Physica A 419 (2015) 301–306

Contents lists available at ScienceDirect

Physica A

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

Whether information network supplements friendship network

Lili Miao [a], Qian-Ming Zhang [a,b], Da-Cheng Nie [a], Shi-Min Cai [a,*]

[a] Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, PR China

[b] Center for Polymer Studies and Department of Physics, Boston University, Boston, 02215, USA

HIGHLIGHTS

- The interaction between friendship network and information network in a multi-relation networked system is of interest to us.
- An inverse investigation is presented to analyzing the effect of users' taste on friendship creation.
- The behavioral information involving with popular objects in bipartite information network improves accuracy of friendship prediction.
- The different usable information in bipartite information network is respectively used in friendship prediction and recommendation system.

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 5 August 2014
Available online 16 October 2014

Keywords:
Link prediction
Homophily
Social network analysis
Resource allocation

ABSTRACT

Homophily is a significant mechanism for link prediction in complex network, of which principle describes that people with similar profiles or experiences tend to tie with each other. In a multi-relationship network, friendship among people has been utilized to reinforce similarity of taste for recommendation system whose basic idea is similar to homophily, yet how the taste inversely affects friendship prediction is little discussed. This paper contributes to address the issue by analyzing two benchmark data sets both including user's behavioral information of taste and friendship based on the principle of homophily. It can be found that the creation of friendship tightly associates with personal taste. Especially, the behavioral information of taste involving with popular objects is much more effective to improve the performance of friendship prediction. However, this result seems to be contradictory to the finding in Zhang et al. (2013) that the behavior information of taste involving with popular objects is redundant in recommendation system. We thus discuss this inconformity to comprehensively understand the correlation between them.

1. Introduction

Connections among individuals are not created blindly but able to be predicted. The reference factors could be found in the properties of individuals and their relationships. Predicting these connections is usually defined as the link prediction problem by treating individuals and connections as nodes and links respectively [1,2]. A large number of methods have been proposed to gain more understanding of link creation [2]. For example, rich-get-richer mechanism suggests that nodes with more neighbors will attract new links with higher probability [3], clustering mechanism declares that two nodes have a high probability of creating a link between them if they share some common neighbors [4], and homophily mechanism

* Corresponding author.
E-mail address: shimin.cai81@gmail.com (S.-M. Cai).

http://dx.doi.org/10.1016/j.physa.2014.10.021
0378-4371/

states the observed tendency that people tie with others of similar profiles or experiences [5]. All these mechanism and their derivation are widely used in link prediction problem.

Here we mainly concern on homophily mechanism that is confirmed by plenty of evidences, such as the observed tendency in an acquaintance network of university members [6], a large-scale instant-messaging network containing 1.8×10^8 individuals [7], friendship networks of a set of American high schools [8], a social network of a cohort of college students in Facebook [9], and performing well in predicting missing links for directed networks co-working with potential theory [10]. Homophily could also be reflected in individual reading and musical tastes, shopping history, interests, and so on. These information are widely used to measure the similarity between users in collaborative filtering recommender systems [11]. Meanwhile, facilitated by online social networks (e.g., www.epinions.com and www.last.fm), users can not only create connections to others, but also choose objects they like. From the perspective of homophily, two users can be considered similar if they have many common friends, or have chosen plenty of same objects. Here we raise the question: are the two kinds of similarity correlated with each other?

Some researchers have already revealed that the preferences of trusted friends were beneficial to improve the effectiveness of recommendation [12–14]. In turn, it is lacking that whether the similarity of personal taste could reflect the intimate friendships. This question arises as the friendship and personal taste commonly characterize the users belonging to an identifiable social network. It is beneficial to check the consistency between the two features of homophily, denoting the relationship type and personal taste. To address this question, we collect two data sets both including users' friendships and behavior information, i.e., reading traces in epinions.com and listening traces in last.fm. The similarities between pairs of users are measured according to their friendships and personal tastes. Experimental results suggest that the similarity of taste is helpful but does not work solely. Further, considering the possible redundant or misleading information in recommender systems [15], we reexamine the effect of similarity of taste through gradually removing the behavior information by different strategies. It interests us that these popular objects, which were considered to reflect common taste and lower personalized information in Ref. [15], play much more important role in improving the effectiveness of link prediction in social network.

The rest of the paper is organized as follows: In Section 2 we introduce the data sets and link prediction model, and in Section 3 present the experimental results and the discussion. Finally, we detail our conclusion in Section 4.

2. Material and method

2.1. Data

We find two available data sets that consist of both preference of users and relationship among them. One is collected from *Lastfm*,[1] an online music system where each user has a list of most listened songs and is allowed to make friends with others online [16]. The other is gathered from *Epinions*,[2] a general consumer review site where users can read reviews about variety of items and allows users to decide whether to trust others [17]. To represent these data clearly, we build two networks for each data set: Relation Network (RN) and Taste Network (TN). Concretely, RN indicates the relationship between users, while TN represents the preference of users to objects. We name them as LastfmRN, LastfmTN, EpinionRN and EpinionTN, respectively.

- LastfmRN contains 1543 online users interconnected by 25 434 connections.
- LastfmTN is a bipartite network which contains 1543 users and 18 745 songs. There are 92 834 links between users and songs in total.
- EpinionRN consists of 49 288 users and 381 035 trust relations.
- EpinionTN is a bipartite network with 664 823 ratings on 139 738 reviews rated by 49 288 users.

2.2. Similarity-based link prediction model

An abundance of indices have been proposed to measure the node similarity in networks [2]. For effectiveness and simplification, we choose resource allocation (RA) index [18,19] identified as a well performed one based on local information. Consider a pair of nodes, x and y, which are not directly connected. The node x can send some resource to y, with their common neighbors playing the role of transmitters. In the simplest case, we assume that each transmitter has a unit of resource, and will averagely distribute to all neighbors. Then, the similarity between x and y can be defined as the amount of resource y received from x,

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)}, \tag{1}$$

where $\Gamma(x)$ and $\Gamma(y)$ are the set of neighbors of x and y respectively and $k(z)$ is the degree of node z. Note that the common neighbor z has different meanings in RN and TN. In RN, z indicates common users; while in TN, the common neighbors of users are obviously the objects.

[1] www.last.fm.

[2] www.epinions.com.