Social Capital in the Blogosphere
A Case Study

Matthew Smith
Nathan Purser
Christophe Giraud-Carrier

Data Mining Lab (http://dml.cs.byu.edu)
Dept. of Computer Science, Brigham Young University
Social Capital

- Concept popularized by Robert Putnam
  - Fosters reciprocity, coordination, collaboration, and communication
  - Researched by many others including Burt, Lin, Coleman, and Bordieu
  - Bonding and bridging
- Social connections are beneficial
  - Individual and group
  - Ex. CEO Compensation, open source projects
- How to measure?
The Blogosphere

- Open community that anyone can join
  (e.g., Blogger, Wordpress, SixApart, your own setup)

- One can blog about anything
  (e.g., fine cuisine, bluegrass music, CS research)

- Both explicit and implicit connections
  (e.g., anchor links, interests)

- Measurable
  (e.g., posts are time-stamped, clickstream available)
Types of Connections

• Explicit Link
  ‣ Direct knowledge, interaction, or communication
  ‣ Ex. friends, web links, and club members
  ‣ Explicit Social Networks (ESNs)

• Implicit Link
  ‣ Inherent similarities or affinities
  ‣ Ex. attributes, hobbies, interests, and background
  ‣ Implicit Affinity Networks (IANs)
Explicit Social Network (ESN)
Links: Friends, Web Links, etc.
Implicit Affinity Network (IAN)
Links: Affinities or inherent similarities
Hybrid Network
ESN overlaid with IAN

Applications: Medical, Political, Blogosphere, etc.
Actual vs. Potential Social Capital

- Potential Social Capital (IAN)
- Actual Social Capital (ESN)
  - Accrues only when explicit links are present

![Table]

<table>
<thead>
<tr>
<th>ESN Link</th>
<th>IAN Link</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Actual Bonding</td>
<td>Actual Bridging</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Potential Bonding</td>
<td>Potential Bridging</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bonding vs. Bridging Social Capital

- **Individual**

  \[
  bonding(i, j) = s_{ij}^{IAN} s_{ij}^{ESN}
  \]

- **Network**

  \[
  bonding = \frac{\sum_{i,j} bonding(i, j)}{\sum_{i,j} s_{ij}^{IAN}}
  \]

  \[
  bridging(i, j) = (1 - s_{ij}^{IAN}) s_{ij}^{ESN}
  \]

  \[
  bridging = \frac{\sum_{i,j} bridging(i, j)}{\sum_{i,j} 1 - s_{ij}^{IAN}}
  \]
Blog Experiment

- Focus
  - Social capital largely unknown
  - Communities centered around topics

- Details
  - Created blog database / Google Reader API
  - 13 million blog entries
  - 38,000+ blogs
  - July 2006 - July 2007 (1 year)
Entry Retrieval Process

- Began with Robert Scoble's blog
- Three step process
  1. Use *pyrfeed* to access blog entries using the unofficial Google Reader API
  2. Extract all links within blog entries
  3. Follow all HTML links to other blogs
Criteria for Implicit Links

- **Topics**
  - Used first level of blog entries
  - Latent Dirichlet Allocation (LDA)
  - Ten topics were extracted (see next slide)

- Implicitly linked by identical topic sets
  - Topic membership assigned when entries contained an $n$-gram from the topic
  - Identical topic sets
<table>
<thead>
<tr>
<th>Topic</th>
<th>Most Likely Topic Components (10 of 20 listed for each topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>de la, regions dash details, la ciudad, de mayo, de abril, de junio, nelson blogcast, de las, de los, distrito federal</td>
</tr>
<tr>
<td>2</td>
<td>elliott back, google news, news articles original, comments office depot featured gadget, platinum system packs, hard drive, geek chic, nvidia geforce, santa rosa, mobile pc</td>
</tr>
<tr>
<td>3</td>
<td>technorati tags, open source, social media, san francisco, windows vista, web site, search engine, years ago, social networking, york times</td>
</tr>
<tr>
<td>4</td>
<td>pdd nos, autism spectrum disorder, autistic children, autistic child, autistic persons, developmental disabilities, ancient greek, michael phelps, autistic son, unstrange minds</td>
</tr>
<tr>
<td>5</td>
<td>fourth quarter, stock symbol, related articles read, etfs type, call transcripts, research stocks, related stocks, net income, cash flow, seeking alpha</td>
</tr>
<tr>
<td>6</td>
<td>lindsay lohan, san francisco, wesmirch permalink, bay area, paris hilton, bed jumping, ice cream, mark pritchard, ed jew, san jose</td>
</tr>
<tr>
<td>7</td>
<td>windows vista, visual studio, net ajax, scott hanselman, download advertisement, windows xp, sql server, windows server, pure evil, web service</td>
</tr>
<tr>
<td>8</td>
<td>feed preferences powered, unified communications, siemens networks, acme packet, mobile convergence, vosky exchange, internet telephony, sip trunking, siemens ag, oliver rist</td>
</tr>
<tr>
<td>9</td>
<td>john mccain, rudy giuliani, white house, mitt romney, homeland security, hillary clinton, fred thompson, al qaeda, real id, barack obama</td>
</tr>
<tr>
<td>10</td>
<td>roxanne darling, ukuule experiment, wines tasted, beach walks, sports racer intros, download quicktimedownload ipoddownload, gary vaennerchuk, shozurobert scoble, discollection hair, joanne colanstony</td>
</tr>
</tbody>
</table>
Criteria for Explicit Links

- Explicitly linked by hyperlink references within blog entries
- 30 reciprocal cross-references
  - Narrowed number of blogs to 224
  - 2358 links, 494 explicit, 1864 implicit
Conclusions

1 of 2

• Bonding relationships exist
  ▸ Explicitly disconnected bloggers writing about the same topics were identified
  ▸ New sub-communities through bonding

• Bridging relationships exist
  ▸ Actual bridging was shown
  ▸ Bridging opportunities were identified
Conclusions
2 of 2

• Methodology
  ▸ Actionable, applicable to online communities
• Mathematical formulation of social capital
  ▸ Utilizes explicit (ESN) and implicit links (IAN)
  ▸ Bonding and bridging vary independently
Future Work

- Affinity and social relationship strengths
  - Which attributes should be used for affinities?
  - What is a significant explicit relationship?
- Further validate social capital metrics
- Suggest potential connections to bloggers
- Pinpoint bloggers with high social capital
  - Adjust the filtering criteria
  - Leverage the long tail
Questions & Comments

Ask me now: ?

Email me: Matthew Smith
smitty@byu.edu

Connect:

Web: http://dml.cs.byu.edu/~smitty
Blog: http://dmine.blogspot.com
LinkedIn: http://linkedin.com/in/smitty