Factors Impacting Community Response in an Interest-Sharing Network

Iris Howley
Carnegie Mellon University
Pittsburgh, PA 15213
ihowley@cs.cmu.edu

Todd Newman
Microsoft Research
Redmond, WA 98052
toddne@microsoft.com

ABSTRACT
The arrival of a new interest-sharing network, So.cl, provides for a new opportunity to explore human behavior as it relates to constructing public contributions and receiving community response. This study looks at archival data in order to better understand how types of shared content receive interaction from others. The results suggest that a So.cl user should include more photos and less links on their post to increase the quantity of likes and comments the community gives to the post, among other discoveries.

Author Keywords
Social networks; Log analysis;

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Group and Organization Interfaces.

INTRODUCTION
The surge of popularity in social networking websites provides an opportunity to better understand human behavior and how knowledge people choose to share with their online community receives or does not receive response from other members. Given the importance of knowledge and information within communities [8] and the importance of response and interaction to retaining membership in these websites [2], it is worthwhile to explore what types of information are being shared on an informal knowledge-sharing website and how that impacts membership response.

So.cl[1] has risen as an interest-networking site for supporting informal learning by combining "browsing the web, sharing links, connecting with people through what they share, and learning and ultimately gaining expertise" [3]. So.cl provides a lightweight sharing mechanism through which members can create posts based upon their search engine and member queries, attach a message, photos, links, and videos. The community then publicly responds with one-click "likes", comment responses, and new posts based on previous, inspirational posts. While So.cl is described as an interest network for informal learning that leverages the social aspects of learning, social tagging, and features of electronic communities and interest networks, there are still many questions to be asked about how members use and respond to these tools. In this paper, we use log analysis to examine what information is being shared and how the community is interacting with that knowledge in order to gain a better understanding of how to improve user engagement in online communities.

CHI 2013, April 27–May 2, 2013, Paris, France.

BACKGROUND
We know from past research on the popular microblogging service, Twitter, that relationships in this social networking site do not mimic known characteristics of human social networks, and that the majority of trending topics shared within Twitter are headline news [5]. However, the Twitter community cultivates knowledge beyond news taglines, as users turn to the Twitter community to do everything from ask advice to filing suggestions to companies [6]. A similar variety of content is being shared on other social networking websites as well, such as Google+. Kairam et al [4] interviewed users about their sharing behaviors and found that users share: information they deem to be valuable, information about themselves, requests for help, attempts to start a discussion, and information to create awareness about a topic, among other purposes. We know that knowledge is being shared on So.cl, but we do not yet know what types of knowledge are being shared.

Knowledge is valuable within communities of practice, making these information-sharing behaviors important for developing an online community of experts. Electronic communities of practice members are motivated to participate because they appreciate discussion around interests and interaction with the community [8], an activity that simple question and answer sites do not always fulfill. While there is community interaction occurring on So.cl, we do not yet know the quantity of interaction, or where that interaction is centered.

Teevan et al [7] explored a related series of research questions regarding information-seeking and community feedback on Facebook. The post features explored included punctuation, number of sentences, and scoping. Stating the information request as a single sentence question and scoping the audience received the best response. These results are similar to those of [2] which explored community response within Usenet groups and only examined self-disclosing introductions and making requests. In this Usenet study, introducing oneself to the group and requests were better correlated with responses than message length or newsgroup traffic.

SO.CL

So.cl[1] is a recently public, Twitter-style interest-sharing network, with a similar public feed and follower model. However, instead of a limit of 140 characters, So.cl members incorporate curated search engine results into their shared “post”, along with multiple photos, links, tags, and longer messages, as shown in Figure 1. Once the post has been published to the public feed, other community members may then click, comment, riff, or like the post.

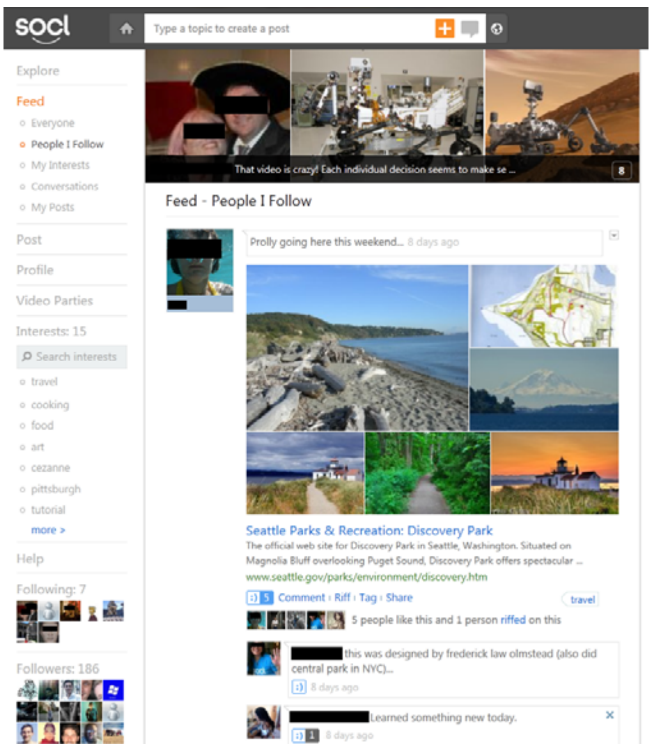

Figure 1. A screenshot of So.cl showing the user’s following feed. Users construct a post using the search bar at the top, and then curate photos and links from the results to include with their post.

When a user performs a search within So.cl, the results include both matches from an Internet query as well as results from other users within the community. From this point, the user may elect to create a post using curated results from the search query, or to click on user-created content that was returned in the So.cl query. Users can easily explore content that is being shared at this moment, as well as filter the global feed based upon users they follow, or upon common interests.

[1] http://www.so.cl/

METHOD

Post data was gathered from So.cl internal data stores from July 13-August 13, 2012. All user actions are stored in these data stores, and they were reorganized for analysis purposes. Gathered data include, amongst other items: queries, post message, links, photos, videos, comments, likes, and clicks on posts.

Types of Posts

We split the types of posts into three dimensions: structural features, content features, and post intentions.

Structural Features

First, we gathered specific information about structural objects included with the So.cl post. This includes:

- Post Type: Whether the post is a query, a status (a post without a query performed), a search (a query without an attached message), a link (a query for a specific URL without a message), or a comment.
- Queries: The Bing queries performed, if any, in curating objects for the post. Members of the So.cl community cannot see this information.
- Tags: All tags added by the original post author, or other users of the site.
- Language: For the purposes of this study, we are limited to English-language only posts.
- Messages: Both the text of the message, as well as the number of words in the message. Not all So.cl posts include a message (see Post Type).
- Photos, Links, and Videos: So.cl splits Bing query results into types of items, including photos, links, and videos. The URLs for each of these items have been incorporated into our dataset.

Content Features

One of the most obvious distinguishing features between posts is the topic of the post. Latent Dirichlet Allocation (LDA) models are used in a variety of situations to automatically discover topics on large quantities of unlabeled data [1]. In our study, we apply LDA topic modeling to find topics that each post is most associated with. A single post may actually represent several topics, and LDA represents this by constructing a distribution with proportions across possible topics for each post. To use the LDA model, we concatenate queries and tags for each post together as they introduce less noise than message text. Additionally, one must specify a number of topics to generate, and as our corpus is rather large, we set this value to 25. However, we used human consensus of two coders to reduce the number of topics to 13, as several automatically generated topics appeared to center around genres such as “food”, “travel”, or “tech”, so our final 13 topics were:

animals, art, events, fashion food, games, movies, music, photography, reading, so.cl, tech, and travel. Any post that had a topic likelihood less than a low threshold was discarded for this analysis.

Post Intention

A more subtle dimension on which posts can be categorized is the author's perceived intention in the post. In order to determine perceived post intention, we developed a coding manual with the following intentions:

- Knowledge-Sharing: The main intent of the post is to share a fact or piece of knowledge.
- Question: The main intent of the post is to ask a question or seek information.
- Humor: The main intent of the post is to be humorous, or cause a laugh.
- Promote: The main intent of the post is to get additional followers/readers of the author's product. Self-promotion.
- Welcome: The intent of the post is to announce one's arrival or departure and other phatic messages.
- Beautiful: The main intent of the post is to share something that is pretty, beautiful, or awe-inspiring.
- Cute: The main intent of the post is to share something adorable or endearing rather than beautiful.
- Advice: The main intent of the post is to share an inspirational saying or piece of advice.
- About Me: The main intent of the post is to discuss happenings in the author's life.
- I Want: The main intent of the post is to share something alluring, that the author wants.

We developed the coding manual through a series of refinements using multiple coders and training data to achieve an inter-rater reliability Cohen's Kappa score of 0.76, which is above the acceptable threshold of 0.7. Only "query" and "status" posts were coded, as they included a written message from the post author.

Community Interaction

Community interaction variables were collected similarly to the Post Objects data, as they are stored similarly. Community interaction data includes:

- Clicks: Every time a user clicks an object associated with a post, that data is stored. These are stored privately.
- Likes: A user may perform a one-click action known as a "like", although it is displayed as an emoticon in So.cl.
- Comments on the post: We gathered the text for the comments on each post.

RESULTS

In this section, we describe our analysis of post features and community response. We analyzed posts from 14,000 active So.cl users in the one month period of July 13-August 13, 2012. An "active user" is any user who either created a post or interacted with another's post during that one month period. On May 18, 2012 So.cl changed its membership policy such that potential members no longer needed an invitation to join the website and memberships were now publicly available. 60% of the public memberships from our sample were created through a Facebook profile and the remaining 40% used their Windows Live account.

Structural Features Results

As a sample of what posts contained, for posts that could include multiple links, photos, and tags (i.e., queries and riffs), the median values for each post included 0 links (0.65µ), 1 photo (2.9µ), and 1 tag (0.96µ) added by the user. These same posts had a median number of 1 like (1.5µ), 0 clicks (0.58µ), and 0 comments (0.34µ). Structural features of a post's ability to predict community response was determined through a step-wise regression. The number of links, photos, and tags on a post were significantly predictive of the number of likes a post would receive, as shown in Table 1, with an R^2 value of 0.19. However, these results are complemented by a significant interaction with Post Intention, described later. Similarly, structural features of a post significantly predict the number of comments and clicks a post receives, but those R^2 values were 0.026 and 0.063, respectively.

Variables	Beta	Std Error	t	p
Intercept	1.10	0.02	50.77	< .0001**
# links	-0.88	0.03	-31.15	< .0001**
# photos	0.87	0.02	52.39	< .0001**
# tags	0.34	0.03	12.97	< .0001**

Table 1. The regression statistics for predicting number of likes on a So.cl post, with a logarithmic transformation performed on continuous independent variables.

Post Intention Results

A sample of 1437 query posts was coded for post intention resulting in: 141 "about me" posts, 163 "beautiful", 26 "cute", 75 "humorous", 25 "inspirational", 77 "I want", 433 "knowledge-sharing", 69 "questions", 284 "recommendations", 49 "self-promotions", and 95 "welcome" posts. Similar to the structural features results, post intention significantly predicts number of likes on the post, $F(10, 1146) = 14.30$, $p < 0.0001$ with an R^2 value of 0.11. A post-hoc analysis shows that beautiful, cute, want, and humor posts receive significantly more likes than about me, question, recommendation, knowledge-sharing, and self-promotion posts, with inspirational posts not really being indistinguishable, likely due to their low number of occurrences in the sample.

Post intention also significantly predicts number of comments, $F(10, 1146) = 18.53$, $p < 0.0001$ with an R^2 value of 0.14. A post-hoc analysis shows that welcome posts get significantly more comments than question, cute, about me, humor, and beauty posts which receive more comments than "I want" posts, recommendations, self-promotion, and knowledge-sharing posts with inspirational posts being indistinguishable again. Similar to the structural features results, the p-value for predicting number of clicks is significant, but the R^2 value is low at 0.05.

*Structural Features * Intention Results*
A significant interaction between number of photos attached to the post and the post intention was found for number of likes, $F(21, 1446) = 21.83$, $p < 0.0001$ with an R^2 value of 0.24. About me and recommendation posts significantly increased in number of likes as number of photos increased. A similar significant interaction was found between number of links and post intention, $F(21, 1446) = 9.5$, $p < 0.0001$ with an R^2 value of 0.12. While most post intentions experienced a positive increase in likes with the increase in links, humorous posts received marginal decreases in likes.

Looking at predictors of number of comments, there was a significant interaction between post intention and number of links, $F(21, 1446) = 12.86$, $p < 0.0001$ with $R^2 = 0.16$. An increasing number of links resulted in a significantly decreasing number of comments for knowledge-sharing, recommendation, and beautiful posts. The interaction term for post intention and number of photos was also significant for determining the number of comments on a post, F (21, 1446) = 12.56, $p < 0.0001$, with an R^2 value of 0.16. Recommendations received a significant increase in number of comments with an increase in number of photos.

The interaction term for predicting number of clicks is also significant, but the R^2 values are below those already reported, and will not be discussed in this paper.

Content Features Results
Our LDA topic modeling algorithm included 13 topics: 190 "animal" posts, 96 "art", 119 "events", 270 "fashion", 465 "food", 471 "game" posts, 92 "movies", 575 "music", 1841 "photography", 86 "reading', 211 "so.cl", 435 "tech", and 473 "travel" posts. There was a significant effect of post topic on number of likes, $F(12, 5323) = 12.48$, $p < 0.0001$, $R^2 = 0.03$. Looking at number of comments and number of clicks, the R^2 values reported were even lower.

DISCUSSION
The goal of this log analysis was to discover what types of posts in an interest-sharing social website experience the most community response. On So.cl, there is a large quantity of knowledge-sharing and recommendation posts, which aligns with the website's goal of informal learning. However, these post intentions interact with the number and kinds of objects attached to the post. Most importantly, posts containing recommendations increase in number of received likes and comments as photos go up, but decrease in number of comments as number of links increases. In short, a So.cl user should include more photos and less links to increase the quantity of likes and comments the community gives to the post. These results can be leveraged to help members of interest-sharing networks increase the community response they receive on their contributions.

It is worth noting that correlations between community response and number of clicks were difficult to find. This suggests that there are other features of posts that may influence other members' clicking behaviors. Additional models in this study, including post topic content have reported low R^2 values, which also suggest other variables impacting community behavior. Future work includes examining possibly important, but omitted, variables including user social networks, homophily with other members, and the interestingness of the post content.

As analyses were performed on archival data and there was not an experimental manipulation, some hidden explanation for these results may exist. Furthermore, the dataset included data from a single interest-networking website and so it may not apply to other social networking communities.

ACKNOWLEDGEMENTS
This work was supported in part by a Training Grant awarded by the Department of Education (#R305B040063).

REFERENCES
1. Blei, D., Ng, A., and Jordan, M. Latent dirichlet allocation. *Journal of Machine Learning Research* (2003), 993-1022.
2. Burke, M., Joyce, E., Kim, T., Anand, V., and Kraut, R. Introductions and requests: Rhetorical strategies that elicit response in online communities. Communities & Technologies (2007).
3. Farnham, S.D., Lahav, M., Raskino, D., Cheng, L., Ickman, S., and Laird-McConnell, T. So.cl: An interest network for informal learning. In *Proc. ICWSM 2012*, AAAI Press (2012).
4. Kairam, S., Brzozowski, M., Huffaker, D., and Chi, E. Talking in circles: selective sharing in Google+. In *Proc. CHI 2012,* ACM Press (2012), 1065-1074.
5. Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? In *Proc. WWW 2010*, ACM Press (2010).
6. Miller, C. C. Putting Twitter's world to use. *The New York Times*. April 13 (2009).
7. Teevan, J., Morris, M.R., and Panovich, K. Factors affecting response quantity, quality, and speed for questions asked via social network status messages. In *Proc. ICWSM 2011* (2011).
8. Wasko, M.M. and Faraj, S. "It is what one does": why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems 9*, (2000), 155-173.