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?
  o people who will do favors to you
  o 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?)

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

Max value of Degree Centrality in the Network

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

\[ C_D = 1.0 \]

\[ C_D = 0.167 \]

\[ \frac{4+4+4+4+4}{5} \times 4 \]
\[ \frac{1+0+0+0+1}{4} \times 3 = \frac{1}{6} \]
\[ \frac{1+1+0+1+0+1+1}{6} \times 5 = \frac{5}{30} \]

\[ C_D = 0.167 \]
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)

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

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

\( g_{jk} = \) total number of shortest paths

Usually normalized by:

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

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.
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

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

\[ C'_c(i) = \left( C_c(i) \right) / (N - 1) \]
Closeness: Example

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

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

\[ Ax = \lambda x \]

Neighbourhood of \( X_v \)

Adjacency Matrix of the graph
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
Katz Centrality

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

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

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

RT: @jowyang If you are watching the debate you’re invited to participate in #tweetdebate. Here is the 411:
http://tinyurl.com/3jdy67
Tweet Crawl

• 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
Discussion Segments
Cluster of tags
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>
Sinks

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

- drinking
- candidates
- wins
- minute
- mccains
- tweet
- moderator
- kennedy
- plan
- hope
- hand
- comment
- moderator
- tie
- home
- audience
- republicans
- cut
- earmarks
- compared
- tax
- joke
- bear
- dollars
- problems
- personal
- american
- screen
- bailout
- energy
- economic
- power
- condoulo
- nuclear
- giving
- freeze
- giving
- hand
- senate
- difference
- winning
- strategy
- iraq
- festooned
- fact
- war
- story
- pulling
- pakistan
- understanding
- drinking
- republican
- strategy
- sounds
- coming
- telling
- story
- lousy
- democracies
- lol
- government
- story
- wars
- league
- times
- iran
- -5
- georgia
- video
- blog
- league
- russia
- ha
- lot
- condescending
- issue
- oil
- bringing
- security
- experience
- 9/11
- safe
- country
- tactics
- attack
- video
- pakistan
- bringing
- presidential
- debates
- minutes
- eisenhower
- financial
- direct
- policy
- news
- mississippi
- university
- street
- greed
- main
- how's
- house
- package
- accountable
- wall
- rewarded
- crisis
- $18
- requests
- earmarks
- gateway
- loopholes
- $5,000
- pork-barrel
- employer
- business
- tax
- programs
- medicare
- cost
- eliminate
- $700
- hard
- decisions
- agency
- rescue
- to
- funding
- winning
- leave
- timetable
- violence
- lessons
- baghdad
- surge
- succeed
- started
- border
- taliban
- prepared
- supported
- muddle
- u.s
- bombing
- pakistan
- Qaeda
- army
- henry
- kissinger
- contacts
- preparation
- legitimize
- table
- iranians
- sanctions
- ahmadinejad
- precondition
- georgia
- international
- putin
- ukraine
- russia
- world's
- offshore
- aggression
- resurgent
- nato
- restore
- knowledge
- missile
- safer
- veterans
- terms
- focused
- earth
- billions
- challenges
Twitter as Reaction

drinking

winning

giving

-5

presidential
debates

minutes
eisenhower
financial
direct
policy
news
mississippi
university

street
greed
how's
house
package
accountable
wall
rewarded

$18
requests
earmarks
gateway
loopholes
$5,000
pork-barrel
employer
business

programs
healthcare
cost
eliminate
$700
hard
decisions
agency
rescue

funding
winning
leave
timetable
violence
lessons
agencies
success

border
taliban
prepared
supported
muddle
u.s
bombing
pakistan
al-qaeda
army

henry
kissinger
preparation
preparation
legitimise

georgia
video
blog
league
russia
9/11
safe
country
tactics
attack
video
pakistan

restore
knowledge
putin
safer
veterans
terms
focused
death
billions

relations
with
russia

terrorist
threat

opening
financial
recovery

solving
financial
crisis

financial
recovery

lessons
of
iraq

troops
in
afghanistan

threat
from
iran

university
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
References

• Kleinberg’s book: Chapter 3.
• Ack: L. Adamic’s slides

• M. Newmann. **Networks**. Oxford University Press. April 2010.