Measures of Centrality in Graphs

François Théberge
theberge@ieee.org

CSE/TIMC

August 2019

Degree

Let $G = (V, E)$ with adjacency matrix A

Entries in A: $a_{ij} > 0 \iff (i, j) \in E$

For an unweighted graph, $a_{ij} \in \{0, 1\}$

For a directed graph, we may have $a_{ij} \neq a_{ji}$.

Definition: The *degree centrality* of node $i \in V$ in an undirected graph is:

$$C_D(i) = d_i = \sum_{j \in V} a_{ij}$$

Degree

Definition: The *in-degree centrality* of node $i \in V$ in a directed graph is:

$$C_D^{in}(i) = d_i^{in} = \sum_{j \in V} a_{ji}$$

Definition: The *out-degree centrality* of node $i \in V$ in a directed graph is:

$$C_D^{out}(i) = d_i^{out} = \sum_{j \in V} a_{ij}$$

The *normalized degree centrality* for unweighted graphs is obtained by dividing the above quantities by $n - 1 = |V| - 1$ assuming no self-loops.

Centrality

We can define centrality as a combination of high degree, and being connected to highly central nodes:

$$c(i) = \sum_{j \in V} a_{ij} c(j)$$

If $\det(A - I) \neq 0$, the unique solution is $c = 0$.

We define eigenvector centrality as being proportional to the neighbour's centrality

Centrality

Definition: The *eigenvector centrality* of node i is a solution of:

$$\lambda c^E(i) = \sum_{j \in V} a_{ij} c^E(j)$$

which we can write in matrix terms $Ac^E = \lambda c^E$.

For directed graphs, we define $A^t c^{E_{in}} = \lambda c^{E_{in}}$ and $Ac^{E_{out}} = \lambda c^{E_{out}}$.

For a connected graph, we measure centrality with respect to the eigenvector corresponding to the largest eigenvalue, which is real and positive (Perron-Frobenius), that is:

$$c^E(i) = u_1(i)$$

where u_1 is the leading eigenvector.

Centrality

Another approach is to consider *shortest paths* between nodes.

A *geodesic* is a shortest path between 2 nodes (smallest number of hops, or smallest sum of edge weights).

Let d_{ij}, the length of a geodesic between nodes i and j, with $d_{ii} = 0$.

For disconnected graphs, we may have $d_{ij} = \infty$; for unweighted graphs, we set $d_{ij} = n$ in practice.

Definition: The *closeness centrality* of node i is defined as:

$$c^C(i) = \left(\sum_{j \neq i} d_{ij} \right)^{-1}$$

Centrality

A more commonly used approach is to consider all geodesics;

let:

n_{jk}: number of geodesics between nodes j and k

$n_{jk}(i)$: number of geodesics between nodes j and k passing through node i.

Definition: The *betweenness centrality* of node i is:

$$c^B(i) = \sum_{j \neq i} \sum_{k \neq i,j} \frac{n_{jk}(i)}{n_{jk}}$$

Centrality

Yet another approach is to consider the impact of removing some node i from G

Let $P(G)$ be some measure of performance on graph G

Let G_i be the graph G obtained by removing all edges adjacent to node i

Definition: The *delta-centrality* of node i is:

$$c^{\Delta}(i) = \frac{P(G) - P(G_i)}{P(G)}$$

Note that we require that all $P(G) - P(G_i) \geq 0$.

Centrality

One possible choice for $P(G)$ is:

Definition: The *efficiency* of graph G is

$$E(G) = \frac{1}{n(n-1)} \sum_{i} \sum_{j \neq i} \frac{1}{d_{ij}}$$

For unconnected graphs, pairs of nodes with $d_{ij} = \infty$ have zero contribution.

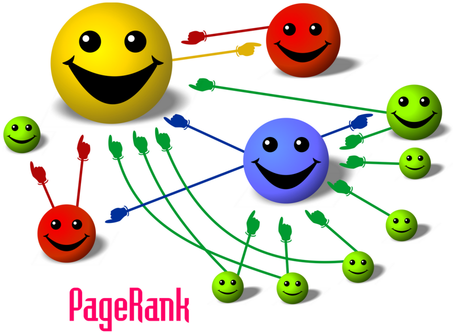
All measures of centrality can be generalized to *group centrality* measures by considering sets of nodes.

For example, degree of a group of nodes g is the number of vertices not in g joined to at least one node in g.

Centrality

Another measure of centrality in the PageRank algorithm used by Google

Idea: important pages are linked to by many important pages.



REF: Wikipedia

Centrality

Imagine a random walk where, given we are at node i:

- with probability $1-d$: jump to a random node $j \in V$.
- with probability d: jump to node j with probability a_{ij}/d_i^{out}.

PageRank values are obtained by solving:

$$R(i) = \frac{1-d}{n} + d \sum_{j \sim i} \frac{R(j)}{d_j^{out}}$$

or in matrix notation:

$$R = (\frac{1-d}{n}\mathbf{1} + dM)R$$

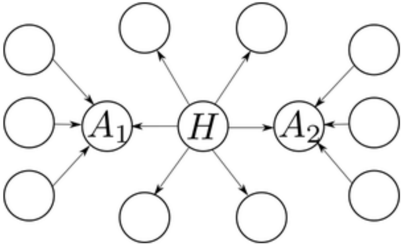
where M is the normalized adjacency matrix.

Solution is obtained by computing the leading eigenvector.

Centrality

In a directed graph, we define a *hub* as a node with high out degree, and an *authority* as a node with high in degree.

In an undirected graph, the two concepts are the same.



REF: blog.scrapinghub.com

Centrality

We define hub centrality for node i as:

$$c^H(i) = \nu \sum_j a_{ij} c^A(j)$$

and authority centrality as:

$$c^A(i) = \mu \sum_j a_{ji} c^H(j)$$

Let $\lambda = \frac{1}{\nu\mu}$, we get:

$$A^t A c^A = \lambda c^A$$

$$A A^t c^H = \lambda c^H$$

Edge Centrality

Betweenness centrality can be re-defined for the edges $e \in E$;

let

n_{jk}: number of geodesics between nodes j and k

$n_{jk}(e)$: number of geodesics between nodes j and k passing through edge e.

Definition: The *betweenness centrality* of edge e is:

$$c^B(e) = \sum_j \sum_{k \neq j} \frac{n_{jk}(e)}{n_{jk}}$$

Correlation

Let p_i, the proportion of nodes of degree i in G, the *degree distribution*

The degree correlation measures the relationship between degree of nodes linked by edges

In an *assortative* network, high degree nodes tend to be more connected to other high degree nodes, and the same for low degree nodes.

In a *disassortative* network, high degree nodes tend to be more connected to low degree nodes, and vice-versa.

Correlation

Definition: The *degree correlation function* for nodes of degree k is given by:

$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$

with $P(k'|k)$ the probability that from an edge incident to a degree k node leads to a degree k' node.

If there is no correlation (neutral network), we get

$$k_{nn}(k) = \frac{\sum_i i^2 \cdot p_i}{\sum_i i \cdot p_i}$$

where p_i: proportion of nodes of degree i.

Correlation

The *friendship paradox* states that the average degree of a node's neighbours is typically higher than its own degree.

This is due to the fact that in many networks:

$$\frac{\sum_i i^2 \cdot p_i}{\sum_i i \cdot p_i} >> \sum_i i \cdot p_i$$

Simply stated, it is more likely to link with hubs (high degree nodes).

Correlation

The degree correlation function can be approximated by

$$k_{nn}(k) = ak^{\mu}$$

where:

- $\mu > 0$ for assortative networks
- $\mu \approx 0$ for neutral networks, and
- $\mu < 0$ for disassortative networks

Centrality

Notebook #1