Social and Technological Network Analysis

Lecture 3: Centrality Measures

Dr. Cecilia Mascolo
(some material from Lada Adamic’s lectures)
In This Lecture

• We will introduce the concept of centrality and the various measures which have been associated to this concept.
• We will show an application.
Centrality

• Finding out which is the most central node is important:
  – It could help disseminating information in the network faster
  – It could help stopping epidemics
  – It could help protecting the network from breaking
Centrality: visually

• Centrality can have various meanings:

- Indegree
- Outdegree
- Betweenness
- Closeness
Degree Centrality

When is the number of connections the best centrality measure?
- people who will do favors to you
- people you can talk to / have a beer with
Normalization

• Divide for the max number of nodes (N-1)
Freeman’s Network Centrality

• How do we calculate the value of centrality of the network
  – To check how much variation there is among the nodes (heterogeneity?)

Max value of the above: when network is a star: 1 node has $C=N-1$ and all others $(N-1)$ have 1.

$$C_D = \sum_{i=1}^{\phi} \frac{[C_D(n^*) - C_D(i)]}{[(N - 1)(N - 2)]}$$
Freeman Network

Centrality Explained

• Explanation of the denominator:
• In the star topology one node has degree N-1 and all other nodes have degree of 1:

\[ 0 + ((n-1)-1) \times (n-1) = (n-2) \times (n-1) \]
Freeman’s Network Centrality

$C_D = 1.0$

$4 + 4 + 4 + 4 + 4 / 5 \times 4$

$1 + 0 + 0 + 0 + 1 / 4 \times 3 = 1 / 6$

$1 + 1 + 0 + 1 + 0 + 1 + 1 / 6 \times 5 = 5 / 30$

$C_D = 0.167$

$C_D = 0.167$
Examples: Financial Networks
When is Degree Centrality not so good?
When is Degree Centrality not so good (2)?

• Ability to broker between groups
• Likelihood that information originating anywhere in the network reaches you...
Betweenness Centrality

• 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 (Formally)

\[ C_B(i) = \sum_{j \neq k} g_{jk}(i) / g_{jk} \]

Where \( g_{jk}(i) \) = the number of shortest paths connecting \( jk \) passing through \( i \)
\( g_{jk} \) = total number of shortest paths

Usually normalized by:

\[ C'_B(i) = \frac{C_B(i)}{\left(\frac{(n-1)(n-2)}{2}\right)} \]

number of pairs of vertices excluding the vertex itself
Betweenness: Example

- 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 alternative paths for these pairs to take, so C gets full credit
Facebook Example (Adamic)

Color (from blue to red) is betweenness
Size is degree.
Closeness Centrality

• What if it is 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 (Formally)

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

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

\[ C'_c(i) = \frac{C_c(i)}{(N - 1)} \]
Closeness: Example

\[ C_c'(A) = \left[ \frac{\sum_{j=1}^{N} d(A, j)}{N - 1} \right]^{-1} = \left[ \frac{1 + 2 + 3 + 4}{4} \right]^{-1} = \left[ \frac{10}{4} \right]^{-1} = 0.4 \]
Examples

\[(1+1+2+3+4+4/6)^{-1}=6/15=0.4\]
Example: Facebook (Adamic)

Degree is the size
Color is closeness
Eigenvector Centrality

- Degree Centrality depends on having many connections: but what if these connections are pretty isolated?
- A central node should be one connected to powerful nodes

\[ x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t \]

- Neighbourhood of \( X_v \)
- Adjacency Matrix of the graph

\[ Ax = \lambda x \]
Eigenvector Centrality
Algorithm

• 1. Start by assigning centrality score of 1 to all nodes ($v_i = 1$ for all $i$ in the network)
• 2. Recompute scores of each node as weighted sum of centralities of all nodes in a node's neighborhood: $v_i = \sum_{j \in N} a_{ij} \cdot v_j$
• 3. Normalize $v$ by dividing each value by the largest value
• 4. Repeat steps 2 and 3 until values of $v$ stop changing.
Example

- Node A: Degree: 3, Eigenvector Centrality: 0.182
- Node B: Degree: 4, Eigenvector Centrality: 0.091
- Node C
- Node D
Katz Centrality

- Closeness counts the number of shortest paths, but one could count the **number of paths**.

\[
C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}
\]

Alpha is an attenuation factor
\(A^k (ij)\) indicates if \(ij\) are connected by \(k-1\) hops.
Application of Centrality Measures to Big Data
People Tweet While They Watch
ANATOMY OF A TWEET

Repeated (retweet) content starts with RT

Address other users with an @

Rich Media embeds via links

Tags start with #

RT @jowyang If you are watching the debate you’re invited to participate in #tweetdebate. Here is the 411 http://tinyurl.com/3jdy67
• Three hashtags: #current #debate08 #tweetdebate
• 97 mins debate + 53 mins following = 2.5 hours total.
• 3,238 tweets from 1,160 people.
  – 1,824 tweets from 647 people during the debate.
  – 1,414 tweets from 738 people post debate.
• 577 @ mentions (reciprocity!)
  – 266 mentions during the debate
  – 311 afterwards.
• Low RT: 24 retweets in total
  – 6 during
  – 18 afterwards.
Volume of Tweets

The graph depicts the volume of tweets over time, with data categorized into 'AI Tweets' and 'Address Tweets'. The timeline is marked with 'Debate On Air' and 'Post Debate' periods.
Discussion Segments
Twitter Mentions

John Tweets: “Hey @mary, my person is winning!”

Large node size = High eigenvector centrality
Automatic Discovery through Centrality

## Centrality

<table>
<thead>
<tr>
<th>Twitter User</th>
<th>Eigenvector Centrality</th>
<th>In Degree</th>
<th>Out Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>@barackobama</td>
<td>0.472</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>@newshour</td>
<td>0.427</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>@johnmccain</td>
<td>0.277</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>@charleswinters</td>
<td>0.223</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>@jeremyfranklin</td>
<td>0.223</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>@saleemkhan</td>
<td>0.223</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>@srubenfeld</td>
<td>0.223</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>@msblog</td>
<td>0.221</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>@frijole</td>
<td>0.175</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>
High in-degree but poor centrality:

@current, who ran the Hack the Debate program, and @jowyang, an employee of Forrester Research who uses Twitter as a personal, not corporately related microblog.
Tweets to Terms
Twitter as Reaction

- drinking
- candidate
- wins
- minute
- Lehrer
- let
- tv
- mccains
- tweet
- moderator
- kennedy
- plan
- hope
- hand
- comment
- moderator
- tie
- home
- audience
- republicans
- cut
- earmarks
- compared
- tax
- joke
- bear
- dollars
- problems
- personal
- americans
- screen
- bailout
- energy
- economic
- power
-_condolence
- nuclear
- fact
- war
- giving
- hand
- pulling
- pakistan
- lousy
- democracies
- government
- l0l
- story
- wars
- league
- times
- ha
- lot
- descending
- issue
- oil
- georgia
- video
- blog
- league
- russia
- 9/11
- safe
- country
- tactics
- attack
- blog
- video
- pakistan
- bringing
- security
- experience
- presidential
- debates
- greed
- main
- how's
- house
- direct
- package
- accountable
- wall
- rewarded
- crisis
- street
- $18
- requests
- earmarks
- gateway
- loopholes
- $5,000
- pork-barrel
- employer
- business
- tax
- programs
- medicare
- cost
- eliminate
- $700
- hard
- decisions
- agency
- rescue
- funded
- winning
- leave
- timetable
- violence
- lessons
- bangladesh
- surge
- succeed
- started
- border
- taliban
- prepared
- supported
- u.s.
- bombing
- iranians
- sanctions
- ahmadinejad
- precondition
- georgia
- international
- putin
- ukraine
- safer
- veterans
- world's
- focused
- earth
- billions
- challenges
- restore
- knowledge
- missile
- putin
- russia
- world's
- focused
- earth
- billions
- challenges
Summary

• We have introduced various measures of centrality and explained the pros and cons
• We have illustrated one example of use of centrality in a Twitter related example
• Kleinberg’s book: Chapter 3.
• Ack: L. Adamic’s slides
