Persistent Homology in Complex Network Analysis

Ulderico Fugacci
Kaiserslautern University of Technology
Department of Computer Science

July 7, 2017
Anything has Shape

“Data has shape and shape has meaning”
Gunnar Carlsson
Anything has Shape

“Data has shape and shape has meaning”
Gunnar Carlsson

Persistent Homology:
From:
Shape Analysis

To:
Chemistry
Neuroscience
Biophysics
Biology
Oncology
Network Analysis
...
“Data has shape and shape has meaning”
Gunnar Carlsson
Complex Networks

Definition:

A network is a complex system consisting of individuals or entities connected by specific ties such as

- Personal Relationship
- Shared Knowledge
- ...

References:
M. Newman, Networks: An Introduction, 2010
J. Scott, Social Network Analysis, 2017
Complex Networks

A Bunch of Examples:

- Social Networks
Complex Networks

A Bunch of Examples:

- *Social Networks*
- *Sensor Networks*
Complex Networks

A Bunch of Examples:
- *Social Networks*
- *Sensor Networks*
- *Biological Networks*
Complex Networks

A Bunch of Examples:

- Social Networks
- Sensor Networks
- Biological Networks
- Collaborative Networks
Complex Networks

A Bunch of Examples:

- Social Networks
- Sensor Networks
- Biological Networks
- Collaborative Networks
- ...

Outline

Brief Introduction to Complex Network Analysis
Outline

Brief Introduction to Complex Network Analysis

Persistence-based Network Analysis
Network Analysis

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

- *individuals* ↔ *nodes*
- *ties* ↔ *arcs*
Network Analysis

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

- **individuals** ↔ **nodes**
- **ties** ↔ **arcs**

Arcs can be:
- *directed*
- *weighted*
Network Analysis

**Representation:**

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

- **individuals** ↔ **nodes**
- **ties** ↔ **arcs**

Arcs can be:

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

- **individuals** ↔ **nodes**
- **ties** ↔ **arcs**

Arcs can be:
- *directed*
- *weighted*
Network Analysis

A Two-level Analysis:
Network Analysis

A Two-level Analysis:

- **Egocentric**
- **Sociocentric**
Network Analysis

A Two-level Analysis:

- Egocentric
- Sociocentric
Identifying Key Players

What is the most important individual?
Identifying Key Players

What is the most important individual?
Identifying Key Players

What is the most important individual?
Identifying Key Players

What is the most important individual?
Identifying Key Players

Centrality Measures:

Different criteria to underline different roles:

Key players
Brokers
Bridges
Isolated
...

A function $F : V \rightarrow R$ assigning to each node a "centrality" value:

- **Degree centrality**
- **Betweenness centrality**
- **Closeness centrality**
- **Eigenvector centrality**
- **Erdős distance**
Identifying Key Players

Degree Centrality:

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

$$D(v) := \# \{u \in V \mid (u, v) \in E\}$$
Given a node $v$ of $G=(V, E)$,

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

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:

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

**Closeness Centrality:**

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

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

**Eigenvector Centrality:**

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

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

where $\lambda$ is constant and

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

i.e., the $v^{th}$ entry of the eigenvector of

$$Ax = \lambda x$$

$x>0$ implies $\lambda$ must be the largest eigenvalue of $A$ and $x$ the corresponding eigenvector.
Identifying Key Players

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
## Centrality Measures:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Degree</strong></td>
<td>How many individuals can ( v ) reach directly?</td>
</tr>
<tr>
<td><strong>Betweenness</strong></td>
<td>How likely is ( v ) to be the most direct route between two individuals?</td>
</tr>
<tr>
<td><strong>Closeness</strong></td>
<td>How fast can ( v ) reach everyone in the network?</td>
</tr>
<tr>
<td><strong>Eigenvector</strong></td>
<td>How well is ( v ) connected to other well-connected individuals?</td>
</tr>
<tr>
<td><strong>Erdös</strong></td>
<td>How far is ( v ) from a specific individual?</td>
</tr>
</tbody>
</table>
Structural Analysis

Sociocentric Networks:

✦ **Structural Metrics:**
  - *Average* of a Centrality Measure
  - *Diameter*
  - *Density*
  - *Transitivity*
  - ...

✦ **Community Decomposition:**
  - *Atomic Communities*
  - *Clustering Techniques*
Structural Analysis

Structural Metrics:

- How far are two individuals at most?

Diameter:
The longest shortest path between any two nodes

Diameter(G) = 2
Structural Analysis

Structural Metrics:

- How close is $G$ to being an “everyone knows everyone” network?

**Density:**

\[
\text{Density}(G) = \frac{\text{Number of edges of } G}{\text{Number of all possible edges}}
\]

\[\text{Density}(G) = \frac{4}{6} = 0.67\]
Structural Analysis

Structural Metrics:

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

Transitivity:
\[
\frac{\text{Number of closed triplets of nodes}}{\text{Number of connected triplets}}
\]

Transitivity($G$) = $1/3 = 0.33$
Structural Analysis

Community Decomposition:

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

Image from [Fortunato 2009]
Structural Analysis

Community Decomposition:

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

Clique:
maximal subgraph whose nodes are all adjacent to each other

Image from [Fortunato 2009]
Structural Analysis

Community Decomposition:

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

\[ n \text{-Clique:} \]
maximal subgraph such that the distance of each pair of its nodes is not greater than \( n \)

Image from [Fortunato 2009]
Structural Analysis

Community Decomposition:

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

$k$-Plex:
maximal subgraph in which each node is adjacent to all other nodes of the subgraph except at most $k$ of them

Image from [Fortunato 2009]
Structural Analysis

Clustering Techniques:

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

approach based on

- Centrality Measures
- Atomic Communities
- Quality Functions
Structural Analysis

Clustering Techniques:

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

Centrality Measures

- Atomic Communities
- Quality Functions

Girvan-Newman Algorithm:

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

Image from [Fortunato 2009]
Structural Analysis

Clustering Techniques:

Agglomerative (bottom-up)  \{  approach based on  \}

Divisive (top-down)  \{  Centrality Measures  \}

Atomic Communities

Quality Functions

Clique Percolation:

\( k \)-adjacency: two clique 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

\textit{Decomposition} in \( k \)-clique communities

\( k = 4 \)

Image from [Palla et al. 2005]
Structural Analysis

Clustering Techniques:

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

Modularity-based Algorithm:

- **Modularity**: measure for clustering quality
- **Iterated aggregation** of communities of nodes whose merging increases modularity

- **Centrality Measures**
- **Atomic Communities**
- **Quality Functions**

Image from [Blondel et al. 2008]
Outline

- Brief Introduction to Complex Network Analysis
- Persistence-based Network Analysis
Persistence-based Network Analysis

Several Application based on Persistent Homology:

- Sensor Networks [De Silva 2013]
- Brain Networks [Lee et al. 2012]
- Collaborative/Co-occurrence 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 $\text{Flag}(G)$ induced by $G$
  - simplices of $\text{Flag}(G)$ $\leftrightarrow$ cliques of $G$
Persistence-based Network Analysis

Several Application based on Persistent Homology:

- **Sensor** Networks [De Silva 2013]
- **Brain** Networks [Lee et al. 2012]
- **Collaborative/Co-occurrence** 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 $\text{Flag}(G)$ induced by $G$
  - *simplices of* $\text{Flag}(G)$ ↔ *cliques of* $G$
Persistence-based Network Analysis

A Common Pipeline in TDA:

Complex Construction  →  \( \text{PH}_k \) module  →  Distance Computation

Topological Summaries have proven to be particularly effective to distinguish shapes

but

It's still hard to give a meaningful interpretation of what homological cycles represent
Outline

- Brief Introduction to Complex Network Analysis
- Persistence-based Network Analysis
Thank you

Ulderico Fugacci

TU Kaiserslautern, Dept. of Computer Science