Social Networks

JProf. Dr. Claudia Wagner
University Koblenz Landau
GESIS – Leibniz Institute for the Social Sciences
6th of June

- 10:30 Status Reports
  - ~7min presentation per team
  - Phrase a question that you can answer!
  - Describe how you answer the question: data and method
  - Related Work

- 14:15 Guest Lecture
  - Jan Lorenz: Agent Based Modeling

13th of June

- 10:30 no exercise class → but personal appointment with Tutor possible
- 14:15 Guest Lecture: Analytical Sociology (Haiko Lietz)
Goal for Today

- Strong and Weak ties (Not all links are equal)
- Social Capital (some nodes are in a better position than others)

Global Network Measures
- Density, shape of degree distribution, average path length, diameter, homophily, degree assortativity, ...

Local Network Measures
- Clustering coefficient, transitivity, balance, network motifs
- Node Position Measures
  - Degree, Centrality
Formally, we can characterize a network through 2 statistics:

- The characteristic path length $L$
  - The average length of the shortest paths connecting any two nodes.
    - This is not the same as the diameter of the graph, which is the maximum shortest path connecting any two nodes

- The clustering coefficient $C$
  - The average local density.

A small world graph is any graph with a relatively small $L$ and a relatively large $C$. 

Slide from James Moody
• Everyone in a cave knows each other
• A few people make connections
• Are C and L high or low?
• C high, L high
In a highly clustered, ordered network, a single random connection will create a shortcut that lowers $L$ dramatically.

*Small world* properties can be created by a small number of shortcuts.

Slide from Lada Adamic
“If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.” (Anatole Rapoport 1953)
(1) We meet our friends through other friends
  - B and C have opportunity to meet through A

(2) B and C’s mutual friendship with A gives them a reason to trust A

(3) A has incentive to bring B and C together to avoid stress
  → if A is friends with two people who don’t like each other it causes stress
A bridge is an edge whose removal places A and B in different components.

If A is going to get new information (like a job) that she doesn’t already know about, it might come from B.
A local bridge is an edge whose endpoints A and B have no friends in common (so a local bridge does not form the side of any triangle)

If A is going to get new information (like a job) that she doesn’t already know about, it might come from B.
Social Embeddedness

- The embeddedness of an edge in a network to be the number of common neighbors the two endpoints have (the numerator in the neighborhood overlap).
- Local bridges are precisely the edges that have an embeddedness of zero.

$$\text{embeddedness}_{ij} = \frac{n_{ij}}{((k_i - 1) + (k_j - 1) - n_{ij})}$$

- $n_{ij}$ is the number of common friends of node $i$ and $j$
- $k_i$ is the degree of node $i$
- $k_j$ is the degree of node $j$
- What is the embeddedness of edge between A and B?
- What is the embeddedness of edge between A and D?
High social embeddedness $\rightarrow$ trusted relationship

Low social embeddedness $\rightarrow$ local or global bridge

Quercia et al. analyse friendships on fb:
- The more socially embedded a tie the less likely it will break

The strength of an interpersonal tie is a (probably linear) combination of

- The amount of time
- The emotional intensity
- The intimacy
- The reciprocal services which characterize the tie

Can you give examples of strong / weak ties?

Mark Granovetter, Stanford University
Weak ties can occur between cohesive groups

- old college friend
- former colleague from work
The Strength of Weak Ties

- Granovetter: How often did you see the contact that helped you find the job prior to the job search
  - 16.7% often (at least once a week)
  - 55.6% occasionally (more than once a year but less than twice a week)
  - 27.8% rarely – once a year or less

- Weak ties will tend to have different information than we and our close contacts do

- Long paths rare
  - 39.1% info came directly from employer
  - 45.3% one intermediary
  - 3.1% > 2 (more frequent with younger, inexperienced job seekers)

- Also in the Milgram small world experiments, acquaintanceship ties were more effective than family, close friends at passing information

Slide from James Moody
Measure Tie Strength

- \( NO(A,B) = 0 \)
- \( NO(A,F) = 1/6 \)

\[
NO(x, y) = \frac{|\text{common neighbors of } x \text{ and } y|}{|\text{neighbors of at least one of } x \text{ or } y|}
\]
The position of a person in a network may give the person some advantage (e.g. early access to information, control access, trust and support)
Two key concepts:

- **Closure**: competitive advantage through trusted relationships that help to manage and reduce risk (Coleman, 1988 and 1990) → a node that is part of a dense cluster maintains trusted relationships

- **Brokerage**: competitive advantage through access control (Burt, 1992) → a node that functions as a bridge has early access to information (brokerage role)
Social Capital
Structural Holes

- Structural holes separate non-redundant sources of information
  - Non-redundant contacts are connected by structural holes

- Network betweenness, proposed by Freeman (1977), is an index that measures the extent to which a person brokers indirect connections between all other people in a network.

\[
C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}
\]

Betweenness centrality: number of shortest paths going through the actor \(\sigma_{st}(i)\)

All shortest path from \(s\) to \(t\) that go through \(i\)

All shortest path from \(s\) to \(t\)
Dijkstra's Algorithm

- Goal: Compute shortest path → select one seed node (here A) and compute shortest path between seed node and all other nodes

- Initialize: seed node gets distance 0, all other nodes get distance infinity, put all nodes in unvisited node list

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm

- Step 1: choose node with smallest distance (here A)
- Step 2: compute distance between selected node (A) and unvisited neighbour nodes (B, D) → update if smaller
- Step 3: add selected node (D) to list of visited nodes

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm

- Update table
- Update lists

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm

- Step 1: choose node with smallest distance (here D) and visit all neighbours
- Step 2: compute distance between selected node (D) and unvisited neighbour nodes (B, D) → update if smaller
- Step 3: add selected node (D) to list of visited nodes

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra’s Algorithm

- Update table
- Update lists


Vertex | Shortest distance from A | Previous vertex
---|---|---
A | 0 | 0
B | 3 | D
C | ∞ | C
D | 1 | A
E | 2 | D

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm


https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm

- Update table
- Update lists

Choose unvisited node with smallest distance (now B)

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
The only unvisited neighbour of B is C

Visited = [A, D, E]  Unvisited = [B, C]

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
- No update of table necessary, but update lists

![Graph](image)

- C has no neighbours to visit, no updates
- Algorithm ends

https://www.youtube.com/watch?v=pVfj6mxhdMw&t=8s
Dijkstra's Algorithm

- Greedy Algorithm since we always select the node with the shortest distance from seed node
- Algorithm makes local optimal decisions

- Pick seed node and initialize distance table
- While len(unvisited nodes) != 0:
  - selected_node = min(distance(unvisited nodes))
  - Compute distance for unvisited neighbours of selected_node
  - If distance of selected_node is smaller than distance in table
    - Update table
  - Add selected node to visited list and remove it from unvisited list
Burt finds empirical evidence for the fact that brokers tend to have good ideas

- Survey data on several hundred managers in a large company.
- Compensation, positive performance evaluations, and promotions were disproportionately given to managers who brokered connections across structural holes.
- Managers whose networks spanned structural holes were more likely to express an idea and to discuss it with colleagues.

Summary

- Triangles (triadic closure) lead to higher clustering coefficients
  - Your friends will tend to become friends

- Local bridges will often be weak ties $\rightarrow$ Information comes over weak ties
  - Social embeddedness identifies weak ties

- Structural holes benefit from their position $\rightarrow$ can control access, have early access to information, may contribute more ideas
  - Structural holes have high betweenness centrality
Network Measures

- **Global Measures**
  - Density, shape of degree distribution, average path length, diameter, homophily, degree assortativity, ...

- **Local Measures**
  - Clustering coefficient, transitivity, balance, network motifs

- **Node Position Measures**
  - Degree, Centrality
Density of Networks

- Density = how many links exist / how many could exist

- How many links could exist in a network with 4 nodes?
  - Handshake Problem: how many handshakes could happen between 4 people: A, B, C, D
    
    • A – B
    • A – C
    • A – D
    • B – C
    • B – D
    • C – D

  - Undirected Network: \((N*(N-1))/2\)
  - Directed Network: \(N*(N-1)\)
- is the **maximum length of all shortest paths** in a network

- Compute the shortest path from every node to all other nodes → the **diameter** is the longest of all the calculated path lengths.

- If a network is disconnected, its diameter is the maximum of all diameters of its connected components.
Wiener Index

- Is a topological index
- Its defined as the sum of the lengths of the shortest paths between all pairs of nodes

Degree Distributions

Slide from Jennifer Goldbeck
What is the shape of a degree distribution?

- Many nodes on the internet have low degree
  - One or two connections
- A few (hubs) have very high degree
- The number $P(k)$ of nodes with degree $k$ follows a power law:

  $$ P(k) \propto k^{-\alpha} $$

  Where alpha for the internet is about 2.1
- I.e., the fraction of web pages with $k$ in-links is proportional to $1/k^2$
“the“ is the most frequent word and accounts for 7% of all word occurrences.

What is the frequency of the second and third most frequent word?
Power-law distribution

- linear scale
- log-log scale

- high skew (asymmetry)
- straight line on a log-log plot

Slide from Lada Adamic
Power-law distributions

- Right skew
  - normal distribution is centered on mean
  - power-law or Zipf distribution is not

- High ratio of max to min

- Power-law distributions have no “scale”:
  - relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities
Where do power laws come from?

- Many different processes can lead to power laws
- There is no one unique mechanism that explains it all

Slide from Lada Adamic
• Price (1965)
  • Citation networks
  • new citations to a paper are proportional to the number it already has
  • each new paper is generated with m citations
  • new papers cite previous papers with probability proportional to their in-degree (citations)
This is a “Rich get Richer” Model

Explanation for various power law effects

1. **Citations**

2. Assume **cities** are formed at different times, and that, once formed, a city grows in proportion to its current size simply as a result of people having children

3. **Words**: people are more likely to use a word that is frequent (perhaps it comes to mind more easily or faster)
Assortative Mixing (Newman 2003)

- Assortative Mixing refers to selective linking of nodes to other nodes who share some common property
  - E.g. degree
    - High degree nodes in a network associate preferentially with other high-degree nodes (this is called degree assortativity)
    - Low degree nodes in a network associate preferentially with high-degree nodes (this is called degree dissortativity)
  - E.g. node attributes such as gender or race
    - nodes of a certain type tend to associate with the same type of nodes (this is also called homophily)
**Assortative Mixing (Newman 2003)**

- $e_{ij}$ defines the fraction of edges in a network that connect a node of type $i$ to one of type $j$
  - In an undirected network: $e_{ij} = e_{ji}$
  - It satisfies the sum rule:
    \[
    \sum_{ij} e_{ij} = 1, \quad \sum_{j} e_{ij} = a_{i}, \quad \sum_{i} e_{ij} = b_{j}
    \]

- **Assortativity coefficient:**
  \[
  r = \frac{\sum_{i} e_{ii} - \sum_{i} a_{i} b_{i}}{1 - \sum_{i} a_{i} b_{i}}
  \]
  - $r = 0$ when there is no assortative mixing
  - $r = 1$ when there is perfect assortative mixing
Assortative Mixing (or Homophily)

- [Newman 2003]

**FIG. 8** Friendship network of children in a US school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not _vice versa_. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.
Network Measures

- Global Measures
  - Density, shape of degree distribution, average path length, diameter, homophily, degree assortativity, ...

- Local Measures
  - Clustering coefficient, transitivity, balance, network motifs

- Node Position Measures
  - Degree, Centrality
Clustering Coefficient (Watts & Strogatz 1998)

\[ C = \text{The fraction of pairs of neighbors of the node that are connected} \]
\[ \text{“What percentage of your friends know each other?”} \]

Let \( n_i \) be the number of neighbors of vertex \( i \)

\[ C_i = \frac{\text{number of connections between } i \text{’s neighbors}}{\text{maximum number of possible connections between } i \text{’s neighbors}} \]

\[ C_i \text{ directed} = \frac{\# \text{ directed connections between } i \text{’s neighbors}}{n_i \times (n_i - 1)} \]

\[ C_i \text{ undirected} = \frac{\# \text{ undirected connections between } i \text{’s neighbors}}{n_i \times (n_i - 1)/2} \]
Triad: a subgroup of three actors

Network Motifs: a subgroup and all possible edges between them

Count how often each pattern shows up

![Diagram of four possible triadic states in a graph: 0 lines, 1 line, 2 lines, and 3 lines.](image)
Local Network Measures

Transitivity

• If actor i „likes“ j, and j „likes“ k, then i also „likes“ k

Balance

• If actor i and j like each other, they should be similar in their evaluation of some k
• If actor i and j dislike each other, they should evaluate k differently

Example 1: Transitivity
Example 2: Balance
Example 3: Balance
Node Position Measures

- **Degree**: how connected is a node?
- **Closeness**: how easy can a node reach other nodes?
- **Betweenness**: how much does a node function as a connector, as a bridget for other nodes?
- **Influence, Prestige, Eigenvector**: how important are your friends? (Page Rank, HITS)
Degree Centrality

- Node with highest degree centrality has degree 7

- Normalize by highest possible degree: \( n-1 = 17 \)
- Normalized degree centrality = \( \frac{7}{17} \)
Closeness Centrality

- How easy can a node reach other nodes?
- Closeness of node $v \rightarrow$ inverse of the sum of the shortest path between $v$ and all other nodes $i$

$$closeness(v) = \frac{1}{\sum_{i \neq v} d_{vi}}$$

- The more central a node is, the lower its total shortest distance to all other nodes.
- Closeness can be regarded as a measure of how long it will take to spread information from $v$ to all other nodes sequentially.
What is the closeness centrality of node C?

1/5
What is the closeness centrality of node D?

1/5
Betweeness Centrality

- Quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

- A node with high betweenness centrality is a broker → can control access to information in the network

**Betweenness centrality:** number of shortest paths going through the actor 

\[ C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \]

- All shortest path from s to t that go through i
- All shortest path from s to t
Example

- Betweenness of node 2 = ?

How many node-pairs to check?
\[(n-1) \times (n-2) / 2 = 10\]

1-3, 1-4, 1-5, 1-6
3-4, 3-5, 3-6
4-5, 4-6
5-6

\[\sigma_{13}(2) / \sigma_{13} = 1/1\]
\[\sigma_{14}(2) / \sigma_{14} = 0\]
\[\sigma_{15}(2) / \sigma_{15} = 0\]
\[\sigma_{34}(2) / \sigma_{34} = 0\]
\[\sigma_{35}(2) / \sigma_{35} = 1/2\]
\[\sigma_{36}(2) / \sigma_{36} = 0\]
\[\sigma_{45}(2) / \sigma_{45} = 0\]
\[\sigma_{46}(2) / \sigma_{46} = 0\]
\[\sigma_{56}(2) / \sigma_{56} = 0\]

Don`t forget to normalize it!
- Two nodes have the same number of friends but the importance of their friends differ.

The friends of this node are more important.

- Eigenvector centrality or Page Rank → importance of a node is proportional to the importance of its neighbours.
Eigenvector Centrality

- The centrality of a node \( i \) is proportional to the centrality of its neighbours.

\[
C_i^e(g) = a \sum_j g_{ij} C_j^e(g)
\]

- Self-referential concept.

\( g_{ij} \) is 1 if \( i \) and \( j \) are friends and 0 otherwise.
Eigenvector Centrality

- We have an equation with multiple unknowns
  - In Matrix notation: $C = agC$

- $C$ is an eigenvector $\rightarrow$ has multiple solutions

- But we look for the eigenvector $C$ that is associated with the largest eigenvalue $g$

- Perron Frobenius Theorem $\rightarrow$ if we deal with a non-negative matrix, the eigenvector that is associated with the largest eigenvalue will have all positive real values
Eigenvector Centralities

This node is more important than…

…this node
Similar Idea: importance of a node is proportional to importance of its neighbours

- Random surfer: do a random walk in the network and remember how often you visited each page
- Teleportation probability
- $PR(X) =$ probability that you are at node $X$

Teleportation probability. What is the probability that we end up at $X$ in teleportation step?

What is the probability that we end up at $X$ in random walk?

$$PR(x) = \frac{1-\lambda}{N} + \lambda \sum_{y \rightarrow x} \frac{PR(y)}{out(y)}$$
- Page ranks sum to 1 (probabilities!)

https://en.wikipedia.org/wiki/PageRank
Any further questions?

See you next week