

# Who Resemble You Better, Your Friends or Co-visited Users<sup>\*</sup>

Jinjing Ma and Yan Zhang<sup>\*\*</sup>

Department of Machine Intelligence, Peking University, Beijing 100871, China  
Jinjing.ma@pku.edu.cn, zhy@cis.pku.edu.cn

**Abstract.** People are alike their graphic neighbors in social networks, which is generally accepted as the basic assumption in interest prediction. But what kind of neighbors is a better information source? This paper aims at answering this question by comparing the results of predicting users' interests in blog social networks with different relationships and parameters. Since social networks usually keep "friends" and "visitors" as basic social roles, we take these two online social relationships as the main information source. In this paper, we discover that (1) combining different information sources might lead to better prediction, and (2) there are many other factors that can affect the results significantly.

**Keywords:** social network, interest prediction.

## 1 Introduction

Interest prediction is widely used in recommendation systems. For example, based on the social correlations among its users, *flickr* provides rankings of contents [3]. And online stores like *dangdang.com* apply collaborative filtering systems to recommend products. In fact, there are two assumptions widely used in recommendation systems.

**Assumption 1.** People share similar interests with their neighbors in social networks;  
**Assumption 2.** Pages clicked by users with similar interests or relation networks tend to share similar contents.

As we can see, these two methodologies are functional in the researches of user interest prediction and applications [1, 2]. Then what will happen if we apply them on the same platform and make a combination? This paper aims at answering this question.

From *Assumption 2*, we have a deduction.

**Deduction 1.** Users visited by the same user probably share similar interests.  
Based on *Deduction 1* we introduce the virtual relation among those visited by the same users (defined as co-visited users).

---

<sup>\*</sup> This work was supported by NSFC with Grant No. 61073081 and 60933004, and HGJ 2010 Grant 2011ZX01042-001-001.

<sup>\*\*</sup> Corresponding author.

To answer the question we put forward at the beginning, we take friendship, 2-degree friendship, and co-visited relation into account. Here, we do prediction with friendship based on *Assumption 1*, with co-visited relation based on *Deduction 1*, and with 2-degree friendship as extension. With which we conclude that the larger number of one's co-visited users gathered information from, the higher possibility we have to infer his overall perspective of interests. We want to make a step in discovering if different relations have different utilities in interest prediction.

In this paper, we choose a blog site, since there are many social networks built with the basic function of a blog site, like posting photos and videos can be regarded as transformed blogs since they are all pointing to expressing. Thus, though SNSs have different information flows when focusing more on user interactions, understanding the structure of a blog site might help us to do more complex researches on social networks like *Facebook*. For example, sharing or forwarding friends' contents can be seen as the combination of accepting and publicizing information. In addition, the relationships on a blog site are clean since there are few relations except friendship. Thus, we can exclude some disturbances in advance.

Within this context we make the following contributions:

- We compare predictions based on different relation networks separately, and we discover that their performances are quite different. When considering different linear combined results of them, we find that some of the combinations lead to improvement but others do not.
- We import some other parameters to modify the results of prediction. And we evaluate their effects on predictions. Thus, we discover that some of the parameters can affect the results to a large extent.

This paper is organized as follows. In Section 2 we review related work to prepare the necessary background for user modeling and prediction method. The methodology including the prediction model and the evaluation principle is presented in Section 3. The way we establish the dataset and comparison experiments performed are discussed in Section 4. In Section 5 we analyze the results of the experiments and conclude with comments and future plans of our work.

## 2 Related Work

Our work, which is focusing on the comparison of predicting results of users' interests with various relationships, is related to prior approaches to interest modeling, social tags as well as friendship in social networks.

In social networks, users may express or hint their interests in various ways. A commonly used source is user-published content, including articles [10], and author-generated tags [5, 6]. And a research classifies the tags into topics and finds that online users use more details to describe the contents they prefer [9]. When comparing different sources for interest prediction, we take advantage of the simplicity of user-generated tags.

Modeling social correlation and network structure is a fundamental work for interest prediction, and has been extensively studied by researchers, including social scientists [11]. When users grow more similar due to social influence and those with