Part 1 - Characteristics of Networks
Content of the course

- **Introduction, background and fundamentals of network analysis** - representing and measuring networks, centrality measures

- **Strategic network formation** - pairwise stability and efficiency, the connections model and its dynamic version, the co-author model, positive and negative externalities in networks, small worlds in an islands-connections model, general tension between stability and efficiency

- **Network games and allocation rules** - Myerson value in network settings, egalitarian allocation rule, component-wise egalitarian allocation rule, flexible network allocation rules, allocation rule for dynamic random networks
Social and economic networks

- Interdisciplinary field: economics, sociology, psychology, mathematics, statistics, computer sciences, physics, biology ...

- Different approaches to the analysis of social networks: theoretical models, empirical works, experiments

- Central role for modeling different phenomena: transmission of information (e.g., job opportunities), learning, influence, opinion formation, contagion, trade of goods and services, business interactions, financial networks, scientific collaboration, political interactions, criminal activities, ...

Bavelas B (1948) A mathematical model for group structure, Human Organizations 7: 16–30


Bonacich PB (1972) Factoring and weighting approaches to status scores and clique identification, *Journal of Mathematical Sociology* 2: 113–120


References (3/8)


Freeman LC (1977) A set of measures of centrality based on betweenness, Sociometry 40: 35–41


Representing Networks
A network is represented by a graph \((N, g)\), where

- \(N = \{1, 2, \ldots, n\}\) set of nodes (agents, players, vertices)
- \(g = [g_{ij}]\) real-valued \(n \times n\) matrix (adjacency matrix)

\(g_{ij}\) - relationship between \(i\) and \(j\) (possibly weighted and/or directed), also referred to as a link \(ij\) or an edge.

\(G =\) collection of all possible networks on \(n\) nodes

We assume that graphs are simple, i.e., \(g_{ii} = 0\) for all \(i \in N\) (no loops) and \(g_{ij} \in [0, 1]\) (no multiple edges).

A network is directed if \(g_{ij} \neq g_{ji}\) for some \(i, j \in N\), and undirected otherwise.

In what follows we consider an unweighted network \(g\) with

\[
  g_{ij} = \begin{cases} 
  1 & \text{if there is a link between } i \text{ and } j \\ 
  0 & \text{otherwise}, 
  \end{cases}
\]

and we assume that \(g_{ij} = g_{ji}\) for all \(i, j \in N\).
Another (equivalent) way of representing a network: 
\((N, g)\), where

- \(g\) is the set of links (subsets of \(N\) of size 2)
- \(ij \in g\) iff \(g_{ij} = 1\)

\[
g = \begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
\end{bmatrix} \iff g = \{12, 23\}
\]

Notation:
- \(g + ij\): network obtained by adding \(ij\) to \(g\)
- \(g - ij\): network obtained by deleting \(ij\) from \(g\)
- \(g' \subset g\) \iff \(\{ij \mid ij \in g'\} \subset \{ij \mid ij \in g\}\)

Given \(S \subset N\) and \(g\), \(g|_S\) denotes \(g\) restricted to \(S\):

\[
[g|_S]_{ij} = \begin{cases} 
1 & \text{if } i \in S, j \in S, g_{ij} = 1 \\
0 & \text{otherwise}, 
\end{cases}
\]

Given \(S \subset N\), \(g^S\) is the complete network on the nodes \(S\) (viewed as network on \(N\)).
N\textsubscript{i}(g) = neighborhood (set of neighbors) of \(i\) in \(g\)

\[ N\textsubscript{i}(g) = \{j \in N : g\textsubscript{ij} = 1\} \]

\(\eta\textsubscript{i}(g)\) = degree of \(i\) in \(g\) = number of \(i\)'s neighbors in \(g\), i.e.,

\[ \eta\textsubscript{i}(g) = |N\textsubscript{i}(g)| \]

A network \(g\) is regular if for some \(\eta \in \{0, 1, ..., n - 1\}\), \(\eta\textsubscript{i}(g) = \eta\) for each \(i \in N\).

\(g^N\) = complete network (regular network with \(\eta = n - 1\))

\(g^\emptyset\) = empty network (regular network with \(\eta = 0\))
How can one node be reached from another one in $g$?

- **Walk** = sequence of links $i_1i_2, \cdots , i_{K−1}i_K$ such that $g_{i_ki_{k+1}} = 1$ for each $k \in \{1, \cdots , K − 1\}$ (a node or a link may appear more than once)
- **Trail** = walk in which all links are distinct
- **Path** = trail in which all nodes are distinct
- **Cycle** = trail with at least 3 nodes in which the initial node and the end node are the same.

- **Geodesic** between two nodes is a shortest path between them.

- **$l_{ij}(g)$** = geodesic distance between $i$ and $j$ in $g$

  If there is a path between $i$ and $j$ in $g$, then

  $$l_{ij}(g) = \text{the number of links in a shortest path between } i \text{ and } j$$

  $$l_{ij}(g) = \min_{\text{paths } P \text{ from } i \text{ to } j} \sum_{kl \in P} g_{kl}.$$  

  If there is no path between $i$ and $j$ in $g$, we set $l_{ij}(g) = \infty$.  

A network is **connected** if there exists a path between any pair of nodes \( i, j \in N \) (i.e., if it consists of a single component).

The **components** of a network are the distinct maximal connected subgraphs.

Two nodes belong to the same **component** if and only if there exists a path between them.

A link \( ij \) is a **bridge** in \( g \) if \( g - ij \) has more components than \( g \).

A **tree** is a connected network that has no cycles.

A **forest** is a network such that each component is a tree.

Any network that has no cycles is a forest.

A **star** is a connected network in which there exists some node \( i \) (center) such that every link in the network involves \( i \).
\( g^k = \text{kth power of } g \) 
\( g^0 := \mathbb{I} \text{ with } \mathbb{I} = n \times n \text{ identity matrix, where} \)
\( g^k_{ij} \) \text{ = number of walks of length } k \text{ that exist between } i \text{ and } j \text{ in } g \).

E.g. walks of length 3 between 1 and 2: (12, 24, 42), (13, 34, 42), (12, 21, 12), (13, 31, 12).

\[
g = \begin{bmatrix}
0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0
\end{bmatrix} \quad g^2 = \begin{bmatrix}
2 & 0 & 0 & 2 \\
0 & 2 & 2 & 0 \\
0 & 2 & 2 & 0 \\
2 & 0 & 0 & 2
\end{bmatrix} \quad g^3 = \begin{bmatrix}
0 & 4 & 4 & 0 \\
4 & 0 & 0 & 4 \\
4 & 0 & 0 & 4 \\
0 & 4 & 4 & 0
\end{bmatrix}
\]
Measuring Networks
While small networks can be easily illustrated, large networks are more difficult to describe.

It is important to be able to compare networks and classify them according to their properties.

Some characteristics of a network:

- Degree distribution
- Diameter and average path length
- Cliquishness and clustering
- Centrality
The **degree distribution of a network** is a description of the relative frequencies of nodes that have different degrees.

$P(\eta) = \text{fraction of nodes that have degree } \eta \text{ under a degree distribution } P$, where $P$ can be a frequency distribution (if describing data) or a probability distribution (for random networks).

E.g., A network is regular of degree $k$ if $P(k) = 1$ and $P(\eta) = 0$ for all $\eta \neq k$.

The **diameter** of a network is the largest distance between any two nodes in the network.

How diameter can vary across networks with (almost) the same number of nodes and links?
Some characteristics of networks (3/5)

- **Average path length** between nodes - the average is taken over geodesics; it is bounded above by the diameter, sometimes can be much shorter than the diameter.
- For networks that are not connected, one often reports the diameter and the average path length in the largest component and specifies if it is a **giant component** (unique largest component, if there is one).
- A **clique** is a maximal completely connected subnetwork (%3 nodes) of a given network.
- **Can a node be part of several cliques? Yes!**
- One measure of cliquishness is to count the number and size of the cliques in a network.
- **The clique structure is very sensitive to slight changes in a network.**
The most common way of measuring some aspect of cliquishness is based on transitive triples or clustering.

The individual clustering for a node $i$

$$Cl_i(g) = \frac{\{jk \in g \mid k \neq j, j \in N_i(g), k \in N_i(g)\}}{\{jk \mid k \neq j, j \in N_i(g), k \in N_i(g)\}}$$

We set $Cl_i(g) = 0$ if $i$ has no more than one link.

$$Cl_i(g) = \frac{\{jk \in g \mid k \neq j, j \in N_i(g), k \in N_i(g)\}}{\eta_i(g)(\eta_i(g) - 1)/2}$$

The overall clustering

$$Cl(g) = \frac{\sum_i \{jk \in g \mid k \neq j, j \in N_i(g), k \in N_i(g)\}}{\sum_i \{jk \mid k \neq j, j \in N_i(g), k \in N_i(g)\}}$$
The average clustering coefficient

\[ Cl^{Avg}(g) = \frac{\sum_i Cl_i(g)}{n} \]

- \( Cl(g) \) and \( Cl^{Avg}(g) \) can be very different.
- Components, cliques, clusters - what a difference?
- See the file with the examples.
Centrality Measures
Given nodes that represent agents (players) and links that represent relationships between the agents (communication, influence, dominance ...), the following questions may appear:

- How central is a node (player) in the network?
- What is his position and prestige?
- How influential is his opinion?
- To which degree is the agent successful and powerful in collective decision making?

Centrality measures can be useful for the analysis of the information flows, bargaining power, infection transmission, influence, etc.

The aim of this part is to present the main (basic) centrality and prestige measures.
Standard measures of centrality

- The concept of centrality captures a kind of prominence of a node in a network.
- Since the late 1940’s a variety of different centrality measures that focus on specific characteristics inherent in prominence of an agent have been developed.
- Measures of centrality can be categorized into the following main groups (Jackson (2008)):
  1. Degree centrality - how connected a node is
  2. Closeness centrality - how easily a node can reach other nodes
  3. Betweenness centrality - how important a node is in terms of connecting other nodes
  4. Prestige- and eigenvector-related centrality - how important, central, or influential a node’s neighbors are.

Degree centrality of a node

- The degree centrality (Shaw (1954), Nieminen (1974)):
  How connected is a node in terms of direct connections?
- The degree centrality $C_i^d(g)$ of node $i$ in network $g$ is given by

$$C_i^d(g) = \frac{\eta_i(g)}{n-1} = \frac{|N_i(g)|}{n-1} \in [0, 1]$$

- Index of the node’s communication activity: the more ability to communicate directly with others, the higher the centrality.

```
C_i^d(g) = 0.5 \text{ for } i \in \{3, 5\}, \quad C_i^d(g) = 0.33 \text{ for } i \notin \{3, 5\}.
```
Degree centrality of a network

Let $i^*$ be a node which attains the highest degree centrality $C_{i^*}^d(g)$ in $g$. The degree centrality $C^d(g)$ of network $g$ is

$$C^d(g) = \frac{\sum_{i=1}^n [C_{i^*}^d(g) - C_i^d(g)]}{\max_{g' \in G} \left[ \sum_{i=1}^n [C_{i^*}^d(g') - C_i^d(g')] \right]} = \frac{\sum_{i=1}^n [C_{i^*}^d(g) - C_i^d(g)]}{n - 2}$$

$C^d(g) = 1$ if $g$ is a star, and $C^d(g) = 0$ if $g$ is a regular network.

$$C_i^d(g) = \frac{1}{2} \text{ for } i \in \{3, 5\}, \quad C_i^d(g) = \frac{1}{3} \text{ for } i \notin \{3, 5\}$$

$$C^d(g) = \frac{1}{5} \cdot 5 \cdot \left( \frac{1}{2} - \frac{1}{3} \right) = \frac{1}{6}$$
Closeness centrality of a node

- The closeness centrality (Beauchamp (1965), Sabidussi (1966)) is based on proximity:
  How easily can a node reach other nodes in a network?
- The closeness centrality $C_i^c(g)$ of node $i$ in network $g$ is

  $$C_i^c(g) = \frac{n - 1}{\sum_{j \neq i} l_{ij}(g)}$$

- Measure of the node’s independence or efficiency: the possibility to communicate with many others depends on a minimum number of intermediaries.

- $C_4^c(g) = 0.60$, $C_3^c(g) = C_5^c(g) = 0.55$, $C_i^d(g) = 0.4$ otherwise.
Let \( i^* \) be a node which attains the highest closeness centrality \( C_{i^*}^c(g) \) in \( g \). The closeness centrality \( C^c(g) \) of network \( g \) is

\[
C^c(g) = \frac{\sum_{i=1}^{n} [C_{i^*}^c(g) - C_i^c(g)]}{\max_{g' \in G} \left[ \sum_{i=1}^{n} [C_{i^*}^c(g') - C_i^c(g')] \right]} = \frac{\sum_{i=1}^{n} [C_{i^*}^c(g) - C_i^c(g)]}{(n - 2)(n - 1)/(2n - 3)}
\]

\( C^c(g) = 1 \) if \( g \) is a star, and \( C^c(g) = 0 \) if \( g \) is a cycle.

\[
C_4^c(g) = 0.60, \quad C_3^c(g) = C_5^c(g) = 0.55 \\
C_i^c(g) = 0.4 \text{ for } i \in \{1, 2, 6, 7\} \\
C^c(g) = 0.33.
\]
We introduce a decay parameter $\delta$, with $0 < \delta < 1$, and consider the proximity between a given node and each other node weighted by the decay.

The decay centrality of node $i$ in network $g$ is

$$C_{i}^{dc}(g, \delta) = \sum_{j \neq i} \delta^{l_{ij}(g)}$$

For $\delta = 0.5$, $C_{i}^{dc}(g, 0.5) = 2$ for $i \in \{3, 4, 5\}$, $C_{i}^{dc}(g, 0.5) = 1.5$ for $i \in \{1, 2, 6, 7\}$. 

\[7\] \[2\] \[5\] \[4\] \[3\] \[6\] \[1\]
Betweenness centrality of a node (1/2)

- The betweenness centrality (Bavelas (1948), Freeman (1977, 1979)):
  How important is a node in terms of connecting other nodes?

- The betweenness centrality $C^b_i(g)$ of node $i$ in network $g$ is

$$C^b_i(g) = \frac{2}{(n-1)(n-2)} \sum_{k \neq j : i \notin \{k, j\}} \frac{P_i(kj)}{P(kj)}$$

$P_i(kj) =$ number of geodesics between $k$ and $j$ containing $i \notin \{k, j\}$

$P(kj) =$ total number of geodesics between $k$ and $j$

- Index of the potential of a node for control of communication: the possibility to intermediate in the communications of others is of importance.

- If $g$ is a star, then $C^b_i(g) = 1$ for $i$ being the center and $C^b_i(g) = 0$ otherwise.
Betweenness centrality of a node (2/2)

\[
C^b_i(g) = \frac{2}{(n - 1)(n - 2)} \sum_{k \neq j: i \notin \{k, j\}} \frac{P_i(kj)}{P(kj)}
\]

\[
\begin{align*}
C^b_4(g) &= 0.60 \\
C^b_3(g) &= C^b_5(g) = 0.53 \\
C^b_i(g) &= 0 \text{ for } i \in \{1, 2, 6, 7\}
\end{align*}
\]
Let \( i^* \) be a node which attains the highest betweenness centrality \( C_{i^*}^b(g) \) in \( g \).

The betweenness centrality \( C^b(g) \) of network \( g \) is

\[
C^b(g) = \frac{\sum_{i=1}^n [C_{i^*}^b(g) - C_i^b(g)]}{n - 1}
\]

\[
C_4^b(g) = 0.60, \quad C_3^b(g) = C_5^b(g) = 0.53
\]

\[
C_i^b(g) = 0 \text{ for } i \in \{1, 2, 6, 7\}
\]

\[
C^b(g) = 0.42.
\]
Katz prestige

- Measures of centrality that are based on the idea that a node’s importance is determined by the importance of its neighbors.
- The Katz prestige $C_{i}^{PK}(g)$ of node $i$ in $g$ is defined as

$$C_{i}^{PK}(g) = \sum_{j \neq i} g_{ij} \frac{C_{j}^{PK}(g)}{\eta_{j}(g)}$$

If $j$ has more relationships, then $i$ gets less prestige from being connected to $j$. This definition is self-referential.

- Calculating $C^{PK}(g)$ - finding the unit eigenvector of $\tilde{g}$:

$$C^{PK}(g) = \tilde{g} C^{PK}(g)$$

$$(\mathbb{I} - \tilde{g}) C^{PK}(g) = \mathbf{0}$$

$\tilde{g}$ - the normalized adjacency matrix $g$ with $\tilde{g}_{ij} = \frac{g_{ij}}{\eta_{j}(g)}$, we set $\tilde{g}_{ij} = 0$ for $\eta_{j}(g) = 0$.

$C^{PK}(g)$ - the $n \times 1$ vector of $C_{i}^{PK}(g), i \in N$,

$\mathbb{I}$ - the $n \times n$ identity matrix, $\mathbf{0}$ - the $n \times 1$ vector of 0’s.
Eigenvector centrality

- If we do not normalize $g$, we get the eigenvector centrality $C^e(g)$ associated with $g$ (Bonacich (1972)).
- The centrality of a node is proportional to the sum of the centrality of its neighbors.

$$\lambda C^e_i(g) = \sum_j g_{ij} C^e_j(g)$$

$$\lambda C^e(g) = gC^e(g)$$

and thus $C^e(g)$ is an eigenvector of $g$ and $\lambda$ is the corresponding largest eigenvalue of matrix $g$.

- The Katz prestige can be seen as a kind of eigenvector centrality with the network adjacency matrix being weighted.
Second prestige measure of Katz

- $C^{PK2}(g, a) =$ the second prestige measure of Katz (1953)
- Introducing an attenuation parameter $a$ to adjust the measure for the lower ‘effectiveness’ of longer walks in a network.
- The prestige of a node is a weighted sum of the walks that emanate from it, and a walk of length $k$ is of worth $a^k$, where $0 < a < 1$. The vector of prestige of nodes is

$$C^{PK2}(g, a) = ag1 + a^2g^21 + \cdots + a^kg^k1 + \cdots$$

where $1$ is the $n \times 1$ vector of 1’s.
- Each entry of the vector $g^k1$ is the total number of walks of length $k$ that emanate from each node, and $g1$ is simply the vector of degrees of nodes.
- For $a$ sufficiently small, $C^{PK2}(g, a)$ is finite and

$$C^{PK2}(g, a) - agC^{PK2}(g, a) = ag1$$

$$C^{PK2}(g, a) = (I - ag)^{-1}ag1.$$
Bonacich centrality

- A two-parameter family of prestige measures which can be seen as a direct extension of $C^{PK2}(g, a)$.
- An agent can have some status which does not depend on its connections to others.
- Bonacich centrality (Bonacich (1987)) is given by
  \[ C^B(g, a, b) = ag1 + abg^21 + \cdots + ab^kg^{k+1}1 + \cdots \]
  \[ C^B(g, a, b) = (I - bg)^{-1} ag1 \]
  where $a$ and $b$ are parameters, and $b$ is sufficiently small.
  - $b$ captures how the value of being connected to another node decays with distance.
  - $a$ captures the base value on each node.
  - For $b = 0$, $C^B(g, a, b)$ takes into account only walks of length 1 and reduces to $ad_i(g)$.
  - For $b > 0$, $C^B(g, a, b)$ takes into account more distant interactions.
  - Obviously $C^{PK2}(g, a)$ and $C^B(g, a, b)$ coincide when $a = b$. 
**Example (ctd)**

- **Centrality measures**
- **Nodes** → 1,2,6,7 | 3,5 | 4

<table>
<thead>
<tr>
<th>Centrality measures</th>
<th>1,2,6,7</th>
<th>3,5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree, Katz prestige</td>
<td>0.33</td>
<td>0.50</td>
<td>0.33</td>
</tr>
<tr>
<td>Closeness</td>
<td>0.40</td>
<td>0.55</td>
<td>0.60</td>
</tr>
<tr>
<td>Decay centrality, $\delta = 0.5$</td>
<td>1.5</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Decay centrality, $\delta = 0.75$</td>
<td>3.1</td>
<td>3.7</td>
<td>3.8</td>
</tr>
<tr>
<td>Decay centrality, $\delta = 0.25$</td>
<td>0.59</td>
<td>0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>Betweenness</td>
<td>0</td>
<td>0.53</td>
<td>0.60</td>
</tr>
<tr>
<td>Eigenvector centrality</td>
<td>0.47</td>
<td>0.63</td>
<td>0.54</td>
</tr>
<tr>
<td>Second Katz prestige, $a = 1/3$</td>
<td>3.1</td>
<td>4.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Bonacich centrality, $a = 1$, $b = 1/3$</td>
<td>9.4</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Bonacich centrality, $a = 1$, $b = 1/4$</td>
<td>4.9</td>
<td>6.8</td>
<td>5.4</td>
</tr>
</tbody>
</table>
In this example, the most central are
- either agent 4
- or agents 3 and 5.

Different centrality measures capture different aspects of centrality, and therefore can have highest values for different individuals.