Glossary of Network Characteristics

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Size](#_Toc129094439)

[Total Size](#_Toc129094440)

[Effective Size](#_Toc129094441)

[K-core size](#_Toc129094442)

[Communication Frequency](#_Toc129094443)

[Degree Centrality](#_Toc129094444)

[Out-degree Centrality](#_Toc129094445)

[In-degree Centrality](#_Toc129094446)

[Density](#_Toc129094447)

[Communication Efficiency](#_Toc129094448)

[Efficiency (Burt, 1995)](#_Toc129094449)

[Local efficiency (Latora, Vito, and Massimo Marchiori., 2001)](#_Toc129094450)

[Embeddedness](#_Toc129094453)

[Closeness Centrality](#_Toc129094454)

[Harmonic Centrality](#_Toc129094455)

[Number of Triangles](#_Toc129094456)

[Connectedness](#_Toc129094457)

[Eigenvector Centrality](#_Toc129094458)

[Second Order Centrality](#_Toc129094459)

[PageRank](#_Toc129094460)

[Reach](#_Toc129094461)

[Eccentricity](#_Toc129094462)

[Local Reaching Centrality (Reach)](#_Toc129094463)

[Brokerage](#_Toc129094464)

[Constraint (non-redundancy)](#_Toc129094465)

[Betweenness centrality](#_Toc129094466)

[Information Centrality](#_Toc129094467)

[Communicability betweenness centrality](#_Toc129094468)

[Assortativity/homophily](#_Toc129094469)

[Assortativity coefficient (Newman, 2003)](#_Toc129094470)

Size

* Total Size
* Effective Size
* K-core size

Communication Frequency

* Degree Centrality
* Out-degree Centrality
* In-degree Centrality
* Density

Communication Efficiency

* Efficiency (Burt, 1995)
* Local efficiency (Latora, Vito, and Massimo Marchiori., 2001)

Embeddedness

* Closeness Centrality
* Harmonic Centrality
* Number of Triangles

Connectedness

* Eigenvector Centrality
* Second Order Centrality
* PageRank

Reach

* Eccentricity
* Local Reaching Centrality

Brokerage

* Constraint (non-redundancy)
* Betweenness centrality
* Information Centrality
* Communicability betweenness centrality

Assortativity/homophily

* Assortativity coefficient (Newman, 2003)

Size

Total Size

Effective Size

K-core size

The *total size* of a network is a count of all the nodes in the network. In an ego network, this can be calculated to include all neighbors 1 or 2 steps away from the ego node. Given that many of the nodes that are 1-2 steps away from the ego node are connected to each other and therefore may serve as duplicate sources of information, some nodes included in the total network size may be considered redundant contacts. The *effective size* of an ego network accounts for this redundancy. For example, if an ego is connected to 3 nodes that are all connected to each other, the total size of the network is 3, but the effective size is 1. Networkx uses this formula,

where t is the number of ties in the ego network (not including ties to ego), and n is the number of nodes (including ego).

k-core size is the count of all nodes in a k-core. The k-core can be used to identify subgroups of employees who are particularly well-connected within the organization. A k-core is a maximal subgraph that contains nodes of degree k or more. To compute the k-core of a social network, nodes with a degree less than k are removed from the network, iteratively (starting with 1 unless otherwise specified), until no such nodes remain.

Communication Frequency

* Degree Centrality
* Out-degree Centrality
* In-degree Centrality
* Density

The density of a node refers to the ratio of the number of existing edges to the number of possible edges.

Communication Efficiency

* Efficiency (Burt, 1995)
* Local efficiency (Latora, Vito, and Massimo Marchiori., 2001)

*Efficiency* of an employee’s network refers to the amount of non-redundant communication. It is calculated by dividing the effective size (defined above) by the degree centrality (defined above) of a node.

*Local efficiency* is a measure of how well connected a node's immediate neighbors are to each other when the node itself is removed from the network. More formally, if we let G be the original network and G(i) be the subnetwork consisting of the immediate neighbors of node i (i.e., the subnetwork obtained by removing node i and its edges from G), then the local efficiency of node i is given by:

E\_loc(i) = [1 / (n\_i(n\_i - 1))] \* sum(d(j, k))

where n\_i is the degree of node i (i.e., the number of immediate neighbors of node i), and d(j, k) is the shortest path length between nodes j and k in the subnetwork G(i). Note that local efficiency is a measure of how efficiently information can flow within a local neighborhood of a node, while global efficiency (which is computed differently) is a measure of how efficiently information can flow across the entire network.

Embeddedness

Closeness Centrality

Harmonic Centrality

Number of Triangles

*Closeness centrality* is the reciprocal of the sum of the shortest path lengths from a given employee to all other employees in the network. It is represented as a number between 0 and 1, where higher numbers mean greater closeness (lower average distance). Employees with high closeness centrality are those who can quickly reach other employees through a few email exchanges.

*Harmonic centrality* is calculated by summing the harmonic mean of the distances between a given employee and all other employees in the network. The harmonic mean of the distances between a given employee and all other employees is the reciprocal of the average of the reciprocals of the distances. Employees with high harmonic centrality are those who are reachable to many other employees in the network. Yannick Rochat, in a talk at Application of Social Network Analysys (ASNA 2009) observed that in an undirected graph with several disconnected components, the inverse of the harmonic mean of distances offers a better notion of centrality than closeness, because it weights less those elements that belong to smaller components. Tore Opsahl made the same observation in a March 2010 blog posting. Raj Kumar Pan and Jari Saram¨aki deviated from the classical definition of closeness in [Pan and Saram¨aki 11], using, in practice, harmonic centrality, with the motivation of better handling disconnected nodes (albeit with no reference to the harmonic mean).

*Number of triangles.* For unweighted graphs, the clustering of a node is the fraction of possible triangles through that node that exist. This should represent the cliquiness of a network.

Connectedness

Eigenvector Centrality

Second Order Centrality

PageRank

*Eigenvector centrality* computes the centrality for a node based on the centrality of its *direct* neighbors. An actor who is high on eigenvector centrality is connected to many actors who are themselves connected to many actors. *Second order centrality* is based on eigenvector centrality and takes into account the weight of edges. To calculate second-order centrality, one approach is to first calculate the eigenvector centrality for each node in the network. Then, for each node, the second-order centrality is calculated as the sum of the eigenvector centralities of its neighbors, weighted by the strength of the connections between them. *Local reaching centrality* takes into account both the number of neighbors a node has and the extent to which those neighbors themselves have many connections. The local reaching centrality of a node in a directed graph is the proportion of other nodes reachable from that node. A node with high local reaching centrality is one that is well-connected to other well-connected nodes, and is therefore well-positioned to quickly disseminate information or influence others. Whereas eigenvector centrality is derived from degree centrality while local reaching centrality is derived from shortest paths.

*PageRank* is a measure of importance that was originally developed by Google for ranking web pages in search results. It is based on the idea that a page is important if it is linked to by other important pages. PageRank calculates a score for each node in the network based on the incoming links to that node, as well as the PageRank scores of the nodes that link to it. The calculation is based on the idea of random walks, where a hypothetical user randomly clicks on links in the network and the PageRank score of a node is proportional to the probability of ending up at that node in the long run. In contrast to eigenvector centrality, PageRank takes into account not only the importance of a node's direct neighbors, but also the importance of the nodes that link to them. This means that a node with few connections to other nodes may still have a high PageRank score if it is linked to by other important nodes. Eigenvector centrality, on the other hand, only takes into account the importance of a node's direct neighbors, and does not consider the importance of the nodes that link to them.

Another difference is that PageRank can be more robust to certain types of network structures, such as networks with many nodes that have no or few connections, while eigenvector centrality can be more sensitive to such structures.

Reach

Eccentricity

Local Reaching Centrality (Reach)

The *eccentricity* of each node is the maximum number of steps it takes to reach any other node in the network (i.e., the node’s longest shortest path in graph). In nontechnical terms, eccentricity refers to the extent to which an employee is distant from other employees in the network. More specifically, the eccentricity of a node is the maximum number of steps it takes to reach any other node in the network. In the case of email communication, this means that an employee with high eccentricity sends and receives emails to and from employees who are further away in the network, rather than having direct communication with many of their colleagues. High eccentricity can indicate that an employee is isolated or has a unique role in the network that requires them to communicate with a specific subset of employees. Alternatively, it can also suggest that the employee is a central point of communication between different groups or departments in the organization, which can be valuable for information flow and coordination.

*Local reaching centrality* is based on the idea that a node is important if it can easily reach other nodes in the network. It calculates a score for each node in the network based on the number of nodes that can be reached from that node within a certain number of steps (i.e., the node's "reachability" in the network). The calculation is typically based on the concept of a "local neighborhood" around each node, with nodes that have larger local neighborhoods being assigned higher centrality scores. Whereas PageRank takes into account the importance of a node's direct and indirect neighbors, local reaching centrality only considers the node's direct neighbors and their reachability in the network. This means that a node with few direct connections to other nodes may still have a high PageRank score if it is linked to by other important nodes, while a node with a large local neighborhood but few direct connections may have a high local reaching centrality score but a low PageRank score. Also, PageRank is typically based on a global analysis of the entire network, while local reaching centrality is based on a local analysis of each node's neighborhood. This means that PageRank may be more appropriate for identifying the most important nodes in large, complex networks, while local reaching centrality may be more appropriate for identifying the most important nodes in smaller, more tightly-knit networks.

Brokerage

Constraint (non-redundancy)

Constraint is the extent to which an employee is connected to other employees who are not directly connected to each other. Often, a high constraint size indicates that an employee is positioned in an influential role in the network, as they are connecting people who would not otherwise be connected. In general however, constraint may be considered a measure of non-redundancy. For example, if employee A sends emails to employee B and employee C, and employee B also sends emails to employee C, then there is redundancy in the relationships between these employees. However, if employee A sends emails to employee B and employee C, but employee B does not send emails to employee C, then employee A is positioned in a way that creates a constraint because they are the only connection between employee B and employee C.

Betweenness centrality

Information Centrality

*Information centrality* is calculated by measuring the number of shortest paths between pairs of employees that pass through a given employee. Employees with high information centrality are those who are on many of the shortest paths between pairs of other employees in the network. As such, information centrality reflects the extent to which an employee is a gatekeeper or a bottleneck in the flow of information within the network. In general, actors with higher information centrality are predicted to have greater control over the flow of information within a network; highly information-central individuals tend to have a large number of short paths to many others within the social structure. Because the raw centrality values can be difficult to interpret directly, rescaled values are sometimes preferred (see the rescale option).

Communicability betweenness centrality

*Communicability betweenness centrality* takes into account the number and strength of all paths between pairs of nodes in the network. It is based on the concept of communicability, which measures the ease with which information can flow between nodes in a network. To calculate communicability betweenness centrality for a node, one approach is to first compute the communicability matrix for the network. This matrix contains the pairwise communicabilities between all nodes in the network, which can be calculated using matrix exponentiation. Then, for each node, the communicability betweenness centrality is calculated as the sum of the communicability contributions of all pairs of nodes that pass through that node. The communicability contribution of a pair of nodes is defined as the ratio of the number of paths between those nodes that pass through the node in question, to the total number of paths between those nodes. Unlike traditional betweenness centrality measures, which focus on shortest paths, communicability betweenness centrality takes into account all paths in the network and their relative strengths.

Assortativity/homophily

Assortativity coefficient (Newman, 2003)

The assortativity coefficient can be used to identify patterns of homophily or heterophily in the network based on the distribution of degrees among nodes. Homophily refers to the tendency for nodes with similar attributes to be connected to one another, while heterophily refers to the tendency for nodes with dissimilar attributes to be connected. The assortativity coefficient is a correlation coefficient measuring the relationship between the number of links between nodes of one attribute (e.g., females) and the number of links among nodes of another attribute (e.g., males). The assortativity coefficient ranges from -1 to 1, with values closer to 1 indicating strong assortative mixing and values closer to -1 indicating strong disassortative mixing. A value of 0 indicates random mixing, where nodes are connected to other nodes with degrees that are independent of their own degree.