Link Recommendation for Promoting Information Diffusion in Social Networks

Dong Li, Zhiming Xu, Sheng Li, Xin Sun
Harbin Institute of Technology
Harbin, 150001, P.R. China
{hitlidong, xuzm, lisheng, sunxin}@hit.edu.cn

Anika Gupta, Katia Sycara
Carnegie Mellon University
Pittsburgh, PA-15213, USA
anikag@andrew.cmu.edu, katia@cs.cmu.edu

ABSTRACT
Online social networks mainly have two functions: social interaction and information diffusion. Most of current link recommendation researches only focus on strengthening the social interaction function, but ignore the problem of how to enhance the information diffusion function. For solving this problem, this paper introduces the concept of user diffusion degree and proposes the algorithm for calculating it, then combines it with traditional recommendation methods for reranking recommended links. Experimental results on Email dataset and Amazon dataset under Independent Cascade Model and Linear Threshold Model show that our method noticeably outperforms the traditional methods in terms of promoting information diffusion.

Categories and Subject Descriptors
H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *information networks*

Keywords
link recommendation; information diffusion; diffusion degree

1. INTRODUCTION
With the rapid development of networking sites (e.g., Facebook, Twitter and LinkedIn), online social networks have drawn substantial attention. They mainly have two functions: social interaction and information diffusion. Based on these two functions, users can make friends and obtain information in social networks. Link recommendation can help users find suitable friends based on user's preferences and network structure, which can enhance the social interaction function. Furthermore, link recommendation is also closely related with the information diffusion function, because it can speed up network evolution and change network structure, which directly impacts information diffusion. Most link recommendation methods ignore this point.

Link recommendation algorithms should not only focus on accurately estimating the proximity based on user's preferences and social relations, but also consider that the new connections added to network by recommendation algorithms should maximize information diffusion. This last problem has been neglected in the literature. This problem is different from the influence maximization problem [2]. The objective in [2] is to select the *top-K* influential nodes in the social network that can be targeted to maximize influence spread over the network. In contrast, our problem is to add *K* new links for each target user so as to let the updated network maximize information diffusion.

To address this problem, Chaoji et al. [1] proposed a greedy algorithm based on RMPP diffusion model. But in their RMPP model, information spreads from user i to j only along the path with maximum probability among all paths form i to j containing at most one edge added by recommendation methods. Obviously, this is inconsistent with real information diffusion process. And the greedy algorithm also needs a lot of calculation, it is not easily scalable to huge networks. In this paper, we propose a scalable heuristic algorithm. Specifically, we introduce the concept of user diffusion degree and propose its corresponding calculation algorithm. Then we combine it with proximity between users estimated by traditional recommendation methods for reranking links. In contrast with [1] limited to the RMPP model, we implement the proposed method on two common diffusion model, Linear Threshold Model and Independent Cascade Model [2], which can simulate information diffusion process more realistically than RMPP model. And our heuristic algorithm requires less calculation compared with greedy algorithm, it can be applied in larger datasets.

2. METHODS
A social network can be represented as a weighted graph $G=(V,E)$. For each edge $(i,j)\in E$, $w(i,j)$ denotes the weight or propagation probability of the edge. Recommendation methods [4, 5] usually calculate proximity $p(i,j)$ for each user pair $(i,j)\notin E$ and recommend links based on this proximity between users. These methods can satisfy user's preference and enhance growth of social networks. For the purpose that the updated network can maximize information diffusion, we introduce the user diffusion degree which is a measure of the influence a user has over the spread of information through the network. Let $d(i)$ represent the diffusion degree of user i. The main purpose of this paper is not to propose a new method of computing proximity between users, but to combine the proposed diffusion degree with proximity for recommendation to facilitate information diffusion. This diffusion degree can work with any recommendation method proposed in previous literature.

Diffusion degree is different from node betweenness, which is a measure of a node's centrality in a network, and is normally calculated as the number of shortest paths between nodes pairs that run through the node of interest. Betweenness can measure diffusion degree of users in some sense, but it has a significant drawback. For example, when user u and user v have the same betweenness, they may in different situations: node pairs connected by u mainly belong to one group, and node pairs connected by v distribute in many different groups. Obviously, in this case u can only promote information diffusion in that local group, while v can trigger information exchange and spread between different groups as a hub, hence user v should have a higher diffusion degree than user u. In this paper, we propose a method of computing user's diffusion degree, which, in contrast with betweenness, takes the group properties of users into account.

WWW 2013 Companion, May 13–17, 2013, Rio de Janiero, Brazil.
ACM 978-1-4503-2038-2/13/05.

Algorithm 1 User Diffusion Degree Calculation Algorithm

Given a social network G
Use community detection algorithm to get the set of different groups $\{G_1, G_2 \cdots, G_m\}$. For each user $i \in V$, $1 \le g(i) \le m$ means the group number user *i* belongs to.
For $i = 1$ to $|V|$ do
  Pick up user *i*, set $d(i) = 0$,
  Let $S(i) = \{(j,k) \mid (j,i) \in E \,\&\&\, (i,k) \in E \,\&\&\, (j,k) \notin E\}$
  For each user pair $(j,k) \in S(i)$ do
    $d(i) = (|G_{g(i)} \cup G_{g(j)} \cup G_{g(k)}| / |G_{g(i)}|) \times (w(j,i) + w(i,k)) + d(i)$
  End for
End for

Algorithm 1 summarizes the algorithm of calculating diffusion degree. The main idea of this algorithm is: for a user *i*, if he has a large number of neighbor pairs who are not directly connected and the users connected to user *i* belong to many different groups, then user *i* gets a high diffusion degree. In Algorithm 1, we use community detection algorithm to get user's group number [3]. After getting the diffusion degree of users, we adopt multiplication to combine proximity with diffusion degree: i.e. for user i and user j, $rank(i,j) = p(i,j) * d(j)$, $j \in N(i)$. $N(i)$ is a recommendation candidate set of user *i*, which can be got by any traditional recommendation method. Then we recommend new links to target users based on the final ranking score.

3. EXPERIMENTS

In our experiments, we adopt two public realistic graph datasets[1]: Enron-Email dataset (36692 nodes and 367662 edges) and Amazon dataset (334863 nodes and 925872 edges). For each dataset, we randomly select about 1% users as target users, and recommend *K* links to each target user. We adopt widely applied Independent Cascade Model and Linear Threshold Model to simulate information diffusion process, and randomly select about 1% users as initial active users in each dataset. Since we do not have the data to infer weights of edges, we use TRIVALENCY model [2] to assign the weights for edges. Community detection algorithm used in this paper is the Girvan-Newman algorithm [3], which considers that one user only belongs to one group.

In this part, we focus on evaluating the diffusion performance of the updated graphs where new links were added by different recommendation methods. We use lift as the evaluating metric:

$$Lift(X) = \frac{f(E \cup X) - f(E)}{f(E)} \times 100$$

where X is the set of links recommended by different methods. $f(.)$ is the information diffusion function and E is edges set of original graph G. $f(E)$ means the number of users activated by the diffusion information in the graph G. We use two diffusion models to simulate diffusion process to get the value of $f(.)$.

We use the number of common neighbors [4] between two users as proximity metric and combine it with the diffusion degree to recommend links. Actually, the diffusion degree we proposed can be combined with any proximity calculation method, and here we just take the common neighbors as an example to illustrate that our method can get better result than that using only proximity for link recommendation (the first baseline method), so as to prove that our method can help any link recommendation method to enhance the information diffusion function of social networks. Furthermore, we replace the diffusion degree in our combination method with user betweenness to get another heuristic method as the second baseline for comparison.

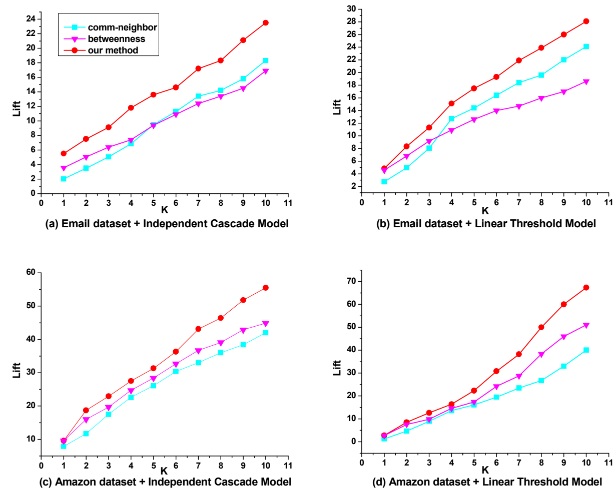

Figure1: Diffusion performances on Email and Amazon datasets under Independent Cascade Model and Linear Threshold Model.

Figure 1 shows the results of all the compared methods on different datasets under two diffusion models, the number of added new links *K* ranges from 1 to 10. As we can see, our method can get better diffusion performance than other methods under both two diffusion models and two datasets. Bad performance of user betweenness on Email dataset shows its drawback in some sense. As *K* increases, the increase in Lift value of our method is larger than that of betweenness, which illustrates our method can improve the drawback of betweenness and facilitate information spread between different groups.

4. CONCLUSION

In this paper, we proposed a method for solving the problem of recommending links with the explicit objective of maximizing information diffusion in the social networks. We introduced the concept of user diffusion degree and proposed its corresponding calculation algorithm. It can be combined with any traditional recommendation method for reranking recommended links. Experimental results on Email and Amazon datasets under two diffusion models have validated the effectiveness of our methods.

Acknowledgements. This work is supported by the Natural Science Foundation of China (No. 61173074) and the ZTE cooperation project (No. MH20120428).

5. REFERENCES

[1] Chaoji,V., Ranu,S., Rastogi, R., and Bhatt, R. Recommendations to boost content spread in social networks. In *WWW*, pages 529-538, 2012.
[2] Chen, W., Wang, C., and Wang, Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, pp. 1029-1038, 2010.
[3] Girvan, M. and Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA 99*, 7821-7826(2002)
[4] Liben-Nowell, D., and Kleinberg, J. The link prediction problem for social networks. In *CIKM*, pp. 556-559, 2003.
[5] Yin, D., Hong, L., and Davison. B. D. Structural link analysis and prediction in microblogs. In *CIKM*, pp 1163-1168, 2011.

[1]http://snap.stanford.edu/data/index.html