

Centrality

Which nodes are most important or central in network?

Emergence from dynamics and flows.

Eigenvector Centrality

Using degree as notion of centrality, award one point for every neighbor a node has.

But not all nodes are equivalent, since connecting to important nodes makes one more important.

Use a dynamics framework to give score proportional to sum of neighbors.

1. Make initial guess about centrality $x_i$ of each node $i$, e.g. all $x_i = 1$.
2. Update this to $x' = x_i$ as

$$x' = \sum_j A_{ij} x_j,$$

where $A_{ij}$ is element of adjacency matrix.

So in matrix vector notation:

$$x' = Ax.$$

After $t$ steps, vector is

$$x(t) = A^t x(0).$$

Write $x(0)$ as linear combination of eigenvectors $v_i$ of $A$:

$$x(0) = \sum_i c_i v_i,$$

for some constants $c_i$. Then:

$$x(t) = A^t \sum_i c_i v_i = \sum_i c_i \lambda_i^t v_i = \lambda_i^t \sum_i c_i \left[ \frac{\lambda_i}{\lambda_2} \right] v_i,$$

where $\lambda_i$ are the eigenvalues of $A$, and $\lambda_2$ is the largest in abs.

Since $\lambda_i / \lambda_2 < 1$ for all $i \neq 1$, all terms in sum other than first decay exponentially as $t$ becomes large, and so in the limit $t \to \infty$,

$$x(t) \to c_1 \lambda_2^t v_1.$$
So leading vector of centrality is just proportional to first eigenvector of \( A \):

\[
Ax = \lambda x.
\]

So \( x_i \) is proportional to sum of centrality of \( i \)'s neighbors:

\[
x_i = k_i \sum_j A_{ij} x_j.
\]

and can be large either because many neighbors or important neighbors.

This is good for undirected networks, but what about directed networks:

→ generally not symmetric, so two sets of eigenvectors, left, right.

→ use right eigenvector since centrality arises from others pointing to you.

One difficulty that remains with directed networks is that

only vertices in a strongly connected component of two or more vertices,

or the out-component of such a component can have non-zero eigenvector centrality.

(unfortunately DAGs have no strongly connected components of more than one vertex)

How to deal with this issue?

→ give each vertex a little centrality for free

Katz Centrality:

\[
x_i = \alpha \sum_j A_{ij} x_j + \beta,
\]

where \( \alpha, \beta \) are positive constants

with the second term, even vertices with zero in-degree get centrality \( \beta \).

In matrix-vector form:

\[
x = \alpha Ax + \beta 1.
\]

\[
x = (I - \alpha A)^{-1} \beta 1
\]

→ in matrix-vector form:

\[
x = (I - \alpha A)^{-1} \beta 1.
\]
There are distinct approaches for choosing the parameters.

One can, of course, also use Katz centrality for undirected networks.

In many cases it means less if a vertex is only one among many that we point to.

Modify Katz centrality to get PageRank by making centrality proportional to neighbors' centrality but divided by their out-degree.

$$x'_i = \alpha \sum_j A_{ij} \frac{x_j}{k_{out}^j} + \beta$$

A small mathematical difficulty in dividing by 0 when $k_{out}^j = 0$.

Just artificially set $k_{out}^j = 1$ for all such vertices, so vertices with no out-going edges contribute zero to centrality of other nodes.

$$x = \alpha A D^{-1} x + \beta I$$

where $D$ is diagonal matrix with elements $D_{ii} = \max(k_{out}^i, 1)$.

$$x = \beta (I - \alpha A D^{-1})^{-1} I$$, and setting $\beta = 1:

$$x = (I - \alpha A D^{-1})^{-1} I = D(D - \alpha A)^{-1} I.$$

Now use distance/flow rather than dynamics to define centrality.

**Closeness Centrality**

Mean distance from a node to other nodes.

If $d_{ij}$ is length of geodesic path from $i$ to $j$ then,

$$d_i = \frac{1}{n} \sum_{j} d_{ij}$$

This gives small values for more central vertices and large values for less central ones.

So consider

$$C_i = \frac{1}{d_i} = \frac{n}{\sum d_{ij}}.$$
An alternative is to use the harmonic mean:

\[ C_i = \frac{1}{n-1} \sum_{j} \frac{1}{d_{ij}} \]

which actually deals with infinite distances among different concepts.

**Betweenness Centrality (flow):**

measure the extent to which a vertex lies on paths between other vertices.

\[ \Rightarrow \text{ amount of flow passing through each vertex is proportional} \]

\[ \text{to number of geodesic paths the vertex lies on.} \]

This is betweenness centrality.

First consider an undirected network in which at most one geodesic path between any two nodes.

\[ \Rightarrow \text{consider all geodesic paths in such a network} \]

\[ \Rightarrow \text{betweenness centrality is the fraction of those paths that pass through } i. \]

Let \( n_{ij}^e \) be 1 if vertex \( i \) lies on geodesic path from \( st \).

\[ 0 \text{ else} \]

Then \( X_i = \sum_{st} n_{ij}^e. \)

What if multiple geodesic paths; give weight inversely proportional to this number.

Letting \( N_{ij} \) be total \# geodesic paths from \( st \).

\[ X_i = \frac{\sum_{st} n_{ij}^e}{N_{ij}}. \]

Alternatively is to consider all paths for flow rather than just shortest.

**Flow betweenness.**

Since maximum flow between \( s \) and \( t \) is also number of edge-independent paths between them, so flow betweenness is \# of independent paths that go through \( i \).