



## Structural link prediction based on ant colony approach in social networks



Ehsan Sherkat<sup>a,b,\*</sup>, Maseud Rahgozar<sup>a</sup>, Masoud Asadpour<sup>b</sup>

<sup>a</sup> Database Research Group, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

<sup>b</sup> Social Networks Lab, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

### HIGHLIGHTS

- This algorithm employs ant colony approach for link prediction in graphs.
- It outperforms some of the unsupervised structural link prediction algorithms.
- A new perspective for link prediction is introduced.

### ARTICLE INFO

#### Article history:

Received 25 April 2014

Received in revised form 8 August 2014

Available online 18 October 2014

#### Keywords:

Structural link prediction

Ant colony algorithm

Social network analysis

Complex networks

### ABSTRACT

As the size and number of online social networks are increasing day by day, social network analysis has become a popular issue in many branches of science. The link prediction is one of the key rolling issues in the analysis of social network's evolution. As the size of social networks is increasing, the necessity for scalable link prediction algorithms is being felt more. The aim of this paper is to introduce a new unsupervised structural link prediction algorithm based on the ant colony approach. Recently, ant colony approach has been used for solving some graph problems. Different kinds of networks are used for testing the proposed approach. In some networks, the proposed scalable algorithm has the best result in comparison to other structural unsupervised link prediction algorithms. In order to evaluate the algorithm results, methods like the top-*n* precision, area under the Receiver Operating Characteristic (ROC) and Precision–Recall curves are carried out on real-world networks.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction and background

This paper is dedicated to the problem of link prediction using ant colony approach. In this section, first the definition, applications and existing solutions for link prediction are introduced then the definition and applications of ant colony approaches for solving social network problems are provided. Finally, a brief introduction of the proposed method based on the ant colony approach is discussed.

*Link prediction definition.* As the size and number of online social networks are increasing day by day, social network analysis has become a popular issue in many branches of science. One of the emerging topics in social network analysis is link prediction. Prediction of a new connection or link between two nodes based on attributes of existing nodes and links in the

\* Corresponding author at: Database Research Group, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran. Tel.: +98 3134401856.

E-mail addresses: [e.sherkat@ut.ac.ir](mailto:e.sherkat@ut.ac.ir) (E. Sherkat), [rahgozar@ut.ac.ir](mailto:rahgozar@ut.ac.ir) (M. Rahgozar), [asadpour@ut.ac.ir](mailto:asadpour@ut.ac.ir) (M. Asadpour).

graph is called link prediction [1]. In other words, by considering the snapshot of a graph in time  $t$ , the goal of link prediction is to predict new links in the graph at time  $t + 1$  [2]. The time  $t + 1$  can be a week, a month, a year or even years after the snapshot  $t$  being taken. In 2005, a new scientific issue called link mining was formally presented to the scientific society, and the link prediction is considered as a subset of link mining [3].

*Link prediction applications.* Link prediction has a different variety of applications. In biological networks like protein–protein interaction and metabolic networks, discovering interactions between proteins is time-consuming and costly because it needs doing many experiments in the laboratory. Instead of doing experiments in the laboratory, if a precise link prediction algorithm predicts interactions, the costs and the number of experiments would be considerably decreased. This problem may also be called as missing link prediction [4]. In many networks, some unwanted links could be created because of noisy information or data. By the help of link prediction methods, these false links can also be detected. This application of link prediction can be used in communication surveillance systems [5]. The evolutionary model of a network can be obtained with link prediction, and it gives scientists a better insight of a network. Understanding the evolving mechanism of a network is one of the key applications of link prediction. In Ref. [6], a novel likelihood approach based on link prediction algorithms is introduced. In medicine, the prevalence of contagious disease can be modeled by a graph, and each predicted link shows a potential infected area in the society [1]. In online social networks, new friends can be offered by predicting common interests among users, and this will increase the loyalty of the members to that social network [7]. Predicting a customer purchase and suggesting it to the customer is another application of link prediction. Sometimes, this problem also called as recommender systems in which a prediction is made on customer's purchases in e-commerce websites [8].

*Link prediction methods.* Link prediction methods can be divided into two general categories of supervised and unsupervised methods. Unsupervised approaches predict links without any prior knowledge or training phases. On the other hand, supervised ones use a model trained and prior knowledge gained from the network, for predicting links.

Unsupervised approaches mostly use similarity or proximity measures and structural attributes of network for predicting links without using any training phase. The length of the shortest path or the number of common neighbors between two nodes are some structural attributes, which are used in these approaches [2]. Jaccard's coefficients [9] approach is similar to common neighbors, except that in this similarity measure, two nodes are more close to each other if they have more common neighbors and less uncommon ones. Node degree is the key structural attribute for prediction of new links. That means, two nodes with a higher degree would have more probabilities to interact with each other in the future. This approach is called preferential attachment [10,11]. Studying the evolution of especial subgraphs is a recent approach that has been used in some link prediction algorithms. As the proposed approach can be categorized in this category of approaches, more details about these approaches are provided in Section 2.

There are also some supervised link prediction approaches. Maximum likelihood based approaches are an example of supervised link prediction algorithms. In Ref. [12], a likelihood approach for inferring hierarchical structure from network data called Dendrogram introduced and showed that the existence of hierarchy can explain how commonly observed topological properties of networks would be reproduced. Probability models are also considered as supervised link prediction algorithms [1]. In these approaches, probability models and network's distribution are the basis for predicting new links. A binary classifier can also be used as a link prediction algorithm. In Ref. [13], some binary classifiers are used for link prediction. Although supervised approaches consider the especial characteristics of each network, they may be time-consuming and may not be suitable for large networks [1].

*Ant colony algorithm.* In the early of the 1990s, ant colony optimization (ACO) was introduced by Dorigo and his colleagues [14]. The goal of this approach was to solve hard combinatorial problems. ACO is based on the foraging behavior of ants. During searching for food, ants randomly search their environment. When the food is found, the quantity and quality of it, is tested by the ant, and then it takes a piece of food. On the way home, a chemical substance known as a pheromone is released by the ant in its path. The amount of released pheromone by the ant is proportional to the quality and quantity of the food source. The pheromone is a kind of stigmergy between ants, and it enables them to find the shortest path from food source to the home [14].

All the problems should be modeled by a graph in which the ACO could be applicable to be used. The ACO is consisting of three main steps. First of all, the probability of traversal of an ant  $k$  from node  $i$  to node  $j$  in the graph is  $p_k^{ij}$ .

$$P_{ij}^k = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{l \in N_i^k} (\tau_{il}^\alpha \eta_{il}^\beta)}. \quad (1)$$

In Eq. (1),  $\tau_{ij}$  is an amount of edge pheromone, and  $\eta_{ij}$  is the cost of traveling from node  $i$  to node  $j$ .  $N_i$  is the neighbors set of node  $i$ . The coefficients  $\alpha$  and  $\beta$  are general parameters of the algorithm. All the edges have the initial amount of pheromone.

The second step is whenever the ant finds a food. The food is the desired solution to the problem. Each ant releases pheromone in the path from the food source to the home based on Eq. (2):

$$\Delta \tau_{ij}^k = \begin{cases} \frac{1}{c^k} & (i, j) \in T^k \\ 0 & \text{otherwise} \end{cases}, \quad \tau_{ij} = \tau_{ij} + \sum_{k=1}^m \Delta \tau_{ij}^k. \quad (2)$$

دريافت فوري

متن كامل مقاله



- ✓ امكان دانلود نسخه تمام مقالات انگلیسي
- ✓ امكان دانلود نسخه ترجمه شده مقالات
- ✓ پذيرش سفارش ترجمه تخصصي
- ✓ امكان جستجو در آرشيو جامعى از صدها موضوع و هزاران مقاله
- ✓ امكان دانلود رايگان ۲ صفحه اول هر مقاله
- ✓ امكان پرداخت اينترنتى با کليه کارت های عضو شتاب
- ✓ دانلود فوري مقاله پس از پرداخت آنلاين
- ✓ پشتيباني كامل خريد با بهره مندي از سيسitem هوشمند رهگيری سفارشات