Network Metrics

This overview is from Sosa, Sueur, and Puga-Gonzalez (2021).

General Principles

Network metrics are mathematical calculations to quantify specific features of a network, including global, nodal, and polyadic measures. Unlike other chapters, here we will present a suite of the most commonly used network metrics and the corresponding BI functions under the class `m.net.`, implemented with JAX and usable within any bi model. This section is inspired by XXX, and users willing to dig further are invited to read and cite this paper.

Nodal metrics

Nodal metrics* enable the assessment of nodes' social heterogeneity and the understanding of underlying mechanisms such as individual characteristics (e.g., the ageing process), ecological factors (e.g., demographic variation), and evolutionary processes (e.g., differences in social styles). Node measures are calculated at a nodal level and assess, in different ways and with different meanings, how an individual is connected. Connections can be ego's* direct links only (e.g., degree, strength), its alters'* links as well (e.g., eigenvector, clustering coefficient), or even all the links in the network (e.g., betweenness). Node measures can also be used to describe the overall network structure through distributions, means, and coefficients of variation.

Degree and strength

The degree `m.net.degree` measures the number of links of a node. When computed on an undirected network, the degree represents the number of alters of an ego. When the network is directed, *it represents the number of either incoming* or outgoing* links of an ego, and it is then called in-degree `m.net.indegree` or out-degree `m.net.outdegree`, respectively. Note that degree can also be computed in directed networks; in this case, it represents the sum of incoming and outgoing links and not the number of alters.

$$D_i = \sum_{j=1}^{N} a_{ij}$$

Where a_{ij} is the value of the link between nodes i and j. Isolated node(s) can be considered as zero(s).

Strength (or weighted degree) `m.net.strength` is the sum of the links' weights in a weighted network*. When the network comprises directed links, then it is also possible to differentiate between in-strength `m.net.instrength` (the sum of weights of incoming links) and out-strength `m.net.outstrength` (the sum of weights of outgoing links). While degree and strength can be considered correlated, it may not always be the case, as individuals can interact frequently with a few social partners or vice versa (Liao, Sosa, Wu, & Zhang, 2018). Therefore, it is necessary to test their correlation prior to the analysis.

$$S_i = \sum_{j=1}^{N} a_{ij} w_{ij}$$

Where a_{ij} is the value of the link between nodes i and j. Isolated node(s) can be considered as zero(s).

Eigenvector centrality

Eigenvector centrality `m.net.eigenvector` is the first non-negative eigenvector value obtained by transforming an adjacency matrix linearly. It can be computed on weighted, binary, *directed, or undirected networks. It measures centrality* by examining the connectedness of an ego as well as that of its alters. Thus, a node's eigenvector value can be linked either to its own degree or strength or to the degrees or strengths of the nodes to which it is connected. Eigenvector may be interpreted as the social support or social capital of an individual (Brent, Semple, Dubuc, Heistermann, & MacLarnon, 2011), that is, the real or perceived availability of social resources.

$$\lambda c = W c$$

Where λ is the largest eigenvalue of the adjacency matrix W. Isolated node(s) can be considered as zero(s).

Local clustering coefficient

The local clustering coefficient `m.net.cc` measures the number of closed triplets* over the total theoretical number of triplets (i.e., open and closed), where a triplet is a set of three nodes that are connected by either two (open triplet) or three (closed triplet) edges. This measure aims to examine the links that may exist between the alters of an ego and measures the cohesion of the network. The main topological effect of closed triplets is the clustering of the network, generating cohesive clusters, and is thus strongly related to modularity (see corresponding section). The local clustering coefficient can be computed in a binary network by measuring the proportion of links between the nodes of an ego-network* divided by the number of potential links between them. In weighted networks, several versions exist, such as those from Barrat, Barthelemy, Pastor-Satorras, and Vespignani (2004) or Opsahl and Panzarasa (2009).

Binary Local Clustering Coefficient

$$C_i^b = \frac{2L}{N_i(N_i - 1)}$$

Where L is the number of links in the ego-network of node i.

Barrat's Local Clustering Coefficient

$$C_i^W = \frac{1}{S_i(D_i - 1)} \sum_{j \neq h \in N} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}$$

Where S_i and D_i are the strength and the degree of node i, respectively. w_{ij} and w_{ih} are the weights of the links, and a_{ij}, a_{ih}, a_{jh} are the links between the nodes.

Opsahl's Local Clustering Coefficient

$$C^W(G) = \frac{\sum_{\tau_\Delta} w}{\sum_\tau w}$$

Where τ_Δ represents closed triplets, and w is the chosen weighting scheme (maximum, minimum, arithmetic, or geometric mean).

Betweenness

Betweenness (WIP) is the number of times a node is included in the shortest paths (geodesic distances) generated by every combination of two nodes. The value of the betweenness indicates the theoretical role of a node in social transmission (information, disease, etc., see Figure 1), as it indicates to what extent a node connects subgroups, as a bridge, and thus is likely to spread an entity across the whole network (Newman, 2005).

$$b = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those paths that pass through v. As no paths go through isolated nodes, their betweenness value can be considered zero.

Polyadic metrics

Patterns of interactions (how and with whom individuals interact) can be examined using specific network measures* that analyse local-scale interactions within a network and make it possible to test hypotheses about the mechanisms underlying network connectivity. These types of measures are generally used to test mechanistic biological questions, such as what factors (e.g., ecological as well as sociodemographic) affect individuals' interactions/associations.

Assortativity

Assortativity (Newman, 2003) (WIP) is probably the most used measure to study homophily (preferential associations or interactions among individuals sharing the same characteristics; Lazarsfeld & Merton, 1954). Assortativity values range from −1 (total disassortativity, i.e., all the nodes associate or interact with those with the opposite characteristic, such as males interacting exclusively with females) to 1 (total assortativity, i.e., all the nodes associate or interact with those with the same characteristic, such as males interacting only with males). The assortativity coefficient measures the proportion of links between and within clusters of nodes with the same characteristics. Individuals' characteristics can be continuous (e.g., age, individual network measure, personality) or categorical features (e.g., sex, matriline belonging; Figure 2). Assortativity does not consider directionality* and can be measured in weighted (Leung & Chau, 2007) or binary (Newman, 2003) networks using categorical or continuous characteristics (Figure 2). The use of one or the other assortativity variant depends on the type of characteristics being examined and, whenever possible, the weighted version should be preferred since it is more reliable than the binary version (Farine, 2014).

Binary Assortativity

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

Where e_{ii} is the proportion of specific links, a_i is the proportion of outgoing links, and b_i is the proportion of incoming links.

Weighted Continuous Assortativity

$$r = \frac{\sum_i e_{ii}^w - \sum_i a_i^w b_i^w}{1 - \sum_i a_i^w b_i^w}$$

Where e_{ii}^w is the proportion of weighted links, and a_i^w, b_i^w are the proportions of weighted outgoing and incoming links.

Transitive triplets

Transitive triplets (WIP) are closed triplets where the links among the nodes follow a specific temporal pattern of creation, that is, when the establishment of links between nodes A and B and between nodes A and C is followed by the establishment of a link between nodes B and C. This network measure can be computed in directed, binary, or weighted networks. These types of connections can be studied over time based on the creation of links. From a static perspective, directionality can be considered by calculating the number of transitive triplets divided by the number of potential transitive triplets, and weights can also be considered by using Opsahl's variants, which are discussed in the section on local clustering coefficient (Opsahl & Panzarasa, 2009). While transitivity is importantly related to the clustering coefficient (the clustering coefficient includes transitive triplets), not all closed triplets are transitive. Transitive triplets are one of the 16 possible configurations of a triplet considering open and closed triplets as well as link directionality (i.e., triad census).

Global metrics

The structure of this section is based on the distinction between network connectivity and social diffusion (information or disease spread). However, the social diffusion section contains measures specifically designed to study theoretical (i.e., considering the diffusion is perfectly related to network links and link weights) social diffusion features based on geodesic distances (see corresponding section). Aspects of the structure and properties of a group (e.g., cohesion, sub-grouping) can be quantified using global network measures. *For instance, one may*

quantify properties such as network resilience (see Diameter), network clustering* (see Modularity) through network connectivity analysis, or network transmission efficiency* (see Global efficiency) through network theoretical social diffusion analysis.

Density

The density `m.net.density` is the ratio of existing links to all potential links in a network. This measure is easy to interpret; it assesses how fully connected a network is. Density considers neither directionality nor link weights.

$$D = \frac{2|L|}{|N|(|N| - 1)}$$

Where L is the number of links and N is the number of nodes. Isolated node(s) can be considered as zero(s).

Geodesic Distance

Geodesic distance `m.net.geodesic_distance` is the shortest path considering all potential dyads in a network. This measure thereby indicates the fastest path of diffusion. Geodesic distance can be calculated in binary, weighted*, directed, or undirected networks. In weighted networks, it can be normalized (by dividing all links by the network's mean weight), and the strongest or the weakest links can be considered as the fastest route between two nodes. This great number of variants of geodesic distance can greatly affect the results and interpretations. Researchers must thus have knowledge of the variants and know which one is the most appropriate according to their research question (Opsahl, Agneessens, & Skvoretz, 2010).

The computation uses algorithms like breadth-first search, depth-first search, or Dijkstra's algorithm. None handle isolated nodes.

Diameter

The diameter `m.net.diameter` of a network represents the longest of the shortest paths in the network. The diameter is used in ASNA to examine aspects such as network cohesion and the rapidness of information or disease transmission. While global efficiency measures the theoretical social diffusion spread, diameter informs on the maximum path length of diffusion required to reach all nodes.

Global efficiency

Global efficiency (WIP) is the ratio between the number of individuals and the number of connections multiplied by the network diameter. It provides a quantitative measure of how efficiently information is exchanged among the nodes of the network. As global efficiency gives a probability of social diffusion, it may help to better understand social transmission phenomena in the short and long term (Migliano et al., 2017). Pasquaretta et al. (2014) found a positive correlation between the neocortex ratio and global efficiency in primate species with a higher neocortex ratio. By drawing a parallel between cognitive capacities and social network efficiency, this study showed that in species with a higher neocortex ratio, individuals may adjust their social relationships to gain better access to social information and thus optimize network efficiency. Alternatively, studies on epidemiology in ant colonies showed that ants adapt their interaction rate to decrease network efficiency when infected by a pathogen (Stroeymeyt et al., 2018).

Modularity

Modularity (WIP) is a measure designed to quantify the degree to which a network can be divided into different groups or clusters, and its value ranges from 0 to 1. Networks with high modularity have dense connections within the modules but sparse connections between them. Modularity can be computed in weighted, binary, directed, or undirected networks.

$$Q = \sum_{s=1}^{m} \left[\frac{l_s}{|E|} - \left(\frac{d_s}{2|E|} \right)^2 \right]$$

Where l_s is the number of edges in the s-th community, and d_s is the sum of the degrees of the nodes in the community.

Global Clustering Coefficient

The global clustering coefficient (WIP), like the local clustering coefficient, evaluates how well the alters of an ego are interconnected and measures the cohesion of the network. Its main topological effect is the clustering of the network, generating cohesive clusters, and is thus strongly related to modularity. However, it becomes highly correlated with density and less so with modularity as density grows. Several variants of the global clustering coefficient can be found: (a) the ratio of closed triplets to all triplets (open and closed), and (b) the binary local mean clustering coefficient derived from the node level (see Local clustering coefficient). The binary local mean clustering coefficient allows us to consider node heterogeneity and thus should be preferred over the first variant. Weighted versions also exist and are based on the same variants described in the section on the local clustering coefficient and require the same considerations.

$$C^b(G) = \frac{\sum \tau_\Delta}{\sum \tau}$$

Where τ is the total number of triplets and τ_Δ represents closed triplets.

Reference(s)

Sosa, Sebastian, Cédric Sueur, and Ivan Puga-Gonzalez. 2021. "Network Measures in Animal Social Network Analysis: Their Strengths, Limits, Interpretations and Uses." *Methods in Ecology and Evolution* 12 (1): 10–21. https://doi.org/https://doi.org/10.1111/2041-210X.13366.