



Social and Technological Network Analysis

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

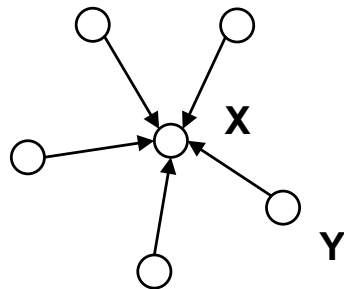


- Finding out which is the most central node is important:
 - It could help disseminating information in the network faster
 - It could help stopping epidemics
 - It could help protecting the network from breaking

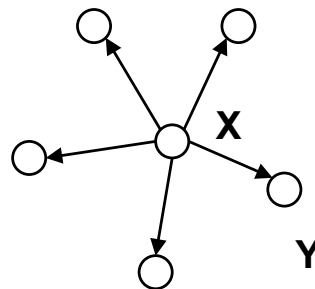
Centrality: visually



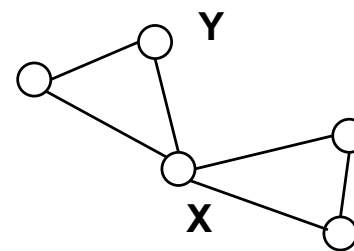
- Centrality can have various meanings:



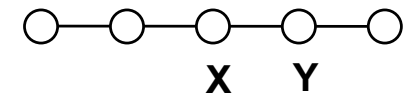
indegree



outdegree

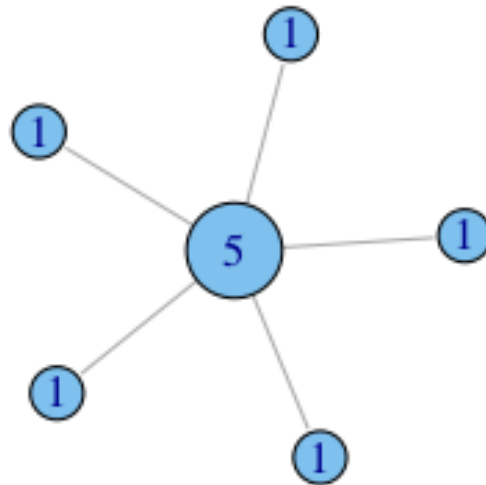


betweenness



closeness

Degree Centrality



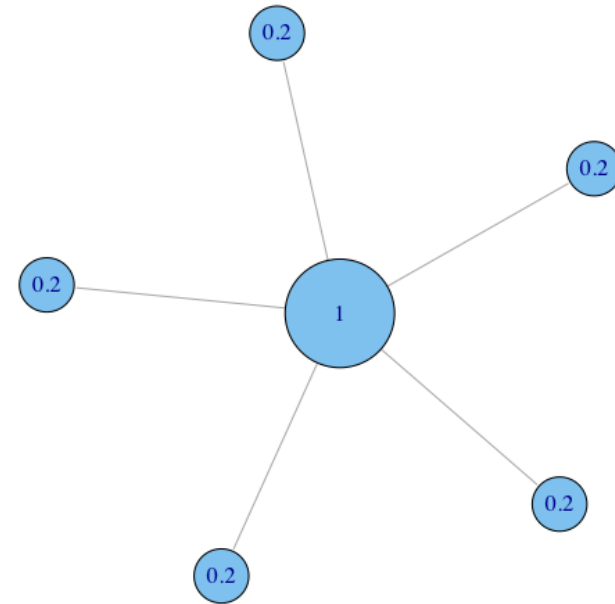
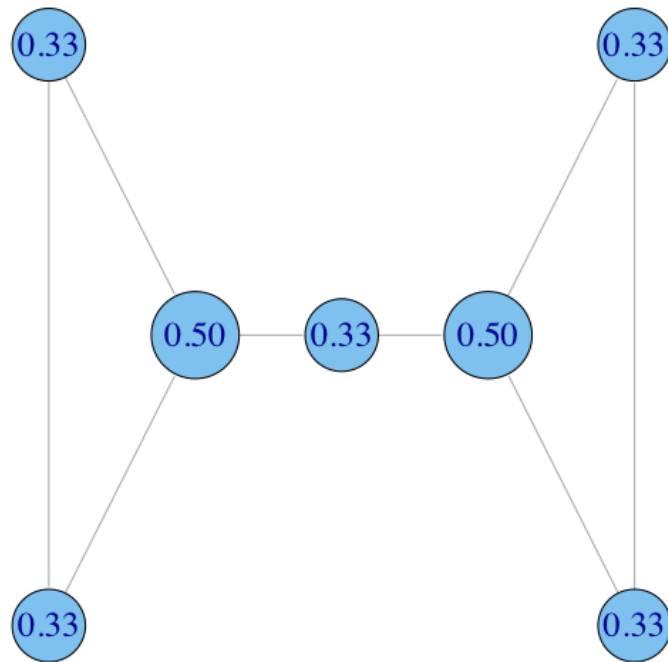
When is the number of connections the best centrality measure?

- 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?)

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

Max value of Degree Centrality in the Network

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

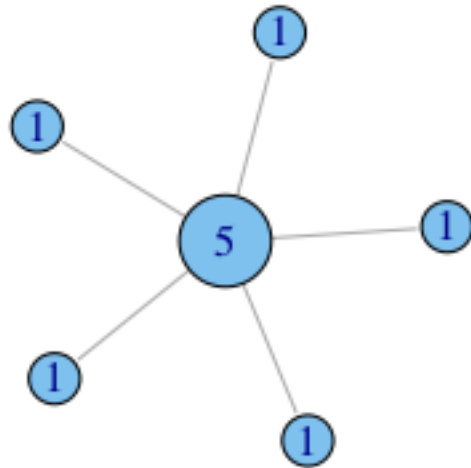
Freeman Network Centrality Explained



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

$$0 + ((n-1) - 1) * n - 1 = (n-2) * (n-1)$$

Freeman's Network Centrality

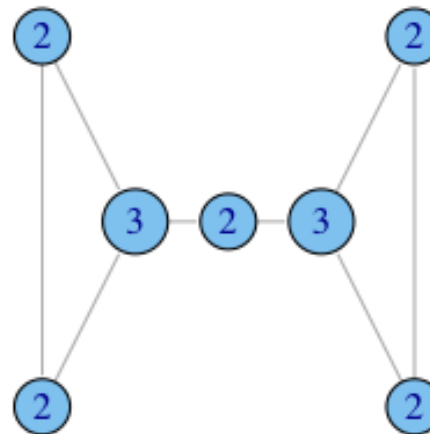


$$C_D = 1.0$$

$$4+4+4+4+4/5*4$$
$$1+0+0+0+1/4*3=1/6$$
$$1+1+0+1+0+1+1/6*5=5/30$$

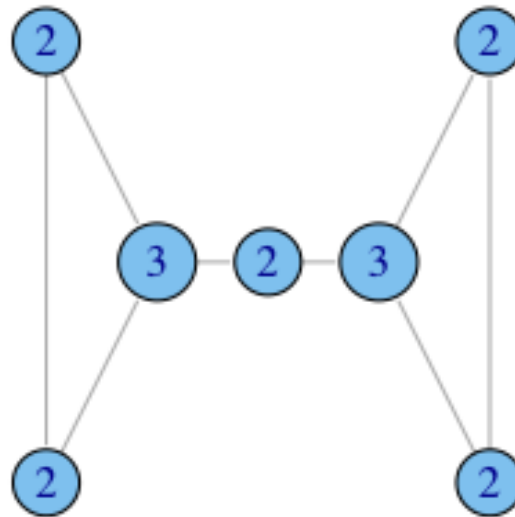


$$C_D = 0.167$$

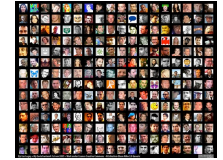


$$C_D = 0.167$$

When is Degree Centrality not so good?



When is Degree Centrality not so good (2)?

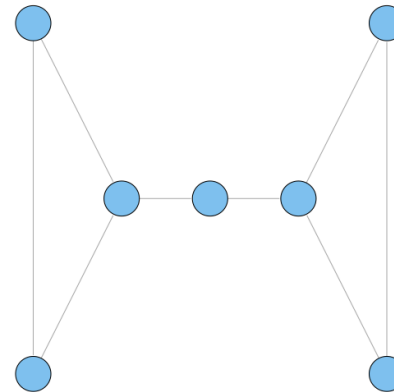
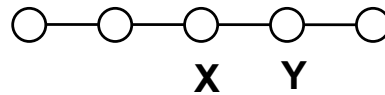
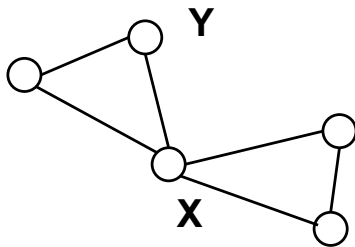


- Ability to broker between groups
- Likelihood that information originating anywhere in the network reaches you...

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:

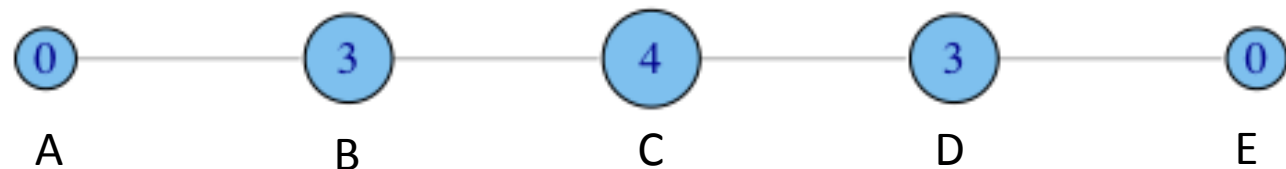
$$C'_B(i) = C_B(i) / [(n-1)(n-2)/2]$$

number of pairs of vertices
excluding the vertex itself

Betweenness: Example of nominators



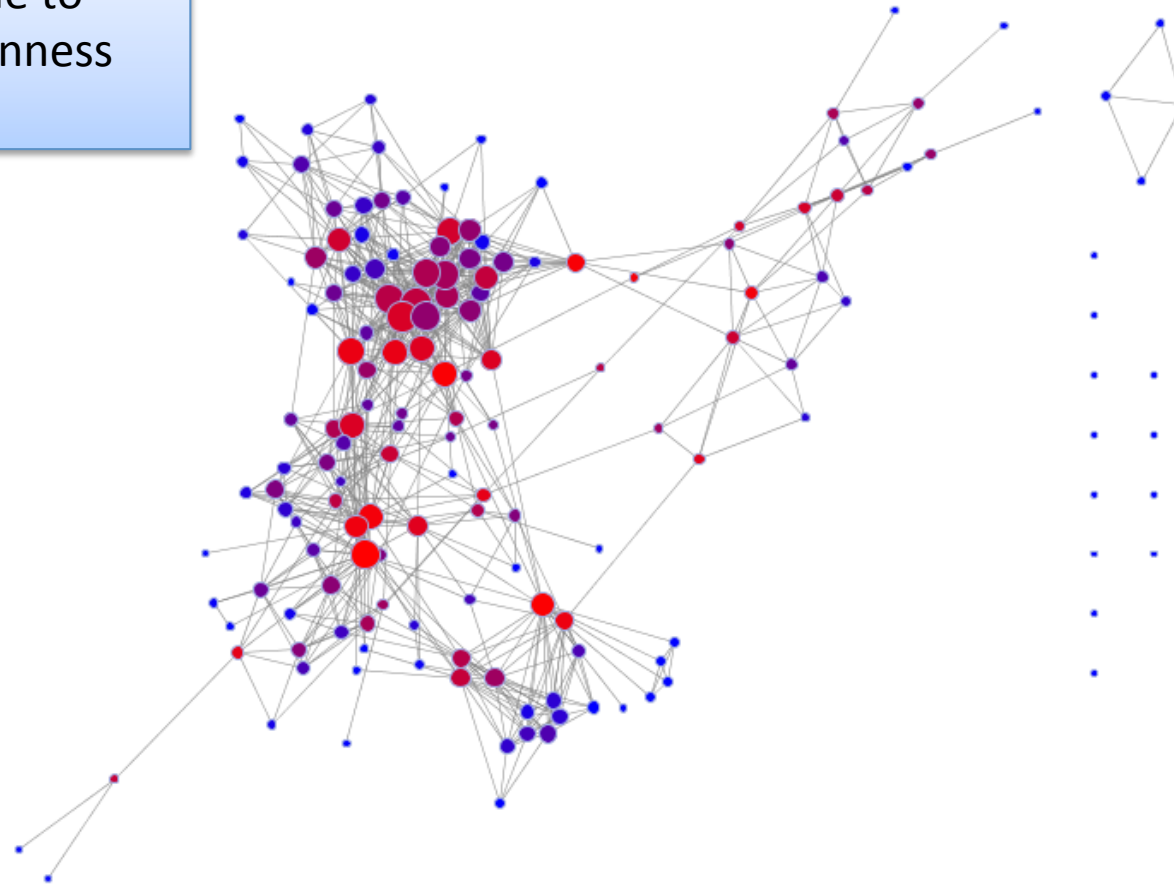
- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E), (B,D),(B,E)
- note that there are no alternative paths for these pairs to take, so C gets full credit



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)

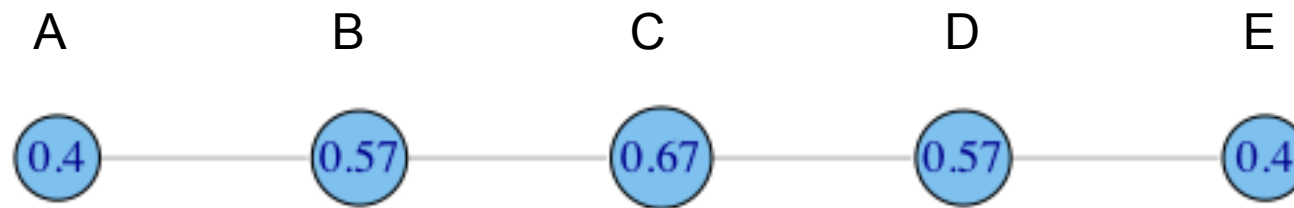


- Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

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

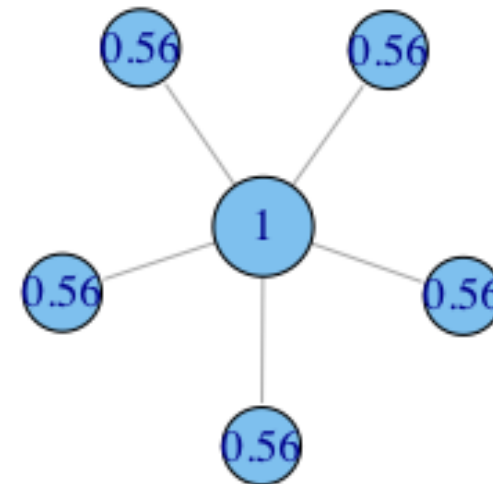
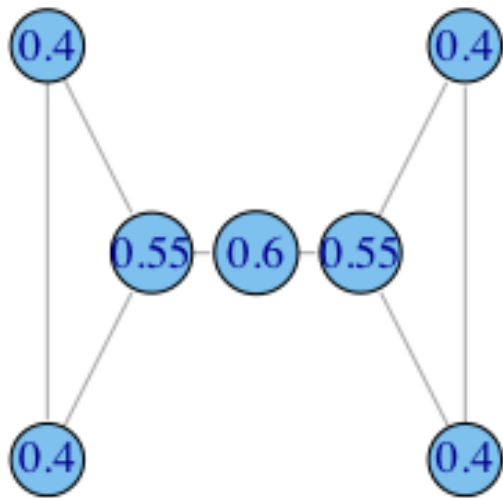
$$C'_c(i) = (C_c(i)) / (N - 1)$$

Closeness: Example



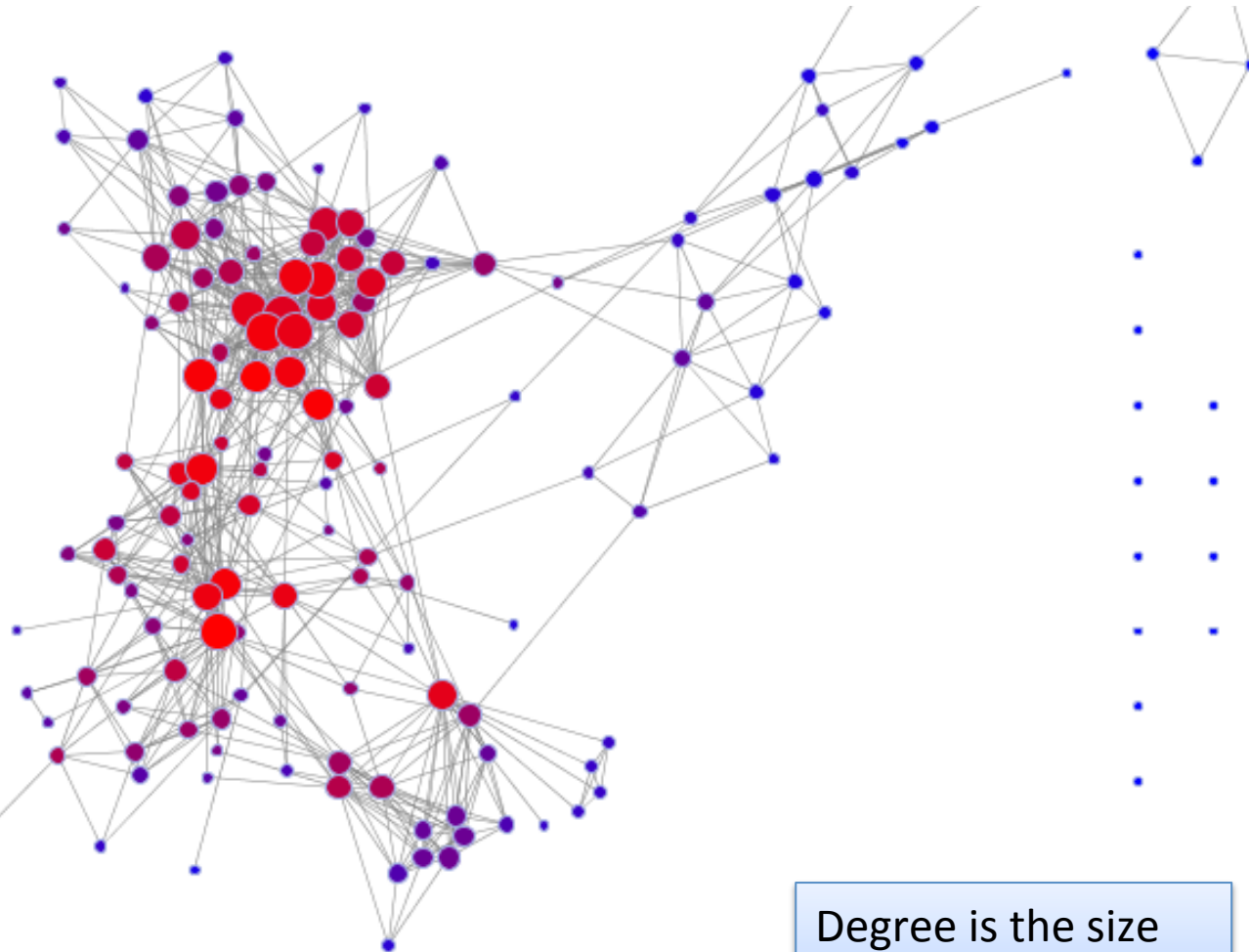
$$C'_c(A) = \left[\frac{\sum_{j=1}^N d(A, j)}{N-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$

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

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in E} a_{v,t} x_t$$

Neighbourhood of x_v

$$A\mathbf{x} = \lambda\mathbf{x}$$

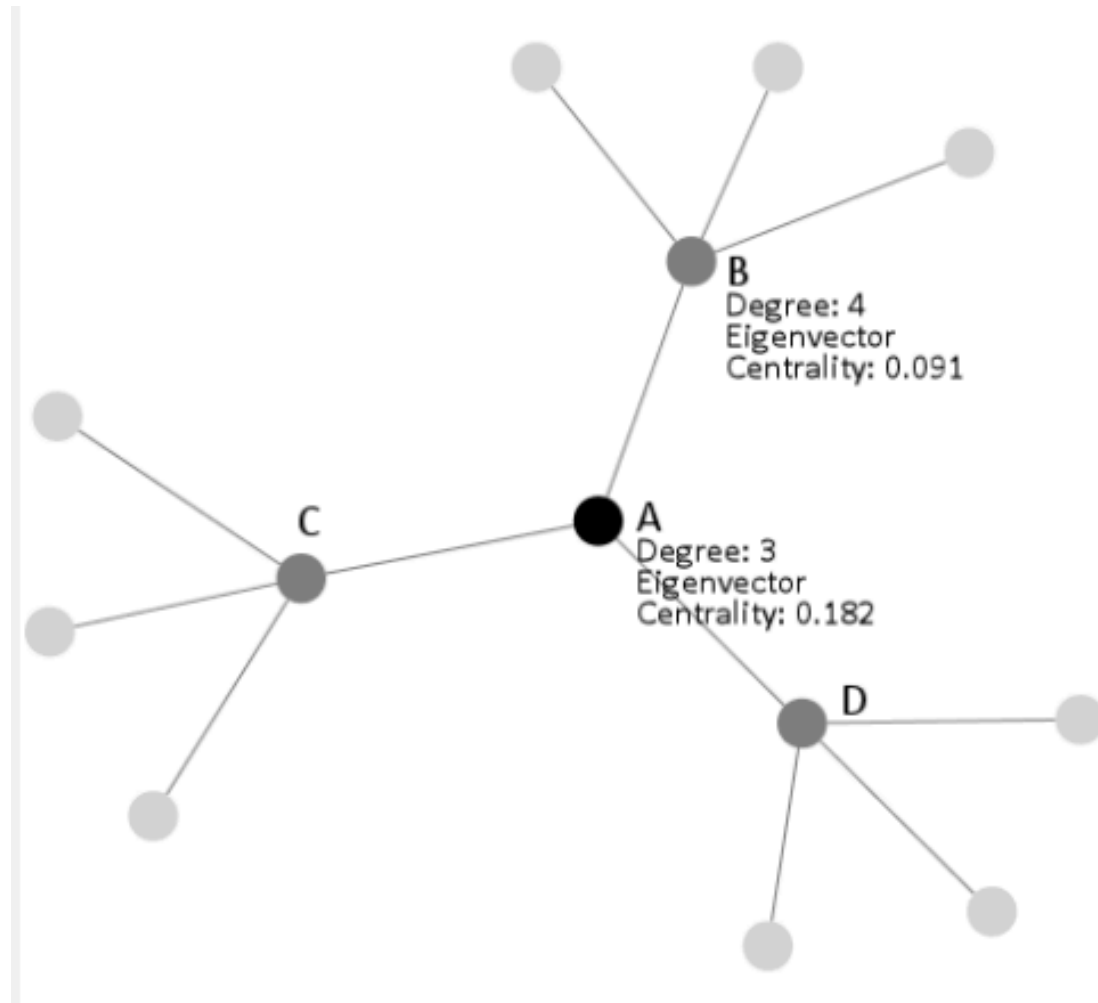
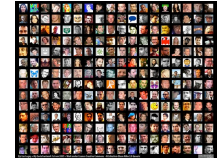
Adjacency Matrix of the graph

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: $v_i = \sum_{j \in N} a_{ij} * v_j$
- 3. Normalize v by dividing each value by the largest value
- 4. Repeat steps 2 and 3 until values of v stop changing.

Example



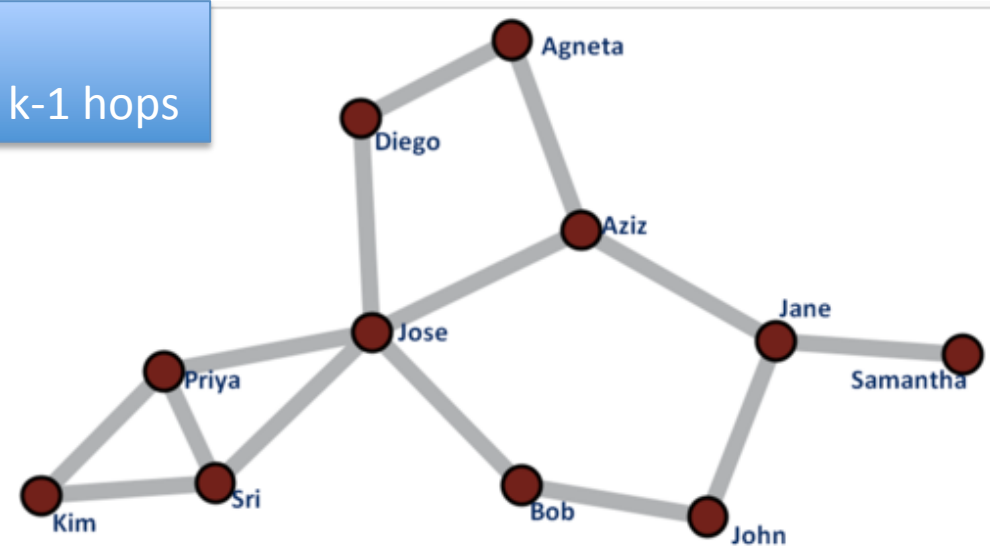
Katz Centrality



- 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



First 2008 Presidential Debate (Full Video)

LIVE

QUESTION: As President, what priorities will you have to give up to pay for the financial recovery plan?

CAMPAGN 2008 CSPAN 2

6:33 pm PT

0:31:58 / 1:36:44

HQ

★★★★★ 4,641 ratings 1,505,936 views

[Favorite](#) [Share](#) [Playlists](#) [Flag](#)

[Send Video](#) [MySpace](#) [Facebook](#) [\(more share options\)](#)

Statistics & Data

Video Responses (74)

[Post a Video Response](#)

[View All](#) - [Play All](#)

Text Comments (20,454) [Options](#)

[Post a Text Comment](#)

- [minimumgrade](#) (2 hours ago) [Reply](#) | [Spam](#) [0](#) [0](#) [0](#)
- Obama sure has fixed the economy! Thank God we got rid of capitalism!
- [kingbigbut](#) (2 hours ago) [Reply](#) | [Spam](#) [0](#) [0](#) [0](#)
- he won that is good

More From: CSPAN

Related Videos

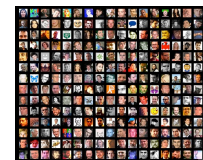
[More Popular News Videos](#)

People Tweet While They Watch



The screenshot shows the Twitter home page. At the top left is the Twitter logo. To the right are navigation links for 'Home', 'Profile', and 'Find People'. Below this is a text input field with the placeholder 'What are you doing?' and a character count of '140'. Below the input field is a text label 'Latest: mac mini = :-> - boxee about 18 hours ago' and an 'update' button. The main content area displays a list of tweets, each with a profile picture, a name with a handle, the tweet text, and a timestamp. The tweets are as follows:

- Raratequila @BigRed561** oh nah, where is dat?
less than 20 seconds ago from web in reply to BigRed561
- o_o** **tabligherooz** VPS از 12.000 تومان/کیگ هاست <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 @duecebigz26 @andrejayp @gogoquedawg @jrdafuture7 @silkybutmilky @clemfuturistic #4thqtr-** aint nuffin star.. @magicdoom04 ??
less than 20 seconds ago from UberTwitter
- lauramdempsey** don't forget, kids: throw your support behind @clrochelle tonight as she embarks on the world's first known #firsttwitterdate. :)
less than 20 seconds ago from web
- easthawaiiarts** While sending tweets, always remembering words of early founder Mrs. F. Sherrard: "Art is for EVERYBODY!!!"
www.ehcc.org #art #hawaii



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>

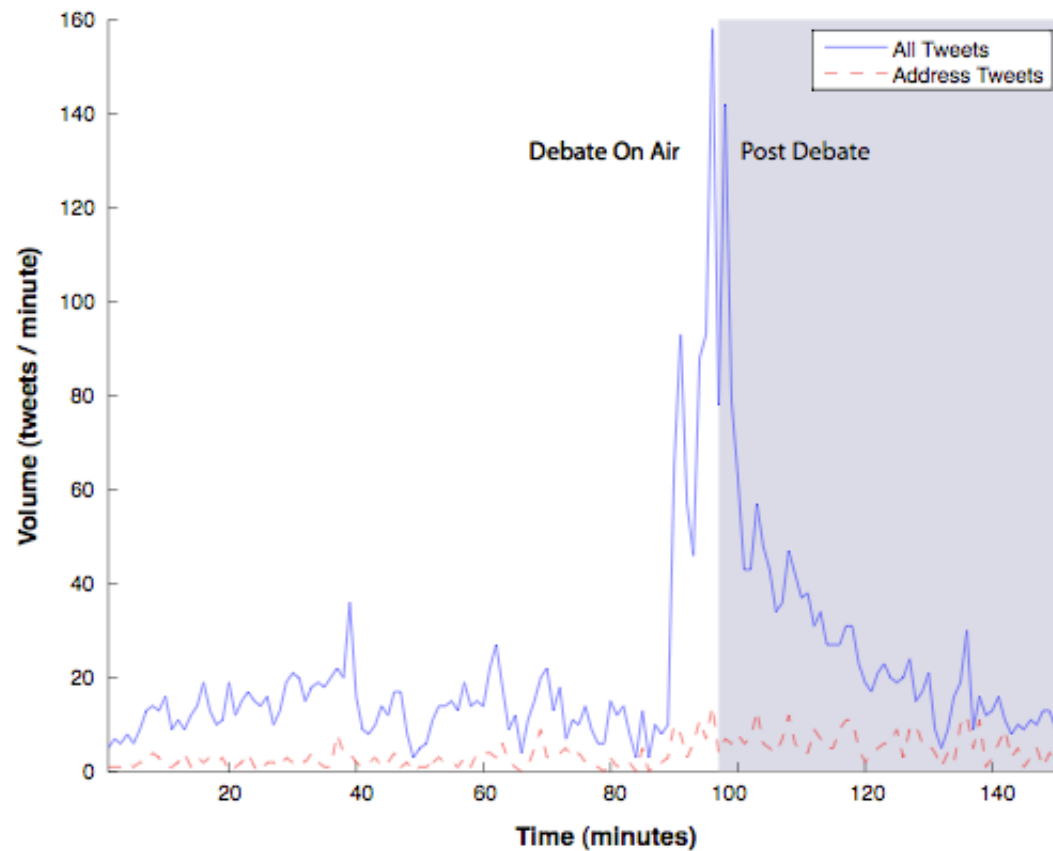
ANATOMY OF A TWEET

Tweet Crawl

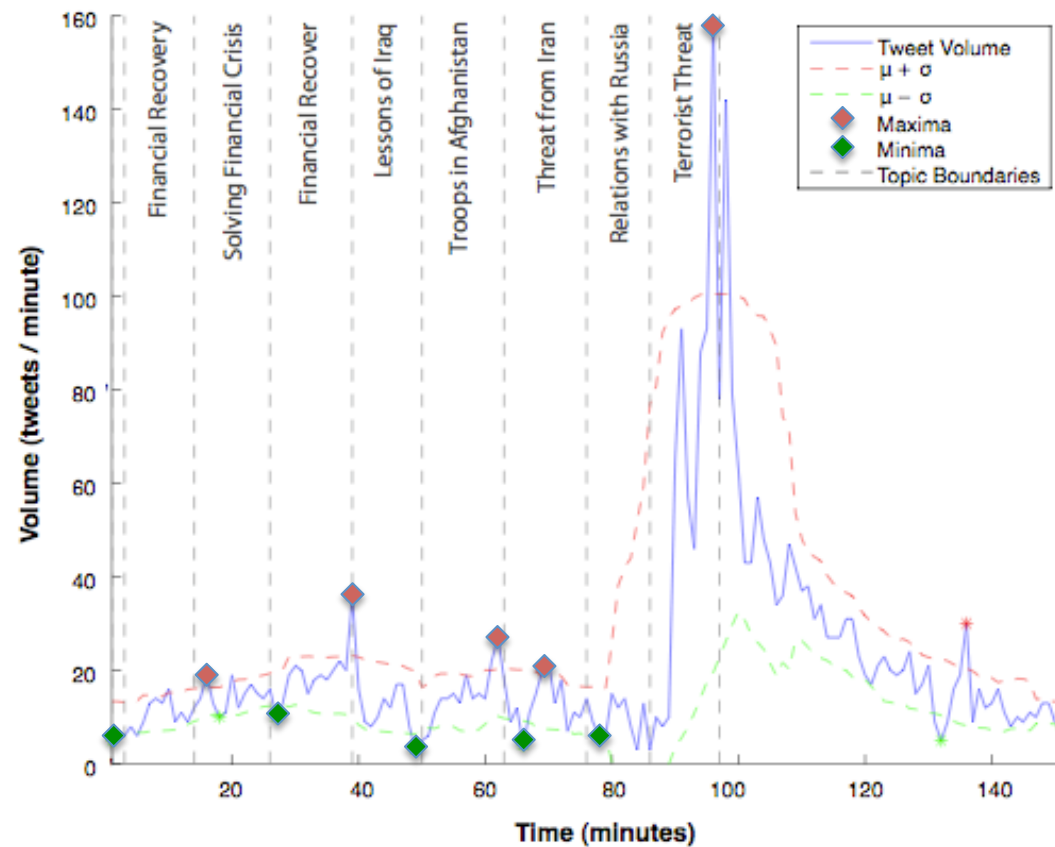


- Three hashtags: ***#current #debate08 #tweetdebate***
- 97 mins debate + 53 mins following = ***2.5 hours*** total.
- ***3,238 tweets*** from ***1,160 people***.
 - 1,824 tweets from 647 people during the debate.
 - 1,414 tweets from 738 people post debate.
- 577 @ mentions (*reciprocity!*)
 - **266** mentions **during** the debate
 - **311** afterwards.
- Low RT: **24 retweets** in total
 - 6 during
 - 18 afterwards.

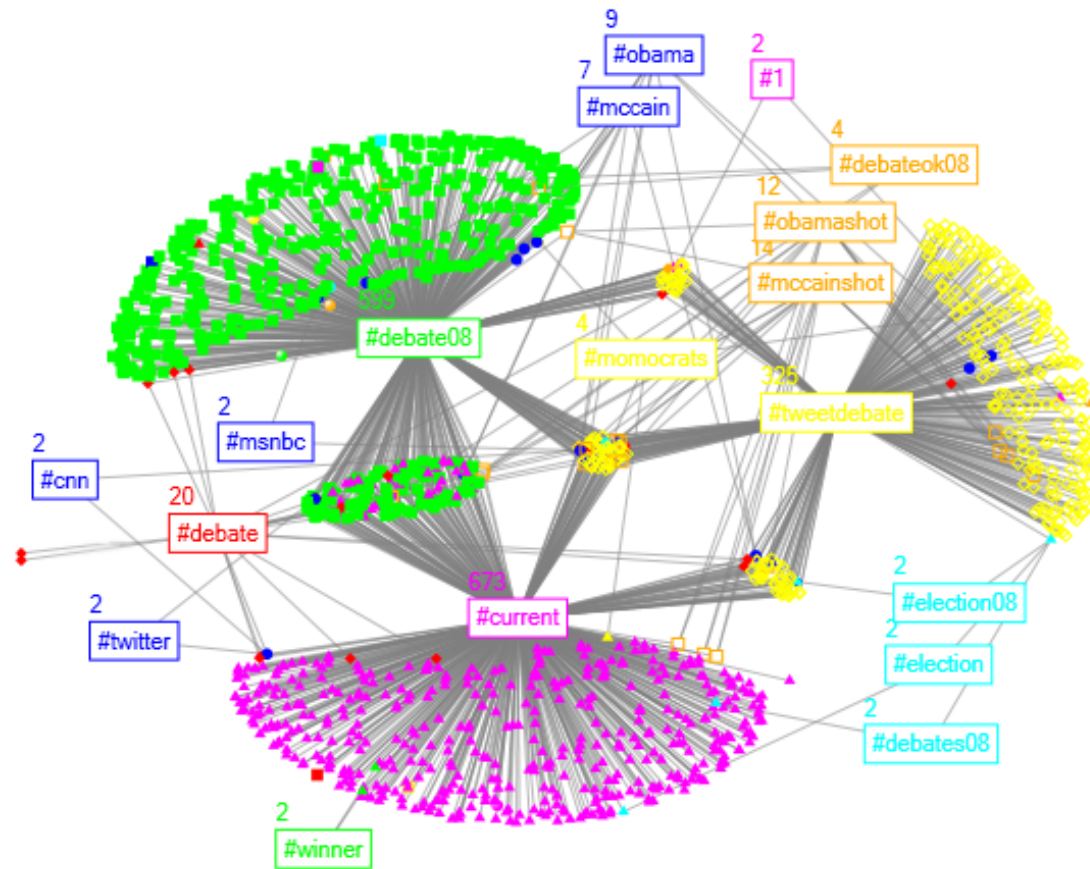
Volume of Tweets



Discussion Segments



Cluster of users and tags

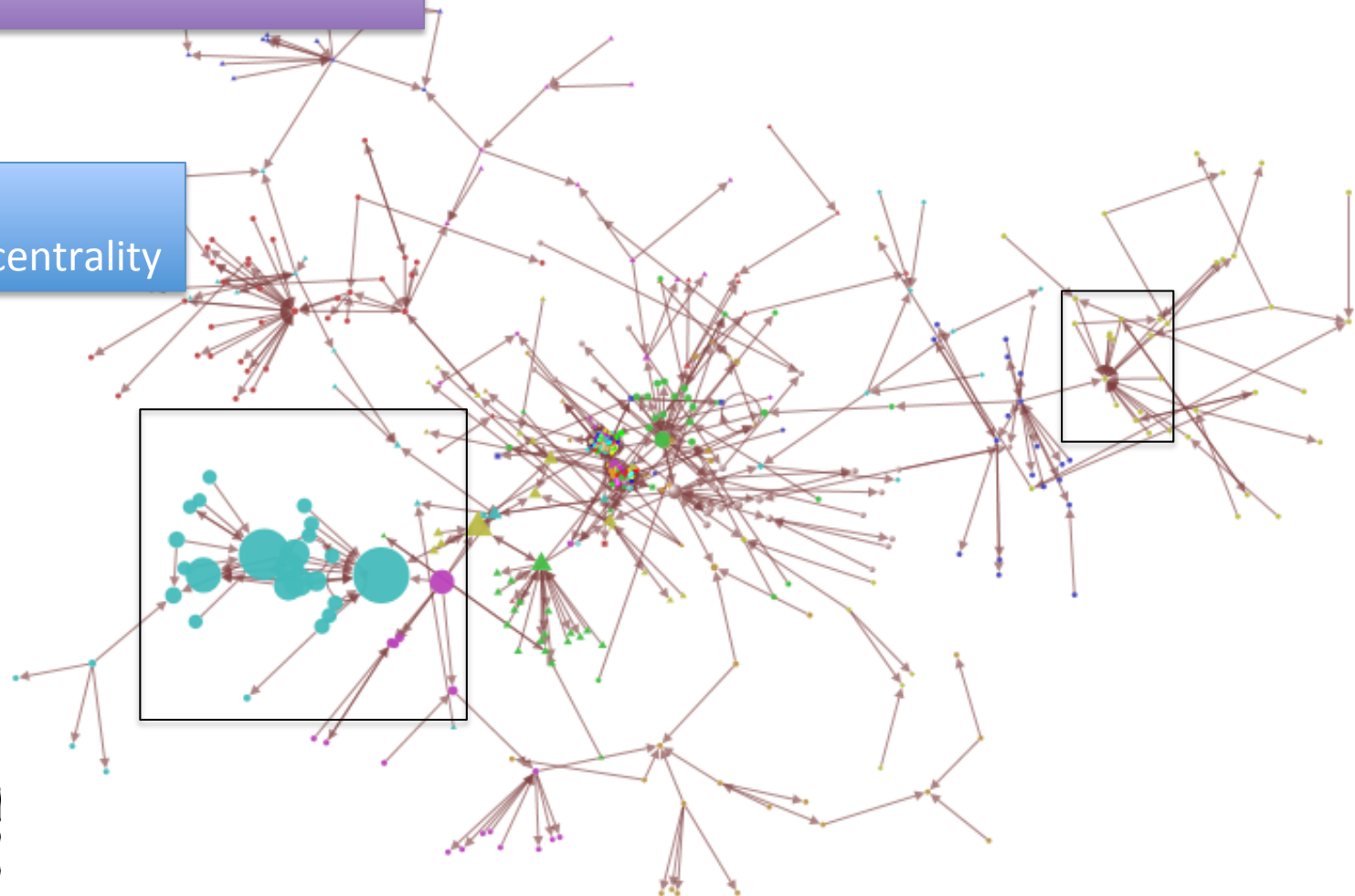


Twitter Mentions



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

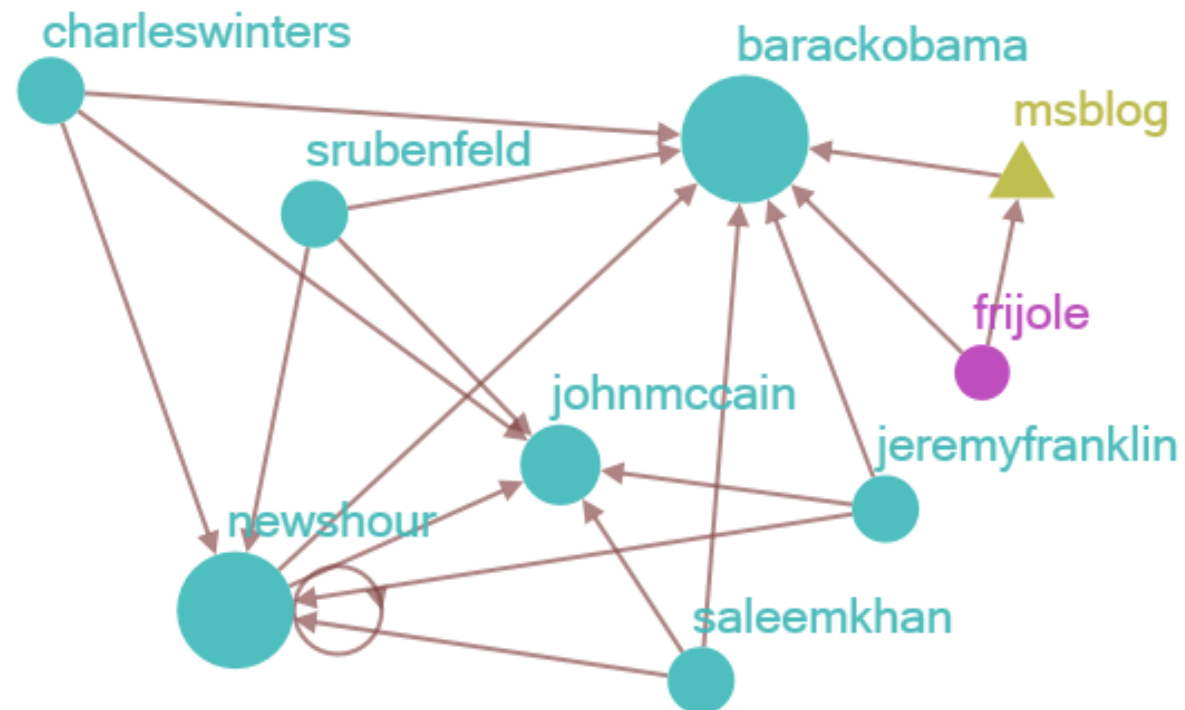
Large node size=
High eigenvector centrality



Automatic Discovery through Centrality



High Eigenvector Centrality Figures on Twitter from the First US Presidential Debate of 2008.



Centrality



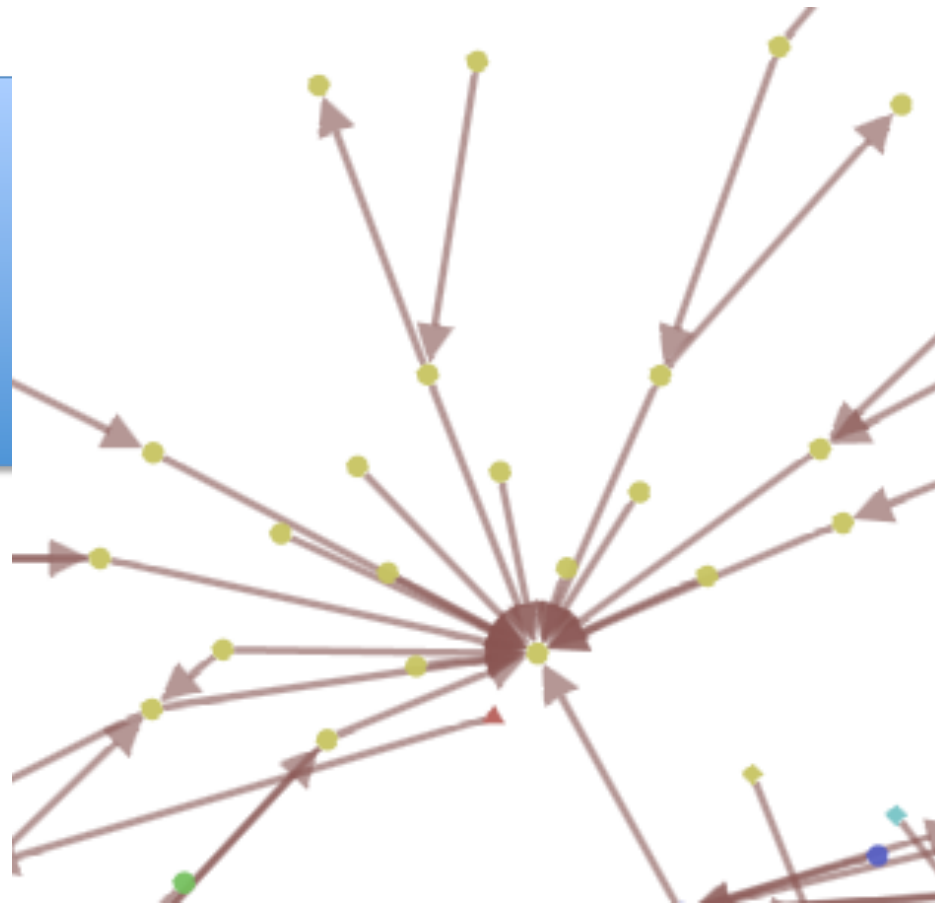
Twitter User	Eigenvector Centrality	In Degree	Out Degree
@barackobama	0.472	15	0
@newshour	0.427	11	5
@johnmccain	0.277	6	0
@charleswinters	0.223	0	3
@jeremyfranklin	0.223	0	3
@saleemkhan	0.223	0	3
@srubinfeld	0.223	0	3
@msblog	0.221	5	6
@frijole	0.175	0	7

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.

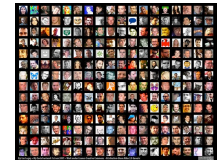


Tweets to Terms



Twitter	drinking candidates wins minute lehrer 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 condoulo 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 condescending issue oil	security experience 9/11 safe country tactics attack video pakistan bringing
Topic	Opening	Financial Recovery	Solving Financial Crisis	Financial Recovery	Lessons of Iraq	Troops in Afghanistan	Threat from Iran	Relations with Russia	Terrorist Threat
Debate	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 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	georgia international putin ukraine russia world's offshore aggression resurgent nato	restore knowledge missile safer veterans terms focused earth billions challenges

Twitter as Reaction



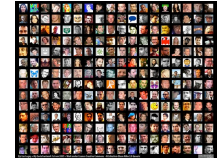
Twitter	drinking candidates wins <i>minute</i> lehrer letting tv mccains tweet moderator	kennedy plan hope hand comment moderator tie home audience republicans	cut <i>earmarks</i> compared <i>tax</i> joke bear dollars problems personal american	screen bailout energy economic power condoulo nuclear giving freeze looked	senate difference <i>winning</i> strategy iraq festooned fact war giving hand	pulling <i>pakistan</i> understand drinking republican strategy sounds coming telling story	lousy democracies lol government story wars league times Iran -5	<i>georgia</i> video blog league <i>russia</i> ha let condescending issue oil	security experience 9/11 <i>safe</i> country tactics attack video pakistan bringing
Topic	Opening	Financial Recovery	Solving Financial Crisis	Financial Recovery	Lessons of Iraq	Troops in Afghanistan	Threat from Iran	Relations with Russia	Terrorist Threat
Debate	presidential debates <i>minutes</i> eisenhower financial direct policy news mississippi university	street greed main how's house package accountable wall rewarded crisis	\$18 requests <i>earmarks</i> gateway loopholes \$5,000 pork-barrel employer business <i>tax</i>	programs medicare cost eliminate \$700 hard decisions agency rescue tom	funding <i>winning</i> leave timetable violence lessons baghdad surge succeed started	border taliban prepared supported muddle u.s bombing <i>pakistan</i> qaeda army	henry kissinger contacts preparation legitimize table <i>iranians</i> sanctions ahmadinejad precondition	<i>georgia</i> international putin ukraine <i>russia</i> world's offshore aggression resurgent nato	restore knowledge missile <i>safer</i> veterans terms 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

References



- Kleinberg's book: Chapter 3.
- **Ack: L. Adamic's slides**
- M. Newmann. **Networks**. Oxford University Press. April 2010.
- David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill. 2009. *Tweet the debates: understanding community annotation of uncollected sources*. In Proceedings of the first SIGMM workshop on Social media (WSM '09). ACM, New York, NY, USA