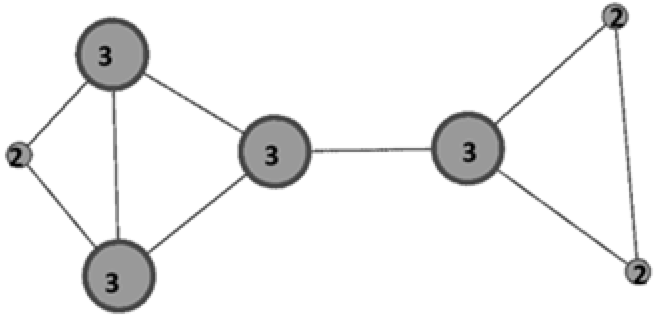
Análise de Redes Avançada

Aula 4

Medidas (continuação)

Centralidade de Grau

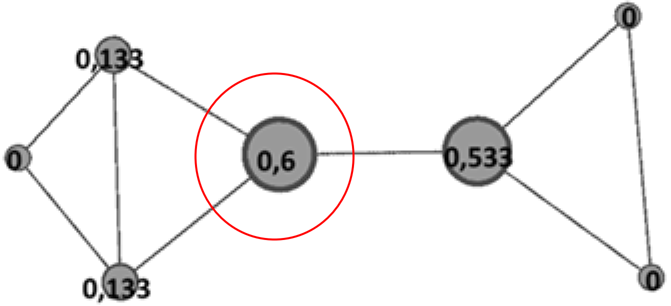


$$c_D(i) = \frac{\deg(i)}{n-1}$$

- How "connected" is a node?
- Degree captures connectedness
- Normalize by n-1
- Which are the most active entities?
- If it is receiving denotes prestige, if it connects it denotes influence. In any way it denotes the magnitude of available resources.

Degree centrality is simply a normalized node degree, i.e., the actual degree divided by the maximal degree possible (n-1). For directed networks, you can define ***in-degree centrality*** and ***out-degree centrality*** separately.

Centralidade Intermediação (*Betweeness*)

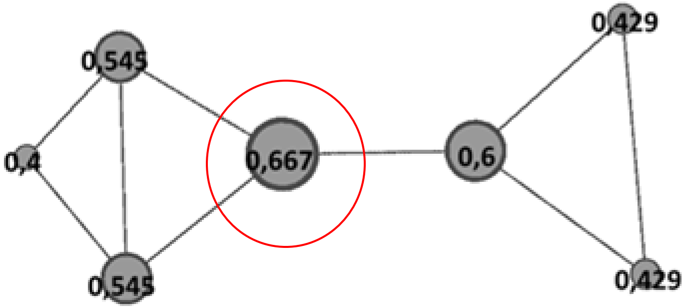


$$c_B(i) = \frac{1}{(n-1)(n-2)} \sum_{j \neq i, k \neq i, j \neq k} \frac{N_{\mathrm{sp}}(j \xrightarrow{i} k)}{N_{\mathrm{sp}}(j \rightarrow k)}$$

- How important is ***i*** as an intermediary?
- Number of shortest paths that crosses each vertice.
- Entities that connect several communities.
- Actors that have mediation power. They represent however weak breakpoints for the all community.

Betweenness centrality of a node is the probability for the shortest path between two randomly chosen nodes to go through that node. This metric can also be defined for edges in a similar way, which is called edge betweenness.

Centralidade de Proximidade (*Closeness*)



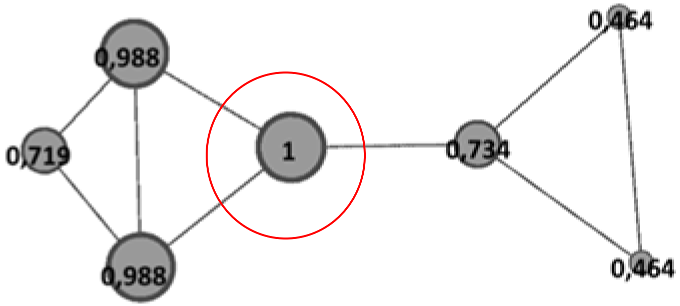
$$c_C(i) = \left(\frac{\sum_j d(i \to j)}{n-1} \right)^{-1}$$

- How close to others a node is?
- Sum of the reciprocal of the distance to all the other vertices.
- Entities with <u>better autonomy</u> and <u>better visibility</u>.
- Quick access to other actors and greater knowledge of the network.

This is an inverse of the average distance from node i to all other nodes. If $cC(i) = 1$, that means you can reach any other node from node i in just one step. For directed networks, you can also define another closeness centrality by swapping i and j in the formula above to measure how accessible node i is from other nodes.

Centralidade de Vector Próprio (*Eigenvector*)

$$c_E(i) = v_i$$



v_i is the i-th element of the dominant eigenvector v of the network's adjacency matrix. Eigenvector v is usually chosen to be a non-negative unit vector ($v_i >= 0$, $|v| = 1$).

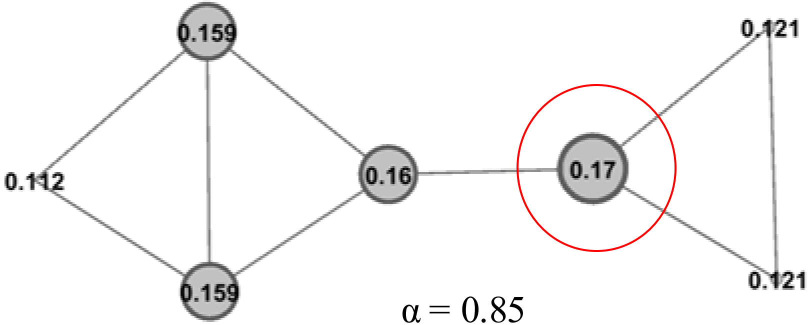
Eigenvector centrality measures the "importance" of each node by considering each incoming edge to the node an "endorsement" from its neighbor. This differs from degree centrality because, in the calculation of eigenvector centrality, endorsements coming from more important nodes count as more.

Another completely different, but mathematically equivalent, interpretation of eigenvector centrality is that it counts the number of walks from any node in the network that reach node i in t steps, with t taken to infinity.

Centralidade *Page Rank*

Like Eigenvector centrality but instead of A the following matrix is used:

$$T = \alpha A D^{-1} + (1-\alpha)\frac{J}{n}$$



D is a diagonal matrix whose i-th diagonal component is $1/deg(i)$,

J is an n x n all-one matrix, and α is the damping parameter (= 0.85 is commonly used by default).

PageRank measures the asymptotic probability for a random walker on the network to be standing on node i, assuming that the walker moves to a randomly chosen neighbor with probability α or jumps to any node in the network with probability $1-\alpha$, in each time step. Eigenvector v is usually chosen to be a non-negative unit vector ($v_i \geq 0$, $|v| = 1$) which gives a probability distribution.

Measure - Clustering Coefficient

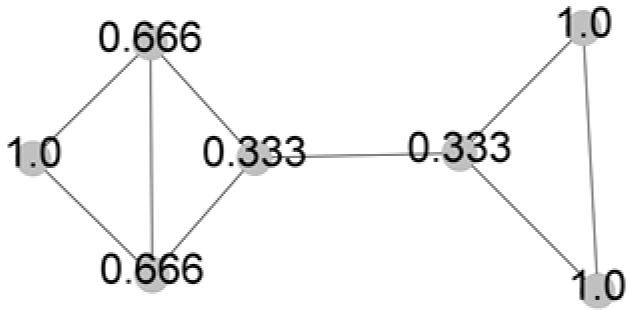
$$C(i) = \frac{|\{\ \{j,k\} \mid d(i,j) = d(i,k) = d(j,k) = 1\ \}|}{\deg(i)(\deg(i)-1)/2}$$

The denominator is the total number of possible node pairs within node i's neighborhood, while the numerator is the number of actually connected node pairs among them. Therefore, the clustering coefficient of node i calculates the probability for its neighbors to be each other's neighbors as well. Note that this metric assumes that the network is undirected.

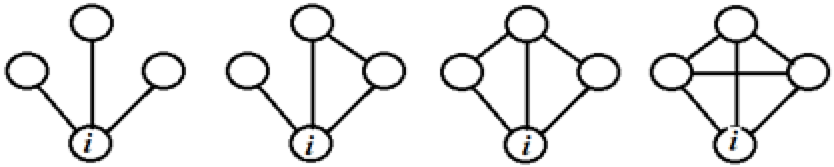
The following average clustering coefficient is often used to measure the level of clustering in the entire network:

$$C = \frac{\sum_i C(i)}{n}$$

Coeficiente de *Clustering*



C = 5/7 = 0.714



$Ci = 0$ $Ci = 1/3$ $Ci = 2/3$ $Ci = 1$

- Quantifies how close a vertex is to belonging to a clique

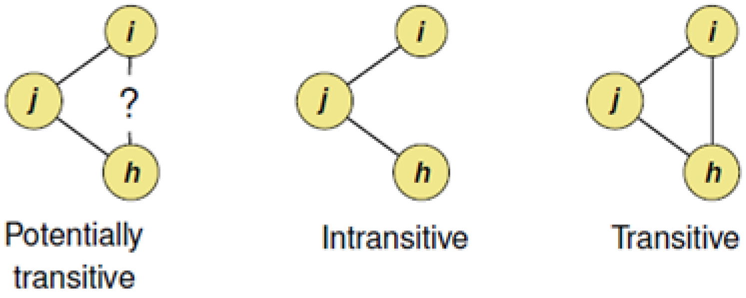
- A graph is considered a small-world if its average clustering coefficient is higher than a random graph with the same number of vertices.

Medida de Transitividade

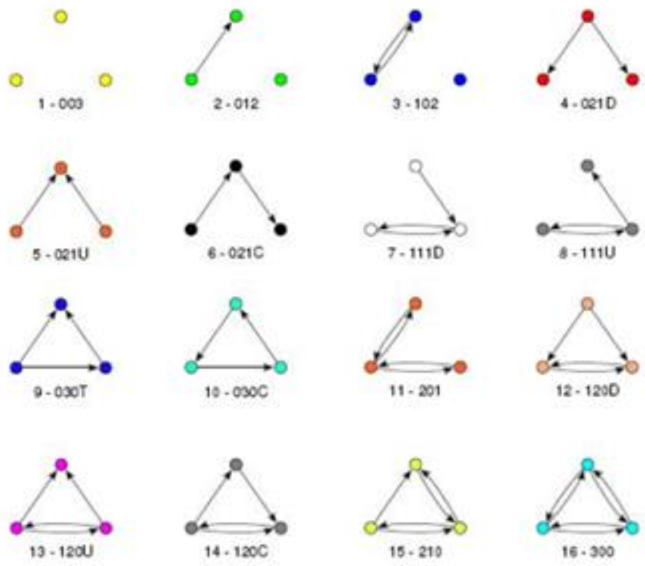
$$C_T = \frac{\left|\{\ (i,j,k) \mid d(i,j) = d(i,k) = d(j,k) = 1\ \}\right|}{\left|\{\ (i,j,k) \mid d(i,j) = d(i,k) = 1\ \}\right|}$$

This is very similar to clustering coefficients, but it is defined by counting connected node triplets over the entire network. The denominator is the number of connected node triplets (i.e., a node, i, and two of its neighbors, j and k), while the numerator is the number of such triplets where j is also connected to k. This essentially captures the same aspect of the network as the average clustering coefficient, i.e., how locally clustered the network is, but the transitivity can be calculated on directed networks too. It also treats each triangle more evenly, unlike the average clustering coefficient that tends to underestimate the contribution of triplets that involve highly connected nodes.

Medida de Transitividade



<u>"Friends of my friends are also my friends"</u>

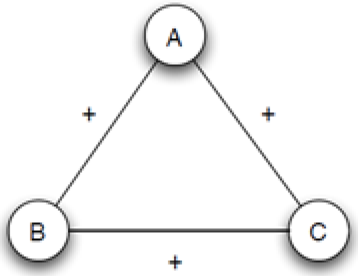


$$\text{Transitive Index} = \frac{\#\text{Transitive triads}}{\#\text{ Potentially transitive triads}}$$

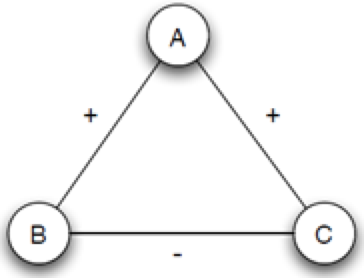
Permit statistics over all the network

Balanço Estrutural

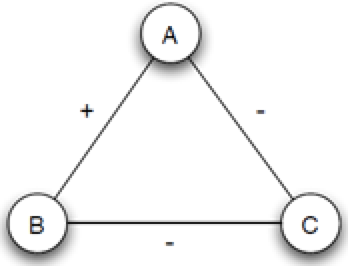
- Harary’s Theorem - a complete signed graph is balanced if and only if the nodes can be partitioned into two sets so that all ties within sets are positive, and all ties between sets are negative.



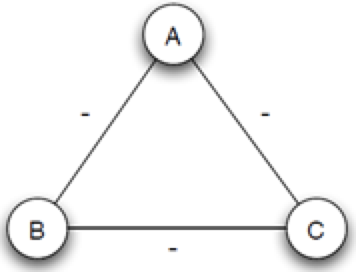
Balanced



Not Balanced



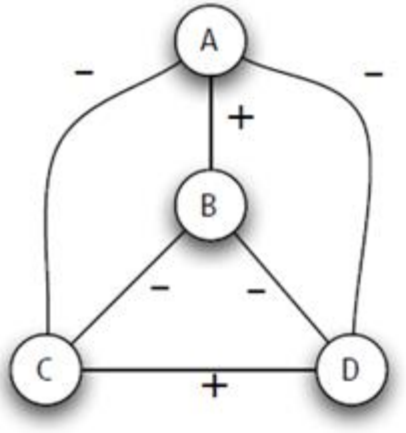
Balanced



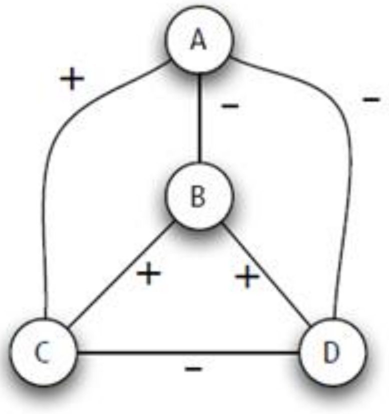
Not Balanced

Balanço Estrutural

- Harary’s Theorem - a complete signed graph is balanced if and only if the nodes can be partitioned into two sets so that all ties within sets are positive, and all ties between sets are negative.



Balanced



Not Balanced

Assortatividade

Assortativity coefficient

$$r = \frac{\sum_{(i,j)\in E}(f(i)-\bar{f}_1)(f(j)-\bar{f}_2)}{\sqrt{\sum_{(i,j)\in E}(f(i)-\bar{f}_1)^2}\sqrt{\sum_{(i,j)\in E}(f(j)-\bar{f}_2)^2}} \qquad \bar{f}_1 = \frac{\sum_{(i,j)\in E} f(i)}{|E|}, \qquad \bar{f}_2 = \frac{\sum_{(i,j)\in E} f(j)}{|E|}.$$

where E is the set of directed edges (undirected edges should appear twice in E in two directions)

The assortativity coefficient is a Pearson correlation coefficient of some node property f between pairs of connected nodes. Positive coefficients imply assortativity, while negative ones imply disassortativity.

If the measured property is a node degree (i.e., $f = deg$), this is called the degree assortativity coefficient. For directed networks, each of $f1$ and $f2$ can be either in-degree or out-degree, so there are four different degree assortativity you can measure: in-in, in-out, out-in, and out-out.