Psychological Networks Summer School

Day 2, part 3: Descriptive Network Analysis

Sacha Epskamp

University of Amsterdam
Department of Psychological Methods

05-07-2016

Prevent the outbreak:
`http://vax.herokuapp.com/`

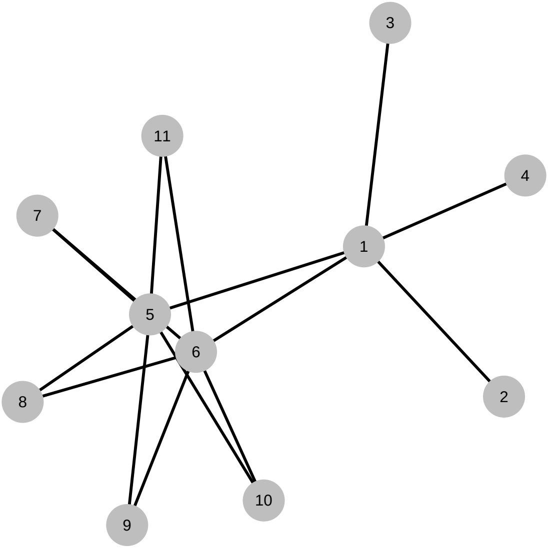
- How fast will the disease spread?
  - Connectivity
- What nodes should be vaccinated?
  - Centrality
- Which parts of the network should be quarantined?
  - Clustering

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 $\boldsymbol{A}$:

$$a_{uv} = a_{vu} = \begin{cases} 1 & \text{if } \{u, v\} \in E \\ 0 & \text{Otherwise} \end{cases}$$


3
11
4
7
1
5
6
8
2
10
9

Connectivity

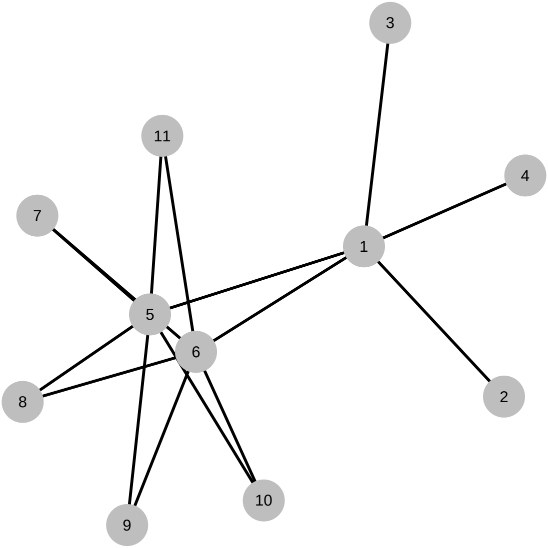
The shortest path length between nodes v and u, $\mathrm{dist}(v, u)$, is defined in an unweighted graph as the minimum number of steps you need to take from node v to reach node u:

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

- 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

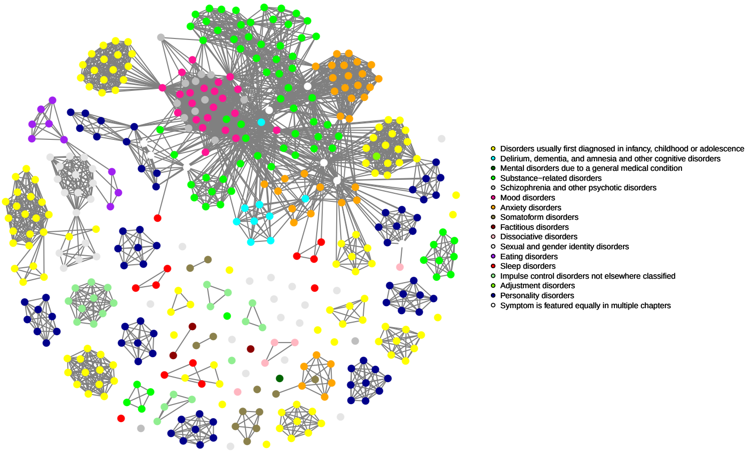
The mean shortest path length is called the average shortest path length (APL) and is an important measure for how well connected a graph is:

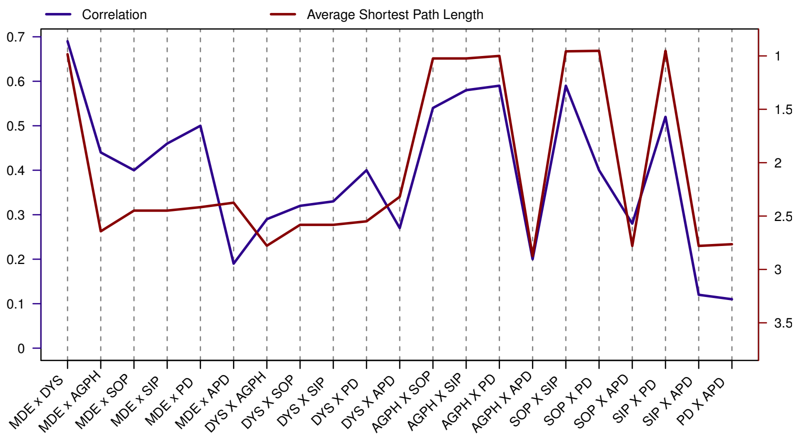
$$\mathrm{APL}(G) = \frac{\sum_{v,u} \mathrm{dist}(v, u)}{|V|\left(|V| - 1\right)/2}$$


3
11
4
7
1
5
6
8
2
10
9

Shortest paths (average: 2):

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1	1	1	1	2	2	2	2	2
2	1	0	2	2	2	2	3	3	3	3	3
3	1	2	0	2	2	2	3	3	3	3	3
4	1	2	2	0	2	2	3	3	3	3	3
5	1	2	2	2	0	2	1	1	1	1	1
6	1	2	2	2	2	0	1	1	1	1	1
7	2	3	3	3	1	1	0	2	2	2	2
8	2	3	3	3	1	1	2	0	2	2	2
9	2	3	3	3	1	1	2	2	0	2	2
10	2	3	3	3	1	1	2	2	2	0	2
11	2	3	3	3	1	1	2	2	2	2	0


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


Correlation
Average Shortest Path Length
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
1.5
2
2.5
3
3.5
MDE x DYS
MDE x AGPH
MDE x SOP
MDE x SIP
MDE x PD
MDE x APD
DYS X AGPH
DYS X SOP
DYS X SIP
DYS X PD
DYS X APD
AGPH X SOP
AGPH X SIP
AGPH X PD
AGPH X APD
SOP X SIP
SOP X PD
SOP X APD
SIP X PD
SIP X APD
PD X APD

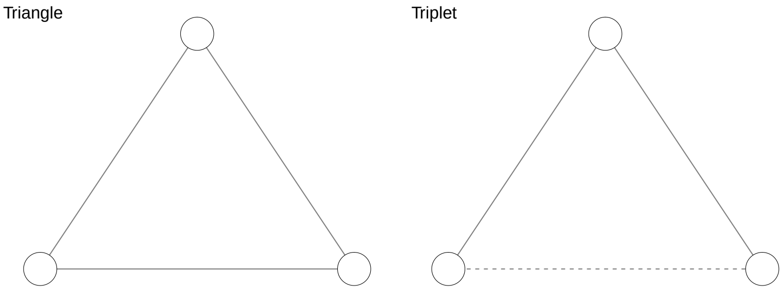
Clustering

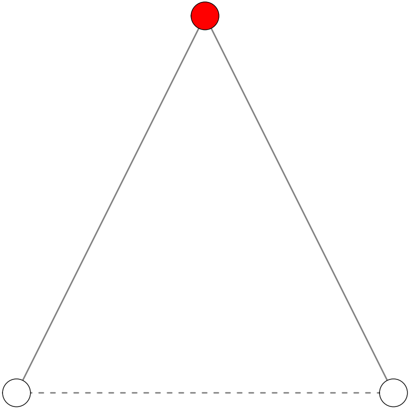
Are two connected nodes also connected to each other? Or more general, does a graph exhibit cliques?

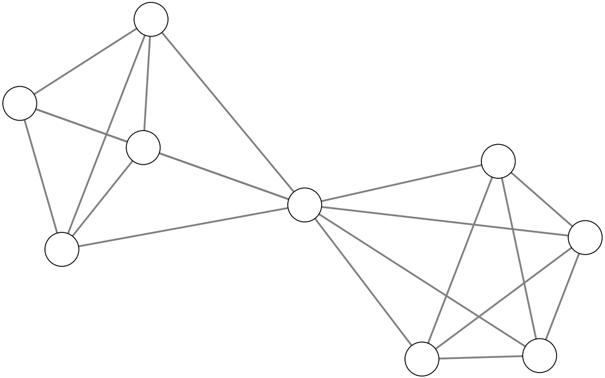
The local clustering coefficient, $\mathrm{cl}(v)$, gives for node v the proportion of neighbors of v that are also connected to each other.
This corresponds for dividing the number of “triangles” of which node v is part, $\tau_{\Delta}(v)$ to the number of possible triangles of which v could be part: $\tau_3(v)$:

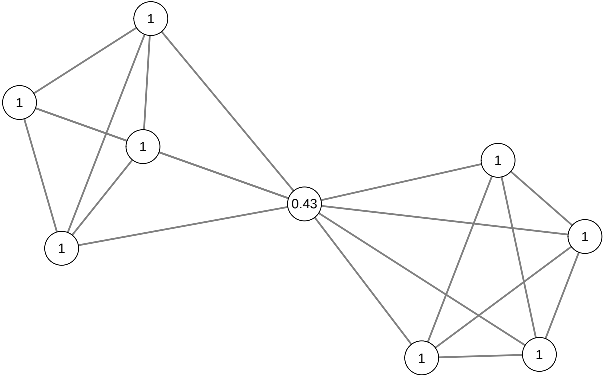
$$\mathrm{cl}(v) = \frac{\tau_{\Delta}(v)}{\tau_3(v)}$$

$\tau_3(v)$ also corresponds to the number of *triplets* in which node v is the middle node.


Triangle
Triplet






1
1
1
0.43
1
1
1
1
1

- 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

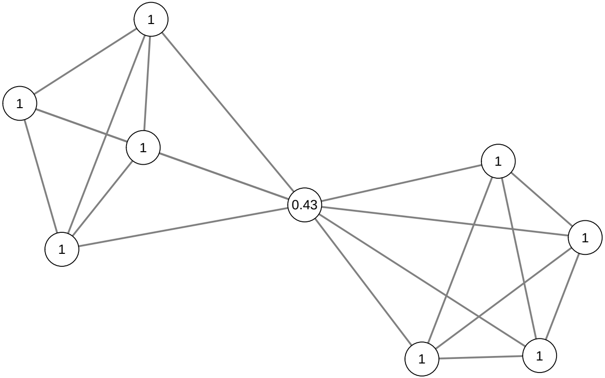
Traditionally, a global clustering coefficient for the whole graph can be obtained by averaging all the local clustering coefficients:

$$\mathrm{cl}(G) = \frac{1}{|V|} \sum_{i=1}^{|V|} \mathrm{cl}(i)$$

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

$$\begin{aligned}
\mathrm{cl}_T(G) &= \frac{\sum_{i=1}^{|V|} \tau_\Delta(i)\mathrm{cl}(i)}{\tau_3(i)} \\
&= \frac{3\tau_\Delta(G)}{\tau_3(G)}
\end{aligned}$$

This is the more modern clustering coefficient, also termed transitivity

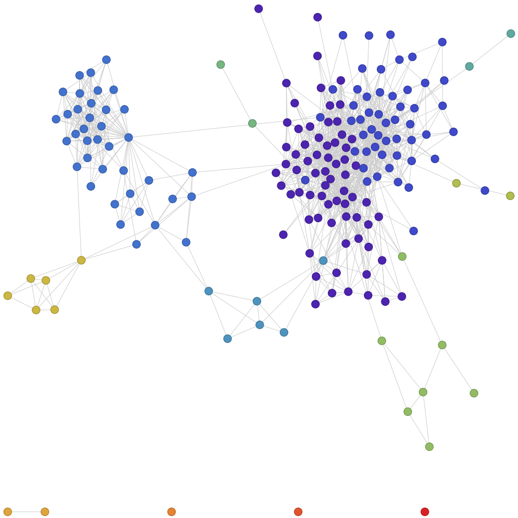


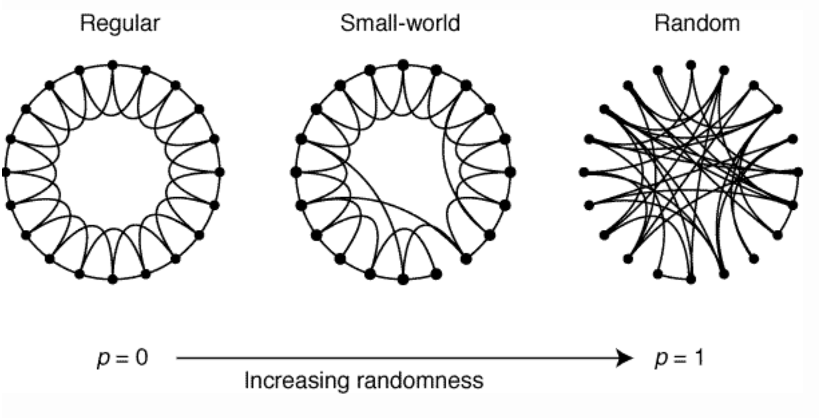
Global clustering: 0.79; Transitivity: 0.94

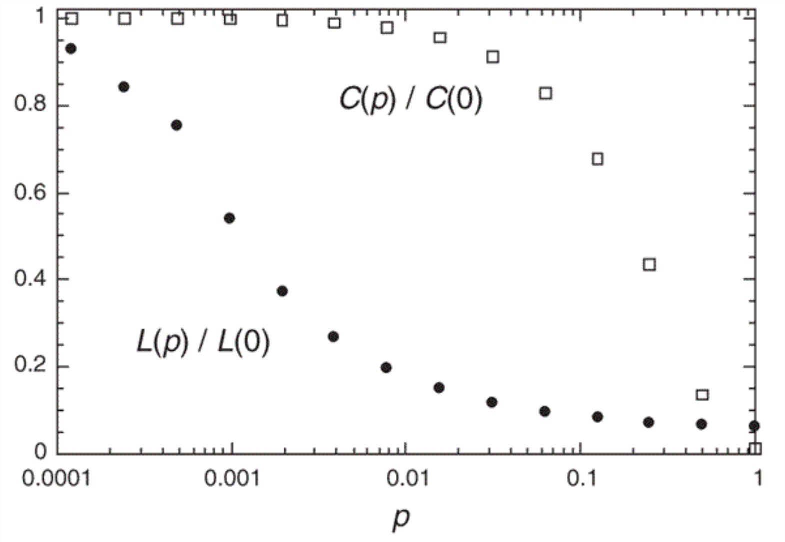
Small-world Networks

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

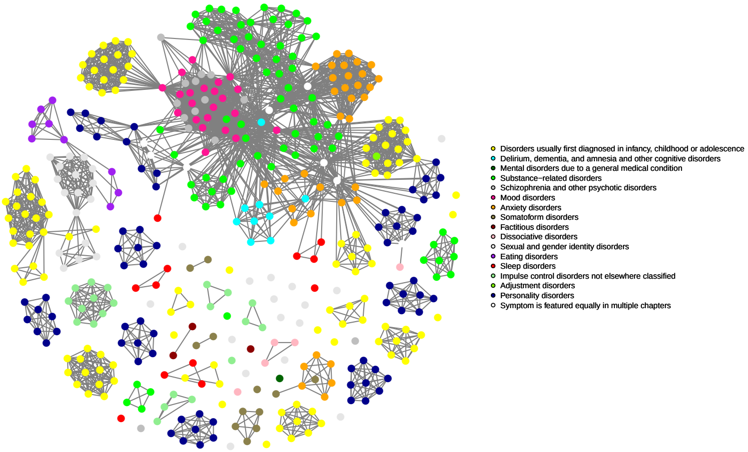



Regular
Small-world
Random
$p = 0$
Increasing randomness
$p = 1$


C(p) / C(0)
L(p) / L(0)
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1
p

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.

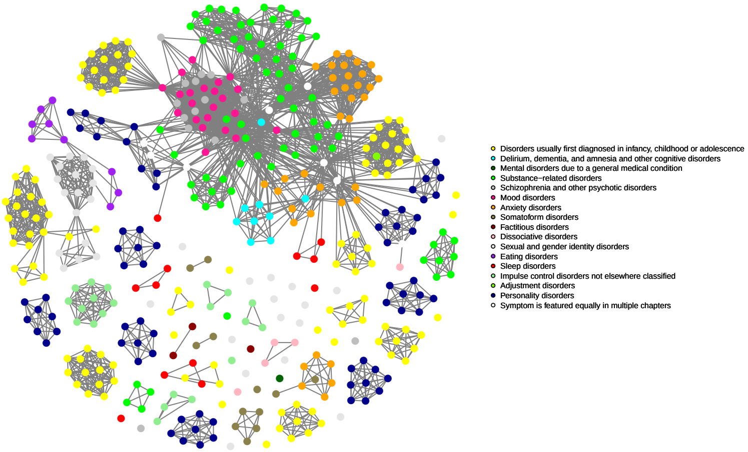
The *small world index* captures this in a single number. Higher than 3 is typically interpreted as a small world


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

Centrality

Centrality measures assign numeric values to the importance of nodes in the graph and answer the question “what is the most central node?”.

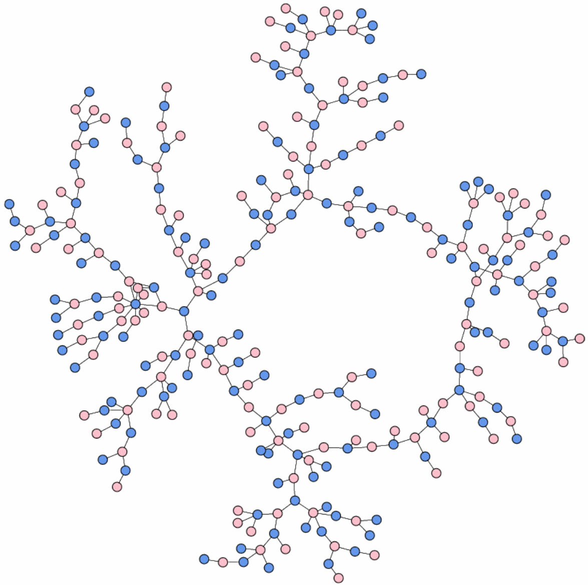
- Degree
- Closeness
- Betweenness


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 $\boldsymbol{A}$:

$$C_D(v) = \sum_{i=1}^{|V|} a_{iv} = \sum_{j=1}^{|V|} a_{vj}$$

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:

$$C_C(v) = \frac{1}{\sum_{i=1}^{|V|} \mathrm{dist}(v, i)}$$

This is only an interesting measure for fully connected graphs or components.

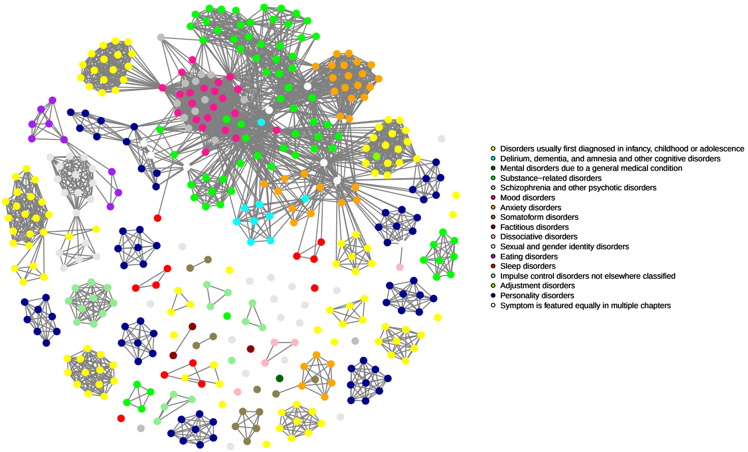
Betweenness of node v is defined as the sum of proportions of the number of shortest paths between all pairs of nodes that go through node v:

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

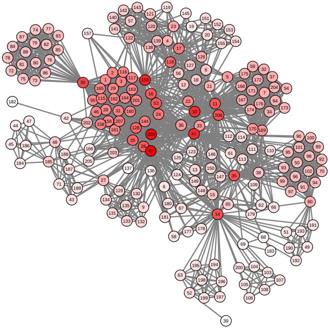
Where $\sigma(i, j)$ is the total number of shortest paths between any two nodes and $\sigma(i, j \mid v)$ the number of those paths that go through v.

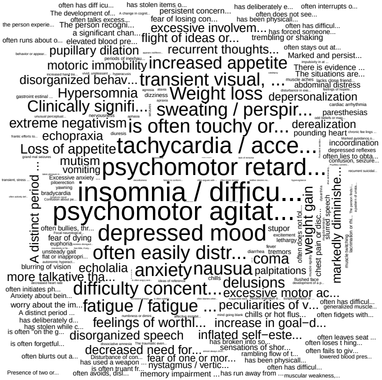
Centrality recap

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?


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

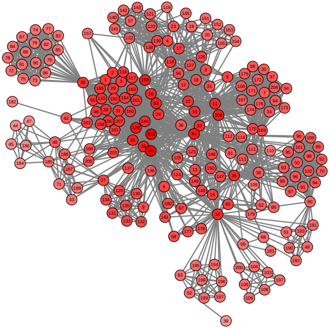
Degree





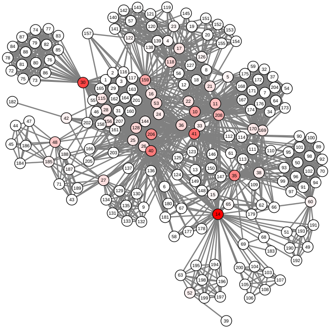
insomnia / difficu...
psychomotor agitat...
psychomotor retard...
tachycardia / acce...
depressed mood
often easily distr...
is often touchy or...
transient visual, ...
Weight loss
sweating / perspir...
increased appetite
Clinically signifi...
difficulty concent...
fatigue / fatigue ...
anxiety
nausua
feelings of worthl...
increase in goal−d...
disorganized speech
inflated self−este...
decreased need for...
peculiarities of v...
excessive motor ac...
delusions
coma
echolalia
more talkative tha...
extreme negativism
Loss of appetite
mutism
vomiting
echopraxia
Hypersomnia
disorganized behav...
motoric immobility
pupillary dilation
recurrent thoughts...
flight of ideas or...
excessive involvem...
depersonalization
derealization
abdominal distress
palpitations
markedly diminishe...
weight gain
A distinct period ...
fear of one or mor...
nystagmus / vertic...

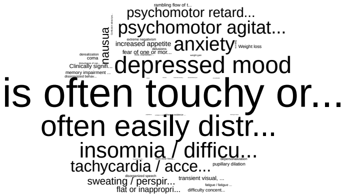
Closeness




echopraxia
extreme negativism
palpitations
pupillary dilation
excitement
Hypersomnia
vomiting
yawning
ataxia
delusions
Weight loss
diarrhea
stupor
mutism
difficulty concent...
chills
weight gain
nausua
fever
tremors
depressed mood
anxiety
increased appetite
insomnia / difficu...
psychomotor agitat...
psychomotor retard...
transient visual, ...
fatigue / fatigue ...
sweating / perspir...
is often touchy or...
often easily distr...
echolalia
disorganized speech
incoordination
unexpected travel ...
flushed face
feelings of worthl...
muscle aches
coma
lethargy
tachycardia / acce...
disorganized behav...
muscular weakness,...

Betweenness





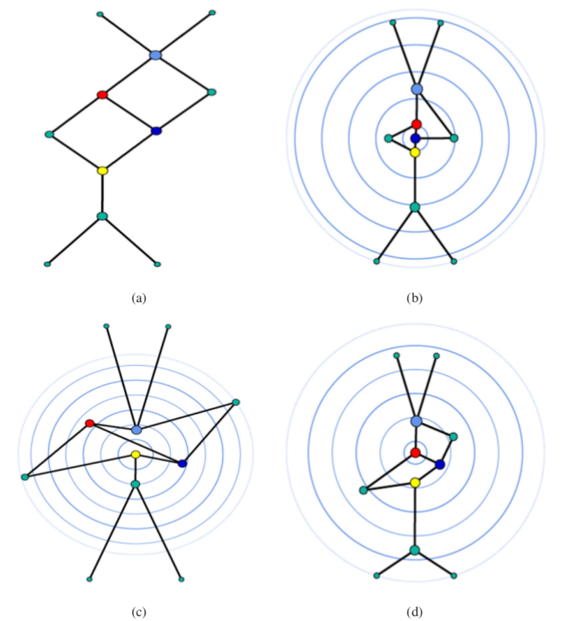
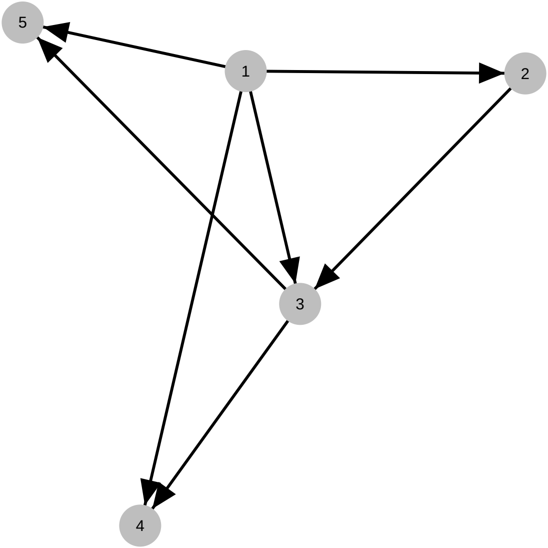


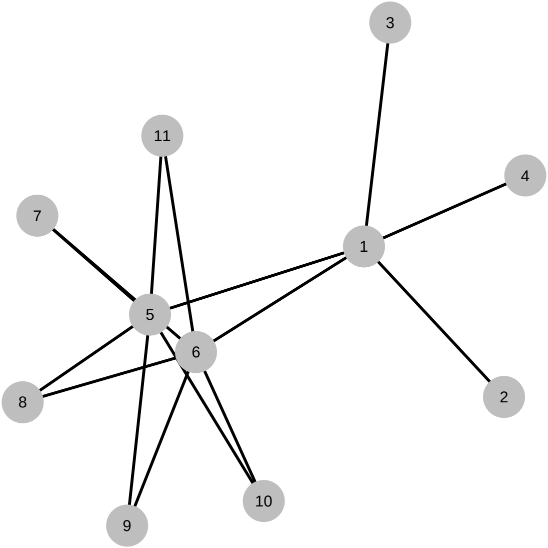
Fig. 4.4 Illustration of (b) closeness, (c) betweenness, and (d) eigenvector centrality measures on the graph in (a). Example and figures courtesy of Ulrik Brandes.

Weighted and Directed networks

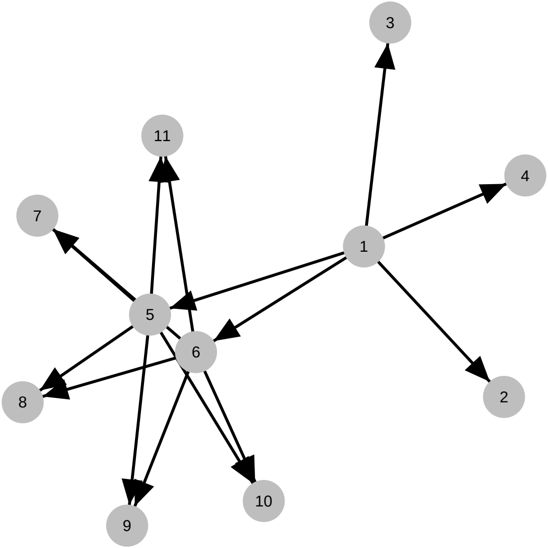
What is the most central node?



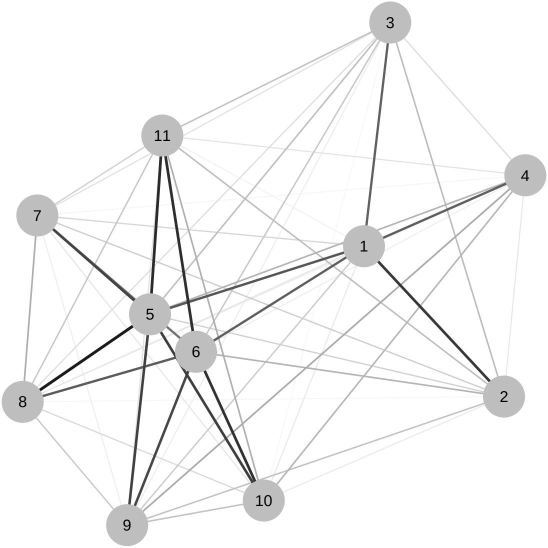
What is the most central node?



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:

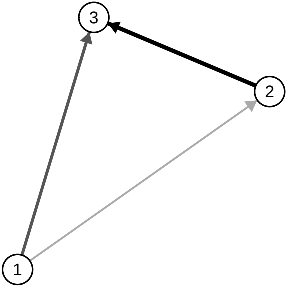
$$l_{ij} = \begin{cases} \frac{1}{|w_{ij}|} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

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

$$\lim_{w_{ij} \to 0} \frac{1}{|w_{ij}|} = \infty$$

$$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{L} = \begin{bmatrix} 0 & 1 & 1/2 \\ \infty & 0 & 1/3 \\ \infty & \infty & 0 \end{bmatrix}$$



The weigthed version of degree is called 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:

$$s_i^{(in)} = \sum_{j=1}^{n} |w_{ji}|$$

$$s_i^{(out)} = \sum_{j=1}^{n} |w_{ij}|$$

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:

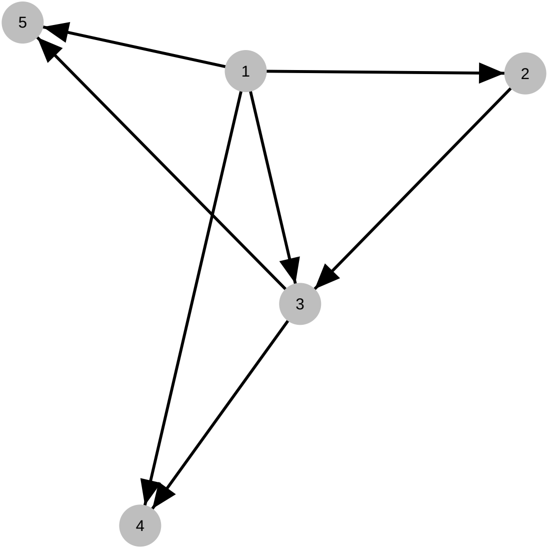
$$d(i,j) = \min\left(a_{ih} + \ldots + a_{hj}\right)$$

which can be obtained through Dijkstra's algorithm (Dijkstra, 1959) with weights fixed to 1. Similarity, 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:

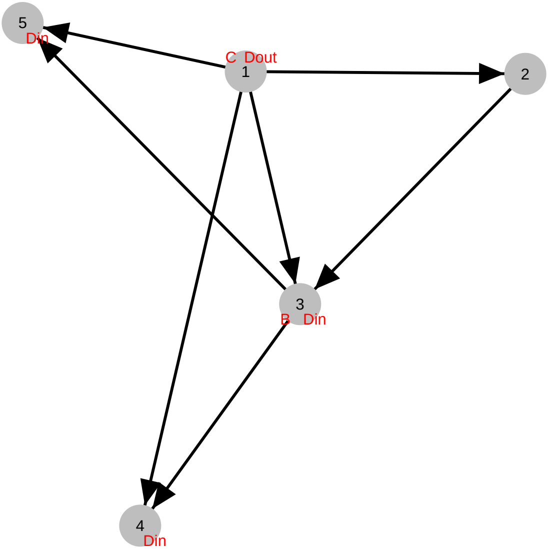
$$d(i,j) = \min\left(\frac{1}{|w_{ih}|} + \ldots + \frac{1}{|w_{hj}|}\right) = \min\left(l_{ih} + \ldots + l_{hj}\right)$$

Closeness and Betweenness naturally follow using this definition of shortest path length

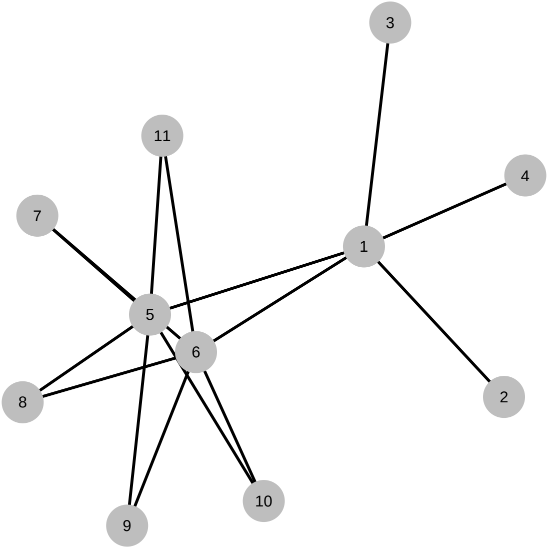
What is the most central node?



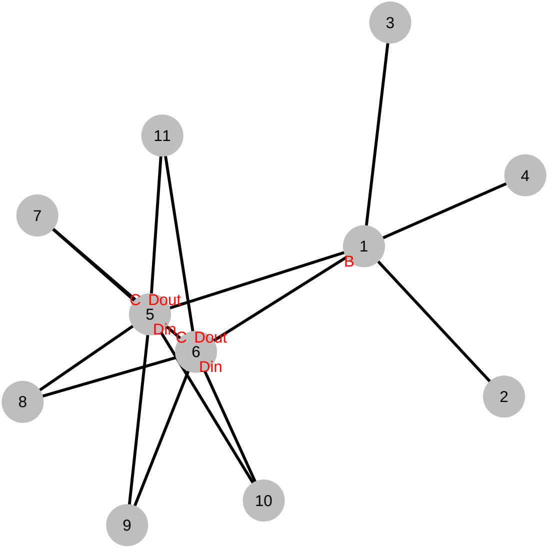
What is the most central node?



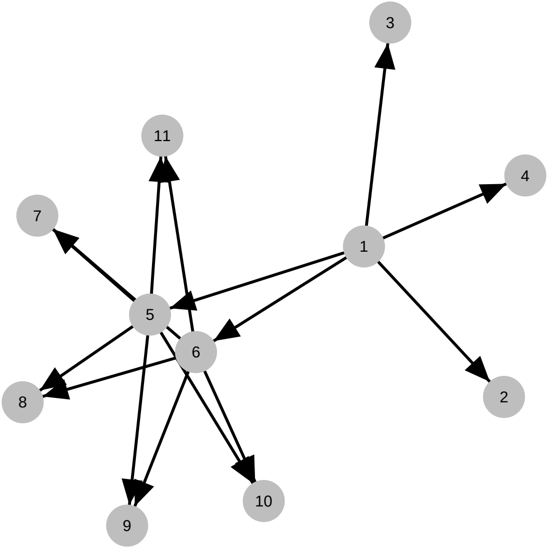
What is the most central node?



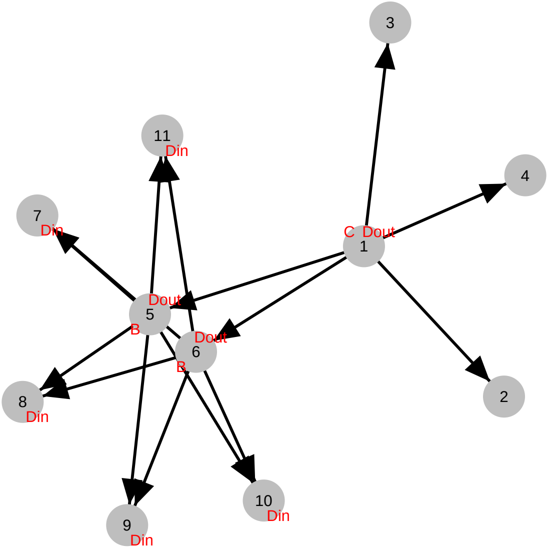
What is the most central node?



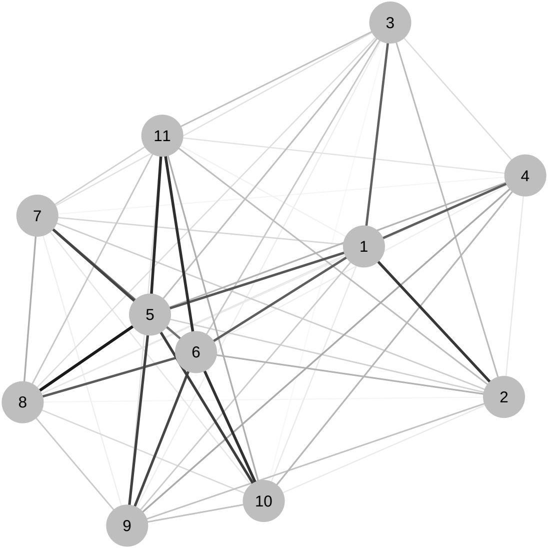
What is the most central node?



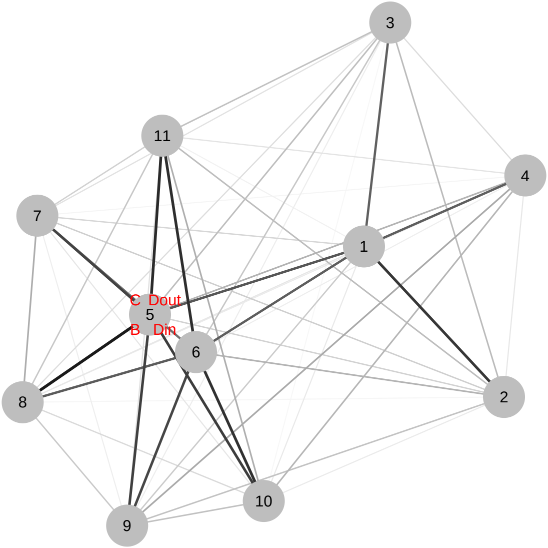
What is the most central node?



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!

References I

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, *1*(1), 269–271.

Watts, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, *393*(6684), 440–442.