The Early Stages of Social Capital:
A Twitter Case Study

Matthew Smith
http://m.smithworx.com
Christophe Giraud-Carrier
cgc@cs.byu.edu

http://dml.cs.byu.edu
March 2010
Introduction

- Social Capital - leverage our online social networks
- Develop framework that quantifies the value
- Allows for theories to be tested

This paper focuses on the following (in Twitter):
  - “Bonding is more likely to occur than bridging” - Lin 01
  - “Closure is the most obvious force” - Burt 05
  - Homophily principle: “Birds of a feather flock together” - Smith-Lovin 87
  - “Similarity begets friendship” -
Motivation

• friendship
• web links
• following
• etc.

---

ESN
Motivation

- commonality
- similarity
- share attributes
Motivation
How does knowing more about the new user help?
Twitter Basics - Home

What’s happening? 140


Tweets (by those I follow)

CNNbrk Haiti children taken by American missionaries reunite with families http://on.cnn.com/cLoYWM half a minute ago via web

larrybrauner 30 Tips for Using Social Media in Your Business http://ping.fm/5GHdH half a minute ago via Ping.fm

ashbuckles #chromeitup 1 minutes ago via TweetDeck

larrybrauner How to Use Social Networking Sites to Drive Business http://amplify.com/u/3bx1 3 minutes ago via Amplify

Hypothesis

Utilizing a following strategy motivated by a desire to maximize potential bonding social capital produces higher returns (or social success) than utilizing other strategies.
Let \( A \) be the set of new Twitter accounts (\(|A|=9\))

Assign each a following strategy

Each behave identically (except for their following strategy)

Let \( U \) be the set of sampled Twitter users to follow (\(|U|=5000\))

tweeted “data mining” or “social networks” in Sept. ’09

Compare stats for each strategy; all accounts in \( A \)

Test if strategy \( A \) (\( max. \ pot. \ bonding \)) results in higher returns
# Experimental Setup - Strategies

- **(A) max. potential bonding**
- **(B) max. potential bridging**
- **(C) median affinities**
- **(D) random**
- **(E) min. following/ers diff.**
- **(F) max. following/ers diff.**
- **(G) median num. followers**
- **(H) min. num. followers**
Experimental Setup - Selection

\( u : \) Pool of 5000

\( \alpha : \)

Begin Selection...
Experimental Setup - Selection

Stop When Each Account is Following 500 Users

\( u : \)

\( \alpha : \)
Experimental Setup - Tracking

Hi, I'm Jon. Thanks for visiting my homepage! I love the Internet, Twitter, and life. If you would like to contact me, you can send me a message on Twitter or send email to: jon at twittercapital.com
## Experimental Results

<table>
<thead>
<tr>
<th>rank&lt;sub&gt;f&lt;/sub&gt;</th>
<th>strategy</th>
<th>following</th>
<th>follow-backs</th>
<th>↓ followers</th>
<th>rejects</th>
<th>churn</th>
<th>follow&lt;sub&gt;total&lt;/sub&gt;</th>
<th>follower&lt;sub&gt;total&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>max. potential bonding (A)</td>
<td>500</td>
<td>158 (32%)</td>
<td>202 (40%)</td>
<td>12</td>
<td>127</td>
<td>512</td>
<td>329</td>
</tr>
<tr>
<td>2</td>
<td>max. following/ers diff. (F)</td>
<td>500</td>
<td>84 (17%)</td>
<td>172 (34%)</td>
<td>12</td>
<td>324</td>
<td>512</td>
<td>496</td>
</tr>
<tr>
<td>3</td>
<td>random (D)</td>
<td>500</td>
<td>118 (24%)</td>
<td>154 (31%)</td>
<td>20</td>
<td>103</td>
<td>520</td>
<td>257</td>
</tr>
<tr>
<td>4</td>
<td>median affinities (C)</td>
<td>500</td>
<td>99 (20%)</td>
<td>123 (25%)</td>
<td>25</td>
<td>93</td>
<td>525</td>
<td>216</td>
</tr>
<tr>
<td>5</td>
<td>max. potential bridging (B)</td>
<td>500</td>
<td>99 (20%)</td>
<td>120 (24%)</td>
<td>25</td>
<td>91</td>
<td>525</td>
<td>211</td>
</tr>
<tr>
<td>6</td>
<td>min. following/ers diff. (E)</td>
<td>500</td>
<td>87 (17%)</td>
<td>99 (20%)</td>
<td>50</td>
<td>55</td>
<td>550</td>
<td>154</td>
</tr>
<tr>
<td>7</td>
<td>median num. followers (G)</td>
<td>500</td>
<td>63 (13%)</td>
<td>86 (17%)</td>
<td>31</td>
<td>51</td>
<td>531</td>
<td>137</td>
</tr>
<tr>
<td>8</td>
<td>min. num. followers (H)</td>
<td>500</td>
<td>33 (07%)</td>
<td>42 (08%)</td>
<td>79</td>
<td>29</td>
<td>579</td>
<td>71</td>
</tr>
<tr>
<td>9</td>
<td>follow nobody (I)</td>
<td>0</td>
<td>0 (&lt;— %)</td>
<td>3 (&lt;— %)</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 1: **Follower Statistics**: Each of the nine accounts are listed by strategy and ranked by the number of followers obtained during the experiment, denoted rank<sub>f</sub>. The following column is the number of users that the account was following at the end of the experiment. The follow-backs column reports the number of users followed that were following the account back at the end of the study, the percent of following is supplied for reference (i.e., follow-backs/following). The followers column is the number of users following the account at the end of the experiment (including those that were never followed by the account). The percent of following is also supplied for reference (i.e., followers/following). The rejects column reports the number of users that could not be followed on Twitter at the time (e.g., account was protected, user attempting to follow was blocked, or user was suspended). The churn statistic reports the number of users that followed the account for a time, but were no longer following the account at the end of the experiment. The follow<sub>total</sub> is the total number of users that were followed by the account, i.e., the sum of following and rejects. The follower<sub>total</sub> is the total number of users that followed the account during the experiment, i.e., the sum of followers and churn.
Results: Followback-to-Following

<table>
<thead>
<tr>
<th>strategy</th>
<th>significantly different</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) max. potential bonding</td>
<td>B, C, E, F, G, H</td>
</tr>
<tr>
<td>(B) max. potential bridging</td>
<td>A, H</td>
</tr>
<tr>
<td>(C) median affinities</td>
<td>A, G, H</td>
</tr>
<tr>
<td>(D) random</td>
<td>G, H</td>
</tr>
<tr>
<td>(E) min. following/ers diff.</td>
<td>A, H</td>
</tr>
<tr>
<td>(F) max. following/ers diff.</td>
<td>A, H</td>
</tr>
<tr>
<td>(G) median num. followers</td>
<td>A, D, C</td>
</tr>
<tr>
<td>(H) min. num. followers</td>
<td>A, B, C, D, E, F</td>
</tr>
</tbody>
</table>

Table 3: **Followback-to-Following**: Pairwise Proportion Test Results. ($\alpha = 0.01$, Bonferroni corrected $p$-values)
## Results - Clicks & Mentions

<table>
<thead>
<tr>
<th>rank(_c)</th>
<th>rank(_f)</th>
<th>strategy</th>
<th>clicks(_t)</th>
<th>clicks(_p)</th>
<th>↓ total clicks</th>
<th>mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>random (D)</td>
<td>900</td>
<td>9</td>
<td>909</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>max. potential bonding (A)</td>
<td>882</td>
<td>15</td>
<td>897</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>min. num. followers (H)</td>
<td>850</td>
<td>16</td>
<td>866</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>max. following/ers diff. (F)</td>
<td>849</td>
<td>7</td>
<td>856</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>median affinities (C)</td>
<td>846</td>
<td>9</td>
<td>855</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>min. following/ers diff. (E)</td>
<td>821</td>
<td>19</td>
<td>840</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>max. potential bridging (B)</td>
<td>773</td>
<td>11</td>
<td>784</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>follow nobody (I)</td>
<td>775</td>
<td>1</td>
<td>776</td>
<td>1</td>
</tr>
</tbody>
</table>
Results - Clicks vs. Followers

Figure 3: Clicks vs. Followers. The linear model shown by the regression line (dashed) poorly fits the data having an $R^2$ value of 0.28. There is, however, a positive Pearson correlation of 0.62, yet it is not as high as might be expected.
Results - Follow-backs Retained

Figure 3: Follow-backs Over Time: Follow-backs obtained by accounts in $\mathcal{A}$ throughout the duration of the study. Days in which following rounds occurred (i.e., accounts in $\mathcal{A}$ followed users in $\mathcal{U}$) are marked in the row labeled f. Days that new status updates were posted to the accounts in $\mathcal{A}$ are marked in the row labeled s.
Conclusions

- New users that requested to follow others having similar profile descriptions (i.e., maximize potential bonding) significantly increased the number of follow-backs.

- This empirical evidence confirms the prevalent sociological beliefs.

- This research assists new users to determine who to connect to in order to grow their network.
Questions & Comments

Ask me now:

Email or Call me:
Matt Smith
smitty@byu.edu
(801) 788-4920

Connect:

Web: http://m.smithworx.com
Blog: http://dmine.blogspot.com
Facebook: http://facebook.com/smithworx
LinkedIn: http://linkedin.com/in/smitty
Twitter: http://twitter.com/smithworx