BI

Nordic Centre
for Internet and Society

Visualisations and Network Theory

Network Statistics and Measures

Lecture 3

February 3rd and February 4th, 2020

Agenda

Metrics and Measures
Global and Individual

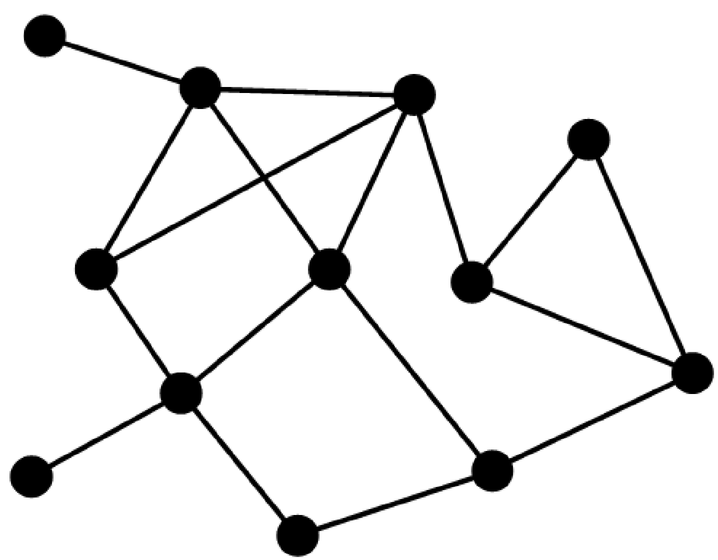
Exercise
R igraph Tutorial 1

1

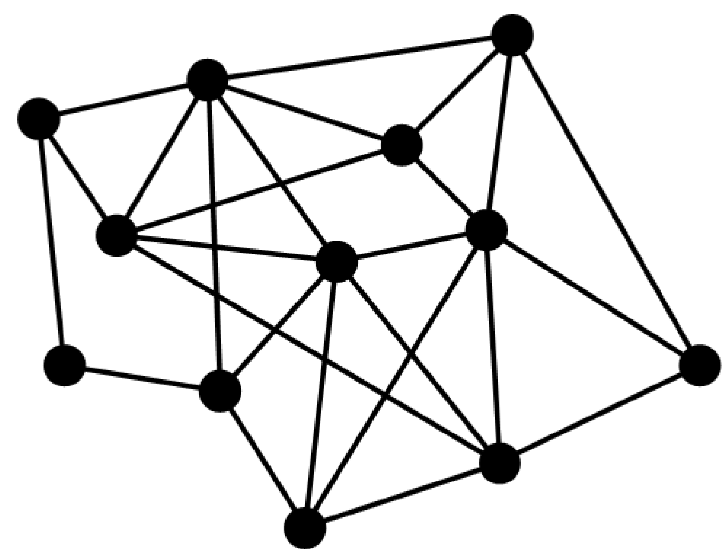
Metrics and Measures
Whole Networks

Density

- Number of ties, expressed as percentage of the total number of possible ties



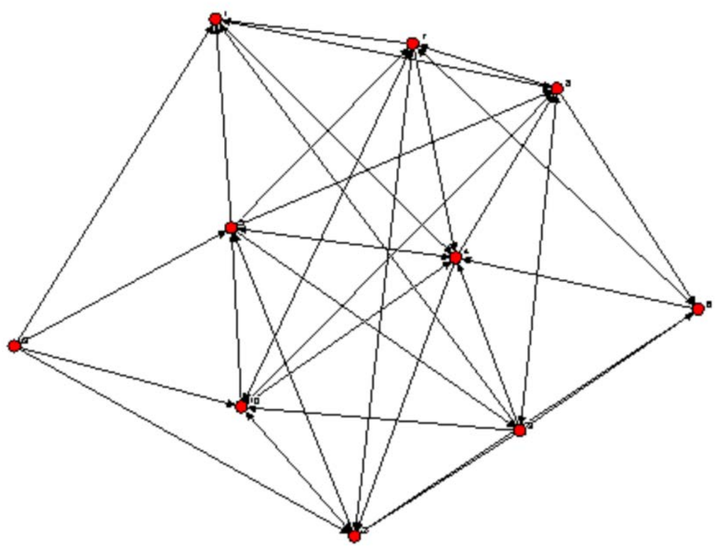
low density: 25%



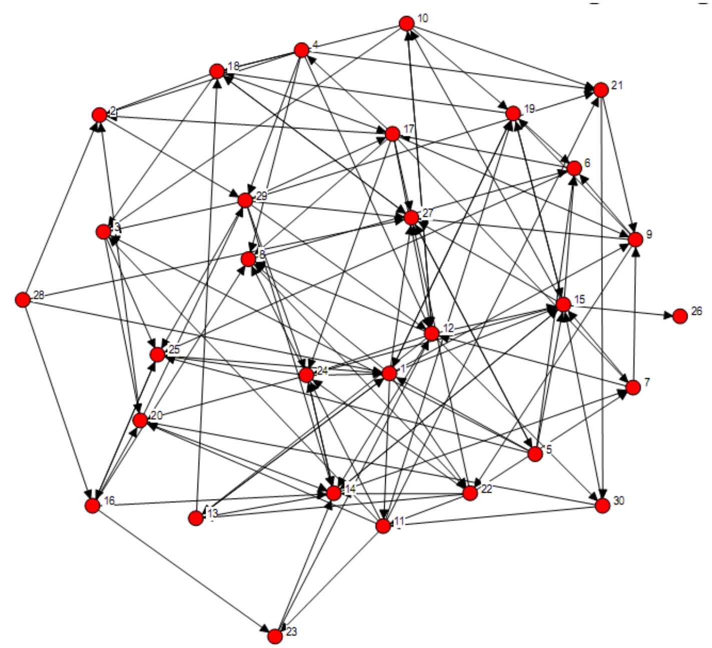
high density: 39%

Average Degree

- Average number of ties in a network



Density: 0,47
Average Degree: 4

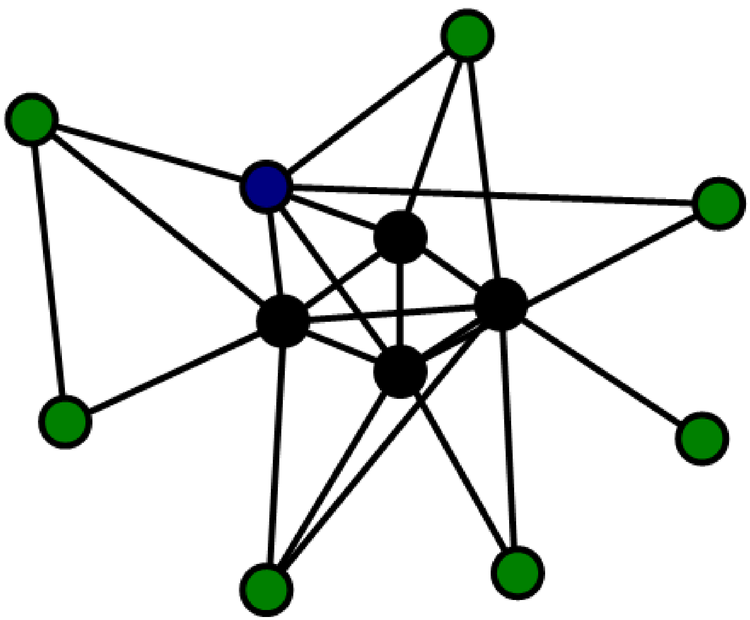


Density: 0,14
Average Degree: 4

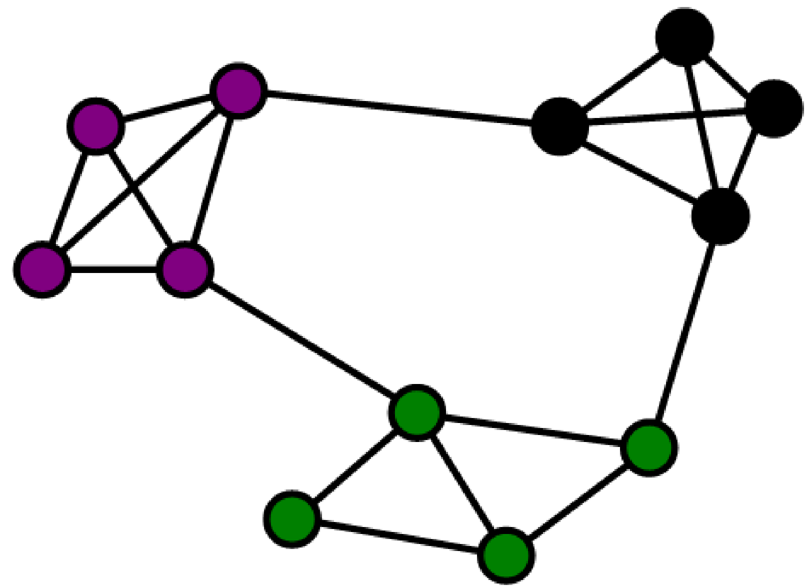
BI

Average Distance

- Average geodesic distance between all pairs of nodes



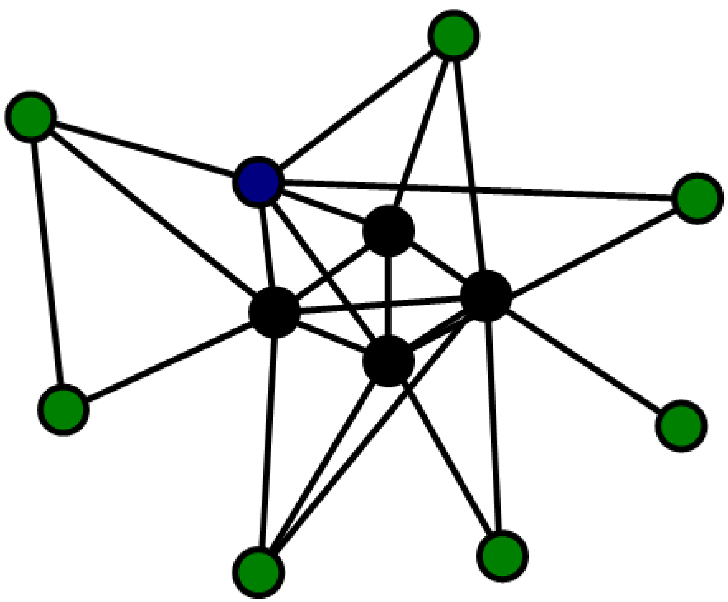
avg. distance 1.9



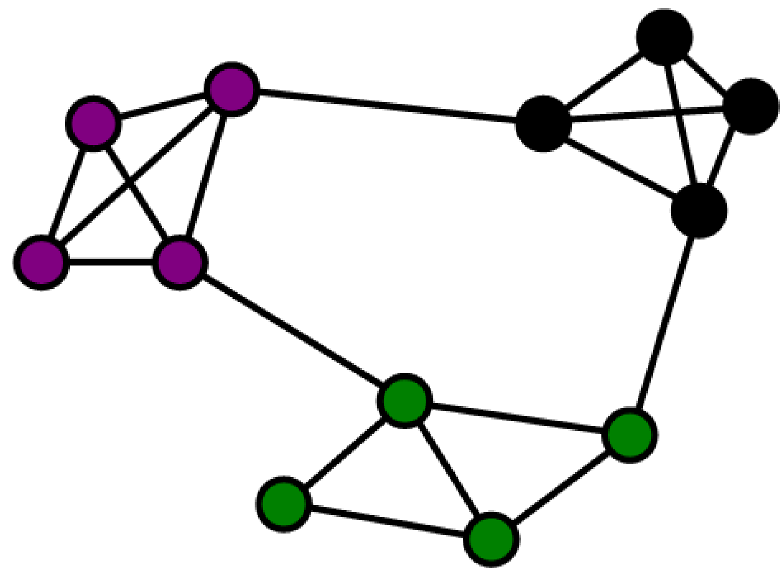
avg. distance 2.4

Diameter

- Maximum distance (length of the shortest path between the two most distant nodes)



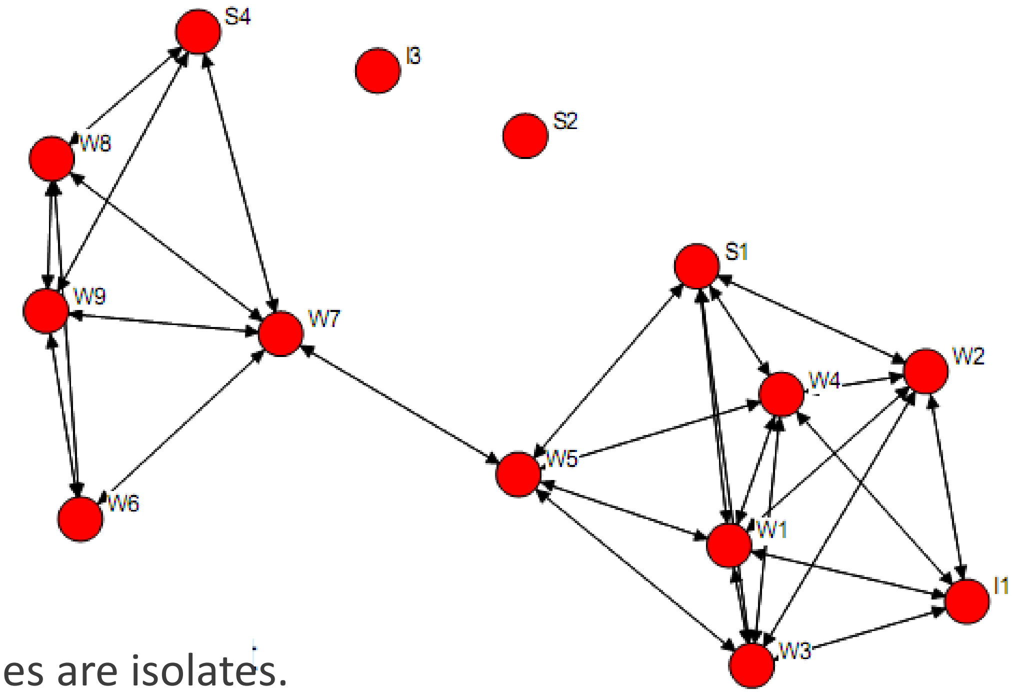
diameter 3



diameter 4

Number of Components

- Component ratio: number of components minus 1 divided by number of nodes minus 1



CR is 1 when all nodes are isolates.
CR is 0 when all nodes are in one component.

CR: (3-1)/(14-1) = 0.154

2

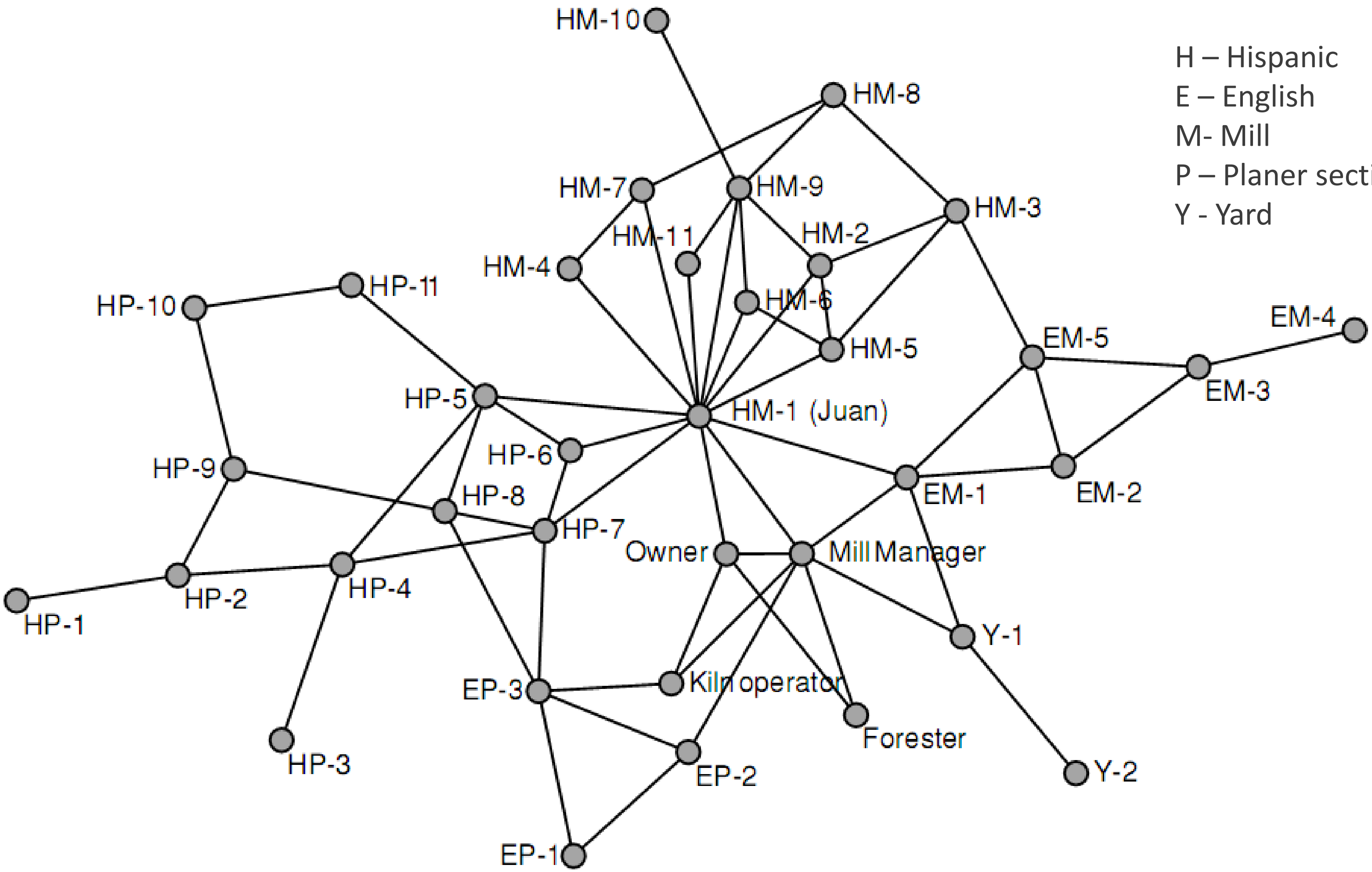
Metrics and Measures

Actors and Centrality

Centrality Measures

- Degree Centrality
- Degree Prestige
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Centrality & PageRank

Example – Communication ties within a sawmill



H – Hispanic
E – English
M- Mill
P – Planer section
Y - Yard

Distance

- The larger the number of sources accessible to a person, the easier it is to obtain information. Social ties constitute social capital that may be used to mobilize social resources.

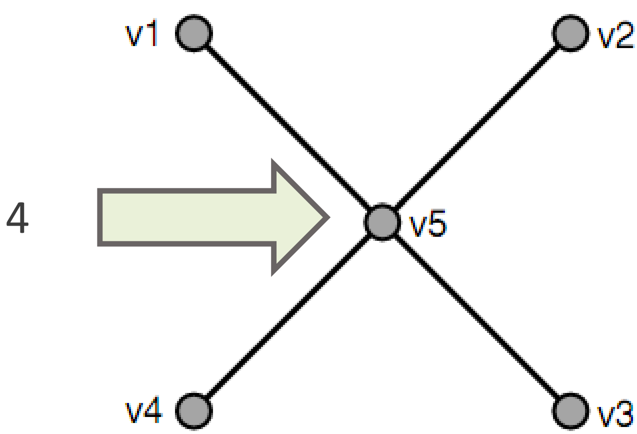
A **geodesic** is the shortest path between two vertices.

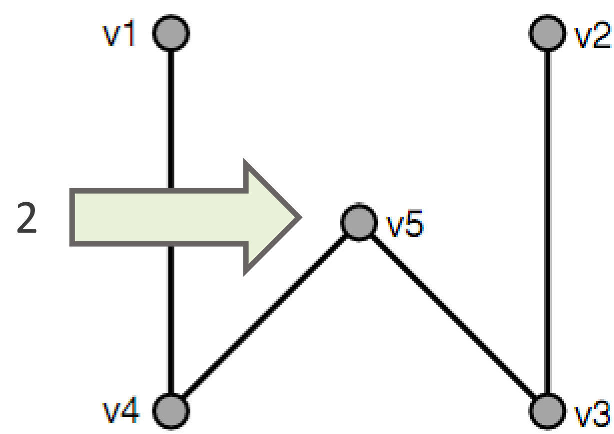
The **distance** from vertex u to vertex v is the length of the geodesic from u to v.

Degree Centrality

- The simplest indicator of centrality is the number of its ties (degree in a simple undirected network)

The **degree centrality** of a node is its degree.

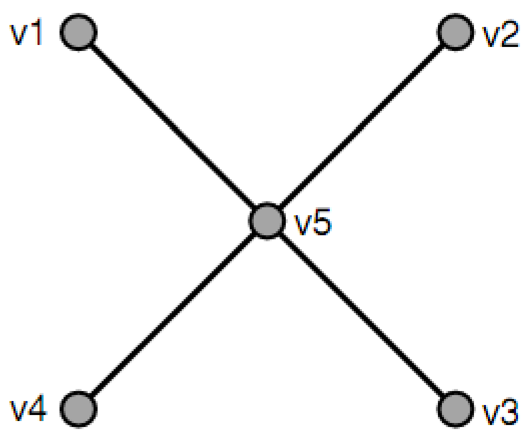




Degree Centrality for whole networks

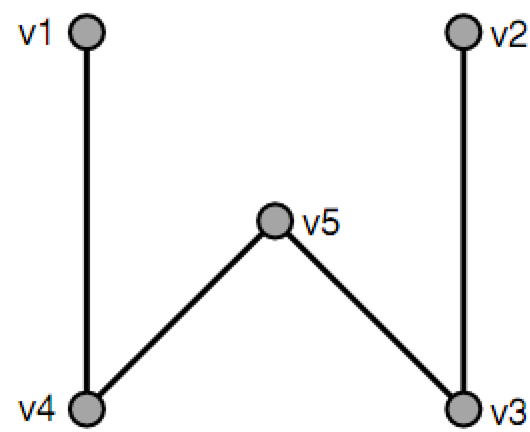
Degree centralization of a network is the variation in the degrees of vertices divided by the maximum degree variation which is possible in a network of the same size.

Network A



Degree Centralization = 1

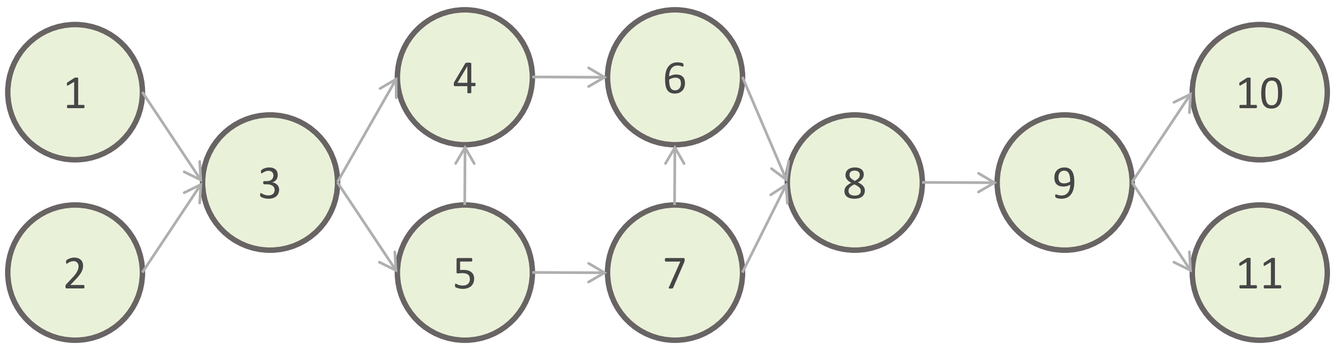
Network B



Degree Centralization = 0.17

Prestige Centrality = Indegree

- Prestige can be expressed as the relative indegree of an actor (***degree prestige***)



Prestige of node 3:

$$P_d(n_3) = 2_{+3} / (11 - 1) = 0,2$$

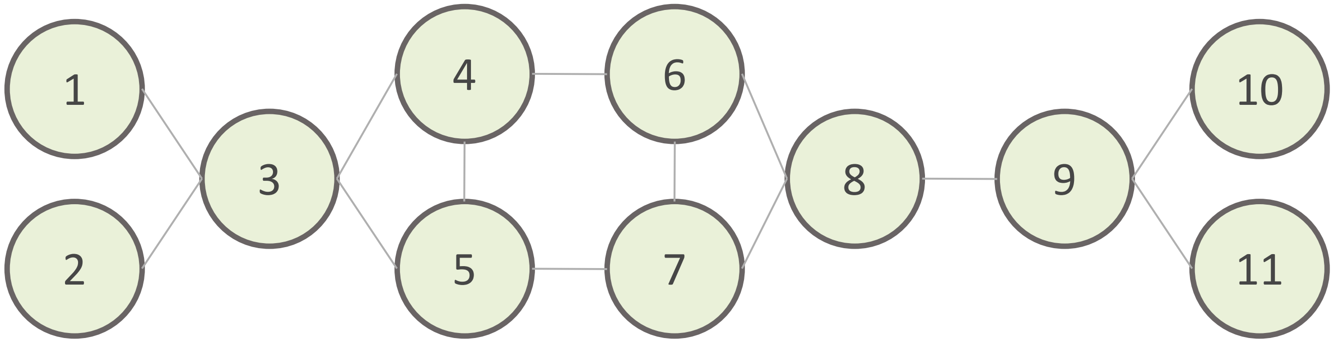
$$P_d(n_j) = x_{+j} / (g - 1)$$

Closeness Centrality

- **Closeness centrality** : A person is always then central, if that person regarding to the network relation is very close to all other persons. Such a central position allows to improve the efficiency of the communication of an actor. Such an actor is able to desseminate and receive information fast.

$$C_c(n_i) = \frac{g-1}{\sum_{j=1}^{g} d(n_i, n_j)}$$

Closeness Centrality



ni	nj	d
3	1	1
3	2	1
3	4	1
3	5	1
3	6	2
3	7	2
3	8	3
3	9	4
3	10	5
3	11	5
		25

$$C_c(n_3) = \frac{11-1}{25} = 0,4$$

Notice: We are only analyzing symmetrical relations and fully connected networks.

n	Cc
1	0.27
2	0.29
3	**0.4**
4	0.45
5	0.45
6	0.45
7	0.45
8	0.45
9	0.37
10	0.27

Betweenness Centrality

- Nodes that connect two in other respects unconnected subpopulations are actors with a high ***betweenness centrality*** score.
- Betweenness centrality measures bridging or being a bridge.
- Notice: We are assuming that information always travels on the shortest paths!

$$C_b(n_i) = \frac{\displaystyle\sum_{\substack{j \neq k \\ i \neq j,k}} \frac{g_{jk}(n_i)}{g_{jk}}}{(g-1)(g-2)}$$

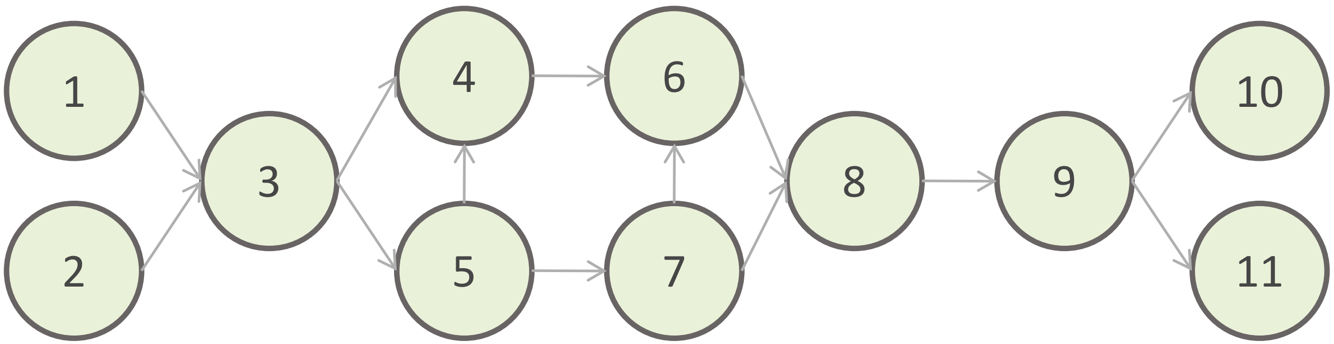
Normalization

BI

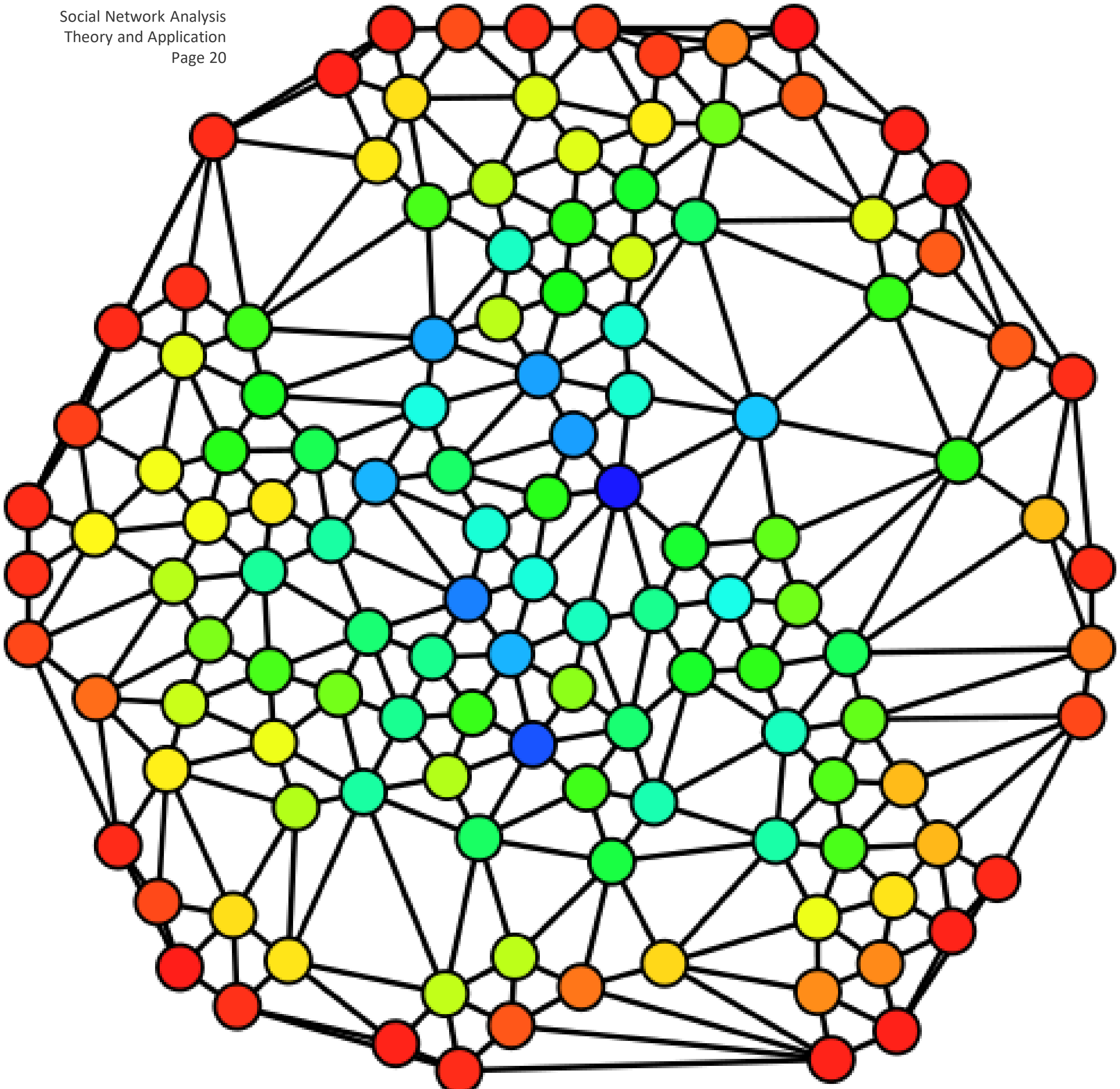
Where g_{jk} = the number of shortest paths connecting *jk*
$g_{jk}(i)$ = the number that actor *i* is on.

Betweenness centrality

- Notice: In directed networks it is possible that some actors are not reachable by others, but are themselves able to reach other nodes by themselves.



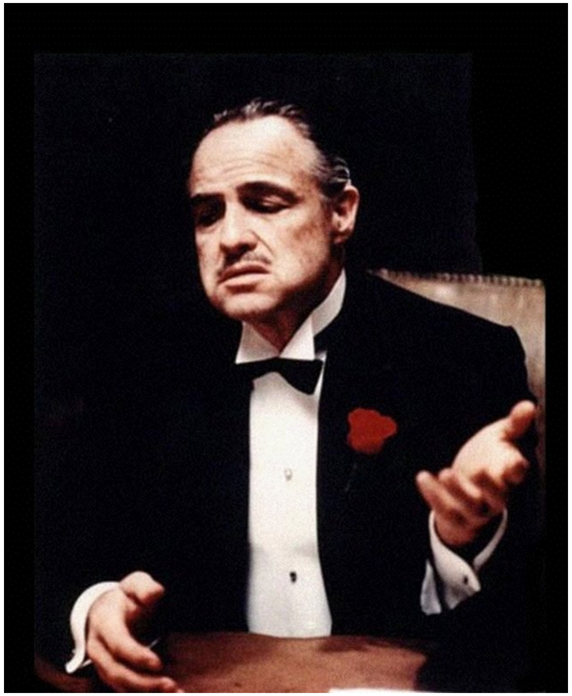
1	2	3	4	5	6	7	8	9	10	11
0	0	0.37	0.22	0.22	0.22	0.22	**0.48**	0.37	0	0



Hue (from red=0 to blue=max) shows the node betweenness.

Eigenvector Centrality

Don Corleone did not have many strong ties. He was a man of few words, yet he could make an offer you can't refuse. Don Corleone surrounded himself with his sons and his trusted *capos*, who in turn, handled the day to day management issues of the family.



Eigenvector-like centrality measures look at the **value** of connections, not only the number of connections.

Rationale: Being connected to well connected friends is better than being connected to friends without many connections.

Example: Google PageRank Algorithm

- Google's PageRank algorithm ranks websites based on an eigenvector-type centrality logic
- Sites that receive a lot of links from important sites are ranked highly

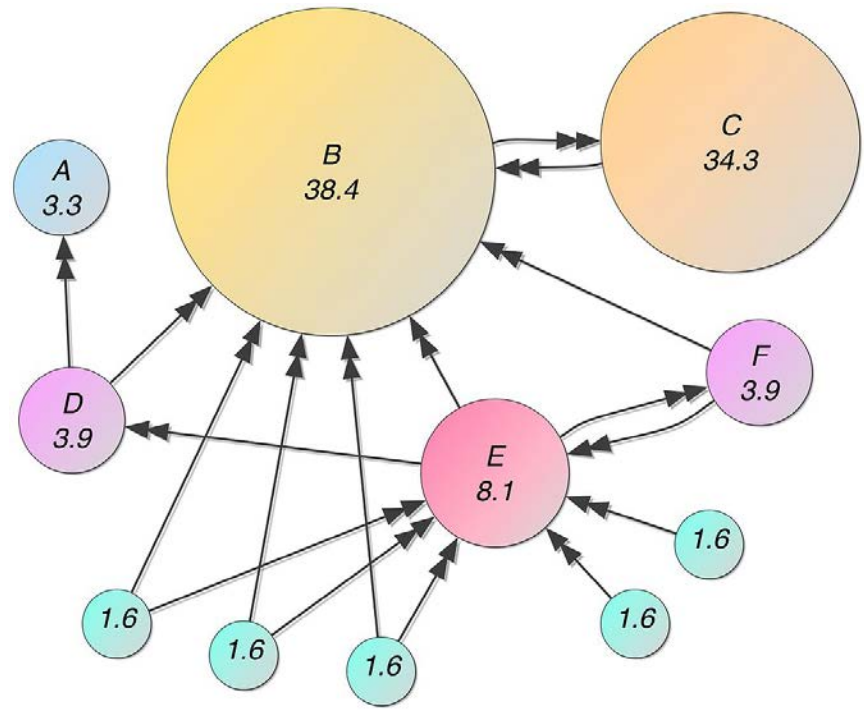
ELSEVIER

Computer Networks and ISDN Systems
Volume 30, Issues 1–7, April 1998, Pages 107-117

The anatomy of a large-scale hypertextual Web search engine ☆

Sergey Brin, Lawrence Page

Computer Science Department, Stanford University, Stanford, CA 94305, USA

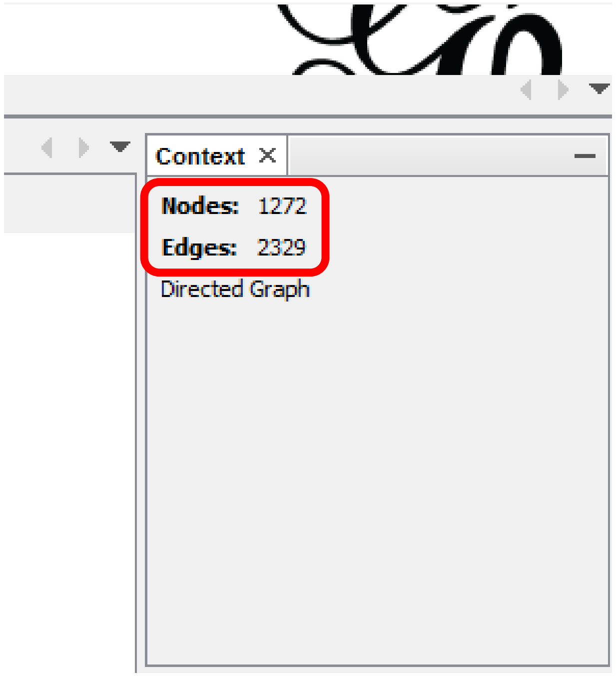


Centralities in Comparison

- **Degree**: How many people can this person reach directly?
- **Betweenness**: How likely is this person to be the most direct route between two people in the network?
- **Closeness**: How fast can this person reach everyone in the network?
- **Eigenvector:** How well is this person connected to other well-connected people?

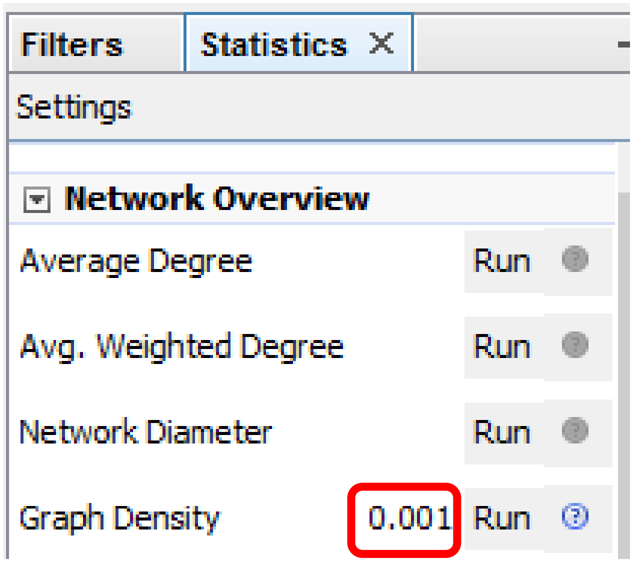
3 Network Measures and Where to Find Them

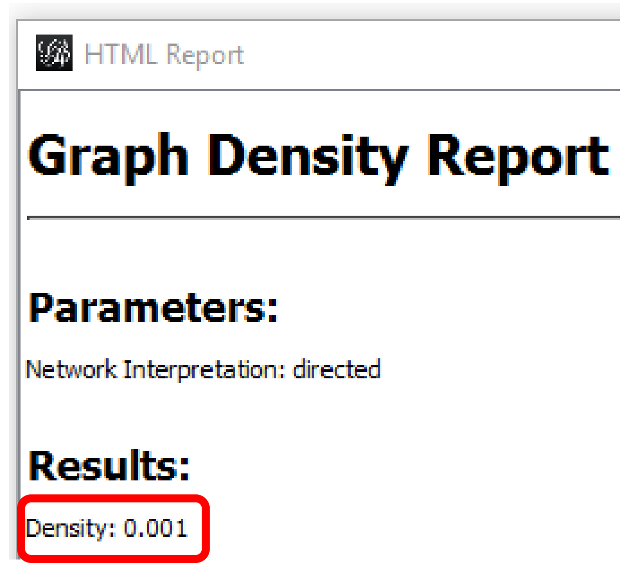
Number of Nodes and Ties



- Gives information about the **size** and **connectendness** of the network.
- The higher these numbers, the bigger and more complex the network.
- Depends a lot on the **phenomenon**: What type of nodes and ties do we have?

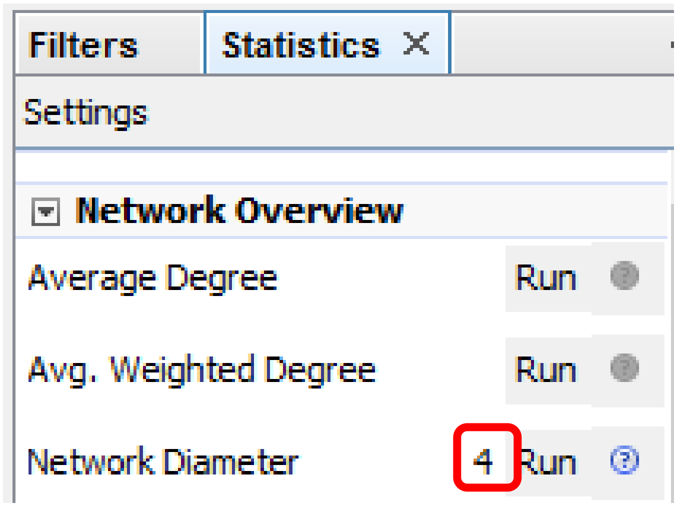
Density

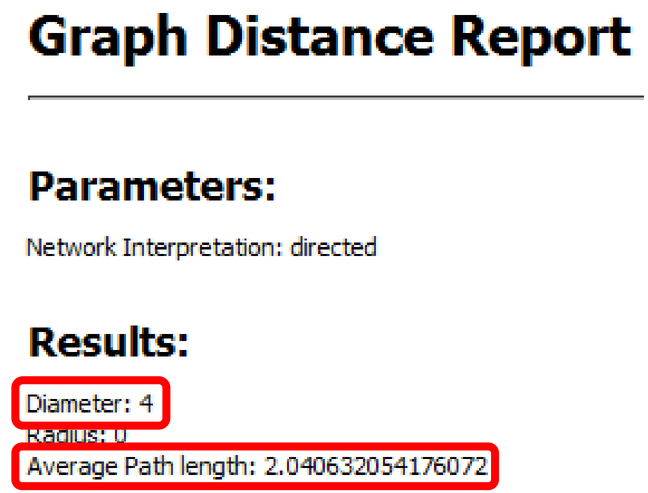




- **Density** = number of realized ties as fraction of number of possible ties [0,1].
- Density tells us **how connected** a graph is, how much interaction or connection there is on average.
- **Larger networks** are often less dense. Interpret density with size and type of your network in mind. Would it be realistic to have a dense network?
- With **social networks**, there is often a limit to the number of connections (resource constraints) => Social networks often have low density.

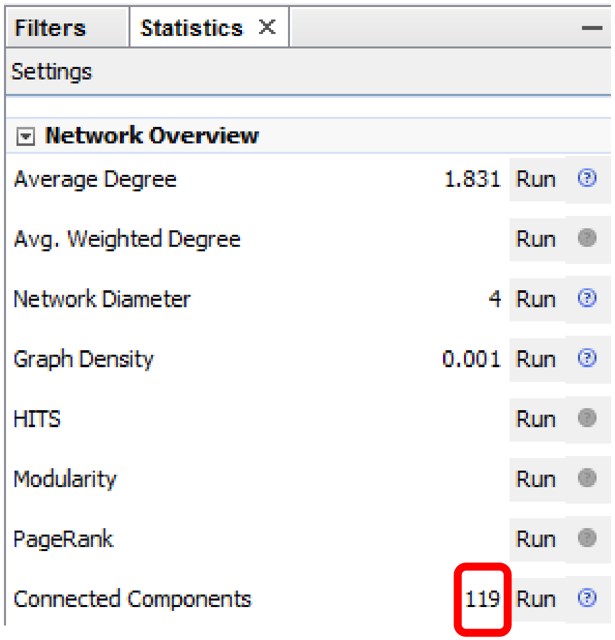
Diameter and Average Path Length

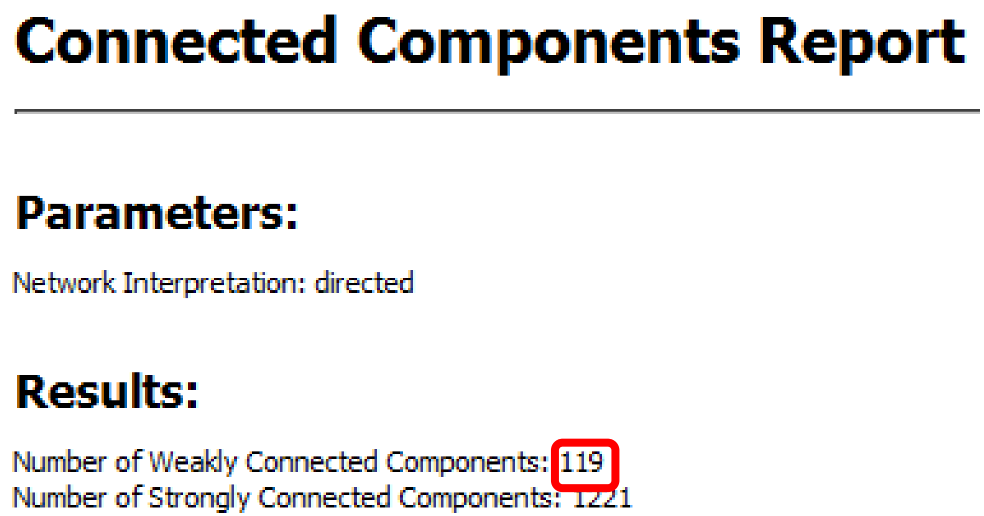




- **Diameter** = Shortest path between the two nodes that are furthest away
- **Average Path Length** = How many steps it takes for two nodes to reach each other
- The lower the diameter and average path length, the more **small world** and bridged the network
- High diameter and average path length indicate a dispersed, chain-like network
- Interpret average path length and diameter with size and type of your network in mind.

Number of Connected Components and Isolates





- **Components** are the connected parts of a network. Often, there is one giant connected component and few, if any, smaller components.
- This is quite visible in the network **visualization**.
- **Isolates** (nodes with degree 0) are counted as components.
- We are interested in the number of **weakly connected components**, again in relation to the size of the network.

thank you for your time

Norwegian Business School (BI)
Nordic Center for Internet & Society
Nydalsvn. 37 / N-0442 Oslo
bi.edu/ncis @BI_NCIS