

# BiNet: Trust Sub-network Extraction using Binary Ant Colony Algorithm in Contextual Social Networks

Xiaoming Zheng  
Department of Computing  
Macquarie University  
Sydney, NSW 2109, Australia  
xiaoming.zheng@mq.edu.au

Yan Wang  
Department of Computing  
Macquarie University  
Sydney, NSW 2109, Australia  
yan.wang@mq.edu.au

Mehmet A. Orgun  
Department of Computing  
Macquarie University  
Sydney, NSW 2109, Australia  
mehmet.orgun@mq.edu.au

**Abstract**—Online Social Networks (OSNs) have become an integral part of daily life in recent years. OSNs contain important participants, the trust relations between participants, and the contexts in which participants interact with each other. All of these have a great influence on the prediction of the trust between a source participant and a target participant, which is important for a participant’s decision-making process in many applications, such as seeking service providers. However, predicting the trust from a source participant to a target one based on the whole social network is not really feasible. Thus, prior to trust prediction, the extraction of a small-scale sub-network containing most of the important nodes and contextual information with a high density rate could make trust prediction more efficient and effective. However, extracting such a sub-network has been proved to be an NP-Complete problem. To address this challenging problem, we propose BiNet: a social context-aware trust sub-network extraction model to search for near-optimal solutions effectively and efficiently. In this model, we first capture important factors that affect the trust between participants in OSNs. Next, we define a utility function to measure the trust factors of each node in a social network. At last, we design a novel binary ant colony algorithm with newly designed initialization and mutation processes for sub-network extraction incorporating the utility function. The experiments, conducted on two popular datasets of Epinion and Slashdot, demonstrate that our approach can extract sub-networks covering important participants and contextual information while keeping a high density rate. Our approach is superior to the state-of-the-art approaches in terms of the quality of extracted sub-networks within the same execution time.

**Keywords**-Trust; Sub-network extraction; Trust prediction;

## I. INTRODUCTION

In recent years, a large and growing number of users have joined e-commerce, online recruitment and social network web sites, such as Facebook, LinkedIn and Renren<sup>1</sup>. Moreover, online social networks (OSNs) have proliferated to be the preferred platforms for a variety of rich activities, such as seeking employees and jobs, and making trustworthy recommendations for products and services. For example, according to a survey on 2600 hiring managers in 2009 by CareerBuilder<sup>2</sup>, 45% of those managers used social

<sup>1</sup><http://www.renren.com/>

<sup>2</sup><http://www.careerbuilder.com/>

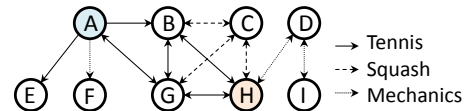


Figure 1. A social network

networking sites to investigate potential employees. In 2012, the ratio increased to 92%. In the context of such activities, *trust* is the commitment to a future action based on a belief that it will lead to a good outcome, despite the lack of ability to monitor or control the environment [5]. Trust is critical for the decision making of users, for example, in a scenario of selecting a service provider based on the recommendations by others. Therefore, effective and efficient approaches and mechanisms are needed to predict the trust between two participants who may not know each other in large-scale OSNs.

OSNs are usually represented as graphs as shown in Fig. 1. A node in the graph represents a participant in an OSN while the edge pointing from one node to an adjacent node corresponds to their real-world or online interactions (e.g.,  $A \rightarrow B$  in Fig. 1). Different types of edges represent different contexts, which refer to any information available for characterizing the participants and the situations of interactions between them [23], e.g., a solid line refers to the relationship in playing tennis, a dashed line refers to squash and a dotted line refers to mechanics in Fig. 1. The trust can be explicitly given by one participant to another based on their history of interactions.

In Fig. 1, suppose  $A$  is looking for a service provider such as a tennis coach and  $H$  is recommended to  $A$  as a tennis coach. But,  $A$  does not know  $H$  before. Here we assume, playing tennis is the *target context*, i.e., the context in which the trust between a source node and a target node needs to be predicted. In  $A$ 's mind,  $B$ ,  $E$  and  $G$  are good tennis players.  $B$ ,  $G$  and  $H$  trust each other and  $G$  also trusts  $A$  regarding tennis playing.  $C$  trusts  $B$ ,  $G$  and  $H$  regarding squash playing and vice versa.  $F$ ,  $D$ ,  $I$  and  $H$  are good machinists. In order to predict if  $H$  will be a good tennis coach in  $A$ ' mind, it is unnecessary to use the whole social network in Fig. 1, because  $F$ ,  $D$  and  $I$  are only good at mechanics while tennis player  $E$  has no knowledge of

others. Let us assume this social network is only constructed in three contexts: tennis, squash and mechanics. In order to boost the efficiency and effectiveness of trust prediction regarding the target context tennis, the social sub-network in Fig. 2 is extracted from Fig. 1 by removing the social relations in mechanics and keeping only the important social relations for the prediction of trust between  $A$  and  $H$  on tennis and squash playing, because mechanics is irrelevant to the target context tennis while squash is relevant to tennis.

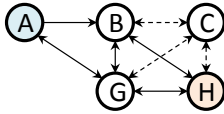


Figure 2. Sub-Network

In the literature, there are some existing works focusing on *trust prediction*, i.e., the process of estimating a new pairwise trust relationship between two nonadjacent participants. Most of these works predict trust based on social graphs (e.g., Fig. 2) using inference approaches [5], [17], while a few of them predict trust from a *trust matrix*, i.e., a representation of social networks in matrices rather than graphs, using matrix factorization approaches [7], [29]. However, all these works assume that the trust network has already been extracted or even the whole dataset is directly used.

An extracted sub-network (e.g., Fig. 2) needs to satisfy the following requirements: (1) it should contain the source node, the target node and most of the nodes which are important for trust prediction between the source node and the target node; (2) the scale of the sub-network/matrix is kept relatively small; and (3) a source participant may introduce constraints of trust relations or contextual information into the sub-network extraction process for various purposes, such as employee recruitment and movie recommendation, which makes the problem more challenging.

Such an extracted trust sub-network can help improve the effectiveness and efficiency of trust prediction [10], [15]. However, extracting such a trust sub-network/matrix is a multi-objective optimization problem, which is known to be NP-complete [1], [15].

In this paper, we propose a social context-aware trust sub-network extraction model, called *BiNet*, which can be used to extract a contextual sub-network for the specific purpose of predicting the trust from a source node to a target node in the target social context of an item to be recommended. The contributions in our proposed model can be summarized as follows: (a) we first present the factors that impact trust between participants in a complex contextual social network, which includes the role impact factor, credibility, preference similarity, social intimacy and existing trust; (b) we propose a trust utility function which takes these impact factors into account to illustrate the attributes of each node in the social network; and (c) inspired by the ant colony foraging process [2], [4], we design a novel binary ant colony algorithm (NBACA) for the sub-network extraction problems, by adding an initialization process and a mutation process

and improving the path selection and pheromone update processes of the conventional binary ant colony algorithm (BACA). The experiments conducted on two popular social network datasets, Epinion<sup>3</sup> and Slashdot<sup>4</sup>, demonstrate the superior performance of our proposed approach over the state-of-the-art approaches.

## II. RELATED WORK

In this section, we briefly review the related works in two categories of social trust prediction and sub-network extraction.

### A. Trust Prediction in Social Networks

The existing works on trust prediction are dominated by inference approaches, as people would naturally seek help or suggestions from trusted friends or friends' friends which forms a trust path based on the trust relationship between them. Golback et al. [5] propose a trust inference mechanism in social networks based on an averaging strategy along trust paths. Li et al. [13], [16] propose a trust inference approach adopting the multiplication strategies to propagate trust along a trust path. Wang et al. [23] propose a probabilistic approach for contextual trust inference.

Recently, trust prediction approaches based on matrix factorization have been proposed and led to better performance [20]. In addition, the participants' similar habits, contexts and profiles can impact the performance of trust prediction. Yao et al. [26] propose a model to predict trust between two participants from important properties including transitivity, multi-aspect and trust bias. Zheng et al. [29] add a propagation and similarity regularization term to matrix factorization incorporating both personal properties and interpersonal properties and achieve better trust prediction performance. Some studies [18], [16] have suggested to predict trust by taking into account some kind of social contextual information.

All of these trust prediction models assume that the trust network or matrix including the source and the target participants has already been extracted. However, sub-network extraction is a necessary step prior to trust prediction as it provides the foundation for trust prediction and has an impact on the performance of trust prediction.

### B. Network Extraction

In the literature, there are only very a few approximation algorithms proposed for the NP-Complete sub-network extraction problem for trust prediction in online social networks. As the resource discovery problem in P2P networks has some similar properties as the trust network extraction problem, some search strategies from P2P networks can be applied to trust network extraction. These approaches can be divided into two groups: *traversal methods* and *heuristic methods*. For the application on small scale datasets, the

<sup>3</sup><http://www.epinions.com/>

<sup>4</sup><http://slashdot.org/>

traversal methods are able to search for the best sub-network by adopting methods, such as breadth-first search and depth-first search. However, on large datasets, it is computationally unfeasible to find the optimum solution, and therefore, heuristic methods can be used to find a near-optimal solution.

Hintsanen et al. [6] propose a model to find the most reliable sub-network. They treat social networks as Bernoulli random graphs and extract a sub-graph by adding paths to the extracted sub-graph one by one till the most reliable status is reached. Liu et al. [14] propose a model to find  $K$  optimal social trust paths for the selection of trustworthy service providers in complex social networks. The  $K$  paths selected from a source participant to a target one actually form a sub-network. Liu et al. [15] propose a social context-aware trust network extraction model, which applies an optimized Monte Carlo method to extract an optimal trust network from the source to the target participants, under user-given constraints of trust network utility yielding the highest utility. These existing works rely on trust paths and do not perform well for matrix factorization based trust prediction approaches as the density of extracted sub-network is not considered and even the state-of-art sub-network extraction models still need improvement in the aspects of both efficiency and effectiveness.

Jang et al. [8] propose a binary ant colony algorithm (BACA) in which the initial pheromone on paths is equally distributed and the path selection only depends on the existing pheromone. This BACA can also be utilized to find the most reliable sub-network. However, when compared with the above approaches, the performance of BACA is not improved much, as the scale of the selected sub-network is affected by the initialization.

### III. INTRODUCTION OF BASIC ACA

As ant colony algorithm (ACA) is a type of efficient approaches with robustness and global searching ability for solving multi-objective optimization problems [25], [4], our proposed sub-network selection approach is based on ant colony algorithms. This section briefly introduces the basic idea of ACA.

ACA was first inspired by the observation of ant colonies in the early 1990's [2], [4]. Ants are social insects and live in colonies. Their behaviors are revealed by the whole colony instead of individuals. When the ants go out to search for food, initially, every ant randomly explores the area around its nest and leaves chemical pheromone on the path it travels. Once an ant finds a food source, it takes part of the food back to the nest leaving pheromone along all the way. When other ants come near the paths with pheromone, they tend to choose the path, with a high probability, with the strongest pheromone which guides other ants to follow the same path. Meanwhile, the pheromone on the path is strengthened by each ant traveling along the path. Thus in the end most ants will follow the same path to the food source from the nest. However, when choosing a path,

individual ants can make incidental mistakes. Therefore, they have a certain probability not to choose the path with the highest pheromone, forming new paths and enabling most ants to find the shortest path [2].

### IV. SUB-NETWORK EXTRACTION USING NBACA

The ant colony algorithm simulates the process of foraging in an ant society. When applying ant colony algorithms to the sub-network extraction problem, a few aspects need to be addressed in the design. In this subsection, we propose our NBACA based model to find a sub-network in a social network including the source and target nodes. While the number of nodes in the extracted sub-network should be as small as possible, it has to contain as many important nodes as possible. Thus, the final solution is a trade-off between a high node importance and the size of the sub-network.

#### A. Pre-processing

As the extracted sub-network is specific for the subsequent trust prediction from a source node to a target node, the information affecting trust prediction must be considered in the sub-network extraction process. In order to utilize social information about participants and their interactions for the sub-network extraction process, in each context, the social information needs to be organized into several aspects, which are called trust impact factors. The details of each trust impact factor are described below.

##### 1) Trust impact factors:

**Role impact factor:** In a certain context, the role impact factor (denoted as  $RIF$ ) illustrates the impact of a participant's social position and expertise on his/her trustworthiness based on the fact that a person who has expertise in a domain is more credible than others with less knowledge [15].

**Reliability:** In a certain context, reliability ( $RLB$ ) measures the rate of a participant's suggestions accepted by others [9]. A participant with high reliability is likely to be sought suggestions from, which can affect the trust towards the participant. The reliability is calculated as one minus the deviation between the predicted rating and the actual ratings of a participant in [9].

**Preference similarity:** It is illustrated in Social Psychology [27] that a participant can trust and have more social interactions with another participant, with whom he/she shares more preferences (e.g., both of them like playing squash). Preference similarity ( $PS$ ) between two participants' preferences can impact the trust between them to some extent [23]. Here,  $PS_{i,j} = PS_{j,i}$  for participants  $i$  and  $j$ .

**Social intimacy:** Social intimacy ( $SI$ ) refers to the frequency of connections between participants in a social network. The degree of social intimacy can impact trust as people tend to trust those participants with more intimate social relationships between them [3].

**Existing trust:** Trust is a belief that an entity, such as a person or an organization, will behave in an expected

manner, despite the lack of ability to monitor or control the environment in which it operates [21]. It can be impacted by all the above properties and the trust value can be greatly different between the same two participants in different interaction contexts [24], [28]. Let  $T_{i,j}$  denote the existing trust participant  $i$  gives to participant  $j$ . A higher  $T_{i,j}$  indicates more trust to  $j$  in  $i$ 's mind. Here, trust between two participants in a given context is not symmetrical, which means  $T_{i,j}$  may not be equivalent to  $T_{j,i}$ .

2) *Utility*: For the extraction of a sub-network which is specific for the prediction of the trust from a source participant to a target one, for each node in the sub-network, we propose a *node utility* (denoted by  $u_i$ ) which is the weighted sum of all the above trust impact factors in a sub-network. It can be formulated by:

$$u_i = F \cdot W' \quad (1)$$

where,  $W$  is a coefficient vector given by users; and  $F = [RIF_i, RLB_i, SI_{s,i}, PS_{s,i}, T_{s,i}, SI_{i,t}, PS_{i,t}, T_{i,t}]$ , is a vector containing all the factors that affect trust between the source participant and target participant. These factors can be divided into three groups: (a)  $RIF_i$  and  $RLB_i$  are independent factors from the source and target participants; (b)  $SI_{s,i}$ ,  $PS_{s,i}$  and  $T_{s,i}$  are the factors revealing the relationship with the source participant; and (c)  $SI_{i,t}$ ,  $PS_{i,t}$  and  $T_{i,t}$  reveal the relationship with the target participant.  $u_i \in [0, 1]$  as the range of each factor is  $[0, 1]$ .

### B. Formulation of the Problem

The problem can be described as finding a certain number of nodes (say  $m$ ,  $0 < m \leq n$ ) out of the  $n$  nodes to compose a sub-network which increases the objective function value as much as possible, the weighted sum of average node utility of the extracted sub-network and the density of the extracted sub-network. The objective function is formulated as:

$$G(X) = \zeta \frac{\sum_{i=1}^n u_i x_i}{\sum_{i=1}^n x_i} + \tilde{\zeta} D(X), \quad (2)$$

where  $X = \langle x_i | i = 1, \dots, n \rangle$  is a vector representing all the nodes of the selected sub-network;  $x_i = 1$  means the  $i$ th node is selected while  $x_i = 0$  means the  $i$ th node is not selected.  $\sum_{i=1}^n x_i$  is the number of the selected nodes.  $D(X)$  is the density of the current sub-network.  $u_i (i = 1, 2, \dots, n)$  is the utility of node  $i$  calculated by Equation (1).  $\zeta$  and  $\tilde{\zeta}$  are the weights.

Therefore, the sub-network extraction problem can be formulated as:

$$\max G(X) \quad s.t. \begin{cases} x_i \in \{0, 1\}, 1 \leq i \leq n \\ x_s + x_t = 2 \\ u_i x_i \geq K_t \end{cases} \quad (3)$$

where,  $x_s$  and  $x_t$  are the selection of the source node and the target node;  $x_s + x_t = 2$  means both of the source node and the target node must be selected;  $K_t$  is a threshold value of the judgment of the important nodes.

### C. The Design of NBACA

For the trust sub-network extraction problem, the conventional BACA [8], [11] has two main disadvantages: (a) the number of the selected nodes in the extracted sub-network fluctuates around the mathematical expectation of the number of the selected nodes in initialization; and (b) the path selection process is only determined by pheromone information without any heuristic function from prior knowledge. These two disadvantages slow down the convergence speed of BACA. To overcome these disadvantages, a novel binary ant colony algorithm (NBACA) is designed.

Fig. 3 shows the designed weighted graph containing  $n+1$  knots (stops) arranged in the order from 1 to  $n+1$ . Ant movement starts from knot 1. At each knot  $i (i = 1, 2, \dots, n)$ , there are 2 directed paths  $a[i, j] (j \in \{0, 1\})$  connecting to knot  $i+1$ . On each path  $a[i, j] (i = 1, 2, \dots, n; j = 0, 1)$ , there is a value  $u_i$ , representing the utility of node  $i$ , and  $U = \{u_i | i = 1, \dots, n\}$ . Therefore, an ant  $k$  going via path  $a[i, 1]$  means the  $i$ th node is selected (noted by  $x_i^k = 1$ ) by the ant  $k$  for the sub-network, while the ant  $k$  going via path  $a[i, 0]$  means the  $i$ th node is not selected ( $x_i^k = 0$ ). The sub-network selected by ant  $k$  is called solution  $k$ , represented by  $X^k = \langle x_i^k | i = 1, \dots, n \rangle$ .

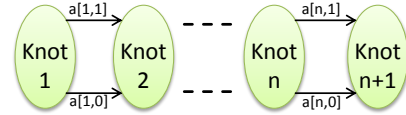


Figure 3. Construction of the weighted graph

Different from the conventional ACA and BACA, in our designed NBACA, the pheromone on path  $a[i, j]$  is represented by a percentage value, denoted by  $\tau_{ij}(t)$ . Therefore, only one of the two paths from knot  $i$  to  $i+1$  is needed to be stored in each iteration, as there are only two paths connecting knot  $i$  and knot  $i+1$  as well as  $\tau_{i0}(t) = 1 - \tau_{i1}(t)$ . The detailed processes in NBACA are summarized in the following subsections.

1) *Initialization*: In ants' natural world, when an ant selects a path for foraging, its selection is affected by the pheromone on each available path. And the path with more pheromone has a higher probability to be selected. Our proposed initialization process is to produce the pheromone of each path at the beginning time  $t = 0$ . The initial pheromone on path  $a[i, 1]$  can be formulated as:

$$\tau_{i1}(0) = \varphi \chi_i + \psi \eta_i \quad (4)$$

where,  $\chi_i \sim \mathcal{N}(\mu, \sigma^2)$  represents a random distribution of the pheromone on each path. As there are two paths connecting two adjacent knots, the mathematical expectation  $\mu$  of the distribution is fixed to 0.5 where variance  $\sigma$  is adjustable.  $\eta_i$  is an expectation heuristic function representing nodes' importance in a social network. Here,  $\eta_i = u_i / \max\{u_j | j = 1, \dots, n\}$  is the rate of node  $i$ 's utility to the maximum node utility in the whole social network. Coefficients  $\varphi$  and  $\psi$  represent the weights of  $\chi_i$  and  $\eta_i$

respectively. In addition, the initial pheromone on path  $a[i, 0]$  is  $\tau_{i0}(0) = 1 - \tau_{i1}(0)$  in our NBACA.

2) *Path Selection:* On path  $a[i, j]$  at time  $t$  ( $t = 0, 1, 2, \dots$ ), the probability of the path to be selected is determined by both the pheromone  $\tau_{ij}(t)$  and the heuristic function value  $\eta_i$ . At time  $t = 0$ ,  $y$  ants are created and put on knot 1. Then each ant selects a path and moves to the next knot according to both the pheromone on each path and the heuristic function. This process continues till the ant reaches the terminal knot  $n + 1$ . The solutions are represented by  $\{X^k | k = 1, \dots, y\}$ .

At time  $t$ , the transition probability of ant  $k$  ( $k = 1, 2, \dots, y$ ) moving from knot  $i$  ( $i = 1, 2, \dots, n$ ) to knot  $i + 1$  via path  $a[i, 1]$  is:

$$p_{i1}^k(t) = \alpha \tau_{i1}(t) + \beta \eta_i \quad (5)$$

where,  $\eta_i$  is a heuristic function value which is the same as that in the initialization process. The larger  $\eta_i$  is, the more likely ant  $k$  selects node  $i$  (goes via the path  $a[i, 1]$ ).  $\alpha$  and  $\beta$  are the weights of the pheromone and the heuristic function when ants select the paths. In addition, the transition probability of moving via  $a[i, 0]$  is:

$$p_{i0}^k(t) = 1 - p_{i1}^k(t). \quad (6)$$

3) *Mutation:* The mathematical expectation of the number of paths  $a[i, 1]$  (selected nodes from a social network) is affected by the initialization[11]. In order to increase the variance of the number of selected nodes in each iteration, we propose a mutation process: an ant has a certain probability to make mistakes in remembering the paths it has passed through. This process is able to produce new solutions especially with a different number of selected nodes, which broadens the scope of the search for a solution.

In the mutation process, ants can more easily forget the selection of paths corresponding to the unimportant nodes in a social network. For each ant  $k$ , this process is formulated as:

$$X_-^k = \langle x_i^k * 1\{p_{i1}^k > \lambda_i\} | i = 1, \dots, n \rangle \quad (7)$$

$$X_+^k = \langle 1\{(x_i^k + 1\{p_{i1}^k > \lambda_i\}) > 0\} | i = 1, \dots, n \rangle \quad (8)$$

where  $\lambda_i \sim \mathcal{U}(0, 1)$  is a normal number obtained from a continuous uniform distribution.  $1\{\cdot\}$  is a boolean function. It is equal to 1 if the condition in  $\{\cdot\}$  is satisfied. Otherwise it equals 0. Eqs. (7) and (8) are able to generate solutions with less and more numbers of selected nodes from a social network respectively. Finally, an ant  $k$  obtains 3 different solutions after finishing its trip  $\{X^k, X_-^k, X_+^k\}$ .

4) *Upgrade of Pheromone:* When all the  $y$  ants reach the terminal knot,  $3y$  feasible solutions  $\{X^k, X_-^k, X_+^k | k = 1, \dots, y\}$  can be obtained. If the best solution in the current iteration (denoted by  $X'_{best}$ ) is better than that of all the past iterations (the best-so-far solution, denoted by  $X_{best}$ ), then,  $X_{best} = X'_{best}$  and the pheromone on each path  $i$  is upgraded accordingly. The pheromone upgrade procedure consists of two parts: (a) the pheromone evaporation procedure which is applied to path  $i$  if node  $i$  is not selected for the so-far-

best solution ( $x_i = 0$  and  $x_i \in X_{best}$ ) and is formulated as:

$$\tau_{i1}(t+1) = (1 - \rho)\tau_{i1}(t) \quad (9)$$

and (b) the pheromone intensification procedure which is only available for path  $i$  if node  $i$  is selected for the best-so-far solution ( $x_i = 1$  and  $x_i \in X_{best}$ ) and formulated as:

$$\tau_{i1}(t+1) = \tau_{i1}(t) + \rho |1 - \tau_{i1}(t)| \quad (10)$$

where  $\rho$  ( $0 < \rho < 1$ ) is a parameter called the pheromone evaporation rate and  $(1 - \rho)$  is the pheromone residue rate;  $\rho$  ( $0 < \rho < 1$ ) is the pheromone increment rate on path  $a[i, j]$  in the current iteration.

Because  $p_{i0}^k(t) = 1 - p_{i1}^k(t)$  and the pheromone in our NBACA is presented as a percentage, the evaporation and intensification processes need to be conducted only on the paths of  $a[i, 1]$  ( $i = 1, 2, \dots, n$ ).

**Algorithmic process:** The main process of finding the sub-network in a social network using NBACA is described in Algorithm 1. In each iteration, the best solution, among the solution sets  $X^{(NC)} = \{X^k, X_-^k, X_+^k | k = 1, \dots, y\}$  obtained from the  $y$  ants in the current iteration, is used to update the pheromone, if it is better than the so-far-best solution. Then, the next iteration starts using the updated pheromone information. Finally, the iteration process ends returning the best solution, when  $NC$ , the number of iterations already run, reaches the maximum value  $NC_{max}$  or  $NF$ , the number of iterations where the best-so-far solution stayed the same, reaches the preset maximum value  $NF_{max}$ .

**Summary:** Different from the conventional BACA, the design of our NBACA is greatly improved in each step. Its characteristics are as follows: (a) An initialization process is introduced. The pheromone on each path is initialized by both a random value from a Gaussian distribution and the node utility proposed in Section IV-A, which improves the initial probability of each path to be selected in ants' initial movement. (b) In our designed NBACA, the pheromone information is represented by the percentage value, which reduces the storage of pheromone information and simplifies path selection and pheromone update processes resulting in a reduction of execution time. (c) A mutation strategy is first introduced into conventional BACA, which increases the solution range in each iteration and speeds up the convergence of the NBACA. And (d) our designed NBACA is not limited to the application of sub-network extraction problems and it can be applied to any case where the conventional ACA is applicable.

## V. EXPERIMENTS

We have conducted experiments on two popular social network datasets Epinion and Slashdot [12], and compared the performance of our BiNet with two state-of-the-art approaches SCAN [15] and FDRS [6] and a baseline approach BACO [8].

---

**Algorithm 1:** Binary Ant Colony Algorithm

---

**Data:**  $\varphi, \psi, \alpha, \beta, \rho, \varrho, y, n, NC_{max}, NF_{max}$ **Result:** The best solution when the iteration ends

```
1 begin
2   Initialize  $NC$  &  $NF$  &  $X_{best}$ ;
3   while  $NC < NC_{max}$  &  $NF < NF_{max}$  do
4     Produce  $y$  ants and put them on knot 1;
5     Initialize  $\tau_{i1}(0) (i = 1, 2, \dots, n)$ ;
6      $NC = NC + 1$ ;
7     for each ant  $k (k = 1, 2, \dots, y)$  do
8       for each movement  $i (i = 1, \dots, n)$  do
9         Select next path  $a[i, j]$  via Eq.(4);
10         $x_i^k = j$ ;
11        Return  $X^k = \langle x_i^k | i = 1, \dots, n \rangle$ ;
12        Get  $X_-^k$  &  $X_+^k$  via Eq.(7&8);
13       $X^{(NC)} = \{X^k, X_-^k, X_+^k | k = 1, \dots, y\}$ ;
14       $X'_{best} = \text{arcmax}G(X^{(NC)})$ ;
15      if  $G(X'_{best}) > G(X_{best})$  then
16         $X_{best} = X'_{best}$ ;
17        Upgrade pheromone via Eq. (9&10);
18      else
19         $NF = NF + 1$ ;
20    Return  $X_{best}$ ;
```

---

### A. Dataset Description

Although, there are a number of studies on mining a single impact factor in social networks [19], [22], there is no dataset in place that contains all the contextual values we need. Thus, the experiments are conducted on semi-synthetic datasets which consist of the datasets from real social networks of **Epinion** (131,828 nodes and 841,372 edges) and **Slashdot** (82,144 nodes and 549,202 edges), and synthetic trust impact factor values. In order to demonstrate that the performance of our model is not data sensitive, 10 groups of the trust impact factor values are randomly generated for both Epinion and Slashdot datasets in the experiments respectively.

### B. Comparisons

In order to evaluate the performance of our proposed model BiNet, we compare it with two state-of-the-art models SCAN and FDRS and a baseline model BACO:

**SCAN** is a social context-aware trust network discovery approach which considers social contextual impact factors and finds the context-aware trust network under certain constraints of each trust impact factor, by adopting a Monte Carlo search method with optimization strategies [15].

**FDRS** is a fast discovery approach of reliable sub-networks which treats a social network as a Bernoulli random graph and builds up the sub-graph by incrementally adding paths from a source node to a target node to an

initially empty sub-graph until the addition of any paths will not increase the objective function value of the whole sub-network [6].

**BACO** is a binary ant colony optimization algorithm which is based on the concept and principles of ant colony optimization to solve the binary and combinatorial optimization problems. It can be applied, as a baseline approach, to extract the trust sub-network when the selection of a node in a social network is treated as the selection from binary paths in ants' movements [8].

In addition, all three models are coded and executed in Matlab R2012B on a desktop empowered with an Intel i7-2600 CPU and an 8G memory running Windows 7 64-bit Professional.

### C. Experiment Setting

In order to compare the differences in efficiency and effectiveness between our proposed model BiNet and each of SCAN, FDRS and BACO, experiments are conducted on both Epinion and Slashdot datasets enhanced with synthetic trust impact factor values respectively to find the near optimal sub-network for each of 10 pairs of nodes which are randomly selected, with different social connection degrees, as the source-target node pairs.

In the experiments on the Epinion dataset, the sub-network extraction for each source-target node pair is performed on the trust relationships of the Epinion dataset with each of the 10 groups of trust impact factor values. Likewise, in the experiments on the Slashdot dataset, the sub-network extraction for each source-target node pair is performed with each of the 10 groups of trust impact factor values.

Then, each of BiNet, SCAN, FDRS and BACO is run for 10 times as cross validation. In total, each model is run for 2000 times (2 datasets  $\times$  10 groups of impact factor values  $\times$  10-time cross validation  $\times$  10 source-target pairs). The average results on both datasets are plotted in Figs. 4-5 respectively.

Parameters, such as the trust factor constraints, only affect the utility values obtained in the experiments but do not affect the performance comparison between different models, as all models are compared on the same datasets. This type of parameters are given by users in application. Here, equal weights are adopted in the experiments. Other parameters, such as the ones in the ant colony algorithm, are determined by the experiments of the models, where  $\zeta = \varsigma = 0.5$ ,  $\varphi = 0.2$ ,  $\psi = 0.8$ ,  $\alpha = \beta = 0.5$ ,  $\rho = 0.1$ ,  $\varrho = 0.1$ ,  $K_t = 0.5$ ,  $NC_{max} = 400$  and  $y = 40$ .

### D. Results and Analysis

**Results:** BiNet, SCAN and BACO are all iterative algorithms whose results get better as the time goes. In Figs. 4-5, we present the mean results over each group of datasets delivered within the first 40 seconds time limitation which are sufficient to demonstrate the performance of each model, as in real applications, we cannot really execute the models for such a long time. The best, mean and worst results on



Table I  
THE RESULTS AT 40<sup>th</sup> SECOND

Dataset	Cases	BiNet	SCAN	FDRS	BACO	BiNeT v.s. SCAN	BiNet v.s. FDRS	BiNet v.s. BACO
Epinion	Min	0.575	0.540	0.485	0.294	6.1% higher	15.7% higher	48.9% higher
	<b>Mean</b>	0.631	0.587	0.541	0.301	<b>6.9% higher</b>	<b>14.3% higher</b>	<b>52.3% higher</b>
	Max	0.69	0.612	0.581	0.312	11.3% higher	15.8% higher	54.8% higher
Slashdot	Min	0.529	0.503	0.513	0.279	4.9% higher	3.0% higher	47.4% higher
	<b>Mean</b>	0.611	0.559	0.561	0.285	<b>8.5% higher</b>	<b>8.2% higher</b>	<b>53.4% higher</b>
	Max	0.657	0.599	0.598	0.292	8.8% higher	9.0% higher	55.6% higher

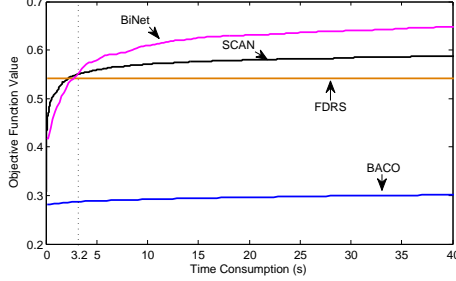


Figure 4. Results on Epinion Dataset

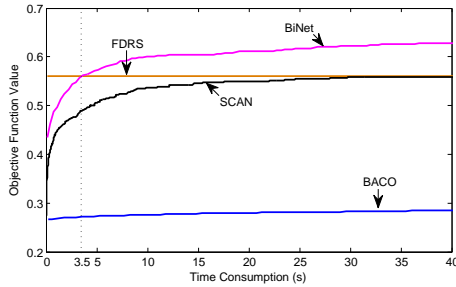


Figure 5. Results on Slashdot Dataset

each group of datasets at the 40<sup>th</sup> second are presented in Table I. As the FDRS model is not an iterative model, it yields one fixed result on each dataset using over 100 seconds.

Fig. 4 shows the average objective function values (Eq.(2)) of the sub-networks extracted by all the four models on the Epinion dataset within the first 40 seconds. As time goes on, the best-so-far solutions of BiNet, SCAN and BACO become better, while the result of FDRS keeps unchanged as it is not an iterative algorithm. Our proposed BiNet outperforms all other three models after 3.2 seconds. At the 40<sup>th</sup> second, the average objective function value of the sub-network delivered by BiNet is 6.9% higher than the one delivered by SCAN, 14.3% higher than the one delivered by FDRS, and 52.3% higher than the one delivered by BACO.

Fig. 5 shows the average objective function values of the sub-networks extracted by the four models on the Slashdot dataset within the first 40 seconds. On this dataset, our BiNet outperforms both SCAN and BACO from the very beginning and outperforms FDRS after 3.5 seconds. At the 40<sup>th</sup> second, the average result of BiNet is 8.5% higher than that of SCAN, 8.2% higher than that of FDRS, and 53.4%

higher than that of BACO.

**Analysis:** The differences between our BiNet model and SCAN, FDRS and BACO on both datasets, especially, the significant improvement between BiNet and BACO, mainly come from the following aspects: (a) the initialization process of our proposed algorithm set up our model using both random values from normal distribution and a heuristic function from nodes utilities, which makes our model able to find solutions with a relatively high utility from the first iteration without losing diversity; (b) the mutation process generates an extra solution with a smaller number of selected nodes and another extra solution with a larger number of selected nodes in each iteration, and thus can broaden the scope of search in each iteration; (c) the percentage representation of pheromone information only needs to record the pheromone on half the number of paths, and thus can save memory and reduce execution time; (d) our designed model selects nodes from both pheromone information and a heuristic function, which speeds the convergence and betters the performance within a fixed time; (e) pheromone update process is kept simple, which makes our model more efficient; and (f) the big improvement between BiNet and BACO demonstrates that our proposed model significantly overcomes the conventional disadvantages of BACAs.

In addition, our proposed model outperforms the other models in all the three cases presented in Table I shows that our model is not data sensitive and applies to a wide scope of applications.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a social context-aware trust sub-network extraction model to search near-optimal solutions effectively and efficiently. In this model, first, we have discussed the impact factors that affect the trust of a participant in another participant's opinion in online social networks. Then, we have proposed a trust utility function that takes these impact factors to illustrate the attribute of each node. After that, we have proposed a novel binary ant colony algorithm incorporating the utility function. In this newly designed algorithm, an initialization process and a mutation process are added, while the conventional path selection and pheromone update process are improved for sub-network extraction. The experiments, conducted on Epinion and Slashdot datasets enhanced with synthetic trust impact factor values, demonstrate that our proposed model outper-

forms the existing comparable heuristic methods in terms of the quality of extracted trust sub-networks. In particular, our newly designed NBACA overcomes the disadvantages of the conventional BACA for sub-network extraction. Moreover, our proposed NBACA has the same scope of applications as the conventional BACA, but achieves significantly better performance.

In the future, we plan to further improve the performance of the sub-network extraction model as well as the binary ant colony algorithm for better efficiency and effectiveness.

#### REFERENCES

- [1] S. Baase and A. V. Gelder. *Computer Algorithms: Introduction to Design and Analysis*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1999.
- [2] C. Blum. Ant colony optimization: Introduction and recent trends. *Physics of Life Reviews*, 2(4):353–373, 2005.
- [3] S. Brehm. *Intimate relationships*. Random House, 1985.
- [4] M. Dorigo. *Optimization, Learning and Natural Algorithms*. PhD thesis, Politecnico di Milano, Italy, 1992.
- [5] J. Golbeck and J. A. Hendler. Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, 6(4):497–529, 2006.
- [6] P. Hintsanen, H. Toivonen, and P. Sevon. Fast discovery of reliable subnetworks. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 104–111, 2010.
- [7] J. Huang, F. Nie, H. Huang, Y. Lei, and C. Ding. Social trust prediction using rank-k matrix recovery. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI'13*, pages 2647–2653. AAAI Press, 2013.
- [8] S. Jang, J. Roh, W. Kim, T. Sherpa, J. Kim, and J. Park. A novel binary ant colony optimization: Application to the unit commitment problem of power systems. *Journal of Electrical Engineering & Technology*, 6(2):174–181, 2011.
- [9] D. Jia, F. Zhang, and S. Liu. A robust collaborative filtering recommendation algorithm based on multidimensional trust model. *JSW*, 8(1):11–18, 2013.
- [10] C. Jones and B. Randell. Dependable pervasive systems. In R. Mansell and B. S. Collins, editors, *Trust and Crime in Information Societies*. Edward Elgar, 2005.
- [11] M. Kong, P. Tian, and Y. Kao. A new ant colony optimization algorithm for the multidimensional knapsack problem. *Computer and Operations*, 35:2672–2683, 2008.
- [12] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Signed networks in social media. In *28th ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1361–1370, 2010.
- [13] L. Li, Y. Wang, and E.-P. Lim. Trust-oriented composite services selection and discovery. In L. Baresi, C.-H. Chi, and J. Suzuki, editors, *ICSOC/ServiceWave*, volume 5900 of *Lecture Notes in Computer Science*, pages 50–67, 2009.
- [14] G. Liu, Y. Wang, and M. Orgun. Finding k optimal social trust paths for the selection of trustworthy service providers in complex social networks. In *2011 IEEE International Conference on Web Services (ICWS)*, pages 41–48, July 2011.
- [15] G. Liu, Y. Wang, and M. A. Orgun. Social context-aware trust network discovery in complex contextual social networks. In *AAAI*, pages 101–107, 2012.
- [16] G. Liu, Y. Wang, M. A. Orgun, and E.-P. Lim. A heuristic algorithm for trust-oriented service provider selection in complex social networks. In *Services Computing (SCC), 2010 IEEE International Conference on*, pages 130–137, July 2010.
- [17] G. Liu, Y. Wang, M. A. Orgun, and E.-P. Lim. Finding the optimal social trust path for the selection of trustworthy service providers in complex social networks. *IEEE Transactions on Services Computing*, 6(2):152–167, 2013.
- [18] X. Liu. Towards context-aware social recommendation via trust networks. In X. Lin, Y. Manolopoulos, D. Srivastava, and G. Huang, editors, *WISE (1)*, volume 8180 of *Lecture Notes in Computer Science*, pages 121–134. Springer, 2013.
- [19] A. McCallum, X. Wang, and A. Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *JAIR*, 30(1):249–272, Oct. 2007.
- [20] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, volume 20, pages 1–8, 2008.
- [21] W. Sherchan, S. Nepal, and C. Paris. A survey of trust in social networks. *ACM Computing Surveys*, 45(47):47:1–47:33, 2013.
- [22] C. Wang, J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. Mining advisor-advisee relationships from research publication networks. In *KDD'10*, pages 203–212, 2010.
- [23] Y. Wang, L. Li, and G. Liu. Social context-aware trust inference for trust enhancement in social network based recommendations on service providers. *World Wide Web Journal (WWWJ)*, 18(1):159–184, 2015.
- [24] Y. Wang and V. Varadharajan. Role-based recommendation and trust evaluation. In *CEC/EEE*, pages 278–288. IEEE Computer Society, 2007.
- [25] X. Xiao, D. Xiao, J. Lin, and Y. Xiao. Overview on multi-objective optimization problem research. *Application Research of Computers*, 28(3):806–808, 2011.
- [26] Y. Yao, H. Tong, X. Yan, F. Xu, and J. Lu. Matri: a multi-aspect and transitive trust inference model. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1467–1476, 2013.
- [27] R. B. Zajonc. Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, 10(6):224–228, 2011.
- [28] H. Zhang, Y. Wang, and X. Zhang. Efficient contextual transaction trust computation in e-commerce environments. In *IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pages 318–325, 2012.
- [29] X. Zheng, Y. Wang, M. A. Orgun, Y. Zhong, and G. Liu. Trust prediction with propagation and similarity regularization. In *AAAI*, pages 237–243, Quebec City, Quebec, Canada, July 27-31 2014.