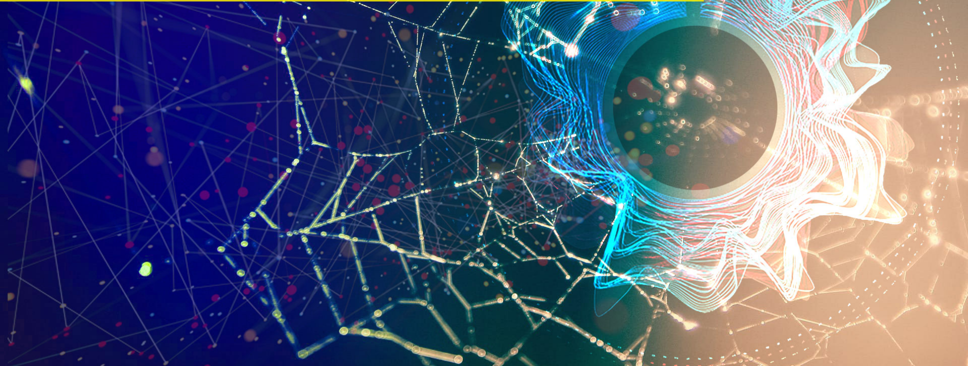
Centrality in networks
Finding the most important nodes

Sergio Gómez



Centrality in networks

Finding the most important nodes

S. Gómez[1]

[1]Dept. Computer Engineering and Mathematics
Rovira i Virgili University, Spain.

08th Dec. 2017

Outline

Centrality in networks

Datasets in business and consumer analytics can be frequently represented in the form of networks, in which the nodes represent any kind of item, e.g. products, consumers, brands, firms, etc., while the links represent relationships between them. For example, in co-purchasing networks, the links could account for pairs of products bought together, whereas in international trade networks the edges could represent the amount of a product which is exported from one country to another one.

The possibilities are infinite, and the extraction of information from these networks is the object of study in several fields, from complex networks and complex systems to data science, among others. Here we aim at finding the most important nodes in a network, which could be crucial in many business applications.

The importance of a node in a complex network depends on the structural characteristic or dynamic behavior we could be interested in. As a consequence, the literature is full of different definitions, all of them perfectly meaningful under specific set-ups. Our objective is to explain the rationale behind the most widely used centrality measures, to be able to decide which one is the more adequate for our needs.

Most of them are easy to describe and understand, some are also easy to calculate with the appropriate tools, while others represent a computational challenge which requires the use of complex algorithms which are not easy to implement. Fortunately, there exist several software applications and packages which simplify the finding of the centralities of the nodes in complex networks.

Mathematical Notation

Centrality in networks

The main mathematical object in the study of complex networks is the adjacency matrix $A = (a_{ij})$, which encodes the full topology of the network or graph: $a_{ij} = 1$ if there is an edge from node i to node j, and $a_{ij} = 0$ otherwise. We suppose the network has N nodes, thus $i, j \in \{1, \ldots, N\}$, and that there are no self-loops, i.e. $a_{ii} = 0$. If the direction of the links is not important, the network is called undirected, and the adjacency matrix is symmetric, $A = A^T$, where A^T denotes the transpose of matrix A. For undirected networks, the degree k_i of a node is its number of neighbors, and is calculated as

$$k_i = \sum_{j=1}^{N} a_{ij} \,. \tag{1}$$

Mathematical Notation

Centrality in networks

Directed networks require the distinction between the links that arrive to a node and those that depart from it, therefore it is convenient to distinguish between the output and input degrees:

$$k_i^{\text{out}} = \sum_{j=1}^{N} a_{ij} \,, \quad (2)$$

$$k_i^{\text{in}} = \sum_{j=1}^{N} a_{ji} \,. \quad (3)$$

Of course, if the network is undirected, $k_i^{\text{out}} = k_i^{\text{in}} = k_i$. We will try to describe the centrality measures in the general case of directed networks, since undirected networks can be considered just as particular cases. However, there are definitions of centrality which do not make sense or cannot be calculated for certain kinds of networks, thus we will explicitly explain the applicability of each centrality type. We will also suppose there are no self-loops in the network, thus all the diagonal elements of the adjacency matrix are zero, $a_{ii} = 0$.

Mathematical Notation

Centrality in networks

The number of edges is calculated by just taking the sum of all the components of the adjacency matrix:

$$2L = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \, . \tag{4}$$

The number of edges is L for undirected networks, but $2L$ for directed ones. The reason is that the adjacency matrix of undirected networks counts every edge twice, $a_{ij} = a_{ji} = 1$.

Centrality in networks - Content outline

- Degree centrality
- Closeness centrality
- Betweenness centrality
- Eigenvector centrality
- Katz centrality
- Hubs and Authorities centrality
- PageRank centrality
- Random walk centralities

Degree centrality

The first and simplest proposal of a centrality measure for the nodes in a network is the degree,

$$C_i^{(\mathrm{deg})} = k_i \, . \tag{5}$$

This is a concept which was developed in the context of social networks long time ago [53, 29]. The idea was that a person having many direct connections to other people must be central with respect to the communication between them, acquiring a sense of being in the mainstream of information. On the contrary, people with low degree could miss most of the information flowing in the network, thus playing a residual role.

Degree centrality

Nodes with high degree, clearly above the average in the network, are called hubs. The discovery that many real-world networks have power-law degree distributions [4], with only a few hubs collecting a great proportion of the overall links in the network, was in fact one of the cornerstones in the development of the actual theory of complex networks.

Sometimes it is useful to normalize the centralities considering their maximum value, which for the degree equals $N - 1$, thus

$$C_i^{(\text{deg,norm})} = \frac{k_i}{N-1} \,. \tag{6}$$

However, normalization is usually not needed, since what matters is the rank of the nodes after sorting them according to the selected centrality measure (which does not change with normalization). Several additional centrality measures were defined as variants of the degree (see e.g. [53, 31, 48, 44]), but they have become outdated, so we just skip them.

Degree centrality

The degree is a simple an effective centrality measure for undirected networks, but not for directed ones, in which we have to distinguish between incoming and outgoing links. A possible approach could be to take as centrality the sum, $k_i^{\text{in}} + k_i^{\text{out}}$, i.e. the total number of connections discarding their directionality, or the average of both input and output degrees, $(k_i^{\text{in}} + k_i^{\text{out}})/2$; the average is more convenient because it coincides with the degree when applied to undirected networks:

$$C_i^{(\text{deg,avg})} = \frac{k_i^{\text{in}} + k_i^{\text{out}}}{2}. \qquad (7)$$

Another alternative consists in defining two degree centralities, one for the incoming and the other for the outgoing links, since they measure different things: a node with high input degree centrality represents a node which is in good position to receive information, while large output degree centralities correspond to important sources of information:

$$C_i^{(\mathrm{deg,out})} = k_i^{\mathrm{out}}, \tag{8}$$

$$C_i^{(\mathrm{deg,in})} = k_i^{\mathrm{in}}, \tag{9}$$

Now, it becomes clear why the importance of a node is closely related to the process or property we are interested in, since even degree centrality admits several diverging interpretations in directed networks.

Closeness centrality

Centrality in networks

If you have items distributed within a circle, its center has the property that all the items are at a distance equal or smaller than the radius, while other positions may be as much as twice that distance. This suggests that a measure of centrality in networks could consider the distances to the rest of nodes, and thus central nodes would be close to all of them. The advantage of being central in this way comes from the possibility of sending or broadcasting information, being sure the time needed to reach the whole network is as short as possible.

Closeness centrality

Closeness centrality is based on this idea: for each node, you calculate the distance to all the other vertices in the network, and define a centrality in which shorter distances imply higher closeness centrality, and vice versa. There are several ways of expressing mathematically this concept. First, let us call d_{ij} the distance between nodes i and j. The distance in a graph is defined as the minimum number of hops (following links) needed to move from one node to another, or, in other words, the length of the shortest path between them.

Closeness centrality

Centrality in networks

Then, the closeness centrality [9, 50] reads

$$C_i^{(\mathrm{clos})} = \frac{1}{\sum_{j=1}^{N} d_{ij}}\,, \tag{10}$$

which can be normalized [10] as

$$C_i^{(\mathrm{clos,norm1})} = \frac{N-1}{\sum_{j=1}^{N} d_{ij}}\,, \tag{11}$$

or also as

$$C_i^{(\mathrm{clos,norm2})} = \frac{N}{\sum_{j=1}^{N} d_{ij}}\,. \tag{12}$$

Closeness centrality

The difference between using $N-1$ or N is irrelevant for the ranking of the nodes. The $N-1$ makes sense since the distance from a node to itself is always zero, $d_{ii}=0$, but the N provides simpler expressions for certain analytic derivations. Here we are supposing the network is connected (strongly connected if directed), otherwise some of the distances are infinity and the closeness centrality of all nodes becomes zero. To avoid these infinities, a simple heuristic consists in replacing each infinite distance by N, i.e. a value larger than all the finite distances.

Closeness centrality

Centrality in networks

An alternative definition which maintains the infinities and works even if the network is not connected is found by just swapping the reciprocal and sum operations [24]:

$$C_i^{(\mathrm{clos2})} = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{d_{ij}}, \tag{13}$$

where, by convenience, $d_{ij} = \infty$ if there is no path between i and j, i.e. $1/d_{ij} = 0$. The term $1/d_{ii}$ is explicitly excluded from the sum to avoid the corresponding infinity. Equation (13) may be viewed as a centrality based on the harmonic mean of the distances, and has the advantage that most of the contribution comes from the distances to the closer nodes.

Closeness centrality

Likewise degree centrality, closeness centrality also admits output and input versions for directed networks, depending on whether the distances are computed from or to the reference node, respectively. Note that distances are not symmetric in directed networks.

Since we already have several definitions for the closeness centrality, the addition of input and output closeness centralities multiplies the options. This is important to be aware of, since different software may choose and implement centralities in distinctive ways, thus being not exactly comparable.

Betweenness is another of the traditional centrality measures developed in the study social science. Here we fix our attention in the nodes which are crossed when you follow shortest paths. A node which falls in the communication paths between many pairs of nodes plays an important role, since it can control the flow of information. Formally, the standard measure for this property is called betweenness centrality [2, 28], and is defined as

$$C_i^{(\text{betw})} = \frac{1}{(N-1)(N-2)} \sum_{\substack{s,d=1 \\ s \neq d \neq i}}^{N} \frac{\sigma_{sd}(i)}{\sigma_{sd}} . \tag{14}$$

The sum covers all source/target pairs of nodes, excluding node i, σ_{sd} represents the number of shortest paths from source node s to destination node d, and $\sigma_{sd}(i)$ is the number of those paths that include node i. In other words, the betweenness is the average fraction of paths that cross a node. This expression of the betweenness is valid for both directed and undirected networks, and includes the optional normalization factor.

If there are no paths between the origin s and the destination d (disconnected graph), then $\sigma_{sd} = 0$ and it becomes necessary to define $\sigma_{sd}(i)/\sigma_{sd} = 0$. An example of a node with high betweenness would be a node which is a bridge between two disconnected parts of the network: to go from one part of the network to the other you are forced to cross the bridge, no matter if this node has just a few links.

Betweenness naturally appears in communication dynamics on top of complex networks, e.g. it can be shown that the onset of congestion in a simple model of routing is related to the maximum betweenness of the system [33].

Betweenness centrality

Centrality in networks

The calculation of both closeness and betweenness centrality can be very costly, since the standard Floyd-Warshall algorithm to find all the shortest paths in a graph scales as $O(N^3)$ [26]. Fortunately, we may apply the Brandes' algorithm, with a cost $O(NL + N^2 \log N)$, which is reduced to $O(NL)$ for the unweighted networks we have considered so far [16].

Betweenness centrality

There exist some variations on the definition of betweenness, the most remarkable one being the possibility of including node i as both source s and destination d [43], which we have forbidden in our previous definition. The decision of including or not the end-points of the paths when calculating the betweenness depends on the particular dynamics you may be interested in.

For example, in routing dynamics in which a queue is attached to each node, it is possible to decide between putting the created packets in the queue of the source node [55], or skipping this queue and enqueuing them directly to the first neighbor in the path [33].

Both alternatives are acceptable, but they lead to slightly different values of the betweenness. Another variant of betweenness is the one which calculates the number of shortest paths at the level of edges, thus defining a link betweenness, the natural extension to links of the vertex betweenness. We are not going to consider link centralities in the rest of this chapter, but it may be useful for the reader to know of their existence and one of their paradigmatic examples.

Eigenvector centrality

All the previous centrality measures take into account the topological position of nodes in the network, but not the importance of the nodes themselves. It could be desirable, for example, that a node be considered as important if its neighbors are also important. This leads to a recursive definition of centrality, in which the centrality of a node depends on the centralities of the neighbors, which are also unknown. Fortunately, it is possible to write self-consistent equations which can be easily solved using linear algebra techniques.

Eigenvector centrality

The simplest of this kind of approaches consists in defining the centrality of a node as proportional to the sum of the centralities of the neighbors, so as the larger the importance of the neighbors, the more central the node is [15, 13, 14]. In mathematical terms,

$$\lambda C_i^{(\mathrm{eig})} = \sum_{j=1}^{N} a_{ji} C_j^{(\mathrm{eig})} , \tag{15}$$

where λ is the proportionality constant. The a_{ji} term emphasizes that node i receives the contribution to centrality from its neighbors through the incoming links.

For example, in the World Wide Web network, building a website with many links to important sites is easy to build and has no cost, so it gives no information at all. However, receiving hyperlinks from relevant sites is a good indicator of quality, and can be used to measure the centrality of the website.

Equation (15) is expressed in matrix form as

$$A^T \mathbf{C}^{(\mathrm{eig})} = \lambda \mathbf{C}^{(\mathrm{eig})}, \tag{16}$$

which means the vector of centralities $\mathbf{C}^{(\mathrm{eig})}$ is an eigenvector of A^T (or equivalently, a left-eigenvector of A) with eigenvalue λ. Since the components of the adjacency matrix are all non-negative, we can apply the Perron-Frobenius theorem [47, 30], which ensures that, if the matrix is irreducible, there exists a unique solution of Eq. (16) in which all the centralities $C_i^{(\mathrm{eig})}$ are positive (up to positive common factors), and which corresponds to the largest eigenvalue $\lambda > 0$.

The matrix is irreducible if the graph is strongly connected (or simply connected, if the network is undirected). For directed networks this condition is difficult to be fulfilled, thus eigenvector centrality is basically used only for undirected networks. Some variants of the eigenvector centrality, such as Katz, HITS or PageRank, are more adequate for directed networks.

Eigenvector centrality

Centrality in networks

The calculation of the eigenvector centrality can be easily performed by power iteration: initialize all the centralities to one, multiply by A^T, normalize the vector, and repeat the multiplication-normalization steps until convergence. Common normalizations used are those in which the sum of all centralities are either 1 or N. Again, the normalization does not affect the ranking of the nodes, thus any choice is equally acceptable.

Katz centrality is a proposal that lays between degree and eigenvector centrality. It was introduced as a way of generalizing the degree centrality, taking into account not only the immediate neighbors but also the nodes reachable in larger number of steps [35]. Since you want that the closer the nodes, the larger their influence, a decay parameter $\alpha < 1$ is introduced to weight the contributions of nodes at increasing path lengths. It is defined in this way:

$$C_i^{(\text{katz})} = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k (A^k)_{ji} \,. \tag{17}$$

The power matrix A^k accounts for the number of paths between every pair of nodes, e.g. $(A^3)_{ji} = \sum_r \sum_s a_{jr} a_{rs} a_{si}$, where the paths start at node j, then go to node r, next to s and finally arrive to i, for all possible values of the intermediate nodes r and s.

Denoting I the identity matrix of order N, and $\mathbf{1}$ the vector of length N with all components equal to 1, we can write

$$\mathbf{C}^{(\text{katz})} = \sum_{k=1}^{\infty} (\alpha A^T)^k \mathbf{1} = \left((I - \alpha A^T)^{-1} - I \right) \mathbf{1} \,, \qquad (18)$$

which, after some algebra, becomes

$$\mathbf{C}^{(\text{katz})} = \alpha A^T (\mathbf{C}^{(\text{katz})} + \mathbf{1}) \,, \qquad (19)$$

or in components

$$C_i^{(\text{katz})} = \alpha \sum_{j=1}^{N} a_{ji} (C_j^{(\text{katz})} + 1) \,. \qquad (20)$$

Katz centrality

Centrality in networks

Equations (19) and (20) are closely related to the eigenvector centrality Eqs. (16) and (15), respectively. Basically, the Katz centrality of a node is related to the centralities of the incoming neighbors, likewise eigenvector centrality, but with the addition of one unit per neighbor. This means all nodes have a minimum level of centrality, different from zero, which helps to avoid the problems of eigenvector centrality with non-strongly connected components.

Of course, the α parameter has to be small enough to ensure the convergence of Eq. (17), and of the iteration process. It can be shown that proper values of the parameter must be in the interval $0 < \alpha < 1/\lambda$, where λ is the maximum eigenvalue of the adjacency matrix A.

Katz centrality can be extended by replacing the vector **1** by any other set of constants:

$$\mathbf{c}^{(\text{katz2})} = \alpha A^T \mathbf{c}^{(\text{katz})} + \beta\,, \tag{21}$$

Katz centrality

Centrality in networks

This is useful to allow each node i to have a minimum centrality β_i, which could be set even from external information of the nodes, unrelated to the network structure.

When α approaches zero most of the contribution to the Katz centrality comes from the constant term β, while α values close to its upper bound $1/\lambda$ give the dominant role to the eigenvector term. In practice, most of the authors use values of the parameter near the upper bound.

In directed networks, nodes can have very different roles if we consider only the input or output links. The idea of the Hyperlink-Induced Topic Search (HITS) approach, also known as hubs and authorities' algorithm [37], is to assign to each node a couple of scores: a hub centrality, which takes into account the role of the node in sending links, and an authority centrality, measuring the capacity of the node to receive links. Using the same approach that eigenvector centrality, the importance as authority depends on the relevance of the hubs that send the incoming links, and the other way around, important hubs give more weight as authorities to the receiver nodes.

Hubs and Authorities centrality

Centrality in networks

Denoting $C_i^{(\text{hub})}$ and $C_i^{(\text{auth})}$ the hub and authority centralities of node i, the following recursive definition holds:

$$C_i^{(\text{auth})} = \alpha \sum_{j=1}^{N} a_{ji} C_j^{(\text{hub})} , \qquad (22)$$

$$C_i^{(\text{hub})} = \beta \sum_{j=1}^{N} a_{ij} C_j^{(\text{auth})} . \qquad (23)$$

Hubs and Authorities centrality

In matrix form,

$$\mathbf{c}^{(\text{auth})} = \alpha A^T \mathbf{c}^{(\text{hub})}, \tag{24}$$

$$\mathbf{c}^{(\text{hub})} = \beta A \mathbf{c}^{(\text{auth})}, \tag{25}$$

which can be combined to form two decoupled equations:

$$A^T A \mathbf{c}^{(\text{auth})} = \gamma \mathbf{c}^{(\text{auth})}, \tag{26}$$

$$A A^T \mathbf{c}^{(\text{hub})} = \gamma \mathbf{c}^{(\text{hub})}, \tag{27}$$

where $\gamma = (\alpha\beta)^{-1}$. Applying the Perron-Frobenius theorem as for the eigenvector centrality, and realizing that matrices $A^T A$ and AA^T are symmetric, then the authorities and hubs centralities are given by the leading eigenvector of their respective matrices. Moreover, it can be shown that the eigenvalues of $A^T A$ and AA^T are exactly the same, thus the two equations are consistent and γ is the maximum eigenvalue of any of them.

Additionally, multiplying both sides of the first equation by A and of the second equation by A^T, we get

$$AA^T(A\mathbf{C}^{(\text{auth})}) = \gamma(A\mathbf{C}^{(\text{auth})}), \quad (28)$$

$$A^TA(A^T\mathbf{C}^{(\text{hub})}) = \gamma(A^T\mathbf{C}^{(\text{hub})}), \quad (29)$$

which means that hubs and authorities centralities are related in the following way:

$$\mathbf{C}^{(\text{auth})} = A^T\mathbf{C}^{(\text{hub})}, \quad (30)$$

$$\mathbf{C}^{(\text{hub})} = A\mathbf{C}^{(\text{auth})}. \quad (31)$$

Hubs and Authorities centrality

Centrality in networks

This framework was designed to rank web pages, but is perfectly valid for all kinds of directed networks, e.g. citations or trade networks. When the network is undirected the distinction between hubs and authorities disappears, and their centralities coincide with those obtained by eigenvector centrality.

PageRank centrality

PageRank has become a notorious centrality measure since it lays at the core of the Google search engine. When you make a search query, the PageRank score of each web page is used to sort the results, which are then presented to the user. Of course, PageRank is in fact used in conjunction with other heuristics and criteria, but at least it provides a good starting point.

The rationale behind PageRank is similar to eigenvector centrality, but with a relevant distinction: when a node receives a link from an important source, it is not the same if that site has many links or just a few. If the number is large, the contribution is diluted, and should be penalized. Thus, it seems reasonable to normalize the score of a node by its number of outgoing links, before adding it to the score of the receiver.

The full equation for the PageRank centrality is the following [18]:

$$C_i^{(\mathrm{pr})} = \alpha \sum_{j=1}^{N} a_{ji} \frac{C_j^{(\mathrm{pr})}}{k_j^{\mathrm{out}}} + \frac{1-\alpha}{N} \,. \tag{32}$$

The constant term plays an equivalent role as in Katz centrality, ensuring the equation has a unique and non-trivial solution for directed networks, while parameter α, known as the dumping factor, controls the fraction of contribution between the eigenvector and constant terms.

Note that PageRank is already normalized, $\sum_i C_i^{(\text{pr})} = 1$, as can be easily checked by summing both sides of Eq. (32) for all the nodes i. For nodes with no outbound links, $k_j^{\text{out}} = 0$, but the numerator is also zero, thus a simple solution is to replace k_j^{out} by $\max(k_j^{\text{out}}, 1)$; otherwise, the terms $0/0$ are just supposed to be 0.

We may also write Eq. (32) in matrix form:

$$\mathbf{C}^{(\mathrm{pr})} = \alpha A^T D^{-1} \mathbf{C}^{(\mathrm{pr})} + \frac{1-\alpha}{N}\mathbf{1}\,, \tag{33}$$

where D is the diagonal matrix with elements $D_{ii} = \max(k_j^{\mathrm{out}}, 1)$.
In this way, the solution is given by:

$$\begin{aligned} \mathbf{C}^{(\mathrm{pr})} &= \frac{1-\alpha}{N}(I - \alpha A^T D^{-1})^{-1}\mathbf{1} \\ &= \frac{1-\alpha}{N} D(D - \alpha A^T)^{-1}\mathbf{1}\,. \end{aligned} \tag{34}$$

Anyhow, the common way of solving Eq. (32) is by iteration, as explained above. The dumping factor was set by the authors to $\alpha = 0.85$, but this is a quite arbitrary selection which can be tuned as desired.

Random walk centralities

Centrality in networks

Looking at Eq. (32) for the PageRank, a new interpretation comes out when we realize that

$$P_{ij} = \frac{a_{ij}}{k_i^{\text{out}}} \tag{35}$$

represents the probability that a random walker follows a link from node j to node i [39, 45, 57]. Matrix P, which may be written as

$$P = D^{-1}A\,, \tag{36}$$

is right stochastic, since $\sum_j P_{ij} = 1$ for all rows i, i.e. $P\mathbf{1} = \mathbf{1}$. Using P, the PageRank equation becomes

$$\mathbf{C}^{(\text{pr})} = \alpha P^T \mathbf{C}^{(\text{pr})} + \frac{1-\alpha}{N}\mathbf{1}\,. \tag{37}$$

This equation corresponds to the dynamics of a random walker which, with probability α follows a random link of the current node, and with probability $1 - \alpha$ jumps to a random node; this behavior justifies why the second term is also referred to as the teleportation term, and it is necessary to escape from nodes without output links.

Moreover, $\mathbf{C}^{(\mathrm{pr})}$ turns out to be the occupation probability of this random walker, thus providing a physical interpretation: PageRank centrality is equal to the probability of the random walker being found at each of the nodes.

If we remove the teleportation term by setting the dumping factor to $\alpha = 1$, the PageRank equation is simplified to $\mathbf{C}^{(\mathrm{pr})} = P^T \mathbf{C}^{(\mathrm{pr})}$, which has a simple solution for unweighted networks: $\mathbf{C}^{(\mathrm{pr})} = \mathbf{k} = \mathbf{C}^{(\mathrm{deg})}$, i.e. the PageRank becomes the degree. In the general case of directed networks and with teleportation this solution does not hold, but it suggests that PageRank is a kind of modified version of the degree centrality.

We have shown so far that a random walk dynamics on complex networks gives an alternative explanation of PageRank to the one inspired by eigenvector and Katz centrality. However, this is not the only centrality measure that can be defined using random walks. In fact, random walks constitute a good proxy for the spreading of information in networks, and we can take advantage of it to introduce new measures of the importance of nodes. In particular, we are going to briefly describe random-walk closeness centrality and random-walk betweenness centrality [40].

In the definition of betweenness centrality given in Sect. 3, only nodes crossed by shortest paths are considered. This makes sense for certain dynamics, e.g. vehicles trying to reach their destination minimizing the travel distance, or servers dispatching packets using the standard Internet protocols. The same can be said about closeness centrality, which implicitly assumes that shortest-path distances are the way to go from one node to another.

Random walk centralities

Centrality in networks

However, if we consider rumors, news, fads or epidemics, to name a few, their spreading is more random, and for sure they do not follow shortest paths. This is where random walkers stand out, as an alternative and often better model of information spreading, that can help in the introduction of additional measures of centrality. In fact, real propagation usually lays somewhere in-between shortest paths and random walks, the two extreme cases.

A measure of random-walk betweenness centrality requires the computation of the probability that a random walk crosses a certain node while traveling between all other pairs of nodes. This is accomplished by introducing a new transition matrix $Q^{[d]}$ with an absorbing state at the destination node d (when the random walker arrives to d, it is removed from the system),

$$Q_{ij}^{[d]} = \begin{cases} 0, & \text{if } i = d \\ P_{ij}, & \text{otherwise}, \end{cases} \tag{38}$$

Calculating the expected number of times the random walker arrives to node i (in any number of steps) when starting at node s and before reaching the destination d,

$$p_{si}^{[d]} = \sum_{n=0}^{\infty} \left[(Q^{[d]})^n \right]_{si} = \left[(I - Q^{[d]})^{-1} \right]_{si} . \tag{39}$$

and finally averaging over all possible origins and destinations,

$$C_i^{(\text{rwbetw})} = \frac{1}{N(N-1)} \sum_{\substack{s,d=1 \\ s \neq d}}^{N} p_{si}^{[d]} . \tag{40}$$

In this case we have allowed i to be in the end-points of the paths, unlike for shortest-path betweenness.

Random-walk closeness centrality follows the same idea: the distance between two nodes s and d is replaced by the average time needed by a random walker to reach d when starting the walk at s. This quantity receives the name of mean first-passage time (MFPT), and has the property of not being symmetric even for undirected networks. The MFPT in which origin and destination are the same node is known as mean return time.

The calculation of MFPT_{sd} is quite involved [43], so we skip it, but once we have obtained them, the average first-passage time becomes

$$h_d = \frac{1}{N} \sum_{s=1}^{N} \text{MFPT}_{sd} \,, \tag{41}$$

and the random-walk closeness centrality is just

$$C_i^{(\text{rwclos})} = \frac{1}{h_i} \,. \tag{42}$$

Note that we have based the definition on the paths arriving to the node for which we are calculation the centrality, thus using the same choice as for the PageRank and other centralities.

Weighted networks are those for which a certain value is assigned to each of the edges. The standard interpretation is that the larger the weight, the more connected or related are the nodes. Flows, similarities, strengths of social ties, capacities, correlations, intensities and proximities are examples of this kind of weighted relationships.

The matrix of weights w_{ij} may be seen as a generalization of the adjacency matrix, in the sense that we may consider that a null weight corresponds to the absence of a link, and in many cases we may just replace the adjacency matrix by the weights matrix to obtain generalizations of the unweighted concepts, centrality being one of them [5, 1].

Note also that the adjacency matrix is recovered if we suppose all the weights are equal to 1. The natural generalization of the degree is called the strength of the node and is given by

$$w_i = \sum_{j=1}^{N} w_{ij} \,. \tag{43}$$

Centrality in Weighted Networks

Centrality in networks

Directed networks require the distinction between input and output strengths,

$$w_i^{\text{out}} = \sum_{j=1}^{N} w_{ij} , \quad (44)$$

$$w_i^{\text{in}} = \sum_{j=1}^{N} w_{ji} , \quad (45)$$

and the total strength of the network reads

$$2w = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} . \quad (46)$$

With these ingredients, the generalization of the degree centrality would be the strength centrality, which could be normalized using the maximum strength. In the same way, eigenvector, hubs and authorities, and PageRank centralities are obtained by simple substitution of the adjacency matrix components and the degrees by weights and strengths, respectively. The Katz centrality also admits this treatment in its interpretation as an eigenvalue problem, but it is questionable the meaning of the powers of the weights matrix.

For the random-walk centralities, the weights allow to have different transition probabilities from a node to each of its neighbors,

$$P_{ij} = \frac{w_{ij}}{w_i^{out}}, \tag{47}$$

and once they are determined, the definitions of PageRank, random-walk betweenness and random-walk closeness remain the same.

The problem arises when we want to generalize centralities based on distances, like closeness or betweenness. The first option consists in discarding the weights, something which also applies to the cases above. However, when the relationship between nodes represent distances, they cannot be ignored.

For example, in geographical and transportation networks we may have available the distances between connected nodes. Now, the shortest path between two nodes is not the path with the least number of hops, but the path for which the sum of the distances of the edges (the length of the path) is the smallest one.

In these cases, the definition of closeness and betweenness centralities do not need to be changed, but the algorithms to calculate them require important modifications. For instance, while a breadth-first traversal is enough to find the distances in unweighted networks, a Dijkstra's algorithm is necessary to cope with the distances of the edges.

Another important class of networks which deserves special treatment with regards to centrality is that of interconnected multilayer networks [36, 12]. In multilayer networks the nodes are distributed in layers, with intra-layer and inter-layer links connecting nodes in the same and different layers, respectively. If every node represents a different entity, no matter in which layer it is located, it is perfectly meaningful to calculate the centralities of the nodes as if the network were not multilayer, i.e. disregarding the structure in layers.

Alternatively, we could just find the centralities of the nodes inside the layers, just by considering each layer as a separate network, ignoring the inter-layer links. These procedures lead to two centralities per node, one global and the other local to the layer. Thus, a node can be at the same time very central in a layer, but not so important for the whole multilayer network.

In interconnected multilayer networks the same node may be present in several layers at the same time, and this fact affects the definition of centrality itself. If one node has a different centrality in each layer, how do we have to aggregate them to produce a single centrality for the node?

There have been several proposals of ways to define eigenvector centralities [54, 34] and PageRank [8] for multiplex networks, which are the particular case of multilayer interconnected networks in which inter-layer links only connect instances of the same node in different layers, but not different nodes.

A more general framework makes use of the tensorial formulation of multilayer networks [22], which has allowed a grounded development of the extension of centrality measures to general multilayer networks [23, 56]. The remarkable finding is that centrality in interconnected multilayer networks reveals the most versatile nodes, in the sense that the highest centrality (versatility) is assigned to nodes which are not necessarily very central in any layer but which are fundamental for the cohesiveness and integration of the whole structure [23].

We are not going to develop all the theory of centrality (versatility) for multilayer networks, but it is easy to show the main ideas with eigenvector centrality. First, the replacement of the adjacency matrix for multilayer networks is the adjacency tensor $M_{j\beta}^{i\alpha}$, representing the links between nodes i in layer α and nodes j in layer β. The eigentensor equation becomes:

$$\sum_{i=1}^{N}\sum_{\alpha=1}^{U} M_{j\beta}^{i\alpha} C_{i\alpha}^{(\text{eigvers})} = \lambda C_{j\beta}^{(\text{eigvers})}, \quad (48)$$

where U is the number of layers.

After solving this equation for the largest eigenvalue, the final eigenvector centrality (versatility) is obtained by summing up the contributions at each layer:

$$C_j^{(\text{eigvers})} = \sum_{\beta=1}^{U} C_{j\beta}^{(\text{eigvers})}. \quad (49)$$

Note that Eq. (48) takes into account the complete structure of the multilayer network, unlike some approaches in which layers are analyzed as isolated layers, thus losing the information of the inter-layer connectivity.

In a similar way, centralities based on distances or random walkers make use of the full structure of the network, but at the same time the multiplicity of the nodes in the different layers pose restrictions on the paths.

For example, although paths may change layer crossing inter-layer links, it is natural to consider that shortest paths from an origin to a destination must start and end, respectively, in the layers that minimize the distance.

As a consequence, shortest paths in multilayer networks cannot be found by iterating over all pairs of nodes, ignoring the multilayer structure. This demonstrates the fundamental differences between standard and interconnected multilayer networks, and how they affect the structural and dynamical properties on top of them.

We have designed a couple of small networks, one undirected and the other directed, to grasp the differences between the most central nodes according to each of the definitions of centrality we have elaborated above. Figures 3 and 4 show them for the undirected network, while Figures 7 and 9 for the directed network.

In addition, Tables 1 and 2 enumerate the most, second most, and third most central nodes for the undirected and directed networks, respectively. These networks have been designed in such a way that each centrality measure leads to different most central nodes, with few coincidences, to emphasize the topological features which distinguish them. Note that the simmetries present in the networks are responsible of the existence of several distinct nodes with exactly the same centrality.

Table 1: Most central nodes of undirected network in Figs. 3 and 4.

Centrality	Most central	Second most central	Third most central
Degree	28	19, 24	6, 7, 10, 11, 16, 20, 22
Closeness	18	16	24
Eigenvector	19	20, 22	21, 23
Katz	28	19	20, 22
Betweenness	28	24	18
PageRank	28	24	36, 39
Random-walk betweenness	28	19, 24	6, 7, 10, 11, 16, 20, 22
Random-walk closeness	18	16	24

Table 2: Most central nodes of directed network in Figs. 7 and 9.

Centrality	Most central	Second most central	Third most central
Input degree	8, 14	25, 27	23
Output degree	13	1, 12	23, 27
Input closeness	14	22	20
Output closeness	13	19	16
Eigenvector	23	27	25
Katz	27	25	28, 29, 30, 31
Hub	13	12	1
Authority	14	8	7
Betweenness	20	23	8
PageRank	31	30	27

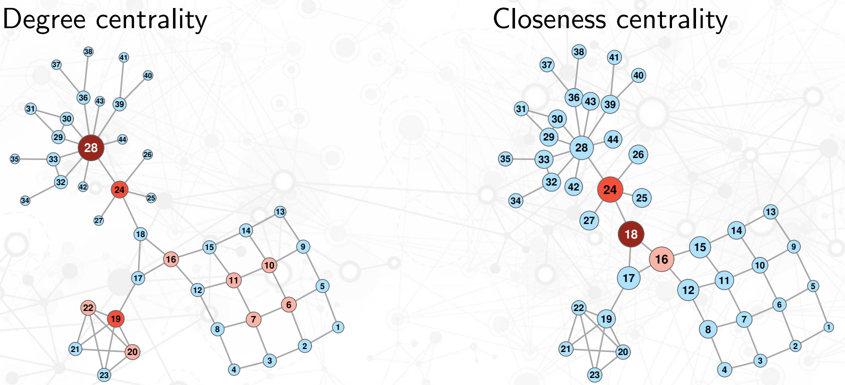


Figure 1: Centralities for an undirected network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks

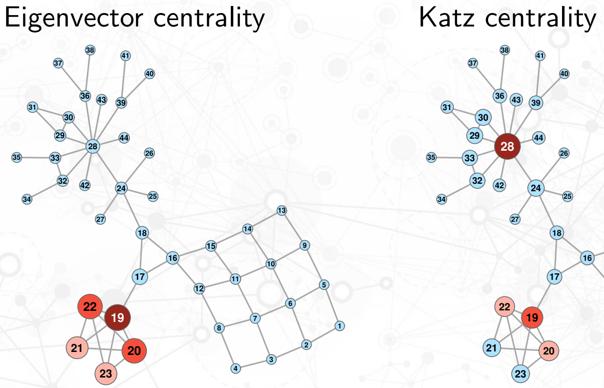


Figure 2: Centralities for an undirected network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

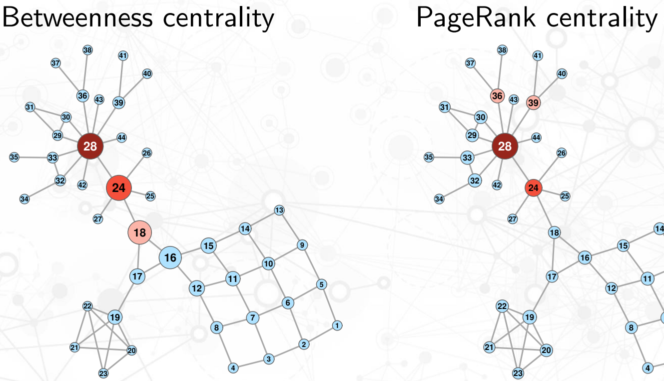


Figure 3: Centralities for an undirected network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Random-walk betweenness centrality and Random-walk closeness centrality

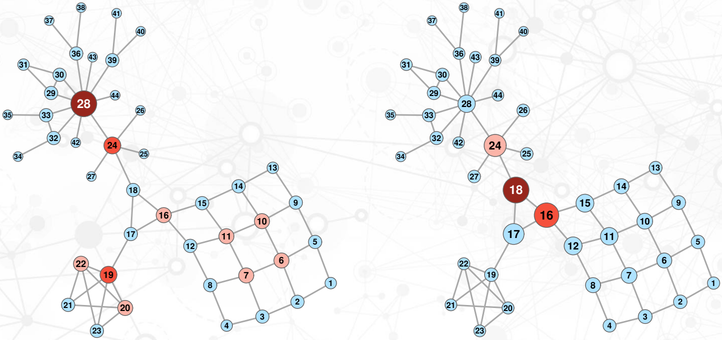


Figure 4: Centralities for an undirected network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks

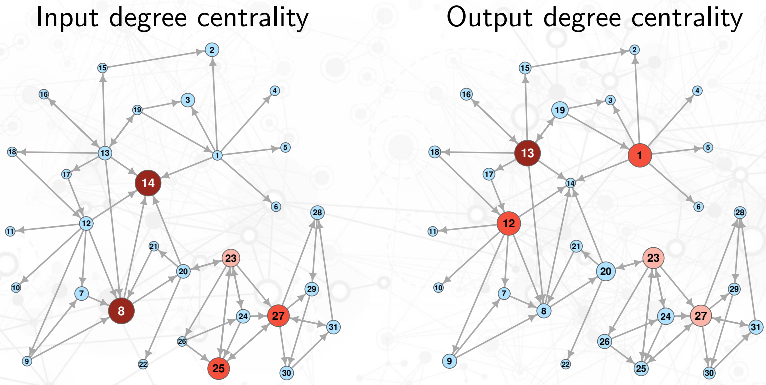


Figure 5: Centralities for a directed network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks

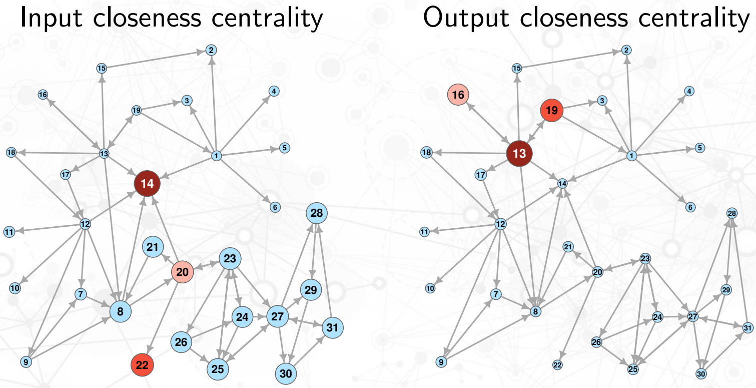
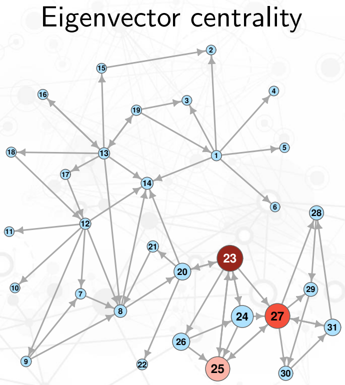


Figure 6: Centralities for a directed network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks



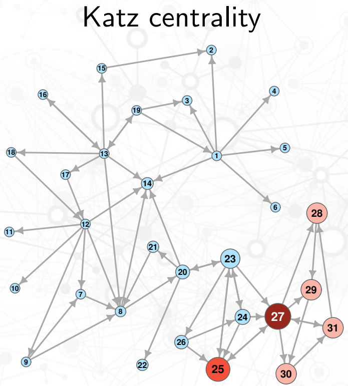


Figure 7: Centralities for a directed network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks

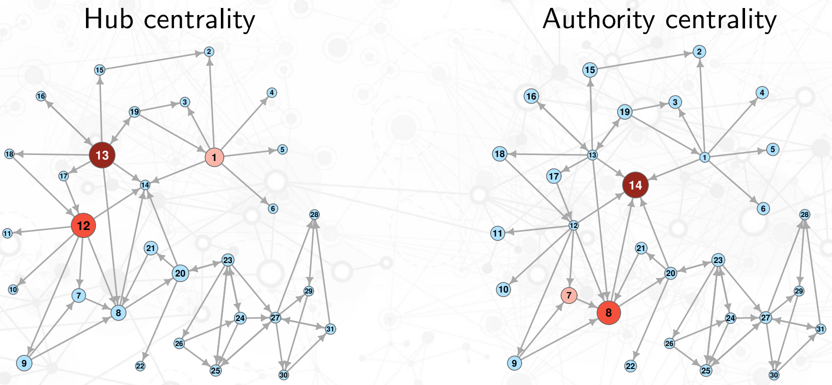


Figure 8: Centralities for a directed network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Examples

Centrality in networks

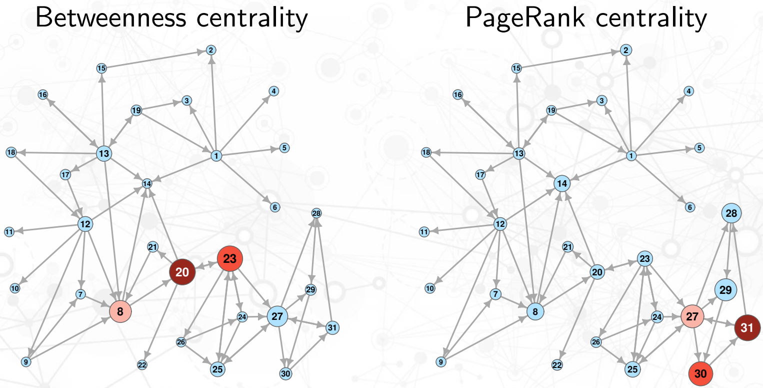


Figure 9: Centralities for a directed network. Nodes with highest centrality in dark red (and white node label), second largest centrality in red, third largest centrality in light red, and rest of nodes in blue. Sizes proportional to centrality with an offset.

Although not directly related to the main topic of the book, we are going to analyze now a real network which is easy to recognize for a large audience, and whose results help in the understanding of the different definitions of centrality in networks.

This is the Network of Thrones [11], a network compiled from the third volume "A Storm of Swords" of the book series "A Song of Ice and Fire", written by the novelist and screenwriter George R. R. Martin, and widely popularized by the HBO TV series "Games of Thrones", created by David Benioff and D. B. Weiss.

Examples

Centrality in networks

The network contains the 107 characters of "A Storm of Swords" connected with 353 weighted edges. Two characters (nodes) are linked when their names are found in the book separated by at most 15 words, meaning they have interacted in some way. The weight counts the number of this kind of interactions.

Since the Network of Thrones is weighted, we have opted here to use the weighted versions of several centrality measures, namely: strength, weighted closeness, weighted betweenness, weighted eigenvector and weighted PageRank. For the weighted closeness and betweenness, we have replaced the original weights w_{ij} by distances defined as $d_{ij} = 1/w_{ij}$, to take into account that the larger the weight, the smaller the distance (or dissimilarity) between the nodes.

In Table 3 we show the 12 most central nodes for the degree centrality and the five weighted centralities mentioned above. Unlike the previous synthetic networks, several characters are always among the most central nodes, with Tyrion Lannister in top of them, followed by Sansa Stark, Jaime Lannister and Robb Stark, and to lower extend Jon Snow, Tywin Lannister and Cersei Lannister.

Figures 11 to 15 show the centralities of the Network of Thrones as proportional to the size of the nodes (and of the font of the names). The colors of the nodes correspond to the seven modules found using two different community detection approaches [20, 27], which produce exactly the same partition: modularity optimization [42] (using a combination of extremal optimization [25], tabu search [3] and fast algorithm [41]) and Infomap [49]. These communities are highly correlated with the different locations where the action takes place.

Table 3: Most central nodes of the Network of Thrones, using weighted centralities.

Rank	Degree	Strength	Closeness	Betweenness	Eigenvector	PageRank
1	Tyrion	Tyrion	Tyrion	Robb	Tyrion	Tyrion
2	Sansa	Jon	Sansa	Tyrion	Sansa	Jon
3	Jon	Sansa	Jaime	Sansa	Jaime	Daenerys
4	Robb	Jaime	Robb	Jon	Joffrey	Jaime
5	Jaime	Bran	Tywin	Jaime	Cersei	Sansa
6	Tywin	Robb	Cersei	Robert	Robb	Robb
7	Cersei	Samwell	Brienne	Daenerys	Tywin	Bran
8	Arya	Arya	Joffrey	Stannis	Bran	Samwell
9	Catelyn	Joffrey	Catelyn	Samwell	Arya	Arya
10	Joffrey	Daenerys	Arya	Tywin	Brienne	Joffrey
11	Robert	Cersei	Margaery	Arya	Catelyn	Cersei
12	Samwell	Tywin	Bran	Bran	Margaery	Tywin

Examples

Centrality in networks

Degree centrality

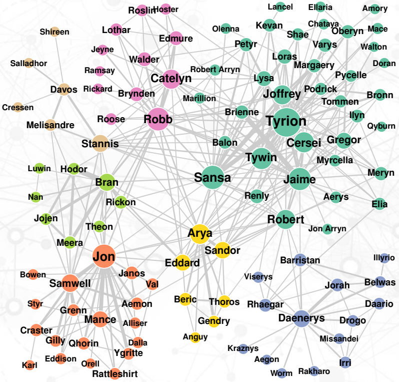


Figure 10: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Strength centrality

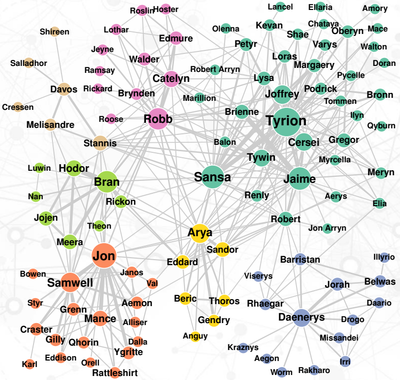


Figure 11: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Examples

Centrality in networks

Weighted closeness centrality

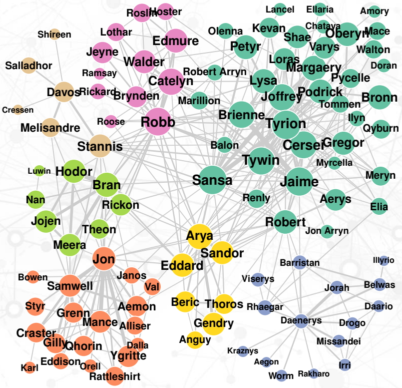


Figure 12: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Examples

Centrality in networks

Weighted betweenness centrality

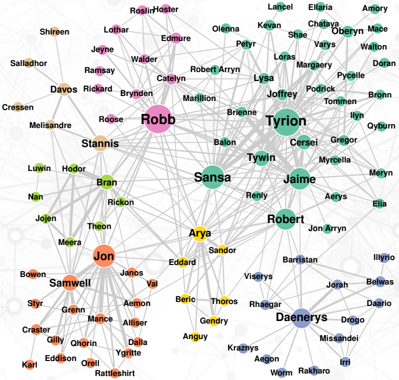


Figure 13: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Examples

Centrality in networks

Weighted eigenvector centrality

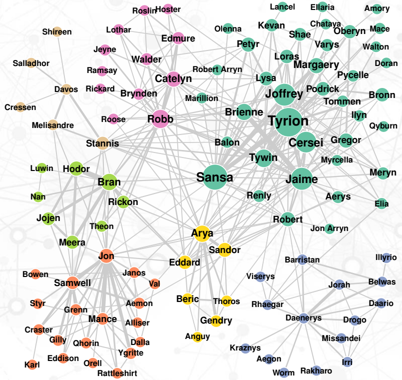


Figure 14: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Examples

Centrality in networks

Weighted PageRank centrality

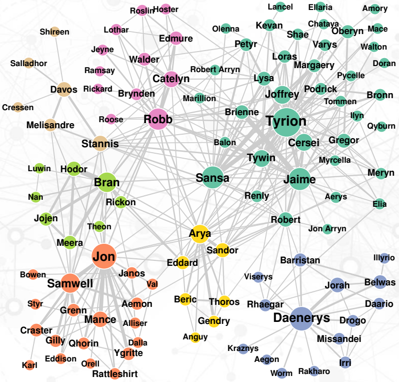


Figure 15: Centralities for the Network of Thrones. Nodes are colored according to the modules found by community detection algorithms. Sizes of nodes proportional to centrality with an offset. Width of links proportional to weights.

Here comes a list of software tools which can be used to calculate centralities in complex networks:

- Pajek[1]: Analysis and visualization tool for Windows (can be run under Linux and MacOS using Wine) [7]. Allows the calculation of several centralities: degree, strength, closeness, betweenness, hubs and authorities (HITS), and a few additional ones not described above.
- Gephi[2]: Visualization and exploration software [6]. Calculates degree, strength, eigenvector, HITS and PageRank centralities.

[1] Pajek: http://mrvar.fdv.uni-lj.si/pajek
[2] Gephi: https://gephi.org

- Radatools[3]: Set of programs for the analysis of complex networks, with main attention to community detection and the finding of structural properties [32]. Calculates degree, strength, betweenness (weighted and unweighted, directed and undirected, for nodes and edges) and other centralities.
- Cytoscape[4]: Originally designed for biological research, now it is a general platform for complex network analysis and visualization [52]. It does not directly calculate centralities, but there are plug-ins which can be used to find some of them.
- igraph[5]: Collection of network analysis tools with the emphasis on efficiency, portability and ease of use [19]. Calculates degree, strength, betweenness, closeness, eigenvector, HITS and PageRank centralities.

[3] Radatools: http://deim.urv.cat/~sergio.gomez/radatools.php
[4] Cytoscape: http://www.cytoscape.org
[5] igraph: http://igraph.org

- NetworkX[6]: Python software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks [51]. Calculates degree, strength, closeness, betweenness, eigenvector, HITS, Katz and PageRank centralities, and a few additional ones.
- SNAP[7]: General purpose, high performance system for analysis and manipulation of large networks [38]. Calculates degree, strength, closeness, betweenness, eigenvector and HITS centralities.
- Visone[8]: Tool for the analysis and visualization of social networks [17]. Calculates degree, strength, closeness, betweenness, eigenvector, HITS and PageRank centralities, and a few additional ones.

[6] NetworkX: http://networkx.github.io
[7] SNAP: http://snap.stanford.edu/snap
[8] Visone: https://www.visone.info

- MuxViz[9]: Framework for the multilayer analysis and visualization of networks [21]. Calculates the generalizations of centralities to multilayer networks (versatilities), including degree, eigenvector, Katz, HITS and PageRank centralities.
- graph-tool[10]: Efficient Python module for manipulation and statistical analysis of graphs [46]. Calculates PageRank, betweenness, closeness, eigenvector, Katz, HITS and other centralities.

The integration of some tools with Python (igraph, NetworkX, graph-tool) and R (igraph, MuxViz) allows a high-level implementation of the missing centralities without too much effort.

[9] MuxViz: http://muxviz.net

[10] graph-tool: https://graph-tool.skewed.de

Conclusion

We have seen how it is possible to find the most important items in a data set, provided we transform this data into a complex network. The definition of “most important” is not unique, there exist several complementary ways, each one concentrated in one structural characteristic of the network. Degree centrality allows to find the most connected nodes. Closeness centrality finds the nodes which are in the “middle” of the network, i.e. at a shortest average distance to the rest of the nodes.

Conclusion

Betweenness centrality is specialized in the nodes which are "bridges" between separated parts of the network. Eigenvector centrality looks for nodes whose importance is given by the sum of the centralities of the nodes which send links to it, thus becoming a recursive definition which is expressed as an eigenvector and eigenvalue problem. Katz centrality represents a balance between closeness and eigenvector centralities. Finally, the dynamics of random walkers in the network is the basis for several centralities, standing out PageRank, the well-known measure originally used to rank web pages by the Google search engine.

Conclusion

We have also considered how centralities must be adapted for the different kinds of network, e.g. by taking into account the directionality of the links, their weights, or the multilayer structure. In summary, centrality integrates a large set of definitions and tools to analyze the relevance of the nodes in networks, being able to identify the most important ones, which may constitute the first step in many marketing and business applications, where targeted actions increase their success rate and reduce the overall cost.

References I

[1] S.E. Ahnert et al. “Ensemble approach to the analysis of weighted networks”. In: *Physical Review E* 76.1 (2007), p. 016101.

[2] Jac M Anthonisse. “The rush in a directed graph”. In: *Stichting Mathematisch Centrum. Mathematische Besliskunde* BN 9/71 (1971), pp. 1–10.

[3] Alex Arenas, Alberto Fernandez, and Sergio Gomez. “Analysis of the structure of complex networks at different resolution levels”. In: *New Journal of Physics* 10.5 (2008), p. 053039.

[4] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *Science* 286.5439 (1999), pp. 509–512.

[5] Alain Barrat et al. “The architecture of complex weighted networks”. In: *Proceedings of the National Academy of Sciences USA* 101.11 (2004), pp. 3747–3752.

[6] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. “Gephi: an open source software for exploring and manipulating networks.” In: *ICWSM* 8 (2009), pp. 361–362.

[7] Vladimir Batagelj and Andrej Mrvar. “Pajek – Program for large network analysis”. In: *Connections* 21.2 (1998), pp. 47–57.

[8] Federico Battiston, Vincenzo Nicosia, and Vito Latora. “Structural measures for multiplex networks”. In: *Physical Review E* 89.3 (2014), p. 032804.

References II

[9] Alex Bavelas. "Communication patterns in task-oriented groups". In: *The Journal of the Acoustical Society of America* 22.6 (1950), pp. 725–730.

[10] Murray A Beauchamp. "An improved index of centrality". In: *Behavioral Science* 10.2 (1965), pp. 161–163.

[11] Andrew Beveridge and Jie Shan. "Network of thrones". In: *Math Horizons* 23.4 (2016), pp. 18–22.

[12] Stefano Boccaletti et al. "The structure and dynamics of multilayer networks". In: *Physics Reports* 544.1 (2014), pp. 1–122.

[13] Phillip Bonacich. "Factoring and weighting approaches to status scores and clique identification". In: *Journal of Mathematical Sociology* 2.1 (1972), pp. 113–120.

[14] Phillip Bonacich. "Power and centrality: A family of measures". In: *American Journal of Sociology* 92.5 (1987), pp. 1170–1182.

[15] Phillip Bonacich. "Technique for analyzing overlapping memberships". In: *Sociological Methodology* 4 (1972), pp. 176–185.

[16] Ulrik Brandes. "A faster algorithm for betweenness centrality". In: *Journal of Mathematical Sociology* 25.2 (2001), pp. 163–177.

[17] Ulrik Brandes and Dorothea Wagner. "Analysis and visualization of social networks". In: *Graph drawing software* (2004), pp. 321–340.

References III

[18] Sergey Brin and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine". In: *Computer Networks and ISDN Systems* 30.1 (1998), pp. 107–117.

[19] Gabor Csardi and Tamas Nepusz. "The igraph software package for complex network research". In: *InterJournal, Complex Systems* 1695.5 (2006), pp. 1–9.

[20] Leon Danon et al. "Comparing community structure identification". In: *Journal of Statistical Mechanics: Theory and Experiment* 2005.09 (2005), P09008.

[21] Manlio De Domenico, Mason A. Porter, and Alex Arenas. "MuxViz: a tool for multilayer analysis and visualization of networks". In: *Journal of Complex Networks* 3.2 (2015), p. 159. DOI: 10.1093/comnet/cnu038.

[22] Manlio De Domenico et al. "Mathematical formulation of multilayer networks". In: *Physical Review X* 3.4 (2013), p. 041022.

[23] Manlio De Domenico et al. "Ranking in interconnected multilayer networks reveals versatile nodes". In: *Nature Communications* 6 (2015).

[24] Anthony Dekker. "Conceptual distance in social network analysis". In: *Journal of Social Structure (JOSS)* 6 (2005).

[25] Jordi Duch and Alex Arenas. "Community detection in complex networks using extremal optimization". In: *Physical review E* 72.2 (2005), p. 027104.

References IV

[26] Robert W Floyd. “Algorithm 97: shortest path”. In: *Communications of the ACM* 5.6 (1962), p. 345.

[27] Santo Fortunato. “Community detection in graphs”. In: *Physics reports* 486.3 (2010), pp. 75–174.

[28] Linton C Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977), pp. 35–41.

[29] Linton C Freeman. “Centrality in social networks conceptual clarification”. In: *Social Networks* 1.3 (1979), pp. 215–239.

[30] Georg Frobenius. “Über Matrizen aus nicht negativen Elementen”. In: *Sitzungsber. Königl. Preuss. Akad. Wiss.* (1912), pp. 456–477.

[31] W. L. Garrison. “Connectivity of the interstate highway system”. In: *Papers and Proceedings of the Regional Science Association* 6 (1960), pp. 121–137.

[32] Sergio Gómez and Alberto Fernández. *Radatools software, Communities detection in complex networks and other tools*. 2011.

[33] Roger Guimerà et al. “Optimal network topologies for local search with congestion”. In: *Physical Review Letters* 89.24 (2002), p. 248701.

[34] Arda Halu et al. “Multiplex PageRank”. In: *PLOS ONE* 8.10 (2013), e78293.

[35] Leo Katz. “A new status index derived from sociometric analysis”. In: *Psychometrika* 18.1 (1953), pp. 39–43.

References V

[36] Mikko Kivelä et al. “Multilayer networks”. In: *Journal of Complex Networks* 2.3 (2014), pp. 203–271.

[37] Jon M Kleinberg. “Authoritative sources in a hyperlinked environment”. In: *Journal of the ACM (JACM)* 46.5 (1999), pp. 604–632.

[38] Jure Leskovec and R Sosič. *SNAP: Stanford network analysis platform*. 2013.

[39] L. Lovász. “Random walks on graphs: A survey”. In: *Combinatorics, Paul Erdos is Eighty* 2.1 (1993), pp. 1–46.

[40] Mark EJ Newman. “A measure of betweenness centrality based on random walks”. In: *Social Networks* 27.1 (2005), pp. 39–54.

[41] Mark EJ Newman. “Fast algorithm for detecting community structure in networks”. In: *Physical review E* 69.6 (2004), p. 066133.

[42] Mark EJ Newman and Michelle Girvan. “Finding and evaluating community structure in networks”. In: *Physical review E* 69.2 (2004), p. 026113.

[43] Mark Newman. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010. ISBN: 0199206651, 9780199206650.

[44] J. Nieminen. “On the centrality in a directed graph”. In: *Social Science Research* 2 (1973), pp. 371–378.

[45] Jae Dong Noh and Heiko Rieger. “Random walks on complex networks”. In: *Physical Review Letters.* 92.11 (2004), p. 118701.

References VI

[46] Tiago P Peixoto. “The graph-tool python library”. In: *figshare* (2014).

[47] Oskar Perron. “Zur Theorie der Matrices”. In: *Mathematische Annalen* 64.2 (1907), pp. 248–263. ISSN: 1432-1807.

[48] F. R. Pitts. “A graph theoretic approach to historical geography”. In: *The Professional Geographer* 17 (1965), pp. 15–20.

[49] Martin Rosvall and Carl T Bergstrom. “Maps of random walks on complex networks reveal community structure”. In: *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1118–1123.

[50] G Sabidussi. “The centrality index of a graph”. In: *Psychometrika* 31 (1966), pp. 581–603.

[51] Daniel A Schult and P Swart. “Exploring network structure, dynamics, and function using NetworkX”. In: *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*. Vol. 2008. 2008, pp. 11–16.

[52] Paul Shannon et al. “Cytoscape: a software environment for integrated models of biomolecular interaction networks”. In: *Genome research* 13.11 (2003), pp. 2498–2504.

[53] M. E. Shaw. “Group structure and the behavior of individuals in small groups”. In: *Journal of Psychology* 38 (1954), pp. 139–149.

[54] Luis Solá et al. “Eigenvector centrality of nodes in multiplex networks”. In: *Chaos* 3 (2013), p. 033131.

References VII

[55] Albert Solé-Ribalta, Sergio Gómez, and Alex Arenas. "A model to identify urban traffic congestion hotspots in complex networks". In: *Royal Society Open Science* 3.10 (2016), p. 160098.

[56] Albert Solé-Ribalta et al. "Random walk centrality in interconnected multilayer networks". In: *Physica D: Nonlinear Phenomena* 323 (2016), pp. 73–79.

[57] Shi-Jie Yang. "Exploring complex networks by walking on them". In: *Physical Review E* 71.1 (2005), p. 016107.