigenvector of

$$A\,x = \lambda\,x$$



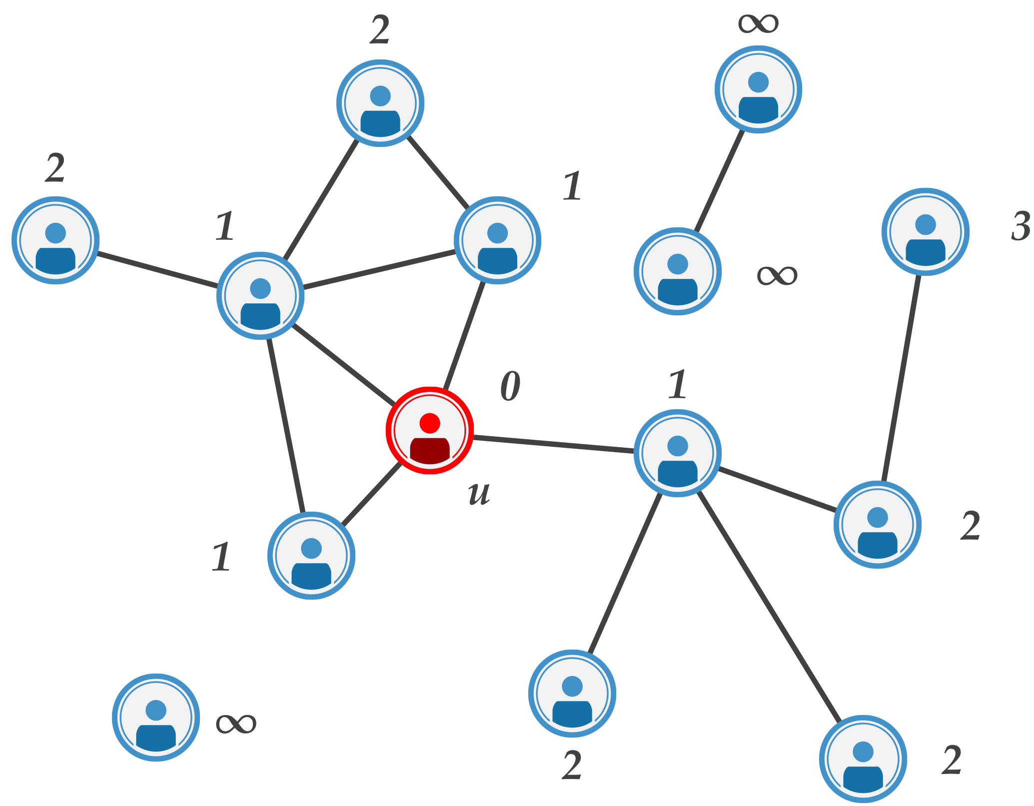
x* > *0 implies λ must be the largest eigenvalue of *A* and ***x*** the corresponding eigenvector

Persistence and Complex Networks

Erdös Distance:

Given two nodes u, v of G = (V, E),

$$E_u(v) := d(u, v)$$



Named after ***Paul Erdös***,

- one of the most prolific mathematicians of the 20th century

Persistence and Complex Networks

Centrality Measures:

A centrality measure for any query!

Degree	***How many individuals can v reach directly?***
Betweenness	***How likely is v to be the most direct route between two individuals?***
Closeness	***How fast can v reach everyone in the network?***
Eigenvector	***How well is v connected to other well-connected individuals?***
Erdös	***How far is v from a specific individual?***

Persistence and Complex Networks

Sociocentric Networks:

- ***Structural Metrics***:
  - *Average of a Centrality Measure*
  - *Diameter*
  - *Density*
  - *Transitivity*
  - ...
- ***Community Decompositions***:
  - *Atomic Communities*
  - *Clustering Techniques*

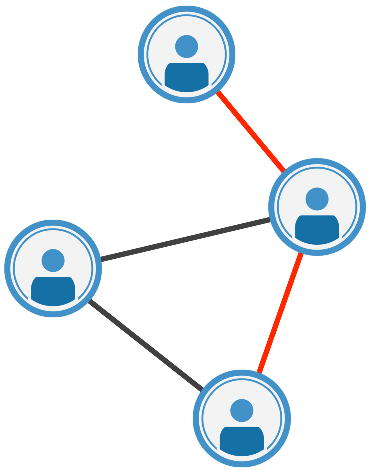
Persistence and Complex Networks

Structural Metrics:

How far are two individuals at most?

Diameter:

The longest shortest path between any two nodes



Diameter(G) = 2

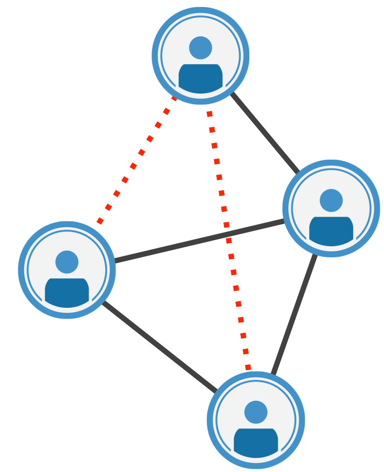
Persistence and Complex Networks

Structural Metrics:

How close is G to being an "everyone knows everyone" network?

Density:

$$\frac{\textit{Number of edges of } G}{\textit{Number of all possible edges}}$$



Density(G) = 4/6 = 0.67

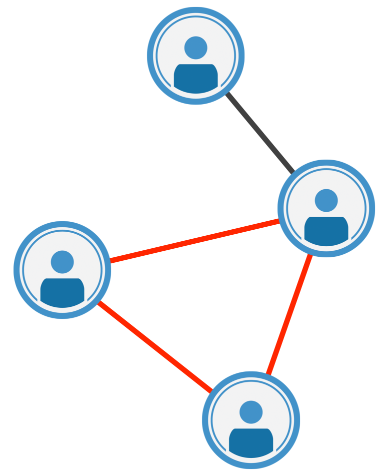
Persistence and Complex Networks

Structural Metrics:

How likely are two individuals connected to an individual v connected to each other?

Transitivity:

$$\frac{\textit{Number of closed triplets of nodes}}{\textit{Number of connected triplets}}$$



Transitivity(G) = 1/3 = 0.33

Persistence and Complex Networks

Community Decompositions:

- ***Atomic Communities:***
  - ***Clique***
  - ***n-Clique***
  - ***n-Clan***
  - ***n-Club***
  - ***k-Plex***
  - ***k-Core***
  - ...

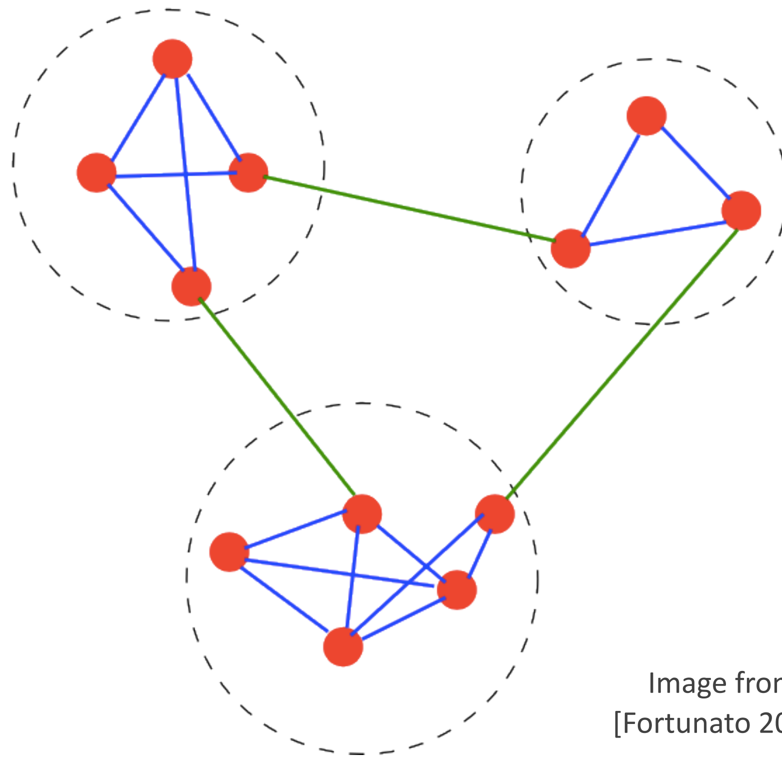


Image from
[Fortunato 2009]

Persistence and Complex Networks

Community Decompositions:

- ***Atomic Communities:***
  - *Clique*
  - *n-Clique*
  - *n-Clan*
  - *n-Club*
  - *k-Plex*
  - *k-Core*
  - ...

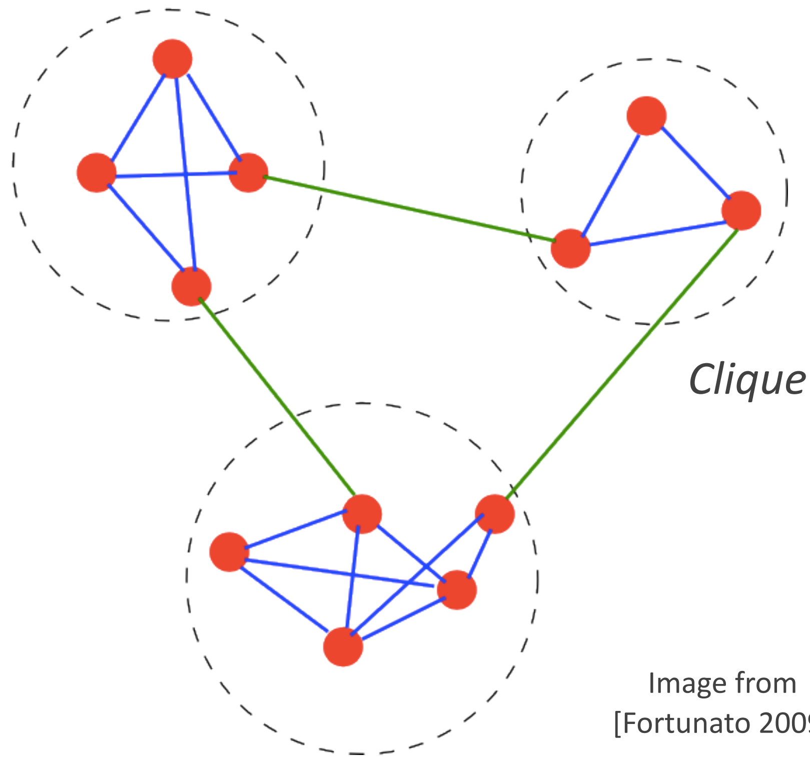


Image from [Fortunato 2009]

Clique:

A maximal subgraph whose nodes are all adjacent to each other

Persistence and Complex Networks

Community Decompositions:

- ***Atomic Communities:***
    - ***Clique***
    - ***n-Clique***
    - ***n-Clan***
    - ***n-Club***
    - ***k-Plex***
    - ***k-Core***
    - ...

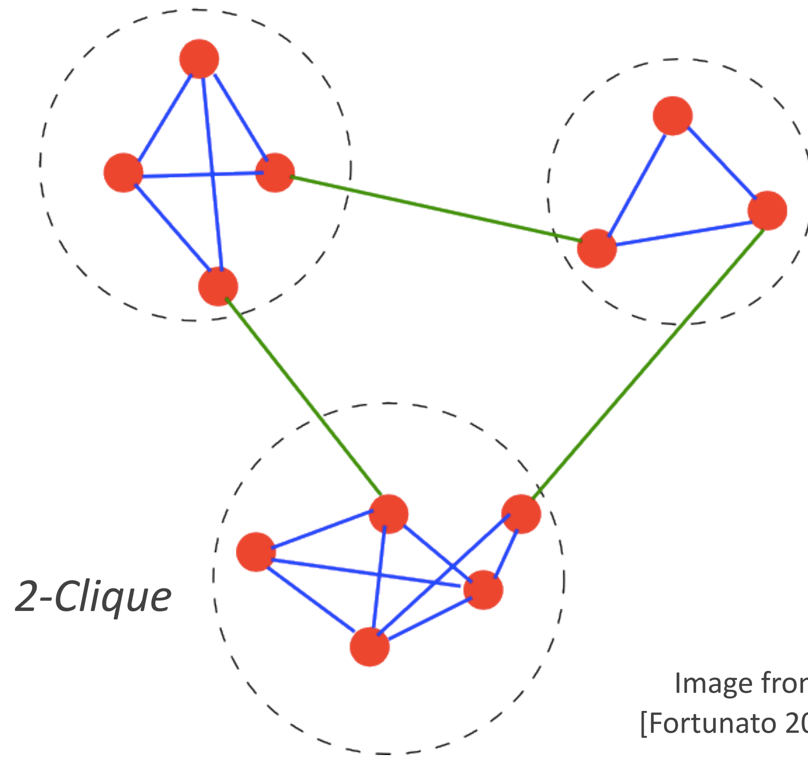


Image from [Fortunato 2009]

n-Clique:

A maximal subgraph such that the distance of each pair of its nodes is not greater than n

Persistence and Complex Networks

Community Decompositions:

- ***Atomic Communities:***
  - ***Clique***
  - ***n-Clique***
  - ***n-Clan***
  - ***n-Club***
  - ***k-Plex***
  - ***k-Core***
  - ...

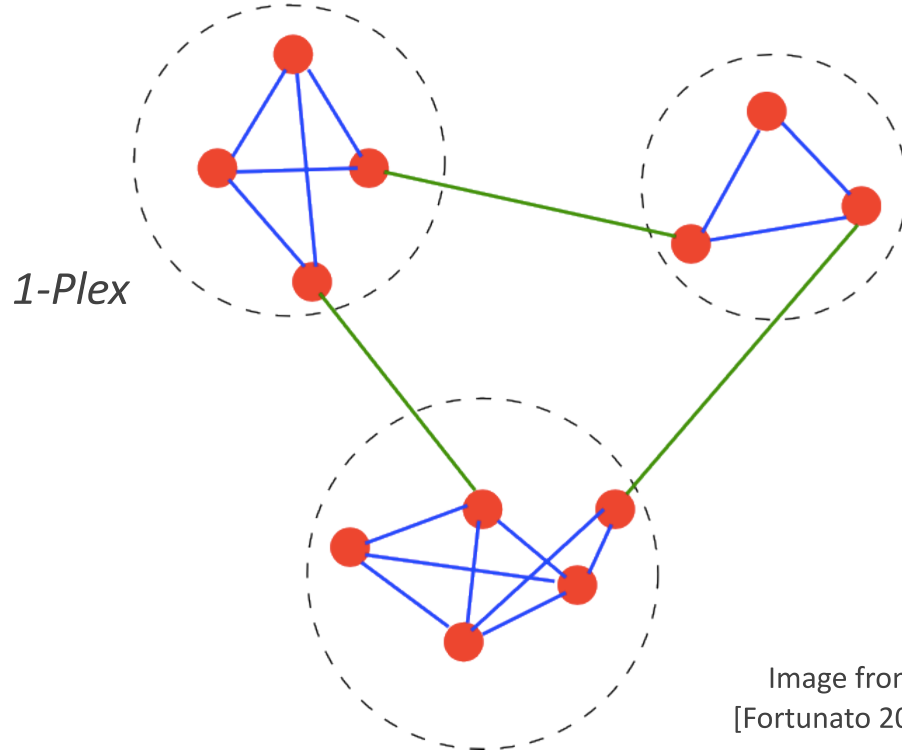


Image from [Fortunato 2009]

k-Plex:

A maximal subgraph in which each node is adjacent to all other nodes of the subgraph except at most k of them

Persistence and Complex Networks

Clustering Techniques:

Agglomerative (bottom-up)	approach based on	Centrality Measures	
		Atomic Communities	
Divisive (top-down)		Quality Functions	

Persistence and Complex Networks

Clustering Techniques:

Agglomerative (bottom-up)

Divisive (top-down)

approach based on

Centrality Measures

Atomic Communities

Quality Functions

Girvan-Newman Algorithm:

Iterated removal of the edge with largest ***betweenness centrality***

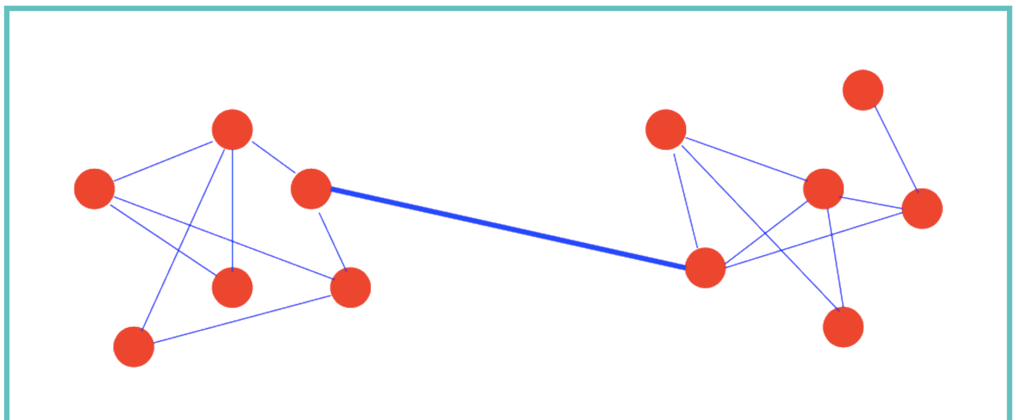


Image from [Fortunato 2009]

Persistence and Complex Networks

Clustering Techniques:

Agglomerative (bottom-up)

Divisive (top-down)

approach based on

Centrality Measures

Atomic Communities

Quality Functions

Clique Percolation:

k-adjacency: two cliques of size *k* are *k*-adjacent if they share *k-1* nodes

k-clique community: maximal union of cliques of size *k* pairwise connected by a sequence of *k*-adjacent cliques

Decomposition in *k*-clique communities

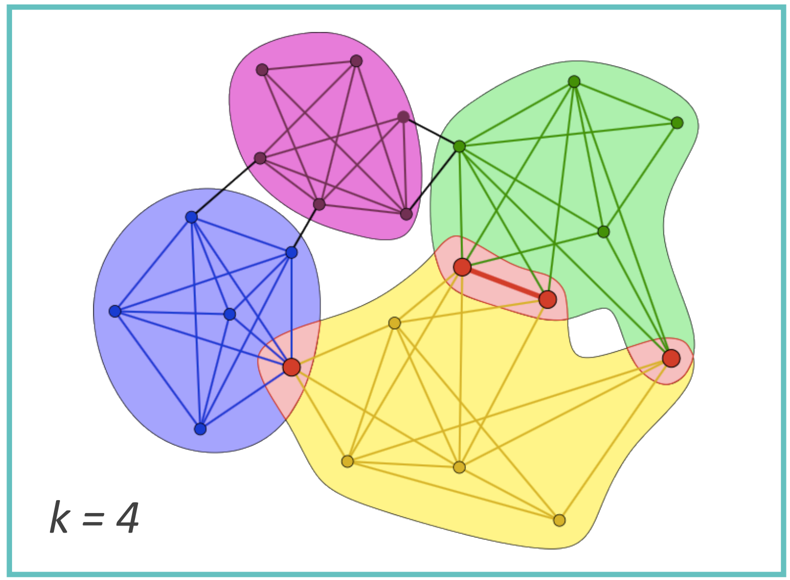


Image from [Palla et al. 2005]

Persistence and Complex Networks

Clustering Techniques:

Agglomerative (bottom-up)

Divisive (top-down)

approach based on

Centrality Measures

Atomic Communities

Quality Functions

Modularity-based Algorithm:

Modularity: measure for clustering quality

Iterated aggregation of communities of nodes whose merging ***increases modularity***

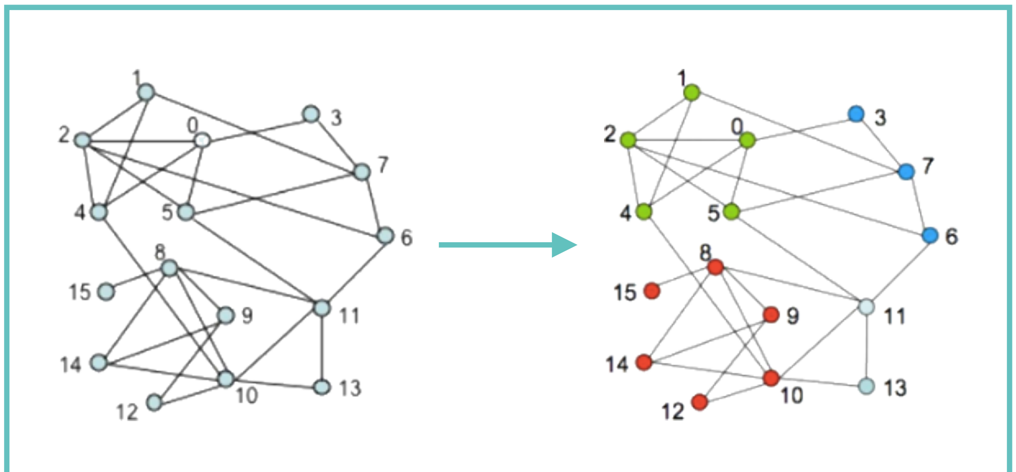


Image from [Blondel et al. 2008]

Persistence and Complex Networks

Several Application based on Persistent Homology:

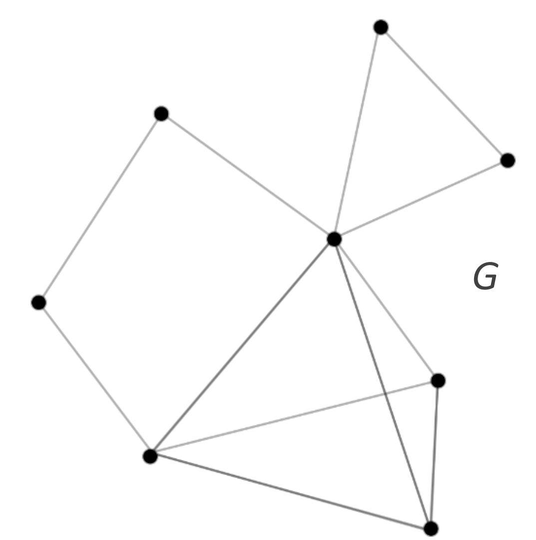
- ***Sensor*** Networks [De Silva 2013]
- ***Brain*** Networks [Lee et al. 2012]
- ***Collaborative/Co-occurence*** Networks [Carstens et al. 2013; Rieck et al. 2016]
- ***Geolocalized*** Networks [Fellegara et al. 2016]
- …

Simplicial Complex Representation:

A network is represented through:

- Simplicial complex ***Flag(G)*** induced by *G*

Simplices of Flag(G) ⟷ *Cliques of G*



Persistence and Complex Networks

Several Application based on Persistent Homology:

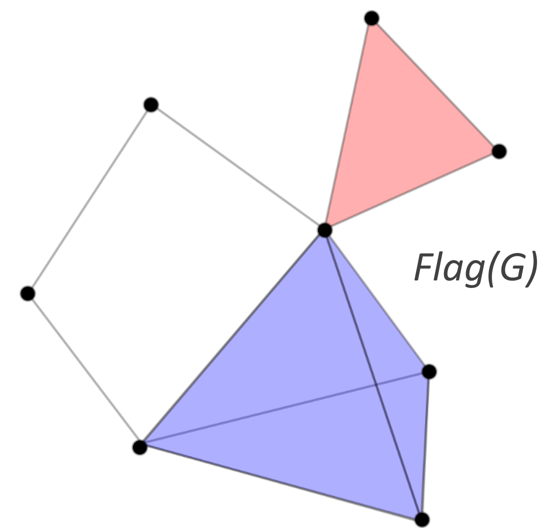
- ***Sensor*** Networks [De Silva 2013]
- ***Brain*** Networks [Lee et al. 2012]
- ***Collaborative/Co-occurence*** Networks [Carstens et al. 2013; Rieck et al. 2016]
- ***Geolocalized*** Networks [Fellegara et al. 2016]
- ...

Simplicial Complex Representation:

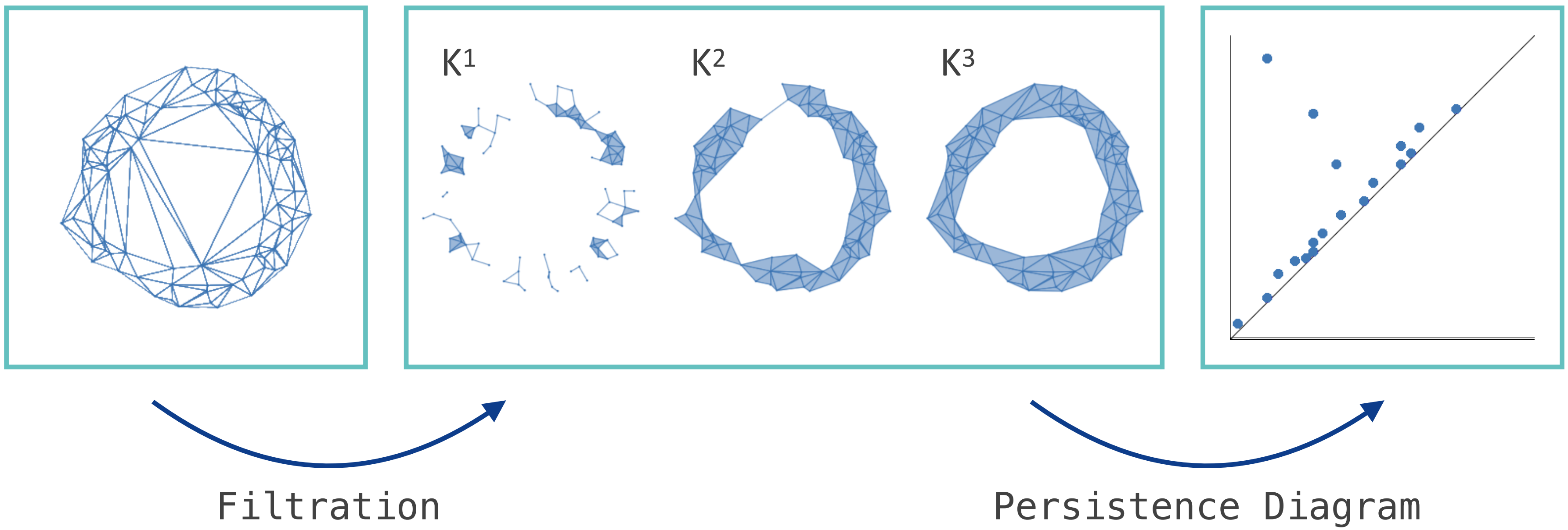
A network is represented through:

- Simplicial complex ***Flag(G)*** induced by *G*

Simplices of Flag(G) ⟷ *Cliques of G*



Persistence and Complex Networks

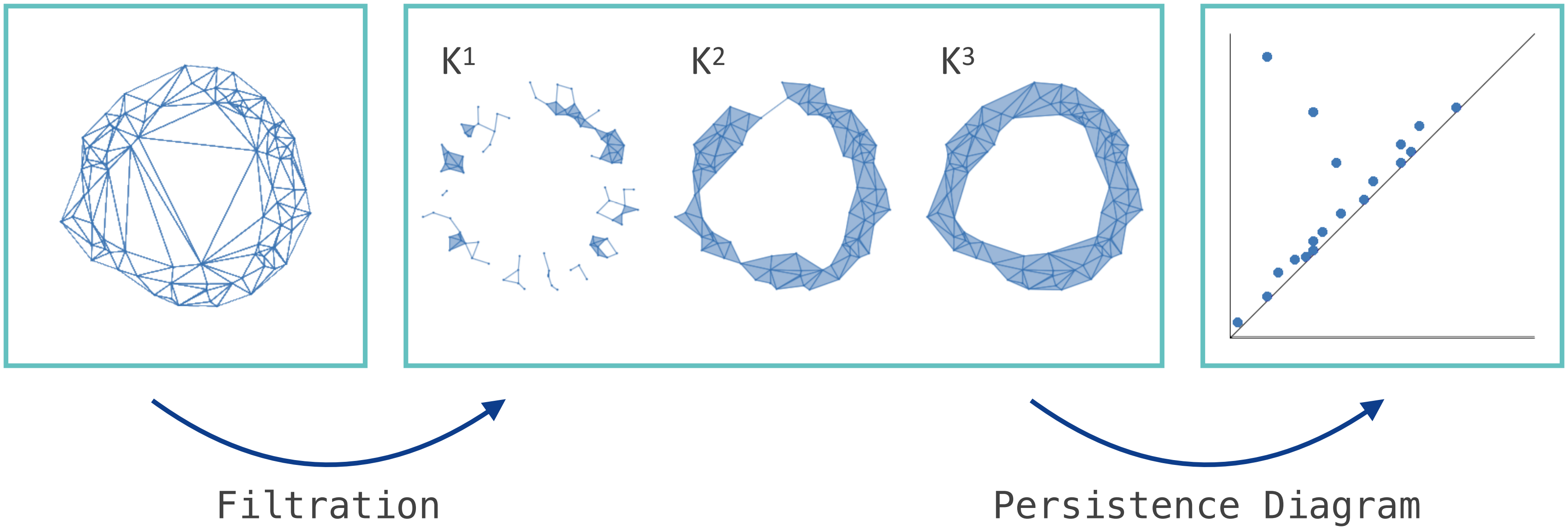


Topological summaries have proven to be particularly effective to ***distinguish shapes***

but

It is hard to obtain a ***meaningful interpretation*** for homological cycles

Persistence and Complex Networks



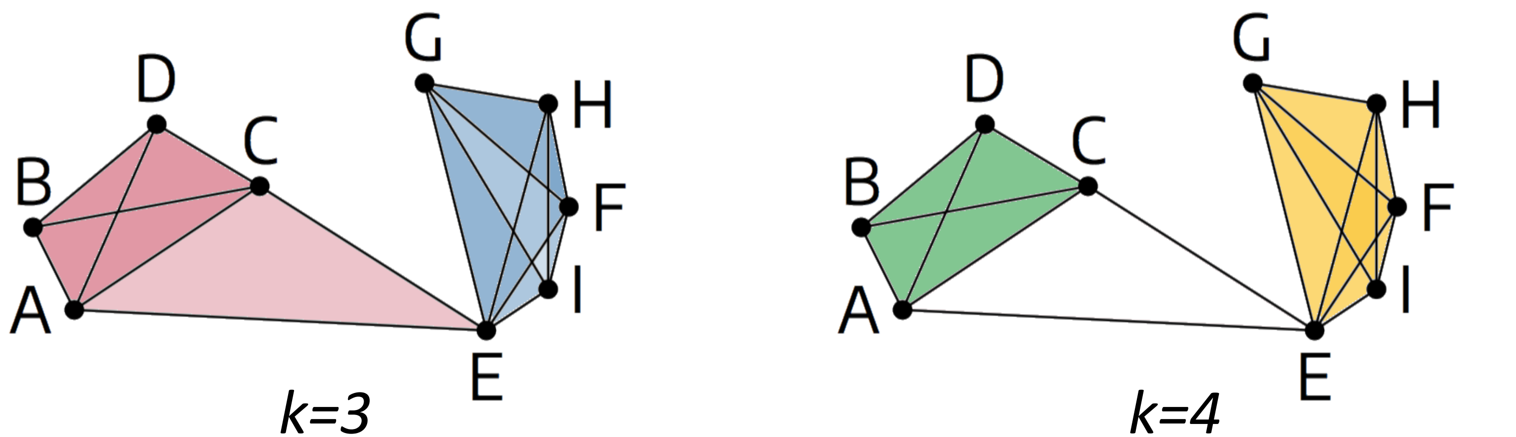
What if …

we study the persistence of something more meaningful than homological cycles?

Persistence and Complex Networks

k-Clique Community:

A ***maximal union*** of *k*-cliques ***pairwise connected*** by a ***sequence of k-adjacent cliques***

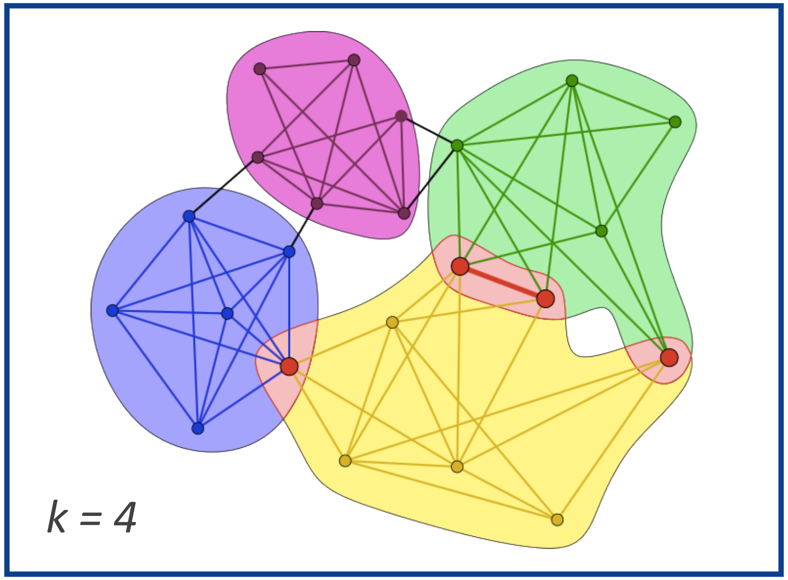


k-Adjacency:

Two *k*-cliques are ***k-adjacent*** if they ***share k-1 nodes***

Persistence and Complex Networks

k-Clique Community Decomposition:

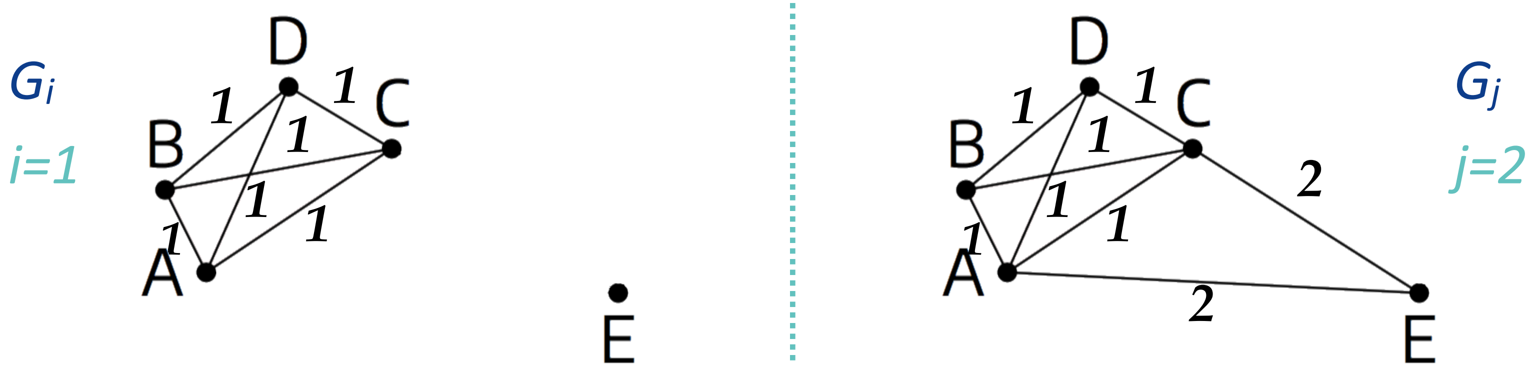


- Reveal ***highly connected*** communities
- Allow ***overlaps***
- Have a ***hierarchical structure***

Persistence and Complex Networks

Clique Communities and Weighted Networks:

Given a weighted network ***G*** and two threshold values ***i*** < ***j***,

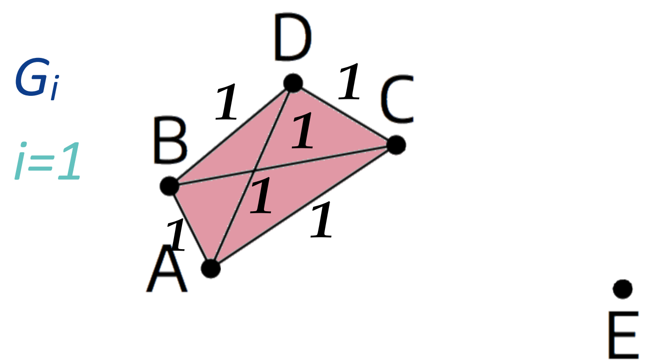


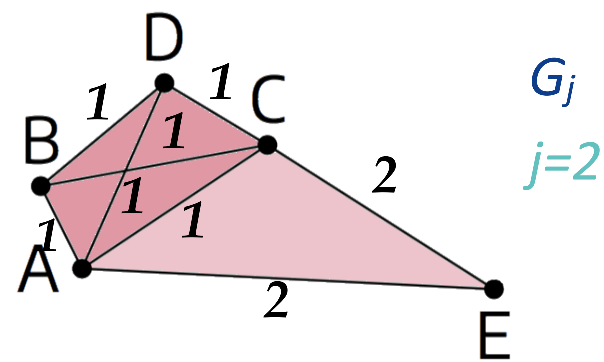
G_i is contained in G_j

Persistence and Complex Networks

Clique Communities and Weighted Networks:

Given a weighted network ***G*** and two threshold values ***i*** < ***j***,





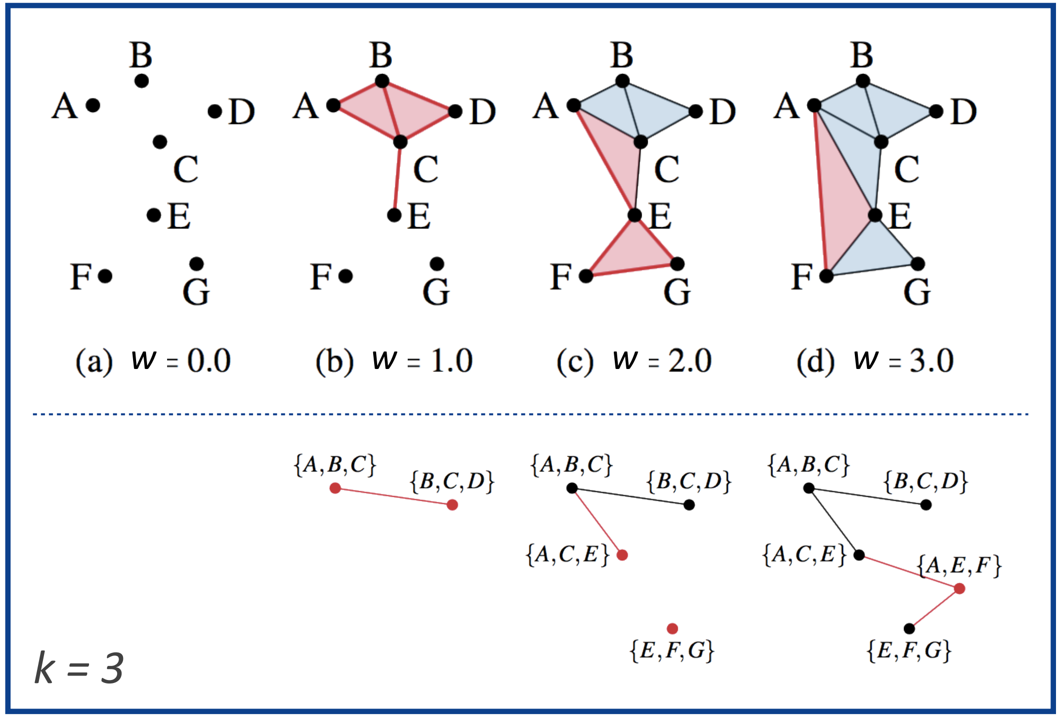
Each k-clique community of G_i is ***contained*** in ***exactly one*** k-clique community of G_j

Persistence and Complex Networks

Clique Community Persistence:

Fixing a value for k and varying the edge-weight threshold, the persistence of **k-clique communities** of G can be tracked by:

- Building a sequence of *k*-dual graphs:
  - ***vertices ↔ k-cliques***
  - ***edges ↔ adjacent k-cliques***
- Tracking the ***connected components*** of the sequence of *k*-dual graphs



Persistence and Complex Networks

Clique Community Persistence:

The presented approach allows for designing tools for:

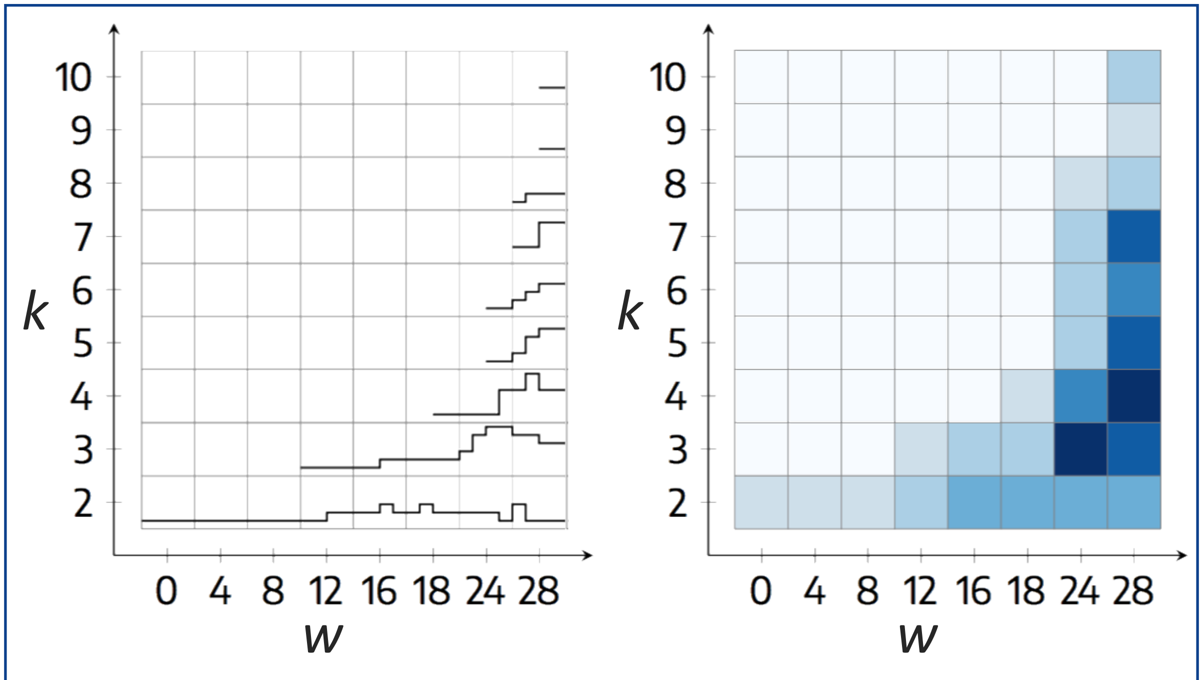
- ***Network Comparison***
  - ***Comparison Measures***
    - ***Persistence Indicator Function (PIF)***
    - ***PIF-based distance***
  - Clique Community ***Centrality Measure***
- ***Single Network Analysis***
  - ***Interactive Visualization Tool*** based on Nested Graphs

Persistence and Complex Networks

Persistence Indicator Function: Defined as the function $f_k : \mathbb{R} \longrightarrow \mathbb{N}$

assigning:

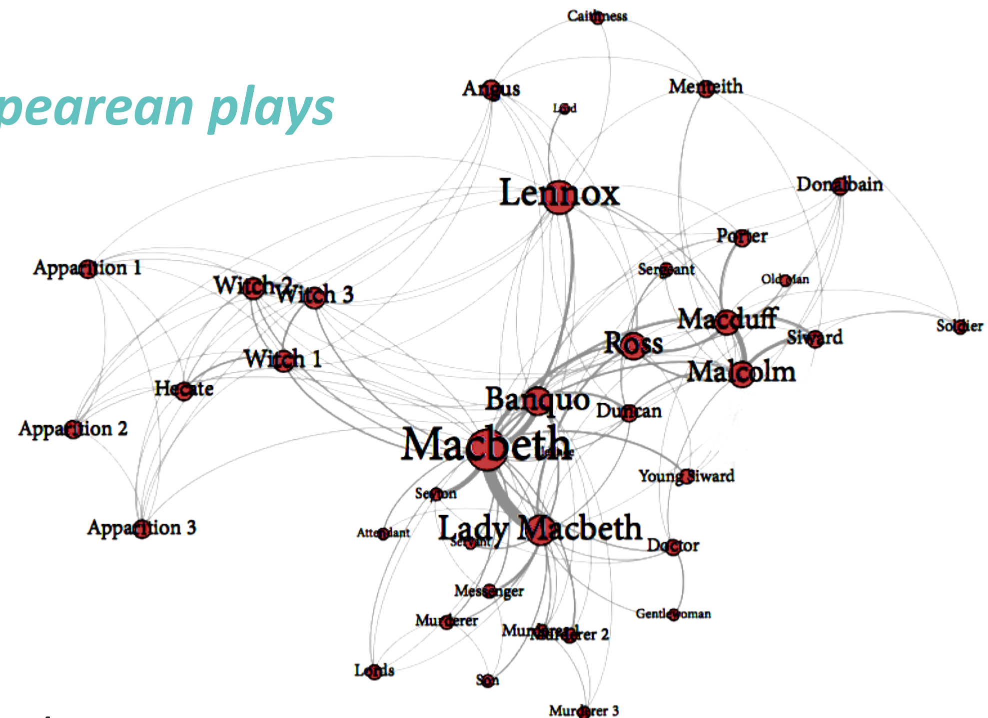
w $\longmapsto$ ***# k-cliques communities "alive" at threshold w***



Persistence and Complex Networks

Persistence Indicator Function:

- ***Co-occurrence networks*** of ***Shakespearean plays***
  - *37 plays considered*

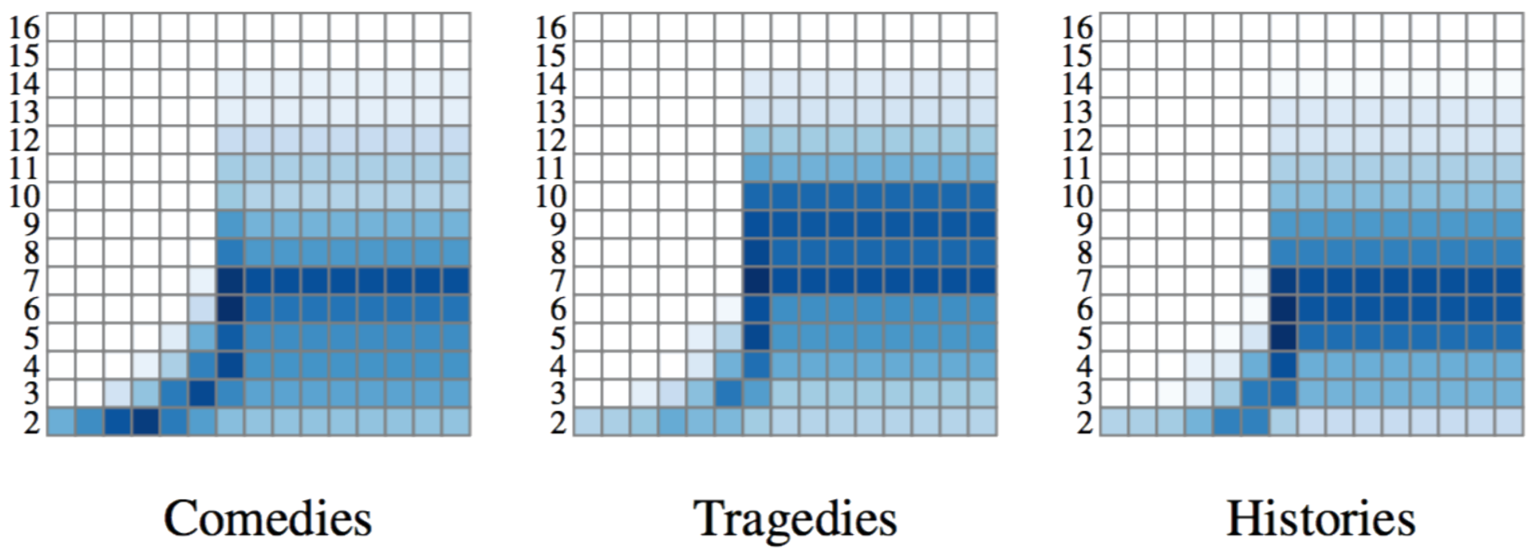


- In each network:
  - ***nodes*** ↔ ***characters*** *of the play*
  - ***edges*** ↔ *characters appearing in the* ***same scene***
  - ***edge weight*** ↔ *inverse of the* ***number of interactions***

Persistence and Complex Networks

Persistence Indicator Function:

PIF enables a comparison of structural differences between groups of plays



Persistence and Complex Networks

PIF-Based Distance:

Given two persistence indicator functions f and g,

PIF-based distance is defined to be the L^p distance between f and g:

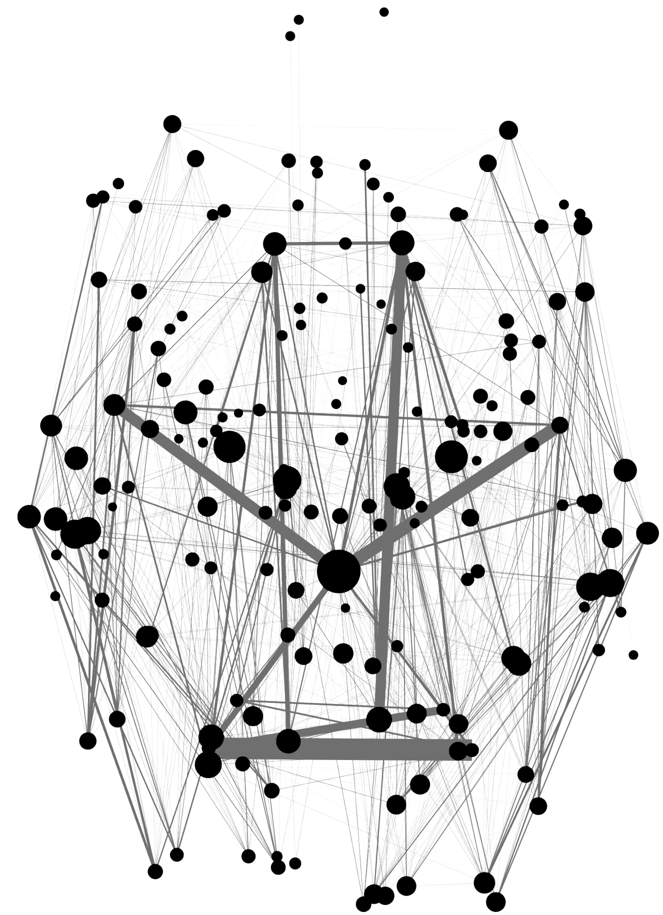
$$dist(f,g) = \left(\int_{\mathbb{R}} |f(x) - g(x)|^p \, dx \right)^{\frac{1}{p}}$$

- Quantifies dissimilarities between PIFs
- Easier to be computed than Wasserstein and bottleneck distances
- Highly correlated to Wasserstein distance

Persistence and Complex Networks

PIF-Based Distance:

- ***Biological networks*** representing variants of ***human brain connectivity***
    - *9 instances considered*

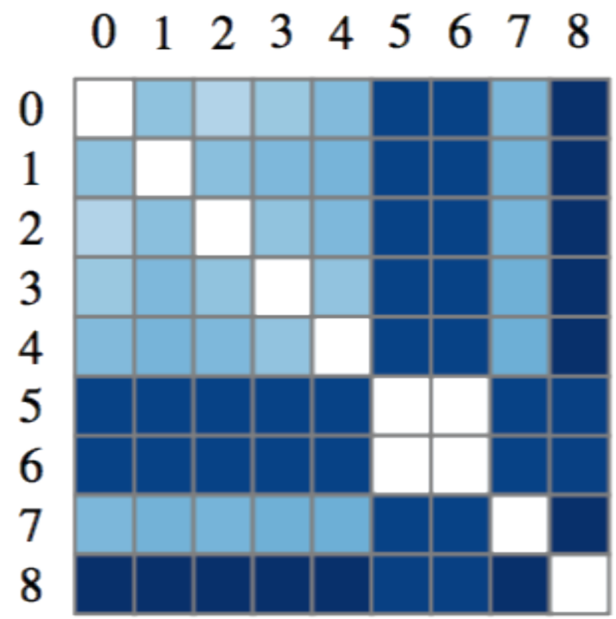


- In each network:
    - ***nodes*** ⟷ ***brain areas***
    - ***edges*** ⟷ ***fibers connecting different areas***

Persistence and Complex Networks

PIF-Based Distance:

Variant	Density	Diam. (weighted)	Avg. degree (weighted)
0	0.125	4 (60.0)	21.21 (2300.3)
1	0.124	4 (60.0)	21.06 (2296.0)
2	0.124	4 (60.0)	21.13 (2295.2)
3	0.124	4 (60.0)	21.16 (2282.0)
4	0.124	4 (60.0)	21.15 (2279.3)
5	0.125	4 (60.0)	21.19 (2264.0)
6	0.125	4 (60.0)	21.19 (2264.0)
7	0.124	4 (60.0)	21.16 (2279.6)
8	0.125	4 (60.0)	21.20 (2257.5)



PIF-based distance reveals differences between networks that common graph measures are incapable of detecting

Persistence and Complex Networks

Clique Community Centrality:

Clique community centrality of a node *v* is defined as

$$centrality(v) = \sum_{C \ni v} pers(C)$$

where:

- C is any clique community containing *v*
- ***pers(C)*** is the "lifespan" of C

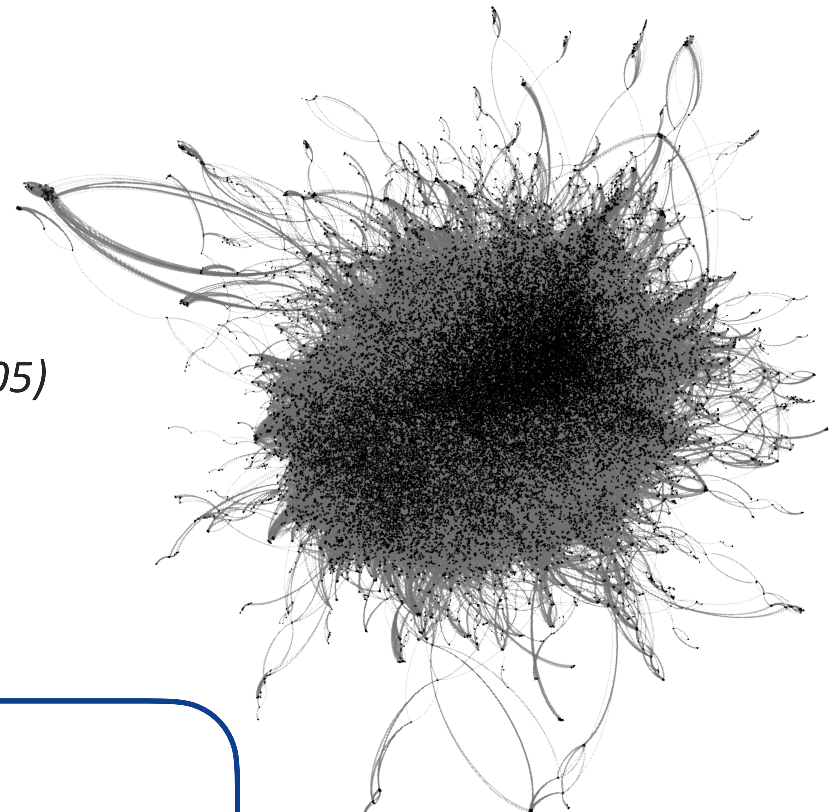
*Nodes belonging to **high-persistence communities** are identified as **relevant***

Persistence and Complex Networks

Clique Community Centrality:

- ***Collaborative networks*** describing ***scientist co-authorship of the "Condensed Matter"*** arXiv category
  - *3 snapshots in time considered (1999, 2003, 2005)*

- Network sizes:
  - *16K - 40K nodes*
  - *47K - 175K edges*



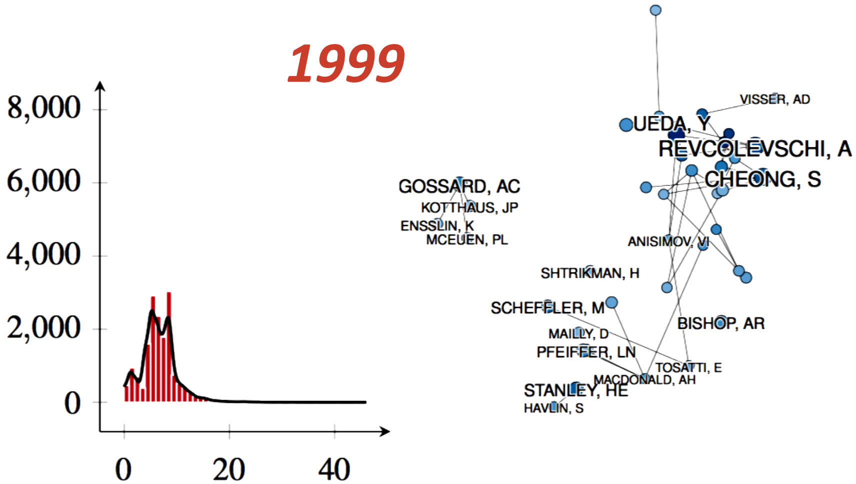
Clique community centrality allows for

- *evaluating the **evolution of network connectivity***
- *filtering away the **less relevant nodes***

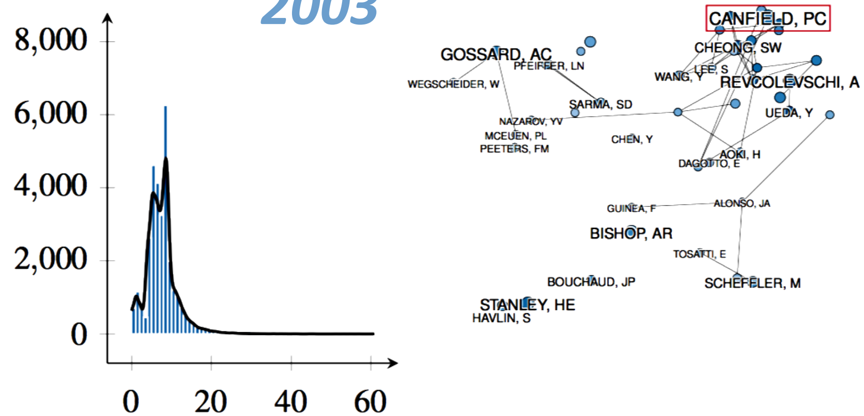
Persistence and Complex Networks

Clique Community Centrality:

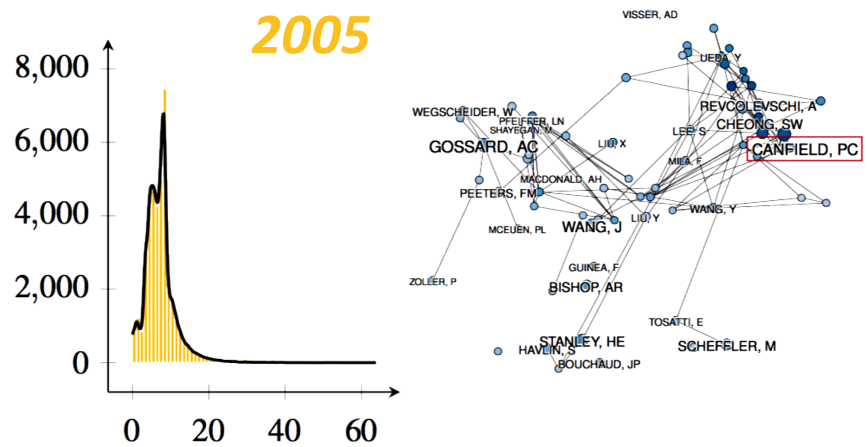
1999



2003



2005



Density estimates of the clique community centrality values

Persistence and Complex Networks

Clique Community Persistence:

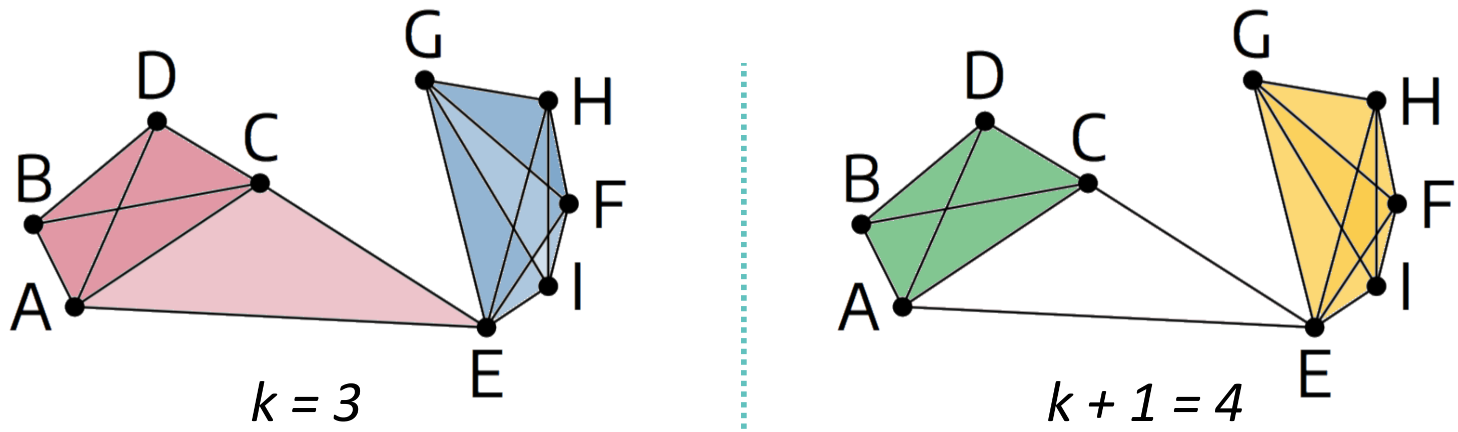
The presented approach allows for designing tools for:

- ***Network Comparison***
  - ***Comparison Measures***
    - ***Persistence Indicator Function (PIF)***
    - ***PIF-based distance***
  - Clique Community ***Centrality Measure***
- ***Single Network Analysis***
  - ***Interactive Visualization Tool*** based on Nested Graphs

Persistence and Complex Networks

Clique Communities and Multiple k-Values:

Given a weighted network ***G*** and any threshold value ***i***,

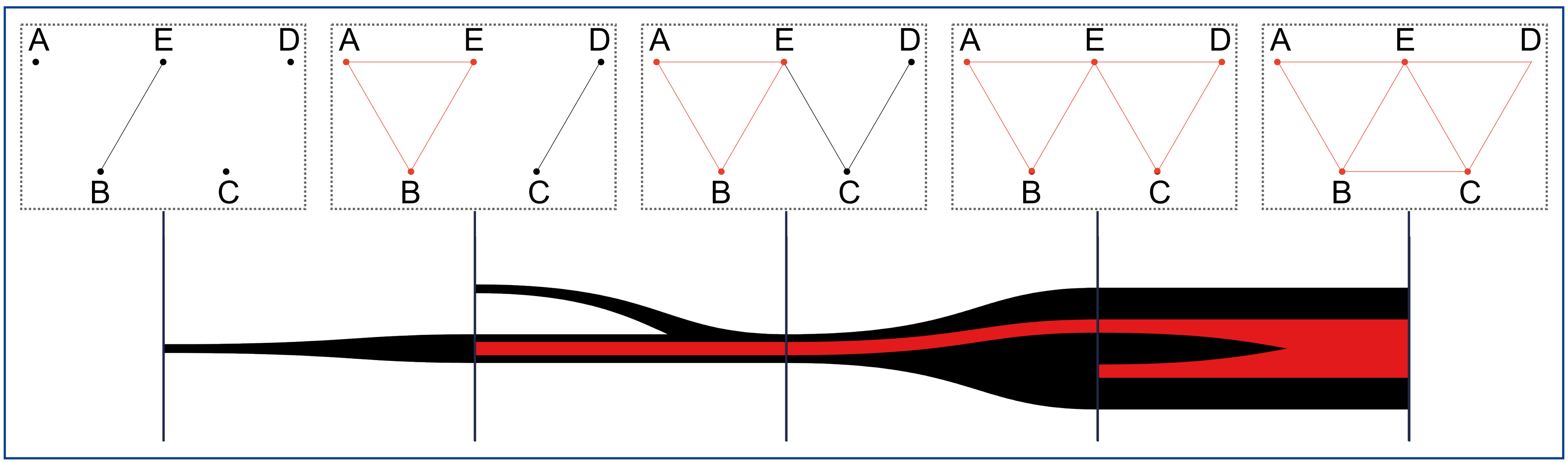


Each *(k+1)*-clique community of G_i is **contained** in ***exactly one*** *k*-clique community of G_i

Persistence and Complex Networks

Nested Graph:

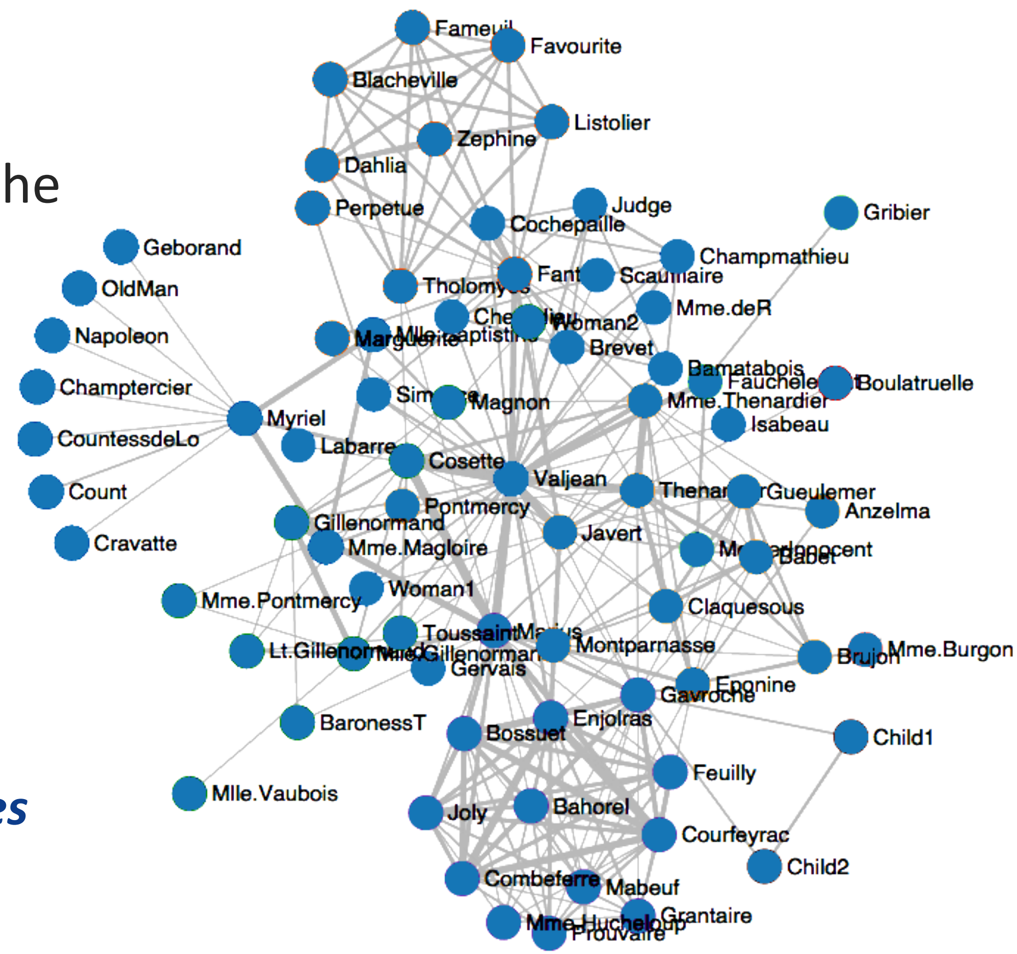
- Originally defined for connected components in scalar fields [Lukasczyk et al. 2017]
- Illustrates ***evolutions across two parameters***



Persistence and Complex Networks

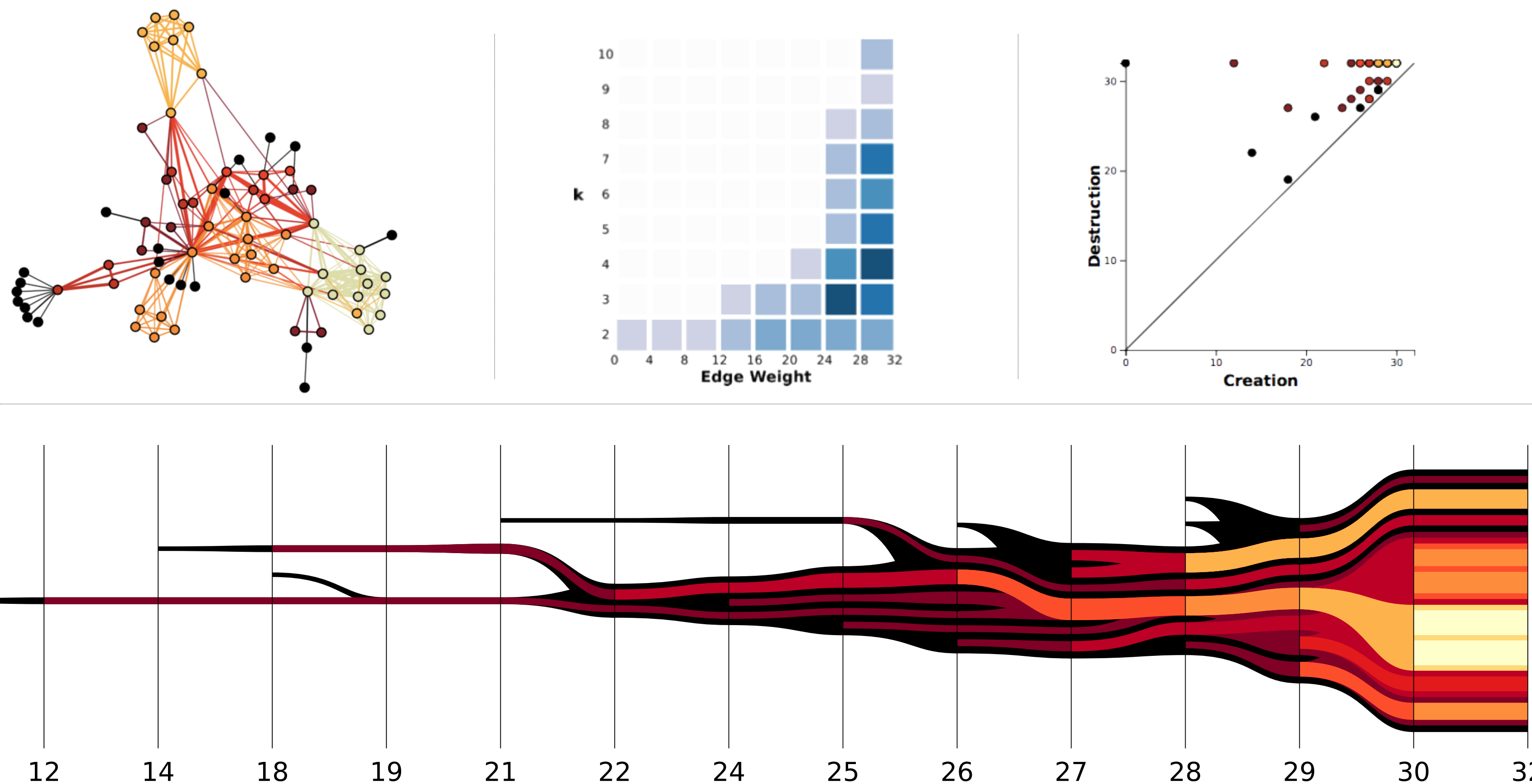
Nested Graph:

- ***Co-occurrence network*** between the characters in **Victor Hugo's novel** ***"Les Misérables"***
  - *77 nodes*
  - *254 edges*



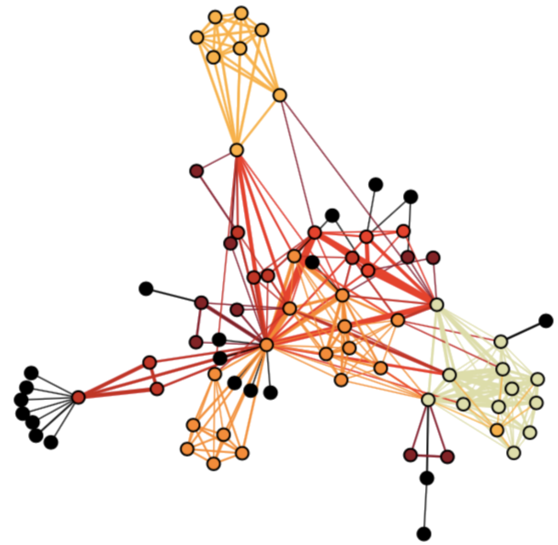
- ***edge weight*** $\leftrightarrow$ ***1 / # co-occurrences***

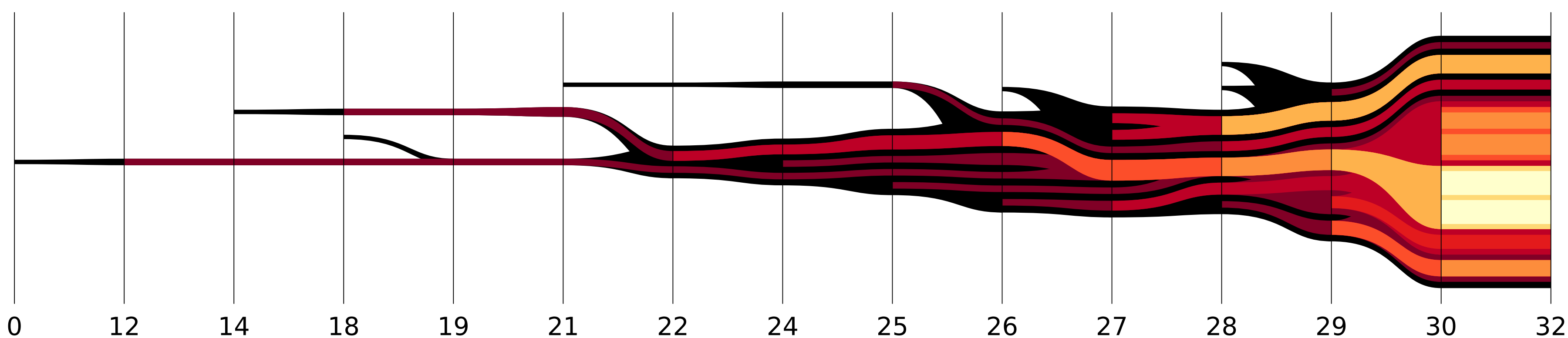
Persistence and Complex Networks



Persistence and Complex Networks

Nested-based visualization tool allows the user for

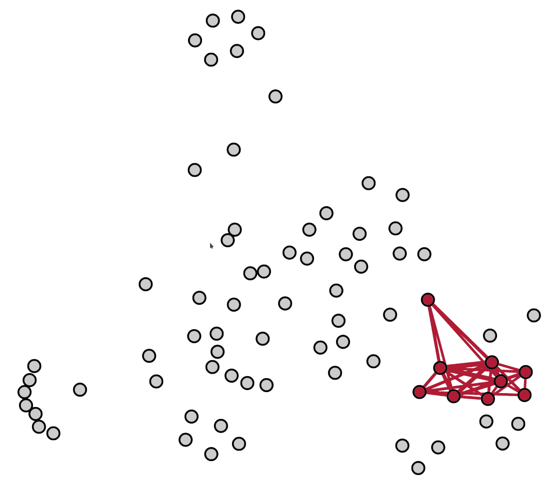




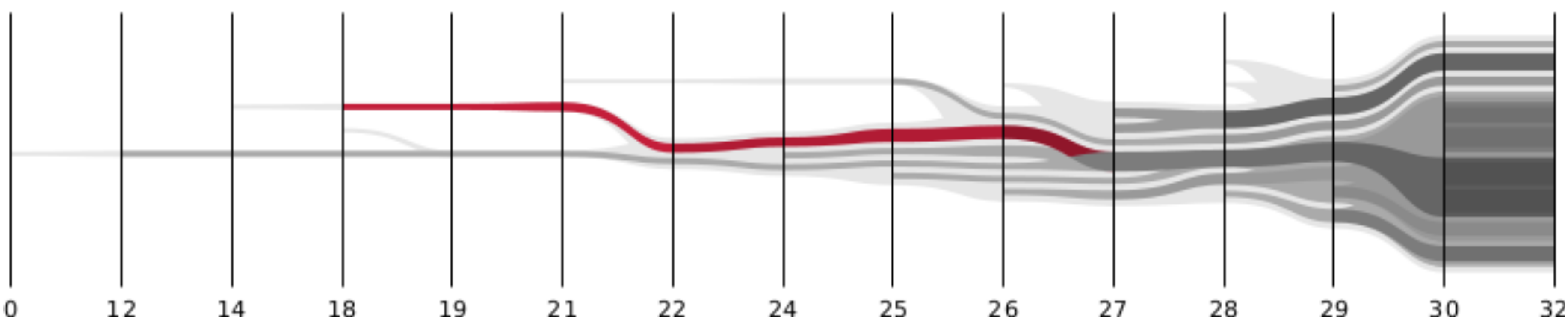
Persistence and Complex Networks

Nested-based visualization tool allows the user for

- ***focusing on*** *the evolution of a* ***specific clique community***
- ***selecting*** *and* ***interactivity exploring different edge weights*** *and* ***clique degrees***



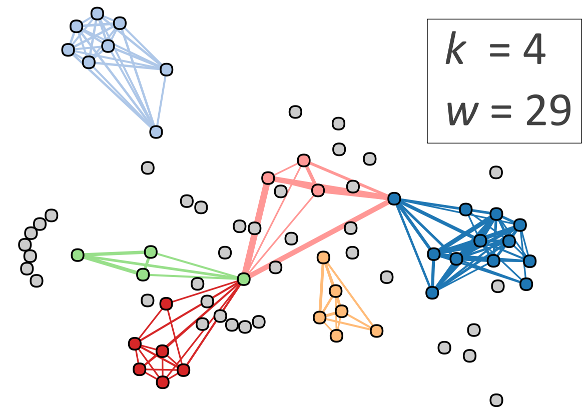
while the force-directed graph layout and the nested graph ***change accordingly***



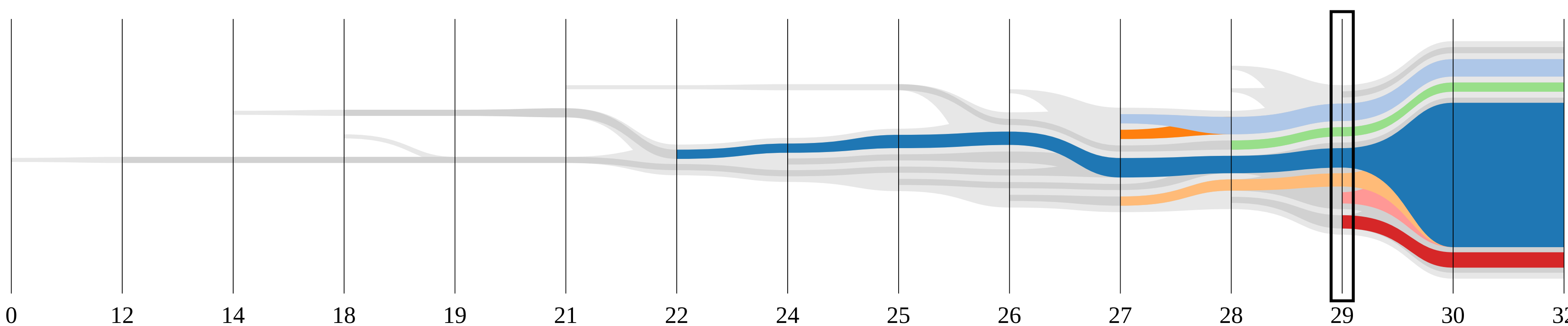
Persistence and Complex Networks

Nested-based visualization tool allows the user for

- ***focusing on*** *the evolution of a* ***specific clique community***
- ***selecting*** *and* ***interactivity exploring different edge weights*** *and* ***clique degrees***



while the force-directed graph layout and the nested graph ***change accordingly***



Persistence and Complex Networks

Nested-based visualization tool allows the user for

- ***focusing on*** *the evolution of a* ***specific clique community***
- ***selecting*** *and* ***interactivity exploring*** ***different edge weights*** *and* ***clique degrees***

while the force-directed graph layout and the nested graph ***change accordingly***

Intuitively:

edge-weight *variation*	↔	*reveal the* ***core part of a community***
clique-degree *variation*	↔	*analyze* ***community according to different granularities***

Bibliography

General References:

- ***Books on TDA:***
  - H. Edelsbrunner, J. Harer. ***Computational topology: an introduction***. American Mathematical Soc., 2010.
  - R. W. Ghrist. ***Elementary applied topology***. Seattle: Createspace, 2014.

- ***Papers on TDA:***
  - G. Carlsson. ***Topology and data***. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.

Bibliography

Today's References:

- ***Kernels:***

- P. Donatini, P. Frosini, A. Lovato. ***Size functions for signature recognition.*** Proc. of SPIE, 3454, pages 178-183, 1998.
- M. Ferri, P. Frosini, A. Lovato, C. Zambelli. ***Point selection: a new comparison scheme for size functions.*** Proc. 3rd Asian Conference on Computer Vision, vol. I, pages 329-337, 1998.
- P. Bubenik. ***Statistical topological data analysis using persistence landscapes.*** Journal of Machine Learning Research, 16.1, pages 77-102, 2015.
- J. Reininghaus, S. Huber, U. Bauer, R. Kwitt. ***A stable multi-scale kernel for topological machine learning***. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pages 4741-4748, 2015.
- G. Kusano, Y. Hiraoka, K. Fukumizu. ***Persistence weighted Gaussian kernel for topological data analysis.*** International Conference on Machine Learning, pages 2004-2013, 2016.
- M. Carrière, M. Cuturi, S. Oudot. ***Sliced Wasserstein kernel for persistence diagrams.*** Proc. of the 34th International Conference on Machine Learning, 70, pages 664-673, 2017.
- H. Adams et al. ***Persistence images: a stable vector representation of persistent homology.*** Journal of Machine Learning Research, 18.1, pages 218-252, 2017.

Bibliography

Today's References:

- ***Networks:***

  - G.Palla, I.Derényi, I. Farkas, T. Vicsek. ***Uncovering the overlapping community structure of complex networks in nature and society.*** Nature, 435.7043, pages 814-818, 2005.
  - S. Fortunato. ***Community detection in graphs.*** Physics Reports, 486(3-5), pages 75-174, 2010.
  - J. Scott. ***Social network analysis.*** SAGE Publications Ltd, 2017.
  - M. Newman. ***Networks.*** Oxford university press, 2018.
  - B. Rieck, U. Fugacci, J. Lukasczyk, H. Leitte. ***Clique community persistence: a topological visual analysis approach for complex networks.*** IEEE Transactions on Visualization and Computer Graphics, 24.1, pages 822-831, 2018.