NETWORK MEASURES

Social Network Analysis

Piyawat L Kumjorn

1

From Social Media Mining: An Introduction.
By Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu.
Cambridge University Press, 2014.

Network Measures

- Centrality
  - How important a node is within a network
  - User Influence
- Transitivity and Reciprocity
  - How links (edges) are formed in a social graph
  - Link Prediction
- Similarity (Structural, Regular)
  - Compute similarity between two nodes in a network
  - Community Analysis, Behavior Prediction

Centrality

How important a node is within a network

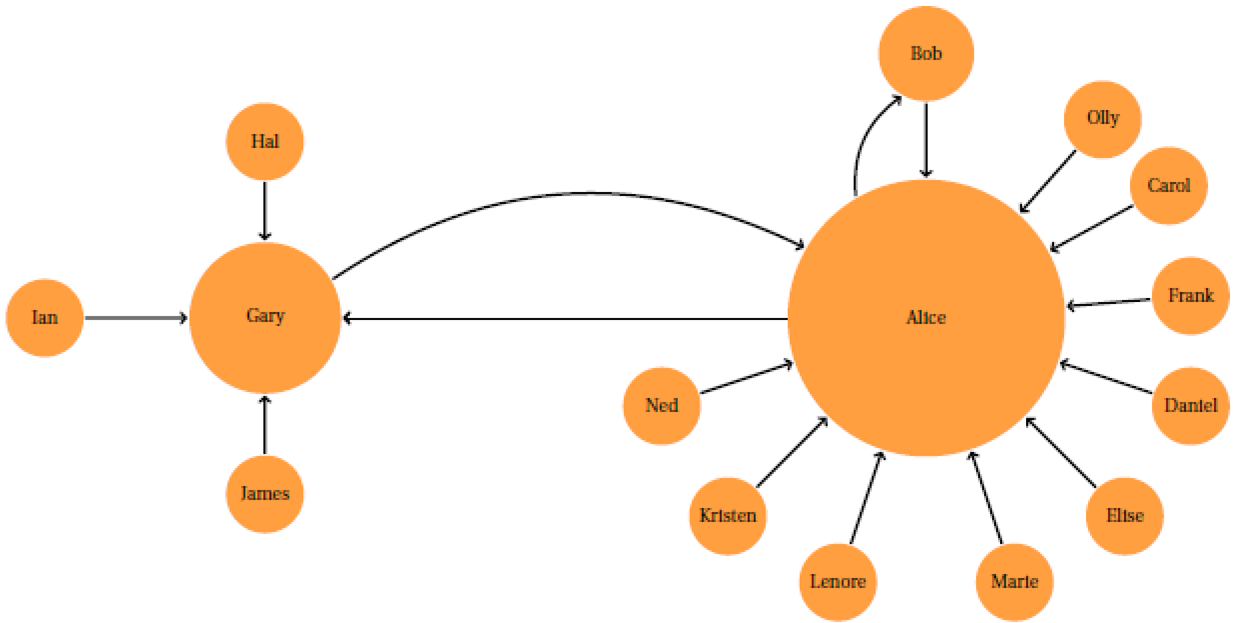
Centrality Measures

- Single Node Centrality
  - Degree Centrality
  - Eigenvector Centrality
  - Katz Centrality
  - PageRank
  - Betweenness Centrality
  - Closeness Centrality
- Group Centrality
  - Degree Centrality
  - Betweenness Centrality
  - Closeness Centrality

DEGREE CENTRALITY

$$
\begin{aligned}
C_d(v_i) &= d_i^{\text{in}} && (\textit{prestige}),\\
C_d(v_i) &= d_i^{\text{out}} && (\textit{gregariousness}),\\
C_d(v_i) &= d_i^{\text{in}} + d_i^{\text{out}}.
\end{aligned}
$$

- Count the number of links attached to the node
- The key question was “how many people retweeted this node?”



Normalizing Degree Centrality

Simple normalization methods include normalizing by the maximum possible degree,

$$C_d^{\text{norm}}(v_i) = \frac{d_i}{n-1}, \tag{3.5}$$

where n is the number of nodes. We can also normalize by the maximum degree,

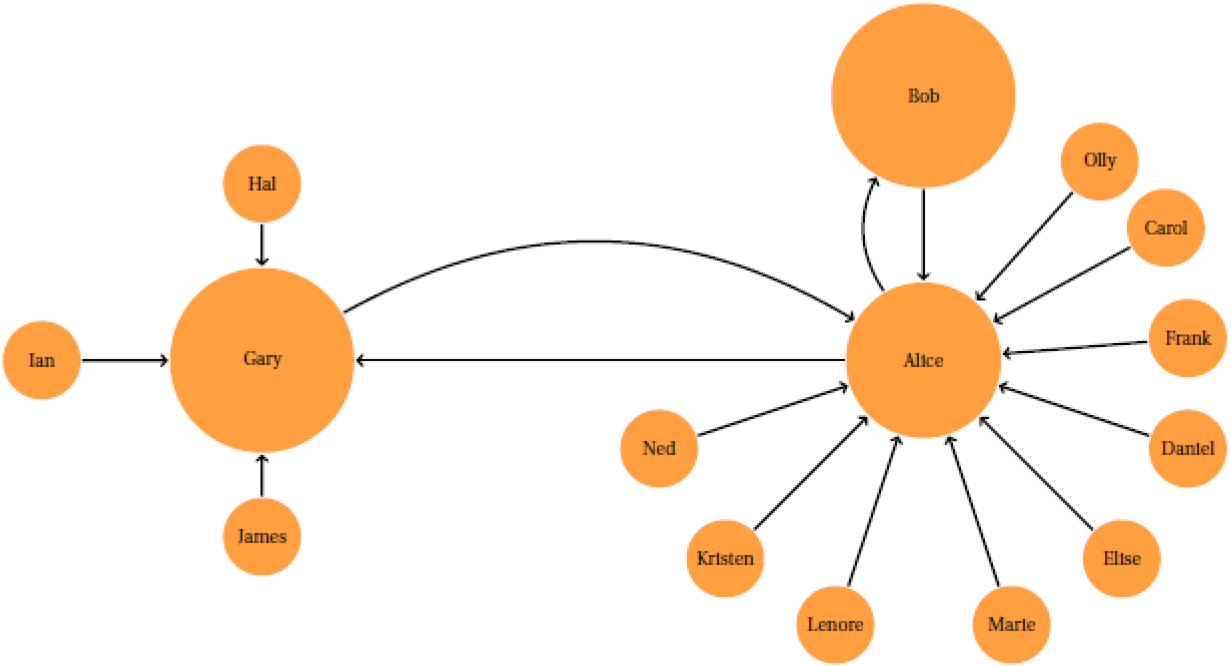
$$C_d^{\max}(v_i) = \frac{d_i}{\max_j d_j}. \tag{3.6}$$

Finally, we can normalize by the degree sum,

$$C_d^{\text{sum}}(v_i) = \frac{d_i}{\sum_j d_j} = \frac{d_i}{2|E|} = \frac{d_i}{2m}. \tag{3.7}$$

Eigenvalue Centrality

- Eigenvector centrality tries to generalize degree centrality by incorporating the importance of the neighbors (focus on incoming neighbors)
- Eigenvector Centrality builds upon this to ask "how important are these retweeters?"



Eigenvalue Centrality

- To keep track of neighbors, we can use the adjacency matrix A of a graph.
- Let $c_e(v_i)$ denote the eigenvector centrality of node v_i.

$$c_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} c_e(v_j),$$

- where lambda is some fixed constant.
- Assuming $\mathbf{C}_e = (C_e(v_1), C_e(v_2), \ldots, C_e(v_n))^T$ is the centrality vectors for all nodes

$$\lambda \mathbf{C}_e = A^T \mathbf{C}_e.$$

Eigenvalue Centrality

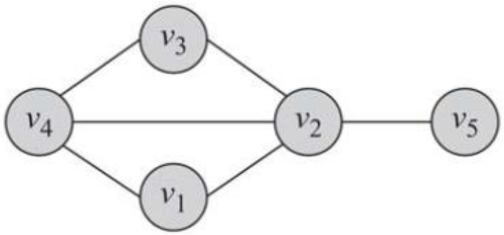
- From $\lambda \mathbf{C}_e = A^T \mathbf{C}_e$.

This basically means that $\mathbf{C}_e$ is an eigenvector of adjacency matrix A^T (or A in undirected networks, since $A = A^T$) and λ is the corresponding eigenvalue.

- To have positive centrality values, we can compute the eigenvalues of A and then select the largest eigenvalue
- The corresponding eigenvector is $\mathbf{C_e}$

  (Eigenvector Centrality of the graph)

Example 3.3. *For the graph shown in Figure 3.2(b), the adjacency matrix is as follows:*



$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3.17}$$

The eigenvalues of A are $(-1.74, -1.27, 0.00, +0.33, +2.68)$. For eigenvector centrality, the largest eigenvalue is selected: 2.68. The corresponding eigenvector is the eigenvector centrality vector and is

$$\mathbf{C}_e = \begin{bmatrix} 0.4119 \\ 0.5825 \\ 0.4119 \\ 0.5237 \\ 0.2169 \end{bmatrix}. \tag{3.18}$$

Based on eigenvector centrality, node v_2 is the most central node.

Betweeness Centrality

- For a node v_i, compute the number of shortest paths between other nodes that pass through v_i,

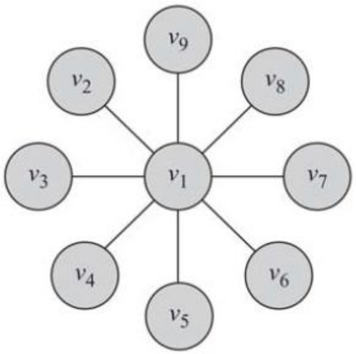
$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}},$$

where σ_{st} is the number of shortest paths from node s to t (also known as *information pathways*), and $\sigma_{st}(v_i)$ is the number of shortest paths from s to t that pass through v_i. In other words, we are measuring how central v_i's role is in connecting any pair of nodes s and t. This measure is called *betweenness centrality*.

Betweeness Centrality

- Possible maximum of betweeness centrality is

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}} = \sum_{s \neq t \neq v_i} 1 = 2\binom{n-1}{2} = (n-1)(n-2).$$



- Hence, to normalize betweeness centrality,

$$C_b^{\text{norm}}(v_i) = \frac{C_b(v_i)}{2\binom{n-1}{2}}.$$

Closeness Centrality

- The more central nodes are, the more quickly they can reach other nodes.
- Formally, these nodes should have a smaller average shortest path length to other nodes.

$$C_c(v_i) = \frac{1}{\bar{l}_{v_i}},$$

where $\bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_j \neq v_i} l_{i,j}$ is node v_i's average shortest path length to other nodes.

Group "Degree" Centrality

- Let S denote the set of nodes to be measured for centrality. Let V – S denote the set of nodes not in the group.
- Group degree centrality is defined as the number of nodes from outside the group that are connected to group members.

$$C_d^{\text{group}}(S) = |\{v_i \in V - S | v_i \text{ is connected to } v_j \in S\}|.$$

- In-degree centrality, Out-degree centrality, and Normalization can also be applied. (Maximum value = |V – S|)

Group "Betweeness" Centrality

$$C_b^{\text{group}}(S) = \sum_{s \neq t, s \notin S, t \notin S} \frac{\sigma_{st}(S)}{\sigma_{st}}$$

- where $\sigma_{st}(S)$ denotes the number of shortest paths between s and t that pass through members of S.
- Maximum value, $2\binom{|V-S|}{2}$ can be used for normalization.

Group "Closeness" Centrality

Closeness centrality for groups can be defined as

$$C_c^{\text{group}}(S) = \frac{1}{\bar{l}_S^{\text{group}}}, \tag{3.45}$$

where $\bar{l}_S^{\text{group}} = \frac{1}{|V-S|}\sum_{v_j \notin S} l_{S,v_j}$ and l_{S,v_j} is the length of the shortest path between a group S and a nonmember $v_j \in V - S$. This length can be defined in multiple ways. One approach is to find the closest member in S to v_j:

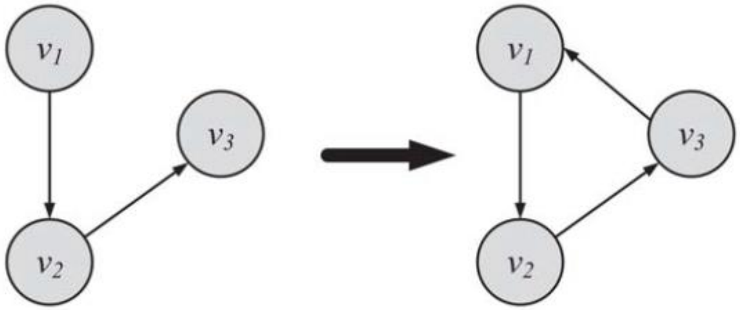
$$l_{S,v_j} = \min_{v_i \in S} l_{v_i,v_j}. \tag{3.46}$$

One can also use the maximum distance or the average distance to compute this value.

Transitivity and Reciprocity

How links (edges) are formed in a social graph

TRANSITIVITY



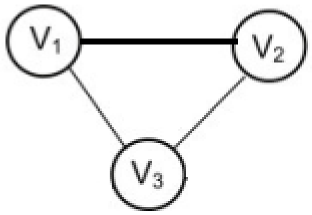
- Transitivity is when a friend of my friend is my friend.
- Higher transitivity in a graph results in a denser graph, which in turn is closer to a complete graph.
- Thus, we can determine how close graphs are to the complete graph by measuring transitivity.
- 2 types:
  - [global] clustering coefficient
  - local clustering coefficient

Global Clustering Coefficient

The clustering coefficient analyzes transitivity in an undirected graph. Since transitivity is observed when triangles are formed, we can measure it by counting paths of length 2 (edges (v_1, v_2) and (v_2, v_3)) and checking whether the third edge (v_3, v_1) exists (i.e., the path is closed).

$$C = \frac{|\text{Closed Paths of Length 2}|}{|\text{Paths of Length 2}|}$$

$$C = \frac{(\text{Number of Triangles}) \times 6}{|\text{Paths of Length 2}|}.$$



1 triangle contains
6 closed path of length 2

(v1,v2)	*(v2,v3)*	*(v3,v1)*
(1,2)	(2,3)	(3,1)
(1,3)	(3,2)	(2,1)
(2,1)	(1,3)	(3,2)
(2,3)	(3,1)	(1,2)
(3,1)	(1,2)	(2,3)
(3,2)	(2,1)	(1,3)

Global Clustering Coefficient

- The global clustering coefficient can also be defined as

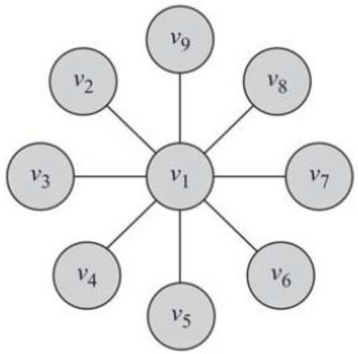
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}}$$

$$= \frac{\text{number of closed triplets}}{\text{number of connected triplets of vertices}}.$$

- A triple is an ordered set of three nodes, connected by two (i.e., open triple) or three (closed triple) edges.
- Two triplets are different when
  - their nodes are different, or
  - their nodes are the same, but the triplets are missing different edges.
- So, one triangle creates three different connected and closed triplets

Local Clustering Coefficient

- If a vertex v_i has k_i neighbors, $k_i(k_i\text{-}1)/2$ edges can exist among the vertices within the neighborhood.



- The clustering coefficient is defined as

$$C_i = \frac{2\, n_i}{k_i\,(k_i - 1)}$$

where n_i denotes the number of links connecting the k_i neighbors of node i.

- The average clustering coefficient is given by

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

RECIPROCITY

- Reciprocity is "If you become my friend, I'll be yours."
- Reciprocity counts the number of reciprocal pairs in the graph. Any directed graph can have a maximum of | E | / 2 pairs.
- Reciprocity can be computed using the adjacency matrix A

$$R = \frac{\sum_{i,j,i<j} A_{i,j} A_{j,i}}{|E|/2} = \frac{1}{|E|} \mathrm{Tr}(A^2),$$

where $\mathrm{Tr}(A) = A_{1,1} + A_{2,2} + \cdots + A_{n,n} = \sum_{i=1}^{n} A_{i,i}$

Similarity

Compute similarity between two nodes in a network

Structural Equivalence

- The size of this neighborhood defines how similar two nodes are.

$$\sigma(v_i, v_j) = |N(v_i) \cap N(v_j)|.$$

  Let $N(v_i)$ and $N(v_j)$ be the neighbors of nodes v_i and v_j, respectively
- By normalization,

$$\sigma_{\text{Jaccard}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|},$$

$$\sigma_{\text{Cosine}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}}.$$

Structural Equivalence

$$\sigma_{\text{significance}}(v_i, v_j) = \sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j),$$

$$\begin{aligned}
\sigma_{\text{pearson}}(v_i, v_j) &= \frac{\sigma_{\text{significance}}(v_i, v_j)}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}} \\
&= \frac{\sum_k (A_{i,k} - \bar{A}_i)(A_{j,k} - \bar{A}_j),}{\sqrt{\sum_k (A_{i,k} - \bar{A}_i)^2}\sqrt{\sum_k (A_{j,k} - \bar{A}_j)^2}},
\end{aligned}$$

Structural Cohesion Measures

Name:	Description:
Density (Harary 1969)	The proportion of group members who are tied (with a "positive" relation, such as friendship, respect, acquaintance, past collaboration, etc.).
Average or maximum Distance (Harary 1969)	The average (or maximum) graph-theoretic distance between all pairs of members
Centralization/Core-Periphery Structure (Freeman 1979; Borgatti & Everett 1996)	The extent to which the network is NOT divided into cliques that have few connections <u>between</u> groups
Homophily* (Marsden 1988)	The extent to which members of the group have their closest ties to members who are similar to themselves