Information Sciences 309 (2015) 102–118

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Friend recommendation with content spread enhancement in social networks

Zhi Yu, Can Wang *, Jiajun Bu, Xin Wang, Yue Wu, Chun Chen

Zhejiang Provincial Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 12 October 2013
Received in revised form 20 January 2015
Accepted 9 March 2015
Available online 18 March 2015

Keywords:
Friend recommendation
Social network
Algebraic connectivity
Content spread
Information dissemination

ABSTRACT

Social network is becoming an increasingly popular media for information sharing. More and more people are interacting with others via major social network sites such as Twitter and Flickr. An important aspect of a social network is its capability in efficiently spreading content, not only within a small circle but also in the whole network. However, most existing methods for recommending friends in social networks only aim at achieving high recommendation success rate. The network grown from such recommendations is not optimized for content spread. In this paper, we propose a novel friend recommendation method ACR-FoF (algebraic connectivity regularized friends-of-friends) that considers both success rate and content spread in the network. Using the *algebraic connectivity* of a connected network to estimate its capability for spreading contents, our recommendation method naturally extends existing friend recommendation algorithms such as FoF to achieve both recommendation relevance and content spread in a social network. Experimental results on simulated and real social network data sets show that our method can significantly improve content spread in a social network with only a very tiny compromise on friend recommendation success rate.

1. Introduction

The past several years witnessed an explosive growth of social network sites and applications. People are increasingly relying on various social networks for sharing contents and interacting with each other. Popular sites such as Twitter and Facebook are seeing a great amount of tweets or posts generated by active users all around the world every day. Social media nowadays has become one of the most important information sources [33] and we intuitively expect information spread in a viral fashion in a social network.

However, recent works indicate information or content in social networks may not spread as efficiently as people believe. Cha et al. explain in [3] the formation of *content locality* in Flickr as they observed from a Flickr dataset ranging over 104 consecutive days, which shows that even popular photos many only circulate within a small clique and result in a quick burnout in content spread. Bakshy et al. echo in [2] with similar discoveries by noting increasing *homophily* in social networks. They point out that individuals with similar characteristics tend to associate with each other, leading to greater

* Corresponding author.
E-mail addresses: yuzhirenzhe@zju.edu.cn (Z. Yu), wcan@zju.edu.cn (C. Wang), bjj@zju.edu.cn (J. Bu), xinwang@zju.edu.cn (X. Wang), wy1988@zju.edu.cn (Y. Wu), chenc@zju.edu.cn (C. Chen).

http://dx.doi.org/10.1016/j.ins.2015.03.012
0020-0255/

opportunities of aligned information source and less possibilities for accessing novel and diverse information. Homophily not only impairs people's motivation for sharing contents, but also may slow down audience growth in social networks.

On the other hand, it is generally believed that the value and revenue of a social network are closely related with its capability in efficiently spreading contents. Content spreading plays an important role in motivating users to express their opinions through the social network because the goal of sharing information is for others to receive [35,21]. Quicker information dissemination can drive up user engagement and in turn improve user retention. Wide outreach of contents in a social network also improves the access to novel information for users. Most social network sites are now offering connectivity boosting features such as "*friend recommendation*" to users to improve content spread. However, traditional friend recommendation methods such as friends-of-friends (FoF) mainly consider number of common neighbors or similarity of user profiles in recommending new friends. One drawback of these methods is that they usually increase connectivity only within a small range of neighborhood and may not necessarily lead to improved content spread capability of the whole network.

Although connecting unknown people seems to be an effective method in discovering novel contents and reducing homophily in social networks, most users lack enough motivation to contact unknown people and thus making the recommendations fail. Considering the limitations of these traditional friend recommendation methods, Chaoji et al. recently proposed in [4] a connection recommendation algorithm that can also improve content spread in a social network. A two-step process is used in their recommendation by first selecting a candidate set of similar users based on the number of common neighbors, similarity of user profiles, etc. and then recommending a certain number of friends in the social network (no more than k friends for each user) that maximize the content spread. The limitation of their method is that the user is not likely to accept all the k recommended friends, which makes the added connections less optimal for content spread. Also the relatively high computational cost of their algorithm makes it less scalable in practical applications. But as the first attempt to incorporate content spread capability into friend recommendation, their work inspires us to seek solutions that better balance the two objectives.

The content spread in a network is usually modeled as a stochastic process, e.g. independent cascade (IC) model [17]. An edge $e(u,v)$ in this model is associated with a probability $p(u,v)$, which is the possibility in step $t+1$ the node u independently propagate to node v the content it receives in step t. Metrics based on IC model, such as the expected amount of content received by nodes [4], are used to evaluate the content spread. But directly optimizing such metric is NP-hard [4]. Even approximation algorithms will incur expensive computational cost. Some existing approaches use the *largest eigenvalue* of the **Adjacency Matrix** [31,26,29] to analyze or improve content spread in a network. In this paper, we use the *algebraic connectivity* of a connected network's **Laplacian Matrix** to improve its capability in spreading content. We think the algebraic connectivity of a network is more closely related to its content spread capability than the largest eigenvalue of the adjacency matrix, which is related more closely to the threshold of an epidemic contagion. Using the algebraic connectivity as a regularizer, we develop a new friend recommendation algorithms called *algebraic connectivity regularized friends-of-friends* (ACR-FoF), which effectively take into account both relevance and content spread in friend recommendation. We summarize the major contributions as follows:

- To the best of our knowledge, this is the first time the algebraic connectivity of a network is used to optimize its content spread capability.
- We propose a new evaluation metric EnPair (enhanced pairs) for evaluating improvement in content spread in social networks and the reduction in homophily.
- We develop a new friend recommendation algorithm ACR-FoF that considers both relevance and content spread in a social network.

The rest of the paper is organized as follows: Section 2 introduces to the related works. Our new friend recommendation algorithm is explained in Section 3. In Section 4, we present and discuss experiments and results. Finally, we conclude the whole paper and present some directions for future work in Section 5.

2. Related works

Our work is related to information disseminations and recommendation in social networks. Here we briefly review these related works.

2.1. Information dissemination

Study of information dissemination or content spread in social networks can be traced back to [27], in which Rogers explains how new ideas spread via communication channels over time. He also analyzes how information spread transforms the way people communicate and adopt new ideas. Some more recent works start exploiting combinatorial optimization in minimizing the network diameter and the average shortest path distances [8,23]. However, since not all recommendations will be accepted by users in practice, these algorithms will lead to a suboptimal solution in recommendation. We will provide more analysis into this issue in Section 3.