Structural and Cognitive Bottlenecks to Information Access in Social Networks

Jeon-Hyung Kang and Kristina Lerman
USC Information Sciences Institute
Network, Information and Brokerage

Information in networks is non-uniformly distributed
- Certain network positions to get preferential access to information (i.e. new job/higher compensation search)

Theories generalize to online networks
- How users of the social news aggregator Digg adopt stories recommended by friends
- Analysis by measuring
  - Network position
  - Activity rate

How to maximize information access
- Whether user’s information access increase by linking to active users?

Image from http://www.fmsasg.com/SocialNetworkAnalysis/
### Variables used in the study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network</strong></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>Number of active friends</td>
</tr>
<tr>
<td>ND</td>
<td>Network diversity</td>
</tr>
<tr>
<td><strong>User</strong></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Volume of outgoing info. (# votes by user)</td>
</tr>
<tr>
<td>I</td>
<td>Volume incoming info. (friend recommendations)</td>
</tr>
<tr>
<td>B</td>
<td>Avg friend activity</td>
</tr>
<tr>
<td>uB</td>
<td>User activity (# adopted recommendations)</td>
</tr>
<tr>
<td>TD</td>
<td>Friend topic diversity</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td></td>
</tr>
<tr>
<td>NRI</td>
<td>Novel information</td>
</tr>
<tr>
<td>NRI^frds</td>
<td>Novel information friends are exposed to</td>
</tr>
<tr>
<td>NAR</td>
<td>Fraction of novel information adopted by user</td>
</tr>
<tr>
<td>FNAR</td>
<td>Fraction of novel information adopted by friends</td>
</tr>
</tbody>
</table>
Network: Network Size

• The size of social network is simply the number of contacts with user exchanges at least one message during time t.
• Size is the most familiar network characteristic related to information benefits and is a good proxy for a variety of characteristics, which describe the breadth and range of actors’ networks
• Example
  • Number of friends : 11
  • Network Size : 5
Network: Network Diversity

- Describes the degree to which contacts are structurally ‘non-redundant’, and there are both first order and second order dimensions of redundancy. (undirected graph)

\[ ND_i = 1 - C_i = 1 - \frac{\left\{ e_{jk} : u_j, u_k \in N_i, e_{jk} \in E \right\}}{|N_i|(|N_i| - 1)} \]

Proportion to connected component in network contacts
USer: Outgoing/Incoming info & User/frd activity

- Volume of incoming info
  \[ I_{it} = 18 \]

- Volume of outgoing info.
  \[ O_i = O_i^a + O_i^s = 8 \]

- Average Friend activity
  \[ B_{it} = 18 / 5 \]

- User Activity
  \[ uB_{it} = O_i^a = 6 \]
**User: Friends Topic Diversity**

- Define a user topic interest vector in terms of the Digg-defined categories.
- Each component of represents the fraction of all votes made by user on stories belonging to that category.
- Define topic diversity of a user's network by averaging pair-wise cosine distances of friends' topic interest vectors.

\[
TD_i = \frac{\sum_{j=1}^{N} \sum_{k=1}^{N} (1 - \cos(\theta_j, \theta_k))}{S_i^2}
\]
Information: Novelty & frac. of adopted Novel info

- Total amount of novel information is an important measure of information content of networks.
  - Max Novel Information is not the same as Max Information

- The same message or a piece of information can be recommended multiple times by multiple friends.

- Use unique id to measure novelty
  - Original tweet id (Twitter)
  - Story id (Digg)
Information: Novelty & frac. of adopted Novel info

Friends of friends of i: \( N_{\{Ni\}} \)

Friends of i: \( N_i \)

User i

• Novel information
  \[ NRI_i = \bigcup_{j=1}^{N_i} j \]
  • Novel information friends are exposed to
  \[ NRI_{i\text{frds}} = \bigcup_{j=1}^{N_{\{Ni\}}} j \]
  • Fraction of novel information adopted by user
  • Fraction of novel information adopted by friends
## Experiment Results: Dataset

<table>
<thead>
<tr>
<th></th>
<th>2009 Dataset</th>
<th>2010 Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Jun 2009 (1 month)</td>
<td>Jul – Dec 2010 (6 months period)</td>
</tr>
<tr>
<td># of users</td>
<td>70K active users</td>
<td>12K users</td>
</tr>
<tr>
<td># of links</td>
<td>1.7M</td>
<td>1.3M</td>
</tr>
<tr>
<td># of stories</td>
<td>3.5K (promoted to Front page)</td>
<td>48K</td>
</tr>
<tr>
<td># of votes</td>
<td>2.1M</td>
<td>1.9M</td>
</tr>
</tbody>
</table>

We examine only the votes that the story accrued before promotion to the front page.
Experiment Results: Dataset

We examine only the votes that the story accrued before promotion to the front page.
How structure and activity shape information access: Effect of Network Size

- The study of email communication within a corporate recruiting firm (Aral & Van Alstyne)
  - Total volume of novel info flowing to recruiters and its diversity increased with their network size, network diversity and channel bandwidth.

- For the online social network of Digg.
  - The same conclusion hold for the online social network?
  - Whether Size of network (S), Diverse structure (ND), High friend activity (B) are likely deliver more Novel information (NRI) and Topic Diversity (TD).

- The simplest ways users can control their position within a network is by adding friends.
  - Does having more friends improve access to information in online social networks?

- The highest number of distinct stories any user was exposed to was 29,558 (~80% of the total)
- Adding more friends in an online social network improves access to novel information
- However, after 100 friends, it becomes counter productive
  - Doubling the number of friends raises the volume of novel info only a few percent with info overload!
How structure and activity shape information access: Effect of User Activity

- Does having active friends lead to greater access to novel info?
  - Potential amount of novel info rises quickly as a function of friend activity
  - However, the amount of novel info to which the user is exposed is just a fraction of this max
  - Both decrease as friend activity grows past 2000!

- The fraction of novel information in the stream as a function of the average activity of their friends.
  - The fraction of novel info drops precipitously as friends become more active.
Effect of Network Structure: the “Diversity-Bandwidth Trade-off”

- We study the interplay between network structure and user activity and their impact on access to info.
- “Diversity-Bandwidth Trade-off”: Structurally diverse networks provide greater access to info, their benefits are offset by lower rate of communication along structurally diverse ties.

- Does it exists on Social media?
  - Divide users into different populations based on the total volume of incoming info.
  - Strong negative relationship (-0.47, p<.01) between network diversity (ND) and bandwidth (B):
    - $S > 322$: -0.54, p<.01
    - $322 \geq S > 130$: -0.58, p<.01
    - $130 \geq S$: -0.50, p<.01

- Diversity-Bandwidth Trade-off in Digg
  - Users who place themselves into positions of greater network diversity within the social network on average receive fewer story recommendations from friends that those who place themselves into positions of smaller network diversity.
Effect of Network Structure: the “Diversity-Bandwidth Trade-off

- Contrast to the finding of Aral & Van Alstyne
  - Email communications: users could increase both the topic diversity and amount of novel information by increasing either their network diversity or channel bandwidth
- Topically diverse information (TD)
  - Users in positions of higher network diversity can access more topically diverse information (TD)
- Novel Information (NRI)
  - Users whose friends are more active can access more novel information (NRI)

<table>
<thead>
<tr>
<th>Year</th>
<th>NRI</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.04*</td>
<td>-0.11*</td>
</tr>
<tr>
<td>B</td>
<td>ND</td>
<td>0.48*</td>
</tr>
<tr>
<td>2010</td>
<td>0.69*</td>
<td>-0.83*</td>
</tr>
<tr>
<td>B</td>
<td>ND</td>
<td>0.41*</td>
</tr>
</tbody>
</table>
How structure and activity shape information access

- Characteristics of users who inject new information into their networks by voting for stories they found outside of their friends’ recommendation.
  - The size of symbol represents the relative number of seeded stories.
  - Divide users into two classes based on the number of friends they have.
- Users with many friends who are very active inject relatively more new stories into their network than users with many but less active friends.
  - Also they are more strongly tied to their friends.
  - It seems counterintuitive that these users, who already receive more recommendations, would take the time to look for new information.
- Users with few friends also tend to have less active friends.
Bottlenecks to Information Access

- Why do users in positions of high network diversity receive less novel info even though they are connected to more topically diverse friends?

- The total amount of novel information that friends of user are exposed
  - In most cases, friends are collectively exposed to a large quantity of novel info, almost all of 36,883 distinct stories.
  - Although most of the users could potentially be exposed to nearly all of the information in the network, they receive far less novel information.
  - Friend Novel Info. Adoption Rates shows
    - Friends of users in positions of high network diversity fail to adopt most of the novel info
    - Users with highly active friends are exposed to more novel info because of higher adoption rates of friends.
Bottlenecks to Information Access

- Cognitive constraints
  - User activity as a function of the number of active friends the user has
  - Limit user’s ability to process incoming information.
  - Similar to the limit the number of stable social relationships a human can maintain.
  - It also has been shown to limit the number of conversations users engage and the spread of information.
Related Works

+ Network structure and the intensity of interaction between people
  - The strength of weak ties [Granovetter 73]
  - Empirical studies in job search, business relations, social capital
    [Uzzi 97, Allen 03, Reagans & Zuckerman 01]
  - Communication in mobile phone and email
    [Onnela et al. 07, Iribarren & Moro 11, Aral & Van Alstyne 11]
+ Relationship between weak ties, structure diversity and access to diverse and novel information
  [Centola 10, Grabowicz et al. 11, Centola & Macy 07, Zhao et al. 10]
  - Email communication within an organization [Aral & Van Alstyne 11]
  - Demonstrate a trade-off between network diversity and channel bandwidth [Aral & David 12]

+ Homophily
  - People tends to be friends with other who belong to the same socio-economic class [Feld 81, Mcpherson et al. 01]
  - People tend to follow who have similar interests [Kang & Lerman 12]

+ Cognitive constraints
  - The number of social relations that people can maintain [Dunbar 03]
  - Limit of the number of conversation partners on Social Media [Goncalves et al. 11]
  - Divide of attention on Social Media [Hodas & Lerman 12]
Conclusion

- Digg: Users in positions of high friend activity increase their access to novel info since their friends adopt a large portion of the novel info.
- Digg: Users in positions of high network diversity are exposed to more diverse information, but since their friends have interests that are different from their own, they do not adopt much of the info they are exposed to.
- Difference between our study and the findings of Aral & Van Alstyne.

- Email: Users could increase their access to diverse and novel information by increasing their network diversity or channel bandwidth.
- In addition, in Aral & Van Alstyne’s study, novelty and diversity were not independent variables: non-redundant information was the product of topic diversity and channel bandwidth. Hence, it’s not surprising that both were correlated highly with network diversity and channel bandwidth
  ➔ We treat them independent variables.