Social Networks

JProf. Dr. Claudia Wagner
University Koblenz Landau
GESIS – Leibniz Institute for the Social Sciences
Last Lectures

- Networks
  - Global Measures
    - E.g., density, diameter, homophily, degree assortativity, ...
  - Local Measures
    - E.g., Clustering coefficient, transitivity, network motifs
  - Node Position Measures
    - Degree, Centrality

- Agent Based Modelling

- Analytical Sociology
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 \( i \)

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

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

\[ C_{i \text{ undirected}} = \frac{\text{# undirected connections between } i\text{’s neighbors}}{n_i \times (n_i - 1)/2} \]

Slide from Lada Adamic
Triad: a subgroup of three actors

Network Motifs: a subgroup and all possible edges between them

Count how often each pattern shows up

Fig. 4.3. Four possible triadic states in a graph
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)
- 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
Betweenness 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.

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

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

- **Betweenness of node 2 = ?**

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

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

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

1,5 / 10

Don`t forget to normalize it!
This node has a higher betweenness centrality
Eigenvector centrality or Page Rank → importance of a node is proportional to the importance of it‘s neighbours

The friends of this node are more important
- Importance of a node is proportional to importance of its neighbours
- Page ranks sum to 1 (probabilities!)

https://en.wikipedia.org/wiki/PageRank
Compute Page Rank

- Random surfer: do a random walk in the network and remember how often you visited each page
- Teleportation probability (lamda)
- 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?

\[
PR(x) = \frac{1-\lambda}{N} + \lambda \sum_{y \rightarrow x} \frac{PR(y)}{\text{out}(y)}
\]
The centrality of a node \( i \) is proportional to the centrality of its neighbours.

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

- \( C_i^e(g) \) is the centrality of node \( i \).
- \( g_{ij} \) is 1 if \( i \) and \( j \) are friends and 0 otherwise.
- \( C_j^e(g) \) is the centrality of node \( j \).

Self-referential concept
• 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
So Far

- Static Networks and how to describe them
- Network Dynamics
  - Networks are usually dynamic, they change over time
  - Processes in networks are dynamics
Dynamics in Networks

- Understanding how information/behavior spreads

Mechanisms of Adoption:

- **Social learning**
  - Peers provide information that increases utility of using X or decreases risk of using X
  - If my friend explains me how X works I may start using it

- **Normative Influence**
  - Does not impact the intrinsic value of X
  - Peers punish someone for using X (or not using X)

- **Network externalities**
  - The value of X depends on the number of prior adopters
  - If enough people use X, X becomes more valuable
Agent-based models (ABM) allow to simulate the contagion process

- Implement micro-behavior of each agent
- Re-create and predict the appearance of complex macro-phenomena (e.g., an epidemic)
Basic Types of Contagion Models

Simple Contagion

- Think about how a virus or information spread

Complex Contagion

- Think about how behavior or practices spread, especially if
  - There is a risk in adopting it (uncertainty/risk)
  - People need help in adopting it (complexity)
  - People cannot see who adopted it (observability)
  - The practice has not fully institutionalized (legitimacy)
  - Some practices require continuous peer-support while others are self sustaining (sustainability)
Simple Contagion

- **SI Model**

- Parameter $\beta$ governs contact and transmission (infection rate): probability of contagion after contact per unit time; it also defines number of contacts per unit time.

- $X$ infected nodes $\rightarrow x = X/n$

- $S$ susceptible nodes $\rightarrow s = S/n$

- $x*s$ is the average number of susceptible people that infected nodes meet per unit time.

- $\beta*x*s$ is the average number of susceptible people that become infected from all infected nodes per unit time.
SI Model

- Describe relationship between rates of change and physical quantities (differential equation)

\[ \frac{dx}{dt} = \beta (1 - x)x \]

Proportion of nodes infected per time unit:

\[ s = 1 - x \]

\[ x + s = 1 \]
\[
\frac{dx}{dt} = \beta (1 - x) x
\]

\[
\int_{x_0}^{x} \frac{1}{(1 - x) x} dx = \int_{0}^{t} \beta dt
\]

\[
\int_{x_0}^{x} \frac{1}{(1 - x)} dx + \int_{x_0}^{x} \frac{1}{x} dx = \beta t - \beta 0
\]

\[
\int_{x_0}^{x} \frac{1}{(1 - x)} dx + \int_{x_0}^{x} \frac{1}{x} dx = \beta t
\]

\[
\ln \frac{1 - x_0}{1 - x} + \ln \frac{x}{x_0} = \beta t
\]

\[
\ln \frac{(1 - x_0)x}{(1 - x)x_0} = \beta t
\]
\[ \ln \frac{(1 - x_0)x}{(1 - x)x_0} = \beta t \]

\[ \iff \frac{(1 - x_0)x}{(1 - x)x_0} = e^{\beta t} \]

\[ \iff \frac{x}{1 - x} = \frac{x_0 e^{\beta t}}{1 - x_0} \]

\[ \iff \frac{1}{x} = \frac{1 - x_0}{x_0 e^{\beta t}} + 1 = \frac{1 - x_0 + x_0 e^{\beta t}}{x_0 e^{\beta t}} \]

\[ \iff x = \frac{x_0 e^{\beta t}}{1 - x_0 + x_0 e^{\beta t}} \]

The Logistic Growth Equation is given by:

\[ x(t) = \frac{x_0 e^{\beta t}}{1 - x_0 + x_0 e^{\beta t}} \]

This equation models the growth of a population or a phenomenon that has a carrying capacity, reaching a maximum value as time progresses. The parameter \( \beta \) represents the growth rate, while \( x_0 \) is the initial condition.

For a visual representation of the Logistic growth curve, please refer to the link provided:

SIR and SIS Model

S: Susceptible (healthy)

I: Infected (sick)

R: Removed (immune / dead)

Infection
Recovery
Recovery

Removal
**SIR model**

- **β infection rate:** probability of infection per unit of time; governs contact and probability of infection
- **γ recovery rate:** probability of recovery from infection per unit time

\[
\frac{ds}{dt} = -\beta sx \\
\frac{dx}{dt} = \beta sx - \gamma x \\
\frac{dr}{dt} = \gamma x
\]

- **Proportion of people that move out of S**
- **Proportion of people that move into I and out of I**
- **Proportion of people that move into R**
Reproduction Number

- **Reproduction number R**: additional people that a newly infected person passes the disease onto before they recover

- \( R = \frac{\beta}{\gamma} \)

- If infection rate is higher than removal rate the disease will survive

- In SI: \( \gamma = 0 \)

- Example reproduction numbers:
  - Aids/HIV: 2-5
  - Measles: 12-18
  - Influenza: 2-3
Whenever there's a spread of the disease that reaches most of the population, we say that there was an epidemic.

\[
\frac{\beta}{\gamma} \leq 1 \text{ then no epidemic occurs}
\]
\[
\frac{\beta}{\gamma} > 1 \text{ then epidemic occurs}
\]
\[
\beta = \gamma \text{ is the epidemic transition}
\]

Figure 6.5. Phase transition in the SIR model. When the reproduction rate \( R \) of the disease exceeds one (the epidemic threshold), an epidemic occurs.
SI Model Netlogo

Infection chances

Recovery chances

http://ccl.northwestern.edu/netlogo/models/epiDEMBasic
• Homogenous Mixing: everyone can meet everyone else
  - Set number of people and infection probability

• An infected person has a probability of recovering after reaching their recovery time period, which is also set by the user. The recovery time of each individual is determined by pulling from an approximately normal distribution with a mean of the average recovery time set by the user.

• White individuals are uninfected, red individuals are infected, green individuals are recovered. Once recovered, the individual is permanently immune to the virus.
Examples

\[ \beta = 0.8, \gamma = 0.4 \quad \text{R=2} \]

\[ \beta = 0.4, \gamma = 0.8 \quad \text{R=0.5} \]

Same as SIR model but recovered nodes become again susceptible

\[
\frac{ds}{dt} = \gamma x - \beta sx \\
\frac{dx}{dt} = \beta sx - \gamma x
\]
So far we assumed homogenous mixing. In networks proportion of nodes that infect per time unit depends on

- Size of both populations
- Network structure
- Infection probability of a disease
SI Model for Networks

- Homogeneous network: all nodes have degree close to $k$ (e.g. Erdős Rényi network or regular lattice)

- Equation of dynamics

\[
\frac{dx}{dt} = \beta(1 - x)x
\]

Proportion of nodes infected per time unit (homogeneous mixing)

\[
\frac{dx}{dt} = \beta \langle k \rangle x (1 - x)
\]

Proportion of nodes infected per time unit (network with average degree $k$)
Average node degree = 2.5

Average node degree = 10

The denser the network the faster the diffusion

- Small World Networks (average shortest path length is small)


Shortcuts in the network speed up simple contagion
Summary Simple Contagion

- Dense networks support fast diffusion
- Short path length supports fast diffusion

- But relationship between network topology and epidemics is complex
  - E.g. no epidemic threshold for scale free networks
  - Do simulations to estimate final proportion of infected nodes
Complex Contagion

- **Simple Contagion**: Adoption depend only on 2 nodes (infected node meets susceptible node) → then the susceptible node gets infected with some probability
  - e.g. virus or information spread like this

- **Complex Contagion**: In many situations just seeing one friend is not enough → adoption depends on several nodes (and node-properties) at the same time
  - e.g., adoption of products or behavior
Granovetter‘s Linear Threshold Model

• Each agent has to make a **binary decision** (e.g., going to a riot or not, yes or no)
• Each agent has an **individual threshold** which determines under which circumstances he/she would pick which alternative
• The payoff of one agent depends on **how many other agents have picked what**
• If the payoff **exceeds a threshold** the agent adapts
• The threshold of each agent may depend on his/her properties such as education, age, personality, etc.
  • Conservative agent \(\rightarrow\) 100% threshold
  • Extremist or Individualist \(\rightarrow\) 0% threshold

Example I

- 100 people milling around in a square
- Each person has an individual threshold which describes how many others need to riot before he/she will start
- Assume we distribute thresholds between 0 and 99 uniformly across the 100 people
- What will happen?
  - Person with threshold 0 starts breaking a window
  - This activates the person with threshold 1 who joins
  - This activates the person with threshold 2

• What happens if we change the threshold of the person who had threshold 0 to threshold 1?

• It is very dangerous to infer individual dispositions from aggregated outcomes!!!

Bandwagon Effects

http://modelingcommons.org/browse/one_model/4784#model_tabs_browse_info
If we know the threshold distribution of agents $f(x)$ and the proportion of agents which are “activated” at timestep $t_1$, we can predict how many agents will be active at the next timestep.

- The cdf of $f(x)$ indicates the proportion of agents having a threshold less than or equal $x$.
  - these agents will join the riot in the next time step.
Probabilistic Threshold Models

• Threshold models are a standard deterministic models, i.e. whenever we pass the threshold the behavior is triggered

• Probabilistic variant of threshold models
  • We define a probability that a person engages in behavior
  • The probability is based on the same variables as the threshold
  • However, the emerging behavior is probabilistic
- Agents may be embedded in a network structure

- In the simplest form threshold model assumes that each agent observes behavior of all others

- Different tie strength may determine how much attention they pay to each neighbour

- If node X gets activated depends on the (weighted) sum of his neighbours
Summary

- Spread of disease $\rightarrow$ simple contagion
  - If network has low diameter $X$ spreads faster
  - Random network and small world networks spread faster than lattice

- Spread of behavior $\rightarrow$ complex contagion
  - Social reinforcement is necessary; clustered networks
  - If network has high clustering $X$ spreads faster
  - Lattice and small world network spread faster than random network
Any further questions?

See you next week