The Network Effects of Recommending Social Connections

Elizabeth M. Daly, Werner Geyer, David R. Millen
IBM T.J. Watson Research
Cambridge, MA 02116, USA
{dalyeliz, wgeyer, david_r_millen}@us.ibm.com

ABSTRACT

Social networking sites have begun to be used in the enterprise as a method of connecting employees. Recommender systems may be used to recommend social contacts in order to increase user engagement, encourage collaboration and facilitate expertise discovery. This paper evaluates the effects of four recommendation algorithms on the network as a whole and the social structure. We demonstrate that depending on the basis of the recommendation algorithm the effects on the network vary greatly and their potential impact should be understood. It is hoped this research can be used as guidance for future recommendation algorithms.

Categories and Subject Descriptors

H.5.3 [**Information Systems**]: Collaborative computing; J.2 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Algorithms, Human Factors

Keywords

Social-Networks, Recommender Systems

1. INTRODUCTION

The power of social networks lies in the delicacies and rationale behind the connections. The result is a network graph with 'small world' properties that have high clustering of communities and a short network distance through ties between these communities. Communities with strong social ties foster trust and communication within the community. On the other hand, a lack on connections to new communities may cause isolation of the group and may have important implications in terms of information diffusion, cross division collaboration and knowledge sharing. Much research has focused exploring the relationship of social structure and knowledge creation. People who provide brokering connections between communities are more likely to connect people with different ideas, interests and perspectives. Burt found that cohesive communities within organizations of similar people may reduce creativity and those who intersect with several communities are more likely to be the source of good ideas [1]. Uzzi and Spiro analyzed a network of creative artists who made Broadway musicals, and found a relationship between the small world properties of the collaborative network and the probability of a show's success [12].

Recommender systems can be used to recommend social contacts. Little work has been done to measure the effects of these recommendations on the network as a whole and the social structure. A social network consists of people, and people have limited resources when it comes to maintaining relationships, providing information and dealing with communication. In contrast when recommending items such as books and movies, the limited resource constraint maybe less problematic. The rich-get-richer phenomenon has been acknowledged to reduce diversity and favor older items, resulting in suppressing potentially better matches [4]. However, the implications of a recommendation algorithm leading to a book becoming a best seller has a significantly different impact when we consider the same algorithm applied to people. A heavily recommended person in an organization that becomes the most connected may gain unfair advantage, become overworked, overexposed or a bottleneck. As a result, people recommendation algorithms in social networks need to consider the upper limits of the number of social ties a person can reasonably maintain [3]. The importance of social structure is widely recognized, however, the 'ideal' structure is of much debate. If a recommender system is to influence the social structure, a delicate balance of these different concerns should be kept in mind when creating or encouraging additional social connections in a network graph.

Previous research on people recommendations has focused on user acceptance, perceived user satisfaction, and impact on site activity. Guy et al. measure the impact of people recommendations on the number of friends. Their data demonstrates a significant increase in social connections and shows that presenting evidence is a crucial element for user satisfaction and acceptance [8]. Chen et al. evaluate four different recommender algorithm in an enterprise social network site [2]. They found that algorithms using social network information have higher acceptance rates than content-based user matches based on online profile information. The latter algorithms however, are better at helping users to discover new friends. This phenomenon is similar to the over-

RecSys2010, September 26–30, 2010, Barcelona, Spain.

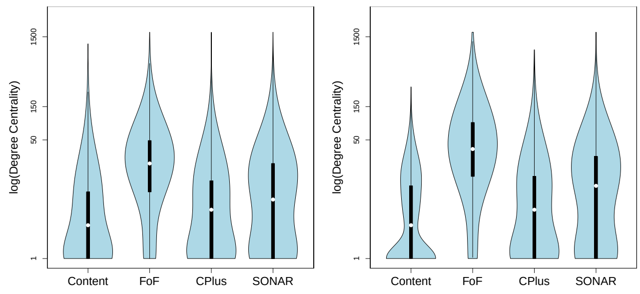
specialization in recommender systems [13], i.e. social algorithms recommend more of the same (i.e. in this case known people) whereas content-based approaches add diversity at the cost of lower acceptance. Freyne et al. [6] analyze the effect of people and content recommendation on retention rates of newcomers and show that people recommendations can increase long-term retention and user activity, in particular when recommending active users. Terveen and McDonald [11] provide a comprehensive, general overview of social matching systems with the goal of creating a framework that helps researchers understand the design space. Our research aims at determining the impact of different recommendation algorithms on the social network structure and addresses some of the questions raised in [11]. We evaluate the after effect of four recommendation algorithms previously deployed in the IBM Research social networking site, Social Blue (formerly known as beehive) [2].

2. RECOMMENDING CONNECTIONS

Our analysis is based on a dataset from a large user study conducted by Chen et al. in July 2008 [2]. They deployed four different recommendation algorithms to four different groups of IBM's SocialBlue social network site. The dataset used in this paper consists of 600 users per recommendation group. At the time of the study, the site had more than 38,000 registered users with an average of 8.2 friends per user. The four different algorithms differ in the amount of social network information used [2]: ***Content Matching*:** Users are recommended based on their similarity to other users on the site. Similarity of two users is measured by their cosine similarity of the TF-IDF weights of keywords derived from their online profile content.***Content-plus-Link*:** This algorithm enhances the Content Matching algorithm with social link information derived from the social network structure, i.e. in addition to the keyword match the algorithm also shows the network path to the recommended person if it's shorter than or equal to 3 links. ***Friend-of-Friend*:** This popular algorithm (used on facebook.com and other social media sites) is based solely on social network structure. It assumes that "if my friend likes someone, I might be interested in that person too." ***SONAR*:** The people recommendations are generated by the social network service SONAR [7], which aggregates relationship information from a multitude of different public data sources within IBM including the organizational chart, patent and publications, several friending systems, a people tagging and a blogs.

3. MEASURING THE SOCIAL NETWORK

We utilize social network analysis techniques to evaluate and explore the results of the social recommender experiment in [2]. '*Degree*' centrality is measured as the number of contacts of a given user [5]. A user with high degree centrality maintains numerous contacts and can be seen as popular. As a result, a central user occupies a structural position that may act as a conduit for information exchange. In contrast, peripheral users with few or no contacts are located at the margins of the network and do not have much ability to propagate information. '*Betweenness*' centrality measures the extent to which a user lies on the shortest paths linking other users [5]. Betweenness centrality can be regarded as a measure of the extent to which a user has control over information flowing between others. A user with a high betweenness centrality has a capacity to facilitate interactions between others that are otherwise separated. '*Modularity*' is a measure that defines the quality of a particular division of a network [10]. The modularity based on a division is a value between -1 and 1 that measures the density of links inside communities as compared to links between communities.



(a) All (b) Accepted

Figure 1: Degree of Recommendations Connections

3.1 Evaluation Methodology

The social graph of the entire network before the recommendation experiment is computed. The connections created as a result of each recommendation algorithm are identified and four post-recommendation graphs are built to represent the effect of each recommendation algorithm.

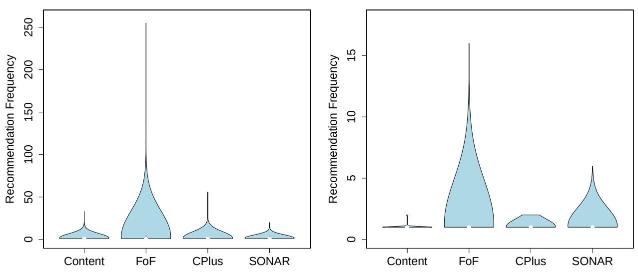
3.2 Degree Distribution

Figure 1 a) shows a violin plot [1] of the degree centrality distribution for all recommended friends using the different algorithms. The Friend-of-Friend algorithm has a higher portion of recommended users with a large number of friends. The Content Matching algorithm on the other hand has a high portion of users with few or no contacts at all. The impact of the Friend-of-Friend favoring of users with an already large social network is also seen in the accepted recommendations shown in figure 1 b). All other three algorithms enabled connections to users with a relatively lower number of social contacts.

3.3 User Recommendation Bias

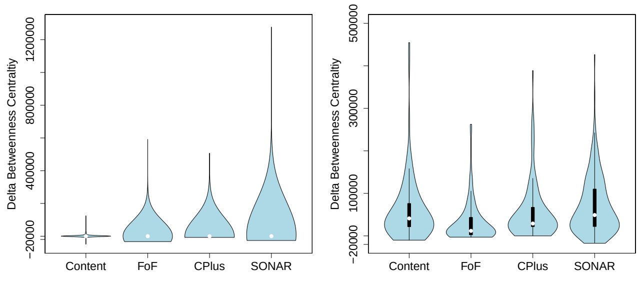
Depending on the matching criteria for recommendation algorithms, certain users may have a clear advantage over others. For example, in the Friend-of-Friend algorithm, a user with a wider social circle has a higher probability of friendship overlap and therefore may be recommended frequently to many different users. This bias to already well connected individuals will result in the recommended users becoming even more connected. In figure 2 a) we see that a single user was recommended to a total of 255 users, which is more than 50% of the Friend-of-Friend experiment group. This single user gained 16 new contacts from being recommended as as shown in figure 2. Content based recommendations, in contrast, have produced more personalized recommended contacts for each participant with the majority

[1] A violin plot is a combination of a box plot and a kernel density plot.
See: cran.r-project.org/web/packages/vioplot/vioplot.pdf



(a) All (b) Accepted

Figure 2: Frequency of User Recommended as a Connection



(a) Recommended Users (b) Connecting Users

Figure 3: The delta in betweenness centrality for users

only being recommended to 10 users or less and only a maximum of 2 new contacts for a single user arose from being recommended. As a result, no single user becomes a clear winner in this algorithm and distributes the recommmendations more evenly throughout the member space.

3.4 Effect on Betweenness Centrality

Figure 3 a) shows the delta in betweenness centrality of users that were recommended. Content Matching has the least impact on the recommended users. The other three algorithms have similar distributions, however, SONAR resulted in a single user increasing their betweenness centrality by the highest amount.

Figure 3 b) shows that the Content Matching algorithm shows the highest increase in betweenness centrality for a connecting user, i.e. the new connections gained by the user placed the participant in a role of connecting users that were previously disconnected. Though the Friend-of-Friend algorithm resulted in the most new connections, it shows the least overall increase in betweenness centrality for the group participants. This is a result of recommending connections that are already indirectly member of the user's social circle. The user that shows the highest increase in betweenness centrality was recommended to %50 of the participants, and 16 users accepted this recommendation. Therefore it took a large number of new connections to increase this users social capital. In contrast, the users whose betweenness centrality was most affected by the Content Matching algorithm only added a single connection, meaning this one new connection had a high impact on the network structure and provided a link to previously unconnected users. The SONAR algorithm also proved to have significant impact on the participating users, again encouraging links between diverse communities.

3.5 Effect on Clustering

Table 2 shows that the Friend-of-Friend and SONAR algorithms recommend less users from the same countries than recommendations based on content. Given the IBM members in beehive span 64 different countries, recommendations based on Content Matching lead to a surprising bias towards recommending users from the same country. All algorithms showed a reduced modularity and therefore in clustering based on country location. The SocialBlue members belong to over 300 different organizational divisions. All algorithms showed a bias in recommending users that are within the same division. However SONAR is the only algorithm that explicitly takes into account organizational hierarchy and management levels. This leads to a clear bias towards recommending users from the same division. As shown in figure 4, both Friend-of-Friend and SONAR have the effect of increasing the modularity of the network based on division, thus strengthening connections within the division. We can also see that users show a higher probability of accepting recommended users that have a similar division. Meaning the recommendations have strengthened the connections within the company divisions. The content based algorithms have reduced modularity, meaning it has improved connections between divisions.

Table 1: Bias of Recommendations % from Same Country

Algorithm	%Recommended	%Accepted
Content Matching	41.38	52.45
Friend-of-Friend	18.25	30.66
Content-plus-Link	41.06	52.42
SONAR	28.83	31.95

Table 2: Bias of Recommendations % from Same Division

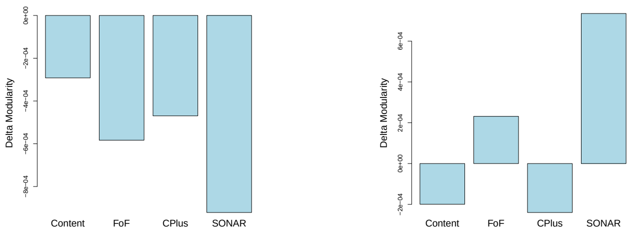
Algorithm	%Recommended	%Accepted
Content Matching	25.63	55.22
Friend-of-Friend	37.56	55.18
Content-plus-Link	32.66	55.48
SONAR	77.19	75.53

3.6 Activity Levels of New Links

Table 3 shows the activity levels of the resulting connections. As can be seen in the 12 months following the experiment, all algorithms show a low uptake in maintaining these relationships. However the algorithm that shows the least percentage of throw away links is the Content Matching algorithm with %18. SONAR however results in the most active links with an average of 3 events per active link, the Content Matching algorithm results in the second most active links. Interestingly, Friend-of-Friend shows the lowest level of activity events per link. The directionality of the activity is also shown and Friend-of-Friend shows the least

Table 3: Number of Links Activity over following 12 months

Algorithm	Created Links	Active Links	%Active	Ave. Event per Active Link	%Participant to Rec	%Rec to Participant
Content Matching	406	74	18.23	2.82	68.92	45.95
Friend-of-Friend	1257	173	13.76	2.35	70.52	34.68
Content-plus-Link	558	66	11.83	2.45	56.06	51.51
SONAR	1638	229	13.98	3.38	64.19	48.47



(a) Country (b) Division

Figure 4: The delta in Modularity of the Network Clustering

activity from the recommended user, suggesting the already popular users may be less interested in maintaining these new connections. The other three algorithms show %50 of the recommended users interacted with the new connection. The Content-plus-Link algorithm shows a surprisingly low level of activity from the participating users. This suggests the users accepted the connections, but did not have significant interest in the users they connected to.

4. CONCLUSION

This paper has explored the network effects of four different recommendation algorithms and their implications for the social network. We argue that recommendation algorithms in social networks should not be evaluated on user acceptance rate alone. A recommendation algorithm that identifies all users in a user's immediate social graph may have a high acceptance rate but may suppress showing contacts that might be interesting, relevant and valuable, but are too far away in the social graph. The choice of recommendation algorithm has an impact on the delicate balance of small world social networks.

If the aim is to promote popularity, then the choice of Friend-of-Friend algorithm may be satisfactory, although in the long-term, this algorithm leads to the 'rich getting richer' and thus may decrease the value of the network. If the goal is to promote long range links that connect otherwise disconnected communities, the Content Matching algorithm could be used. While this algorithm leads to an increase in betweenness centrality and to an increase in long-term user activity, it has the lowest acceptance rates since it recommends mostly unknown users [2]. One possible solution to improve user acceptance of recommendations is to highlight, with every recommendation, how it will affect a users reach and access to new communities and people important for their job. Future work should evaluate how a user's awareness of the network effects impacts acceptance rates compared to showing only content matching evidence.

If the optimal and desired network structure is arguable and unless the aim is clear, then a possible solution is to use a variety of different algorithms. This argument supports the findings by McDonald stating that '*users naturally want the system to augment and assist, not replace their natural behavior*' [9]. As a result, the recommender system could aim to provide relevant and useful options to allow users to make natural selections rather than being overly influenced by a single network view.

5. REFERENCES

[1] R. S. Burt. Structural holes and good ideas. *The American Journal of Sociology*, 110(2):349–399, 2004.

[2] J. Chen, W. Geyer, C. Dugan, M. Muller, I. Guy. Make new friends, but keep the old: recommending people on social networking sites. In *CHI '09*, pages 201–210, NY, USA, 2009. ACM.

[3] R. Dunbar. Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–735, 1993.

[4] D. M. Fleder and K. Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Social Science Research Network Working Paper Series*, April 2007.

[5] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.

[6] J. Freyne, M. Jacovi, I. Guy, W. Geyer. Increasing engagement through early recommender intervention. In *RecSys '09*, pages 85–92, NY, USA, 2009. ACM.

[7] I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, S. Farrell. Harvesting with sonar: the value of aggregating social network information. In *CHI '08*, pages 1017–1026, NY, USA, 2008. ACM.

[8] I. Guy, I. Ronen, E. Wilcox. Do you know?: recommending people to invite into your social network. In *IUI '09*, pages 77–86, USA, 2009. ACM.

[9] D. W. McDonald. Recommending collaboration with social networks: a comparative evaluation. In *CHI '03*, pages 593–600, NY, USA, 2003. ACM.

[10] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. 2003.

[11] L. Terveen and D. W. McDonald. Social matching: A framework and research agenda. *ACM Trans. Comput.-Hum. Interact.*, 12(3):401–434, 2005.

[12] B. Uzzi and J. Spiro. Collaboration and creativity: The small world problem. *American Journal of Sociology,*, 111:447–504, 2005.

[13] C Yu, L Lakshmanan, S Amer-Yahia. It takes variety to make a world: diversification in recommender systems. In *EDBT '09*, 368–378, NY, USA, 2009.