Psychological Networks Summer School
Day 2, part 3: Descriptive Network Analysis

Sacha Epskamp

University of Amsterdam
Department of Psychological Methods

05-07-2016
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Connectivity</th>
<th>Clustering</th>
<th>Centrality</th>
<th>Weighted and Directed networks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prevent the outbreak:

http://vax.herokuapp.com/
• How fast will the disease spread?
  • Connectivity
• What nodes should be vaccinated?
  • Centrality
• Which parts of the network should be quarantined?
  • Clustering
We will first analyze only the unweighted, undirected simple graph $G$:

$$G = (V, E)$$

With $|V|$ nodes and $|E|$ edges, encoded using $|V| \times |V|$ adjacency matrix $A$:

$$a_{uv} = a_{vu} = \begin{cases} 1 & \text{if } \{u, v\} \in E \\ 0 & \text{Otherwise} \end{cases}$$
Connectivity
The shortest path length between nodes $v$ and $u$, $\text{dist}(v, u)$, is defined in an unweighted graph as the minimum number of steps you need to take from node $v$ to reach node $u$:

$$\text{dist}(v, u) = \min (a_{vx} + \ldots + a_{yu})$$

- Can be computed using Dijkstra’s algorithm (Dijkstra, 1959) with weights fixed to 1.
- Commonly referred to as the shortest path length or geodesic distance

The mean shortest path length is called the average shortest path length (APL) and is an important measure for how well connected a graph is:

$$\text{APL}(G) = \frac{\sum_{v, u} \text{dist}(v, u)}{|V| (|V| - 1)/2}$$
### Shortest paths (average: 2):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
- Disorders usually first diagnosed in infancy, childhood or adolescence
- Delirium, dementia, and amnesia and other cognitive disorders
- Mental disorders due to a general medical condition
- Substance-related disorders
- Schizophrenia and other psychotic disorders
- Mood disorders
- Anxiety disorders
- Somatoform disorders
- Factitious disorders
- Dissociative disorders
- Sexual and gender identity disorders
- Eating disorders
- Sleep disorders
- Impulse control disorders not elsewhere classified
- Adjustment disorders
- Personality disorders
- Symptom is featured equally in multiple chapters
Clustering
Are two connected nodes also connected to each other? Or more general, does a graph exhibit cliques?
The **local clustering coefficient**, $c_l(v)$, gives for node $v$ the proportion of neighbors of $v$ that are also connected to each other. This corresponds for dividing the number of “triangles” of which node $v$ is part, $\tau_\Delta(v)$ to the number of possible triangles of which $v$ could be part: $\tau_3(v)$:

$$c_l(v) = \frac{\tau_\Delta(v)}{\tau_3(v)}$$

$\tau_3(v)$ also corresponds to the number of *triplets* in which node $v$ is the middle node.
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Connectivity</th>
<th>Clustering</th>
<th>Centrality</th>
<th>Weighted and Directed networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐</td>
<td>☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐</td>
</tr>
</tbody>
</table>

The diagram illustrates a network with nodes and edges. Each node represents a point of interest, and the edges between nodes indicate connections or relationships. The clustering aspect of network analysis is highlighted by the groupings of nodes, suggesting communities or clusters within the network. The centrality measure is represented by the prominence of certain nodes, which play a key role in the network's structure and function. Weighted and directed networks consider the strength and direction of connections, which are not explicitly shown in the visual representation but are implied in the context of network analysis.
• The local clustering coefficient can also be seen as a measure for redundancy
• A node with clustering of 1 has connected neighbors. Deleting this node will not hugely change the structure of a network
Traditionally, a **global clustering coefficient** for the whole graph can be obtained by averaging all the local clustering coefficients:

\[
cl(G) = \frac{1}{|V|} \sum_{i=1}^{\frac{|V|}{2}} cl(i)
\]

This is an average of averages, and should be properly weighted to obtain a more informative coefficient:

\[
cl_\tau(G) = \frac{\sum_{i=1}^{\frac{|V|}{2}} \tau_\Delta(i)cl(i)}{\tau_3(i)} = \frac{3\tau_\Delta(G)}{\tau_3(G)}
\]

This is the more modern clustering coefficient, also termed **transitivity**
Global clustering: 0.79; Transitivity: 0.94
Small-world Networks

The famous paper of Watts and Strogatz (1998)—already cited 23623 times—describes the “small world” principle that frequently occurs in natural graphs.

• “Six degrees of separation”
• High clustering and low average path length
A graph exhibits a **small world structure** if it has a much higher clustering than a random graph of the same dimensions while still having a low APL.

The *small world index* captures this in a single number. Higher than 3 is typically interpreted as a small world
Introduction

Centrality

Clustering

Connectivity

Weighted and Directed networks

● Disorders usually first diagnosed in infancy, childhood or adolescence
● Delirium, dementia, and amnesia and other cognitive disorders
● Mental disorders due to a general medical condition
● Substance-related disorders
● Schizophrenia and other psychotic disorders
● Mood disorders
● Anxiety disorders
● Somatoform disorders
● Factitious disorders
● Dissociative disorders
● Sexual and gender identity disorders
● Eating disorders
● Sleep disorders
● Impulse control disorders not elsewhere classified
● Adjustment disorders
● Personality disorders
● Symptom is featured equally in multiple chapters

Symptom is featured equally in multiple chapters
Adjustment disorders
Personality disorders
Sleep disorders
Eating disorders
Sexual and gender identity disorders
Dissociative disorders
Factitious disorders
Somatoform disorders
Mood disorders
Anxiety disorders
Delirium, dementia, and amnesia and other cognitive disorders
Mental disorders due to a general medical condition
Substance-related disorders
Schizophrenia and other psychotic disorders
Eating disorders
Sleep disorders
Impulse control disorders not elsewhere classified
Adjustment disorders
Personality disorders
Symptom is featured equally in multiple chapters
Centrality
Centrality measures assign numeric values to the importance of nodes in the graph and answer the question “what is the most central node?”. 

- Degree 
- Closeness 
- Betweenness
Introduction

- Centrality
- Connectivity
- Clustering

Weighted and Directed networks

- Disorders usually first diagnosed in infancy, childhood or adolescence
- Delirium, dementia, and amnesia and other cognitive disorders
- Mental disorders due to a general medical condition
- Substance-related disorders
- Schizophrenia and other psychotic disorders
- Mood disorders
- Anxiety disorders
- Somatoform disorders
- Factitious disorders
- Dissociative disorders
- Sexual and gender identity disorders
- Eating disorders
- Sleep disorders
- Impulse control disorders not elsewhere classified
- Adjustment disorders
- Personality disorders
- Symptom is featured equally in multiple chapters
The **degree** of node $v$, $C_D(v)$ is simply the number of edges connected to node $v$, which we can compute by either summing over row $v$ or column $v$ of $A$:

$$C_D(v) = \sum_{i=1}^{\mid V \mid} a_{iv} = \sum_{j=1}^{\mid V \mid} a_{vj}$$
High School dating
Closeness $C_C(v)$ defines that a node is central if it is ‘close’ to other nodes. This can be computed by taking the inverse of the sum of all path lengths going from node $v$ to all other nodes:

\[
C_C(v) = \frac{1}{\sum_{i=1}^{\mid V \mid} \text{dist}(v, i)}
\]

This is only an interesting measure for fully connected graphs or components.
Betweenness of node $v$ is defined as the sum of proportions of the number of shortest paths between all pairs of nodes that go through node $v$:

$$C_B(i) = \sum_{i \neq j \neq k \in V}^{n} \frac{\sigma(i, j \mid v)}{\sigma(i, j)}$$

Where $\sigma(i, j)$ is the total number of shortest paths between any two nodes and $\sigma(i, j \mid v)$ the number of those paths that go through $v$. 
Centrality recap

**Degree**  How well connected is a node?

**Closeness**  How easy is it to reach all other nodes from a node?

**Betweenness**  How well does a node connect other nodes?
Betweenness

is often touchy or...

often easily distr...

insomnia / difficu...

tachycardia / acce...

tremors

anxiety

depressed mood

psychomotor retard...

psychomotor agitat...

increased appetite

flat or inappropriate

increased appetite

psychomotor retardation

psychomotor agitation

tachycardia / acceleration

tremors

anxiety

depressed mood

psychomotor retardation

psychomotor agitation

tachycardia / acceleration

tremors
Fig. 4.4 Illustration of (b) closeness, (c) betweenness, and (d) eigenvector centrality measures on the graph in (a). Example and figures courtesy of Ulrik Brandes.
Weighted and Directed networks
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
We define the length of an edge from node $i$ to node $j$ as the inverse of the absolute weight:

$$l_{ij} = \begin{cases} \frac{1}{|w_{ij}|} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

Because the denominator is always positive, we will take the limit in the case of a weight of 0:

$$\lim_{w_{ij} \to 0} \frac{1}{|w_{ij}|} = \infty$$
\[ W = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \]

\[ L = \begin{bmatrix} 0 & 1 & 1/2 \\ \infty & 0 & 1/3 \\ \infty & \infty & 0 \end{bmatrix} \]
The weighted version of degree is called \textit{strength}, which simply sums the absolute weights of edges connected to a node. In directed networks, we can distinguish between in-strength and out-strength:

\[
\begin{align*}
    s_{i}^{(in)} &= \sum_{j=1}^{n} |w_{ji}| \\
    s_{i}^{(out)} &= \sum_{j=1}^{n} |w_{ij}|
\end{align*}
\]
The shortest path length between nodes $i$ and $j$, $d(i, j)$ is defined in an unweighted graph as the minimum number of steps you need to take from node $i$ to node $j$:

$$d(i, j) = \min (a_{ih} + \ldots + a_{hj})$$

which can be obtained through Dijkstra's algorithm (Dijkstra, 1959) with weights fixed to 1. Similarly, for weighted graphs the shortest path length is defined as the minimum number of distance needed to cross on the graph to reach node $i$ from node $j$:

$$d(i, j) = \min \left( \frac{1}{|w_{ih}|} + \ldots + \frac{1}{|w_{hj}|} \right) = \min (l_{ih} + \ldots + l_{hj})$$

Closeness and Betweenness naturally follow using this definition of shortest path length.
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
What is the most central node?
How to win VAX

- Delete nodes that have
  - High centrality
  - Low clustering
- Reduce small-worldness
  - Increase the average shortest path length
Thank you for your attention!