Social Media and Social Influence

Nihar Shah · Peter Tu

#cs286r

7 November 2012
Twitter rumors

V. Kolokoltsev
@MemInterRussia

RUSSIAN GOVERNMENT CONFIRMS: BASHER AL ASSAD, PRESIDENT OF SYRIA, HAS BEEN KILLED TODAY IN LATAKYJA WITH HIS WIFE AND TWO PEOPLE.

August 6, 2012
Twitter rumors

V. Kolokoltsev
@MiniInterRussia

RUSSIAN GOVERNMENT CONFIRMS: BASHAR AL ASSAD, PRESIDENT OF SYRIA, HAS BEEN KILLED TODAY IN LATAKYJA WITH HIS WIFE AND TWO PEOPLE.

August 6, 2012

Figure: Price of Crude Oil WTI September 2012 Futures. Source: Bloomberg, 2012.
The xx album

Figure: Release of the xx album Coexist. Source: thexx.com.
**Figure:** Network of 2,200 individuals from the Framingham Heart and Health Study. *Source:* Christakis and Fowler 2007.
A formal definition of social influence

Social influence occurs when one's emotions, opinions, or behaviors are affected by others.
A formal definition of social influence

Social influence occurs when one’s emotions, opinions, or behaviors are affected by others.¹

¹http://qualities-of-a-leader.com/personal-mbti-type-analysis/ by way of Wikipedia.
Why do we care about social influence?

Why do we care about social influence?

From papers: voting and marketing.
Why do we care about social influence?

From papers: voting and marketing.

Others?

<table>
<thead>
<tr>
<th>Music/food/book recommendations</th>
<th>Shared interest groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public health campaigns</td>
<td>Upending dictatorships</td>
</tr>
<tr>
<td>Crowdsourcing</td>
<td></td>
</tr>
</tbody>
</table>
I. Story time

II. Definition and application

III. Models of social influence

IV. Does social influence exist in social media?
   1. Cha et al. 2010
   2. Bond et al. 2012
   3. Goel, Watts, and Goldstein 2012

V. The Power of Social Data
   1. Goel, and Goldstein 20–
Models of social influence

1. "Influentials" - a coterie of highly visible individuals or media sources (Rogers 1962). Examples: Oprah Winfrey, Natalie Portman, George Clooney.


3. "Social Networks" - close relationships determine the degree of influence.

Nihar Shah · Peter Tu
Social Media and Social Influence
7 November 2012
“Influentials”

coterie of highly visible individuals or media sources (Rogers 1962).
Ex: Oprah Winfrey, Natalie Portman, George Clooney
“Influentials”
coterie of highly visible individuals or media sources (Rogers 1962).
Ex: Oprah Winfrey, Natalie Portman, George Clooney

“Accidental influentials”
social influence more dependent on the receptibility of public than the
innate magnetism of an influential (Watts and Dodds 2007).
Analogy: a forest fire
1. "Influentials"
coterie of highly visible individuals or media sources (Rogers 1962).
Ex: Oprah Winfrey, Natalie Portman, George Clooney

2. "Accidental influentials"
social influence more dependent on the receptibility of public than the
innate magnetism of an influential (Watts and Dodds 2007).
Analogy: a forest fire

3. "Social Networks"
close relationships determine the degree of influence
Does social influence exist in social media?  

**The Million Follower Hypothesis**

**Hypothesis:** Twitter users with a high number of followers have significant social influence.
**Hypothesis:** Twitter users with a high number of followers have significant social influence.

<table>
<thead>
<tr>
<th>Name</th>
<th># followers</th>
<th># tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Lady Gaga</td>
<td>30,928,017</td>
<td>2,256</td>
</tr>
<tr>
<td>2 Justin Bieber</td>
<td>29,812,748</td>
<td>19,255</td>
</tr>
<tr>
<td>3 Katy Perry</td>
<td>28,558,878</td>
<td>4,429</td>
</tr>
<tr>
<td>4 Rihanna</td>
<td>26,530,025</td>
<td>6,843</td>
</tr>
<tr>
<td>5 Barack Obama</td>
<td>21,681,783</td>
<td>7,682</td>
</tr>
<tr>
<td>6 Britney Spears</td>
<td>21,593,867</td>
<td>1,862</td>
</tr>
<tr>
<td>7 Taylor Swift</td>
<td>20,403,558</td>
<td>1,566</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n_1$ Harvard University</td>
<td>166,261</td>
<td>14,537</td>
</tr>
<tr>
<td>$n_2$ Yale University</td>
<td>49,828</td>
<td>5,562</td>
</tr>
</tbody>
</table>

Testing the hypothesis

Define

\[
\text{indegree} = \# \text{ of followers} \\
\text{retweet influence} = \# \text{ of retweets with one’s username} \\
\text{mention influence} = \# \text{ of mentions with one’s username}
\]

**Subhypothesis:** If the million follower hypothesis is accurate, then we expect those with the highest indegree to also have the highest retweet and mention influence.

**Strategy:** Rank all users by each influence metric. Run correlation between the rankings.
Does social influence exist in social media?

**The Million Follower Fallacy**

If there exist "influentials", those "influentials" are not necessarily the people with the highest indegree (i.e. the most followers). Hence, the million follower fallacy.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>All</th>
<th>Top 10%</th>
<th>Top 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>indegree v. retweets</td>
<td>0.549</td>
<td>0.122</td>
<td>0.109</td>
</tr>
<tr>
<td>indegree v. mentions</td>
<td>0.638</td>
<td>0.286</td>
<td>0.309</td>
</tr>
<tr>
<td>retweets v. mentions</td>
<td>0.580</td>
<td>0.638</td>
<td>0.605</td>
</tr>
</tbody>
</table>

*Source:* Cha et al. 2010.
The Million Follower Fallacy

If there exist “influentials”, those “influentials” are not necessarily the people with the highest indegree (i.e. the most followers). Hence, the million follower fallacy.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>All</th>
<th>Top 10%</th>
<th>Top 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>indegree v. retweets</td>
<td>0.549</td>
<td>0.122</td>
<td>0.109</td>
</tr>
<tr>
<td>indegree v. mentions</td>
<td>0.638</td>
<td>0.286</td>
<td>0.309</td>
</tr>
<tr>
<td>retweets v. mentions</td>
<td>0.580</td>
<td>0.638</td>
<td>0.605</td>
</tr>
</tbody>
</table>

Source: Cha et al. 2010.
Who are the “influentials” then?

Whoever they are...

- some seem able maintain influence across topics.
- swine flu, michael jackson, iran
Who are the “influentials” then?

Whoever they are...
- some seem able maintain influence across topics.
  - swine flu, michael jackson, iran

- “influential” status may be gained through focused tweeting
  - or by starting digg.com.
Evidence that canvassing Get Out the Vote efforts successfully increases turnout
  - Experimental data that face-to-face mobilization campaigns can trigger social influence effects (Nickerson 2008).

But in-person GOTV efforts require sizable resources

Experiments with email mobilization unsuccessful (Huckfeldt and Sprague 1995).

**Hypothesis:**
Leveraging Facebook’s social network can mobilize voting more cost-effectively.
Experimental Design

On election day November 2, 2010, randomly split US users 18+ into three distinct groups:

1. Social Message ($n = 60,055,176$)

2. Informational Message ($n = 611,044$)

3. No message ($n = 613,096$)
Does social influence exist in social media?

Bond et al. 2012

Effects

Effect on Self-Expression
Users in social cohort 2.08% more likely to identify as a voter than informational group.
Does social influence exist in social media?

Effects

Bond et al. 2012

Increased turnout by 340,000 nationally:
- 60,000 directly
- 280,000 through social contagion

Increased % turnout by 0.4% points
- from 37.4% to 37.8%

<table>
<thead>
<tr>
<th>Year</th>
<th>Winner</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>Bill Clinton</td>
<td>8,201,370</td>
</tr>
<tr>
<td>2000</td>
<td>George W. Bush</td>
<td>-543,816</td>
</tr>
<tr>
<td>2004</td>
<td>George W. Bush</td>
<td>3,012,171</td>
</tr>
<tr>
<td>2008</td>
<td>Barack Obama</td>
<td>9,549,105</td>
</tr>
<tr>
<td>2012</td>
<td>Barack Obama</td>
<td>7,008,214</td>
</tr>
</tbody>
</table>

Presidential v. midterm elections?

Nihar Shah · Peter Tu

Social Media and Social Influence

7 November 2012
Effect on Voting

Increased turnout by 340,000 nationally:
- 60,000 directly
- 280,000 through social contagion

Increased % turnout by 0.4% points
- from 37.4% to 37.8%
Does social influence exist in social media?

Bond et al. 2012

**Effects**

**Effect on Voting**

Increased turnout by 340,000 nationally:
- 60,000 directly
- 280,000 through social contagion

Increased % turnout by 0.4% points
- from 37.4% to 37.8%

<table>
<thead>
<tr>
<th>Year</th>
<th>Winner</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>Bill Clinton</td>
<td>8,201,370</td>
</tr>
<tr>
<td>2000</td>
<td>George W. Bush</td>
<td>-543,816</td>
</tr>
<tr>
<td>2004</td>
<td>George W. Bush</td>
<td>3,012,171</td>
</tr>
<tr>
<td>2008</td>
<td>Barack Obama</td>
<td>9,549,105</td>
</tr>
<tr>
<td>2012</td>
<td>Barack Obama</td>
<td></td>
</tr>
</tbody>
</table>
Effect on Voting
Increased turnout by 340,000 nationally:
- 60,000 directly
- 280,000 through social contagion

Increased % turnout by 0.4% points
- from 37.4% to 37.8%

<table>
<thead>
<tr>
<th>Year</th>
<th>Winner</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>Bill Clinton</td>
<td>8,201,370</td>
</tr>
<tr>
<td>2000</td>
<td>George W. Bush</td>
<td>-543,816</td>
</tr>
<tr>
<td>2004</td>
<td>George W. Bush</td>
<td>3,012,171</td>
</tr>
<tr>
<td>2008</td>
<td>Barack Obama</td>
<td>9,549,105</td>
</tr>
<tr>
<td>2012</td>
<td>Barack Obama</td>
<td></td>
</tr>
</tbody>
</table>

Presidential v. midterm elections?
Effect on self-expression and voting both increasing in closeness of friend

This lends support to theory 3: the influence of social networks.
**Analogy:** Social influence as infectious disease.

- There exists a patient zero who infects a small number of people, who in turn, infect others, who in turn infect others, etc.

**The “Virality” of Social Influence**

- Infection as product adoption.
Does social influence exist in social media?

Goel, Watts, Goldstein 2012

Cascade Size

<table>
<thead>
<tr>
<th>Tree Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Depth</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Social Influence
## The Dataset: 7 Online Domains

<table>
<thead>
<tr>
<th>Network</th>
<th>Description</th>
<th>Adoption Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yahoo! Kindness</td>
<td>social network centered around philanthropy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>register on the site via invite</td>
</tr>
<tr>
<td>2</td>
<td>Zync</td>
<td>video plugin for Yahoo! Messenger</td>
</tr>
<tr>
<td></td>
<td></td>
<td>install plugin via invite</td>
</tr>
<tr>
<td>3</td>
<td>The Secretary Game</td>
<td>webgame</td>
</tr>
<tr>
<td></td>
<td></td>
<td>play online via invite</td>
</tr>
<tr>
<td>4</td>
<td>Twitter News Stories</td>
<td>webgame</td>
</tr>
<tr>
<td></td>
<td></td>
<td>play online via invite</td>
</tr>
<tr>
<td>5</td>
<td>Twitter Videos</td>
<td>third-party</td>
</tr>
<tr>
<td></td>
<td></td>
<td>added app and answered questions</td>
</tr>
<tr>
<td>6</td>
<td>Friend Sense</td>
<td>Facebook application</td>
</tr>
<tr>
<td></td>
<td></td>
<td>purchase voice credits</td>
</tr>
<tr>
<td>7</td>
<td>Yahoo! Voice</td>
<td>purchase voice credits</td>
</tr>
</tbody>
</table>
Does social influence exist in social media?

Goel, Watts, Goldstein 2012

Social Media and Social Influence

7 November 2012
Where do viral videos come from then?!
Where do viral videos come from then?!

**Hypothesis 1:** Exceptional nature of large cascades. Not all outbreaks become epidemics (thankfully).

**Hypothesis 2:** Importance of Influentials (theory 1).
Social data can be valuable in its own right for predictions (Goel and Goldstein, mimeo)

This result holds even when social influence does not exist!

Product marketers can use social networks to infer a given customer’s propensity to purchase a product, even if no other customer influences one another.
Goel and Goldstein introduce stochastic band networks (Holland 1976) to demonstrate this.

- Individuals belong to one of several communities or “bands.”
- Individuals within the same band have the same underlying preference for products.
- Individuals can befriend or “connect” with others both inside and outside their bands.
- Marketers cannot directly observe the band of a given individual.
Stochastic Band Networks
Stochastic Band Networks
Stochastic Band Networks
Now we return to the marketer, who cannot observe groups but only underlying purchasing decision
Formally, we can generate a model with four types of parameters:

1. $n$: the number of individuals
2. $K$: the number of “bands”
3. $\theta^+$ and $\theta^-$: the probability of being connected to a given neighbor; $\theta^+$ is for same-band neighbors, and $\theta^-$ is for different-band neighbors
4. $p_i, i \in \{1, 2, \ldots, K\}$: the probability of an individual in band $i$ adopting the product
We now generate a model with the following parameters:

1. $n = 20$
2. $K = 2$
3. $\theta^+ = 0.8$ and $\theta^- = 0.2$
4. $p_1 = 0.9$ and $p_2 = 0.2$

Notice that at no point, does one individual’s purchase influence another individual.
From the networks, marketers can infer which band a new individual likely belongs to, and so whether that individual might purchase.
Stochastic Band Networks

Mathematically, the result can be shown as follows:

\[
P(\text{node } v \text{ adopts } | \text{ node } v \text{ has } m \text{ adopting neighbors}) = \frac{\sum_{k=1}^{K} p_k \mu_{k,m}}{\sum_{k=1}^{K} \mu_{k,m}}
\]

\[
p_k = \text{probability of an individual in band } k \text{ adopting}
\]

\[
\mu_{k,m} = P(\text{node } v \text{ has } m \text{ adopting neighbors }|\text{ node } v \text{ is in block } k)
\]

\[
= \binom{n-1}{m} \left( \frac{p_k \theta^+}{K} + \frac{\theta^-}{K} \sum_{j \neq k} p_j \right)^m \left( 1 - \frac{p_k \theta^+}{K} - \frac{\theta^-}{K} \sum_{j \neq k} p_j \right)^{n-1-m}
\]
In this example, node 20 is “red” so the true adoption probability is 0.2. Using the formula, the marketer generates an estimated probability of adopting of 0.208.
Stochastic Band Networks

How can marketers use this model in practice?

- What assumptions seem realistic?
Stochastic Band Networks

How can marketers use this model in practice?

- What assumptions seem realistic?
- What assumptions seem unrealistic?

What if $\theta + \theta = \theta - \theta$?
How can marketers use this model in practice?

- What assumptions seem realistic?
- What assumptions seem unrealistic?
- What if $\theta^+ = \theta^-$?
Empirical Examples

Goel and Goldstein demonstrate some uses of social data in forecasting consumer preferences, without assuming any casual relationship.

They compare three prediction models:
Empirical Examples

Goel and Goldstein demonstrate some uses of social data in forecasting consumer preferences, without assuming any casual relationship.

They compare three prediction models:

1. A baseline demographic model
Empirical Examples

Goel and Goldstein demonstrate some uses of social data in forecasting consumer preferences, without assuming any casual relationship.

They compare three prediction models:

1. A baseline demographic model
2. A social model

The social data was the Yahoo communications network, and an edge between two users was established if two users mutually exchanged at least one email or IM.
Goel and Goldstein demonstrate some uses of social data in forecasting consumer preferences, without assuming any casual relationship.

They compare three prediction models:

1. A baseline demographic model
2. A social model
3. A model that uses both demographic and social data
Goel and Goldstein demonstrate some uses of social data in forecasting consumer preferences, without assuming any casual relationship.

They compare three prediction models:

1. A baseline demographic model
2. A social model
3. A model that uses both demographic and social data

The social data was the Yahoo communications network, and an edge between two users was established if two users mutually exchanged at least one email or IM.
Empirical Examples

Consider two examples: shopping, and fantasy sports.

**Figure:** Adoption rates for the shopping (A) and recreational league (B) domains.  
*Source:* Goel and Goldstein, mimeo.
Conclusion

Have we influenced you???
Have we influenced you???

Questions?
Conclusion

Have we influenced you???

Questions?

Thank you!