Analysis of Twitter unfollow: How often and why

SocInfo 2011
Singapore Management University

Haewoon Kwak
Hyunwoo Chun
Wonjae Lee
Sue Moon
Two basic processes in network evolution
Building a relationship

Breaking a relationship
Building a relationship

Breaking a relationship

Friends You Are To Break Up Soon

Your Dad
“Dark side of the force”
Lack of relationship breakup data

• Can you capture the breakup from
  – A paper coauthorship network?
  – A mobile phone call network?
  – An e-mail network?
  – A wall message network?

Very hard to define and capture breakup
Canceling *the* friendship online?
Unfortunately (?), people do take online etiquette seriously.
Click on a friend to begin the sacrifice.

Adam Gilad  Adam Sachs  Adam Zand
Adriana C.  Alex Argrow  Alex Mehr
Alex Panelli  Amanda Tsinonis-  Amanda Wheeler
Amiee Parco  Amy Alex Andersen  Amy Canaday

FRIENDSHIP IS STRONG, BUT THE WHOPPER® IS STRONGER.

Use this application and be rewarded with a free flame-broiled WHOPPER® when you sacrifice 10 of your Facebook friends. Each friend will be notified so choose wisely.

Share The Sacrifice
A proxy for relationship breakup

• Disappearance of e-mail exchange

  “Alice & Bob had exchanged e-mails frequently. At some point they didn’t do any more.”
  But they were doing instant messaging.

A proxy is not always accurate!
How about Twitter?
Different from other OSNs

In Twitter
“I follow you”

In most OSNs
“We are friends”
for everyone
http://twitpic.com/135xa – There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.
What is Twitter, a Social Network or a News Media?

Haewoon Kwak  Changhyun Lee  Hosung Park  Sue Moon
Department of Computer Science, KAIST, Korea

19th International World Wide Web Conference (WWW2010)
Unfollow in Twitter

• **Intentional action** to break a relationship
• **No need for an approval**
• **No notification** to the unfollowed
Write a short message

Read others’ tweets in my timeline

Check out trending topics
Four Types of Tweets

Tweet
Last Day of SocInfo!

Reply
@EePeng Thanks for having me!

Mention
I am attending SocInfo 2011 organized by @EePeng

Retweet
Wow lots of fun talks and great people! RT @sbmoon Last Day of SocInfo!
Recent changes in Twitter API
Our Unfollow Study

- Macroscopic statistics
- User interviews
- Explanatory model
No official records for unfollow

Removed relationship

\[ t = 0 \quad t = 1 \quad \ldots \]

• Compare two follow networks and detect removed relationships
Twitter now too big to crawl

• In 2009 it took 3 months with 20+ hosts to crawl and get the entire 40 million user profiles
• Now in 2011 it has more than 400 million users
• How to sample?
Identify a group with common culture

• Collect 1.2M Korean-speaking users identified by Korean in tweets, bio, location, or name
Data collection

• Collect daily snapshots of follow networks
  – G(I): June 25th to July 15th, 2010
  – G(II): August 2nd to August 31st, 2010

• Time resolution = a single day
Korean follow network grew fast

- Increasing # of users
  - G(I): 718,077 $\rightarrow$ 870,057 $+7,599$/day
  - G(II): 956,261 $\rightarrow$ 1,203,196 $+8,515$/day

- Increasing (high) reciprocity
  - G(I): 56 $\rightarrow$ 58%
  - G(II): 61 $\rightarrow$ 62%

- Increasing avg. # of followees
  - 59.7 $\rightarrow$ 75.7
Macroscopic Statistics
People unfollow frequently

- 43% of active users unfollow at least once during 51 days

- Average number of unfollows per person
  - 15.4 in G(I)
  - 16.1 in G(II)

Link removal frequently occurs in a ‘growing’ network
Factors that correlate with unfollow

- Reciprocity of relationships
- Duration of a relationship
- Followee’s informativeness
- Overlap of relationships
One-way relationships are fragile

One-way

P(broken) = 0.1228

“Emotional closeness”

Reciprocal

P(broken) = 0.0529

Conditional P(remaining will be broken | one is broken) = 0.2345
No knowledge when following started

A followee list returned by API is always [B, A, C]

Instead, we have a temporal ‘order’ of relationships in a personal network
Newer relationships are more fragile

Clear positive trend
Informativeness of users

Next year's rebuttals will be limited to 140 characters, counting spaces.
A retweets B

B is informative to A
Non-informative relationships are fragile

Retweeted or favorited users are less likely to be unfollowed
Strong ties & weak ties

Weak ties are fragile

The more overlapped relationships, the less likely to be unfollowed.
No interaction ≠ breakup

• 85.6% of relationships do not involve any single reply, mention, or retweet
  – 96.3% involve 3 or fewer

• People just subscribe to others’ tweets passively

Why our study of ‘unfollow’ is important for the study of breakup
User Interviews
22 online & face-to-face interviews

- Recruited by word-of-mouth
- Semi-structured
- Logging & camera recording

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Std. dev.</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>27.3</td>
<td>27</td>
<td>22</td>
<td>36</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>Favorites</td>
<td>80.7</td>
<td>1</td>
<td>0</td>
<td>851</td>
<td>199.0</td>
<td></td>
</tr>
<tr>
<td>Followers</td>
<td>846.7</td>
<td>164.5</td>
<td>5</td>
<td>8,772</td>
<td>2,053.9</td>
<td></td>
</tr>
<tr>
<td>Followings</td>
<td>600.4</td>
<td>144.5</td>
<td>5</td>
<td>7,103</td>
<td>1,562.7</td>
<td></td>
</tr>
<tr>
<td>Tweets</td>
<td>3,325.8</td>
<td>583.5</td>
<td>5</td>
<td>30,639</td>
<td>7,220.5</td>
<td></td>
</tr>
<tr>
<td>Registered days</td>
<td>449.2</td>
<td>471</td>
<td>14</td>
<td>766</td>
<td>179.1</td>
<td></td>
</tr>
</tbody>
</table>
Top reasons in unfollow

1. Burst tweets
2. Tweets about uninteresting topics
3. Tweets about mundane details of daily life
   • Automatically generated tweets (e.g., 4sq)
4. Tweets about political issues
Explanatory Model
No explanatory model taking structural and interactional properties into consideration.
Interactional properties describe dynamics

• Has a followee sent a reply to a follower?
• Has a user mentioned a follower in any of one’s tweet?
• Do a user and a follower share common topics of interest?
Sociological concepts in persisting ties

Homophily
Reciprocal services
Tie strength
Power and prestige

+ Twitter-specific feature: Informativeness
Multiple logistic regression model

• Binary dependent variable
  – Whether a relationship at $t_0$ will be broken (1) or persisted (0) at $t_1$

• 78 Independent variables
  – From structural properties
    • # of followers, # of followees, # of common followers, …
  – From interactional properties
    • # of replies, # of retweets, # of communication partners, …
Filtered variables

• Removing multicollinearity
  – # of common followers & followees & neighbors
  – # of follower & those who reply to ego
  – # of those who exchange replies & all replies
  – # of received replies & mentions
  – ...

• We filter out 36 variables and 42 remained
• With stepwise regression, we further winnowed down to 39 variables
<table>
<thead>
<tr>
<th>Structure properties</th>
<th>Coefficient</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$ Followees</td>
<td>-4.13e-04</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u$ Followers</td>
<td>5.59e-05</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u$ Followees/Followers</td>
<td>-1.71e-03</td>
<td>0.998***</td>
</tr>
<tr>
<td>$f$ Follow-back ratio</td>
<td>-0.21 (4.02e-03)</td>
<td>0.810***</td>
</tr>
<tr>
<td>$f$ followees</td>
<td>2.67e-05 (2.53e-07)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ followers</td>
<td>-2.68e-06 (8.01e-08)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ Followees/Followers</td>
<td>1.68e-03 (1.28e-04)</td>
<td>1.002***</td>
</tr>
<tr>
<td>$f$ Follow-back ratio</td>
<td>0.82 (4.68e-3)</td>
<td>2.282***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Order of follow</td>
<td>-1.57e-08 (8.41e-11)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Order of follow</td>
<td>2.58e-08 (6.71e-10)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \leftrightarrow f$ Prestige</td>
<td>-1.22e-06 (2.37e-07)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \leftrightarrow f$ Reciprocity of follows</td>
<td>-0.80 (2.50e-03)</td>
<td>0.451***</td>
</tr>
<tr>
<td>$u \leftrightarrow f$ Common followees</td>
<td>-7.40e-05 (9.12e-06)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \leftrightarrow f$ $N$ Common followees</td>
<td>-2.89 (2.78e-02)</td>
<td>0.056***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity properties</th>
<th>Coefficient</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$ Tweets</td>
<td>2.76e-04 (9.40e-07)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u$ URL tweets</td>
<td>-1.40e-04 (7.24e-06)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u$ Auto-generated tweets</td>
<td>5.26e-04 (7.32e-05)</td>
<td>1.001***</td>
</tr>
<tr>
<td>$u$ Popularity</td>
<td>-1.44e-04 (1.07e-06)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ Tweets</td>
<td>-4.27e-05 (1.11e-06)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ URL tweets</td>
<td>-4.32e-05 (8.74e-06)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ Auto-generated tweets</td>
<td>-2.86e-04 (1.14e-04)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$f$ Popularity</td>
<td>-6.04e-06 (1.02e-07)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Replies</td>
<td>2.85e-03 (1.32e-04)</td>
<td>1.003***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Replies</td>
<td>-0.67 (2.22e-02)</td>
<td>0.516***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Mentions</td>
<td>-9.23e-04 (3.70e-04)</td>
<td>0.999*</td>
</tr>
<tr>
<td>$u \rightarrow f$ Retweets</td>
<td>-1.21e-02 (9.65e-04)</td>
<td>0.988***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Favorites</td>
<td>-5.06e-02 (1.95e-03)</td>
<td>0.951***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Favorites</td>
<td>-3.95e-04 (2.53e-02)</td>
<td>1.000***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Replies</td>
<td>-1.58e-02 (3.91e-04)</td>
<td>0.984***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Replies</td>
<td>-1.49 (3.52e-02)</td>
<td>0.225***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Mentions</td>
<td>-0.66 (5.19e-02)</td>
<td>0.516***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Retweets</td>
<td>-5.17e-02 (2.37e-03)</td>
<td>0.950***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Retweets</td>
<td>-0.76 (4.92e-02)</td>
<td>0.469***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Favorites</td>
<td>-6.11e-02 (5.01e-03)</td>
<td>0.941***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Favorites</td>
<td>-0.35 (4.88e-02)</td>
<td>0.706***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Common hashtags</td>
<td>-0.12 (1.29e-03)</td>
<td>0.883***</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Common hashtags</td>
<td>-2.2 (4.20e-02)</td>
<td>0.109***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Days since first comm.</td>
<td>-1.95e-03 (4.8e-05)</td>
<td>0.998***</td>
</tr>
<tr>
<td>$u \rightarrow f$ Days since last comm.</td>
<td>-1.08e-04 (3.38e-06)</td>
<td>1.000***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Removed variables</th>
<th>Coefficient</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u \rightarrow f$ $N$ Mentions</td>
<td>3.49e-02 (3.46e-06)</td>
<td>1.000</td>
</tr>
<tr>
<td>$u \rightarrow f$ $N$ Retweets</td>
<td>6.25e-03 (2.28e-02)</td>
<td>1.000</td>
</tr>
<tr>
<td>$u \leftarrow f$ Mentions</td>
<td>1.09 (5.68e-04)</td>
<td>1.001</td>
</tr>
</tbody>
</table>

* $p < 0.05$  ** $p < 0.01$  *** $p < 0.001$
* Odds ratio is rounded to thousands.
## Interpretation of models

<table>
<thead>
<tr>
<th><strong>Emotional closeness</strong></th>
<th>If not reciprocal, 2.17x times likely to unfollow</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Homophily</strong></td>
<td>Jaccard coeff. of hashtags decreases 0.2, 2.75x</td>
</tr>
<tr>
<td><strong>Tie strength</strong></td>
<td>If overlap of followees decrease 0.2, 2.12x</td>
</tr>
</tbody>
</table>

- **Receiving actions rather than giving ones** more important variable than the opposite
  - People appreciate receiving feedback
  - Consistent with the study of social capital at Facebook
Summary

- Unfollow is a clear sign of breaking a relationship
  - No interaction is NOT a good indicator of breaking a relationship due to user’s passivity in Twitter

- Quantitative and qualitative studies of unfollow
  - Reveal its characteristics and motivations
  - Discover important structural and interactional properties
Application: Practical uses
Burst tweets, automatically generated tweets, ...

<table>
<thead>
<tr>
<th>Time</th>
<th>Username</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 hrs</td>
<td>espn</td>
<td>BOOM, Andre Ethier just singled in the bottom of the 5th to extend his hitting streak to 28: es.pn/kyk0l3</td>
</tr>
<tr>
<td>13 hrs</td>
<td>espn</td>
<td>Nobody’s perfect. The Red Sox handed Jered Weaver his 1st loss of the season tonight: es.pn/kjABko</td>
</tr>
<tr>
<td>13 hrs</td>
<td>espn</td>
<td>@R1cky_F0ntane Don’t worry, we won’t tell on you.</td>
</tr>
<tr>
<td>14 hrs</td>
<td>espn</td>
<td>We have a winner... RT @lindacohn: Bruins scored!</td>
</tr>
<tr>
<td>1 day</td>
<td>mikaelhuss</td>
<td>Whoa ... the king of Sweden just walked into our office spaces here at #scilifelab</td>
</tr>
<tr>
<td>1 day</td>
<td>owlcity</td>
<td>DARTH VADER: &quot;Hey Obi, I made you some toast.&quot; OBI-WAN: &quot;It's a bit on the dark side, isn’t it?!” #LOL</td>
</tr>
<tr>
<td>1 day</td>
<td>RealWizKhallifa</td>
<td>Lol has gone from meaning, &quot;laugh at a joke” to &quot;wiz Khalifa&quot;!</td>
</tr>
</tbody>
</table>
Application: Theoretical models

• Existing models for growing networks deal with link additions only

• Our studies reveal frequent link removals in a growing network
One population - Korean
How can we generalize our results?
Follow ≠ Friend
Thanks!
Questions?