Centrality, Part I

CptS 591: Elements of Network Science

Assefaw Gebremedhin

School of Electrical Engineering and Computer Science
Washington State University

February 23, 2021

Network Analysis

We could think of the subject of network analysis at different levels of granularity:

- Network-level
    - statistical properties of networks
    - models describing formation and evolution of networks
- Group-level
    - group identification
    - strong or similar linkage, etc
- Element-level
    - centrality: How *important* is a given vertex (or an edge) in a network?

Centrality metrics

- There are different metrics for measuring importance — many kinds of centrality — each valid for its own particular context.
- The same network may be meaningfully analyzed with different centrality indices depending on the question to be answered.
- This lecture (and the next) substantiate this by looking at examples of several different centrality metrics (or indices).

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

We can think of three different graph abstractions for this scenario, resulting in the need for different kinds of centrality measures.

1. Vertices are students, and an edge from A to B means A has voted for B. We may call this a *voting network*.

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

We can think of three different graph abstractions for this scenario, resulting in the need for different kinds of centrality measures.

1. Vertices are students, and an edge from A to B means A has voted for B. We may call this a *voting network*. Here a student is the more central the more vote she gets. (Motivates "degree centrality").

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

2. Vertices are students, and an edge from A to B means A has convinced B to vote for his favorite candidate. We may call this an *influence network*.

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

2. Vertices are students, and an edge from A to B means A has convinced B to vote for his favorite candidate. We may call this an *influence network*. Here a student is the more central the more it is needed to "mediate" the opinion of others (i.e. being "between groups" is important). (Motivates "betweenness centrality").

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

3. Vertices are students, edges represent social friendship. We may call this a *social network*.

Motivating Example

Consider a class of students voting to elect a representative, where every student is allowed to vote for one student.

3. Vertices are students, edges represent social friendship. We may call this a *social network*. Here someone who is a friend of an important person could be regarded as more important than someone having friends with low social prestige. Centrality is determined by centrality of adjacent vertices. (Motivates “feedback centrality”).

Elementary common denominator formalization

- The intuition behind a centrality measure is that it denotes an order of importance on the vertices (or edges) of a network by assigning real values to them.
- As stated earlier, there are many different kinds of centralities, but as a minimalistic requirement we could demand that the result of a centrality index depends on only the *structure* of the network.
- The following definition formalizes this demand.

Structural Index

Definition

Let $G = (V, E)$ be a graph. A real-valued function s is called a structural index if and only if: $\forall v \in V : G$ is isomorphic to $H \Longrightarrow s_G(v) = s_H(\phi(v))$. ($\phi$ is a one-to-one mapping such that $(u, v) \in E(G) \Longleftrightarrow (\phi(u), \phi(v)) \in E(H)$).

A centrality index c is required to be a structural index (SI) and thus induces a semi-order. We say x is more central than y if $c(x) > c(y)$.

Centrality around distances and neighborhoods

Definition

Degree centrality The simplest centrality measure is degree centrality $c_D(v)$ of a vertex v, simply defined as the degree $d(v)$ of v, when the graph is undirected. In the directed case, one distinguishes between in-degree centrality $c_{iD}(v) = d_{in}(v)$ and out degree centrality $c_{oD}(v) = d_{out}(v)$.

Clearly, degree $d(v)$ is a SI; it is independent of labeling, so it is invariant under isomorphism.

Eccentricity

- **Motivation**: consider the problem of determining the best location for an emergency facility such as a hospital.
- A common objective of such an emergency *facility location problem* is to find a site that minimizes the maximum response time between the facility and the site of a possible emergency (minimax criterion).

Definition

Suppose the hospital is located at a vertex $u \in V$. The *eccentricity* $e(u)$ of u is the maximum distance from u to any vertex v in the network, i.e., $e(u) = \max\{d(u, v) : v \in V\}$, where $d(u, v)$ represents the distance from u to v. The minimum over all $e(u)$, $u \in V$ solves the emergency FLP.

The quantity $c_E(u) = 1/e(u)$, which clearly is a SI, is called eccentricity centrality.

Closeness/transmission

- **Motivation**: consider the problem of determining the best location for a service facility such as a shopping mall.
- A common objective here is to find a site for the facility that minimizes total travel time to the facility from everyone in the network (minisum criterion).

Definition

Suppose the mall is located at a vertex $u \in V$. Let $t(u) = \sum_{v \in V} d(u, v)$, where $d(u, v)$ denotes the distance from u to v. The minimum over all $t(u)$, $u \in V$ solves the service FLP.

The quantity $c_C(u) = 1/(t(u)$ is clearly a SI. In social network analysis, a centrality index based on this is called closeness. In communication networks, a similar concept is known by the name transmission number.

Betweeness Centrality

- Let $\delta_{st}(v)$ denote the fraction of shortest paths between s and t that contain the vertex v:

$$\delta_{st}(v) = \sigma_{st}(v)/\sigma_{st}$$

  where σ_{st} denotes the number of all shortest-paths between s and t.
- Then the betweenness centrality $c_B(v)$ of a vertex v is

$$c_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v)$$

- **Motivation:** Measures the control over communication between others.

Katz Index

- An example of feedback centrality – a node is more central the more central its neighbors are.
- Consider an adjacency matrix A representing voting, $A_{ji} = 1 \Longrightarrow j$ voted for i, and the associated directed network.
- Then $A^T e_i =$ number of votes for i.
- Suppose only two individuals k and k' voted for i, but everyone else in the network voted either for k or k'.
- Then it may be that i is the most important person in the network — even if she got only two votes. All others voted for her indirectly.

Katz Index

- The idea behind Katz index is to count additionally all indirect votes where the number of intermediate individuals may be arbitrarily large.
- To account for intermediate values properly, we need to introduce a damping factor α, where $0 < \alpha < 1$, so that the longer the path between vertices j and i, the smaller its impact on the status of i becomes.

Katz Index

- Recalling that $(A^l)_{ji}$ holds the number of paths from j to i with length l, the Katz status of vertex i is

$$c_K(i) = \sum_{l=1}^{\infty} \sum_{j=1}^{n} \alpha^l (A^l)_{ji}$$

if the infinite sum converges.

- In matrix notation,

$$c_K = \sum_{l=1}^{\infty} \alpha^l (A^T)^l e$$

where e is the n-dimensional vector where every entry is 1.

- When does this sum converge?

Katz Index

- Suppose A is a 1-by-1 matrix (scalar). Then we have $c_{K_1} = 1 + \alpha + \alpha^2 + \ldots = 1/1 - \alpha$, if $|\alpha| < 1$. That is, c_{K_1} is a geometric series.
- If we generalize the geometric series to matrices, we get the Neumann series (named after Carl Gottfried Neumann).
- **Neumann series:** $\sum_{l=0}^{\infty} \alpha^l A^l \rightarrow \left(I - \alpha A^T\right)^{-1}$ if $\rho(A) < 1$, where $\rho(A)$ is the spectral radius of A (recall $\rho(A) = \max_i |\lambda_i|$).

Katz Index

- Then we can write Katz index as

$$
\begin{aligned}
c_K = ((I - \alpha A^T)^{-1} e), \text{ if } \rho(\alpha A^T) < 1 \\
(I - \alpha A^T) c_K = e
\end{aligned}
$$

- We can solve this linear system to get the Katz index c_K. This is an inhomogeneous system of linear equations emphasizing the feedback nature of the centrality: the value of $c_K(i)$ depends on the other centrality values $c_K(j), j \neq i$.