CENTRALITY AND PRESTIGE

Measures of centrality and prestige

One of the most important uses of network analysis is identification of 'most central' units in a network.

Measures of centrality and prestige can be defined in two different ways:

- for each unit respectively – **unit centrality** (one number for each unit)

- for the whole network – **network centralisation** (only one number for the whole network).

Unit centrality and prestige

When we talk about centrality we must distinguish between undirected and directed networks:

- the term **centrality measures** is used for **undirected** networks;
  Example: A city is *central*, if a lot of roads are passing through it.

- the term **prestige** is used for **directed** networks. In this case we can define two different types of prestige:
  - one for outgoing arcs (**measures of influence**),
  - one for incoming arcs (**measures of support**).

  Examples:
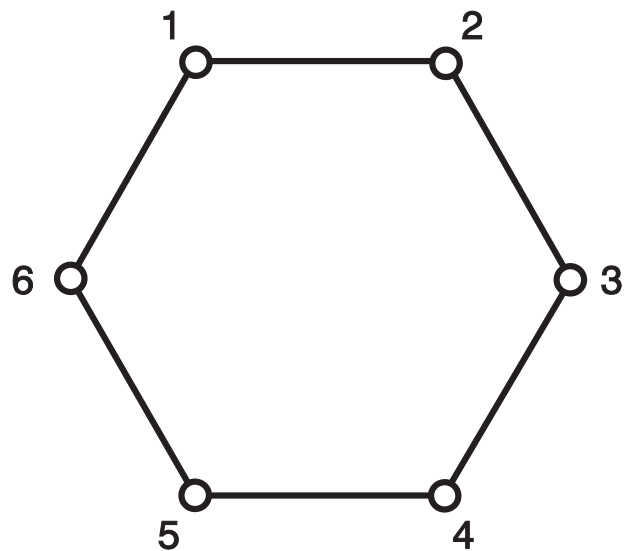 An actor has high *influence*, if he gives commands to several other actors.
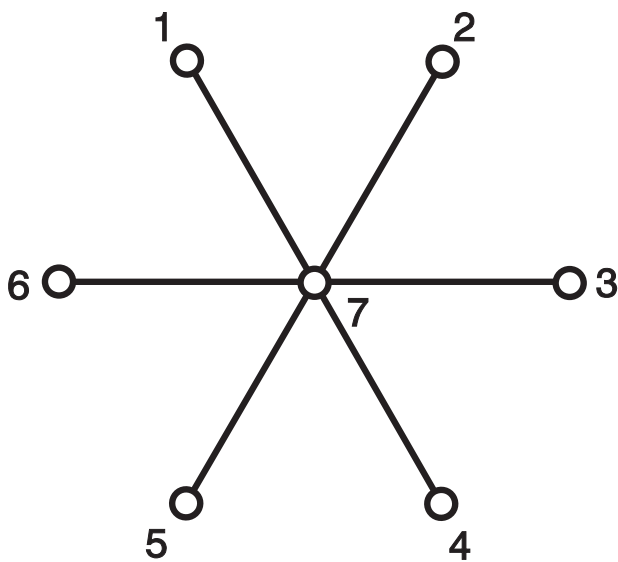 An actor has high *support*, if a lot of people vote for him.

Unit centrality measures

Selected unit is central

- if it has high **degree**,

- if it is easily accessible (**close to**) all other units,

- if it lies on several geodesics (shortest paths) **between** other units.

Star and cycle

According to all criteria mentioned, unit 7 in the star is the most central, while all units in the cycle are equally central.

Degree Centrality

The simplest measure – unit is central in a network, if it is active enough in the sense that is has a lot of links to other units (unit 7 in the star). In the case of cycle all units are equally central.

Degree centrality is defined by a degree of unit x

$$c_D(x) = \text{degree of unit } x$$

Such measures are called **absolute measures of centrality**. Absolute measures cannot be used to compare centralities of networks with different number of units. Therefore such measures are normalised to get values in interval from 0 to 1, where 0 means the smallest possible and value 1 the highest possible centrality. Measures normalised in this way are called **relative measures of centrality**.

Relative degree centrality is

$$C_D(x) = \frac{c_D(x)}{\text{highest degree}} = \frac{c_D(x)}{n-1}$$

if n is number of units in a network, the highest possible degree (in a network without loops) is $n-1$.

The same centrality measure can be used as a measure of *prestige for directed networks*. In this case there are two possibilities

- to choose (*influence* – out degree: number of arcs going out)

- to be chosen (*support* – in degree: number of arcs coming into).

Closeness Centrality

Sabidussi (1966) suggested the measure of centrality according to closeness of unit x:

$$c_C(x) = \frac{1}{\sum_{y \in U} d(x, y)}$$

where $d(x, y)$ is the graph theoretic distance (length of shortest path) between units x and y, U is set of all units.

If network is not strongly connected, we take only reachable vertices into account, but we weight the result with number of reachable vertices.

The most central units according to closeness centrality can quickly interact to all others because they are close to all others.

This measure is preferable to degree centrality, because it does not take into account only direct connections among units but also indirect connections.

Relative closeness centrality is

$$C_C(x) = (n-1) \cdot c_C(x)$$

Explanation: the smallest possible distance of selected unit from all other units is obtained, if the unit has all other units for neighbors. In this case the absolute closeness centrality is $\frac{1}{n-1}$.

Closeness centrality can be computed for undirected and directed networks.

There are two possibilities for directed networks: prestige can be computed according to

- outgoing arcs (how close are all other units from selected one: in how many steps we can reach all other units from the selected one)

- incoming arcs (how close is the selected unit to all others: in how many steps we can reach from all other units the selected one).

Betweenness Centrality

In the case of communication networks the distance from other units is not the only important property of a unit. More important is which units lie on the shortest paths among pairs of other units. Such units have control over the flow of information in the network.

Idea of betweenness centrality measures: unit is central, if it lies on several shortest paths among other pairs of units.

Freeman (1977) defined the centrality measure of unit x according to betweenness in the following way:

$$c_B(x) = \sum_{y < z} \frac{\text{\# of shortest paths between } y \text{ and } z \text{ through unit } x}{\text{\# of shortest paths between } y \text{ and } z}$$

Suppose that communication in a network always passes through shortest available paths:

Betweenness centrality of unit x is the sum of probabilities across all possible pairs of units, that the shortest path between y and z will pass through unit x.

Relative betweenness centrality must be defined for undirected and directed networks separately

- undirected networks:

$$C_B(x) = \frac{c_B(x)}{(n-1)(n-2)/2}$$

- directed networks:

$$C_B(x) = \frac{c_B(x)}{(n-1)(n-2)}$$

Explanation: Suppose that a network consists of n units. For selected unit x number of other, different pairs of units y and z is

- $C_{n-1}^2 = \binom{n-1}{2} = \frac{(n-1)(n-2)}{2}$
  in the case that order of choosing units in pair is not important (undirected network)

- $V_{n-1}^2 = (n-1)(n-2)$
 in the case, that the order of choosing units in pair is important (directed network)

Several other centrality measures exist.

Choosing suitable centrality measure

Different centrality measures can give quite different results. Therefore we must be very careful which centrality measure to choose for a given network:

It can happen, that some units have low degrees, but high betweenness centrality.

Example: units 33, 53, 32 in 48 are the most central according to betweenness centrality, although several other vertices with higher degree centrality exist (unit 61 has degree 6, some units have degree 5). The four units mentioned have degree only 3 or 4.
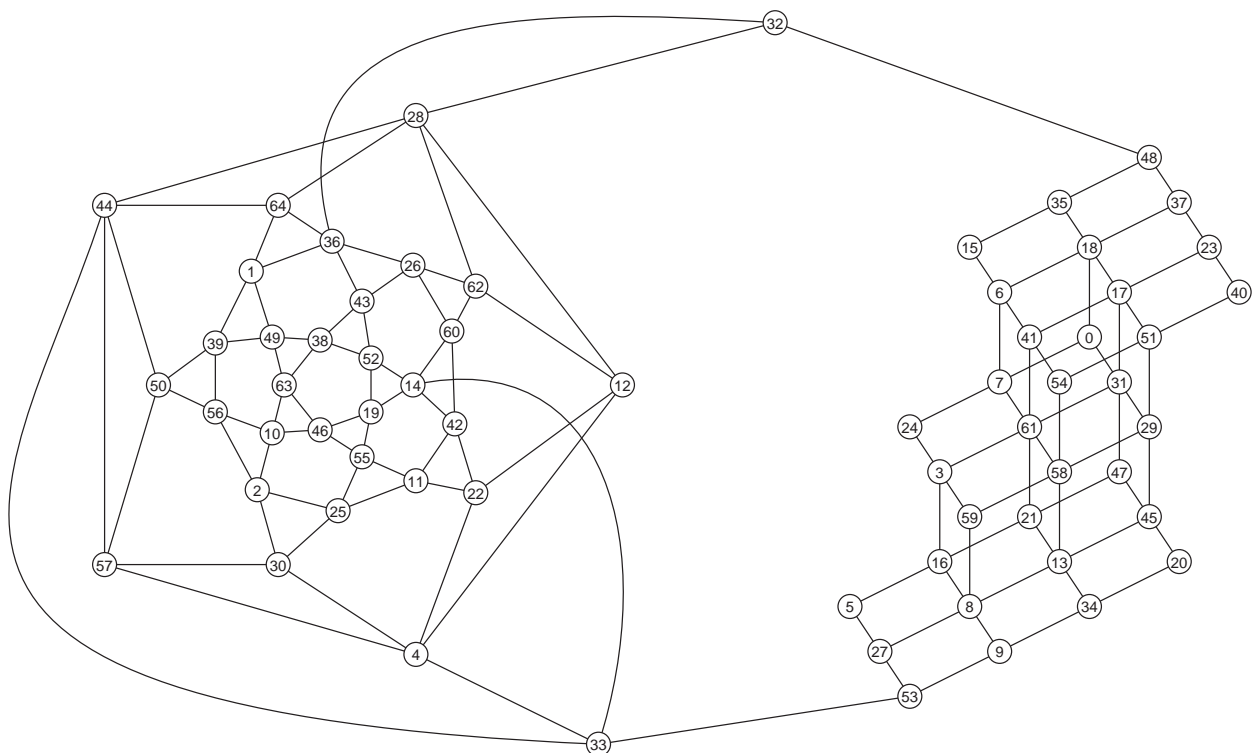
Example: Padgett's Florentine Families
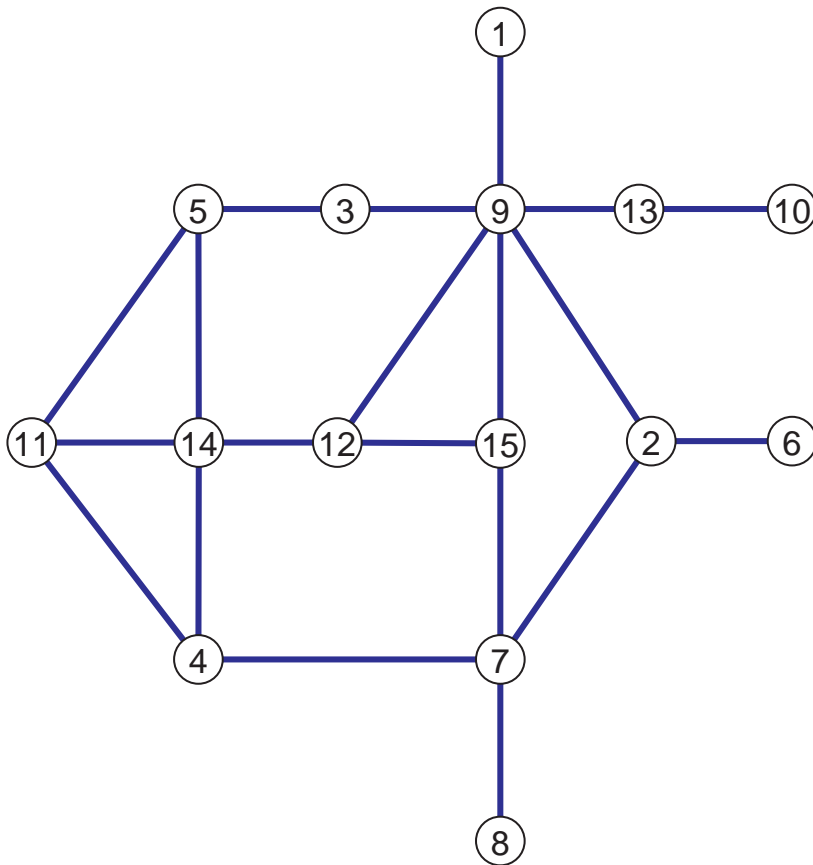

1.	Acciaiuoli	9.	Medici
2.	Albizzi	10.	Pazzi
3.	Barbadori	11.	Peruzzi
4.	Bischeri	12.	Ridolfi
5.	Castellani	13.	Salviati
6.	Ginori	14.	Strozzi
7.	Guadagni	15.	Tornabuoni
8.	Lamberteschi		

Relative centrality measures

No.	Family	C_D	C_C	C_B
1.	Acciaiuoli	0.071	0.368	0.000
2.	Albizzi	0.214	0.483	0.212
3.	Barbadori	0.143	0.438	0.093
4.	Bischeri	0.214	0.400	0.104
5.	Castellani	0.214	0.389	0.055
6.	Ginori	0.071	0.333	0.000
7.	Guadagni	0.286	0.467	0.255
8.	Lamberteschi	0.071	0.326	0.000
9.	Medici	0.429	0.560	0.522
10.	Pazzi	0.071	0.286	0.000
11.	Peruzzi	0.214	0.368	0.022
12.	Ridolfi	0.214	0.500	0.114
13.	Salviati	0.143	0.389	0.143
14.	Strozzi	0.286	0.438	0.103
15.	Tornabuoni	0.214	0.483	0.092

Example

Sampson monastery – liking
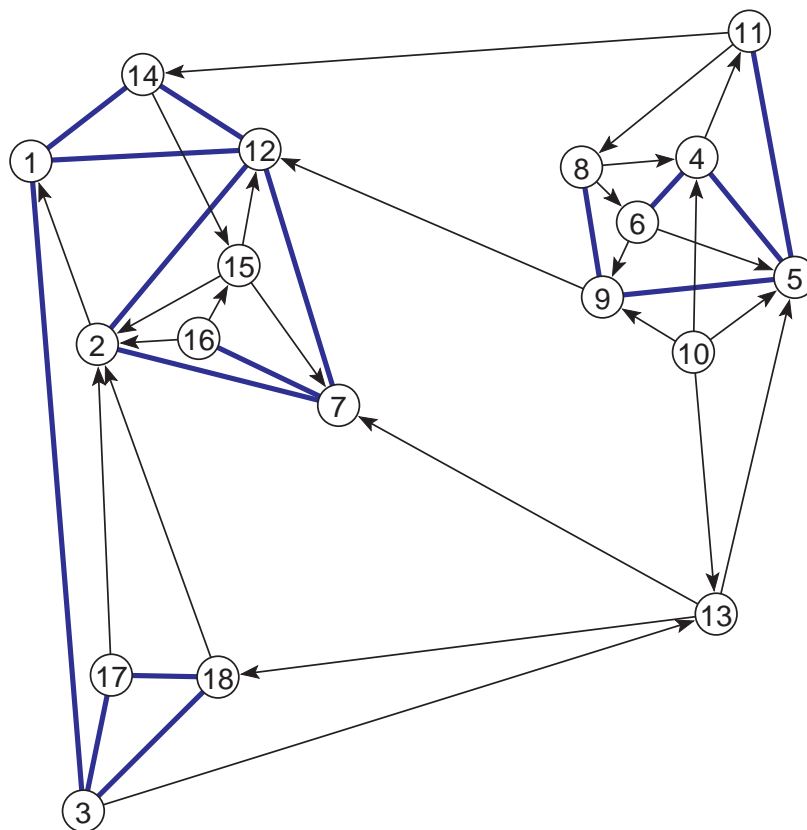
1.	John Bosco	10.	Romuald
2.	Gregory	11.	Louis
3.	Basil	12.	Winfrid
4.	Peter	13.	Amand
5.	Bonaventura	14.	Hugh
6.	Berthold	15.	Boniface
7.	Mark	16.	Albert
8.	Victor	17.	Elias
9.	Ambrose	18.	Simplicius

No.	Name	C_D	C_B	C_C
1.	John Bosco	0.235	0.332	0.472
2.	Gregory	0.353	0.122	0.486
3.	Basil	0.176	0.358	0.378
4.	Peter	0.235	0.065	0.293
5.	Bonaventura	0.353	0.227	0.370
6.	Berthold	0.118	0.007	0.239
7.	Mark	0.294	0.079	0.486
8.	Victor	0.118	0.013	0.239
9.	Ambrose	0.235	0.146	0.293
10.	Romuald	0.000	0.000	0.000
11.	Louis	0.118	0.089	0.283
12.	Winfrid	0.353	0.220	0.586
13.	Amand	0.118	0.250	0.298
14.	Hugh	0.177	0.111	0.447
15.	Boniface	0.118	0.016	0.354
16.	Albert	0.059	0.011	0.333
17.	Elias	0.118	0.002	0.293
18.	Simplicius	0.176	0.021	0.304

Network centralisation

Computed unit centralities in a network can have large or small variance. Network, where one unit (or low number of units) has (have) much higher centrality than other units is highly centralised (example is a star). On the other hand, if unit centrality measures do not differ significantly, the centralisation of a network is low (example is a cycle).

Freeman (1979) defined general network centralisation index

$$C_A = \frac{\sum_{x \in U}(C_A^\star - C_A(x))}{\max \sum_{x \in U}(C_A^\star - C_A(x))}$$

where $C_A^\star$ is the highest value of selected unit centrality measure $C_A(x)$ in the set of units of a network. Network centralisation index is a number between 0 and 1: the index is 0 if all units have equal centrality value (cycle), and 1, when one unit completely dominates all other units (star).

Network Centralisation Indices

Network Degree Centralization

$$C_D = \frac{\sum_{x \in E}(C_D^\star - C_D(x))}{n - 2}$$

Network Closeness Centralization

$$C_C = \frac{\sum_{x \in E}(C_C^\star - C_C(x))}{(n - 1)(n - 2)/(2n - 3)}$$

Network Betweenness Centralization

$$C_B = \frac{\sum_{x \in E}(C_B^\star - C_B(x))}{n - 1}$$

Network Closeness Centralization in Pajek is computed only if network is strongly connected. Otherwise it cannot be computed.

Example: Florentine Families

Network centralisation indices according to marriage relations are:

$$C_D = 0.275$$

$$C_C = 0.322$$

$$C_B = 0.437$$

Centrality measures in Pajek

Pajek computes relative centrality measures, only in the case of degree centrality the absolute measure is computed. Centralities can be found in Network/Create Vector/Centrality.

When computing centrality according to degree and closeness we must additionally select Input, Output or All. If network is undirected we can choose Input or Output (the result is the same).

In the Report window the corresponding network centralisation index is given as well.

List of the selected number of most central vertices can be obtained by Vector/Info.

Centrality measure of each vertex obtained (real number between 0 and 1) can be used to determine the size of a vertex in a layout (Draw/Network + First Vector).

Using Draw/Network + First Vector + Second Vector two vectors (e.g. input and output degree) can be drawn. In this case size in x direction is determined with the first vector (input degree) and size in y direction with the second (output degree).

Examples

1. Compute all centrality measures in undirected network flor.net (Florentine families). Draw picture of network, where sizes of vertices will correspond to selected centrality measure.

2. Compute all centrality measures in directed network sampson.net (Sampson monastery – liking). Draw picture of network, where sizes of vertices will correspond to selected centrality measure.

3. Compute all centrality measures in undirected network write.net. Draw picture of network, where sizes of vertices will correspond to selected centrality measure. Which centrality measure differentiates the vertices the most (standard deviation)?

4. Which are the most important airports in the airlines connections network (usair97.net) according to degree, closeness and betweenness?

Hubs and authorities

Look at another two measures of prestige, which are especially useful in the case of WWW (directed network of home pages).

In directed networks we can usually identify two types of important vertices: hubs and authorities. Each home page describes something (is an *authority*) and because of that other pages point to it. But on the other hand each page points to some other pages (is a *hub*).

Vertex is a *good hub*, if it points to many *good authorities*, and is a *good authority*, if it is pointed to by many *good hubs*.

See: Kleinberg, Jon M.: Authoritative Sources in a Hyperlinked Environment. *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms.* Edited by Howard Karloff (SIAM/ACM-SIGACT, 1998).

For each vertex v we computed two weights $x_v, y_v \in [0, 1]$, which tell how good authority and hub us a given vertex. We say that vertex v is better authority than vertex u, if $x_v > x_u$.

Weights are computed according to network by solving the eigenvector problem of matrices AA^T (hubs) and $A^T A$ (authorities).

Measures of prestige in Pajek

Hubs and authorities are computed in

Network/Create Vector/Centrality/Hubs-Authorities

We get additional question: How many hubs and how many authorities are marked in the partition. Results are:

- Partition, where value 1 (yellow) means, that the vertex is a good authority, value 2 (green) means, that the vertex is a good authority and a good hub, and value 3 (red) means, that the vertex is a good hub.

- Vector with *Hub Weights*, larger value means better hub.

- Vector with *Authority Weights*, larger value means better authority.

Be careful: Algorithm supposes that values on lines represent similarities (larger value means more important choice). In the case that it is the opposite, we must set all values to 1, otherwise results will be wrong. This can be done using Network/Create New Network/Transform/.. ../Line Values/Set All Line Values to 1.

Comparison of prestige measures

Four prestige measures were mentioned:

* input degree;

* input closeness;

* betweenness;

* authorities.

Some of these measures consider prestige in a similar way, some are quite different. For comparing the results of different measures we can use *Pearson correlation coefficient*, which is available in Vectors/Info, but we must first select both vectors. It is easier to export vectors to program package R or SPSS and compute correlations there.

In general we find out that most measures are very similar except betweenness, which measures prestige in a quite different way.

Examples

1. Compute input and output degrees, valued input and output degrees and hubs and authorities in networks representing export of soccer players football.net (Paris 1998) and/or football2002.net (Korea/Japan 2002).