Structural Predictors of Tie Formation in Twitter: Transitivity & Mutuality

Scott A. Golder¹ and Sarita Yardi²

¹ Department of Sociology, Cornell University
² Human Centered Computing, Georgia Institute of Technology

This work was conducted at Microsoft Research New England, Summer 2009.
Introduction

• The problem of discovery:
  – How to meet new people?
  – What are the structural conditions under which new ties are likely to form?
  – Concepts: homophily, balance, triadic closure

• Goals
  – Use behavior in the online world to make discoveries about social life in general.
  – The online world is malleable. Use social science to make the online world better.
Twitter
Twitter

Latest: on a skype call with 8 people and 3 continents. Skype is having a hard time keeping up. 5 days ago

jawar Are we friends on Facebook yet http://facebook.com/ RT
half a minute ago from web

beeblebox One Flu Shot Over the Cuckoo’s Nest
half a minute ago from txt

MarcelaMariana http://bit.ly/tB15l Ciel renueva su imagen como parte de una estrategia que promueve el estilo de vida activo
half a minute ago from web

mattskint @JCLFAULKNER Nah, I just can’t see the point in wasting time with measly wine glasses. I’m gonna drink the bottle anyway…
half a minute ago from Tweetie in reply to JCLFAULKNER

domi_mclean13 #PrayingForBrianLittrell
half a minute ago from Power Twitter

half a minute ago from Twitter Tools

Trending Topics
#MusicMonday
#TmobileSucks
#MM
Halloween
FlashForward
T-Mobile
Gourmet
FTC
Packers
Directed Ties

A → B

vs.

A ← B

A ↔ B
Directed Ties
Conceptual Toolkit

- Homophily - similarity breeds connection”
  - “Structural” vs. “Choice” Homophily
    - neighborhoods, schools, workplaces, clubs, versus social psychology

- Structural Balance and Triadic Closure
  - How do stability and consistency arise in networks?
    - If A – X and X – B, what about A and B?

- Social exchange networks
  - An “attention / information” network.
Four Drivers of Attraction

1. Shared Interests
   - Outlink equivalence

1. Shared Audiences
   - Inlink equivalence

1. Transitivity & Filtering
   - Social Proof vs. Redundancy

1. Reciprocity
   - The Golden Rule
Method & Data

• Web-based field experiment
• Sampled users from the public timeline
  – @user Hi! Would you be willing to help us with a short Twitter experiment? It's easy and takes less than 5 minutes. Thanks!
  – Recruited 2,085; 250 replied and 69 completed.
• Collected 2° egonets and computed network statistics.
Experimental Design

Take a look at the profile below. What parts about this person do you find interesting or not?

Would you want to follow this person? (3/14)

Definitely  Probably  I'm not sure  Probably not  Definitely not!

(this item is missing/broken)

Pixelette

Location: Melbourne
Bio: The way to my heart is with good kerning (and perhaps a cup of tea, some great handwriting, and graphic ephemera from a century ago)

Following: 86    Followers: 117    Updates: 552

do I use margin, or padding? margin or padding, margin or padding . . .
Wed Aug 19 12:35:17 +0000 2009

Junior Mints + Peppermint tea: A pretty naise little break
Wed Aug 19 12:34:44 +0000 2009

Aha! Anonymous slander at risk! Court rules Google must expose identity of slanderous blogger
http://bit.ly/ZJja (via @meetmeatmikes)
Wed Aug 19 12:34:11 +0000 2009

mock mock mock it up
Wed Aug 19 11:18:53 +0000 2009

Remember Clag? What an awesome name for a product. The logo sure is somethin' too.
Tue Aug 18 11:00:08 +0000 2009

@FirstFiveOutlets er.. looking good!!
Experimental Design

Take a look at the profile below. What parts about this person do you find interesting or not?

Would you want to follow this person? (3/14)

- Definitely
- Probably
- I'm not sure
- Probably not
- Definitely not!

pixelette

Location: Melbourne
Bio: The way to my heart is with good kerning (and perhaps a cup of tea, some great handwriting, and graphic ephemera from a century ago)

Following: 86  Followers: 117  Updates: 552

do I use margin, or padding? margin or padding, margin or padding...
Wed Aug 19 12:35:17 +0000 2009

Junior Mints + Peppermint tea. A pretty naiice little break
Wed Aug 19 12:34:44 +0000 2009

Ahah! Anonymous slander at risk! Court rules Google must expose identity of slanderous blogger
http://bit.ly/TJtpa (ha! @meeteamikes)
Wed Aug 19 12:34:11 +0000 2009

mock mock mock it up
Wed Aug 19 11:18:53 +0000 2009

Remember Clag? What an awesome name for a product. The logo sure is somethin' too.
http://tr.im/wBeC?
Tue Aug 18 11:00:08 +0000 2009

@FirstFiveOut.cooh or: looking good!
Experimental Design

Thanks for participating!

We've provided links if you want to actually follow any of the users. This page may take a minute to load.

Flo Foxworthy (flo)

Lacey Wise (laceylove17)

Ryan Thomas (gumby1111)

Lani Nelson (LaniNelson)

Christian (ChrissyAsian)

Adorian (AdorianDeck)
Validity Check

- Subjects' responses do not always match their subsequent behavior.
- Of 69 subjects, 33 followed at least 1 person.
- Most “follows” were for people rated a “Definitely” or “Probably.”
- Low “follow” rates suggest subjects are naïve or overly optimistic in their ratings.
## Summary of Results

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Subject</th>
<th>Alters</th>
<th>Combined</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.078 ***</td>
<td>3.385 ***</td>
<td>3.399 ***</td>
<td>3.364 ***</td>
</tr>
<tr>
<td>Subject’s Friend Count</td>
<td>0.131</td>
<td>0.148</td>
<td>-0.039</td>
<td></td>
</tr>
<tr>
<td>Subject’s Follower Count</td>
<td>0.799 ***</td>
<td>0.730 **</td>
<td>0.803 ***</td>
<td></td>
</tr>
<tr>
<td>Subject’s Tweet Count</td>
<td>-0.308 ***</td>
<td>-0.293 ***</td>
<td>-0.245 †</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>0.357 **</td>
<td>0.360 **</td>
<td>0.354 *</td>
<td></td>
</tr>
<tr>
<td>Subject's account age</td>
<td>-0.024</td>
<td>-0.026</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>Alter’s Friend Count</td>
<td>-0.130</td>
<td>-0.166 *</td>
<td>-0.222 *</td>
<td></td>
</tr>
<tr>
<td>Alter’s Follower Count</td>
<td>0.273 **</td>
<td>0.235 **</td>
<td>0.258 *</td>
<td></td>
</tr>
<tr>
<td>Alter’s Tweet Count</td>
<td>-0.045</td>
<td>-0.013</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Alter’s order in expt.</td>
<td>-0.033 *</td>
<td>-0.036 *</td>
<td>-0.016</td>
<td></td>
</tr>
<tr>
<td>Alter has default photo</td>
<td>-0.067</td>
<td>0.002</td>
<td>-0.011</td>
<td></td>
</tr>
<tr>
<td>Alter includes a bio</td>
<td>0.358</td>
<td>0.336 †</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Alter includes a location</td>
<td>-0.423 †</td>
<td>-0.400 †</td>
<td>-0.439 *</td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-729.133</td>
<td>-733.197</td>
<td>-720.428</td>
<td>-685.852</td>
</tr>
<tr>
<td>Clusters</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td>N</td>
<td>443</td>
<td>443</td>
<td>443</td>
<td>432</td>
</tr>
</tbody>
</table>

† p < 0.10 * p < 0.05 ** p < 0.01
## Network Characteristics

<table>
<thead>
<tr>
<th>Tie type</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity (B → A)</td>
<td>0.348</td>
</tr>
<tr>
<td>(1) A → X → B</td>
<td>-0.053</td>
</tr>
<tr>
<td>(2) A ← X ← B</td>
<td>0.142 *</td>
</tr>
<tr>
<td>(3) A → X ← B</td>
<td>0.068</td>
</tr>
<tr>
<td>(4) A ← X → B</td>
<td>-0.132</td>
</tr>
<tr>
<td>(5) A ↔ X → B</td>
<td>0.222 *</td>
</tr>
<tr>
<td>(6) A ↔ X ← B</td>
<td>-0.155 †</td>
</tr>
<tr>
<td>(7) A → X ↔ B</td>
<td>0.215 **</td>
</tr>
<tr>
<td>(8) A ← X ↔ B</td>
<td>-0.175 *</td>
</tr>
<tr>
<td>(9) A ↔ X ↔ B</td>
<td>0.249 *</td>
</tr>
</tbody>
</table>
Reciprocity

• We did not find a significant effect for reciprocity. 

  – That is, subjects were not significantly more likely to want to follow alters, given that the alters followed them already.

  – Evidence for a primarily structural basis for reciprocity, when it does exist.
Status and Ratings

- Preferential attachment and social proof.
- We show status-based effects based on follower and friend counts.
- In the experiment, high-indegree people were not more likely to be encountered, suggesting a non-structural basis for rating.
- High indegree is in itself desirable.
Location

• Negative and significant effect

• Sharing a location is a kind of similarity:
  _ “Definetly, from houston where I’m from. tweets are worth reading.”
  _ “Same industry as me - Web Designer. In same city as me, Dallas. That’s why I will follow him.”
  _ “He’s funny, lives in my area, and mentions radness in his profile. Definitely adding him.”

• But more people live far from you than near you:
  _ “I would follow this if I were interested in New York. I could see following a twitter dedicated to my neighborhood in Chicago.”
The Best Structural Predictors

- Two conditions must be met:
  - There must be a directed path from A to B (transitivity)
  - At least one tie (A – X or X – B) must be mutual.

- Colloquially speaking:
  - Someone who a close friend pays attention to.
  - A close friend of someone you pay attention to.
  - A close friend of a close friend.

*Mutuality may or may not be a proxy for relational strength.*
• Potential to improve “friend suggestion” algorithms.
  – Twitter just released a “who to follow” feature.

• Questions?

Thanks!

Scott Golder  
sag262@cornell.edu  
@redlog

Sarita Yardi  
syardi3@gatech.edu  
@yardi