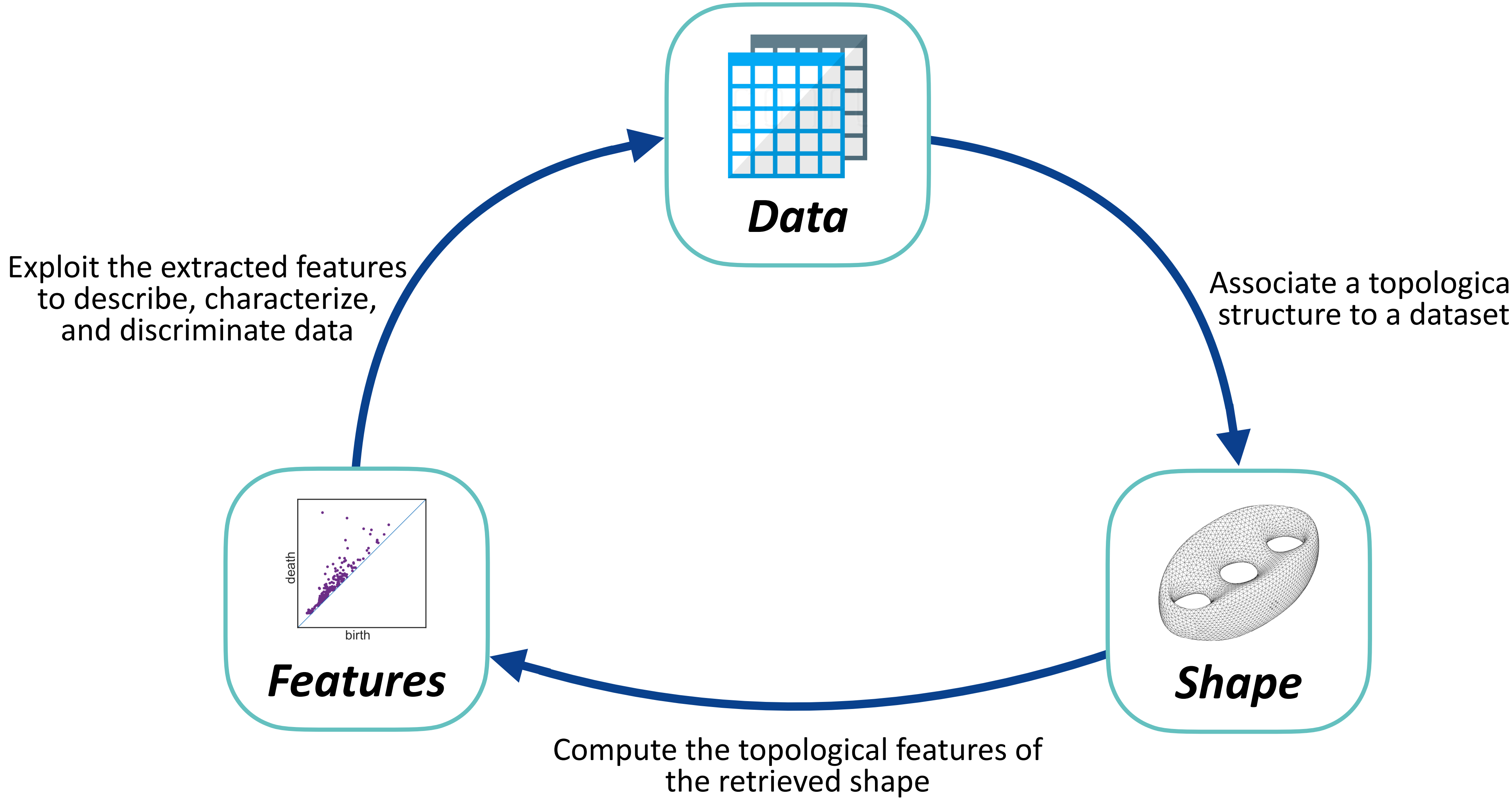
Topological Data Analysis

Persistence & Networks

Ulderico Fugacci

CNR - IMATI

Topological Data Analysis



Persistence and Complex Networks

- *A Primer on Complex Networks*
- *Homological Scaffolds*
- *Clique Community Persistence*

Persistence and Complex Networks

- ***A Primer on Complex Networks***
- *Homological Scaffolds*
- *Clique Community Persistence*

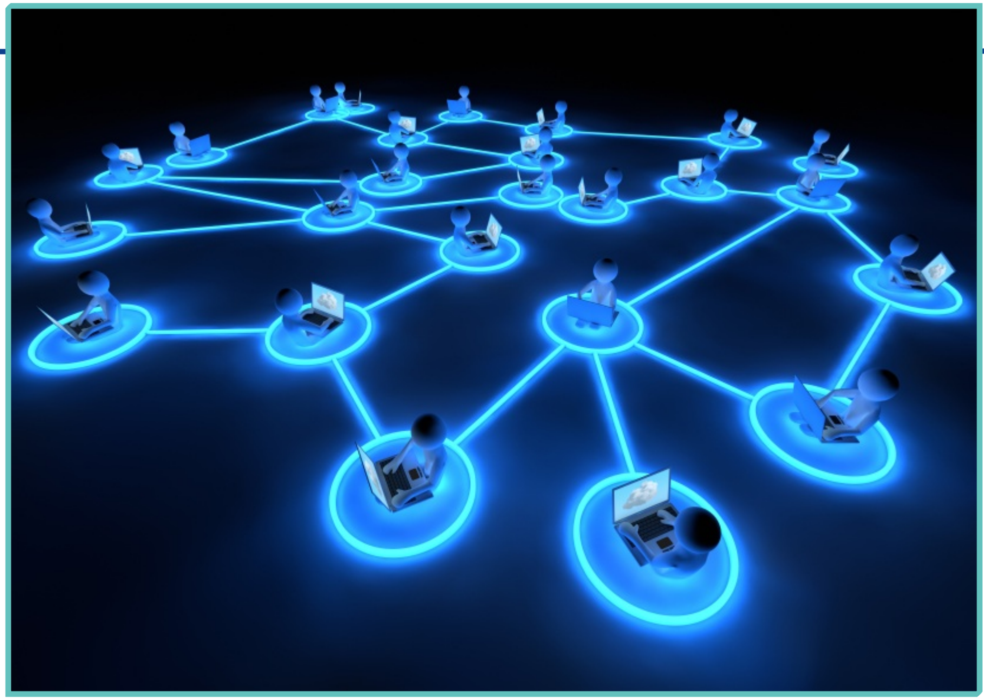
Persistence and Complex Networks

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
- …

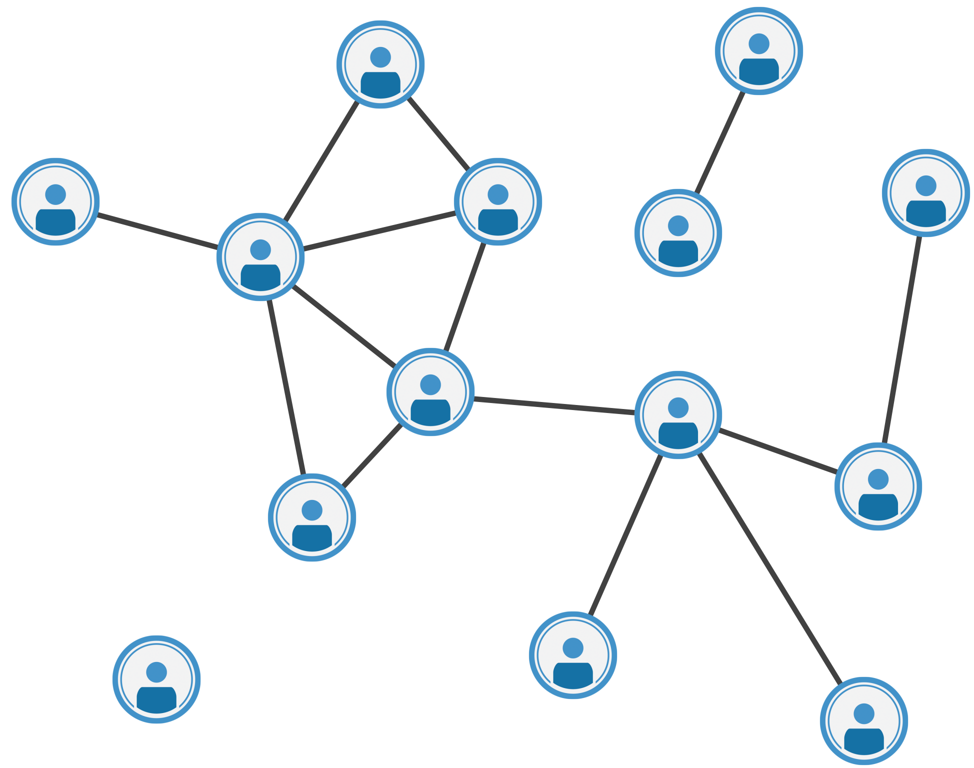


Persistence and Complex Networks

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:

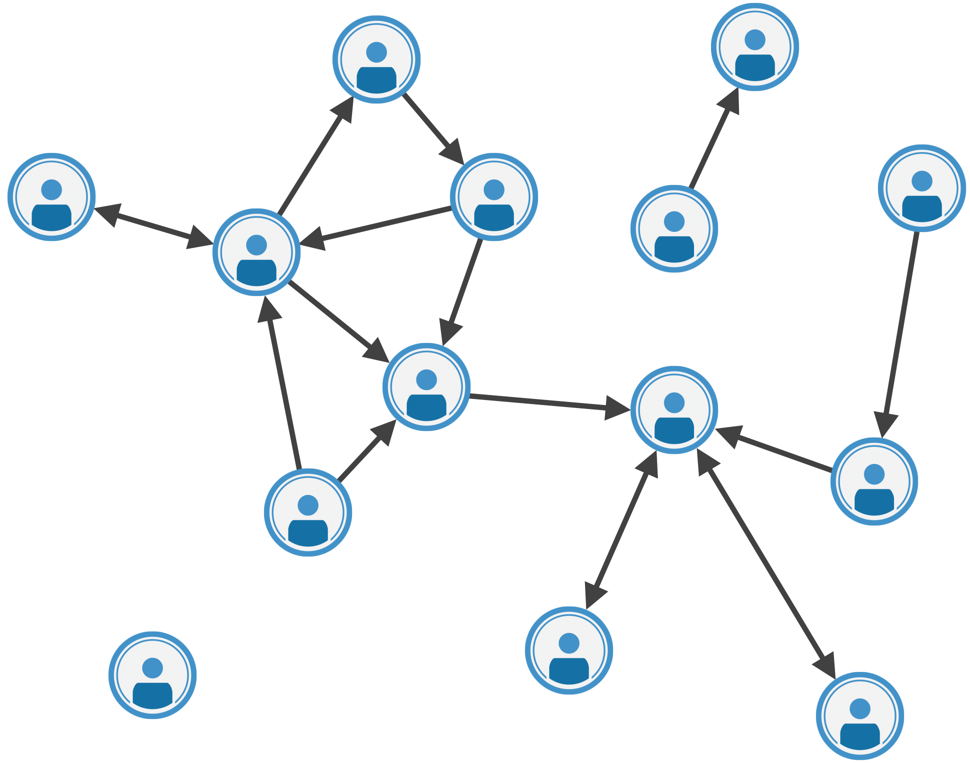
A network can be represented by a ***graph*** ***G = (V, E)*** such that:

Nodes of V ⟺ *Entities or individuals*

Arcs of E ⟺ *Ties between entities*

Arcs can be:

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



Persistence and Complex Networks

Representation:

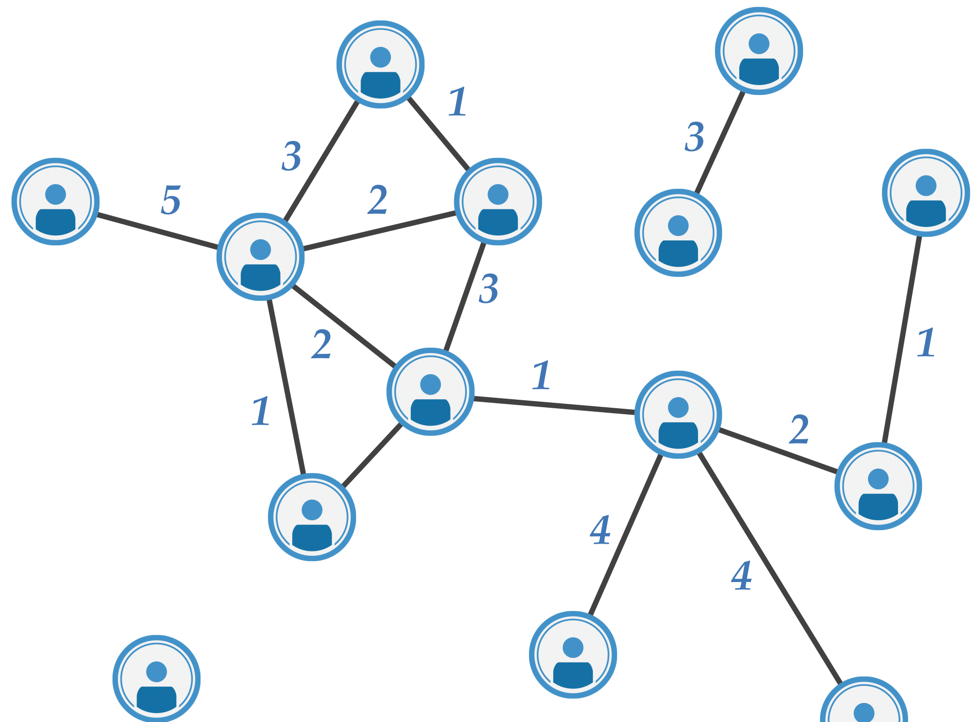
A network can be represented by a ***graph*** ***G = (V, E)*** such that:

Nodes of V ⇔ *Entities or individuals*

Arcs of E ⇔ *Ties between entities*

Arcs can be:

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



Persistence and Complex Networks

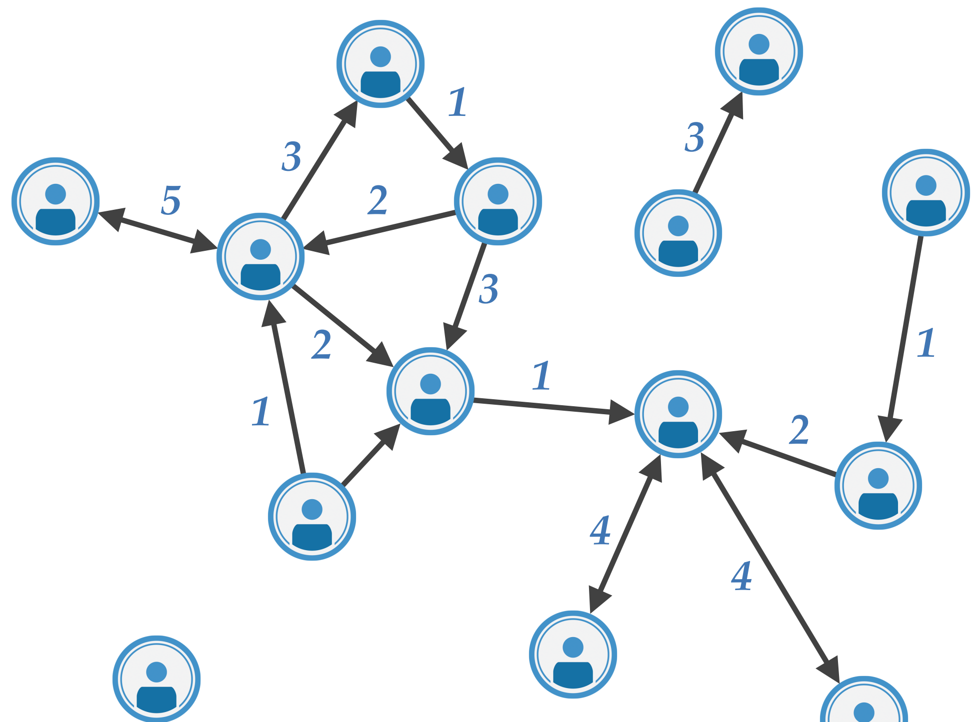
Representation:

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*

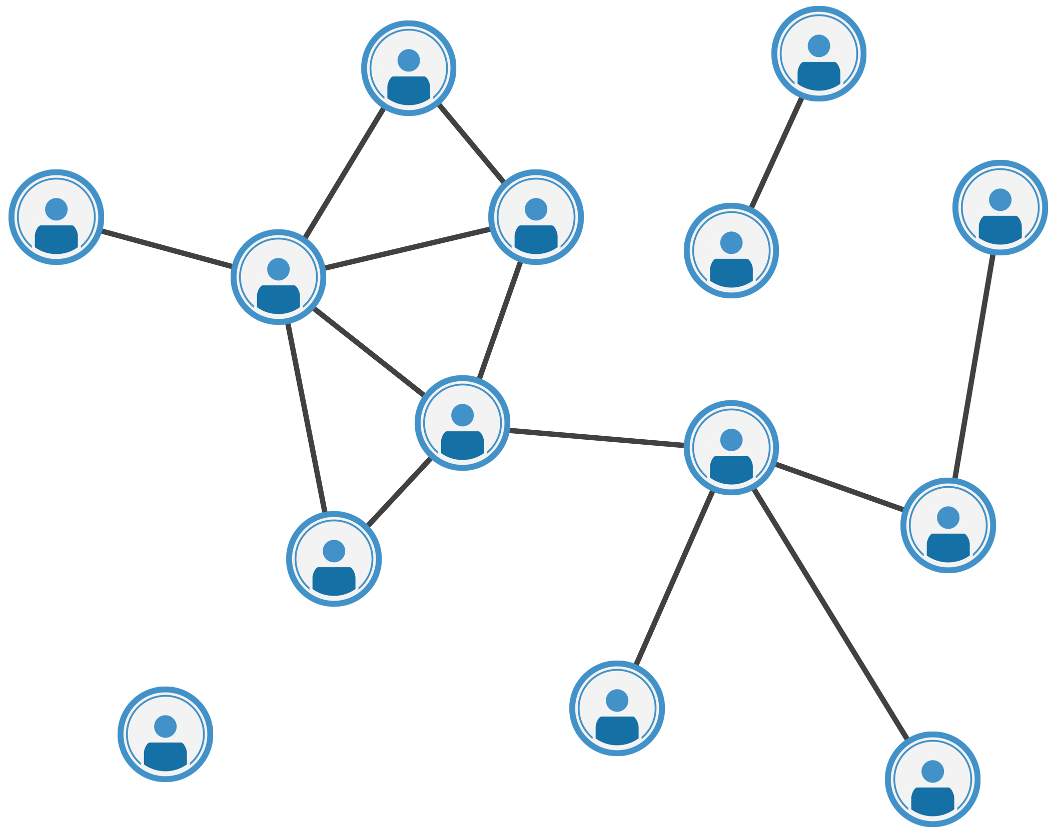
Arcs can be:

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



Persistence and Complex Networks

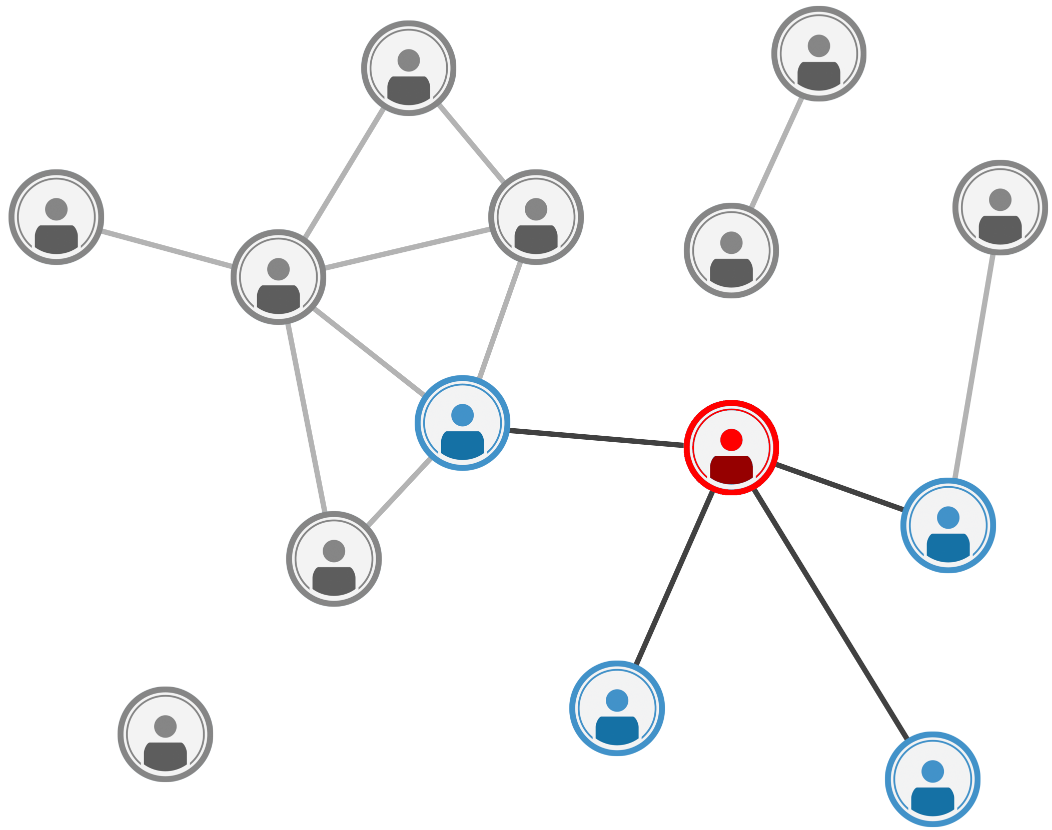
A Two-Level Analysis:



Persistence and Complex Networks

A Two-Level Analysis:

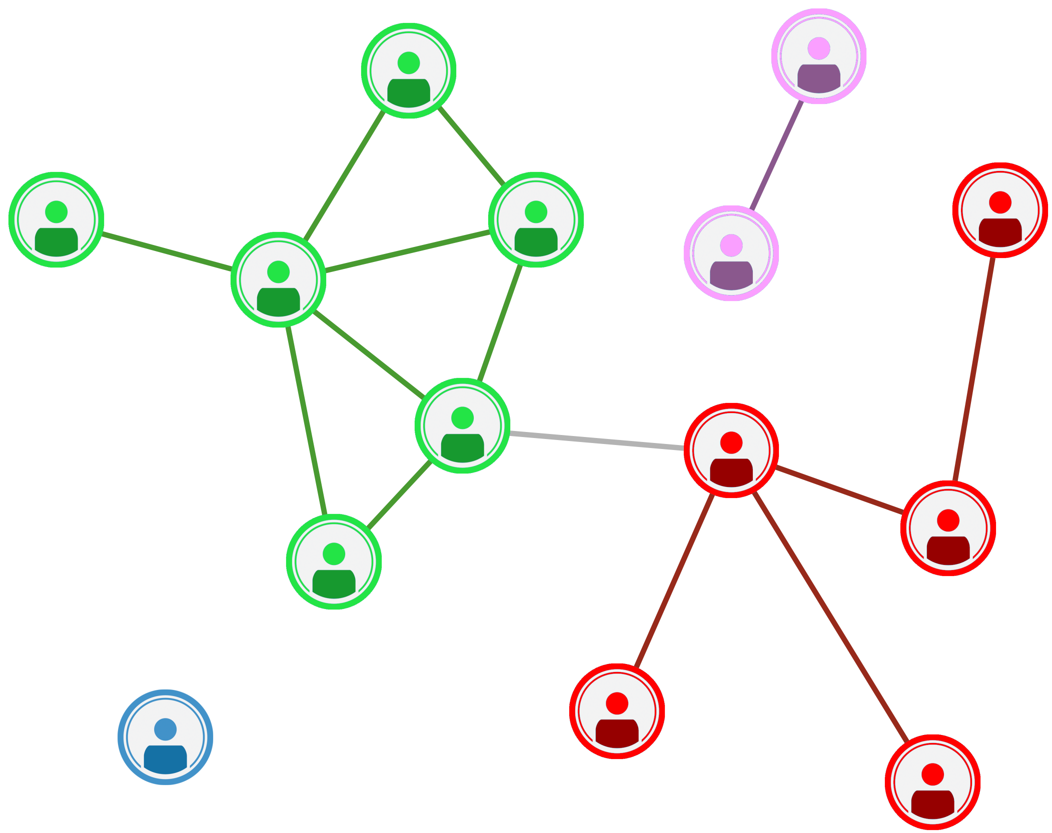
- ***Egocentric***
- *Sociocentric*



Persistence and Complex Networks

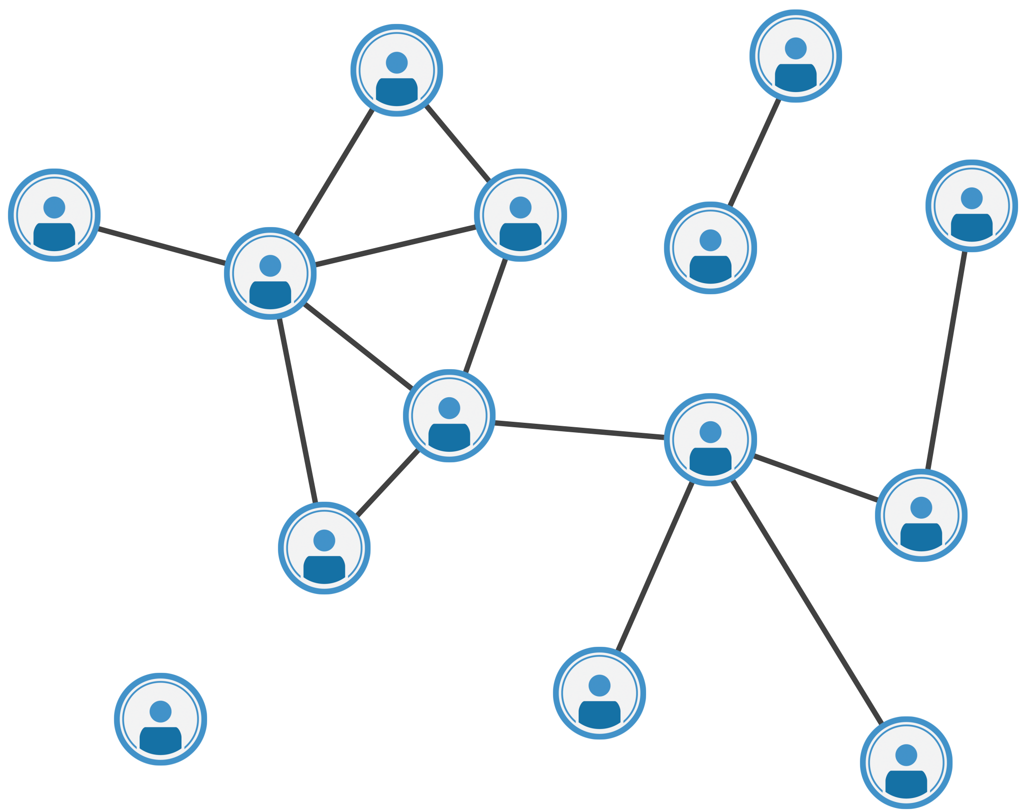
A Two-Level Analysis:

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



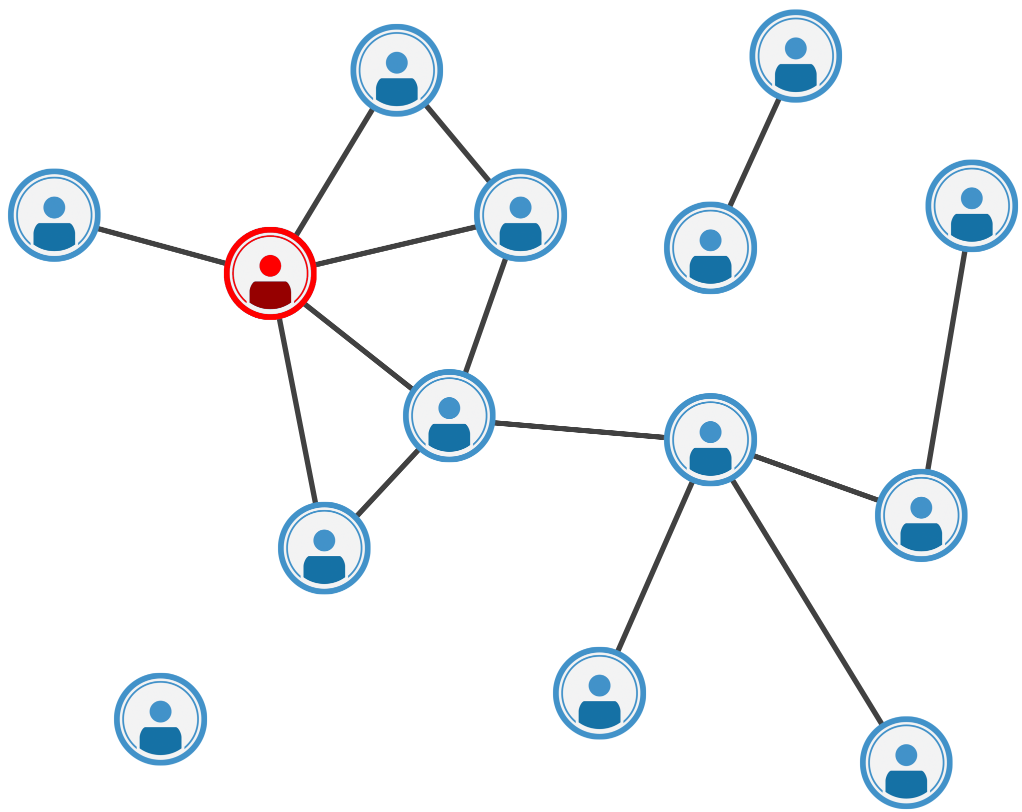
Persistence and Complex Networks

Who is the most important individual?



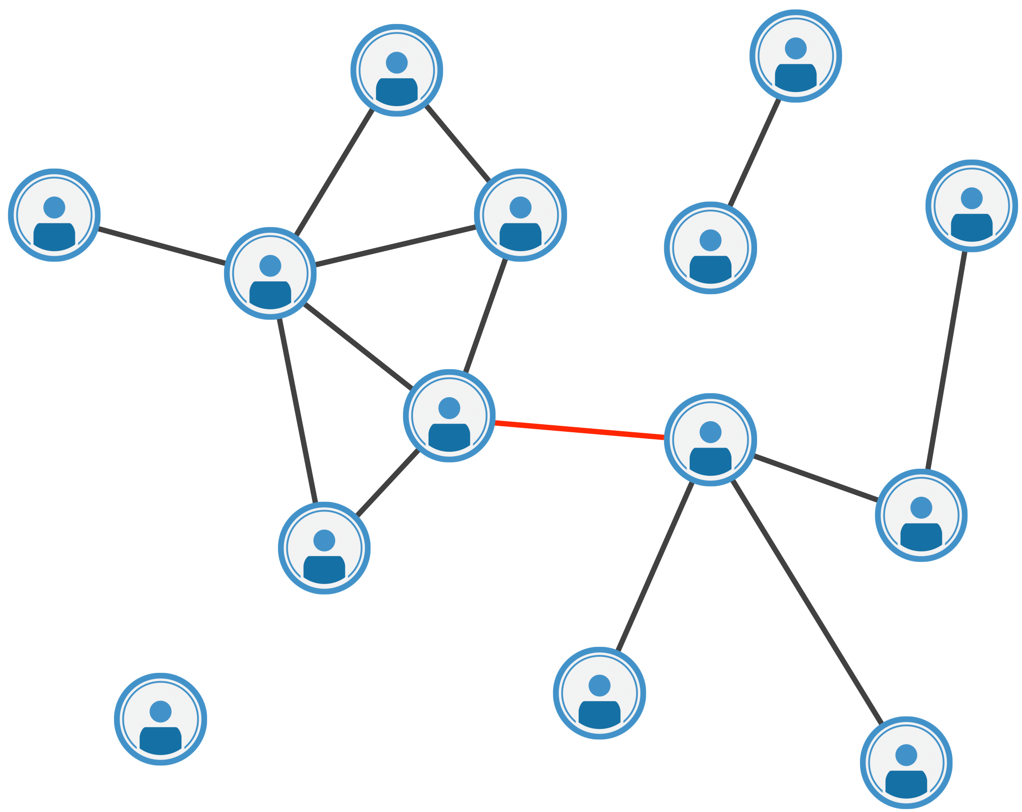
Persistence and Complex Networks

Who is the most important individual?



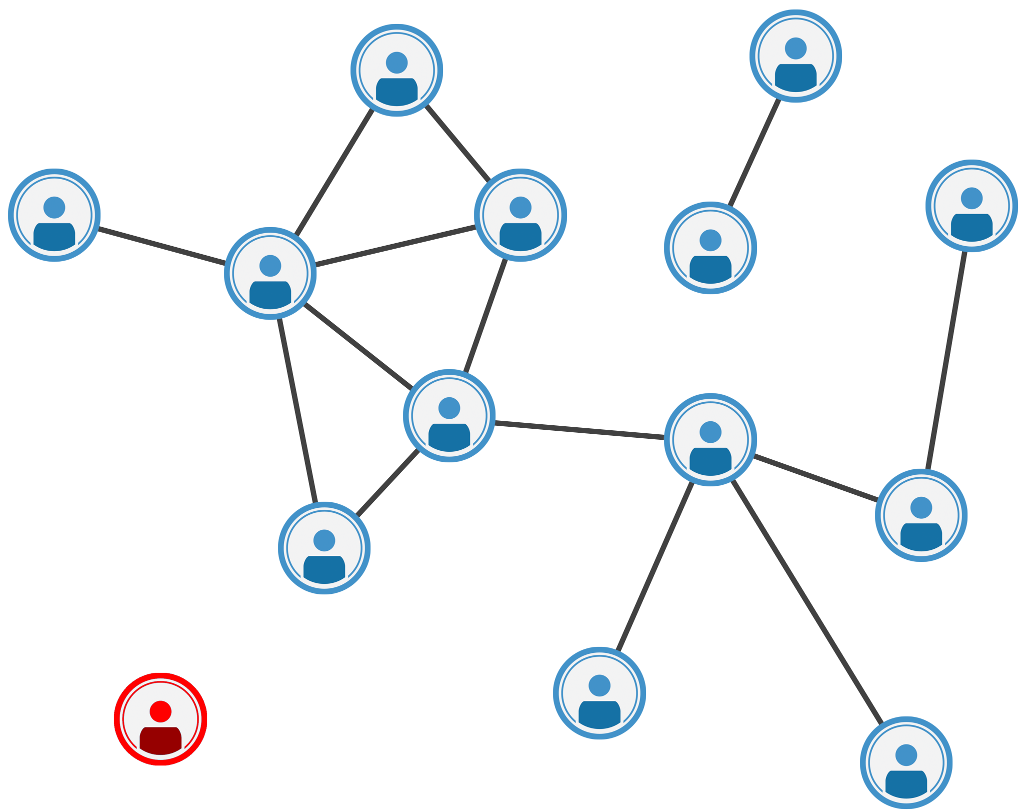
Persistence and Complex Networks

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:*

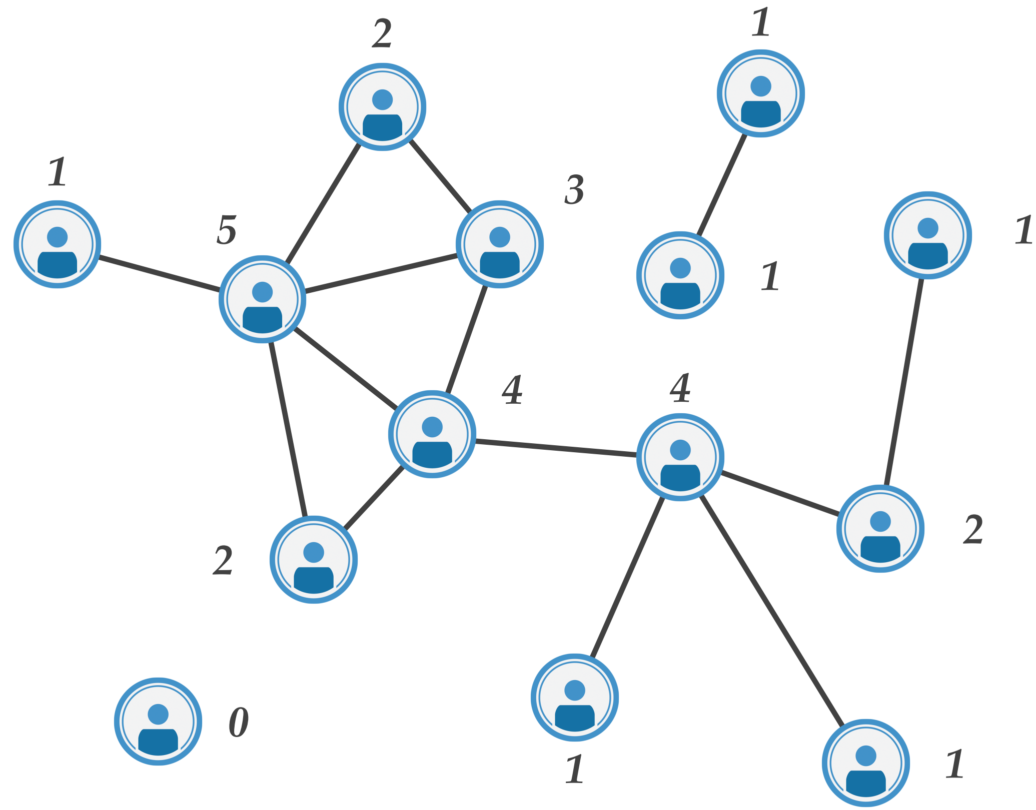
- ***Degree** centrality*
- ***Betweenness** centrality*
- ***Closeness** centrality*
- ***Eigenvector** centrality*
- ***Erdös** distance*

Persistence and Complex Networks

Degree Centrality:

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

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

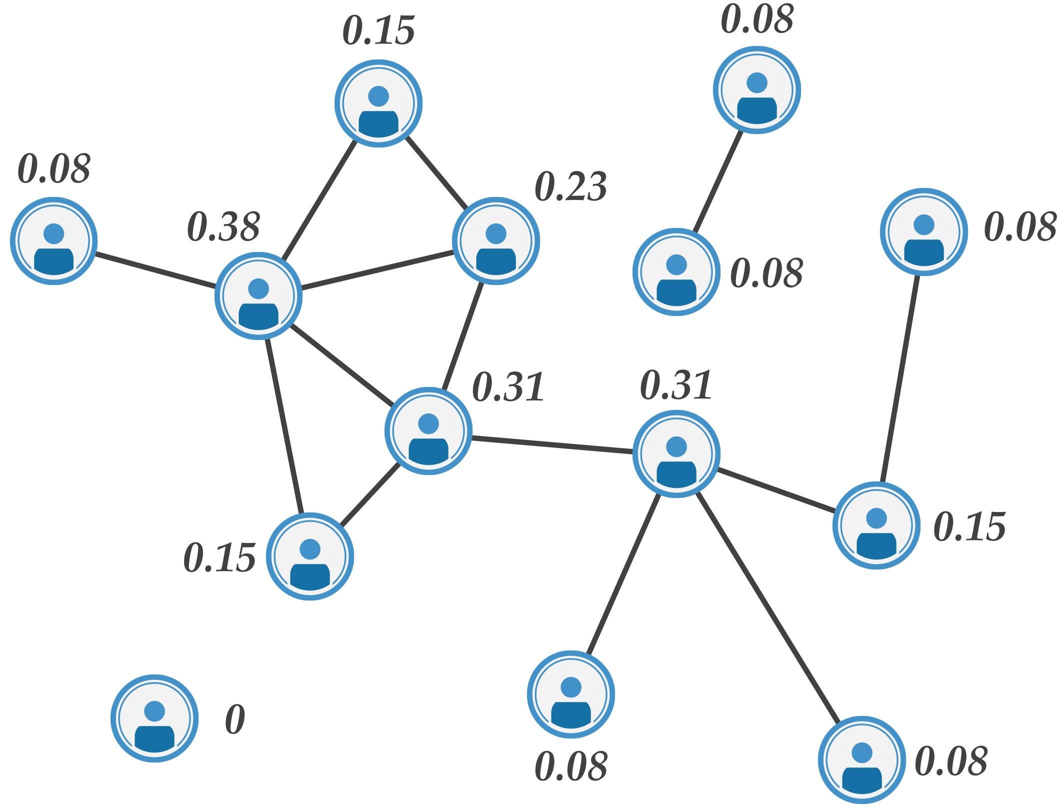


Persistence and Complex Networks

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

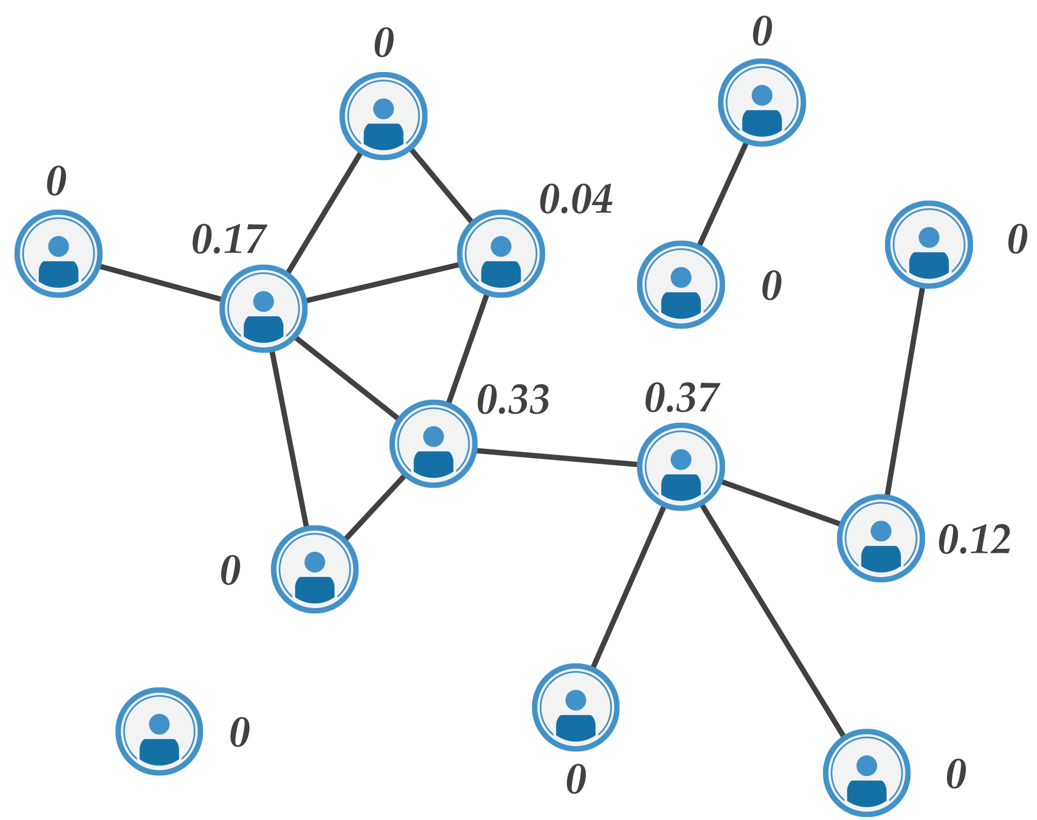
Betweenness Centrality:

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***

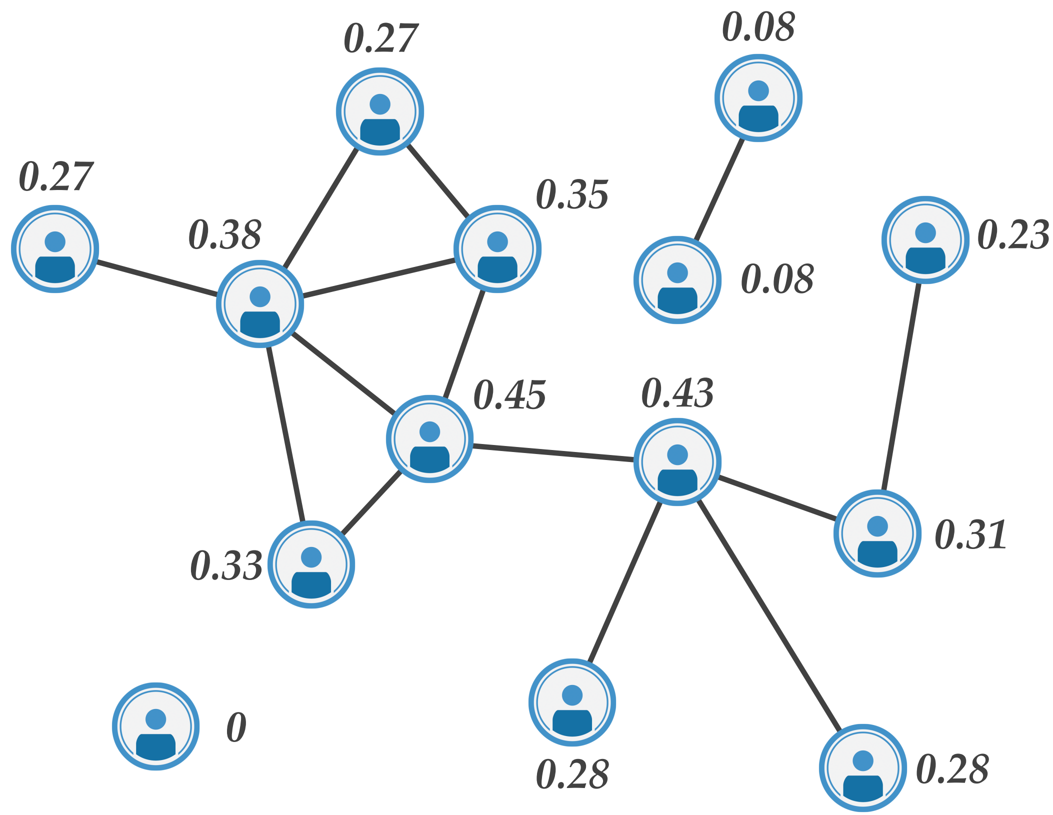


Persistence and Complex Networks

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),

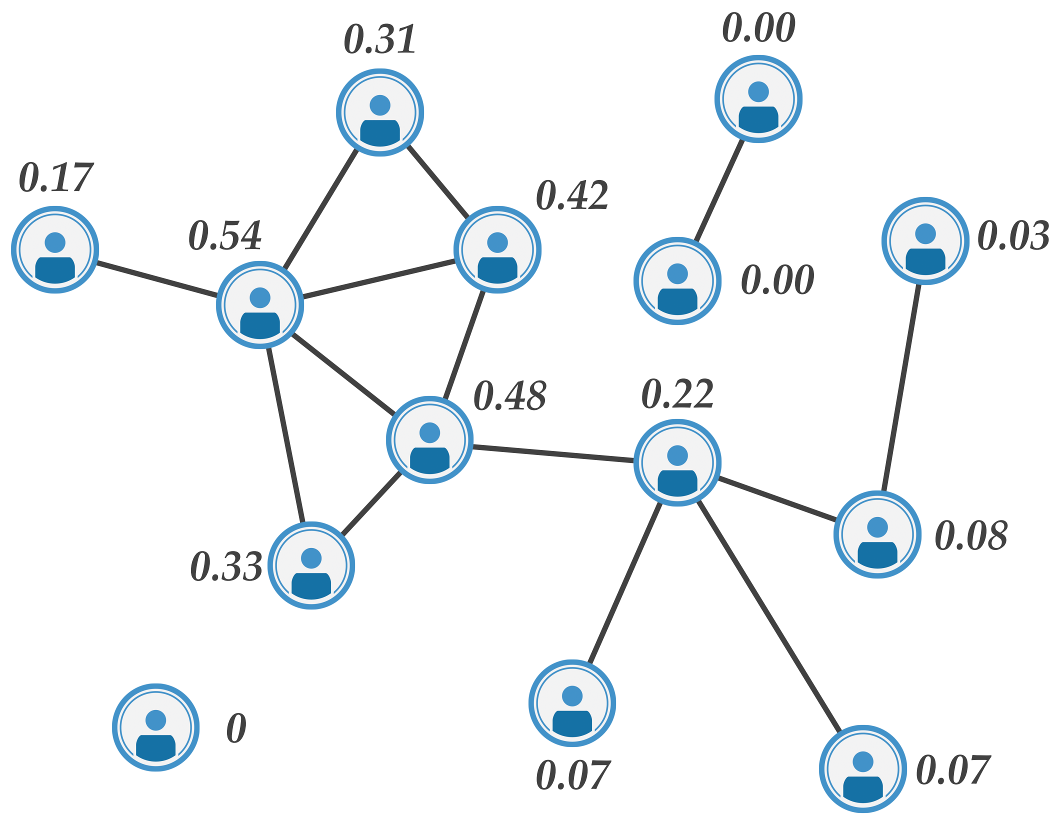
$$x_v := \frac{1}{\lambda} \sum_{u \in V} A_{uv}\, x_u$$

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

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



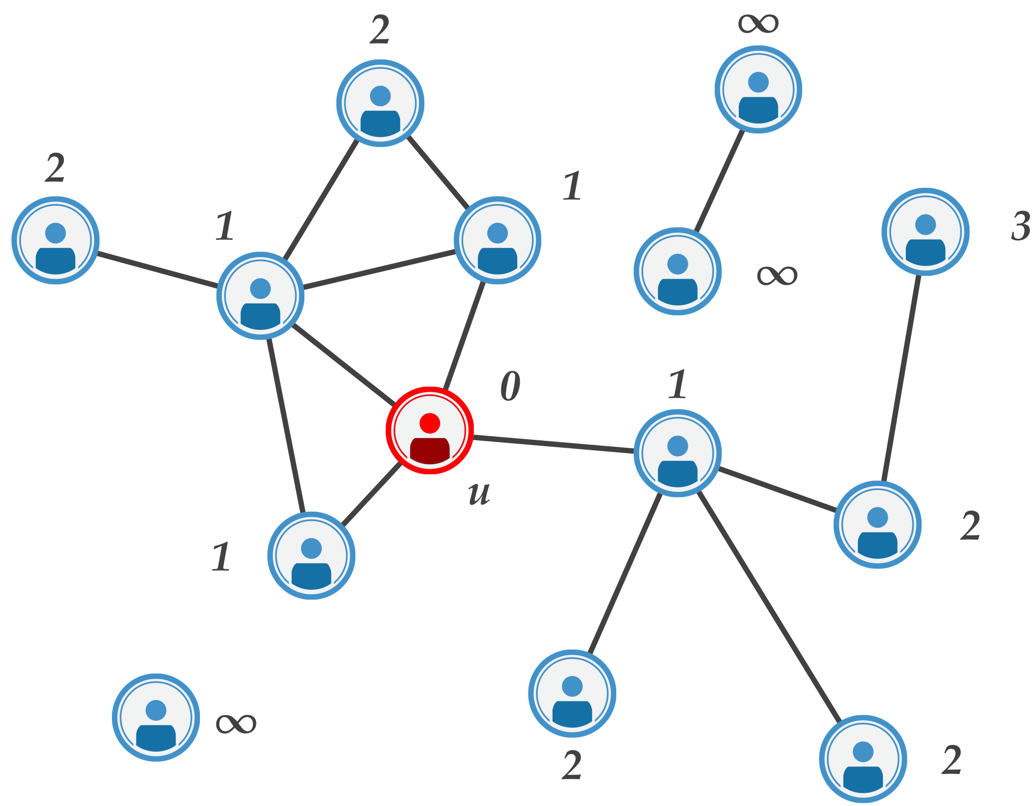
x* > *0 implies λ must be the largest eigenvalue of *A* and ***x*** the corresponding eigenvector

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*

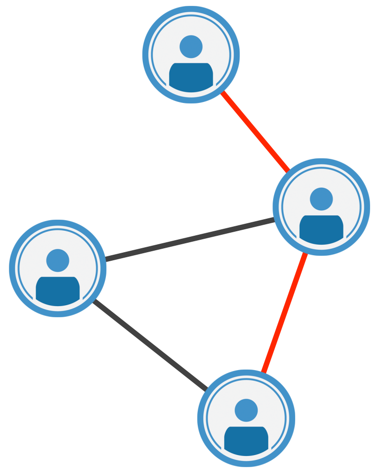
Persistence and Complex Networks

Structural Metrics:

How far are two individuals at most?

Diameter:

The longest shortest path between any two nodes



Diameter(G) = 2

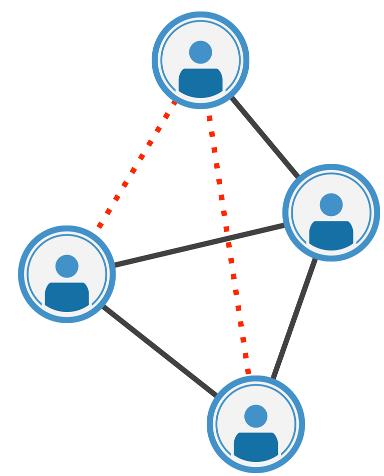
Persistence and Complex Networks

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

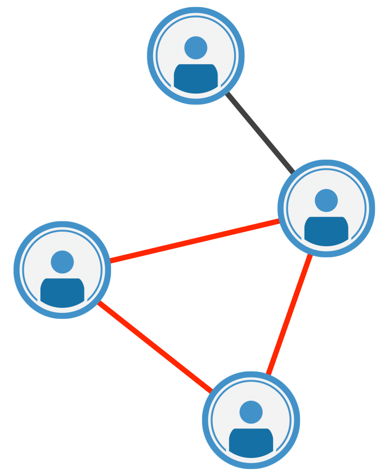
Persistence and Complex Networks

Structural Metrics:

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

Transitivity:

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



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*
  - ...

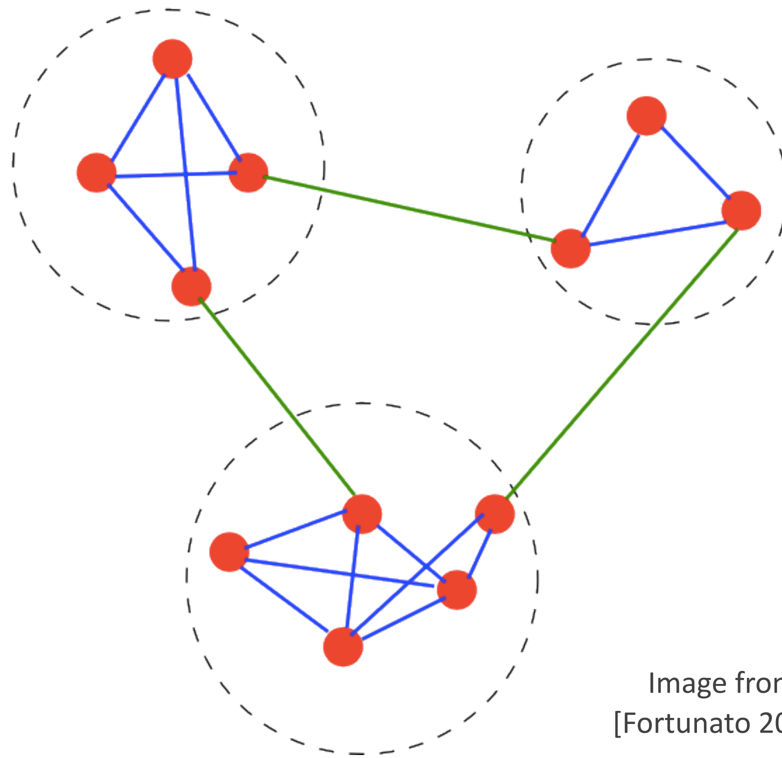


Image from [Fortunato 2009]

Persistence and Complex Networks

Community Decompositions:

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

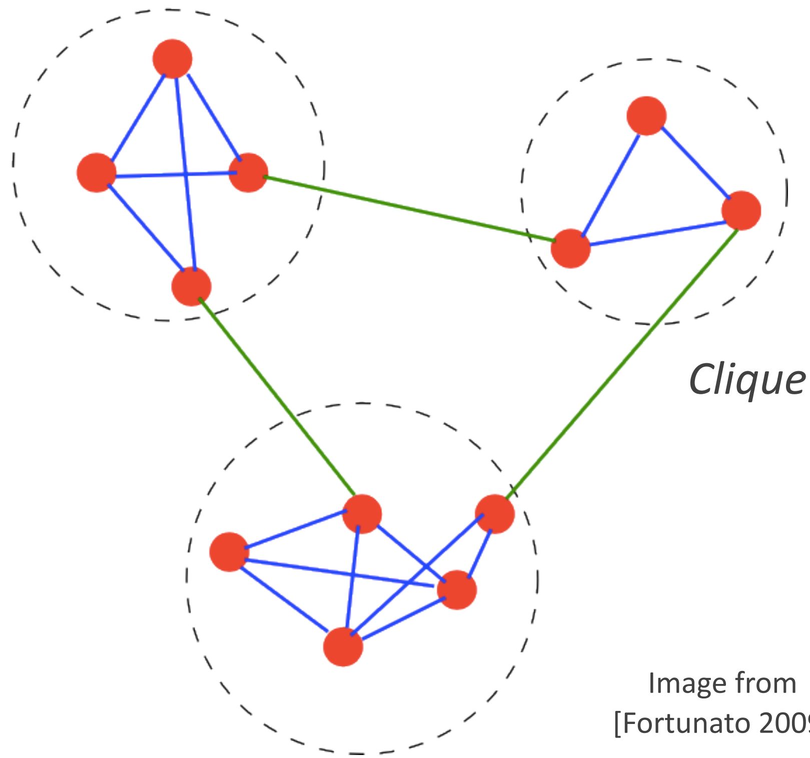


Image from [Fortunato 2009]

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*
  - ...

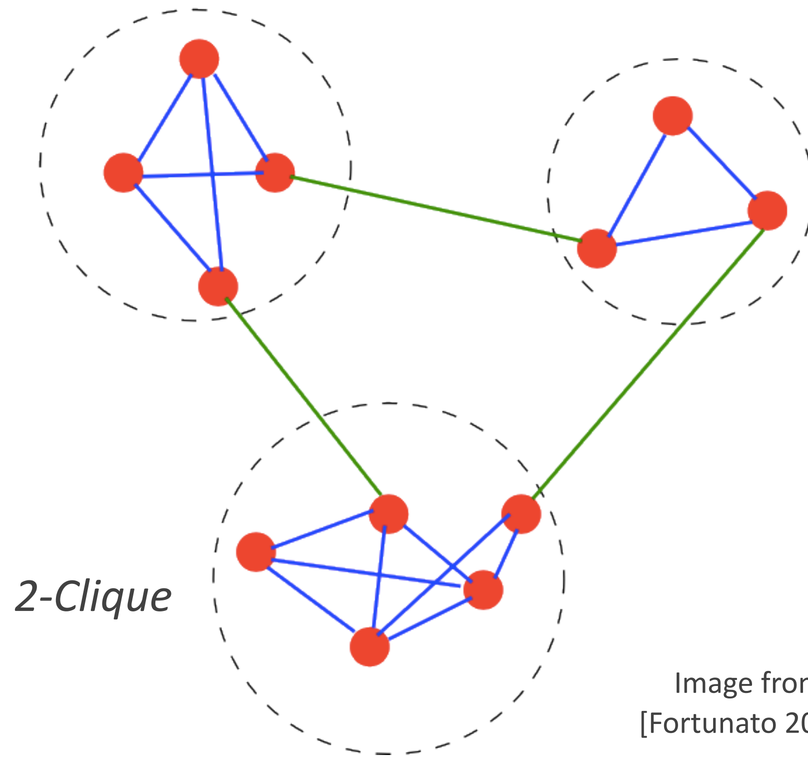


Image from
[Fortunato 2009]

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*
  - ...

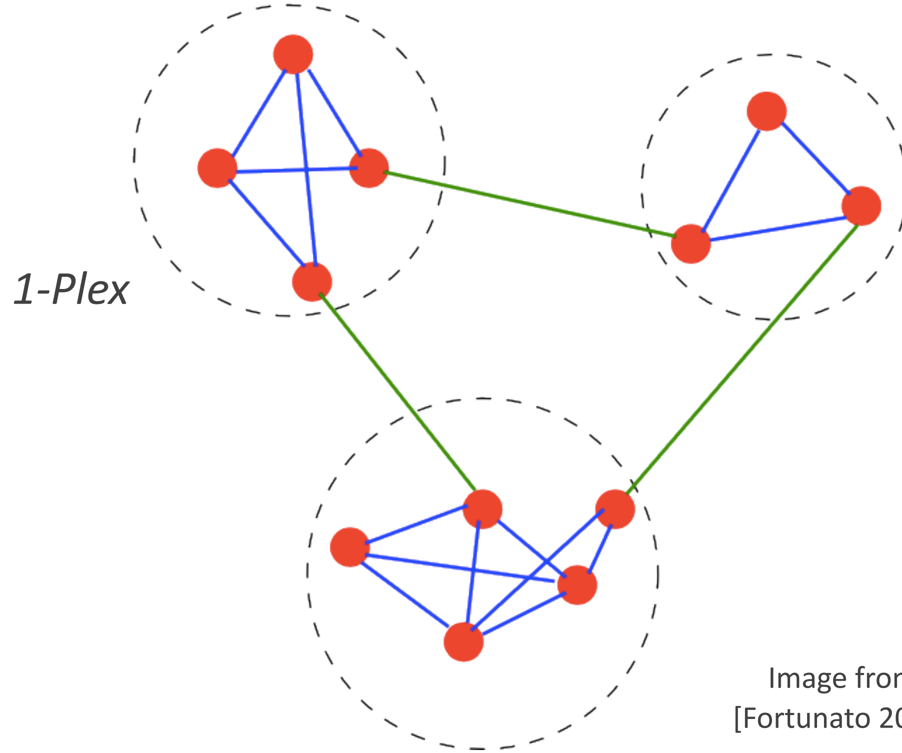


Image from [Fortunato 2009]

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)

Divisive (top-down)

approach based on

Centrality Measures

Atomic Communities

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***

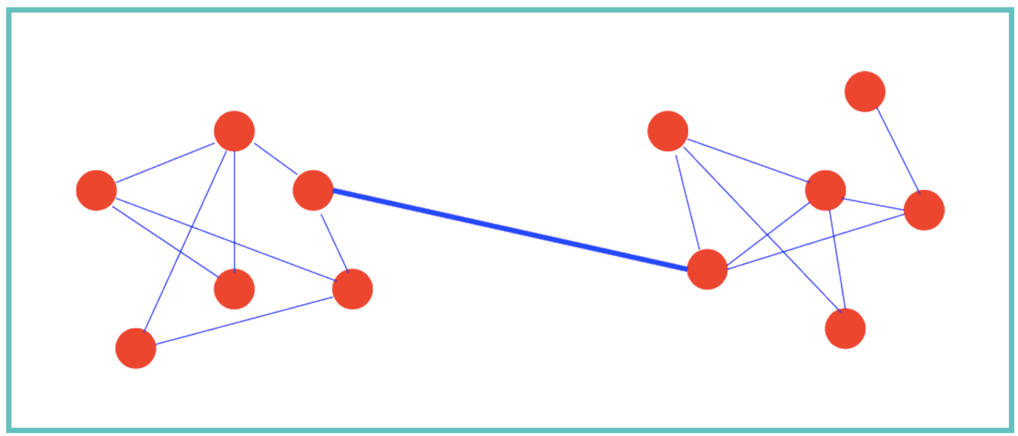


Image from [Fortunato 2009]

Persistence and Complex Networks

Clustering Techniques:

Agglomerative (bottom-up)
Divisive (top-down)

approach based on

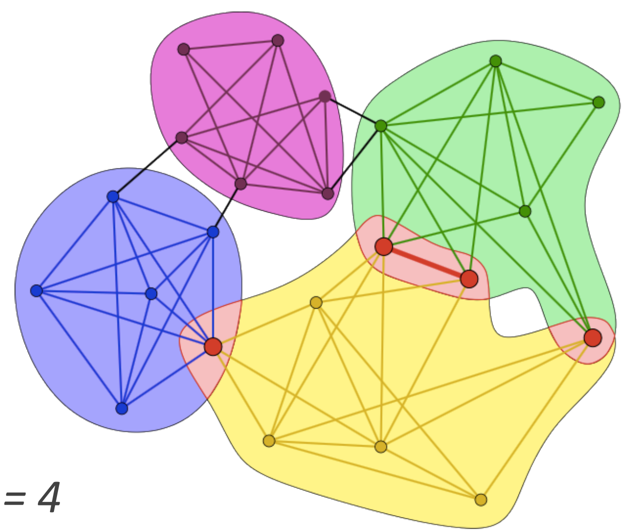
Centrality Measures
Atomic Communities
Quality Functions

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



k = 4

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***

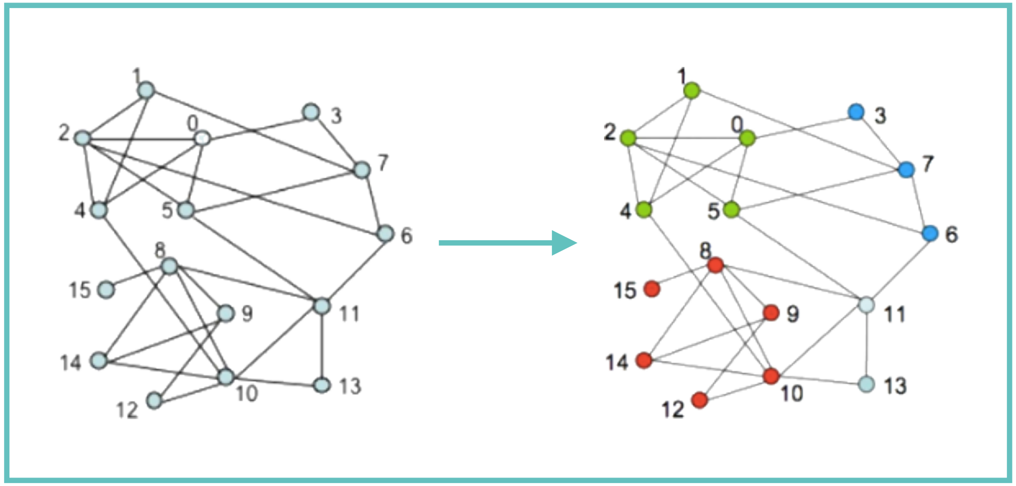


Image from [Blondel et al. 2008]

Persistence and Complex Networks

Several Application based on Persistent Homology:

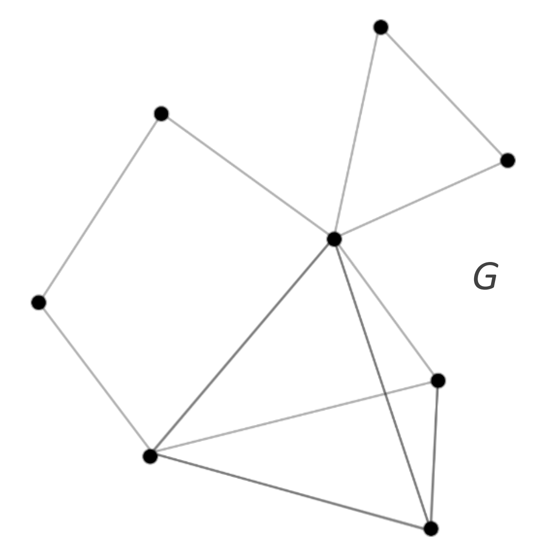
- ***Sensor*** Networks [De Silva 2013]
- ***Brain*** Networks [Lee et al. 2012]
- ***Collaborative/Co-occurence*** Networks [Carstens et al. 2013; Rieck et al. 2016]
- ***Geolocalized*** Networks [Fellegara et al. 2016]
- ...

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:

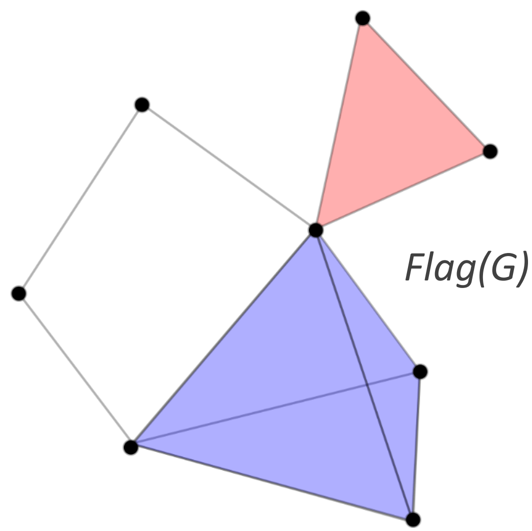
- ***Sensor*** Networks [De Silva 2013]
- ***Brain*** Networks [Lee et al. 2012]
- ***Collaborative/Co-occurence*** Networks [Carstens et al. 2013; Rieck et al. 2016]
- ***Geolocalized*** Networks [Fellegara et al. 2016]
- …

Simplicial Complex Representation:

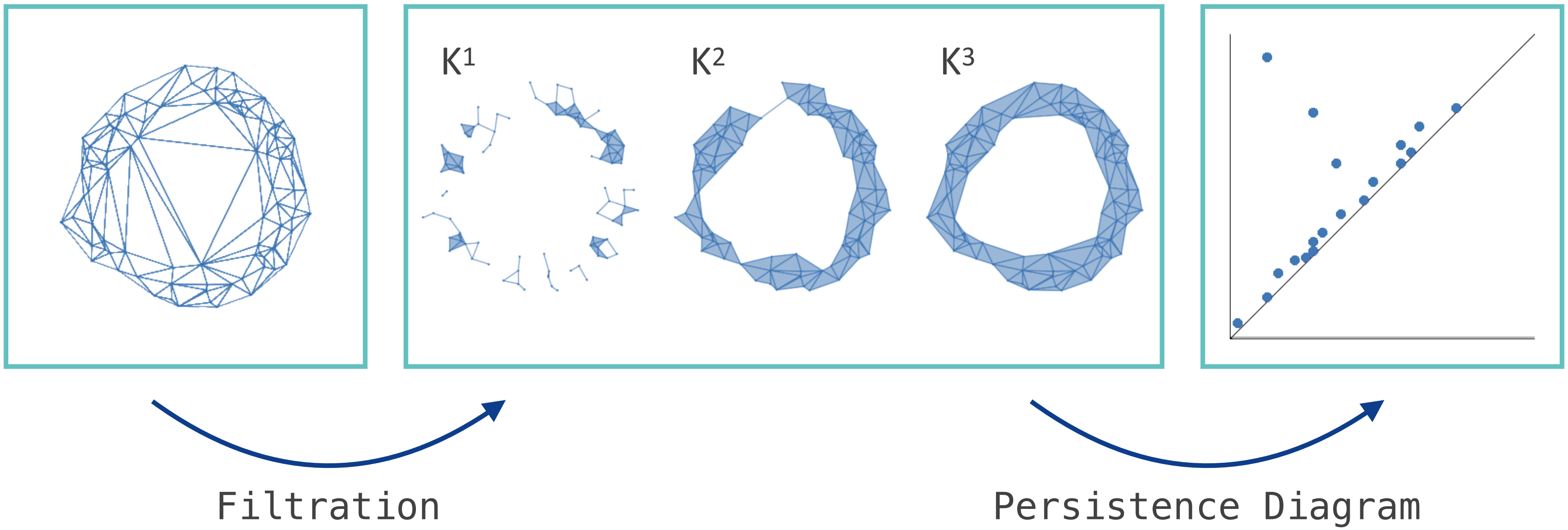
A network is represented through:

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

Simplices of Flag(G) ⟷ *Cliques of G*

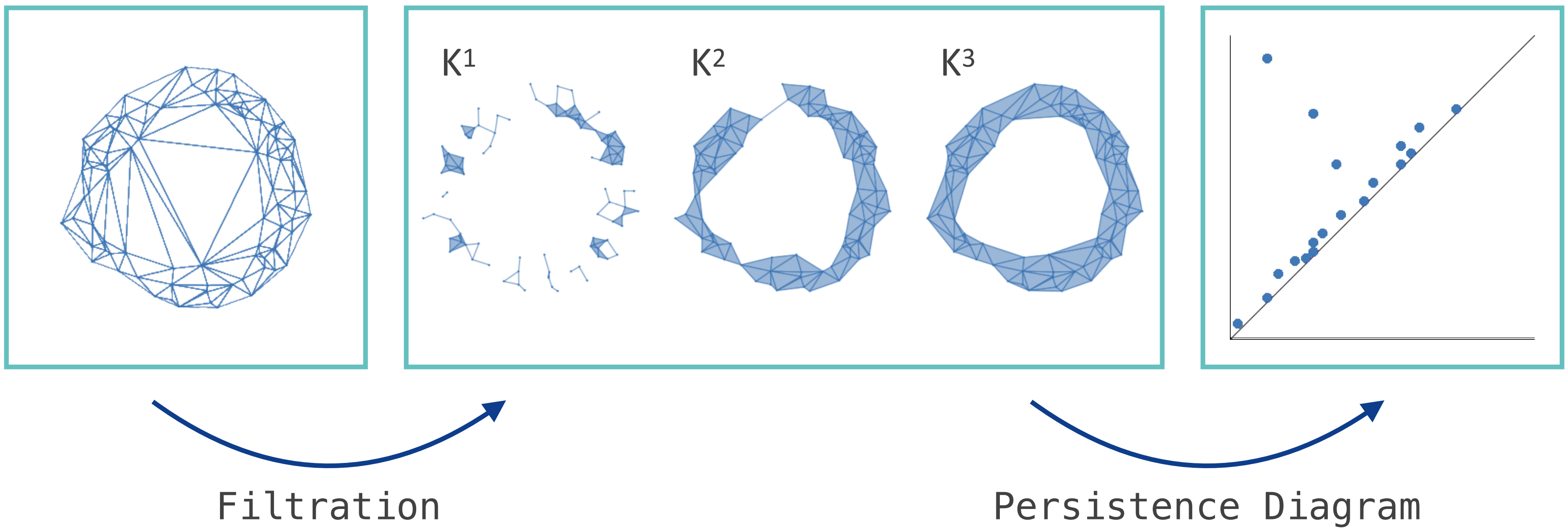


Persistence and Complex Networks



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

Persistence and Complex Networks

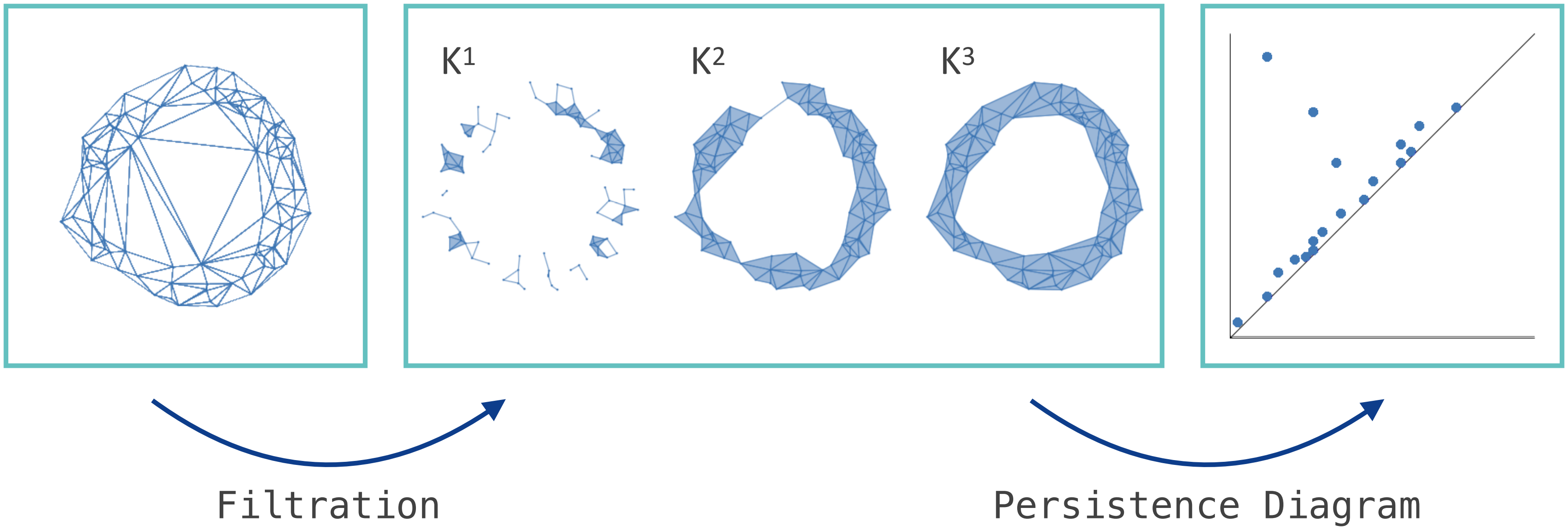


Topological summaries have proven to be particularly effective to ***distinguish networks***

but

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

Persistence and Complex Networks



Can we …

- visualize/localize the homological information through a graph?
- study the persistence of something different than homological cycles?

Persistence and Complex Networks

- *A Primer on Complex Networks*
- ***Homological Scaffolds***
- *Clique Community Persistence*

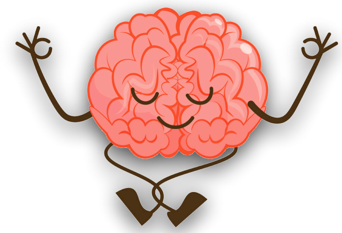
Persistence and Complex Networks

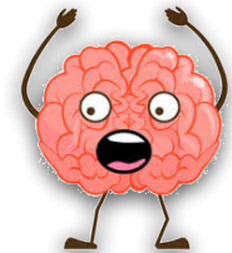
Dataset:

A collection of ***30 weighted graphs*** derived from ***fMRI*** (functional magnetic resonance imaging) obtained by scanning 15 different subjects

For each graph,

Nodes	$\Leftrightarrow$	***169** (Sub)Cortical **Brain Regions***
Arcs Weights	$\Leftrightarrow$	*(Inverse of) **Partial Correlations***





For each subject, 2 graphs obtained on 2 separate occasions, 14 days apart:

- ***Placebo*** (10 ml saline, intravenous injection) in one case
- ***Psilocybin*** (2 mg dissolved in 10 ml saline) in the other one

Goal:

Spot the differences between the two situations

Persistence and Complex Networks

Results: Applying the presented TDA pipeline, one obtains:

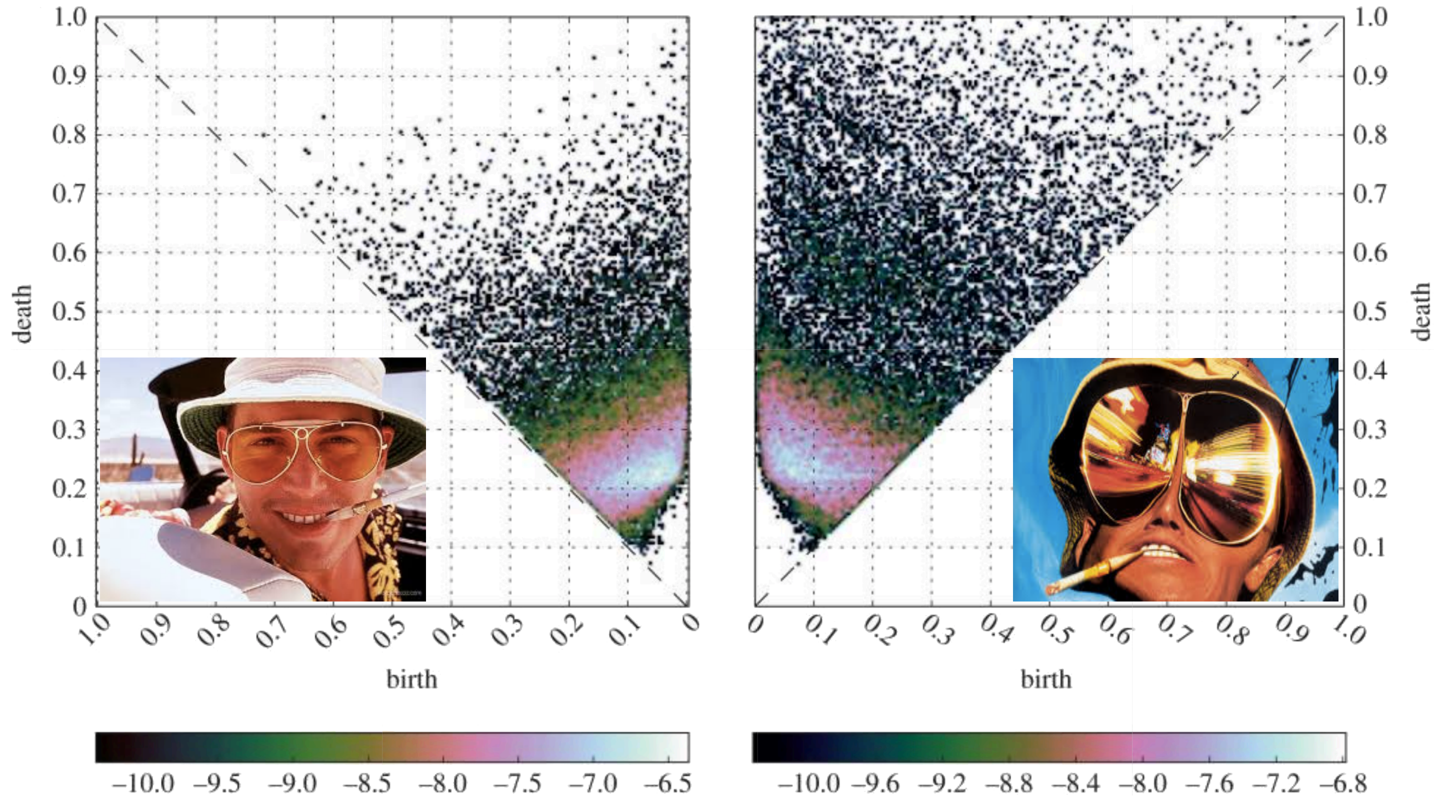


Image from [Petri et al. 2014]

(log-)Probability densities of H_1 for the placebo (left) and the psilocybin (right) groups

Persistence and Complex Networks

Results: Applying the presented TDA pipeline, one obtains:

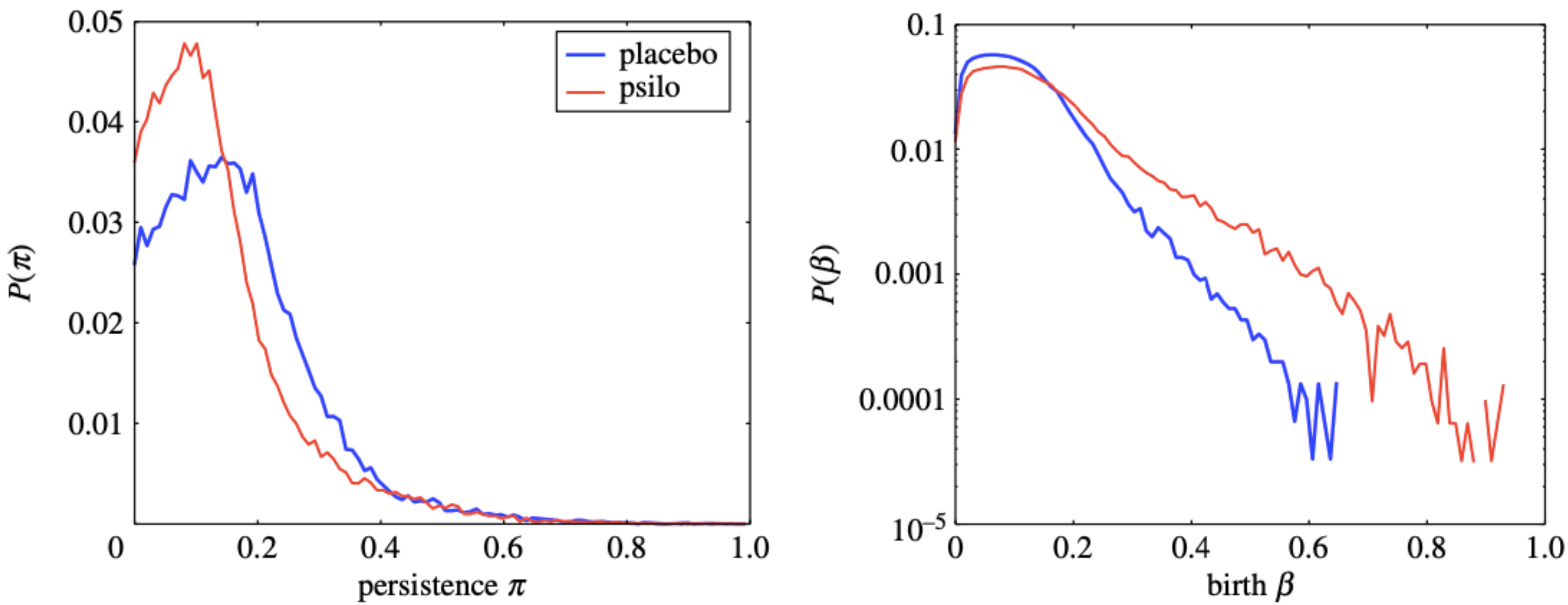


Image from [Petri et al. 2014]

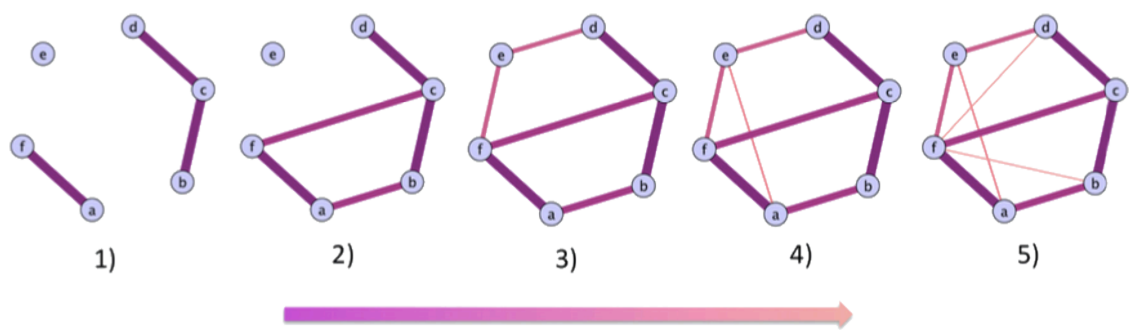
Persistence and birth distributions of H_1 for the placebo (blue) and the psilocybin (red) groups

Persistence and Complex Networks

Homological Scaffolds: *How to visualize/localize the homological information?*

Let $g_1, g_2, \ldots, g_m$ be the representative cycles of H_1 occurring along the filtration of a weighted graph $G = (V, E, w: E \longrightarrow \mathbb{R})$, the ***frequency homological scaffold*** is the graph

$$H^f_G = (V, E, w^f: E \longrightarrow \mathbb{R})$$



defined by

$$w^f(e) = \#\{i \,|\, e \in g_i\}$$

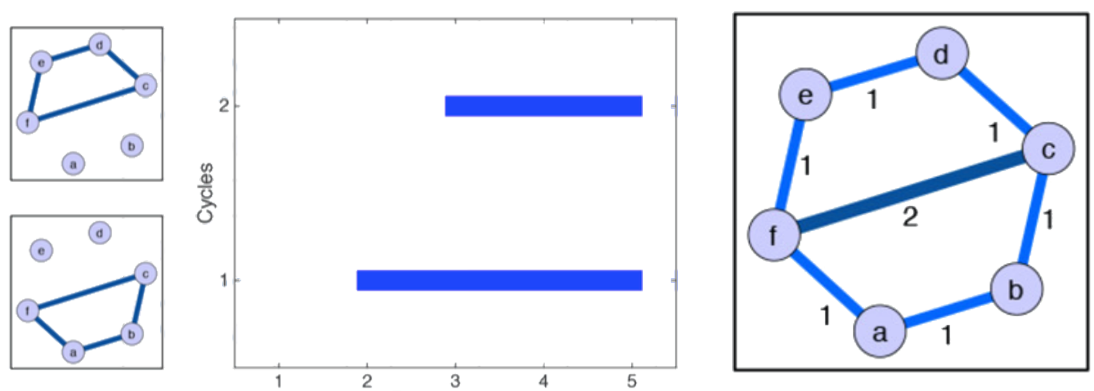


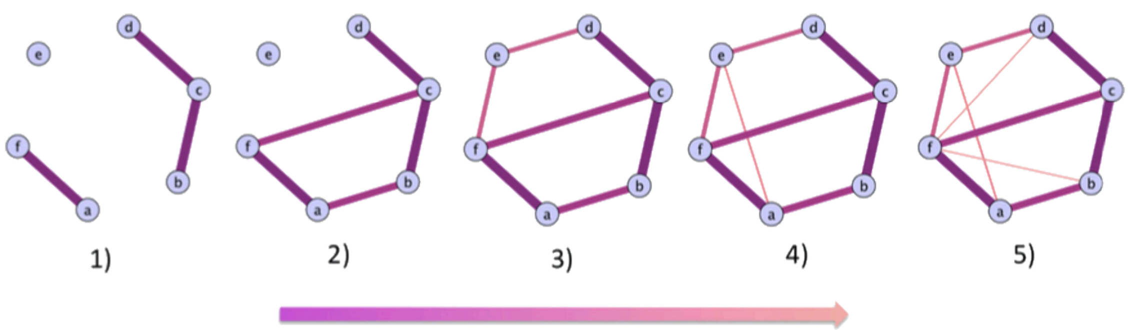
Image from [Lord et al. 2016]

Persistence and Complex Networks

Homological Scaffolds: *How to visualize/localize the homological information?*

Let $g_1, g_2, \ldots, g_m$ be the representative cycles of H_1 occurring along the filtration of a weighted graph $G = (V, E, w: E \longrightarrow \mathbb{R})$, the ***persistence homological scaffold*** is the graph

$$H^p{}_G = (V, E, w^p: E \longrightarrow \mathbb{R})$$



defined by

$$w^p(e) = \sum_{i \,|\, e \in g_i} \pi_{g_i}$$

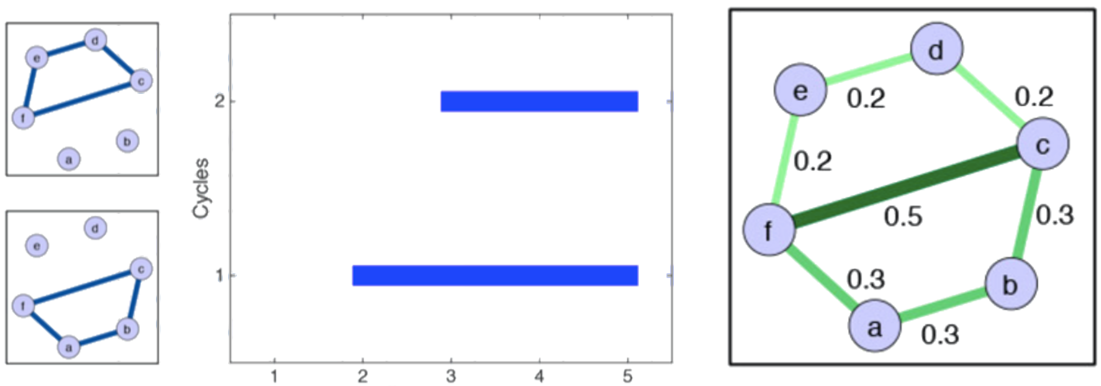


Image from [Lord et al. 2016]

Persistence and Complex Networks

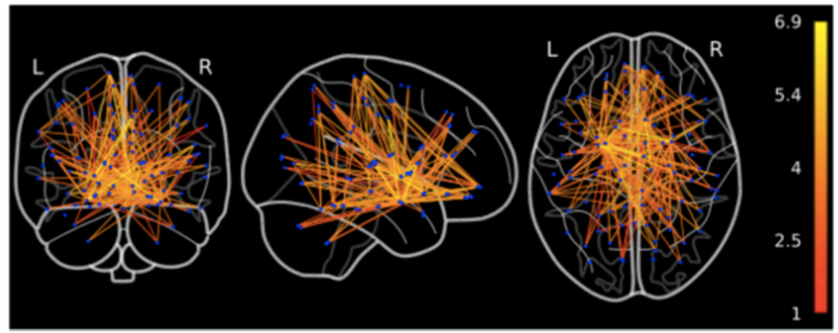
Homological Scaffolds:

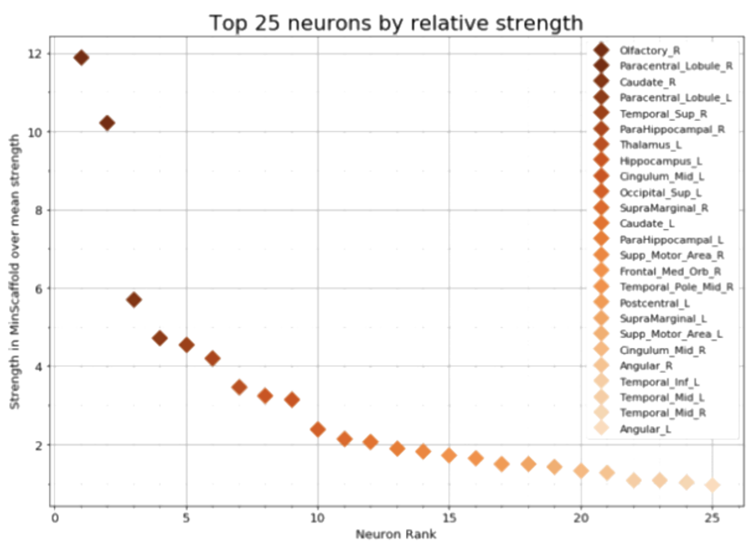
Image from [Petri et al. 2014]

Persistence homological scaffolds for the placebo (left) and the psilocybin (right) groups

Persistence and Complex Networks

Homological Scaffolds:





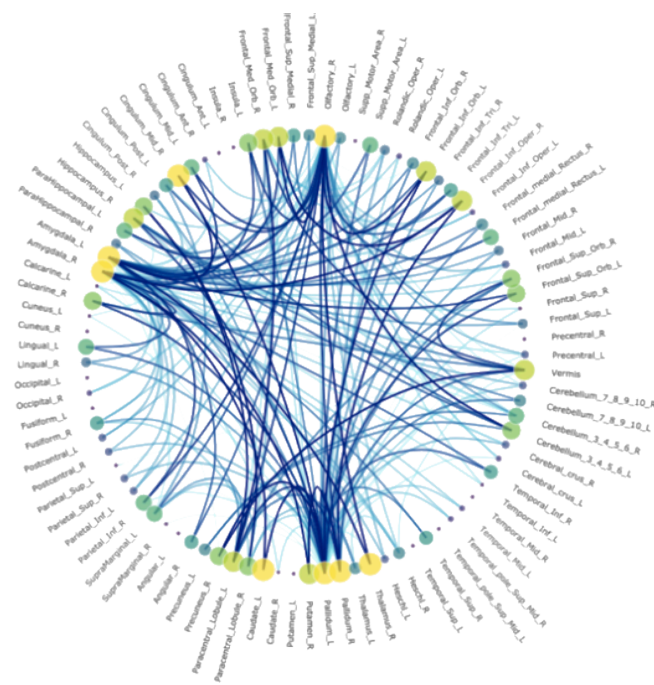


Image from [Guerra et al. 2021]

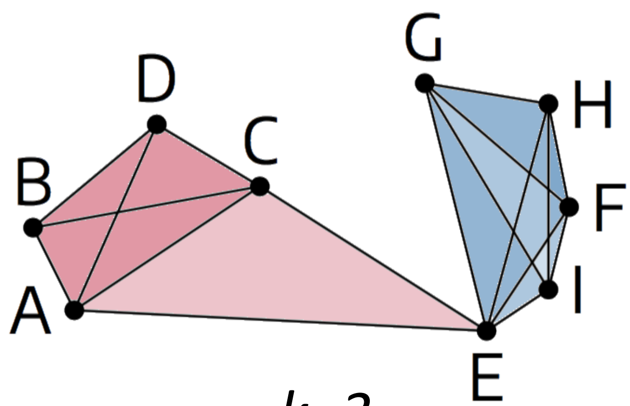
Persistence and Complex Networks

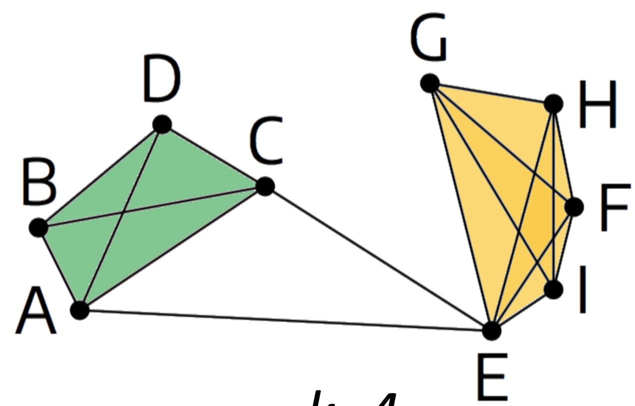
- *A Primer on Complex Networks*
- *Homological Scaffolds*
- ***Clique Community Persistence***

Persistence and Complex Networks

k-Clique Community:

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



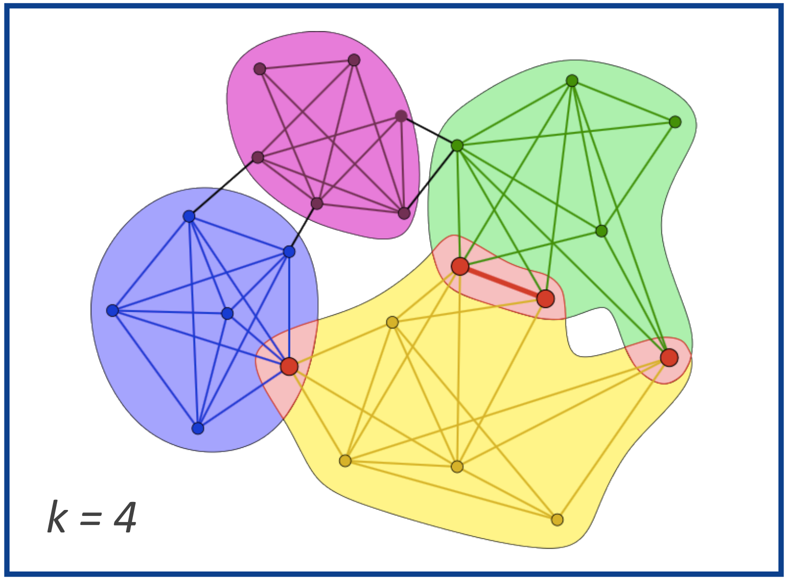


k-Adjacency:

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

Persistence and Complex Networks

k-Clique Community Decomposition:

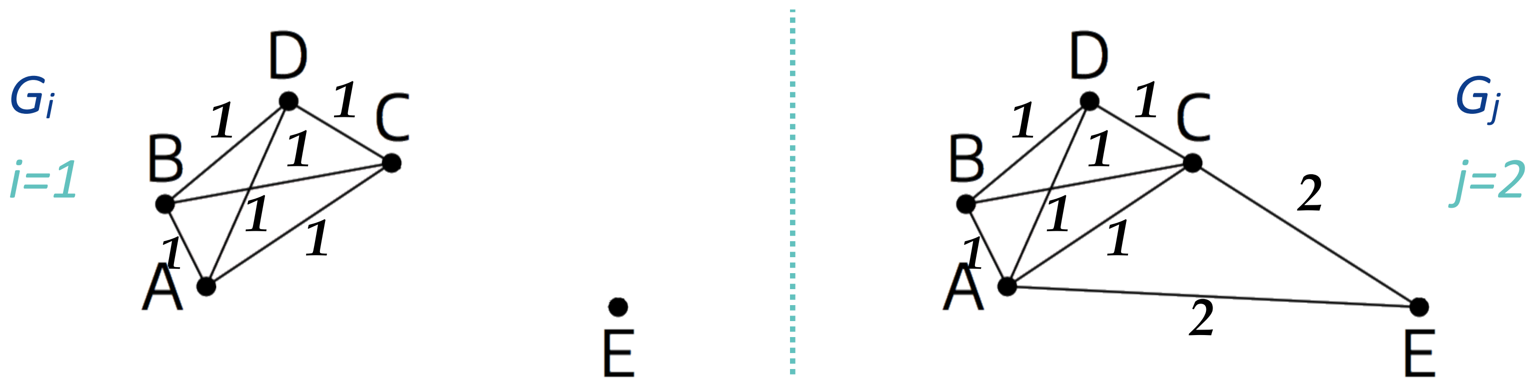


- Reveal ***highly connected*** communities
- Allow ***overlaps***
- Have a ***hierarchical structure***

Persistence and Complex Networks

Clique Communities and Weighted Networks:

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

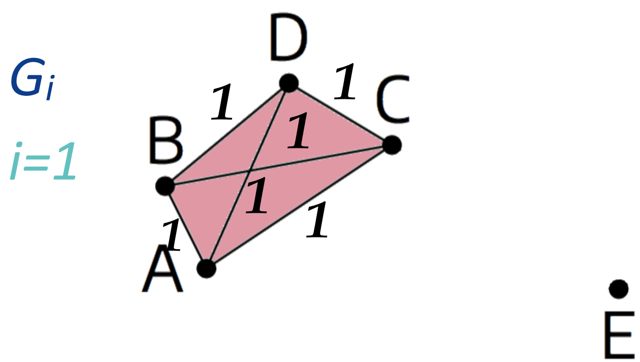


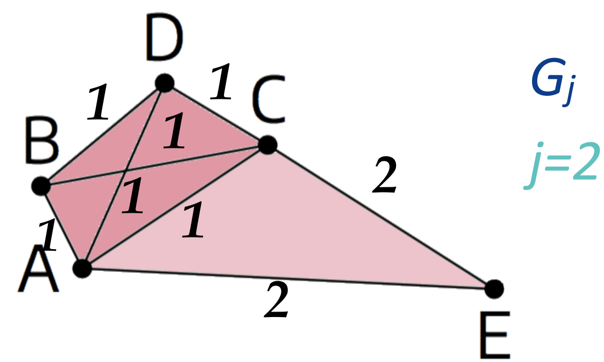
G_i is contained in G_j

Persistence and Complex Networks

Clique Communities and Weighted Networks:

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





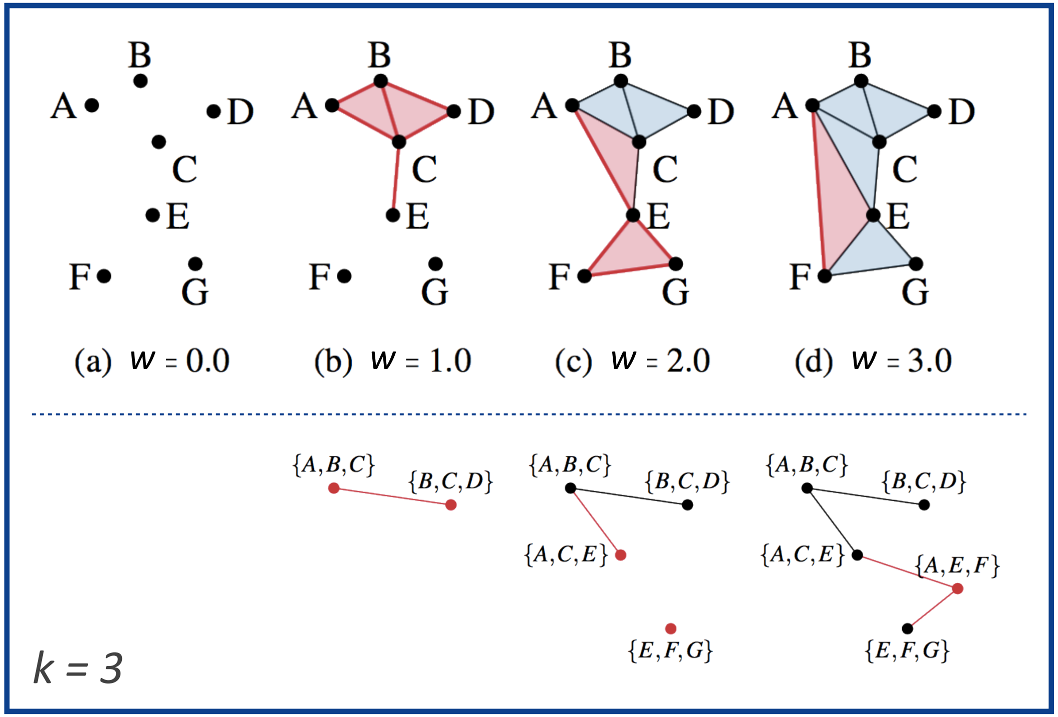
Each k-clique community of G_i is ***contained*** in ***exactly one*** k-clique community of G_j

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:

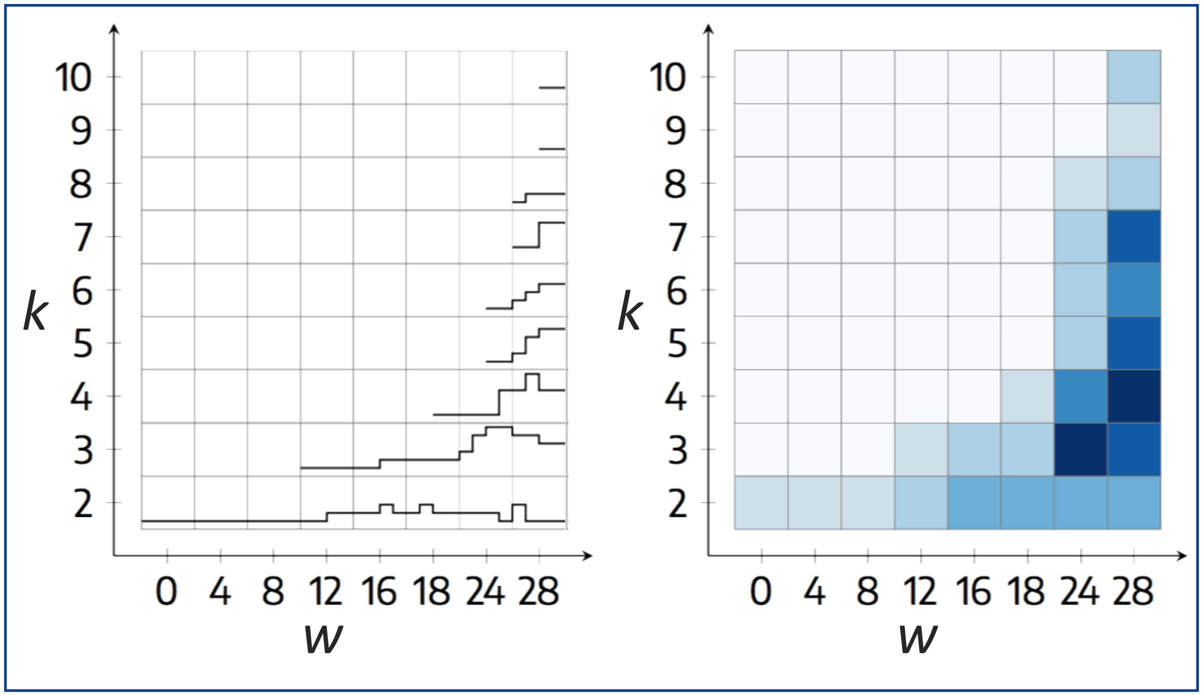
- ***Network Comparison***
  - ***Comparison Measures***
    - ***Persistence Indicator Function (PIF)***
    - ***PIF-based distance***
  - Clique Community ***Centrality Measure***
- ***Single Network Analysis***
  - ***Interactive Visualization Tool*** based on Nested Graphs

Persistence and Complex Networks

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

assigning:

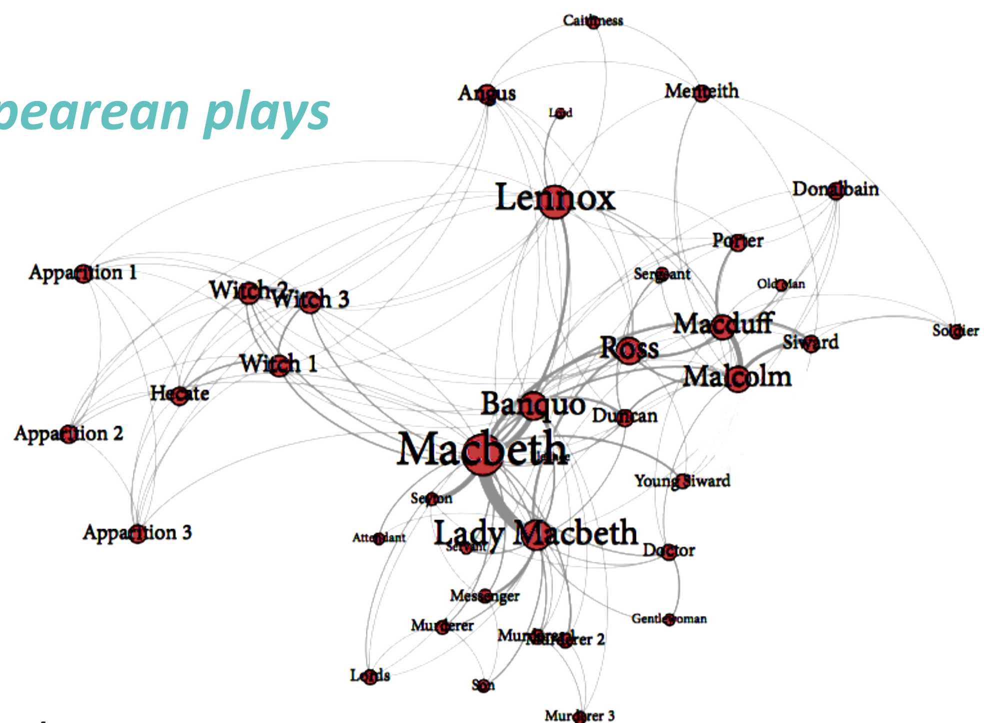
w $\longmapsto$ ***# k-cliques communities "alive" at threshold w***



Persistence and Complex Networks

Persistence Indicator Function:

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

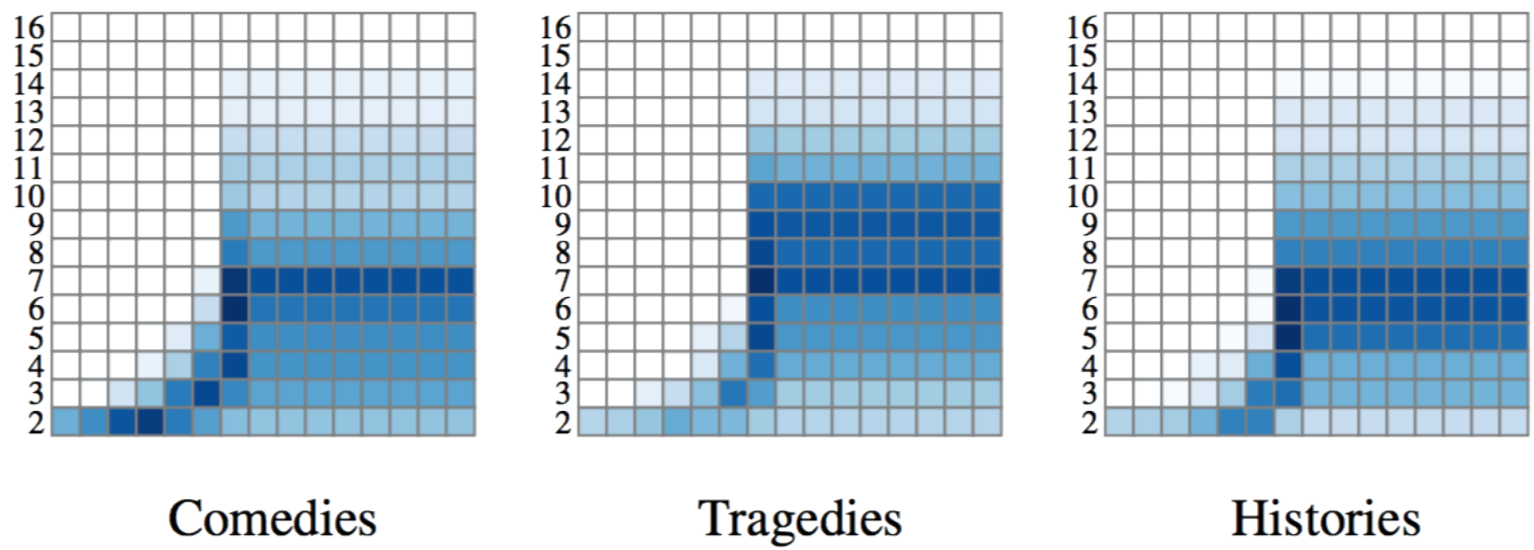


- In each network:
  - ***nodes*** ⟷ ***characters*** *of the play*
  - ***edges*** ⟷ *characters appearing in the* ***same scene***
  - ***edge weight*** ⟷ *inverse of the* ***number of interactions***

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:*

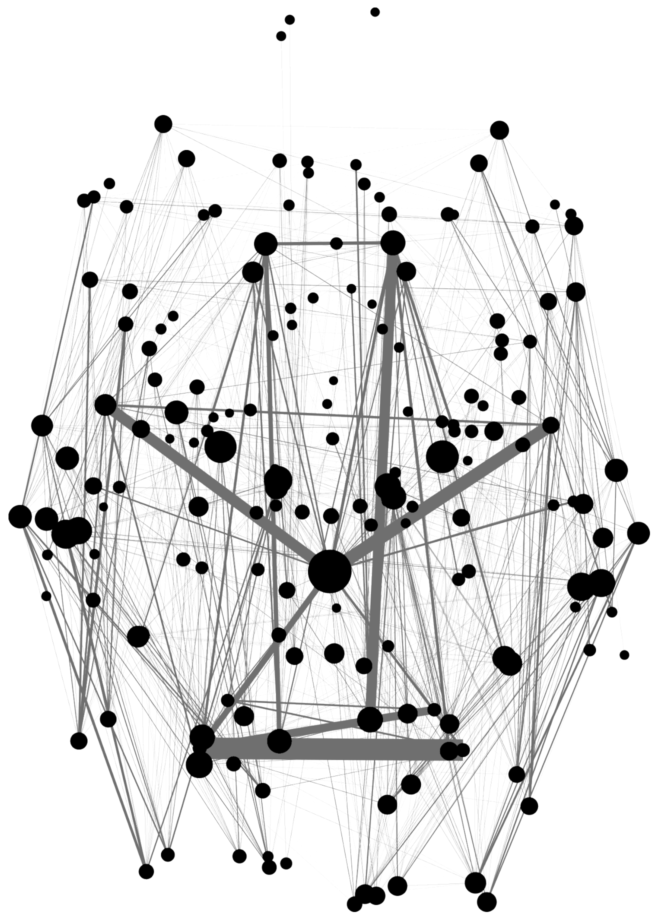
$$dist(f,g) = \left(\int_{\mathbb{R}} |f(x) - g(x)|^p \, dx \right)^{\frac{1}{p}}$$

- 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*

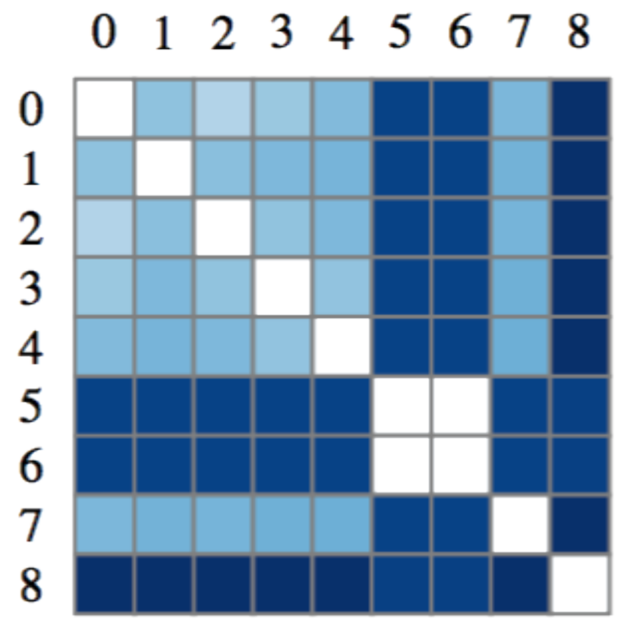


- In each network:
  - ***nodes*** $\leftrightarrow$ ***brain areas***
  - ***edges*** $\leftrightarrow$ ***fibers connecting different areas***

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:

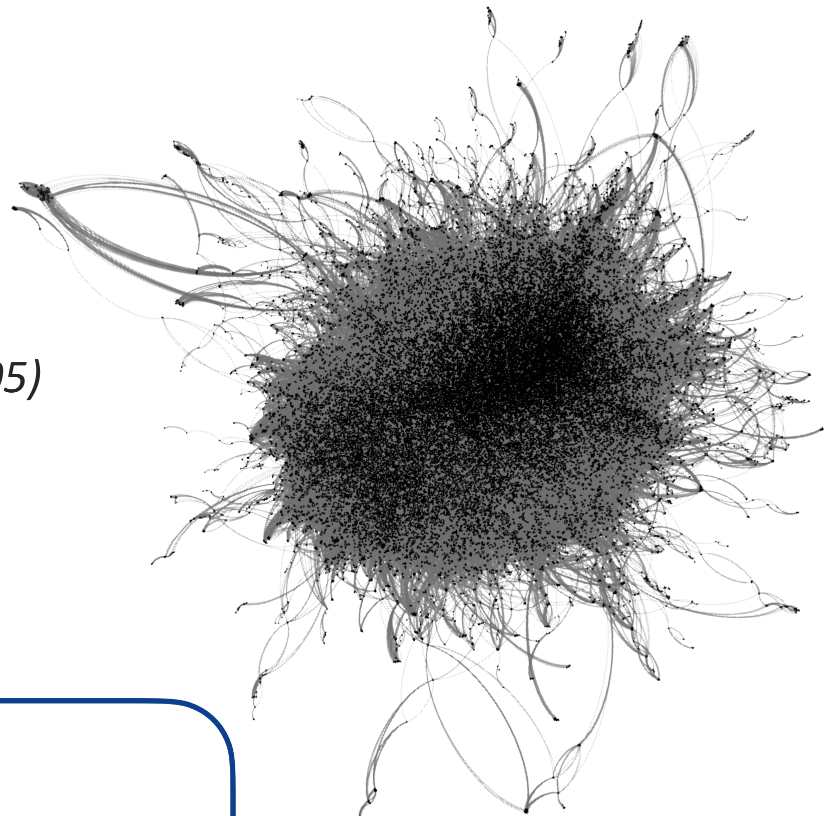
- C is any clique community containing *v*
- ***pers(C)*** is the “lifespan” of C

*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*

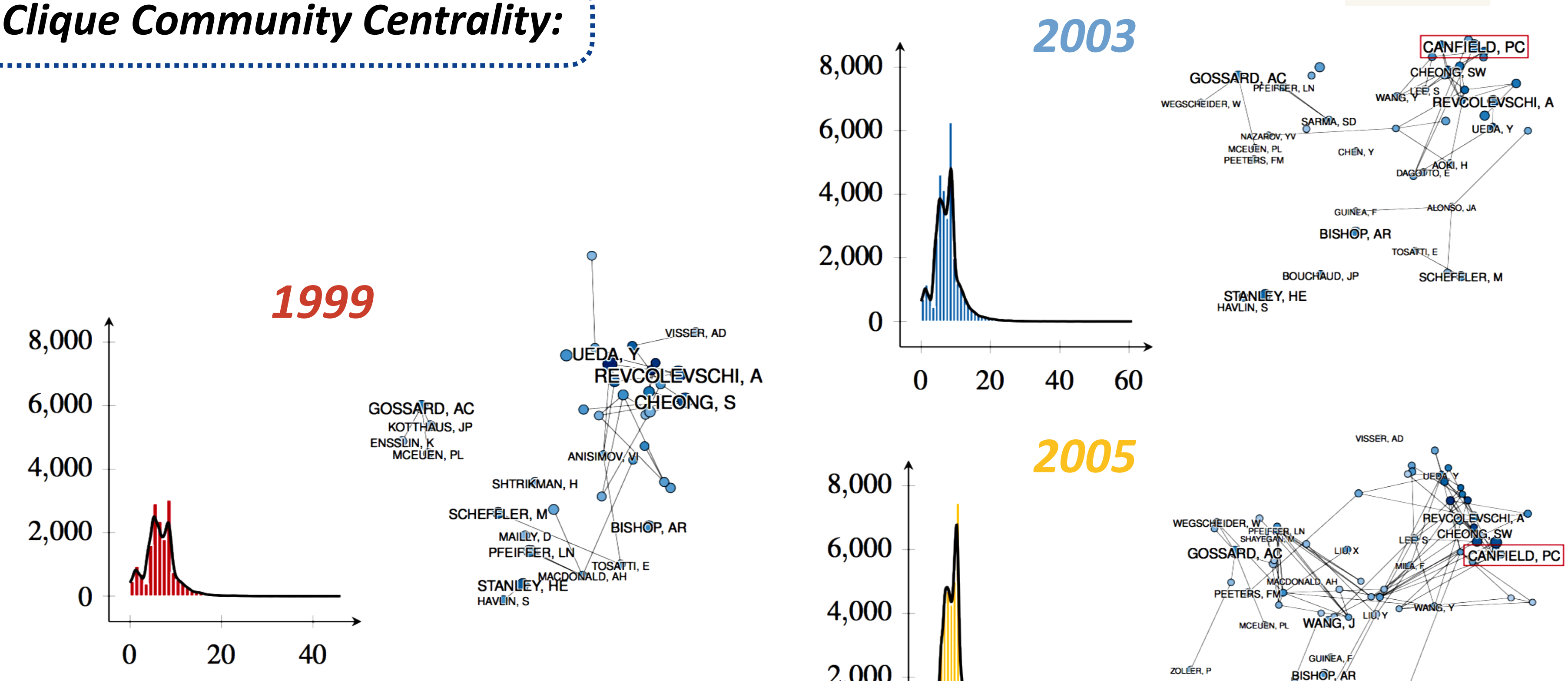


Clique community centrality allows for

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

Persistence and Complex Networks

Clique Community Centrality:



Density estimates of the clique community centrality values

Persistence and Complex Networks

Clique Community Persistence:

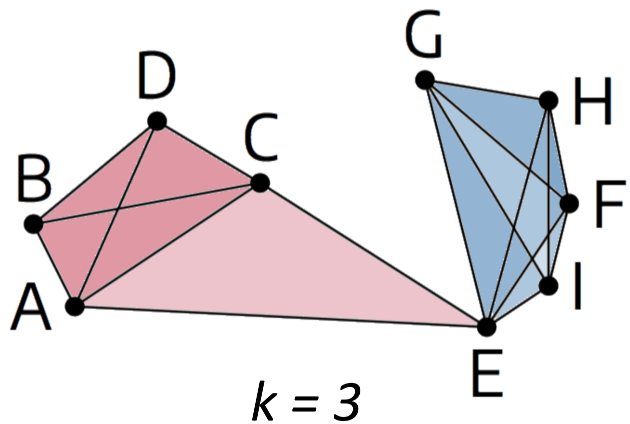
The presented approach allows for designing tools for:

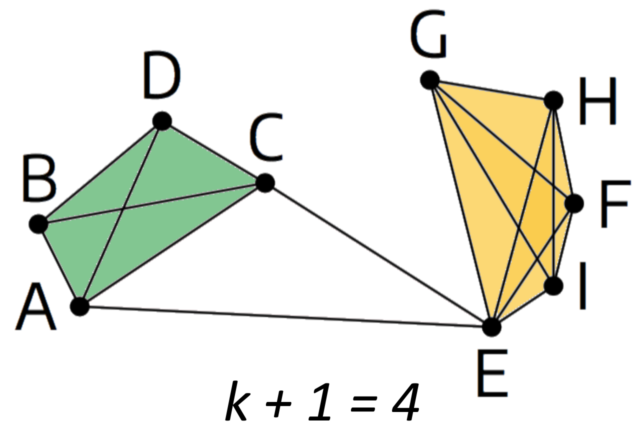
- ***Network Comparison***
  - ***Comparison Measures***
    - ***Persistence Indicator Function (PIF)***
    - ***PIF-based distance***
  - Clique Community ***Centrality Measure***
- ***Single Network Analysis***
  - ***Interactive Visualization Tool*** based on Nested Graphs

Persistence and Complex Networks

Clique Communities and Multiple k-Values:

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



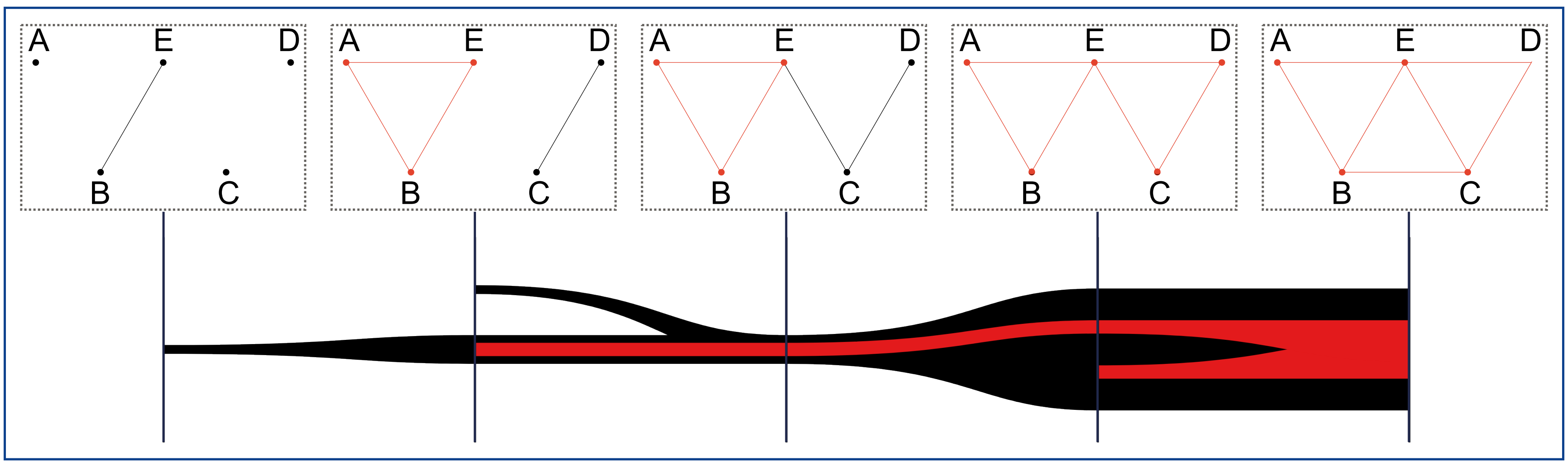


Each *(k+1)*-clique community of G_i is **contained** in ***exactly one*** *k*-clique community of G_i

Persistence and Complex Networks

Nested Graph:

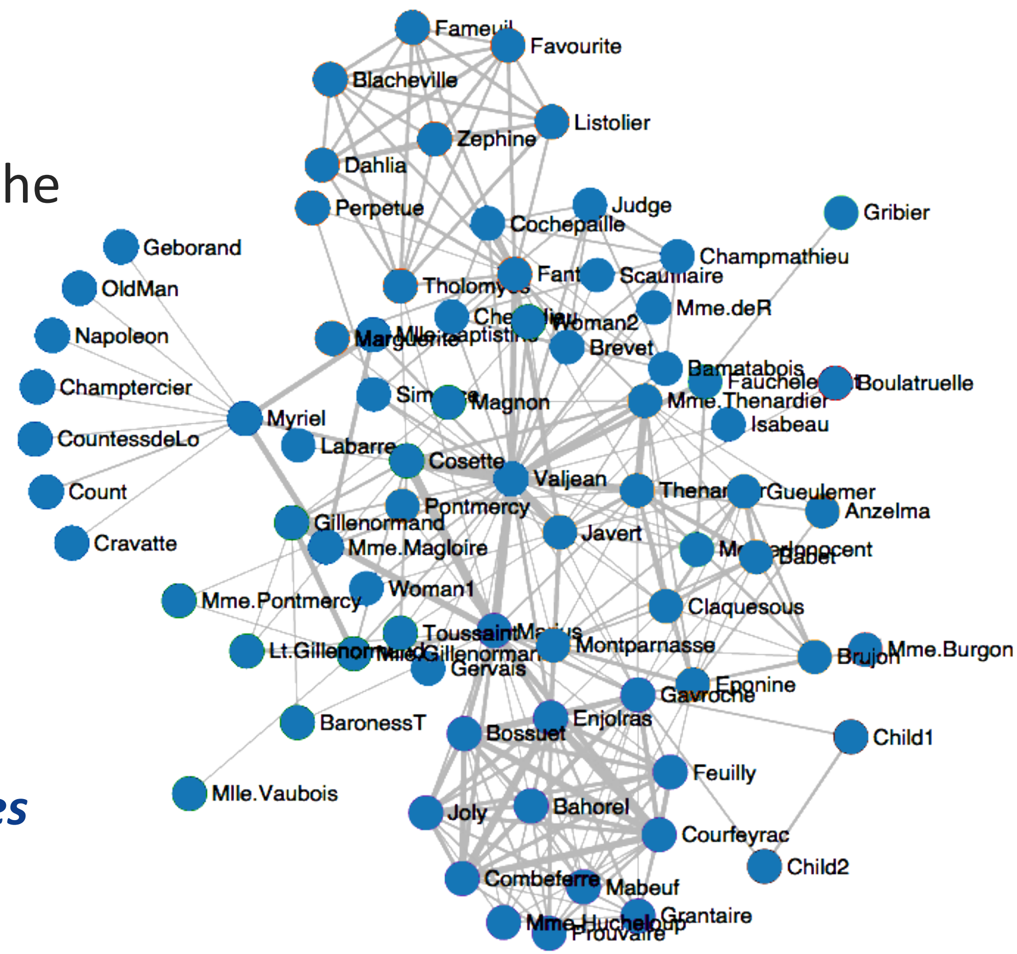
- Originally defined for connected components in scalar fields [Lukasczyk et al. 2017]
- Illustrates ***evolutions across two parameters***



Persistence and Complex Networks

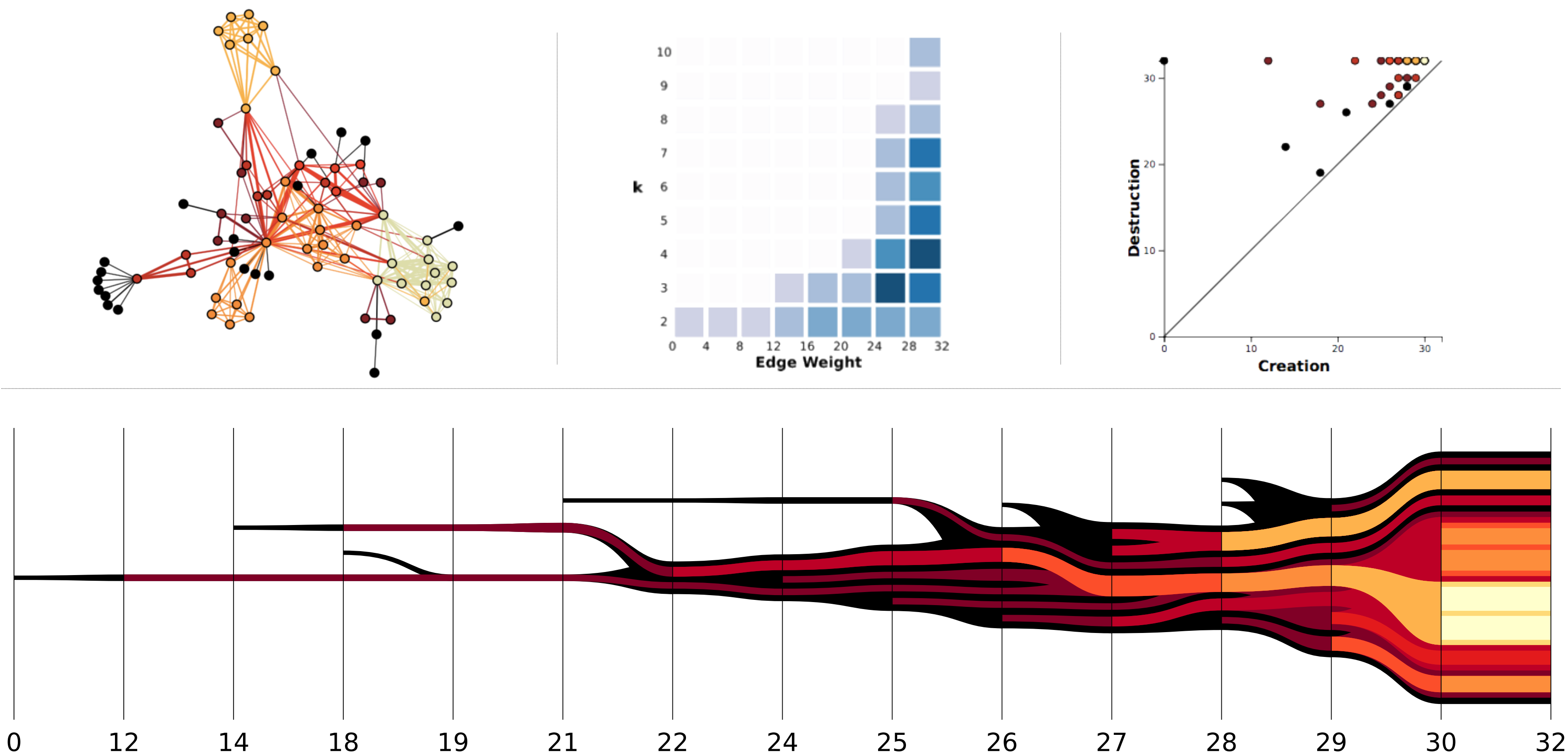
Nested Graph:

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



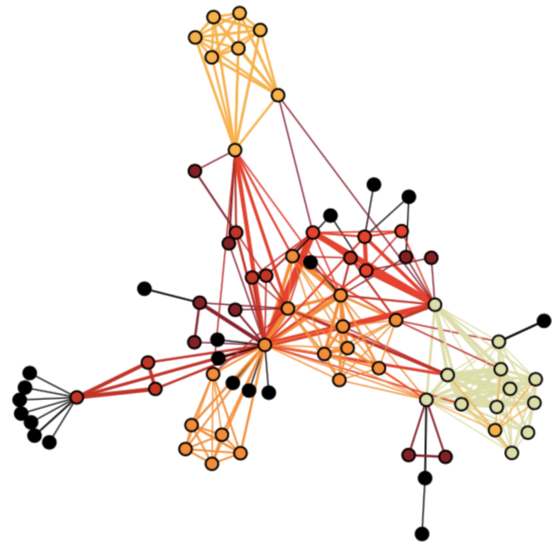
- ***edge weight*** ↔ ***1 / # co-occurrences***

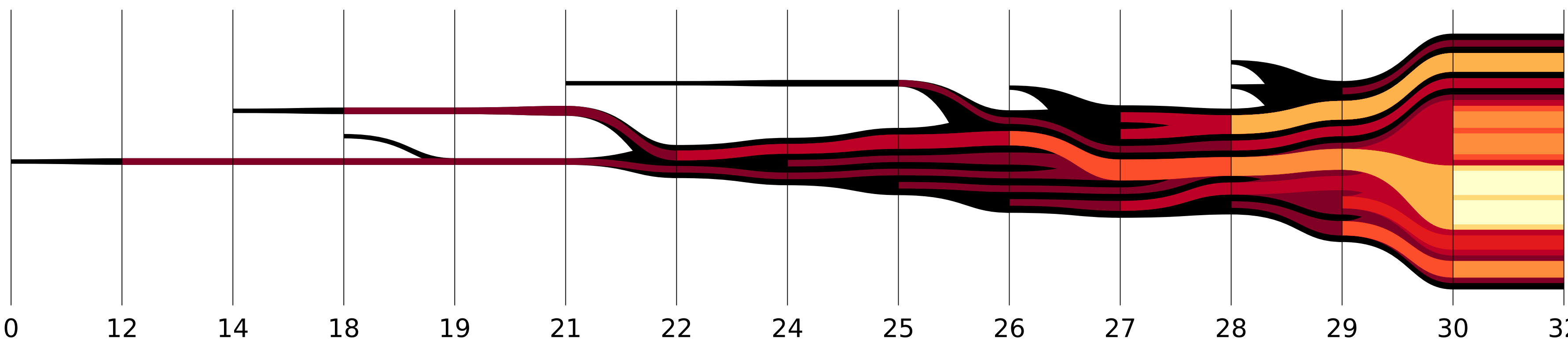
Persistence and Complex Networks



Persistence and Complex Networks

Nested-based visualization tool allows the user for

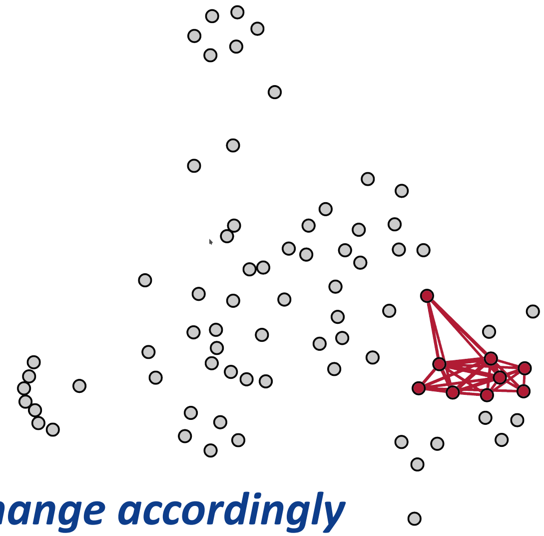




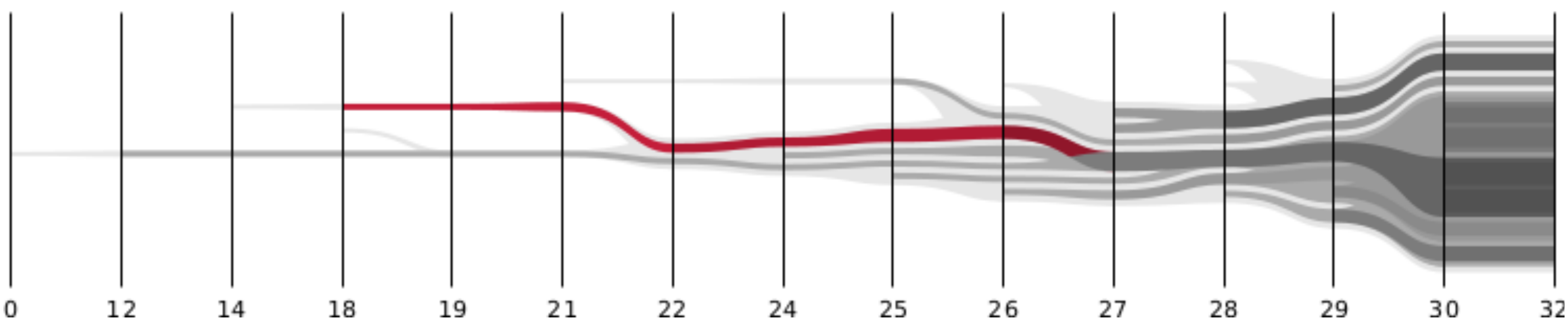
Persistence and Complex Networks

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***



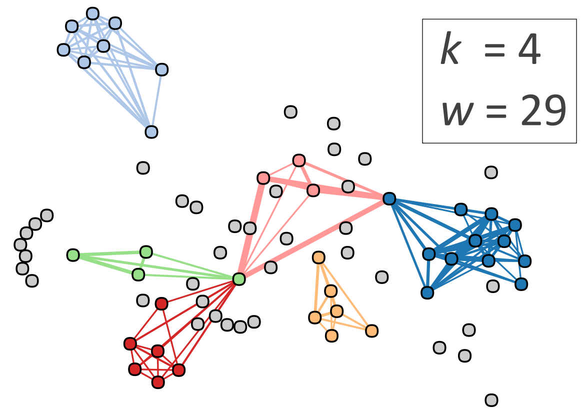
while the force-directed graph layout and the nested graph ***change accordingly***



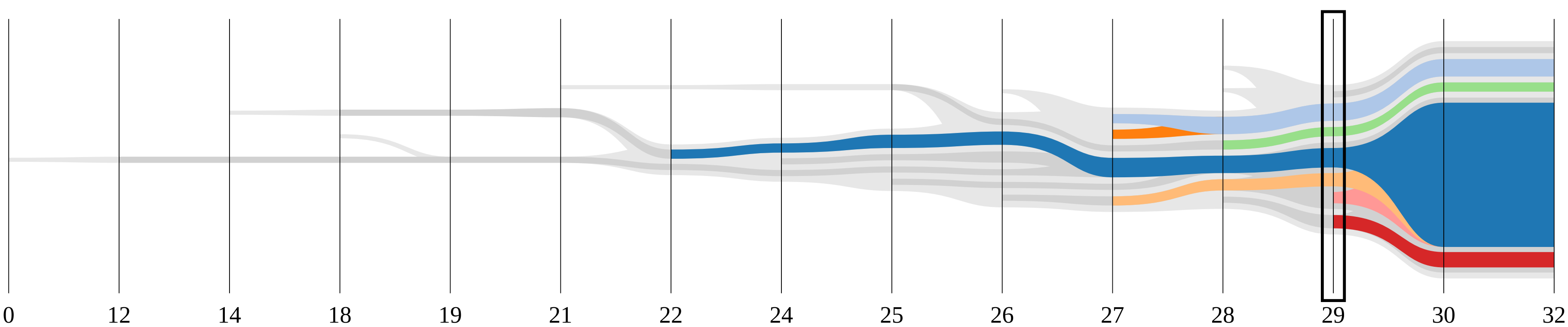
Persistence and Complex Networks

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:*
  - A. J. Zomorodian. ***Topology for computing.*** Cambridge University Press, 2005.
  - H. Edelsbrunner, J. Harer. ***Computational topology: an introduction***. American Mathematical Society, 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.

Today's References:

- *Complex 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.

Bibliography

Today's References:

- ***Homological Scaffolds:***
  - G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, F. Vaccarino. ***Homological scaffolds of brain functional networks.*** Journal of The Royal Society Interface, 11.101, 20140873, 2014.
  - L. D. Lord, et al. ***Insights into brain architectures from the homological scaffolds of functional connectivity networks.*** Frontiers in systems neuroscience, 10.85, 2016.
  - M. Guerra, A. De Gregorio, U. Fugacci, G. Petri, F. Vaccarino. ***Homological scaffold via minimal homology bases.*** Scientific Reports, 11.1, pages 1-17, 2021.
- ***Clique Community Persistence:***
  - 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.