Network Centrality

Based on materials by Lada Adamic, UMichigan
Network Centrality

Which nodes are most ‘central’?

Definition of ‘central’ varies by context/purpose.

Local measure:
  degree

Relative to rest of network:
  closeness, betweenness,
  eigenvector (Bonacich power centrality)

How evenly is centrality distributed among nodes?
  centralization…

Applications:
  Friedkin: Interpersonal Influence in Groups
  Baker: The Social Organization of Conspiracy
Centrality: Who’s Important Based On Their Network Position

In each of the following networks, X has higher centrality than Y according to a particular measure.
He or she who has many friends is most important.

*Degree Centrality (Undirected)*

When is the number of connections the best centrality measure?

- people who will do favors for you
- people you can talk to / have coffee with
Degree: Normalized Degree Centrality

divide by the max. possible, i.e. (N-1)
Centralization: How Equal Are The Nodes?

How much variation is there in the centrality scores among the nodes?

Freeman’s general formula for centralization (can use other metrics, e.g. gini coefficient or standard deviation):

\[
C_D = \sum_{i=1}^{g} \left[ C_D(n^*) - C_D(i) \right] / \left[ (N - 1)(N - 2) \right]
\]

maximum value in the network
Degree Centralization Examples

$C_D = 1.0$

$C_D = 0.167$

$C_D = 0.167$
Degree Centralization Examples

example financial trading networks

high centralization: one node trading with many others

low centralization: trades are more evenly distributed
When Degree Isn’t Everything

In what ways does degree fail to capture centrality in the following graphs?
In What Contexts May Degree Be Insufficient To Describe Centrality?

- ability to broker between groups
- likelihood that information originating anywhere in the network reaches you…
Betweenness: Another Centrality Measure

- Intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?
- Who has higher betweenness, X or Y?
Betweenness Centrality: Definition

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

Where $g_{jk}$ = the number of geodesics connecting $jk$, and $g_{jk}(i)$ = the number of geodesics that actor $i$ is on.

Usually normalized by:

$$C'_B(i) = C_B(i) / \left[ (n-1)(n-2)/2 \right]$$

adapted from a slide by James Moody
Example

Example facebook network: nodes are sized by degree, and colored by betweenness.
Can you spot nodes with high betweenness but relatively low degree?

Explain how this might arise.

What about high degree but relatively low betweenness?
Betweenness On Toy Networks

- non-normalized version:

A lies between no two other vertices
B lies between A and 3 other vertices: C, D, and E
C lies between 4 pairs of vertices (A,D),(A,E),(B,D),(B,E)

- note that there are no alternate paths for these pairs to take, so C gets full credit
Betweenness On Toy Networks

- non-normalized version:
Betweenness On Toy Networks

- non-normalized version:
Betweenness On Toy Networks

- non-normalized version:

  - why do C and D each have betweenness 1?
  - They are both on shortest paths for pairs (A,E), and (B,E), and so must share credit:
    - $\frac{1}{2} + \frac{1}{2} = 1$
  - Can you figure out why B has betweenness 3.5 while E has betweenness 0.5?
All-pairs shortest paths...

“Floyd-Warshall algorithm”

Matrix representation

\[
\begin{array}{cccccc}
\text{FROM} & A & B & C & D & E \\
\text{TO} & A & 0 & 8 & 13 & - & 1 \\
& B & - & 0 & - & 6 & 12 \\
& C & - & 9 & 0 & - & - \\
& D & 7 & - & 0 & 0 & - \\
& E & - & - & - & 11 & 0
\end{array}
\]
All-pairs shortest paths...

$D^0 = (d_{ij}^0)$

\[
\begin{pmatrix}
A & 0 & 8 & 13 & - & 1 \\
B & - & 0 & - & 6 & 12 \\
C & - & 9 & 0 & - & - \\
D & 7 & - & 0 & 0 & - \\
E & - & - & - & 11 & 0 \\
\end{pmatrix}
\]

$d_{ij}^k = \text{shortest distance from } i \text{ to } j \text{ through } \{1, \ldots, k\}$

$D^1 = (d_{ij}^1)$

\[
\begin{pmatrix}
A & 0 & 8 & 13 & - & 1 \\
B & - & 0 & - & 6 & 12 \\
C & - & 9 & 0 & - & - \\
D & 7 & 15 & 0 & 0 & 8 \\
E & - & - & - & 11 & 0 \\
\end{pmatrix}
\]
All-pairs shortest paths...

$$D^2 = (d_{ij}^2)$$

\[
\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & 0 & 8 & 13 & 14 & 1 \\
B & - & 0 & - & 6 & 12 \\
C & - & 9 & 0 & 15 & 21 \\
D & 7 & 15 & 0 & 0 & 8 \\
E & - & - & - & 11 & 0 \\
\end{array}
\]

$$D^3 = (d_{ij}^3)$$

\[
\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & 0 & 8 & 13 & 14 & 1 \\
B & - & 0 & - & 6 & 12 \\
C & - & 9 & 0 & 15 & 21 \\
D & 7 & 9 & 0 & 0 & 8 \\
E & - & - & - & 11 & 0 \\
\end{array}
\]

$$D^4 = (d_{ij}^4)$$

\[
\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & 0 & 8 & 13 & 14 & 1 \\
B & 13 & 0 & 6 & 6 & 12 \\
C & 22 & 9 & 0 & 15 & 21 \\
D & 7 & 9 & 0 & 0 & 8 \\
E & 18 & 20 & 11 & 11 & 0 \\
\end{array}
\]

$$D^5 = (d_{ij}^5)$$

\[
\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & 0 & 8 & 12 & 12 & 1 \\
B & 13 & 0 & 6 & 6 & 12 \\
C & 22 & 9 & 0 & 15 & 21 \\
D & 7 & 9 & 0 & 0 & 8 \\
E & 18 & 20 & 11 & 11 & 0 \\
\end{array}
\]

to store the path, another matrix can track the last intermediate vertex
Floyd-Warshall Pseudocode

Input: $D^0 = (d_{ij}^0)$ (the initial edge-cost matrix)

Output: $D^n = (d_{ij}^n)$ (the final path-cost matrix)

for $k = 1$ to $n$  // intermediate vertices considered
    for $i = 1$ to $n$  // the “from” vertex
        for $j = 1$ to $n$  // the “to” vertex
            $d_{ij}^k = \min\{d_{ij}^{k-1}, d_{ik}^{k-1} + d_{kj}^{k-1}\}$

best, ignoring vertex $k$

best, including vertex $k$
Closeness: Another Centrality Measure

- What if it’s not so important to have many direct friends?
- Or be “between” others
- But one still wants to be in the “middle” of things, not too far from the center
Closeness Centrality: Definition

Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph.

Closeness Centrality:

\[
C_c(i) = \left[ \sum_{j=1}^{N} d(i,j) \right]^{-1}
\]

Normalized Closeness Centrality

\[
C_c^\prime(i) = (C_c(i))/(N - 1)
\]
Closeness Centrality: Toy Example

\[ C'_c(A) = \left[ \frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[ \frac{1 + 2 + 3 + 4}{4} \right]^{-1} = \left[ \frac{10}{4} \right]^{-1} = 0.4 \]
Closeness Centrality: More Toy Examples

![Graph 1](Image)

![Graph 2](Image)
How Closely Do Degree And Betweenness Correspond To Closeness?

- **degree** (number of connections) denoted by size

- **closeness** (length of shortest path to all others) denoted by color
Centrality: Check Your Understanding

- generally different centrality metrics will be positively correlated
- when they are not, there is likely something interesting about the network
- suggest possible topologies and node positions to fit each square

<table>
<thead>
<tr>
<th></th>
<th>Low Degree</th>
<th>Low Closeness</th>
<th>Low Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Closeness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Betweenness</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Generally different centrality metrics will be positively correlated. When they are not, there is likely something interesting about the network. Suggest possible topologies and node positions to fit each square.

<table>
<thead>
<tr>
<th>Centrality</th>
<th>Low Degree</th>
<th>Low Closeness</th>
<th>Low Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embedded in cluster that is far from the rest of the network</td>
<td></td>
<td>Ego's connections are redundant - communication bypasses him/her</td>
<td></td>
</tr>
<tr>
<td>High Closeness</td>
<td>Key player tied to important/active players</td>
<td></td>
<td>Probably multiple paths in the network, ego is near many people, but so are many others</td>
</tr>
<tr>
<td>High Betweenness</td>
<td>Ego's few ties are crucial for network flow</td>
<td>Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others.</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from a slide by James Moody.
We now consider the fraction of all directed paths between any two vertices that pass through a node.

Only modification: when normalizing, we have \((N-1)(N-2)\) instead of \((N-1)(N-2)/2\), because we have twice as many ordered pairs as unordered pairs.

\[
C_B(i) = \frac{\sum g_{jk}(i)}{g_{jk}}
\]

\[
C'_B(i) = \frac{C_B(i)}{[(N-1)(N-2)]}
\]
A node does not necessarily lie on a geodesic from $j$ to $k$ if it lies on a geodesic from $k$ to $j$
Extensions Of Undirected Degree Centrality - Prestige

- degree centrality
  - indegree centrality
    - a paper that is cited by many others has high prestige
    - a person nominated by many others for a reward has high prestige
Extensions Of Undirected Closeness Centrality

- Closeness centrality usually implies:
  - All paths should lead to you
  - And unusually not:
  - Paths should lead from you to everywhere else

- Usually consider only vertices from which the node $i$ in question can be reached
The influence range of $i$ is the set of vertices who are reachable from the node $i$. 
Wrap Up

Centrality

- many measures: degree, betweenness, closeness, ...
- may be unevenly distributed
  - measure via centralization
- extensions to directed networks:
  - prestige
  - influence
  - PageRank
Additional Material

(Not covered in class)
An eigenvector measure:

\[ C(\alpha, \beta) = \alpha(I - \beta R)^{-1} R1 \]

- \( \alpha \) is a scaling vector, which is set to normalize the score.
- \( \beta \) reflects the extent to which you *weight* the centrality of people ego is tied to.
- \( R \) is the adjacency matrix (can be valued)
- \( I \) is the identity matrix (1s down the diagonal)
- \( 1 \) is a matrix of all ones.

Bonachich Power Centrality:
When Your Centrality Depends On Your Neighbors’ Centrality

adapted from a slide by James Moody
The magnitude of $\beta$ reflects the radius of power. Small values of $\beta$ weight local structure, larger values weight global structure.

If $\beta > 0$, ego has higher centrality when tied to people who are central.

If $\beta < 0$, then ego has higher centrality when tied to people who are not central.

With $\beta = 0$, you get degree centrality.
Bonacich Power Centrality: Examples

\( \beta = 0.25 \)

\( \beta = -0.25 \)

Why does the middle node have lower centrality than its neighbors when \( \beta \) is negative?
Centrality When Edges Are Directed
Review: Examples Of Directed Networks

- WWW
- food webs
- population dynamics
- influence
- hereditary
- citation
- transcription regulation networks
- neural networks
Prestige In Directed Social Networks

- when ‘prestige’ may be the right word
  - admiration
  - influence
  - gift-giving
  - trust

- directionality especially important in instances where ties may not be reciprocated (e.g. dining partners choice network)

- when ‘prestige’ may not be the right word
  - gives advice to (can reverse direction)
  - gives orders to (- ” -)
  - lends money to (- ” -)
  - dislikes
  - distrusts
Extensions Of Undirected Degree Centrality - Prestige

- degree centrality
  - indegree centrality
    - a paper that is cited by many others has high prestige
    - a person nominated by many others for a reward has high prestige
Extensions Of Undirected Closeness Centrality

- Closeness centrality usually implies:
  - All paths should lead to you
  - And unusually not:
  - Paths should lead from you to everywhere else

- Usually consider only vertices from which the node $i$ in question can be reached.
Influence Range

- The influence range of $i$ is the set of vertices who are reachable from the node $i$.
Prestige in Pajek

Calculating the indegree prestige

- Net>Partition>Degree>Input
- to view, select File>Partition>Edit
- if you need to reverse the direction of each tie first (e.g. lends money to -> borrows from):
  Net>Transform>Transpose

Influence range (a.k.a. input domain)

- Net>k-Neighbours>Input
  - enter the number of the vertex, and 0 to consider all vertices that eventually lead to your chosen vertex
  - to find out the size of the input domain, select Info>Partition
- Calculate the size of the input domains for all vertices
  - Net>Partitions>Domain>Input
- Can also limit to only neighbors within some distance
Direct nominations (choices) should count more than indirect ones
Nominations from second degree neighbors should count more than third degree ones
So consider proximity prestige

\[
P_p(n_i) = \frac{\text{fraction of all vertices that are in } i\text{'s input domain}}{\text{average distance from } i \text{ to vertex in input domain}}
\]
Prestige vs. Centrality In Diffusion

physician discussion network
nodes are sized by indegree

physician friendship network
nodes are sized by degree
Interested in identifying the structural bases of power. In addition to resources, he identifies:

- Cohesion
- Similarity
- Centrality

which are thought to affect interpersonal visibility & salience
Centrality
Central actors are likely more influential. They have greater access to information and can communicate their opinions to others more efficiently. Research shows they are also more likely to use the communication channels than are periphery actors.
Structural Similarity

• Two people may not be directly connected, but occupy a similar position in the structure. As such, they have similar interests in outcomes that relate to positions in the structure.

• Similarity must be conditioned on visibility. P must know that O is in the same position, which means that the effect of similarity might be conditional on communication frequency.
Cohesion

• Members of a cohesive group are likely to be aware of each other's opinions, because information diffuses quickly within the group.

• Groups encourage (through balance) reciprocity and compromise. This likely increases the salience of opinions of other group members, over non-group members.
Substantive questions: Influence in establishing school performance criteria.

• Data on 23 teachers
• Collected in 2 waves
• Dyads are the unit of analysis (P--> O): want to measure the extent of influence of one actor on another.
• Each teacher identified how much an influence others were on their opinion about school performance criteria.

• Cohesion = probability of a flow of events (communication) between them, within 3 steps.
• Similarity = pairwise measure of equivalence (profile correlations)
• Centrality = TEC (power centrality)
Interpersonal communication matters, and communication is what matters most for interpersonal influence.

Questions: How are relations organized to facilitate illegal behavior?

Pattern of communication maximizes concealment, and predicts the criminal verdict.

Inter-organizational cooperation is common, but too much ‘cooperation’ can thwart market competition, leading to (illegal) market failure.

Illegal networks differ from legal networks, in that they must conceal their activity from outside agents. A “Secret society” should be organized to (a) remain concealed and (b) if discovered make it difficult to identify who is involved in the activity.

The need for secrecy should lead conspirators to conceal their activities by creating **sparse** and **decentralized** networks.
### Figure 1. Concealment Versus Coordination: Theoretical Expectations and Experimental Results

center: good for reaping the benefits
periphery: good for remaining concealed

They examine the effect of Degree, Betweenness and Closeness centrality on the criminal outcomes, based on reconstruction of the communication networks involved.

At the organizational level, low information-processing conspiracies are decentralized high information processing load leads to centralization

At the individual level, degree centrality (net of other factors) predicts verdict.
Wrap Up

Centrality

- many measures: degree, betweenness, closeness, Bonacich
- may be unevenly distributed
  - measure via centralization
- extensions to directed networks:
  - prestige
    - input domain…
    - PageRank (down the road…)

consequences:
- interpersonal influence (Friedkin)
- benefits & risks (Baker & Faulkner)