Neurocomputing 95 (2012) 89–97

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Relationship strength estimation for online social networks with the study on Facebook

Xiaojian Zhao [a,*], Jin Yuan [b], Guangda Li [b], Xiaoming Chen [a], Zhoujun Li [a]

[a] State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, PR China
[b] School of Computing, National University of Singapore, Computing 1, 13 Computing Drive, 117590, Singapore

ARTICLE INFO

Available online 9 April 2012

Keywords:
Online social network
User's profile information
Interaction activity
Activity field
Relationship strength
Latent Dirichlet Allocation
Graphical model

ABSTRACT

Online social network has become a popular way for users to express themselves, connect and share information with each other. However, in online social networks, the connections between different users are all in binary status, which neglects the relationship strengths between them. Meanwhile, the relationship strength between different users is activity field specific. In different activity fields, such as traveling, shopping, and sport, the relationship strengths between the same users may vary significantly. Therefore, in this paper we propose a general framework to measure the relationship strengths between different users, taking consideration not only the user's profile information but also the interaction activities and the activity fields. We conduct the experiments on Facebook dataset and the results show that the proposed framework is promising and can be used to improve the performances of various applications.

1. Introduction

In the past few years, online social network, which is made up of "nodes" (individuals or organizations) connected by one or more specific types of interdependency, such as friendship, kinship, and common interest, is playing an increasingly important role in our daily life. We use social networks on a daily basis to connect with our family, friends and colleagues, to share the content such as photos with others, to gossip, and to obtain real-time up-to-date information of the news and events that are most important to us. Meanwhile, the popular online social networks have hundreds of millions of registered users and are growing at a rapid pace [1]. As these social networks grow and mature, the users have been observed to form hundreds or even thousands of friends. For example, a user in RenRen[1] has on average over 100 friends [2] and the average number of friends in Facebook[2] is over 130 [3]. Certain individuals have much higher degrees than the average; in fact, in Flickr,[3] we found multiple users who have more than 25,000 friends [4].

While all of the online social networks only present a binary state of the friendship, it has been unsurprisingly observed that not all links are created equally. However, the binary relationship only provides a coarse indication of the nature of the friendship, but does not reveal the relationship strength between different users. At the same time, it is important to estimate the relationship strength between different users in online social network rather than simply treat all the friendship as the binary status. In online social networks, the binary relational ties mix the acquaintances and the best friends together. But compared with the other acquaintances, the users are more inclined to contact with their close friends. And this friendship strength could be better inferred using the relationship strength than using the binary relationship status.

To measure the relationship strength between different users, the typical approaches utilize the user's profile information and the interaction activities among different users [5–9]. The user's profile information provides an overview of his/her basic information including the user's hobbies, the religious view, friends, the work experiences, etc. Generally, the users with similar profiles are likely to exhibit greater similarity, therefore, they may have higher relationship strength, and vice verse. On the other hand, numerous interaction activities, such as tagging the friends' pictures, commenting on the friends' posts, and sending a message to the friends, frequently occur through the online social network in our daily life. These interaction activities between different users reflect their relationship closeness, and thus can be used to measure the relationship strength between different users.

Currently, based on these two typical information sources, many approaches [1,10,11] are proposed to estimate an overall relationship strength between different users. However, one problem in these approaches is that they confuse all the interaction activities together

* Corresponding author.
E-mail address: zhaoxj01@gmail.com (X. Zhao).
[1] http://www.renren.com/
[2] http://www.facebook.com/
[3] http://www.flickr.com/

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2011.06.036

and neglects the fact that these interaction activities may belong to different activity fields. For example, some of them may belong to the "diet" field, and others may belong to the "sports" field. Furthermore, instead of estimating the overall relationship strength between different users, the relationship strength on a specific activity field, such as "diet", "sports" and "shopping", seems more reasonable and significant. For example, when a user wants to invite some of his/her friends to have a dinner together, he/she usually sends this invitation to those friends who have a higher relationship strength with him/her on the activity field "diet" (it means that they may have higher probability to have the same interests in the "diet" activity field). Another example is that when a user posts a message about his/her work, the responders are usually his/her working cooperator instead of his/her close friends in the "diet" activity field. Therefore, in real application, measuring the relationship strength on a specific activity field is more useful than measuring the overall relationship strength.

Based on this motivation, in this paper we propose a general framework to measure the relationship strengths between different users on various activity fields. Two information sources, the user's profile information and the interaction activities among different users, are used to estimate the relationship strengths. Our approach consists of two sequential steps: assigning an activity field to each interaction activity document, and measuring the relationship strengths on different activity fields. In the first step, we use Latent Dirichlet Allocation (LDA) clustering algorithm [12] to cluster all the interaction activity documents, where each generated cluster belongs to an activity field with a confidence score. Then the activity field of an interaction activity document could be inferred through the activity field of its clusters as well as the similarity between the interaction activity document and its clusters. In the second step, we propose a graphical inference model to model the relationship among the user's profile, interaction activities and the relationship strengths on different fields. By maximizing the joint probability of the data distribution in the graphical model, we obtain the relationship strengths on various activity fields. The overall relationship strength between two users is calculated by weighted summing of all the strengths on each activity field. We conduct enormous experiments on the Facebook dataset, and the results show that our approach could achieve promising performance as compared to the baseline methods. We summarize the main contributions of this paper as follows:

1. To the best of our knowledge, this is the first work that investigates the relationship strength between different users on various activity fields in online social network.
2. In order to improve the accuracy of assigning an activity field to each interaction activity, we first use Latent Dirichlet Allocation algorithm to cluster all the interaction activity documents, and then for each interaction activity document, its activity field is inferred by considering both the activity fields of the clusters which it belongs to and the probabilities of it belonging to each cluster.
3. We propose a graphical inference model to infer the relationship strengths between different users on various activity fields. In this model, the user's profile information, the interaction activities among different users, and the activity field are all considered.

The rest part of this paper is organized as follows. We review the related works in Section 2. In Section 3, we briefly introduce the framework of measuring the relationship strength between different users on various activity fields in online social network. The details of the proposed approach are elaborated in Section 4. In Section 5, we show the initial experimental results of our approach on the Facebook dataset. Finally, we conclude the paper and discuss the directions of the future works in Section 6.

2. Related works

The growth and popularity of the online social network, such as Facebook, MySpace,[4] RenRen, Flickr, and LinkedIn,[5] has lead to a surge in research focused on modeling the online social networks and their properties [13–15]. Many works focused on the analysis of the social network structure and its growth pattern. For example, Backstrom et al. [16] investigated the evolution of the social network structure and the group membership in MySpace and show that the homophily [17] can be used to improve the predictive models of the group membership. Singla and Richardson [11] investigated the correlation between the individual search topics among the users that interact using instant messaging, and show that not only does a correlation exist but that it increases with the amount of time the user communicate. Crandall et al. [18] studied the temporal evolution of the link structure and the attribute similarity in Wikipedia and propose a mathematical model that included both the influence and the homophily effects to predict the future behavior in the online social network. However, nearly all these approaches focused on the descriptive analysis and the generative models of the link structure, based on the observed structure in the online social network.

Some recent efforts had been made on the link prediction which was a formulization of the problem of the predicting future links in an online social network, given a snapshot of the network at the current time step. These link prediction methods can be generally grouped into two approaches: those that used *topological* features to capture the link structure of the social network and those that used the attribute similarity features in addition to *topological* features. For example, Liben-Nowell and Kleinberg [19] proposed an approach to the link prediction based on the measures for analyzing the "proximity" of nodes in the social network. Adamic and Adar [20] investigated the use of ancillary network information but with the goal of predicting social ties.

Another direction of the related research has focused on estimating the relationship strength between different users in online social network. This is the research area that is most relevant to our work in this paper. Kahanda and Neville [21] proposed a supervised learning approach to predict the link strength from the transactional information. Gilbert and Karahalios [22] presented a predictive model that mapped the social media data to the tie strength. However, these works only considered the binary prediction task of distinguish the strong ties from the weak ties. In other words, these works can only predict the strength as strong ties or weak ties rather than represent the full of spectrum of the relationship strength between different users. Xiang et al. [23] developed an unsupervised model to estimate the relationship strength between different users from the interaction activity (e.g., communication and tagging) and the user similarity. More specifically, they formulated a link-based Latent Variable Model, along with a coordinate ascent optimization procedure for the inference. However, it mixed all the interaction activities from various activity fields together, but did not consider that the relationship strengths between the same user pair may be different in various activity fields. And it is a fact that the relationship strength on a specific activity field, such as "diet", "sports" and "shopping", seems more reasonable and meaningful than an overall relationship strength between different users.

Our work differs from the existing efforts in that we focus on modeling the relationship strength rather than the link existence.

[4] http://www.myspace.com/
[5] http://www.linkedin.com/