Social Media Analytics
Social Media and its impact

- Social networking, blogging, and online forums have turned the Web into a vast repository of comments on many topics, generating a potential source of information for:
  - social science research
  - market and politics forecasts
  - syndromic surveillance
  - information warfare
  - new opportunities for media communication
How big is “social media”? 

• **72% of Internet users** are part of at least one social network, which translates to **940 million** users worldwide.
Impact

• “Social Media Marketing Spending to Hit $3.1 Billion by 2014 (faster than any other form of online marketing)”
  Forrester, 15 September 2009

• “Only 14% of people trust advertisers yet 78% of consumers trust peer recommendations.”
  Erik Qualman’s book “Socialnomics”, 2009
Impact of Social Media
Impact of Social Media on Products

• General Motors cancels ‘Hideous’ Buick SUV after “Would-Be Customers” on Twitter!

• ONE week after announcing a new Buick SUV Christopher Barger, GM’s spokesman for social media said: “The decision was based on customers’ input - face-to-face, blogs and tweets. No matter how they expressed it “they just didn’t like it.”

Impact of Social Media on Products

- Del Monte created a new “hot-selling” dog food snack in 6 weeks
- Used a social community to source for creative ideas (crowdsourcing) and create a new product
- Demonstrates the potential power of social media marketing to influence product sales

http://www.youtube.com/watch?v=yP_3bpCPZaQ
Impact of Social Media on Organisations

Nestlé vs Greenpeace Palm Oil from Destroyed Rainforest

• Nestlé, maker of Kit Kat, uses palm oil from companies that are trashing Indonesian rainforests, threatening the livelihoods of local people and pushing orang-utans towards extinction.
• Nestlé persuaded YouTube to remove Greenpeace’s video
• Storm ensued on Nestlé Facebook page
• TWO months later, Nestlé announced a “zero deforestation” policy in partnership with The Forest Trust (TFT)

“Social media: as you can see we're learning as we go. Thanks for the comments.”
Impact of Social Media on Government

25th Jan 2011 Egypt Blocked Twitter and Facebook!

Egyptian protesters have openly thanked social media's role in the revolution against the country's ruling government.
Social Networks also impact on media and communication

companies, politicians, public body interest in a campaign

Traditional method to reach audience
The social media revolution

companies, politicians, public body interest in a campaign

press releases, interviews

MEDIA

twitter, social media, web sites

AUDIENCE

News, talk shows, entertainment
The ICT revolution and new media

• new media, websites, social media and twitter can be used by audiences, but also by stakeholders and the media (#partitodemocratico, #ballarò,..)

• Audience members can publish their opinions in the new media but are also influenced themselves by opinions of others in the new media

Marco Hanaman @Fare2013Hanaman
«La nostra risposta a Berlusconi è semplice: per noi i voti inutili e dannosi sono quelli a lui, al Pd e a Monti» #Fare2013 #OGiannino
Social Networks Measures

• **Surface Measures**: Based on some properties of specific nodes

• **Graph-based measures**: Based on the graph-structure of the network
Measuring properties of individual nodes (users, web pages..)

Key measurement goals

- Reach
- Quantitative & Qualitative
- Buzz
- Sentiment
- Influence
1. Reach

- Reach
  - Size of your audience
  - How many saw your message
    - E.g. Twitter followers
Reach: Facebook Insights

Monitor and measure your fans, likes,

How do I see Insights for my website?

To see Insights for your website, you must first claim your domain by associating it with a Facebook page or application that you manage, or with your Facebook account. To do this, click on the green "Insights for your Domain" link from the Insights Dashboard. The window that appears will provide a meta tag that must be added to the root of your web page. Once that is done, type in your domain address into the text box and select the account to link it with. Once checked, your claimed domain will appear on the left side navigation bar under the "domains" section.
Reach: Google Analytics

http://www.google.com/analytics/

With Google Analytics tool, you can monitor accesses on your web page. Drill down into site traffic data including source, and region. View sparklines for page views, bounce rates and more.
24.556 persone hanno visitato questo sito

- Visite: 60.545
- Visitatori unici: 24.556
- Visualizzazioni di pagina: 184.624
- Pagine/visita: 3,05
- Durata media visita: 00:02:50
- Frequenza di rimbalzo: 40,35%
- % nuove visite: 37,39%

### Dati demografici

<table>
<thead>
<tr>
<th>Lingua</th>
<th>Visite</th>
<th>% Visite</th>
</tr>
</thead>
<tbody>
<tr>
<td>it</td>
<td>28.418</td>
<td>46,93%</td>
</tr>
<tr>
<td>it-it</td>
<td>24.352</td>
<td>40,22%</td>
</tr>
<tr>
<td>en-us</td>
<td>5.045</td>
<td>8,33%</td>
</tr>
<tr>
<td>en</td>
<td>1.844</td>
<td>3,05%</td>
</tr>
</tbody>
</table>
Reach: Twitter Profile statistics

- Track the number of followers, mentions, lists..
- Do more by comparing keywords over time and Twitter sentiment.
TweetStats
In ur Tweets, Graphin’ Your Stats!

Tweets per Day

Graph your Twitter Stats including
Tweets per hour
Tweets per month
Tweet timeline
Reply statistics
In use by nearly 1,000,000 Twitter-folk!

Enter your Twitter username

Graph My Tweets!

Looking for simple Social Media Monitoring?
Simply Measured
Brought to you by @dacort!

Refresh your stats - stats get updated when you come back after 8 hours and enter your username.
Measures of Social Reach

• **Social reach**: #total followers across all social platforms
• **Growth**: month-over-month social reach growth
• **Engagement**=

\[
\frac{\# \text{ Likes} + \# \text{ Shares} + \# \text{ Retweets} + \# \text{ blog comments}}{\# \text{ of published posts or pieces of content}}
\]
2. Buzz

- **Social Buzz** is the “amplification” of a topic/message through social media: what are people saying about you, where are they saying it, how are they saying it
  - 2 types:
    - Conversation Focus (@RP, reply) vs. Content Focus (#topic)
- Mining **motivations** (text), in addition to data, as a way to understand an audience (either customers, voters, patients, or addressee of a campaign), is an entirely new approach to social analysis (e.g. opinions on #topic).
Buzz Metric example (1)

• Prime Visibility™ showcased their buzz metrics tool by doing their own survey around the online social media elements related to the two U.S. Presidential candidates in 2012.

• Based on three measures:
  – bookmarking,
  – social networking,
  – social knowledge
Buzz Metric example (2)

- **Bookmarking**: Social bookmarking relates to social media websites such as Digg, Del.icio.us, and Reddit. Users submit links to these websites that are of interest to them and other users vote on particular submissions of interest in order to increase their popularity.

- **Social Networking**: Social Networking refers to communities such as MySpace, Facebook, and Friendster. **Account creation, total friends, mentioning of company on a particular page**, and other important factors to determine the total buzz.

- **Social Knowledge**: Social Knowledge refers to informational based websites such as “Yahoo! Answers” and “Wikipedia”. Buzz is calculated differently on each of these websites.

forbes.com

If your head hurts every time you have to think about how to generically refer to a person based on the presence or absence of the Y chromosome, you're in luck. For those not immersed in the study of gender semantics, here is a handy guide for which term to use when.
Haverford College Commencement Speaker Calls Students Arrogant, Immature

In a surprising move, a commencement speaker at Haverford College on Sunday used the celebratory occasion to deliver a sharp rebuke to students who had mounted a campaign against another speaker who had been scheduled to appear but withdrew amid the controversy.

In Letter To Obama, Cisco CEO Complains About NSA Allegations

Warning of an erosion of confidence in the products of the U.S. technology industry, John Chambers, the CEO of networking giant Cisco Systems, has asked President Obama to intervene to curtail the surveillance activities of the National Security Agency.

Preparing For The Worst

Climate change will force us to abandon coastal cities. We better start preparing right now.

That Zero Hedge Article On Unsold Cars Is Bullshit

I would think that most intelligent people would read this and obviously see the flaws but, alas, I've had enough emails about it this morning that I feel the need to refute it. And since it's a [bad] copy of something I wrote, I have an extra responsibility to kill this misunderstanding before it has a chance to spread.

Why 1.5 Million People Are On A Dating Site Called Farmers Only

Dating sites like OKCupid and Match.com are great if you live in a city, where you go on dates with matches and never seen them again if it doesn't work out. But what if you live in a town with, say, 8,000 people?

These Maps Reveal How Slavery Expanded Across The United States

As the hunger for more farmland stretched west, so too did the demand for enslaved labor.
Social Knowledge

Yahoo! Answers

Resolved Question

Honestly, what's a democrat?

Arthur Reeves
2 years ago

Best Answer - Chosen by Asker

I don't know much about politics and I don't know what the democratic party of the past has been like, but I can tell you what I know about it now:
Democrats are into big government:
They want the government to take care of people with welfare, medicaid, unemployment, etc.
Democrats care more about social issues like gay rights, abortion, etc.

That is all I know.

There is a lot more to research before you support one party or the other and I suggest you do.

Source(s):
A democrat because of my social liberalism.

2 years ago

1 person rated this as good

Asker's Rating: *****
Thank you Andi
# Buzz Metric example (3)

<table>
<thead>
<tr>
<th>Website</th>
<th>Bookmarking</th>
<th>Social Networking</th>
<th>Social Knowledge</th>
<th>Total Buzz</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNC.org</td>
<td>272</td>
<td>142</td>
<td>314</td>
<td>728</td>
</tr>
<tr>
<td>GOP.org</td>
<td>28</td>
<td>56</td>
<td>85</td>
<td>169</td>
</tr>
<tr>
<td>Democrats.org</td>
<td>2,963</td>
<td>1,880</td>
<td>1,682</td>
<td>6,525</td>
</tr>
<tr>
<td>DemConvention.com</td>
<td>365</td>
<td>338</td>
<td>230</td>
<td>933</td>
</tr>
</tbody>
</table>

RNC and GOP = Republican official websites
Democrats and DemConvention = same for Democrats
# Buzz Metric example (4)

## Additional (simpler) measures

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Google</th>
<th>Yahoo</th>
<th>MSN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republican</td>
<td>84,300,000</td>
<td>453,000,000</td>
<td>20,700,000</td>
<td>558,000,000</td>
</tr>
<tr>
<td>Republican Party</td>
<td>14,500,000</td>
<td>209,000,000</td>
<td>19,500,000</td>
<td>243,000,000</td>
</tr>
<tr>
<td>Democrat</td>
<td>50,600,000</td>
<td>201,000,000</td>
<td>9,560,000</td>
<td>261,160,000</td>
</tr>
<tr>
<td>Democratic Party</td>
<td>25,900,000</td>
<td>254,000,000</td>
<td>24,400,000</td>
<td>304,300,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Website</th>
<th>Traffic Rank</th>
<th>Links</th>
<th>Page Views</th>
<th>Competitive Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNC.org</td>
<td>161,445</td>
<td>225,532</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>GOP.org</td>
<td>1,036,860</td>
<td>14,288</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Democrats.org</td>
<td>31,738</td>
<td>1,109,598</td>
<td>2.2</td>
<td>1</td>
</tr>
<tr>
<td>DemConvention.com</td>
<td>23,993</td>
<td>154,377</td>
<td>1.7</td>
<td>3</td>
</tr>
</tbody>
</table>
Google Trends
3. Influence

- Your message is valuable when it is repeated and/or commented
  - High probability of others referencing & reproducing what you say
  - E.g. Twitter: reply/mention (@xxx) & retweet (RT)
Twitter as a mean to disseminate information

• Its primary function is not as a social network but perhaps to spread news (including personal news) or other information.
• An unusual feature of Twitter is re-tweeting: forwarding a tweet by posting it again
  Hmmm pretty good incentive.. RT @RT_com: US high school allows Muslims time for prayer if they earn good grades http://on.rt.com/kka96w
• If re-tweeted, a tweet can expect to reach an average of 1000 users (Kwak et al.)
• Another communicational feature of Twitter is the hashtag: a metatag beginning with # that is designed to help others find a post:
  Kelvin Thompson @KelvinThompson5 3 min I missed these ugly 90s fashions. Thank you for this bountiful harvest. #sailormoon pic.twitter.com/v0RlZBs9Q8
Influential Analysis (Twitter)

- **Retweet** and **Reply** features of Twitter is used to enable real-time study

\[
\text{Influential Index} = \frac{n(\text{Reply}) + n(\text{Retweet})}{n(\text{Tweet})}
\]

For example, a tweet:
Verizon will launch iPhone 4 on 10 Feb sent by user ABC

Retweet (think of it as forwarding)
**RT @ABC** Verizon will launch iPhone 4 on 10 Feb

Reply
**@ABC** thanks... I will be there to get one
Influential Analysis: Amplification

• On Twitter:
  – Amplification = # of Retweets Per Tweet

• On Facebook, Google Plus:
  – Amplification = # of Shares Per Post

• On a blog, YouTube:
  – Amplification = # of Share Clicks Per Post (or Video)
Influential Analysis : Applause

• On Twitter:
  – Applause Rate = # of Favorite Clicks Per Post

• On Facebook:
  – Applause Rate = # of Likes Per Post

• On Google Plus:
  – Applause Rate = # of +1s Per Post

• On a Blog, YouTube:
  – Applause Rate = # of +1s and Likes Per Post (or video)
Graph-based measures of social influence

• Previously surveyed measures of influence, such as buzz, applause etc. are based on surface metrics (e.g. number of retweets, etc): graph-based measures go more in-depth.
• Objective: model the social network as a graph
• Use graph-based methods/algorithms to identify “relevant players” in the network
  – Relevant players = more influential, according to some criterion
• Use graph-based methods to identify communities (community detection)
• Use graph-based methods to analyze the “spread” of information
Social Networks Measures

- **Single-node Measures**: Based on some properties of specific nodes
- **Graph-based measures**: Based on the graph-structure of the network
Graph-based measures of social influence

• **Use graph-based methods/algorithms to identify “relevant players” in the network**
  – Relevant players = more influential, according to some criterion

• **Use graph-based methods to identify global network properties and communities (community detection)**

• **Use graph-based methods to analyze the “spread” of information**
NODE = “actor, vertices, points” i.e. the social entity who participates in a certain network
EDGE = “connection, edges, arcs, lines, ties” is defined by some type of relationship between these actors (e.g. friendship, reply/re-tweet, partnership between connected companies..)
SN = graph

• A network can then be represented as a graph data structure
• We can apply a variety of measures and analysis to the graph representing a given SN
• Edges in a SN can be directed or undirected (e.g. friendship, co-authorship are usually undirected, emails are directed)
Social Network

Basic Data Structures

From graphs to matrices

Undirected, binary (0,1)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Directed, binary

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Social Network
Social Network

**Basic Data Structures**

From matrices to lists

\[
\begin{array}{cccccc}
        & a & b & c & d & e \\
 a &  1 &   &   &   &   \\
b &   &  1 &   &   &   \\
c &   &   &  1 &   &   \\
d &   &   &   &  1 &   \\
e &   &   &   &   &  1 \\
\end{array}
\]

**Adjacency List**

\[
\begin{align*}
 a & \rightarrow b \\
b & \rightarrow a, c \\
c & \rightarrow b, d, e \\
d & \rightarrow c \\
e & \rightarrow c, d \\
\end{align*}
\]

**Arc List**

\[
\begin{align*}
 a & \rightarrow b \\
b & \rightarrow a \\
b & \rightarrow c \\
c & \rightarrow b, d, e \\
d & \rightarrow c \\
e & \rightarrow c \\
e & \rightarrow d \\
\end{align*}
\]
In general, a relation can be:
Binary or Valued
Directed or Undirected

Social Network as a graph
Example of directed, valued: Sentiment relations among parties during a political campaign. Color: positive (green) negative (red). Intensity (thikness of edges): related to number of mutual references
Graph-based measures of social influence: key players

Key players

• Using graph theory, we can identify key players in a social network
• Key players are nodes (or actors, or vertexes) with some measurable connectivity property
• Two important concepts in a network are the ideas of centrality and prestige of an actor.
• Centrality more suited for undirected, prestige for directed
Measuring Networks: Centrality

_Centrality_ refers to (one dimension of) _location_, identifying _where_ an actor resides in a network. Mostly used for _undirected_ networks.

• For example, we can compare actors at the edge of the network to actors at the center.

• In general, this is a way to formalize intuitive notions about the distinction between _insiders and outsiders_.
Measuring Networks: **Centrality**

Conceptually, centrality is fairly straightforward: we want to identify **which nodes are in the ‘center’ of the network**. In practice, identifying exactly what we mean by ‘center’ is somewhat complicated.

Three standard centrality measures capture a wide range of “importance” in a network:

- Degree
- Closeness
- Betweenness
Measuring Networks: **Centrality**

1. **Centrality Degree**

The most intuitive notion of centrality focuses on *degree*. Degree is the number of ties, and the actor with the most ties is the most important:

\[ C_D = d(n_i) = X_{i+} = \sum_j X_{ij} \]
Measuring Networks: Closeness Centrality

A second measure of centrality is **closeness centrality**. An actor is considered important if he/she is relatively close to all other actors. Closeness is based on the inverse of the distance of each actor to every other actor in the network.

**Closeness Centrality:**

\[
C_c(n_i) = \left[ \sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}
\]

Normalized Closeness Centrality

\[
C'_c(n_i) = (C_c(n_i))(g - 1)
\]
Measuring Networks: **Centrality**

### Closeness Centrality in the examples

<table>
<thead>
<tr>
<th>Distance</th>
<th>Closeness</th>
<th>normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 1 1 1 1 1 1</td>
<td>.143</td>
<td>1.00</td>
</tr>
<tr>
<td>1 0 2 2 2 2 2 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 0 2 2 2 2 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 2 0 2 2 2 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 2 2 0 2 2 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 2 2 2 0 2 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 2 2 2 2 0 2</td>
<td>.077</td>
<td>.538</td>
</tr>
<tr>
<td>1 2 2 2 2 2 2 0</td>
<td>.077</td>
<td>.538</td>
</tr>
</tbody>
</table>

\[
C_c(n_i) = \left[ \sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}
\]

### Distance  Closeness normalized

<table>
<thead>
<tr>
<th>Distance</th>
<th>Closeness</th>
<th>normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 4 3 2 1</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>1 0 1 2 3 4 3 2</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>2 1 0 1 2 3 4 3</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>3 2 1 0 1 2 3 4</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>4 3 2 1 0 1 2 3</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>4 4 3 2 1 0 1 2</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>3 4 4 3 2 1 0 1</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>2 3 4 4 3 2 1 0</td>
<td>.050</td>
<td>.400</td>
</tr>
<tr>
<td>1 2 3 4 4 3 2 1 0</td>
<td>.050</td>
<td>.400</td>
</tr>
</tbody>
</table>
Measuring Networks: **Centrality**

Closeness Centrality in the examples

<table>
<thead>
<tr>
<th>Distance</th>
<th>Closeness</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6</td>
<td>.048</td>
<td>.286</td>
</tr>
<tr>
<td>1 0 1 2 3 4 5</td>
<td>.063</td>
<td>.375</td>
</tr>
<tr>
<td>2 1 0 1 2 3 4</td>
<td>.077</td>
<td>.462</td>
</tr>
<tr>
<td>3 2 1 0 1 2 3</td>
<td>.083</td>
<td>.500</td>
</tr>
<tr>
<td>4 3 2 1 0 1 2</td>
<td>.077</td>
<td>.462</td>
</tr>
<tr>
<td>5 4 3 2 1 0 1</td>
<td>.063</td>
<td>.375</td>
</tr>
<tr>
<td>6 5 4 3 2 1 0</td>
<td>.048</td>
<td>.286</td>
</tr>
</tbody>
</table>

\[
C_c(n_i) = \left[ \sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}
\]
Measuring Networks: **Betweenness Centrality**

Model based on communication flow: A person who lies on communication paths can control communication flow, and is thus important.

Betweenness centrality counts the number of *geodesic* paths between *i* and *k* that actor *j* resides on. Geodesics are defined as the shortest path between points.
Measuring Networks: **Betweenness Centrality**

\[ C_B(n_i) = \sum_{j<k} g_{jk}(n_i) / g_{jk} \]

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

Usually normalized by:

\[ C'_B(n_i) = C_B(n_i) / [(g - 1)(g - 2) / 2] \]
Measuring Networks: Betweenness Centrality

\[ C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}} \]
Measuring Networks: **Betweenness Centrality**

Actors that appear very different when seen individually, are comparable in the global network.

(Node size proportional to betweenness centrality)
Measuring Networks: **Information Centrality**

It is quite likely that information can flow through paths *other* than the geodesic. The **Information Centrality** score uses *all paths* in the network, and weights them based on their length.
Measuring Networks: **Prestige**

- The term prestige is used for **directed networks** since for this measure the direction is an important property of the relation.

- In this case we can define two different types of prestige:
  - one for outgoing arcs (measures of influence),
  - one for incoming arcs (measures of support).

- Examples:
  - An actor has high influence, if he/she gives hints to several other actors (e.g. in Yahoo! Answers).
  - An actor has high support, if a lot of people vote for him/her (many “likes”)

Measures of prestige

- Influence and support
- Influence domain
- Hubs and authorities
- Brockers
Measuring prestige: influence and support

- **Influence and support**: According to the direction/meaning of a relation, in and outdegree represent support or influence. (e.g., likes, votes for, . . .).

\[
\text{InDegree}(x) = \# \text{ incoming edges}(x)
\]

\[
\text{InDegree}^N(x) = \frac{\# \text{ incoming edges}(x)}{\max_{y \in \text{network}} (\text{InDegree}^N(y))}
\]
Measuring prestige: influence domain

- **Influence domain**: The influence domain of an actor (node) in a directed network is the number (or proportion) of all other nodes which are connected by a path to this node.

All other actors are in influence domain of actor 1: Prest(1)=10/10=1.
Limits of Influence domain

- Influence domain has an important limitation: all the nodes contribute equally to influence.

- Choices by actors 2, 3, and 7 are more important to person 1 than indirect choices by 4, 5, 6, and 8. Individuals 9 and 10 contribute even less to the prestige of 1.
Measuring prestige: Hubs and Authorities, Page Rank

- Hubness is a good measure of influence
- Authority is a good measure of support
- **Kleinberg’s algorithm** (HITS) to compute authority and hubness degree of nodes, same as for link analysis
- **Page Rank** is a good measure of support
- HITS, Page Rank: see previous lessons

\[ a_p = \text{the sum of } h_i \text{ for all nodes } i \text{ pointing to } p \]

\[ h_p = \text{the sum of } a_i \text{ for all nodes } i \text{ pointed to by } p \]
Example

If Mrs. Green is the boss, employees referring directly to her are more important.
High-level scheme

• Hubs and authorities can be computed in sub-communities, i.e. on parts of a large social network graph, or on the entire graph.

• Initial step (create a sub-graph):
  1. Extract from the graph a base set of pages that could be good hubs or authorities (e.g. with many incoming or outgoing links).
  2. From these, identify a small set of top hub and authority pages;
 → using the iterative HITS algorithm.
Measuring prestige: **Brokers** (bridges)

- Network brokerage: Linkages between different groups/communities (for undirected graphs we used betweenness)

---

**Local bridging ties:**
Brokerage of disconnected groups:
“spanning structural holes”

**Local cut points:**
Brokerage through overlapping group membership
Measuring prestige: Brockers

Finding Brockers

- Brockers are "intermediaries", people that create relationships between communities
- As for graph representation, a broker is a node that, if removed from the graph, reduces graph connectivity. For example, it causes the creation of disconnected components (Jenny, Jack and John in the graph)
- Brockers are also called key separators
Example of key separator

Algorithms to identify brokers are all based on some measure of the graph connectivity.
Algorithm for KPP_NEG (Keblady 2010)

• Let $C_G$ be a measure of graph connectivity (e.g. reachability, see later) for a graph $G$; $V$ is the set of actors in $G$ (nodes, vertexes).

• Algorithm KPP-neg (greedy algorithm)

Compute proposed measure of entire graph, $C_G$

$\forall \, v_i \, \in \, V, \, \text{remove} \, v_i \, \text{from} \, \text{the graph}$

Compute $C_{G-\{v_i\}}$ for the graph $G - \{v_i\}$.

Rank the nodes based on $|C_G - C_{G-\{v_i\}}|$ difference. Larger difference ranks higher.

Top ranked nodes are considered as key separators.
KPP-neg (2)

• A measure of connectivity: reachability

Pseudocode 1: $\text{Reach}(v_i) –$ number of nodes reachable from $v_i$

Go to Source vertex $v_i$ and mark it as visited and add to the set $\text{Reach}(v_i)$

For each adjacent vertex, $A$, of $v_i$,

If $A$ is not already visited,

Add adjacent vertex $A$ to the set $\text{Reach}(v_i)$ and mark $A$ as visited

Call $\text{Reach}(A)$

$$C_G = \sum_{i=1}^{n} \text{Reach}(v_i)$$
Example

R(E) = E, C, A, B, D, F
Example (2)

NOTE: node reachability is a more accurate measure than previously seen “REACH”
Graph-based measures of social influence

• Use graph-based methods/algorithms to identify “relevant players” in the network
  – Relevant players = more influential, according to some criterion

• Use graph-based methods to identify global network properties and communities (community detection)

• Use graph-based methods to analyze the “spread” of information