Social Capital
in Online Communities
PhD Proposal
July 27, 2009

Matthew Smith
smitty@byu.edu
Department of Computer Science
Brigham Young University
Introduction

Online communities

Prevalent shift in how people discover information

Social networks dynamic and complex

Social capital within these networks is poorly understood

Rich social network data now available
Motivation

- Finding people to work with
- Recommending who newcomers could connect with
- Understanding how groups of people evolve
What influence does an individual have upon online friends in terms of mobilizing them to act?

Who should one connect with in order to gain access to additional resources?

Which individuals tend to effectively mobilize social resources?

How heterogenous is the network and what bridging opportunities exist?
Modeling

1. Use relationships, attributes, and social resources
2. Identify set of measurable social resources
3. Formalize notion of accessible/mobilized social resources
4. Run experiments to validate formal SC models
Data Availability

Social networking data generated at unprecedented level

- Popular sites: Facebook, MySpace, LinkedIn, Blogger, and Twitter
- Millions of users are contributing to these sites everyday
- Includes ESN, IAN, and some potentially measurable social resources

Data available for public consumption

- Most Twitter data is available, public Blogs are entirely available
- Personal Facebook data is available — more may become more available
Modeling so far...

- Has focused on: relationships and attributes
- The distinction between potential and realized SC
- Bonding and bridging SC are not reciprocal (Putnam’s Puzzle)
- The model is applied to some available community data

Work Published in ITM Journal
Smith M., Giraud-Carrier C., Purser N. (2009)
Potential vs Realized Social Capital

<table>
<thead>
<tr>
<th>ESN Link</th>
<th>IAN Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes: Realized Bonding</td>
</tr>
<tr>
<td>No</td>
<td>Yes: Potential Bonding</td>
</tr>
</tbody>
</table>

Diagram:

- Amy CS
- Bob CA
- Cheryl CS
- Dan S
- Ed A

Relationships:
- Amy CS to Cheryl CS (1.0)
- Amy CS to Dan S (0.5)
- Bob CA to Cheryl CS (0.33)
- Bob CA to Ed A (0.33)
- Cheryl CS to Dan S (0.5)
- Dan S to Ed A (0.5)
Potential Bonding & Bridging

Individual

\[ pb(i) = \sum_{j \in N, j \neq i} s_{ij}^{IAN} \]

\[ pbr(i) = \sum_{j \in N, j \neq i} (1 - s_{ij}^{IAN}) \]

Network

\[ pb = \frac{\sum_{i \in N} pb(i)}{2} \]

\[ pbr = \frac{\sum_{i \in N} pbr(i)}{2} \]
Realized Bonding & Bridging

**Individual**

\[
b(i) = \sum_{j \in N : j \neq i} s_{ij}^{IAN} s_{ij}^{ESN}
\]

**Network**

\[
b = \frac{\sum_{i \in N} b(i)}{2}
\]

\[
br = \frac{\sum_{i \in N} br(i)}{2}
\]
Normalized Metrics

Potential (normalized by possible connections among individuals)

\[ npb = \frac{2}{N(N - 1)}pb \]
\[ npbr = \frac{2}{N(N - 1)}pbr \]

Realized (normalized by potential)

\[ nb = \frac{b}{pb} \]
\[ nbr = \frac{br}{pbr} \]
Social Resources

- **Material Goods**
  - land, houses, car, and money

- **Symbolic Goods**
  - education, memberships in clubs, honorific degrees, nobility or organizational titles, family name, reputation, or fame

Social Resources

Why?
- Measure obtainable benefit

How?
- Identify set of measurable social resources accessible within online communities (e.g., web traffic, comments, group affiliations, education, net worth)
Validation - Experiments
Language Acquisition Experiment

Apply the proposed SC models on Language Acquisition (LA) studies (currently being performed by the BYU Linguistics department)

Studies centered upon effects of social networks

They will use traditional social scientific approaches while we analyze independently

Analyses will then be compared and contrasted
Twitter Experiment #1

Generate \( j \) ego networks (for an assortment of individuals where relative bonding/bridging is known, or at least agreed upon by \( k \) individuals)

Model individual social capital

Compare results using the non-parametric Mann-Whitney rank-sum test
Twitter Experiment #2

Setup a set $A$ of new Twitter accounts

- Each assigned following strategy: random, maximize bonding, maximize bridging, having b. and br. nearest 50%
- Accounts will behave identically (except for their following strategy)

Let $U$, be a set of sampled Twitter users (from the public timeline)

Status updates for these users will be collected

Compare stats: followers, click-thrus, website click-thrus, b. & br. capital

Test if choosing a b./br. strategy produces significantly higher returns
Medical Blogs Experiment

- Select $m$ known medical blogs to seed a network focused on a particular ailment (e.g., autism, alcoholism, cancer)
- Extend network of study (to the ESN $n$ degrees of freedom away from the seed blogs)
- Track ESN and IAN overtime — analyze using the proposed quantitative network SC modeling
- Identify meaningful qualitative examples
Conclusion

- Mathematical model of social capital that incorporates prior social sciences research
- Explicit social network, implicit affinities, social resources
- Rich online data providing opportunities to validate
- Offers new and novel social network analytics
Questions & Comments

Ask me now: ?

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

Connect:
Web: http://dml.cs.byu.edu/~smitty
Blog: http://dmine.blogspot.com
Facebook: http://matt.smithworx.com
LinkedIn: http://linkedin.com/in/smitty
Twitter: http://twitter.com/smithworx
Additional Slides
Approximate Schedule

March-April 2009  Submission of proposal to advisor (first and second drafts)

May-Jun 2009  Submission of proposal to committee members (final draft)

Jun 2009  Schedule Dissertation Proposal

May-Jun 2009  Collect relevant data for Twitter and blog experiments

Develop social capital modeling and research

Assist in developing language acquisition experiments

July-September 2009  Perform Twitter and blog experiments

September 2009  Perform analysis on language acquisition study

Perform analysis on Twitter and blog experiments

October 2009  Prepare dissertation with latest results and findings

December 2009  Submit dissertation to advisor (first and second drafts)

January 2010  Submit dissertation to committee members (final draft)

February 2010  Schedule Dissertation Defense

March 2010  Dissertation Defense


<table>
<thead>
<tr>
<th>Submission Date</th>
<th>Description</th>
</tr>
</thead>
</table>
| Jun 2009        | *Twitter Capital*  
                 | WBBTmine: Wikis, Blogs, and Bookmarking Tools (ECML/PKDD) |
| Nov 2009        | *Measuring Social Resources in Online Communities*  
                 | SBP: Social Computing, Behaviour Modeling, and Prediction  
                 | WWW: World Wide Web Conference |
| Jan 2010        | *Social Capital through Social Media*  
                 | ICWSM: Conference on Weblogs and Social Media  
                 | KDD: Knowledge Discovery and Data Mining |
| May 2010        | *The Latent Value in Social Networks*  
                 | SocialCom: IEEE International Conference on Social Computing  
                 | SNA-KDD: Social Network Mining and Analysis |