Social and Technological Network Data Analytics

Lecture 3: Centrality Measures

Prof. 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?)

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

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.
• 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

$C_D = 1.0$

$C_D = 0.167$

$C_D = 0.167$

\[
\begin{align*}
4 + 4 + 4 + 4 + 4 / 5 \times 4 & = 16\\
1 + 0 + 0 + 0 + 1 / 4 \times 3 & = 1 / 6\\
1 + 1 + 0 + 1 + 0 + 1 + 1 / 6 \times 5 & = 5 / 30
\end{align*}
\]
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:

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

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

• 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

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

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

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

Ax = \lambda x

Neighbourhood of Xv

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

What are you doing? 140

Latest: sec min = :) – boxee about 18 hours ago

Raratequila @BigRed561 oh nah, where is dat?
less than 20 seconds ago from web In reply to BigRed561

tabaligherooz VPS ان 1.200دولار/1فلك لحاسب http://ff.im/6ahAB
less than 20 seconds ago from Friendfeed

jon_cohen @henryb35 no thanks
less than 20 seconds ago from web In reply to henryb35

AndreJayP @HASstyle420 @ducebigz26 @andrejayp @gogoquedawg @jrdafuture7 @silkybutmilky @clrnfuturistic #4thqtr - aint nuffin star... @magicdoom04 ??
less than 20 seconds ago from UberTwitter

lauramemplosey don't forget, kids: throw your support behind @crochelle tonight as she embarks on the world's first known #firsttwitterdate ;)!
less than 20 seconds ago from web

easthawaiarts While sending tweets, always remembering words of early founder Mrs. F. Sherrard: "Art is for EVERYBODY!!!"
www.unicorn.org #art #hawaii

UNIVERSITY OF CAMBRIDGE
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
Tweet Crawl

• Three hashtags: \textit{#current #debate08 #tweetdebate}
• 97 mins debate + 53 mins following = \textbf{2.5 hours} total.
• \textbf{3,238 tweets} from \textbf{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 \textit{during} the debate
  – 311 \textit{afterwards}.
• Low RT: \textbf{24 retweets} in total
  – 6 during
  – 18 afterwards.
Volume of Tweets
Discussion Segments

![Discussion Segments Diagram](image-url)
Cluster of users and tags
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
- drinking candidates wins
- minute
- lehner letting tv 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 condomo nuclear giving freeze looked
- senate difference winning strategy iraq festooned fact
- war giving hand pulling pakistan understand drinking republican strategy sounds coming
telling story
- lousy democracies lol government story wars league times iran -5
- georgia video blog league russia ha lot
cdescending issue oil
- security experience 9/11 safe
country tactics attack video pakistan bringing

## Topic
- presidential debates minutes eisenhower financial direct policy news mississippi university
- street greed main how’s house package accountable wall
- $18 requests earmarks gateway loopholes $5,000 pork-barrel employer business tax
- programs medicare cost eliminate $700 hard decisions agency rescue tom
- 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

## Terrorist Threat
- georgia international putin ukraine safer
- russia world’s offshore aggression resurgent nato
- restore knowledge missile focused
terms earth billions challenges
Twitter as Reaction

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>presidential</td>
<td>street</td>
<td>$18</td>
<td>programs</td>
<td>funding</td>
<td>border</td>
<td>georgia</td>
<td>international</td>
<td>restore</td>
</tr>
<tr>
<td></td>
<td>debates</td>
<td>greed</td>
<td>requests $18</td>
<td>medicare</td>
<td>winning</td>
<td>taliban</td>
<td>video</td>
<td>knowledge</td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td>minutes</td>
<td>how's</td>
<td>earmarks</td>
<td>cost</td>
<td>leave</td>
<td>prepared</td>
<td>blog</td>
<td>contacts</td>
<td>missile</td>
</tr>
<tr>
<td></td>
<td>eisenhower</td>
<td>house</td>
<td>gateway</td>
<td>eliminate</td>
<td>eliminate</td>
<td>supported</td>
<td>league</td>
<td>ukraine</td>
<td>contacts</td>
</tr>
<tr>
<td></td>
<td>financial</td>
<td>package</td>
<td>loopholes</td>
<td>$700</td>
<td>violence</td>
<td>legimize</td>
<td>russia</td>
<td>veterans</td>
<td>terms</td>
</tr>
<tr>
<td></td>
<td>direct</td>
<td>accountable</td>
<td>pork-barrel</td>
<td>hard</td>
<td>u.s</td>
<td>table</td>
<td>safer</td>
<td>world's</td>
<td>focused</td>
</tr>
<tr>
<td></td>
<td>policy</td>
<td>wall</td>
<td>employer</td>
<td>decisions</td>
<td>bombing</td>
<td>iranians</td>
<td>veterans</td>
<td>offshore</td>
<td>billions</td>
</tr>
<tr>
<td></td>
<td>news</td>
<td>rewarded</td>
<td>business</td>
<td>agency</td>
<td>pakistani</td>
<td>sanctions</td>
<td>aggression</td>
<td>terms</td>
<td>challenges</td>
</tr>
<tr>
<td></td>
<td>mississippi</td>
<td>crisis</td>
<td>tax</td>
<td>rescue</td>
<td>al Qaeda</td>
<td>ahmadinejad</td>
<td>aggression</td>
<td>focus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>university</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Keywords:
- drinking
- candidates
- kennedy
- plan
- hope
- hand
- comment
- moderator
- tie
- home
- audience
- republicans
- cut
- earmarks
- compared
- tax
- joke
- bear
- dollars
- problems
- personal
- american
- giving
- freeze
- locked
- pulling
- pakistan
- understood
- drinking
- republican
- strategy
- sound
- coming
- telling
- story
- lousy
- democracies
- government
- lol
- ha
- lot
--condescending
- issue
- oil
- security
- experience
- 9/11
- safe
- country
- tactics
- attack
- video
- pakistan
- iran
- oil
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
