Leaflet on Social Network Analysis – The most important measures, their specification, calculation, and interpretation.

Degree Centrality

The degree-centrality of an *undirected network* is calculated by the sum of all ties of actor n_i to the other actors of the network.

$$C_D(n_i) = d_i = \sum_j x_{ij} = \sum_j x_{ji} \qquad i \neq j$$

To allow for the comparison between networks of different size this indicator is normalized by dividing the C_D by the maximal possible centrality value (n-1).

$$C_D(n_i) = \frac{\sum_j x_{ij}}{(n-1)} = \frac{\sum_j x_{ji}}{(n-1)} = \frac{C_D}{(n-1)} \qquad i \neq j$$

The calculation of degree centrality for *directed graphs* is basically the same except the distinction between incoming and outgoing relationships.

The centrality-measures are used to identify the most central and supposingly 'most important' actors of a network. The concept of centrality is closely related to that of *power* (outgoing ties) and *influence* (incoming ties). Actors in a central position face fewer constraints and more opportunities (e.g. in receiving a certain resource).

The advantage of the degree centrality measure is its applicability even to large networks. However, its shortcoming is that it only accounts for the direct ties and therefore neglects the importance of the position within the network (e.g. critical point of the information flow).

Closeness Centrality

The closeness centrality not only includes the direct ties of an actor but all indirect ties to all other actors in the network. Accordingly, the closeness centrality does not just measure the proximity of an actor to its direct neighbours, but his proximity to all other actors of the network. Closeness centrality therefore is the extent to which an actor lies at short distance to many other actors of the network or.

Analytically closeness centrality of a *connected* network is calculated as the reciprocal value of the *geodesic* distances (shortest paths measured as the number of steps/ties between two actors):

$$C_C(n_i) = \frac{1}{\sum_{j=1}^{n} d(n_i, n_j)} = \left(\sum_{j=1}^{n} d(n_i, n_j)\right)^{-1}$$

The normalised closeness centrality of node n_i is:

$$C'_c(n_i) = \frac{(n-1)}{\sum_{j=1}^{n} d(n_i, n_j)}$$

Closeness centrality is a measure of the *autonomy* and the *speed* of network interactions of an actor. The closer an actor is to the other actors in the network, the quicker and the more independent he is in reaching others. If an actor with a high closeness centrality would leave a network, this would have enormous consequences for the functioning of the overall network structure.

Betweenness Centrality

Betweenness is the percentage of times an actor lies on the shortest path 'between' two other actors.

Analytically the betweenness centrality (C_B) is the probability that the communication between the actors k and j goes via actor i. Therefore the probability b_{jk} for every pair j and k is calculated by dividing the amount of geodesics (i.e. shortest paths) $g_{jk}(n_i)$ between j and k that go via i by the total number of shortest paths g_{jk} between j and k. These probabilities will then be calculated and summed up for every pair of actors in the network.

$$b_{jk}(n_i) = \frac{g_{jk}(n_i)}{g_{jk}}$$

$$C_B(n_i) = \sum_{j<}^{n} \sum_{k}^{n} b_{jk}(n_i) \qquad i \neq j \neq k$$

The normalised betweenness centrality therefore is:

$$C'_B(n_i) = \frac{2C_B(n_i)}{n^2 - 3n + 2}$$

The betweenness centrality is a measure for the control of information flows within the network and the function of single actors as intermediaries. An actor is the more powerful and influential the more indirect ties of other actors are mediated and controlled by him. To have a high betweenness centrality it is not necessary to maintain many direct ties, it is enough to mediate the important ones.

Network Density

The Network Density describes the overall level of linkage among the actors. It is the number of actors who are connected to each other, expressed as a percentage of the maximum possible number of connected actors.

$$Density\Delta_k = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijk}}{n \cdot (n-1)} \qquad i \neq j \neq k$$

With n·(n-1) as the total number of ties possible and k stands as the relation that is under investigation.

Network Density describes the overall *coherence* of a network and therefore allows for implications on the *speed* of diffusion of information and knowledge within the network and the levels of *social capital* and/or *social constraints* that actors have.

Network Cohesion

The Network Cohesion of *directed* graphs is a measure that describes the network as a whole. It is based on the relations between the actors. Formally it is calculated as the number of *mutual* selection of two actors compared to all possible dyads in the network. Groups in which every actor is directly connected to every other actor are called a *'clique'*. Thus, high network cohesion reflects a high degree of homogeneity within the network, since a high percentage of relations between two different actors are chosen mutually. Analytically the network cohesion resembles the network density, only that now it is not the unilateral relations, but the reciprocal ones. This requires the division of the maximal possible connections by the factor of 2:

$$Cohesion\Delta_k = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} + x_{ji})}{\frac{n \cdot (n-1)}{2}} \qquad i \neq j$$

High cohesion is expected to foster the development of mutual *trust* and *understanding* as well as common *norms* and *standards*. Furthermore it enhances the possibility of *monitoring* and *sanctioning* uncooperative behaviour.

Network Clustering (Cliquishness)

Networks could be described as being clustered into cohesive subgroups (neighbourhoods, cliques) that can be defined institutionally, geographically or with respect to the underlying knowledge. The actors of a clique are more closely tied to each other than they are to actors outside the clique. One actor can be associated to different cliques. In graph-theory a clique is a part of the network which exhibits the maximal possible ties between as much actors as possible. The clustering coefficient of any actor is calculated by the proportion of links that exist between the vertices and its neighbourhood divided by the number of links that could possibly exist between them. For an *undirected* we multiply by 2:

$$C_i = \frac{2\sum x_{jk}}{N_i(N_i - 1)}$$

The clustering coefficient for the whole network is then the average of the clustering coefficients of all vertices.

$$C_{Network} = \frac{1}{n}\sum_i Ci$$

Data Collection and Preparation

In general the data collection can either be focused on primary (direct) or secondary (indirect) data.

Three methods of primary data collection can be distinguished:

1) Collection of full network data:
   – Find a list of actors and ask each actor to indicate whether there is a relation and how strong this relation is (e.g. importance, frequency)
   – NOTE: You can repeat it for different types of network relations and resources exchanged.
2) Collection of ego-network data:
   Ask each actor (ego) to mention actors with whom they interact (alters) and how these alter are connected.
   – NOTE: This way you discover only part of the network i.e. only the direct alters of the nodes in your sample and for these alters only the links to the sample actors.
3) The snowball method:
   – Start asking for the links of one or more focal actors; continue asking for the links of mentioned actors; go on until no new actors are added to your list.

Pros: different kinds of links among the same set of actors and data on characteristics of the links.
Cons: high response rate required time- and labour intensive, only static network data.

– Secondary data collection refers to the use of statistical data e.g. patent data, citation data...

Pros: Longitudinal data, less time consuming.
Cons: Only relations that let to outcome; these outcomes and the actor’s behaviour varies between industries and organisations.