1.022 Introduction to Network Models

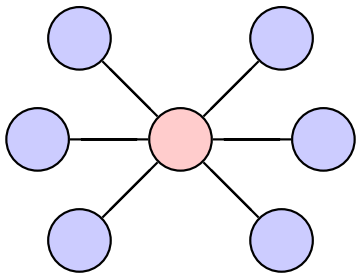
Amir Ajorlou

Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

Lecture 4

The notion of centrality

- Identify the most important nodes in a graph given its topology
  - $\Rightarrow$ Not based on the nature of the particular node
- Different definitions of importance give rise to different centrality measures
  - $\Rightarrow$ Degree, closeness, eigenvector, betweenness, Katz
  - $\Rightarrow$ They induce a centrality ranking on the nodes
- Centrality measures are widely used
  - $\Rightarrow$ Targeted marketing
  - $\Rightarrow$ Network vulnerability to attacks
  - $\Rightarrow$ Epidemiology control
  - $\Rightarrow$ Power in exchange networks



Degree centrality

- Local measure of the importance of a node within a graph
- Sum of the weights of incident edges

$$C_D(i) := \sum_{j|(i,j)\in E} A_{ij}.$$

- High degree centrality value of a given node
  - $\Rightarrow$ The node has a large number of neighbors
  - $\Rightarrow$ Closely related to its neighbors (in weighted similarity graphs)
- For directed networks, both in-degree and out-degree centralities
- Does not capture cascade effects
  - $\Rightarrow$ I am more important if my neighbors are important

Closeness centrality

- How fast information can spread from one node to every other node
- Inverse of the sum of the shortest path lengths s_G to other nodes

$$C_C(i) := \left(\sum_{j \in V} s_G(i,j) \right)^{-1} .$$

- Nodes which are close to others on average have high centrality
  $\Rightarrow$ Such nodes may have more direct influence on others
- Example: Network of movie actors: two actors are connected if they work together
  $\Rightarrow$ Highest centrality 0.4143 for Christopher Lee;Entered into the Guinness Book of World Records in 2007 for most screen credits
  $\Rightarrow$ Lowest centrality 0.1154 for Leia Zanganeh, an Iranian film and theater actress
- Two limitations associated with closeness centrality
  $\Rightarrow$ Spans a very small dynamic range
  $\Rightarrow$ In disconnected graphs, centrality is zero for all nodes

Betweenness centrality

- Control on the optimal flow within a graph

  $\Rightarrow$ Nodes that fall within the shortest path of many pairs of nodes

- The number of shortest paths from j to k is σ_{jk}
- The number of shortest paths from j to k going through i is $\sigma_{jk}(i)$

$$C_B(i) := \sum_{\substack{j,k \in V \\ j \neq i \neq k}} \frac{\sigma_{jk}(i)}{\sigma_{jk}}. \tag{1}$$

- Look at the shortest paths for every two nodes distinct from i

  $\Rightarrow$ Sum the proportion that contain node i

- Betweenness is not a measure of connectivity.

  $\Rightarrow$ For the network of movie actors, highest centrality node is Spanish actor Fernando Rey, with centrality of $7.47x10^8$

  $\Rightarrow$ He starred in many French, American and Spanish movies.

Interpretation in a social network

- What would degree, closeness, and betweenness centrality reveal?
- Degree $\Rightarrow$ Most friends $\Rightarrow$ Most popular person
- Closeness $\Rightarrow$ Can quickly reach the whole group (directly or indirectly)
  $\Rightarrow$ Relevant if we want to quickly spread information in the network
- Betweenness $\Rightarrow$ Power in the transmission of information
  $\Rightarrow$ Relevant if we want to influence communication between groups
- All of them are right, they just reveal different features

Eigenvector centrality

- Linear combination of the centrality of the neighbors

  $\Rightarrow$ Few important neighbors can weigh more than many unimportant ones

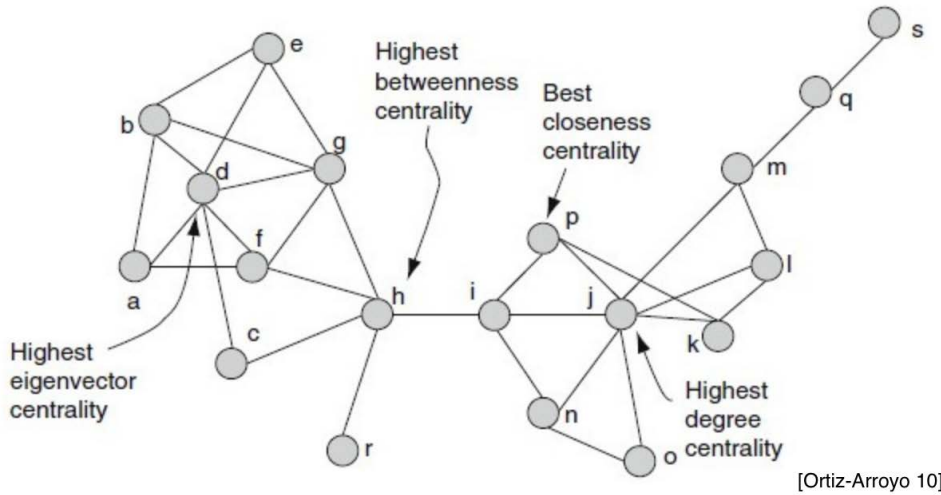
$$C_E(i) := \frac{1}{\lambda} \sum_{(i,j) \in E} A_{ij} C_E(j),$$

- This definition leads to an eigenvector problem
- C_E is the dominant eigenvector of the adjacency matrix

  $\Rightarrow$ Perron-Frobenius guarantees all positive elements

Comparing centrality measures

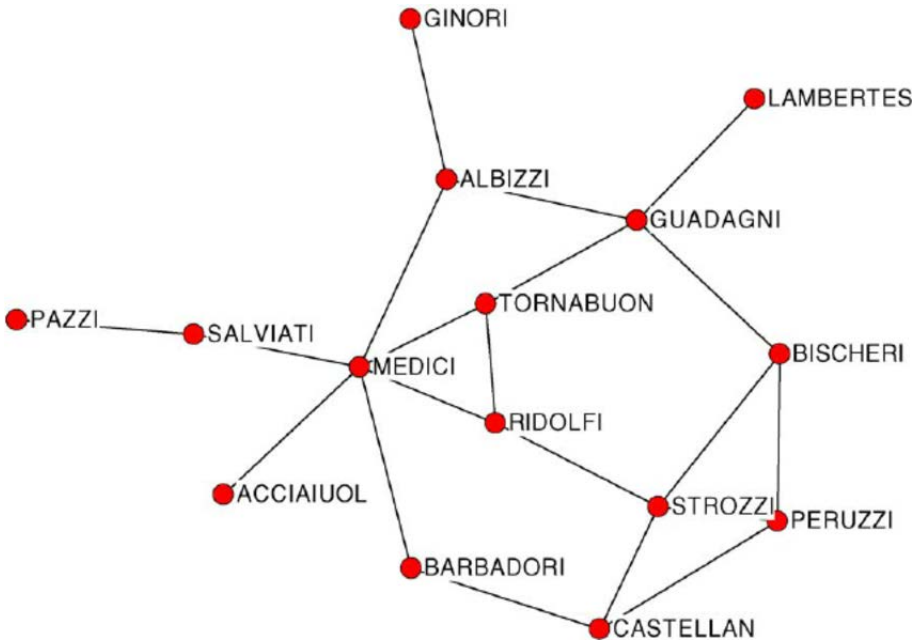
- Different measures target different notions of importance
- In a friendship network, degree centrality would correspond to who is the most popular kid.
- Closeness centrality would correspond to who is closest to the rest of the group,. Useful for spreading information
- Betweenness would tell us about graph "cut points", edges whose deletion will cause multiple connected components



[Ortiz-Arroyo 10]

Centrality and power in networks

- Why were the Medicis the most influential family in 15^{th} c. Florence?
- Political and friendship structure [Padgett & Ansell 93]



Source: Stephen P. Borgatti. "Centrality and Network Flow." *Social Networks* 27 (2005): 55–71.

- Highest betweenness centrality by far

  $\Rightarrow$ Part of many deals between families supported by marriage linkages

Katz centrality

- Degree centrality only depends on the one-hop neighbors of a node
  - $\Rightarrow$ One-hop neighbors more relevant than two-hop neighbors
  - $\Rightarrow$ But indirect relationships are still relevant
- $[\mathbf{A}^k]_{ij}$ contains the number of walks from i to j of length k
  - $\Rightarrow$ Consider the degrees of $\mathbf{A}$, $\mathbf{A}^2$, ..., with a discount factor

$$C_K(i) := \sum_{k=1}^{\infty} \alpha^k \sum_{(i,j)\in E} [\mathbf{A}^k]_{ij} = \left[(\mathbf{I} - \alpha\mathbf{A})^{-1}\mathbf{1}\right]_i$$

  - $\Rightarrow$ where α is small enough to ensure that the series converges
- Katz centrality as a hybrid between degree and eigenvector
  - $\Rightarrow$ Parameter α controls this transition

1.022 Introduction to Network Models

Fall 2018