SI 686 (User Generated Content) Final Report: Analysis of Sumazi

James "Augie" Hill
University of Michigan, Ann Arbor

February 24, 2012

Abstract

Sumazi, a young community established to efficiently realize interpersonal connections, is described and analyzed. The community fills a general yet unfilled niche in the online community ecosystem. Summary statistics of connection requests are reported, suggesting significant free-riding behavior, defined in this context to be posting connection requests without recommending or broadcasting others' requests. Finally, it is recommended that summary measures of user activity are made publicly available to discourage free-riding, and that the cognitive costs of making recommendations are minimized with improved tools.

1 Introduction

Sumazi [1] is a community for interpersonal connection discovery. *Interpersonal connection*, referenced for the remainder of the paper simply as *connection*, is a general term enveloping social, professional, common identity, and potentially other types of relationships between two people. An example of a social connection is a mutual romantic interest. An example of a professional connection is a complementary employment need in which one person has an available job and the other has a need for a job and meets the job qualification requirements. An example of a common identity relationship is a mutual interest in a genre of music. Clearly, the scope of potential connections is quite wide.

The problem Sumazi addresses is formally described by an undirected multi-edge graph in which nodes are people and edges are connections. Each node has certain attributes. Each connection has one boolean attribute: whether it has been realized. Based on a node's attributes, there are a finite number, however numerous, of potential connections. Depending on the type of attribute, these potential connections may be interchangeable, where realizing one or some of the connections causes other potential connections to be removed. Through chance encounters, a small subset of potential connections are ever realized. The goal of Sumazi is to efficiently discover and realize potential connections. While such a graph construction is in actuality highly dynamic, we consider only the case of a static graph, a snapshot of the dynamic graph at some point in time, which generalizes well to the dynamic case by severing connections that become irrelevant and generating new potential connections when attributes change. At any moment, the problem of identifying potential connections is addressable in the static setting.

In Section 2, other communities attempting to address connection discovery are described. In Section 3, the design of Sumazi is described in detail. In Section 4, statistics describing the effectiveness of Sumazi are listed. In Section 5, the design of Sumazi is analyzed from a social psychological and game theoretic perspective. In Section 6, recommendations for modifications to the design of Sumazi are presented.

2 Related Communities

Previously constructed communities have approached the connection discovery problem in a domain-specific manner. Examples of such communities are LinkedIn [2], the domain of which is professional, and OKCupid [3], the domain of which is social-romantic. Interestingly, no existing sites could be found which specifically address the discovery of purely common identity connections. This may be explained by the observation that the type of attachment exhibited in common identity relationships is more strongly directed towards the community of people with a common identity as a whole than towards individual members of that community [4], from which it follows that of greater importance than discovering common identity connections is discovering communities of common identity such as DIY Drones [5], a community of autonomous aerial robot enthusiasts.

The specificity of online communities can be very fine-grained, a result of the negligible transaction costs of global communication on the Internet [6], as is evidenced by a wide variety of specialized romantic connection discovery communities. The reasons these stand-alone hyperspecific communities exist provide insight towards the analysis of Sumazi. The marginal cost of founding a specialized community on an existing more general community system is near zero, yet many communities expend the effort in constructing their own independent systems. A plausible explanation of this behavior is that the general-domain community systems are unable to informationally express the segmentation of the community. That is, the attributes of a specialized community that make it unique cannot be encoded in a more generic community information system. Arbitrarily complex community specialization makes such a task difficult.

The existing popular community with perhaps the most general domain specificity is Craigslist [9], which segments posts hierarchically, with the initial segmentation by geographic area. The segmentation of posts on Craigslist into hyperspecific categories along with the requirement that posts be placed in one of those categories means that any posts that do not fit into one of those categories do not have a place in the community. This is an example of the inability of some communities to design their systems to be expressive enough to enable specialized usage. Another online community, though not purposefully an interpersonal connection discovery community, that has seemingly accomplished such expressiveness is Reddit [7], which boasts over 110,000 specialized communities for sharing Internet content [8]. Perhaps equivalently effective expressiveness is also possible in connection discovery communities, but simply has not yet been implemented.

3 Design

Sumazi leverages the Facebook open development platform [10] with many beneficial effects. Users may sign up to use Sumazi with their Facebook credentials, which reduces the cost of transferring identifying information to only a few mouse clicks. In addition, the interpersonal connections the user has already curated on Facebook are automatically available to the user on Sumazi. Facebook requires that accounts are created using only real names and expends effort ensuring all accounts are linked to real-world identities, so the identities of users on Sumazi are transitively tied to users' real-world identities.

Registered users may post requests for introductions to people that meet some criteria. A post consists of a brief description, a predefined category, and optionally a location. Once created, the post is automatically shared on the posting user's Facebook feed. All posts are publicly accessible from a unique URL.

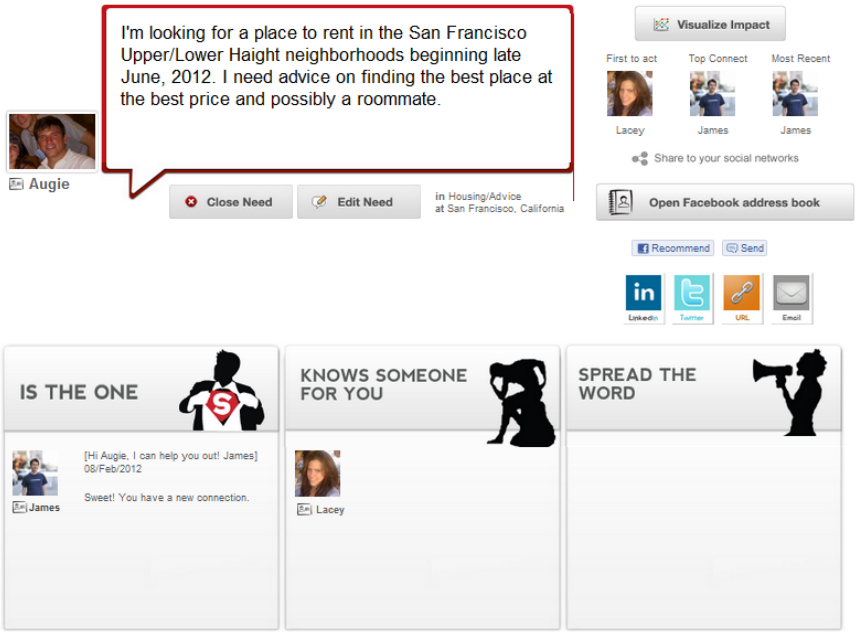

Figure 1: Interface of a post from the perspective of the creator of the post.

The database of posts is publicly searchable and filterable. A user can choose to see only their friend's posts, to see all posts from one of the categories, or to see all posts for a location. When searching the database, posts are presented in a shorter format that includes only the name and photo of the user that created the post and however many characters of the description that fit in the available space. Clicking through to a post-specific page makes available all information on a post.

One of three primary actions may be taken by a user viewing a post. Users may claim they meet the criteria described in the post, may recommend the post to a Facebook friend, or may spread the word by sharing the post on their Facebook feed. There are additional secondary actions available to a viewing user. She may ask a private question of the user that posted the request, and she may also privately flag the post as inappropriate. When a post is flagged, the Sumazi administration team examines the post and possibly removes it.

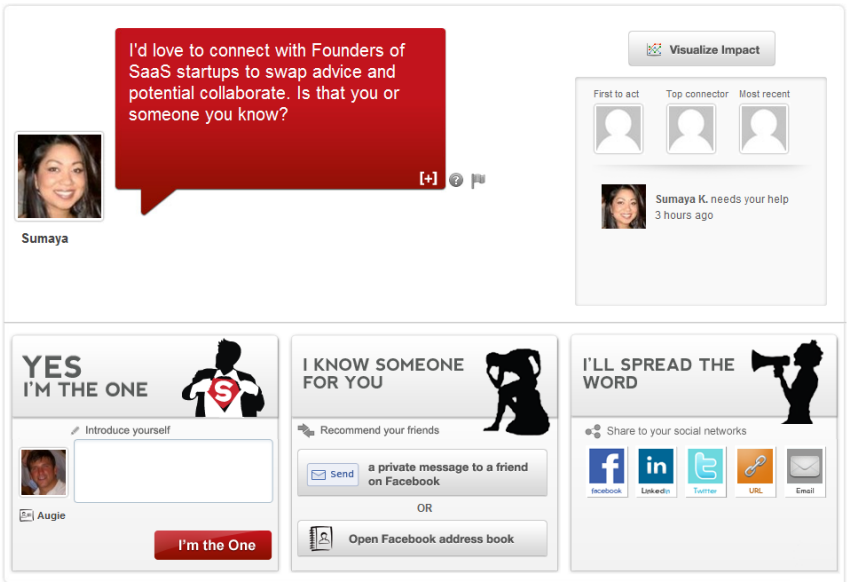

Figure 2: Interface of a post from the perspective of a user that did not create the post.

All primary actions taken on the site are public information. This information is made available in several ways. Information available when viewing a post are the history of actions, the user that most recently acted on the post, the user that shared the post the most, the first user to claim she meets the criteria, and a tree graph representation of the history of actions in which each node is a person and each edge is an action.

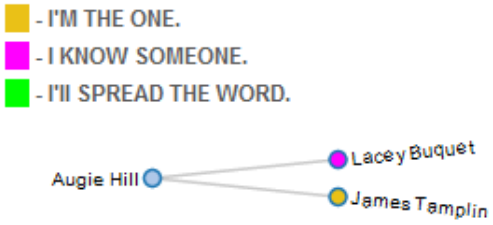

Figure 3: Visualization of the history of primary actions made on a post. In this post, user *Augie Hill* is the creator of the post, *Lacy Buquet* recommended the post to a friend, and *James Tamplin* claimed to meet the criteria described in the post.

A user may edit any aspect of one of the posts she created. She may also close the post, which signals that the post has either been fulfilled or is no longer relevant. When a post is fulfilled, users communicate outside of Sumazi.

4 Statistics

Sumazi launched its Beta tool late September, 2011. I wrote a script to download all of the available posts. The posts that could not be downloaded may or may not have been fulfilled, and will be disregarded for this analysis. 1,184 posts had been created, of which 930 were available.

Sumazi employees have made many contributions to the community during this early phase of its existence. However, their motives are not the same as the motives of ordinary users due to the financial reward they would receive with the success of the community. Thus, Sumazi employees are excluded from the statistics and analysis as much as possible. This is not entirely possible due to the exclusionary effects of some actions, but is nonetheless important to attempt. In this study, posts by employees were included, but actions by employees were not.

Figure 4 describes the actions taken over time by non-employees. Posts were divided into four cohorts, each of size 296, which are graphed on the x-axis. The left-most point of the graph summarizes posts 1 through 296, the second-from left summarizes posts 297 through 592, etc. The number of accessible posts was not equal among the four cohorts, so each statistic is normalized by the number of accessible posts in the cohort, resulting in comparable statistics.

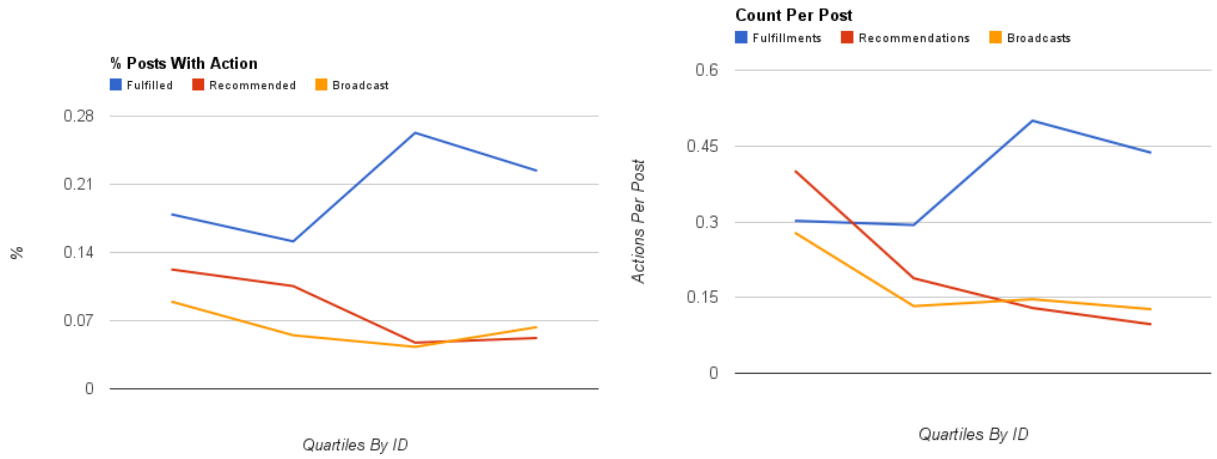

Figure 4: (Left) The percentage of posts in the cohort that have had a certain type of action taken on it. (Right) The number of each type of action taken on posts on average in the cohort.

The following table groups posts by location. The "[Other]" location is a grouping of the long list of locations with few posts.

Location	% Posts (#)	% Fulfilled (#)	% Recommended (#)	% Broadcasted (#)
[Other]	0.47 (435)	0.22 (95)	0.12 (52)	0.124 (54)
	0.22 (204)	0.235 (48)	0.1 (21)	0.07 (14)
San Francisco	0.21 (194)	0.31 (60)	0.19 (37)	0.14 (28)
New York	0.05 (46)	0.24 (11)	0.11 (5)	0.11 (5)
Los Angeles	0.028 (26)	0.19 (5)	0.19 (5)	0.15 (4)
California	0.027 (25)	0.08 (2)	0.24 (6)	0.28 (7)

Figure 5: Table of post statistics grouped by location, sorted by number of posts.

The following table divides posts by top-level category and provides statistics on each. Categories are in reality more fine-grained than shown here, but have been truncated and grouped in to coarser-grained groups for comparison.

Top Category	% Posts (#)	% Fulfilled (#)	% Recommended (#)	% Broadcasted (#)
Business	0.33 (307)	0.23 (71)	0.12 (37)	0.12 (36)
Career	0.25 (232)	0.22 (51)	0.125 (29)	0.1 (24)
Activity	0.126 (117)	0.22 (26)	0.21 (25)	0.15 (18)
Other	0.101 (94)	0.23 (22)	0.25 (24)	0.2 (19)
Nonprofit	0.06 (55)	0.42 (23)	0.36 (20)	0.4 (22)
Academics	0.039 (36)	0.22 (8)	0.14 (5)	0.05 (2)
Housing	0.026 (24)	0.125 (3)	0.17 (4)	0.17 (4)
Journalism	0.017 (16)	0.44 (7)	0.19 (3)	0.19 (3)
Health Care	0.013 (12)	0.25 (3)	0.08 (1)	0.33 (4)
Love	0.012 (11)	0.18 (2)	0 (0)	0 (0)
Orrick Total Access	0.011 (10)	0.3 (3)	0.2 (2)	0.2 (2)
Human Rights Conference	0.005 (5)	0.4 (2)	0 (0)	0 (0)
Taos IT Consulting	0.005 (5)	0.2 (1)	0 (0)	0.2 (1)
Politics	0.004 (4)	0.25 (1)	0 (0)	0.25 (1)
	0.002 (2)	1 (2)	0 (0)	0 (0)

Figure 6: Table of post statistics grouped by category, sorted by number of posts.

5 Analysis

5.1 Players

User roles relating to a particular post are summarized by the following list.

- **Requester**: the user that posts the request.
- **Recommender**: a user that recommends the post directly to a friend. The recommender may know that the recommendee may fulfill the request, or that the recommendee knows others that may fulfill the request.
- **Broadcaster**: a user that broadcasts the post to all of her friends.
- **Fulfiller**: a user that claims to meet the criteria described in the post.

The only strict separation between roles is between the requester and fulfiller; if a requester could fulfill their own request, she would not be posting the request in the first place. A requester is also a broadcaster when she shares the post on her Facebook feed, and may also be a recommender if she knows specific friends that are potential recommenders. A fulfiller may also be a recommender and a broadcaster when the request may be fulfilled by more than one user.

5.2 Requests

Different types of requests have different incentive properties. The dimensions along which requests vary, compiled by examining posts on Sumazi, are the following.

- **Direct Utility Symmetry**: the symmetry of positive utility when the request is fulfilled. Employment is an example of request fulfillment with symmetric positive utility (*symmetrical*), whereas advice is an example of request fulfillment with asymmetric positive utility (*asymmetrical*).
- **Count**: the number of users needed to fulfill the request completely.

It is evident from examining posts that users tend to post from both sides of a symmetrical request. For example, there are posts for employment opportunities and posts in search of employment. Requests with asymmetric utility are apparently only posted by the user that would receive positive utility, as would be expected.

5.3 Identity

By utilizing users' Facebook accounts and transitively users' real-world identities, users are disincentivized to cause harm or contribute low quality content to the community, especially considering that the posts and all actions on posts are public information, which is usually an issue in communities of contributed content [11]. Sumazi utilizes common bond associations, boostrapped from users' Facebook friends, to discover common identity and complementary need connections, which likely incentivizes users to self-police the quality of their recommendations.

5.4 Graph Analysis

Recall the graph described earlier that formally represents the interpersonal connection discovery problem. Each node, representing a person, is described by a set of attributes. Based on the attributes of every person in the graph, potential connections exist between nodes. Nodes have complete information about their own attributes, and partial information about the attributes of the other nodes to which they are connected. The *distance* between two nodes is the length of the connected path from one node to the other. If there does not exist a connected path from one node to the other (the nodes are *disconnected*) then the distance between them is infinite. A more thorough discussion on social graph construction is available in [12].

The likelihood that a post is sufficiently fulfilled is a function of the persons that encounter the post, the *quality* of the persons encountering the post, and the potential connections of the underlying graph. A person's *quality* is a function of the *connectedness* of the person (the number of people to which that person is directly connected), the person's knowledge of neighbor attributes, and distance (negative) to a potential fulfiller. People with higher quality have a higher probability of being able to help fulfill the request. Because the potential connections are a natural result of the node attributes, the only ways to increase the probability of fulfillment are to increase the number of encounters and the quality of people encountering the post. People may encounter posts in one of three ways:

- **Independently**: an encounter on Sumazi independent of actions taken by friends.
- **Broadcasted**: encountered in a friend's broadcast feed.
- **Recommended**: recommended by a friend.

When a requester creates a post, recall that she simultaneously broadcasts it to her friends. She may also recommend the post to particular friends. Without independent encounters, the *reach* of the post is some distance from the requester, a function of the rates of rerecommendation and rebroadcast. That is, the expected distance from the requester to which the request will be passed is finite, and is limited to people connected to the requester, which makes potential connections between disconnected people unrealizable. When independent encounters are introduced, however, infinite distance between potentially connected people no longer precludes request fulfillment, as random independent post encounters may jump the gaps.

Sumazi users are some subset of the nodes in the graph, but each are connected to a larger portion of the underlying graph (assuming the users do not form a clique). An independent encounter of a post can only occur for Sumazi users. Thus, increasing the number of users increases the potential reach of each post. Assuming some number of people (not Sumazi users) that encounter a post by broadcast or recommendation convert to regular Sumazi users, then increasing the number of actions taken by users will increase the number of users. Therefore, increasing both the likelihood that a user will independently encounter a post and the likelihood that a user will act on a post will increase the reach of every post and thus also the likelihood that every post will be fulfilled, resulting in a positive feedback loop commonly termed the *network effect*.

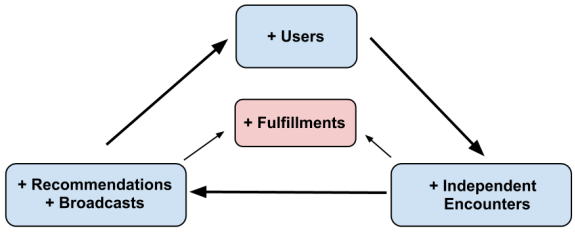

Figure 7: An illustration of the Sumazi positive feedback loop/network effect.

5.5 Individual Incentives

5.5.1 Requester

The requesting user has recognized some need in their lives that can be fulfilled with the help of another person. By posting on Sumazi, the user can only increase the probability that this need is fulfilled, for which the user receives positive utility. Thus, there is no obvious downside to posting a request, and the negative utility of doing so is the cost of making the request. The fact that Sumazi makes the task of posting a request very easy results in a very low cost.

5.5.2 Fulfiller

For symmetric requests, fulfillment returns positive direct utility to the fulfiller. The incentives of the fulfilling user in this case are obvious. For what percentage of posts is there symmetric utility? Unfortunately, an exact answer would require examining all posts manually, but perhaps a heuristc can be used. Consider the statistics presented earlier summarizing posts by category. Categories in which one could presume there are usually symmetric positive rewards are Business, Career, Activity, and Love. Together, these four categories make up about 70% of posts. Cursory review of posts from other categories suggests that the ratio of symmetric requests to asymmetric requests is very high.

For asymmetric requests, fulfillment returns no direct utility to the fulfiller. Why would a user fulfill a request for which she receives no obvious direct utility? Clearly, there must be non-obvious social psychological contributors to the utility function of such a user. Consider the statistics of the Nonprofit category. The rates of fulfillment, recommendation, and broadcast in this category are almost double the rate of every other category with enough posts for comparison, while fulfillers are most certainly not receiving direct utility. Perhaps the fact that all actions taken on posts are public and people are, and like to appear to be, altruistic [13], combine to incentivize users to commit to fulfilling a nonprofit request. Altruism and reputation may also play a less pronounced role in other types of asymmetric requests, such as advice-giving.

5.5.3 Recommender

There is no direct utility for recommending a post to a friend. What kinds of indirect utility can a user receive in return for sharing a post? Common bond theory [4] suggests that expectation of reciprocity may be a motivating factor for recommending a post to a friend. When the recommended friend stands to directly benefit from the post, they may in the future feel obligated to recommend a similarly beneficial post in response. By making a recommendation, a user is also signaling to the friend that they were thinking of the friend and have some intimate knowledge of the attributes of the friend, which serves to strengthen their social bond, and may also motivate a user to recommend a friend that is expected to recommend their friends, a case in which reciprocity may not apply. As with fulfillment, there is also some amount of reputational and altruistic utility received for publicly recommending friends. Consider also the motivations for recommending a post when the requester is a friend of the recommender. The social bond motivation is doubly applicable in such a scenario.

The cost of recommendation is in cognitive effort. A user must consider the attributes of each of her friends, for which an exhaustive mental search may be expensive. It is likely that the current system does not optimally minimize the cognitive cost of recommendation. Based on the statistical figures presented in the previous section, one may conclude that the benefits of making recommendations are not enough to overcome the cost, as the rates of recommendation are in most cases lower relative to the rates of fulfillment.

Two benefits from recommendation compared to broadcast are that the quality of users recommended will generally be higher and the probability that the recommendee encounters the post will be higher because the message specifically targets that person, whereas it is easy for a person to miss or ignore a broadcast post.

5.5.4 Broadcaster

As with recommendations, there is also no direct utility for broadcasting a post. In this case, however, the common bond motivations are lower relative to recommendation due to the impersonality of broadcasting a message. It is also likely that users will only broadcast posts that are in line with their identity, since the post is tied to the user's public persona. Some evidence for this theory comes from the statistics of the Nonprofit category, which has a broadcast

rate that is about four times greater than other categories. Users are apparently selectively broadcasting posts that improve their public image. The cognitive effort cost of broadcasting is comparatively much lower than the cost of recommending, but the benefits are also lower. The statistical rates of broadcast are about as low as the rates of recommendation.

5.6 Free-Riding

In this context, free-riding occurs when users post requests, but do not recommend or broadcast others' requests. The cost of posting a request is low, while the benefits of having a request fulfilled are certainly above the cost of posting. The cost of fulfilling a request, connecting the requesting user with a fulfilling user, is borne by the other users in the community, most of which will receive negative direct utility for recommending and broadcasting the request.

Figure 4 in the previous section suggests that free-riding is common in the community. It clearly shows that many fewer recommendation and broadcast actions are being taken on posts than fulfillments. This result implies that the noted indirect motivational factors for recommending and broadcasting are alone not significant enough to incentivize action.

Could it be that the recommendation and broadcast actions are not required for successful fulfillment? The statistics show that only about a quarter of posts are being fulfilled by non-employees. Even when employee actions are included, the percentage of posts fulfilled is only 45%. The ideal fulfillment rate is, of course, 100%, so the community is clearly falling short when it comes to fulfilling requests. If the rate of recommendation and broadcast are increased, the rate of fulfillment will increase. Thus, correcting the free riding problem will be crucial to the success of the community.

6 Recommendations

In general, the community appears poised to hit a tipping point with the help of several modifications. The first two provided here are technologically simple, while the third is more of a long-term technological effort. The recommendations are as minimal as possible, based on the evidence and analysis provided in previous sections.

6.1 Provide Summary Measures of Activity

A simple yet highly impactful solution to the free-riding problem is the introduction of summary measures of user activity to the community's interface. Specifically, place summarized information about the history of a user's activities somewhere clearly visible alongside each of the user's requests. The information presented should at least include the number of requests, recommendations, broadcasts, and fulfillments. The information should also include a summary measure that is easily comparable between users. For example, a weighted sum of recommendations and broadcasts divided by the number of requests may be a good single metric that is comparable between all users and provides sufficient information to determine whether the requesting user is a free-rider. I do not include fulfillments in this metric because it is ambiguous whether fulfillments are beneficial to the fulfiller. Providing contextualizing information for the summary measures is also important. Make clear how the user compares to the remainder of the community. This will make the summary measures more meaningful and actionable by users encountering the post.

This would modify the behavior of users encountering a request by giving them the information sufficient to determine whether the requesting user is treating the community, and themselves by extension, fairly. It has been shown that people that believe they are being treated unfairly will respond with punishment [14]. In this case, the punishment inflicted on the free-riding user is the refusal to help the requester connect with a fulfiller.

This would modify the behavior of a requester by causing her to consider the reactions of users encountering the request. Knowing that users may punish her for not contributing to the community, she will have an incentive to increase her number of broadcasts and recommendations in order to increase her measures of activty, and as a result increase the probability that her request is fulfilled. This reaction solves the free-riding problem.

It is not necessary to enforce any specific threshold contribution requirements for posting a request. This could actually be damaging to the community when a mixture is chosen that is not in line with what the general community considers to be a fair threshold. I suspect that self-policing behavior will be sufficient to solve the free-riding problem.

This solution is similar to the free-riding solution implemented in the peer-to-peer file sharing network BitTorrent [15], in which the protocol does not explicitly forbid clients from free-riding, but makes available the choice for peers to discriminate against clients that are free-riding. Free-riding in the BitTorrent context is when a client only downloads from its peers, but does not upload. Standard clients employ the tit-for-tat strategy, where uploads are sent reciprocally. Clients quickly learn whether another client is free-riding, after which the free-riding client is shut out of the download.

One would not expect white-washing to be a practical or beneficial strategy around this free-riding solution. Because Sumazi utilizes people's true identities on Facebook, users would need to generate convincing new Facebook identities. Creating such an identity is likely to be more difficult than simply broadcasting and recommending a few posts.

In a more general sense, this one modification would alter the dynamic of the community from one of many independent requests to a set of jointly dependent requests. Instead of a requester focusing on only their request, she will be incentivized to also take helpful actions on others' requests.

6.2 Grow Network With Nonprofit Requests

It is evident that Nonprofit requests have a higher rate of every action relative to other post categories. By increasing the number of requests related to Nonprofit issues, increased recommendations and broadcasts will increase the number of Sumazi users, giving a stronger jump-start to the positive feedback loop described earlier. A secondary benefit would be to improve the public image of Sumazi. While this is not a technical solution to any particular problem in the community, the data show that implementing this recommendation will increase the effectiveness of the community.

6.3 Minimize Cognitive Cost of Recommendation

There is a high cognitive cost to determining which friends would be appropriate for recommendation. If this cost is reduced, the number of recommendations will likely increase. Thus, a goal of the community must be to make recommendation as cognitively easy as possible. Unfortunately, this is likely a technologically hard goal to achieve, but is nonetheless important for the success of the community in the long-term.

An obvious modification that would make recommendations easier is the addition of a tool that automatically suggests to a user the friends to which she may want to recommend the post. Options are cognitively easier than free thought, so this would reduce cognitive cost. Developing a machine learning algorithm that makes such suggestions will be technologically hard due to the sparsity of training samples (the recommendations people make) and high dimensionality of the data space (the semantics of the request description). A heuristic algorithm could simply suggest friends that the user has recommended in the recent past.

7 Conclusions

Sumazi is an interesting new community that seeks to maximize the probability that potential interpersonal connections are realized. Underlying the community is a general but powerful incentive model, which although showing evidence of free-riding behavior, may require only a few simple modifications to correct.

References

[1] Sumazi :: Connect Intelligently. http://sumazi.com

[2] LinkedIn. http://www.linkedin.com

[3] OKCupid. http://okcupid.com

[4] Ren, Yuqing, Robert Kraut, and Sara Kiesler. Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 2009. 28: 377-408.

[5] DIY Drones. http://diydrones.com

[6] Kozinets, Robert V. E-Tribalized Marketing?: The Strategic Implications of Virtual Communities of Consumption. *European management Journal*, 1999. 17(3): 252-264.

[7] reddit: the front page of the internet. http://reddit.com

[8] metareddit - all about reddit. http://metareddit.com

[9] CraigsList. http://craigslist.org

[10] Facebook Developers. http://developers.facebook.com

[11] Jian, L. and J. MacKie-Mason. Incentive-Centered Design For User-Contributed Content. *Oxford Handbook of the Digital Economy*, 2010, in press.

[12] Barnes, J. A. Graph Theory and Social Networks: A Technical Comment on Connectedness and Connectivity *Sociology*, 1969. 3: 215.

[13] Hartley, Scott. Kiva.org: Crowd-Sourced Microfinance & Cooperation in Group Lending. Harvard working paper.

[14] Fehr, Ernst and Klaus M. Schmidt. The Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics*, August 1999.

[15] Cohen, Bram. Incentives Build Robustness in BitTorrent, 2003. http://www.ittc.ku.edu/ niehaus/classes/750-s06/documents/BT-description.pdf Accessed 16 Feb., 2012.