Discovering Trust Networks for the Selection of Trustworthy Service Providers in Complex Contextual Social Networks

Guanfeng Liu, Yan Wang and Mehmet A. Orgun
Department of Computing
Macquarie University
Sydney, NSW, Australia, 2109
{*guanfeng.liu, yan.wang, mehmet.orgun*}*@mq.edu.au*

Huan Liu
Computer Science and Engineering
Arizona State University
Tempe, AZ 85287, USA
huan.liu@asu.edu

Abstract—**Online Social Networks (OSNs) have provided an infrastructure for a number of emerging applications in recent years, e.g., for the recommendation of service providers, where trust is one of the most important factors for the decision-making of service consumers. In order to evaluate the trustworthiness of a service provider (i.e., the target) without any prior interaction with a service consumer (i.e., the source), the trust network from the source to the target need to be extracted firstly before performing any trust evaluation, as it contains some important intermediate participants, the trust relations between the participants, and the social context, each of which has an important influence on trust evaluation. However, the network extraction has been proved to be NP-Complete.**

Towards solving this challenging problem, we first propose a complex contextual social network structure which considers some social contexts, having significant influences on both social interactions and trust evaluation between participants. Then, we propose a new concept called QoTN (Quality of Trust Network) and a social context-aware trust network discovery model. Finally, we propose a Heuristic Social Context-Aware trust Network discovery algorithm (H-SCAN) by adopting the K-Best-First Search (KBFS) method and our optimization strategies. The experimental results illustrate that our proposed model and algorithm outperform the existing methods in both algorithm efficiency and the quality of the extracted trust networks.

Keywords: trust, social networks, service provider selection

I. INTRODUCTION

A. Background and Motivations

Online Social Networks (OSNs), such as Facebook (facebook.com) and LinkedIn (linkedin.com), have already attracted lots of participants, and have been used as a means for a variety of activities. For example, according to a survey on 2600 hiring managers in 2008 by CareerBuilder (careerbuilder.com, a popular job hunting website), 22% of those managers used social networking sites to investigate potential employees. The ratio increased to 45% in June 2009 and 72% in January 2010. In service-oriented environment, OSNs can provide the infrastructure for the recommendation of web service providers or the recommendation of web services. In such an activity, trust is one of the most important factors for the decision making of service consumers [17], requiring approaches and mechanisms for evaluating the trustworthiness between participants who are unknown to each other.

In OSNs, each node represents a participant and each link between two nodes corresponds to a real-world or online
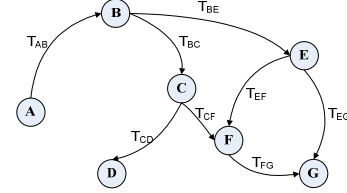

Figure 1. A social network

interaction (e.g., $A \rightarrow B$ and $B \rightarrow C$ in Fig. 1). For adjacent participants (i.e., those nodes with a direct link between them), the trust value between them could be explicitly given by one to the other based on their direct interaction (e.g., T_{AB} and T_{BC}). As each participant usually interacts with many others, multiple trust paths may exist between nonadjacent participants (e.g., path $A \rightarrow B \rightarrow E \rightarrow G$ and $A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$ in Fig. 1) from the source participant (e.g., A) to the target participant (e.g., G), which forms a *trust network*. As indicated in the disciplines of Social Psychology [26] and Computer Science [17, 22], such a trust network can provide the basis for evaluating the trustworthiness of the target. This is because that it contains some important intermediate participants, the trust relations between them and the social context underlying their happened interactions. These information have important influences on trust relationships and trust evaluation. In the literature, there have been some existing trust evaluation approaches that evaluate the trust value between any two nonadjacent participants [17, 23, 24]. However, they all assume the trust network between the two participants have been identified. Namely, extracting such a contextual trust network between two nonadjacent participants is an essential step before performing any trust evaluation between them. The trust network discovery aims to identify a trust network between two nonadjacent participants with less nodes and/or less links but including the important intermediate participants, their trust relations, and the social context. With such a process, trust evaluation methods can work more efficiently and effectively.

However, extracting a trust network between two nonadjacent participants involves finding the longest simple path (a simple path is an acyclic path) in a graph, which has been proved to be an NP-Complete problem [5]. Thus there is no polynomial deterministic algorithm for it. Alternatively, since the resource discovery problem in P2P networks [2] has similar properties, heuristic search strategies developed for P2P network resource discovery can be used. But these methods do not consider the social context in social networks, including social relationships, social positions, residential location and preferences of participants, which have significant influences on both social interactions and trust evaluation [10, 21, 36].

In addition, a source may have different purposes to evaluate the trustworthiness of the target, e.g., looking for a potential employee, or a movie. Thus, to obtain a more trustworthy trust evaluation result, a source participant may specify some constraints of the attributes in social context as the trust evaluation criteria in trust network discovery. However, this capability has not been supported in existing methods. Therefore, it is a significant and challenging problem to extract the contextual trust network which should satisfy the constraints specified by a source and is expected to contain the most trustworthy trust evaluation result.

B. Objectives and Contributions

In this paper, we aim to solve the contextual trust network discovery problem. Our contributions in this paper are summarized as follows.

- As the social network structure in real world scenarios is not well defined, we first propose a new complex contextual social network structure which contains social contextual impact factors, including trust, social intimacy degree, community impact factor, preference similarity and residential location distance. The new network structure reflects the social networks found in the real world better than existing OSNs (*see Section III*).
- Then, we propose several social contextual impact factors and a new concept QoTN (Quality of Trust Network), taking these impact factors as attributes to illustrate the capability of a contextual trust network to guarantee a certain level of trust in trust evaluation. After that we propose a social context-aware trust network discovery model (*see Section IV*).
- Since trust network discovery with QoTN constraints involves finding the longest simple path (a simple path is an acyclic path) in a graph, which has been proved to be an NP-Complete problem [5], we propose a novel Heuristic Social Context-Aware trust Network discovery algorithm, called H-SCAN, by adopting the K-Best-First Search (KBFS) method and our optimization strategies (*see Section V*).
- We have conducted experiments on a real social network dataset, Enron emails[1]. Experimental results demonstrate the superior performance of our proposed model and the H-SCAN algorithm (*see Section VI*).

II. RELATED WORK

A. Social Network Analysis

The studies of social network properties can be traced back to 1960's when the *small-world*[2] characteristic in social networks was validated by Milgram [28]. In an experiment of mail sending, the average path length between two Americans was about 6.6 hops. In recent years, sociologists and computer scientists have started to investigate the characteristics of popular OSNs. According to an experiment at Columbia University [34], the average path length between two people who have interactions through email is about five to seven hops. Mislove *et al.* [29] analyzed several popular online social networks including Facebook, MySpace (myspace.com) and Flickr (flickr.com), and validated the *small-world* and *power-law* characteristics (i.e., in a social network, the probability that a node has degree

k is proportional to k^{-r}, $r > 1$) of online social networks using data mining techniques.

B. Trust in Social Networks

Trust is a critical factor in the decision-making of participants in OSNs [20]. Golbeck *et al.* [17] proposed a trust inference mechanism based on averaging the trust values between the participants along the social trust paths between a source and a target. In addition, Liu et al. [22] have proposed a heuristic algorithm to identify the most trustworthy social trust path between two unknown participants, which helps evaluate the trustworthiness of the target. They further proposed a heuristic algorithm for the selection of K ($K \geq 2$) social trust paths [23] and a novel trust transitivity model in social networks [24]. All of these trust models assume that the trust network from the source to the target already exists. Therefore, the trust network discovery is a necessary step because it provides a foundation to apply the above promising trust evaluation models.

C. Network Discovery

To the best of our knowledge, in the literature, there are no approximation algorithms proposed for the NP-Complete trust network discovery in OSNs. But, the resource discovery problem in P2P networks has the similar properties as the trust network discovery problem. Thus, some search strategies developed for the resource discovery problem can be applied in trust network discovery. These strategies can be classified into three categories.

1) Flooding-Based Search (FBS): The flooding-based mechanism searches the network from the source by using the Breadth First Search (BFS) strategy, which was applied into Gnutella (rfc-gnutella.sourceforge.net). Since this search strategy consumes huge computation time, the Time To Live Breadth First Search (TTL-BFS) method [11, 14] was proposed. In TTL-BFS, the Time To Live (TTL) is set to either a fixed or an unfixed value (e.g., the current TTL is doubled at each layer of search if not finding the target [7]) and its value is decreased by 1 or Vr ($0 < Vr < 1$) after each layer of BFS. During the process, if the target is found, the search terminates. Otherwise, TTL-BFS repeats BFS until $TTL = 0$ or the target is found.

2) Random Walk Search (RWS): RWS [16, 39] firstly searches all the neighboring nodes of the source. If the target is discovered, then the search terminates. Otherwise, the method randomly selects one of the current node's neighbors as the expansion node for the next step of the search.

3) High Degree Search (HDS): The HDS method [2, 38] firstly calculates the outdegree of each of the neighboring nodes of a source and selects the one with the maximal outdegree. If the selected node is the target, then the search terminates. Otherwise, at the selected node, HDS repeats the ourdegree calculation and node selection.

Summary: The above search strategies have good performance in P2P networks which do not contain social contextual information. In addition, they do not support criteria specification in trust evaluation. Thus, existing methods cannot be expected to extract a trust network to deliver a trustworthy trust evaluation result satisfying the constraints in social contexts.

III. COMPLEX CONTEXTUAL SOCIAL NETWORKS

In this section, we propose a new complex contextual social network structure, containing the social contextual impact factors which are not fully considered in existing social network

---

[1] http://www.cs.cmu.edu/ enron/

[2] The average path length between any two nodes is about 6.6 hops in a social network [28].

structures [23, 24], reflecting the social networks in the real world better.

A. Social Contextual Impact Factors

1) **Trust:** Trust is the belief of one participant in another, based on their interactions, in the extent to which the future action to be performed by the latter will lead to an expected outcome [22]. As pointed out in [33, 26], the trust value between two people can be different in different domains. For example, A trusts B in teaching C++, but A may not trust B in repairing a car. In our model, let $T_{AB}^{D_i} \in [0,1]$ denote the trust value that A assigns to B in domain i. If $T_{AB}^{D_i} = 0$, it indicates that A completely distrusts B in domain i, while $T_{AB}^{D_i} = 1$ indicates A completely believes B's future action can lead to the expected outcome in that domain.

2) **Social Intimacy Degree:** As illustrated in Social Psychology [4, 10], a participant can trust and have more social interactions with the participants with whom he/she has more intimate social relationships than those with whom he/she has less intimate social relationships. Let $SI_{AB} \in [0,1]$ denote the *Social Intimacy Degree* between A and B in online social networks. $SI_{AB} = 0$ indicates that A and B have no social relationship while $SI_{AB} = 1$ indicates they have the intimate social relationship.

3) **Community Impact Factors:** Rich activities of participants in social networks can be categorized into different domains (e.g., hiring employees or product sales) based on their characteristics [33]. As illustrated in Social Psychology [3, 12], in a certain domain of interest, an expert's recommendation is more credible than that from a beginner. In addition, as illustrated in Cognitive Science [19] and Computer Science [30], a person who was contacted by many people (i.e., a larger indegree of the node) is more credible than that a person who was contacted by few people. Therefore, let $CIF_A^{D_i} \in [0,1]$ denote the value of the *Community Impact Factor*, illustrating the community impact of participant A in domain i. $CIF_A^{D_i} = 1$ indicates that A is a domain expert and has the greatest impact in domain i while $CIF_A^{D_i} = 0$ indicates that A has no knowledge and has the least impact in that domain.

4) **Preference Similarity:** As illustrated in Social Psychology [25, 37], a participant may trust and have more social interactions with another participant, with whom he/she shares more preferences (e.g., both of them like playing badminton) than those, with whom he/she shares fewer preferences. Let $PS_{AB}^{D_i} \in [0,1]$ denote the value of the *Preference Similarity* between A and B in domain i. When $PS_{AB}^{D_i} = 0$, A and B have no similar preference in the domain. When $PS_{AB}^{D_i} = 1$, they have the same preference in that domain.

5) **Residential Location Distance:** As illustrated in Social Psychology [6, 15], a participant may have more trust and have more social interactions with another whose residential location is close to that of the participant (e.g., in the same neighborhood) than those whose residential locations are far away. Let $RLD_{AB} \in [0,1]$ denote the *Residential Location Distance* between A and B. When $RLD_{AB} = 1$, the residential location of A and B are the same. When $RLD_{AB} = 0$, it indicates that the residential location between them are total different.
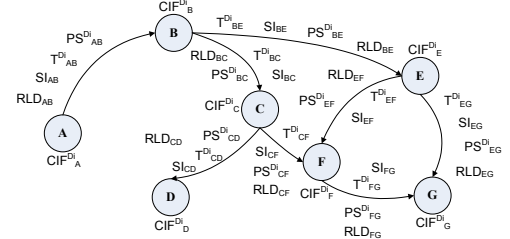

Figure 2.   A complex contextual social network

Although it is difficult to build up comprehensive social intimacy degree, community impact factor, preference similarity and residential location distances in all domains, it is feasible to build them up in some specific social communities by using data mining techniques. [24]. For example, in the email based social networks, through mining the subjects and contents of the emails in Enron Corporation[1], the social relationship between each pair of email sender and receiver (e.g., a CEO and his/her assistant) can be discovered and their roles can be known [27]. Then the corresponding social intimacy degree and community impact factor values can be estimated based on probabilistic models or by using the PageRank model. In addition, in academic social networks formed by large databases of Computer Science literature, the social relationships between scholars (e.g., co-authors, a supervisor and his/her students) can be mined from publications (e.g., from DBLP) and the role of a scholar (e.g., a professor in the field of data mining) can be mined from their homepages. Then the corresponding social intimacy degree and community impact factor can also be calculated. Furthermore, on Facebook, the preference similarity and the residential location distance between two participants can be mined from their profiles [29]. Detailed mining methods of these social impact factor values are out of the scope of this paper.

B. A Complex Contextual Social Network Structure

Based on the above social contextual impact factors, we propose a new structure for complex contextual social networks as depicted in Fig. 2. Based on the new complex contextual social network structure, in the next section, we introduce a novel social context-aware trust network discovery model, where both the impact of social contexts on social interactions and the constraints specification of the above social impact factors are considered.

IV. SOCIAL CONTEXT-AWARE TRUST NETWORK DISCOVERY MODEL

In this section, we first analyze the influence of social contextual impact factors on social interactions, and then propose a new concept Quality of Trust Network (QoTN) and a trust network utility calculation method, all of which are the key components of our social context-aware trust network discovery model.

A. Social Context-Aware Social Interaction

As indicated in Social Psychology [10, 15, 25, 37], some of the social contextual impact factors (i.e., $\varphi = \{SI, PS, T, RLD\}$) have influence on social interactions. For example, based on the statistics of 1000 publications from 18 countries in ISI Web of Knowledge (apps.webofknowledge.com) [35], the first author and the last author had the same address in 54% of the papers, indicating
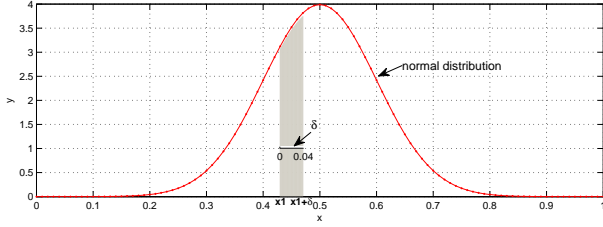
Figure 3.   Normal distribution

that RLD of authors impacts on the social interactions in research. In addition, based on the statistics on Flickr (flickr.com)-an online photo sharing social network [29], any two participants in photo sharing usually have similar preferences, indicating that PS impacts on the social interactions.

B. Social Context-Aware Social Interaction Probability

In the real world, the *normal distribution* has been widely used since the 18th century to model the relative frequency of physical and social phenomena [18]. E.g., the IQ, income and reading skills of people in a general population, the box-office performance of feature films, the output of journal articles by scientists, and the number of violent acts committed by male teenagers [1]. The probability density function of the normal distribution is as Eq.(1) (see the function image in Fig. 3).

$$y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where parameters μ and σ are the mean and standard deviation respectively, controlling the curve of the function image.

In our model, we assume the probability distribution of a social interaction between any two participants with the social contextual impact factors (denoted as $P(A \rightarrow B|X)$, $X \in \varphi$) also follows the normal distribution.

Then, based on mathematical theory of the integration [8], $P(A \rightarrow B|X)$ can be calculated by Eq. (2). In this equation, δ is the length of each small interval (the horizontal axis between 0 and 1 is divided into several small intervals [8]). If X is in one of the intervals (e.g., in the interval $[x_1, x_1 + \delta]$ in Fig. 3), $P(A \rightarrow B|X)$ is the integration of Eq. (1) with a lower limit X and a upper limit $X + \delta$ (in the case shown in Fig. 3, where $X = x_1$). Namely, $P(A \rightarrow B|X)$ is equal to the corresponding area of the trapezoid with curved edges in an interval $[X, X + \delta]$ (e.g., the shadowed area in Fig. 3). In addition, the parameters μ and σ in Eq. (2) can be computed by applying social statistics methods and mathematical theories in a social network [8, 9]. But this problem is out of the scope of this paper. Finally, based on probability theory [8], the aggregated social interaction probability between A and B (denoted as $AP(A \rightarrow B)$) can be calculated by Eq. (3).

$$P(A \rightarrow B|X) = \int_X^{X+\delta} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}} d(X) \quad (2)$$

$$AP(A \rightarrow B) = \prod_{X \in \varphi} P(A \rightarrow B|X) \quad (3)$$

In our model, the aggregated social interaction probability will be considered in the node selection of trust network discovery, where the larger the probability of a node to have a social interaction with the target, the more likely for the node to be selected.

C. Quality of Trust Network (QoTN)

In addition to the influence of social context on social interactions, our model also considers different trust evaluation criteria from a source for trust network discovery. We first propose a new concept, *Quality of Trust Network* as below.

Definition 1: *Quality of Trust Network* (QoTN) is the ability of a contextual trust network to guarantee a certain level of trust in trust evaluation, taking T, SI, CIF, PS, RLD as attributes.

In our model, a source participant can specify multiple constraints for QoTN attributes (i.e., T, SI, CIF, PS and RLD) for intermediate nodes and their links in a trust network, as the requirements of trust network discovery in different domains. Let $QoTN_{v_s,v_t}^{(\eta)}$ ($\eta \in \{T, SI, CIF, PS, RLD\}$) denote the QoTN constraint of η in the trust network from v_s to v_t (throughout this paper, v_s denotes the source and v_t denotes the target in a social network). For example, to get movie recommendation, v_s, a recommendee (i.e., a service consumer) specifies the QoTN constraints for the trust network from the service consumer to the service provider as $\{QoTN_{v_s,v_t}^T > 0.3, QoTN_{v_s,v_t}^{SI} > 0.3, QoTN_{v_s,v_t}^{PS} > 0.3, QoTN_{v_s,v_t}^{RLD} > 0.3, QoTN_{v_s,v_t}^{CIF} > 0.8\}$, if he/she believes the community impact factor of each of the intermediate participants is more important in the service of *movie recommendation*.

D. Trust Network Utility

In our model, we define the utility (denoted as $\mathcal{U}$) as the measurement of the trustworthiness of an extracted trust network. The utility function takes the QoTN attributes T, SI, CIF, PS and RLD as the arguments in Eq. (4)

$$\mathcal{U}(v_s, v_t) = \sum_{i=1}^{N} T_i + \sum_{i=1}^{N} SI_i + \sum_{i=1}^{M} CIF_i + \sum_{i=1}^{N} PS_i + \sum_{i=1}^{N} RLD_i \quad (4)$$

M is the number of the intermediate nodes and N is the number of the corresponding links in the trust network.

The higher the value of $\mathcal{U}(v_s, v_t)$, the higher the quality of the trust network from v_s to v_t. Therefore, the goal of trust network discovery is to extract the optimal trust network from source v_s to target v_t that satisfies multiple QoTN constraints and yields the highest utility.

V. SOCIAL CONTEXT-AWARE TRUST NETWORK DISCOVERY ALGORITHM

To solve the NP-Complete trust network discovery problem with QoTN constraints, we propose a Heuristic Social Context-Aware trust Network discovery (H-SCAN) algorithm, by adopting K-Best-First Search (KBFS) method and our proposed optimization strategies.

A. K-Best-First Search (KBFS)

K-Best-First Search (KBFS) algorithm [13] is based on Best-First Search method, which expands up to the best K nodes in the open list (i.e., the list of all candidates) in each node expansion cycle. KBFS is one of the heuristic algorithms with good efficiency for solving NP-complete problems, such as Number Partitioning problem and n-Puzzle problem [5, 13]. The K best nodes selection method can also be used to select nodes in the trust network discovery. However, KBFS is not designed for the specific trust network discovery problem directly. During the search, KBFS may investigate a node which has no connection with the target and may repeatedly investigate the feasibility of an expansion node's neighboring nodes. These drawbacks

can lead to low efficiency and low effectiveness. Therefore, according to the features of the trust network discovery in complex contextual social networks, we first propose a heuristic *selection probability function* for the K best nodes selection (see Eq. (5)). Secondly, to solve the problems of KBFS in trust network discovery, we propose two optimization strategies in the following subsections. Finally, based on these strategies, we propose a Heuristic Social Context-Aware trust Network discovery algorithm (H-SCAN), where the social contexts of participants are considered in the selection of the K best nodes.

B. H-SCAN

1) Algorithm Description: In H-SCAN, initially, the source participant v_s is regarded as the current expansion node, and H-SCAN searches all the neighboring nodes of v_s (denoted as $v_s.neighbors$) to investigate whether the current node and its corresponding links satisfy the QoTN constraints. If all QoTN constraints can be satisfied, the neighboring node is called *a feasible node* (denoted as v_f). Because the larger the outdegree of a node, the more likely for the node to have a connection with others [2], H-SCAN calculates the *selection probability* of each the feasible nodes v_f (denoted as $SCP(v_f \rightarrow v_t)$) based on $AP(v_f \rightarrow v_t)$ and the outdegree of v_f (let $deg^+(v_f)$ and $deg^-(v_f)$ denote the outdegree and indegree of v_f respectively) by Eq. (5).

$$SCP(v_f \rightarrow v_t) = AP(v_f \rightarrow v_t) \cdot \frac{deg^+(v_f)}{MAX(deg^+)} \quad (5)$$

where $MAX(deg^+)$ is the maximal outdegree of all nodes in a social network.

After that, H-SCAN selects up to K feasible neighboring nodes which have the K maximum selection probabilities, as the next expansion nodes (denoted as v_{exp}), Finally, H-SCAN repeats the above search process at each v_{exp} until it reaches the threshold of search hops (denoted as λ_h, on average $\lambda_h \leq 7$ due to the *small-world* phenomenon of social networks[2]). During the search process, we adopt the following three optimization strategies to improve the efficiency of our algorithm.

Optimization Strategy 1: Avoid Investigating the Intermediate Nodes with $deg^- > 0$ and $deg^+ = 0$. According to the *power-law* characteristic of social networks, most of the nodes in a social network have a small outdegree [29]. Therefore, there can be many nodes with $deg^- > 0$ and $deg^+ = 0$. During KBFS search, a node (denoted as v_x, $v_x \neq v_t$) with $deg^+(v_x) = 0$ and $deg^-(v_x) > 0$ (e.g., v_x in Fig. 4) may be investigated as an candidate of expansion nodes. As there is no social trust path linking v_x and v_t, the investigation of such a node can lead to low efficiency. To avoid this problem, our H-SCAN does not investigate the feasibility of v_x and does not select such a node as an expansion node in the subsequent search. This strategy improves the efficiency and effectiveness of the trust network discovery.
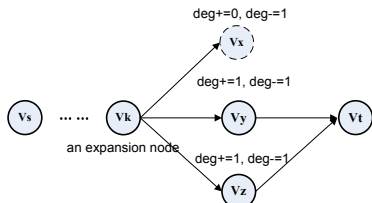

Figure 4.   A case of accessing the node with $deg^+ = 0$

Optimization Strategy 2: Avoid Repeatedly Accessing the

Algorithm 1: H-SCAN

Data: MT, v_s, v_t, $QoTN_{v_s,v_t}^{\eta}$, K, λ_h
Result: $TN(v_s, v_t)$

1 **begin**
2   Set $v_k.expansion = 0$, $v_k.preNumber = 0$, $hops = 1$, $ClosedSet = v_s$, $OpenSet = \emptyset$
3   **while** $hops \leq \lambda_h$ **do**
4   **for** *each* $v_i \in ClosedSet$ **do**
5        $v_i.expansion = 1$
6   **for** *each* $v_j \in adj[v_x]$ **do**
7   **if** $QoTN_{v_i,v_j}^{\eta}$ *can be satisfied* **and** $v_j.expansion = 1$ **then**
8            $v_j.preNumber = v_j.preNumber + 1$
9            $m = v_j.preNumber$
10            $v_j.preNode(m) = v_i$
11   **if** $QoTN_{v_i,v_j}^{\eta}$ *can be satisfied* **and** $v_j.expansion = 0$ **then**
12   **if** $v_j = v_t$ **then**
13              $v_j.preNumber = v_j.preNumber + 1$
14              $m = v_j.preNumber$
15              $v_j.preNode(m) = v_i$
16   **else if** $deg^+(v_j) > 0$ **and** $v_j \neq v_t$ **then**
17              $v_j.preNumber = v_j.preNumber + 1$
18              $m = v_j.preNumber$
19              $v_j.preNode(m) = v_i$
20   Put v_j into $OpenSet$
21      $ClosedSet = $ up to the K Maximal $SCP(v_j \rightarrow v_t)$, $v_j \in OpenSet$
22      $hops = hops + 1$
23   **if** $v_t.preNumber > 0$ **then**
24   Put v_t into $preset$
25   **while** $preSet \neq \emptyset$ **do**
26   **for** *each* $v_m \in preSet$ **do**
27   Delete v_m from $preSet$
28   **for** $i = 1$ *to* $v_m.preNumber$ **do**
29   Add v_m, $v_m.preNode(i)$ and $v_m.preNode(i) \rightarrow v_m$ into $TN(v_s, v_t)$
30   **if** $v_m.preNode(i) \neq v_s$ **then**
31   Put $v_m.preNode(i)$ into $preSet$

end

Neighboring Nodes of An Expansion Node. During KBFS search, an expansion node v_{exp} may be selected more than once in different search hops (e.g., $v_{exp} = v_c$ in Fig. 5). In such a situation, the feasibility of v_{exp}'s neighboring nodes will be repeatedly investigated (e.g., v_d in Fig. 5), leading to low efficiency. To address this issue, upon reaching the same v_{exp} (e.g., v_b) in different search hops, H-SCAN does not investigate its neighboring nodes repeatedly as all of them have been visited in previous search, thus saving execution time.
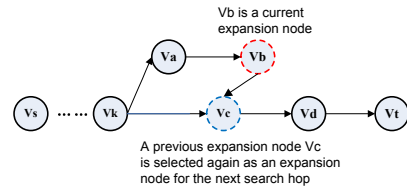

Figure 5.   Repeated selecting an expansion node

2) The Process of H-SCAN: Given a group of QoTN constraints, and a pair of v_s and v_t in a complex contextual social network, the process of H-SCAN includes the following steps. The pseudo-code of H-SCAN is given in Algorithm 1

Initialization: At each node v_k, set $v_k.expansion = 0$, which indicates v_k's neighboring nodes has not been selected as an expansion node. In addition, set two sets to record up to K current expansion nodes (i.e., $ClosedSet$ [31]) and all the candidates for the search of the next hop (i.e., $OpenSet$ [31]), respectively. Furthermore, set the maximal search hop as λ_h,
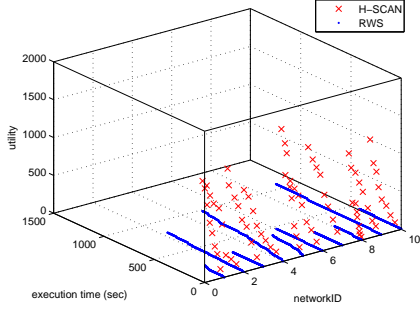
Figure 6. The comparison of utilities extracted with QoTNID=1

Figure 8. The comparison of utilities extracted with QoTNID=3

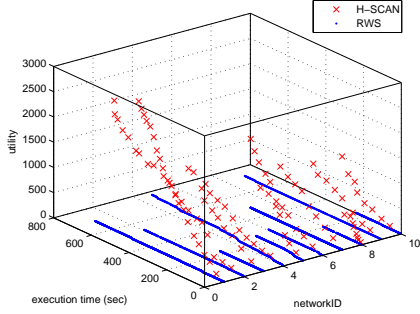Figure 7. The comparison of utilities extracted with QoTNID=2

Figure 9. The comparison of utilities extracted with QoTNID=4

put v_s into $ClosedSet$ and set the number of current search hop as one (denoted as $hops = 1$) *(lines 1-2 in Algorithm 1)*.

Step 1: If there exists any $v_i.expansion = 0$, ($v_i \in ClosedSet$) and $hops \leq \lambda_h$, get v_i from $ClosedSet$, and mark $v_i.expansion = 1$. Otherwise, go to *Step 3 (lines 3-5 in Algorithm 1)*.

Step 2: Investigates v_j, ($v_j \in v_i.neighbors$). If v_j is a feasible node, based on the value of $v_j.expansion$, H-SCAN performs the following search strategies.

> **Situation 1:** If $v_j.expansion = 1$ and the corresponding QoTN constraints can be satisfied, then the number of the preceding nodes of v_j increases by 1 (denoted as $v_j.preNumber = v_j.preNumber + 1$). In addition, let $m = v_j.preNumber$. At v_j, store the mth preceding node of v_j (denoted as $v_j.preNode(m) = v_i$) *(lines 6-10 in Algorithm 1)*.
>
> **Situation 2:** If $v_j.expansion = 0$ and the corresponding QoTN constraints can be satisfied, based on the status of v_j, H-SCAN performs the following search strategies.
>
> > **Case 1:** If $v_j = v_t$, then the number of the preceding nodes of v_j increases by 1 (i.e., $v_j.preNumber = v_j.preNumber + 1$). In addition, let $m = v_j.preNumber$, at v_j, store the mth preceding node of v_j (denoted as $v_j.preNode(m) = v_i$) *(lines 11-15 in Algorithm 1)* .
> >
> > **Case 2:** If $v_j \neq v_t$ and $deg^+(v_j) > 0$, then the number of the preceding nodes of v_j is increased 1 (i.e., $v_j.preNumber = v_j.preNumber + 1$). In addition, let $m = v_j.preNumber$, at v_j, store the mth preceding node of v_j (i.e., $v_j.preNode(m) = v_i$). Finally, put v_j into $OpenSet$ *(lines 16-20 in Algorithm 1)*.

Step 3: Set $hops = hops + 1$. If $hops <= \lambda_h$ and $OpenSet \neq \emptyset$, select up to K candidate (denoted as $v_{cand}(K)$) which have the K maximal $SCP(v_{cand}(K) \rightarrow v_t)$ from $OpenSet$ and put them into $ClosedSet$ *(lines 21-22 in Al-*
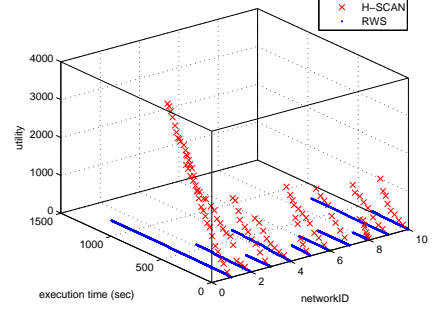
gorithm 1).

Step 4: If $v_t.preNumber > 0$, construct the trust network from v_s to v_t (denoted as $TN(v_s, v_t)$) by searching the preceding nodes from v_t layer by layer until reaching v_s. Otherwise, it fails to return a trust network that satisfies QoTN constraints *(lines 23-31 in Algorithm 1)*.

H-SCAN contains two parts, i.e., network search (*Step 1* to *Step 3*) and trust network construction (*Step 4*). At each search hop, H-SCAN selects up to K expansion nodes; therefore, the time complexity of H-SCAN in network search (i.e., the first part) is $O(K * m * \lambda_h)$, where K is the number expansion nodes selected at each search hop; m is the maximal outdegree of the nodes in a social network, and λ_h is the maximal search hops. In addition, in the worst case, there are $K * (\lambda - 1)$ intermediate nodes and each intermediate node has K preceding nodes. Then the time complexity of H-SCAN in trust network construction (i.e., the second part) is $O(K^2 * \lambda_h)$. In social networks, on average, $\lambda_h < 7$ [28], and $K \leq m$. Thus the time complexity of H-SCAN is $O(Km)$, which is better than TTL-BFS (Time To Live based Breadth First Search) with the time complexity of $O(m^{TTL})$, and the same as both RWS (Random Walk Search) and HDS (High Degree Search). Since H-SCAN considers the social contextual impact factors and adopts our proposed optimization strategies, it can both deliver higher utility and consume less execution time than TTL-BFS, HDS and RWS.

VI. EXPERIMENTAL EVALUATION

A. Experiment Setup

Firstly, in order to evaluate the performance of H-SCAN on trust network discovery, we need a dataset which contains social network structures. The *Enron* email dataset[1] has been proved to possess the small-world and power-law characteristics of social networks, it has been widely used in the studies of social networks [23, 24, 27, 32]. Thus, we select the *Enron* email dataset[1] with 87,474 nodes (participants) and 30,0511 links (formed by sending and receiving emails) as the dataset for our
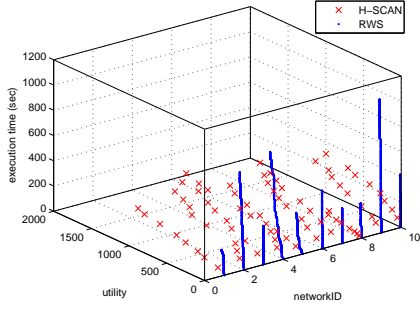
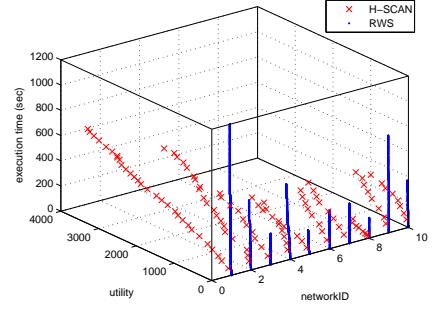Figure 10.   The comparison of execution time with QoTNID=1

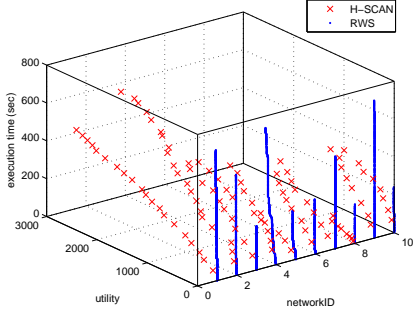Figure 12.   The comparison of execution time with QoTNID=3

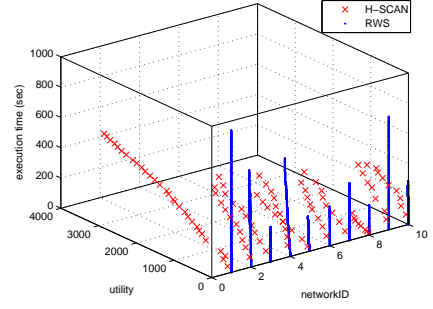Figure 11.   The comparison of execution time with QoTNID=2

Figure 13.   The comparison of execution time with QoTNID=4

experiments. Secondly, we randomly select 10 pairs of source and target nodes from the dataset, and compare our H-SCAN with other methods in all three categories, i.e., TTL-BFS, RWS and HDS (see *Section 2.3*). Thirdly, we set four groups of QoTN constraints as listed in Table I and set the social interaction probability to approximately follow the normal distribution with $\mu = 0.5$ and $\sigma = 0.1$. Fourthly, based on the small-world characteristic of social networks, we set the maximal search hops of all the algorithms as 6. Finally, since the detailed mining method of social contextual impact factor values is out of the scope of this paper, their values are generated by using the function $normrnd(\mu, \sigma)$ in *Matlab*, which generates random numbers in the range of $[0, 1]$, following the normal distribution with μ and σ.

Table I
THE SETTINGS OF QOTN CONSTRAINTS

ID	$QoTN(T)$	$QoTN(SI)$	$QoTN(PS)$	$QoTN(RLD)$	$QoTN(CIF)$
1	0.1	0.1	0.1	0.1	0.1
2	0.15	0.15	0.1	0.1	0.1
3	0.1	0.1	0.15	0.15	0.1
4	0.1	0.1	0.1	0.15	0.15

Each of H-SCAN, RWS and HDS is implemented using Matlab R2008a running on an Lenovo ThinkPad SL500 laptop with an Intel Core 2 Duo T5870 2.00GHz CPU, 3GB RAM, Windows XP SP3 operating system and MySql 5.1.35 relational database. The results are plotted in Fig. 7 to Fig. 13, where the execution time and the utilities of the extracted trust network for each of the algorithms are averaged based on 5 independent runs.

B. Results and Analysis

During the execution of TTL-BFS, in all cases, when $TTL > 2$, TTL-BFS consumes much execution time (i.e., more than 5 hours), making it inapplicable to large-scale social networks. This is because the exponential time complexity of TTL-BFS is $O(m^{TTL})$. In addition, in all cases, even when HDS has more execution time than H-SCAN and RWS, the utilities

of the solutions delivered by HDS are always equal to *zero*. This is because HDS searches nodes based the descending order of their outdegree only without considering the links between these nodes and the target, which could lead to low effectiveness, and thus it cannot be applied for trust network discovery in complex contextual social networks. Therefore, we only compare the utilities of the trust networks extracted by H-SCAN and RWS, and their execution time.

The Comparison of Utilities: Fig. 7 to Fig. 9 plot the extracted trust networks' utilities with different QoTN constraints. From the figures, we could see that our H-SCAN can always deliver much better utilities than RWS under the same execution time. Table II compares the average utilities delivered by H-SCAN and RWS in every second of the execution. Based on the statistics, on average, with the same execution time, H-SCAN delivers utilities that are 48.1 times better than those of RWS. This is because H-SCAN takes into account the influence of social contexts on social interactions, where the larger the probability that a node has a social interaction with v_t, the more likely for the node to be selected. This method increases the probability of finding a good quality trust path from v_s to v_t at each search. In addition, H-SCAN considers the QoTN constrains, and can avoid searching infeasible nodes, improving the effectiveness of search. In summary, H-SCAN greatly outperforms RWS in the quality of the extracted trust networks.

Table II
COMPARISON OF UTILITY

Algorithm	The average ratio of utility and execution time										
	1	2	3	4	5	6	7	8	9	10	total
RWS	0.09	0.08	0.08	0.11	0.12	0.11	0.09	0.09	0.08	0.08	0.93
H-SCAN	5.2	4.5	4.4	4.2	4.5	4.4	4.4	4.4	4.4	4.3	44.7
difference	57.7 times	56.2 times	55 times	38.1 times	37.5 times	40 times	48.8 times	48.8 times	55 times	53.7 times	48.1 times

The Comparison of Execution Time: Fig. 10 to Fig. 12 plot the execution time of H-SCAN and RWS with different QoTN constraints. From the results, we observe that when delivering the same utility, the execution time of H-SCAN is less than that

of RWS in all trust networks. Table III compares the average execution time of H-SCAN and RWS. Based on the statistics, on average, the execution time of H-SCAN to deliver one utility is only 2.02% of that of RWS. This is because H-SCAN adopts KBFS with our proposed *Optimization Strategies 1 and 2*, and thus can avoid (1) repeatedly searching node which has no connection with the target (*by Strategy 1*) and (2) repeated feasibility investigation (*by Strategy 2*), respectively.

Table III
COMPARISON OF EXECUTION TIME

Algorithm	The average ratio of execution time and utility										
	1	2	3	4	5	6	7	8	9	10	total
RWS	19.9	20.8	22.1	17.5	16.3	16.7	20.3	20.4	21.3	21.4	196.7
H-SCAN	0.34	0.39	0.41	0.42	0.39	0.41	0.41	0.39	0.41	0.41	3.98
difference	1.71%	1.87%	1.86%	2.41%	2.39%	2.46%	2.02%	1.91%	1.92%	1.92%	2.02%

Summary: Based on the above experimental results and analysis, we conclude that H-SCAN outperforms the existing methods significantly in both *execution time* and *the quality of the extracted trust networks*. Therefore, H-SCAN is an efficient and effective algorithm for trust network discovery with QoTN constrains in complex contextual social networks.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a complex contextual social network structure containing social contextual impact factors. They all have significant influence on trust evaluation and social interactions, reflecting the social networks in real life better. Then, we have proposed a general concept QoTN (Quality of Trust Network), and a novel social context-aware trust network discovery model. Finally, we have proposed a new Heuristic Social Context-Aware trust Network discovery algorithm by adopting K-Best-First Search (KBFS) method and our proposed optimization strategies. The experimental results demonstrate the superior performance of our proposed H-SCAN algorithm.

In the future work, we plan to incorporate our models and algorithms in a new generation of contextual social network based service provider recommendation systems.

REFERENCES

[1] M. Abell, J. Braselton, and J. Rafter. *Statistics with Maple*. Elsevier Science and Technology, 2012.

[2] L. A. Adamic, R. M. Lukose, and B. A. Huberman. Local search in unstructured networks. In *Handbook of Graphs and Networks*. Wiley, 2003.

[3] P. S. Adler. Market, hierarchy, and trust: The knowledge economy and the future of capitalism. *Organization Science*, 12(12):215–234, 2001.

[4] R. Ashri, S. Ramchurn, J. Sabater, M. Luck, and N. Jennings. Trust evaluation through relationship analysis. In *AAMAS*, pages 1005–1011, 2005.

[5] S. Baase and A. Gelder. *Computer Algorithms Introduction to Design and Analysis*. Addision Wesley, 2000.

[6] E. Barnett and M. Casper. A definition of social environment. *American Journal of Public Health*, 91(3), 2001.

[7] Y. Baryshnikov, E. Coffman, P. Jelenkovic, P. Momcilovic, and D. Rubenstein. Flood search under the california split rule. *Operations Research Letters*, 32, 2003.

[8] M. Bittinger. *Basic Mathematics*. Addison Wesley, 2002.

[9] H. M. Blalock. *Social Statistics*. New York: McGraw-Hill, 1979.

[10] D. J. Brass. *A Socal Network Prespective On Industral/organizational psychology*. Industrial/Organizational Handbook, 2009.

[11] N. Chang and M. Liu. Revisiting the ttl-based controlled flooding search: Optimality and randomization. In *MobiCom*, pages 85–99, 2004.

[12] M. Dalton. *Men Who Manage*. New York: Wiley, 1959.

[13] A. Felner, S. Kraus, and R. E. Korf. Kbfs: K-best-first search. *Annals of Mathematics and Artificial Intelligence*, 39:19–39, 2003.

[14] I. Filali and F. Huet. Dynamic ttl-based search in unstructured peer-to-peer networks. In *CCGrid*, pages 438–447, 2010.

[15] J. Gimpel, K. Karnes, J. Mctague, and S. Pearson-Merkowitz. Distance-decay in the political geography of friends-and-neighbors voting. *Political Geography*, 27:231–252, 2008.

[16] C. Gkantsidis, M. Mihail, and A. Saberi. Random walks in peer-to-peer networks. In *IEEE INFORCOM*, pages 120–130, 2004.

[17] J. Golbeck and J. Hendler. Inferring trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, 6(4):497–529, 2006.

[18] M. Jorge and X. Yi. *Return to RiskMetrics: The Evolution of a Standard*. RiskMetrics, 2001.

[19] R. F. Korte. Biases in decision making and implications for human resource development. *Advances in Developing Human Resources*, 5(4):440–457, 2003.

[20] U. Kuter and J. Golbeck. Sunny: A new algorithm for trust inference in social networks using probabilistic confidence model. In *AAAI'07*, pages 1377–1382, 2007.

[21] S. Lichtenstein and P. Slovic. *The construction of preference*. Cambridge University Press, 2006.

[22] G. Liu, Y. Wang, M. Orgun, and E.-P. Lim. A heuristic algorithm for trust-oriented service provider selection in complex social networks. In *SCC*, pages 130–137, 2010.

[23] G. Liu, Y. Wang, and M. A. Orgun. Finding k optimal social trust paths for the selection of trustworthy service providers in complex social networks. In *ICWS'11*, pages 41–48, 2011.

[24] G. Liu, Y. Wang, and M. A. Orgun. Trust transitivity in complex social networks. In *AAAI'11*, pages 1222–1229, 2011.

[25] N. Luhmann. *Trust and Power*. Wiley, 1979.

[26] R. Mansell and B. Collins. *Trust and crime in information societies*. Edward Elgar Publishing, 2005.

[27] A. Mccallum, X. Wang, and A. Corrada-Emmanuel. Topic and role discovery in social networks with experiments on Enron and academic email. *Journal of Artificial Intelligence Research*, 30(1):249–272, 2007.

[28] S. Milgram. The small world problem. *Psychology Today*, 2(60), 1967.

[29] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *ACM IMC'07*, pages 29–42, 2007.

[30] C. L. Prell. Community networking and social capital: Early investigation. *Journal of Computer Mediated Communication*, 8(3), 2003.

[31] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.

[32] F. L. S. Yoo, Y. Yang and I. Moon. Mining social networks for personalized email prioritization. In *KDD'09*, pages 967–976, 2009.

[33] Y. Wang and V. Varadharajan. Role-based recommendation and trust evaluation. In *IEEE EEE'07*, pages 278–295, 2007.

[34] D. J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton and Company, 2003.

[35] J. Wren, K. Kozak, K. Johnson, S. Deakyne, L. Schilling, and R. Dellavalle. A survey of perceived contributions to papers based on byline position and number of authors. *EMBO Report*, 8(11):988–991, 2007.

[36] R. Zajonc. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 62(3):713–715, 1961.

[37] R. Zajonc. Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, 10(6):224–228, 2011.

[38] H. Zhang, L. Zhang, X. Shan, and V. Li. Probabilistic search in p2p networks with high node degree variation. In *IEEE International Conference on Communications*, pages 1710–1715, 2007.

[39] M. Zhong, K. Shen, and J. Seiferas. The convergence-guaranteed random walk and its applications in peer-to-peer network. *IEEE Transactions on Computers*, 57(5):619–633, 2008.