Computing and Informatics, Vol. 36, 2017, 517–540, doi: 10.4149/cai_2017_3_517

CAT: CONTEXT-AWARE TRUST-ORIENTED WORKER SELECTION IN SOCIAL CROWD

Yang Zhao, Kai Zheng, An Liu, Zhixu Li

School of Computer Science and Technology
Soochow University, China
&
Collaborative Innovation Center of Novel Software
Technology and Industrialization, Jiangsu, China
e-mail: 15237326232@163.com, {zhengkai, anliu, zhixuli}@suda.edu.cn

Guanfeng Liu

School of Computer Science and Technology
Soochow University, China
&
Collaborative Innovation Center of Novel Software
Technology and Industrialization, Jiangsu, China
&
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China

Xiaofang Zhou

School of Computer Science and Technology
Soochow University, China
&
School of Information Technology and Electrical Engineering
The University of Queensland, Brisbane, Australia
e-mail: zxf@suda.edu.cn

Abstract. Finding trustworthy worker is a longstanding issue in crowdsourcing systems. On traditional crowdsouring platforms, like *Amazon Mechanical Turk*[1], the trustworthiness of a worker is usually based on the contextual information, like different *types of tasks* and different *reward amounts* of tasks. However, with the combination of *OSNs* and the crowdsourcing applications in social crowd, in addition to the above mentioned *task based contexts*, the *social contexts* like the *social relationships* and the *social positions* of participants can greatly assist requestors to select trustworthy workers. In order to select the trustworthy workers in social crowd, in this paper, we first present a contextual social network structure which contains complex *social contexts*. Then we propose a trust evaluation model taking both *contexts information* and the requirements of requestors into consideration, which leads to the trust worker selection in social crowd as a classical NP-Complete Multi-Constrained Optimal Path (*MCOP*) selection problem. For solving this challenging problem, we propose a new efficient and effective approximation Context-Aware Trust-Oriented Worker Selection algorithm *CAT*. The results of our experiments conducted on four real *OSN* datasets illustrate the superiority of our method in trustworthy worker selection.

Keywords: Crowdsourcing, social contexts, task based contexts, trustworthy worker

1 INTRODUCTION

1.1 Background

Crowdsourcing [1] has resolved many challenging problems from both industry and academia which are too difficult for computers or too expensive to employ experts such as image tagging [2], entity resolution [3, 9], schema matching [4] and so on. Some well-known crowdsourcing platforms, like *Amazon Mechanical Turk*[1] and *Freelancer*[2], have taken full use of the wisdom of crowd. According to the statistics provided by *Freelancer*[2] in 2015, there are more than 1.5 million workers performing tasks with different types and different reward amounts. On traditional crowdsourcing platforms, workers are anonymous to requestors, and there are little interactions between them, which makes large numbers of malicious workers exist when performing tasks. With the combination of the online social networks and the crowdsourcing websites, there are some social crowd, like *Quora*[3], where the *social contexts* like the *social relationships* and the *social trust* between participants and the *social positions* of participants can assist requestors to select trustworthy workers, as these *social contexts* have significant influence on the trust evaluation [17, 18].

[1] `http://mturk.com`
[2] `https://www.freelancer.com.au`
[3] `https://www.quora.com`

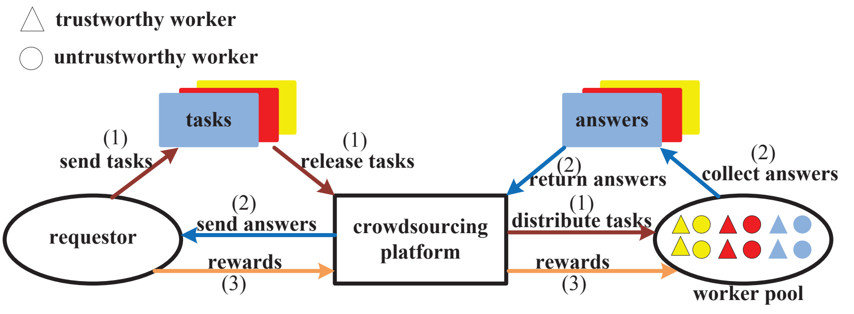

Figure 1. A typical workflow of a crowdsourcing platform

Example 1. Figure 1 depicts the typical operation process of traditional crowdsourcing systems, mainly including the following steps:

1. task owners distribute the tasks on the crowdsourcing platforms and wait for crowd workers to perform;
2. then they collect and aggregate the answers from the workers and estimate the right truth;
3. finally, the right workers receive rewards offered by the requestors via the crowdsourcing platforms.

From Figure 1, we can observe that different kinds of tasks are published on crowdsourcing platforms, which can be different types and different reward amounts. Workers' trustworthiness varies in different context. For each task, both *trustworthy workers* and *untrustworthy workers* may participate to perform it.

1.2 Problem and Challenges

With the sprawl of crowdsourcing systems, workers' trustworthiness has become a prominent problem, since large numbers of dishonest workers may participate in tasks. For example, some dishonest workers aim to receive the maximal rewards by quickly giving plausible answers, and some others aim to boost their trust levels by performing easy tasks [24]. Because of the existing of untrustworthy answers, the requestors usually have to ask more crowd workers to answer the same questions to improve the reliability of the answers, which greatly increases the economic and time cost for the requestors. So selecting trustworthy workers becomes remarkably significant in crowdsourcing systems.

Crowdsourcing platforms, like *Amazon Mechanical Turk*[1] and *Freelancer*[2], adopt the historical records of tasks to evaluate workers' trustworthiness. *Amazon Turk* adopts the overall approval rate to identify the trustworthy workers. However, dishonest workers can easily get high overall approval rates by quickly giving plausible answers or participating easy tasks. In the literature, some researches carry out trust control mechanisms, e.g., Li et al. [10] proposed a general crowd targeting framework that can discover a group of trustworthy workers based on their characteristics. This method must additionally collect the workers' characteristics during the given tasks performed, which makes an additional cost, and sometimes it is impossible to get the complete characteristic information of the crowd workers. In addition, Ye et al. [24] proposed a context-aware trust model for worker selection on traditional crowdsourcing platforms, which considers the contexts like *task types* and *task reward amounts*. Though these methods have considered trust evaluation when selecting trustworthy workers, they neglect the *social contexts*, like *social relationships*, *social trust* and *social positions* in worker selection.

However, with the combination of *OSNs* and crowdsourcing platforms, in addition to the *task based contexts*, like different *types of tasks* and different *reward amounts of tasks* [24], taking the *social contexts* like social relationships and social positions into account is a crucial factor to evaluate the workers' trustworthiness, which has been indicated in the *social science* theories [17, 18]. The following motivation example illustrates the significance of *social contexts* in trustworthy worker selection.

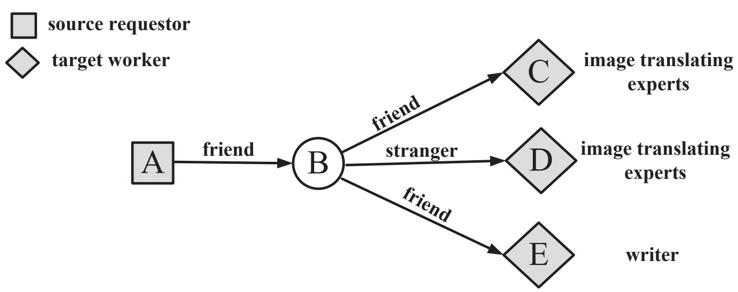

Figure 2. A motivation example

Example 2. In Figure 2, A as the source requestor needs to employ a worker to perform a task of image translation with a certain reward amount. Worker C and worker D are image translating experts, and worker E is a writer. Besides, C and E are friends of B, which means that B is completely familiar with them. While D is a stranger to B, which infers that B has little knowledge about D. Considering the *task based contexts*, E can be identified as an untrustworthy worker, but it cannot differentiate worker C and worker D in such context. From Figure 2, we can see that the source requestor A and the target workers C, D are indirectly linked through B (friend of A) by the social path, i.e., $p_{A \to B \to C}$ and $p_{A \to B \to D}$. The source requestor can evaluate the trustworthiness of the target workers based on the social contexts found in the social path, and the path with trust information linking the source requestor and the target worker is called a *social trust path* [31]. The path $p_{A \to B \to C}$ is more trust than $p_{A \to B \to D}$ since C is more intimate than D to B, therefore, C is more trustworthy than D considering the *social contexts* found in the *social trust path*. So in addition to the *task based contexts*, the *social contexts* can also greatly help requestors select trustworthy workers.

In this paper, in order to find the trustworthy workers, we consider both the *social contexts*, like the *social trust*, the *social relationships* and the *social positions*, and the *task based contexts*, like *task types* and *task reward amounts* in the trustworthiness evaluation, which can greatly help find trustworthy workers for the requestors. Besides, by setting different constraints of contexts values, the requesters can specify different requirements for the workers. This makes the trustworthy worker selection challenging as it becomes a Multi-Constrained Optimal Path (*MCOP*) selection problem, which is a NP-Complete problem [22].

1.3 Contributions

In this paper, we aim to solve the trustworthy worker selection problem in social crowd. Our contributions in this paper are summarised as follows.

1. We first present a contextual social network structure which contains *social contexts* like the *social relationships* and the *social trust* between participants, and the *social positions* of participants.
2. Based on the contextual social network structure, we then propose a context-aware trust evaluation model, which considers both *task based contexts* and *social contexts*. In order to solve the NP-Complete challenging context-aware trustworthy worker selection problem, we propose a new effective and efficient approximation algorithm *CAT*, which can find the trustworthy works based on the requirements of requester. The time complexity of *CAT* achieves in $O(mu)$, where m is the number of simulations; u is the maximal outdegree of nodes.
3. We have conducted experiments on four real datasets of *OSNs* to investigate the performance of our proposed trust evaluation model. The experimental results illustrate that our *CAT* can more effectively select the trustworthy workers than *task based contexts* based worker selection method [24]. Based on the statistics, *CAT* can improve the quality of the trustworthy worker selection by 22.6 % than *CrowdTrust* on average.

The rest of this paper is organized as follows. We first review the related work on worker selection problem in crowdsourcing systems in Section 2. Then we introduce the contextual social network in Section 3. Section 4 presents our proposed trustworthy worker evaluation model, Section 5 presents our contextual-aware worker selection algorithm, Section 6 reports the experimental observation, and Section 7 concludes the paper.

2 RELATED WORK

Crowdsourcing systems are being widely used today in both academia and industry, where the trustworthy worker selection is a longstanding issue. In this section, we introduce the existing studies on crowdsourcing systems, trustworthy worker evaluation, and trustworthy worker selection.

2.1 Crowdsourcing Systems

Crowdsourcing is firstly coined by Jeff Howe for Wired magazine in 2006 [1], which explained the process of distributing works on the internet. Nowadays, crowdsourcing has become a useful tool to address microtasks that are too hard for computers or too expensive to employ experts [14]. Various kinds of crowdsourcing services emergence, such as *Amazon Mechanical Turk*[1] and *Wikipedia* [4]. Amazon Turk is a well-known crowdsourcing platform, which engages large-scale workforce to tackle ten thousands of the HITs, such as the article writing, desicion making and the data entry. And Wikipedia is a remarkable encyclopedia which is continually improved by participants from all over the world. In addition, the crowdsourcing techniques have also been leveraged in the theoretical study area. For example, [3, 9] studied the crowdsourcing techniques on entity resolution, Schema Matching in the paper [4], filtering [15], Tagging [2] and so on. Moreover, some novel kinds of crowdsourced databases are developed, such as CrowdDB [5], CDAS [6], Qurk [7] and Deco [8], compared with traditional database systems, they do not hold the traditional closed-world assumption for human input [5].

2.2 Trustworthy Worker Evaluation

Besides, in the literature, some qualitative models for trust evaluation of crowdsourcing workers have been proposed. Li et al. [10] modeled the worker's trustworthiness as a constant parameter which is a symbol of the correctness of a worker answers the questions, while some other works [16] modeled it as a confusion matrix that aimed to capture the relations between labels in questions and the trustworthiness of workers. In addtion, Ye et al. [24] proposed a context-aware trustworthy worker selection model for traditional crowdsourcing platforms, considering the contexts like *task types* and *reward amounts*.

However, the widespread use of the online social networks makes it possible to make full use of wisdom of crowd. A good example of this idea is success of the website *Quora*[3], which is rapidly growing with social links between users when other sites like Yahoo answers have stalled and begun to shrink. Besides, in the literature, some existing works [11, 12, 13] have taken efforts to capture the crowdsourcing workers from the social media users. Cao et al. [11] aimed to find the optimal subset of workers under a limited budget from the micro-blog followers. Besides,

[4] `https://www.wikipedia.org`

Cao et al. [12] presented Wise Market which can detect the careless answers from sloppy workers and target the high quality answers. Moreover, Forlines et al. [13] focused on the problem of correctly aggregating the individual answers from the social network members for the decision making tasks. However, they all did not consider that workers' trustworthiness could vary in different contexts, including both of the *social contexts*, i.e., the *social relationships* and the *social trust* between participants and the *social positions* of participants, and the *task based contexts*, i.e., *task types* and *reward amounts*. In our work, we aim to model the relationship between contexts and worker's trustworthiness so that we can select workers with high trust levels.

Some previous works proposed by us are for the studies of the trust evaluation in Online Social Networks (OSNs). In [27], we proposed a trust network discovery algorithm to identify the trust network that contains the most important participants between two people in OSNs, and in [28], we further improve the efficiency and effectiveness of our proposed algorithm. In addition, with the extracted trust network, in [29] we proposed a trust-oriented service provider selection method where we considered the influence of social contexts in trust evaluations. Furthermore, in order to improve the efficiency and the effectiveness of the method proposed in [29], in [26] and [30], we proposed social context-aware trust evaluation methods by adopting the Bayes inference and the matrix factorization respectively. These trust models have good performance in trust evaluation in OSNs, but they cannot be directly applied into the social crowd environments, where more contexts like the trust-based contexts are included.

3 CONTEXTUAL SOCIAL NETWORK STRUCTURE

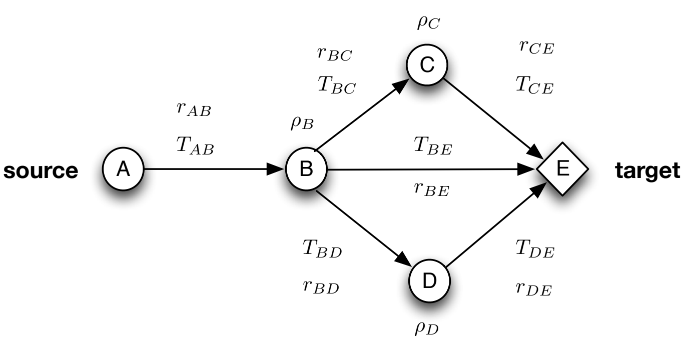

Figure 3. A contextual social network

A contextual social network structure can be modelled as a directed graph $G = (V, E, LV, LE)$, where

- V is a set of vertices,
- E is a set of edges,
- LV is a set of labels for V, for each $v \in V$, $LV(v)$ is the impact of the social position of a worker in a specific area,
- LE is a set of labels for E, For each $e \in E$, $LE(e)$ is the social trust and the social relationships between workers.

Example 3. Figure 3 depicts a *Contextual Social Network*, where each vertex $v_i \in V$ is associated with a *role impact factor*, denoted as $\rho_{v_i} \in [0,1]$, to illustrate the impact of social user v_i, which is determined by the expertise of v_i. $\rho_{v_i} = 1$ indicates that v_i is an expert while $\rho_{v_i} = 0$ indicates that v_i has no knowledge. Moreover, each edge (v_i, v_j) is associated with *social trust*, denoted as $T_{v_i,v_j} \in [0,1]$, and *social intimacy degree*, denoted as $r_{v_i,v_j} \in [0,1]$, to illustrate trust and intimacy social relationships between social users. T, r and ρ are called *social impact factors*, whose values can be extracted by using the data mining techniques [32]. But mining these social impact factor values is another challenging problem, which is out of the scope of this paper.

4 CONTEXT-AWARE TRUSTWORTHY WORKER SELECTION MODEL

In this section, we propose a trust selection model, which considers both the *social contexts* and *task based contexts* in trustworthy worker selection.

4.1 Social Context Based Trust

4.1.1 Quality of Trust (QoT)

Definition 1. *Quality of Trust (QoT)* is the ability to guarantee a certain level of trustworthiness in trust propagation along a social trust path, taking the social trust (T), the social intimacy degree (r), and the role impact factor (ρ) as attributes.

In our model, a requestor can set multiple constraints for QoT attributes (i.e., T, r and ρ) as the requirements of trust evaluation of the social workers, denoted as λ, for example, in Figure 2, the requestor B can set the QoT constraints for the descendant workers as $\lambda = \{\lambda_T > 0.5, \lambda_r > 0.5, \lambda_\rho > 0.5\}$, where λ_T, λ_r, λ_ρ are the constraints of T, r, ρ, respectively.

4.1.2 Social Impact Factor Aggregation

Based on the theories in *Social Psychology* [20], we adopt the multiplication method to aggregate T and r values of a social path from requestor to social worker, and adopt the average method to aggregate the ρ values of the vertices in the path. The details of the aggregation method has been discussed in [23]. The aggregated values of a social path p are denoted as T_p, r_p, ρ_p, respectively.

4.1.3 Social Context Based Utility Function

In the *social context* based trust evaluation model, we define a feasible utility (denoted as U^S) as the measurement of the social trust of the social crowd workers.

$$U^S_{p(a_1,\ldots,a_m)} = \omega_T \times T_{p(a_1,\ldots,a_m)} + \omega_r \times r_{p(a_1,\ldots,a_m)} + \omega_\rho \times \rho_{p(a_1,\ldots,a_m)} \tag{1}$$

where $T_{p(a_1,\ldots,a_m)}$, $r_{p(a_1,\ldots,a_m)}$, $\rho_{p(a_1,\ldots,a_m)}$ are the aggregated social impact factors and ω_T, ω_r and ω_ρ are the weights of $T_{p(a_1,\ldots,a_m)}$, $r_{p(a_1,\ldots,a_m)}$, $\rho_{p(a_1,\ldots,a_m)}$, respectively; $0 < \omega_T, \omega_r, \omega_\rho < 1$ and $\omega_T + \omega_r + \omega_\rho = 1$.

4.2 Task Based Contexts Based Trust

A worker may have different trust levels in different *task based contexts*, including the *task types* and the *task reward amounts*. Based on the two kinds of *task based contexts*, a worker's *TaTrust*, i.e. *task type* based trust, and *RaTrust*, i.e. *task reward amount* based trust, can be calculated by adopting the same methods proposed in [24].

Classification Based on Task Types: A type of HITs (Human Intelligence Tasks) can be decomposed into three dimensions: input, processing and output. Figure 3 is the three-dimensional intelligence space. In the dimension of HIT Input, there are 5 types of tasks in crowdsourcing platforms: figural, symbolic, semantic, audio and video. According to the structure of Guilford's SI model [19], the dimension of HIT Processing also includes 5 types: cognition, memory, divergent production, convergent production and evaluation. While the dimension of HIT Output includes 6 types: units, classes, relations, systems, transformations and implications. Based on the three dimensions, the intelligence space consists of 150 $(5 * 5 * 6)$ cubes for classifying the HITs.

Classification Based on Reward Amounts: In crowdsourcing, a worker who performs well in a range of reward amounts is likely to be trustworthy in the tasks belonging to the same range. i.e., the reward amount of an upcoming HIT is r', then those tasks rewarded between $\alpha r'$ and $\beta r'$ are classified into one type, where α and β are constants. We use the ratio p to classify HITs, which is calculated in Equation (2).

$$p = \begin{cases} 1, & \text{if } 0 < \frac{\max(r', r_i)}{\min(r', r_i)} < 1, \\ 2, & \text{if } 1 < \frac{\max(r', r_i)}{\min(r', r_i)} < 10, \\ 3, & \text{if } 10 < \frac{\max(r', r_i)}{\min(r', r_i)} < 10^2, \\ \ldots, & \ldots \\ h, & \text{if } 10^{(h-1)} < \frac{\max(r', r_i)}{\min(r', r_i)} < 10^h \end{cases} \tag{2}$$

where r_i is the reward amount of a historical HIT record.

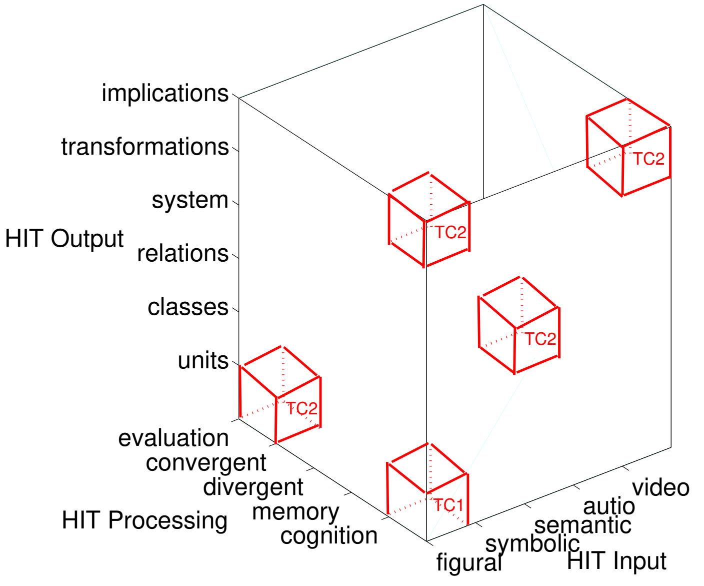

Figure 4. An intelligence space for human intelligence tasks classification

Crowd Context Based Utility Function: In the crowd context based trust evaluation model, we define a feasible utility (denoted as U^C) as the measurement of the crowd trust of the social crowd workers.

$$U^C_{p(a_1,\dots,a_m)} = \omega_{tt} \times TaTrust_{p(a_1,\dots,a_m)} + \omega_{rt} \times RaTrust_{p(a_1,\dots,a_m)} \tag{3}$$

where $TaTrust_{p(a_1,\dots,a_m)}$, $RaTrust_{p(a_1,\dots,a_m)}$ are the aggregated task type based trust and the task reward amount based trust of the social path from the requestor to the worker. ω_{tt} and ω_{rt} are the weights of $TaTrust_{p(a_1,\dots,a_m)}$, $RaTrust_{p(a_1,\dots,a_m)}$, respectively; $0 < \omega_{tt}, \omega_{rt} < 1$ and $\omega_{tt} + \omega_{rt} = 1$.

4.3 Average Utility Function

Here we define an *average utility function* based on *social context* based utility and *task based context* based utility to determine workers' trust levels.

$$U_{p(a_1,\dots,a_m)} = \omega_{U^C} \times U^C_{p(a_1,\dots,a_m)} + \omega_{U^S} \times U^S_{p(a_1,\dots,a_m)} \tag{4}$$

where ω_{U^C} and ω_{U^S} are the weights of $U^C_{p(a_1,\dots,a_m)}$ and $U^S_{p(a_1,\dots,a_m)}$, $0 < \omega_{U^C}, \omega_{U^S} < 1$ and $\omega_{U^C} + \omega_{U^S} = 1$.

4.4 Problem Definition

In order to find trustworthy workers in the social crowd we need to find social trust path from a requestor to a target worker, which can be modeled as a Multi-Constrained Optimal Path (*MCOP*) selection problem. The *MCOP* has been moved to be an NP-Complete problem [22].

Social Trust Path Selection: Given a contextual social network $G(V, E, LV, LE)$, a source requestor R, a target worker pool TW_{set}, a group of QoT constraints $\lambda_T, \lambda_r \lambda_\rho$, and constraints of TaTrust λ_{TT} and of RaTrust λ_{RT}. A social path from requestor R to target worker TW_i can be selected if, and only if the following conditions hold:

1. $T_{p(R,\ldots,TW_i)} \geq \lambda_T$, $r_{p(R,\ldots,TW_i)} \geq \lambda_r$, $\rho_{p(R,\ldots,TW_i)} \geq \lambda_\rho$;
2. $TaTrust_{p(R,\ldots,TW_i)} \geq \lambda_{TT}$, $RaTrust_{p(R,\ldots,TW_i)} \geq \lambda_{RT}$.

When finding social trust path from R to TW_i, the target worker TW_i can be selected if the social path can deliver high values of $U_{p(R,\ldots,TW_i)}$. Moreover, the higher the value of $U_{p(R,\ldots,TW_i)}$, the more trustworthy of the target worker TW_i.

5 CONTEXT-AWARE TRUSTWORTHY WORKER SELECTION ALGORITHM

In this section, we propose a new context-aware trustworthy worker selection algorithm, *CAT*, based on the Monte Carlo method [21], taking both *social contexts* and *task based contexts* into consideration.

Monte Carlo Method: The Monte Carlo method [21] is a computational algorithm which relies on repeated random sampling to compute results. It is also one of the techniques for solving NP-complete problems [21]. Generally, the Monte Carlo method consists of four steps:

1. defining a domain of inputs,
2. generating inputs randomly,
3. performing a computation on each input, and
4. aggregating the results into the final result.

Algorithm Description: *CAT* adopts Monte Carlo method to search a network from v_t (target worker) to v_s (source requestor), and from v_s to v_t respectively. During this process, *CAT* selects up to K candidates at each of the search step. The major idea is depicted as follows.

In our model, we propose an objective function (denoted as δ) to investigate whether a worker can satisfy the requirement from requestor. Obviously, if a social worker satisfies QoT constraints, *TaTrust* and *RaTrust* constraints that requestor sets, it means that each aggregated QoT attribute (i.e., T, r or ρ), *TaTrust* and *RaTrust* of the social path from the requestor to the worker should be larger than the corresponding constraint. And the constraint function is calculated as follows.

$$\delta_{(p)} = \max\left\{\left(\frac{1-T_p}{1-\lambda_T}\right), \left(\frac{1-r_p}{1-\lambda_r}\right), \left(\frac{1-\rho_p}{1-\lambda_\rho}\right),\right.$$
$$\left.\left(\frac{1-\mathit{TaTrust}_p}{1-\lambda_{TT}}\right), \left(\frac{1-\mathit{RaTrust}_p}{1-\lambda_{RT}}\right)\right\} \tag{5}$$

where T_p, r_p, ρ_p are the aggregated *social contexts* values; $\mathit{TaTrust}_p$ and $\mathit{RaTrust}_p$ are the aggregated *task based contexts* values; $\lambda_T, \lambda_r, \lambda_\rho, \lambda_{TT}, \lambda_{RT}$ are the corresponding constraints that the requestor sets. Obviously, the smaller δ value of the social path, the better a worker can satisfy the requirements from requestor.

From Equation (5), we can see that if any aggregated context values of target worker cannot satisfy corresponding constraints, then $\delta_{(p)} > 1$. Otherwise $\delta_{(p)} \le 1$.

1. *Backward Search* aims to investigate whether a worker can satisfy the threshold that the requestor sets (i.e., if $\delta_{(p)} \le 1$), and a few workers that cannot satisfy the corresponding requirements of the requestor could be rejected. At each search step of this procedure, *CAT* calculates δ values of social path from the target worker node to the the neighbouring nodes of the current expansion node, and yield up to K minimum δ values as the next candidate expansion nodes. One of them will be selected as the next expansion node based on the probability calculated by Equation (6). Then the corresponding aggregated context values are recorded at v_{k_i}.

$$\mathit{Pro}^B_{(v_{k_i})} = \frac{\delta_{(p_{v_{k_i}\to v_t})}}{\sum_{i=1}^K \delta_{(p_{v_{k_i}\to v_t})}} \tag{6}$$

where $\mathit{Pro}^B_{(v_{k_i})}$ is the probability of v_{k_i} to be selected as the next expansion node.

The following Theorem 1 illustrates that the social path identified by *Backward Search* procedure can investigate whether there exists a feasible path in the subnetwork.

Theorem 1. The path identified by *Backward Search* procedure with the minimal δ converge to a feasible solution if one exists in a subnetwork.

Proof. Let p_s be a path from v_t to v_s with the minimal δ at v_t delivered by the *Backward Search* procedure. p_* be a feasible solution, and $\delta_{(p_s)} < \delta_{(p_*)}$. Assume p_s is not a feasible solution, then $\exists \varphi \in \{T, r, \rho, \mathit{TaTrust}, \mathit{RaTrust}\}$ that $\varphi_{p_s} < \lambda_\varphi$. Hence, $\delta_{p_s} > 1$. Since p_* is a feasible solution, then $\delta_{(p_*)} < 1$ and $\delta_{(p_s)} > \delta_{(p_*)}$. This contradicts $\delta_{(p_s)} < \delta_{(p_*)}$. Therefore, p_s is a feasible solution. $\square$

2. *Forward Search* intends to compute how trustworthy a worker is (i.e., how large the U value of the social path can get). This procedure uses the information provided by *Backward Search* process. At each of the neighbouring nodes of the current expansion node, *CAT* calculates the aggregated T, r, ρ, *TaTrust*, *RaTrust* values of the social path from v_s to an intermediated node v_m (denoted as path p_m^F). Let p_m^B denotes the path from v_m to v_t identified by the *Backward Search* process. Then a foreseen path from v_s to v_t via v_m (denoted as $p_{fm} = p_m^F + p_m^B$) can be identified. The aggregated contexts values of p_{fm} can be calculated by the same method depicted in Section 4. Then, *CAT* calculates U values of the social path from the source requestor to the neighbouring nodes of the current expansion node, and yield up to K maximum U values of social path of the target workers as the next candidate expansion node. One of them that has the largest Pro^F will be selected as the next expansion worker based on Equation (7) as blow. The following Theorem 2 illustrates that the social trust path identified by the *Forward Search* procedure converges to the optimal solution.

Theorem 2. If K is not less than the maximal outdegree of a subnetwork, the solution p_t identified by the *Forward Search* procedure converges to the optimal solution with the increase of the Forward simulation times.

Proof. Assume the optimal solution in the subnetwork is denoted as p_o, and the path identified by the *Forward Search* procedure is denoted as p_t. if $U_{(p_o)} > U_{(p_t)}$, then $\exists v_i \in p_o$ and $\exists v_j \in p_t (v_i \neq v_t, v_j \neq v_t)$, $U_{(p_{v_s \to v_i})} = 0$ and $U_{(p_{v_s \to v_j})} = 1$. As $T_{p_{v_s \to v_i}} = 0$, $r_{p_{v_s \to v_i}} = 0$, $\rho_{p_{v_s \to v_i}} = 0$ and $TaTrust_{p_{v_s \to v_i}} = 0$, $RaTrust_{p_{v_s \to v_i}} = 0$. Then $T_{p_o} = 0$ and $r_{p_o} = 0$, and thus cannot satisfy the constraints $\lambda_\varphi \in (0, 1)$, $(\varphi \in \{T, r, \rho\})$. Then p_o is an infeasible solution, which contradicts p_o is an optimal solution. Therefore, $U_{(p_o)} = U_{(p_t)}$. So, Theorem 2 is correct. □

Backward Selection Process:

Step 1: Select an unvisited node v_t from B_{set} and mark v_t as visited;
Step 2: Select up to K neighbours of the current expansion node v_t, which have K minimum δ values based on Equation (6);
Step 3: Choose one of them as the next expansion node (denoted as $v_{(k_i)}$) based on Equation (7), and store the corresponding aggregated context values at $v_{(k_i)}$;
Step 4: If $v_{(k_i)}$ is not the source requestor, go to step 1. Otherwise, if $v_{(k_i)}$ is the source requestor and $\delta_{v_s \to v_t} \leq 1$, start *Forward Selection Process*; else if $\delta_{v_s \to v_t} > 1$, delete the worker in the worker pool.

Forward Selection Process: Step 5: Select an unvisited node v_s from F_{set} and mark v_s as visited;
Step 6: Select up to K neighbouring nodes of the current expansion node v_s that have K maximum U values based on Equation (4);

Step 7: Choose one of them as the next expansion node (denoted as $v_{(k_j)}$) based on Equation (8), and store the corresponding aggregated context values at $v_{(k_j)}$;

Step 8: If $v_{(k_j)}$ is not the target worker, go to step 1. Otherwise, return U values of the social path from the source requestor to the target worker.

Summary: The time complexity of *CAT* is $O(mulK)$, where m is the number of simulations, u is the maximal outdegree of nodes, l is the average length of the social trust path from the requestor to the worker, K is the argument specified for K-path selection. In social networks, usually $l < 7$ according to the *small-world* characteristic [25] and K is a constant. Thus the time complexity of *CAT* is $O(mu)$. This work focuses on how to evaluate the trustworthiness of the workers in social crowd environments. As our proposed new trust evaluation model is based on the algorithm proposed in [23], where the efficiency of the baseline algorithm adopted in this work has been validated.

6 EXPERIMENTS

6.1 Experimental Setting

Datasets: In the experiment, in order to evaluate the performance of our proposed trust model in the subnetworks of different scales and structures, we extract 5 subnetworks from each of the 4 large-scale real-world social graphs available at *snap.stanford.edu*, which have been widely used in the literature for social network analysis, by randomly selecting 5 pairs of source and target nodes from each of the 4 large-scale datasets. The details of these datasets are shown in Table 1.

Experimental Settings:

1. As we have discussed in Section 3, the *social contexts* values can be mined from the existing social networks using data mining techniques, which is another very challenging problem, but out of the scope of this work. In addition, the *TaTrust* and *RaTrust* values can be obtained by using the method proposed in [24]. Thus, without loss of generality, we randomly set the T, r, ρ, *TaTrust*, *RaTrust* values by using the function *rand()* in SQL.
2. In addition, as we have discussed in Section 3, a requestor can set constraints for different contexts. Therefore, in the experiments, we specify a set of relative low constraints (i.e., $\lambda_T = 0.005$, $\lambda_r = 0.005$, $\lambda_\rho = 0.005$, $\lambda_{TT} = 0.005$, $\lambda_{RT} = 0.005$) to ensure the high possibility of having one feasible solution. Otherwise, no solution might be delivered by the algorithm so that we cannot compare the performance difference.
3. Moreover, to the best of our knowledge, *CrowdTrust* [24] is the state-of-the-art context-aware trust evaluation method for crowdsourcing systems in the literature which considers the *task based contexts*, i.e., the *task types* and the

Network	Subnetwork ID	Vertices	Edges
Twitter	T1	783	2 385
	T2	638	1 705
	T3	747	2117
	T4	562	1 500
	T5	549	1 536
DBLP	D1	298	5 680
	D2	309	3 958
	D3	162	4 609
	D4	157	3 639
	D5	213	5 021
Epinions	E1	4 551	16 939
	E2	2 476	5 080
	E3	2 696	6 173
	E4	2 559	5 418
	E5	2 499	5 780
Slashdot	S1	5 156	10 307
	S2	3 543	5 966
	S3	3 592	6274
	S4	4 010	7 486
	S5	5 049	9 859

Table 1. The experimental datasets

task reward amounts. So we compare the performance difference between *CrowdTrust* and our proposed *CAT* algorithm. In *CrowdTrust* model, if a worker can satisfy the constraints of the *TaTrust* and the *RaTrust* that requestor sets, he can be selected as a trustworthy worker.

Experimental Environment: All the experiments are implemented using Matlab R2013b on a machine powered by two Intel Core i5-3470 CPU, 3.20 GHz, 8 GB RAM, Windows 7 operating system and MySQL 5.6 database. All the experimental results are averaged based on two independent runs.

6.2 Experimental Results

In order to investigate the performance of *CrowdTrust* [24] and *CAT*, we evaluate the average utility values of the social path from source requestor and target worker with different simulation times varying from 500 to 2 500.

Result 1. Figure 4 to Figure 7 plot the comparison of average utility values of the social path that from source requestor to target worker in 20 subnetworks based on *CrowdTrust* and *CAT*. From the figures, we can see that in all cases, *CAT* can deliver higher average utility than *CrowdTrust*. This is because that *CAT* considers both *social contexts* and *task based contexts* when evaluating workers' trust levels,

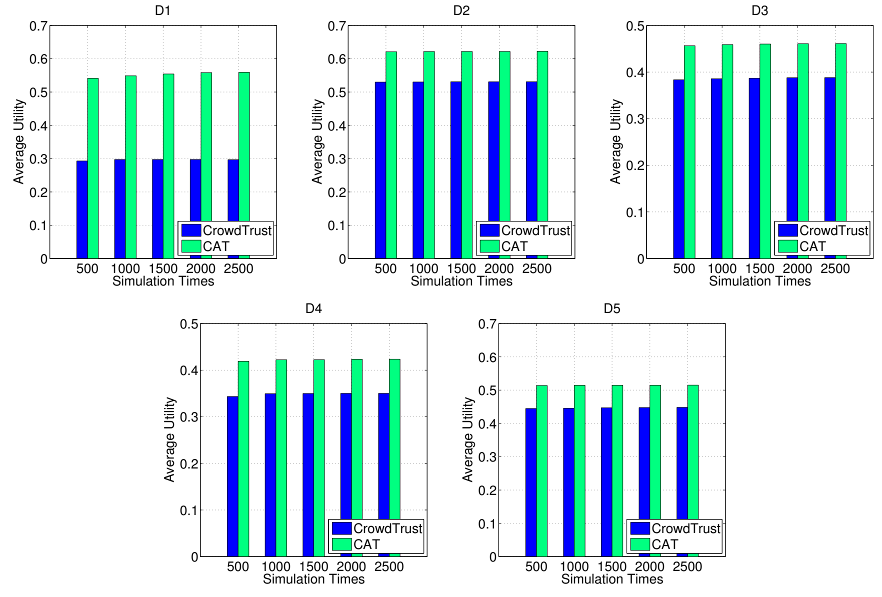

Figure 5. The comparison of average utility on DBLP

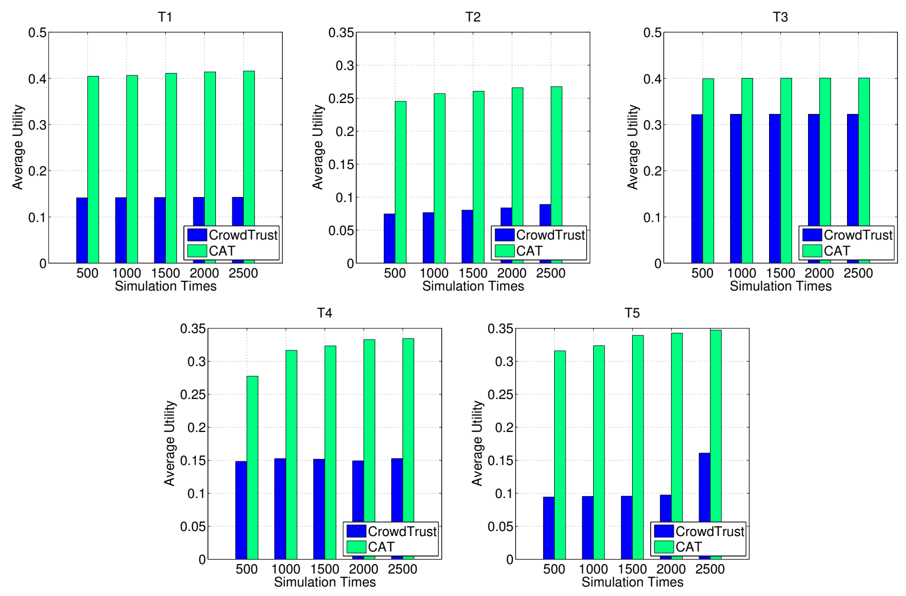

Figure 6. The comparison of average utility on Twitter

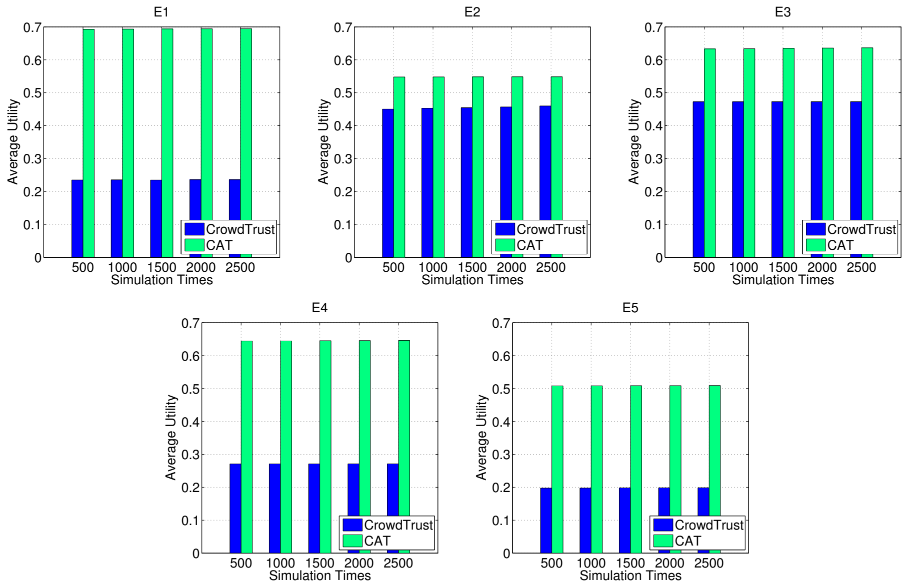

Figure 7. The comparison of average utility on Epinions

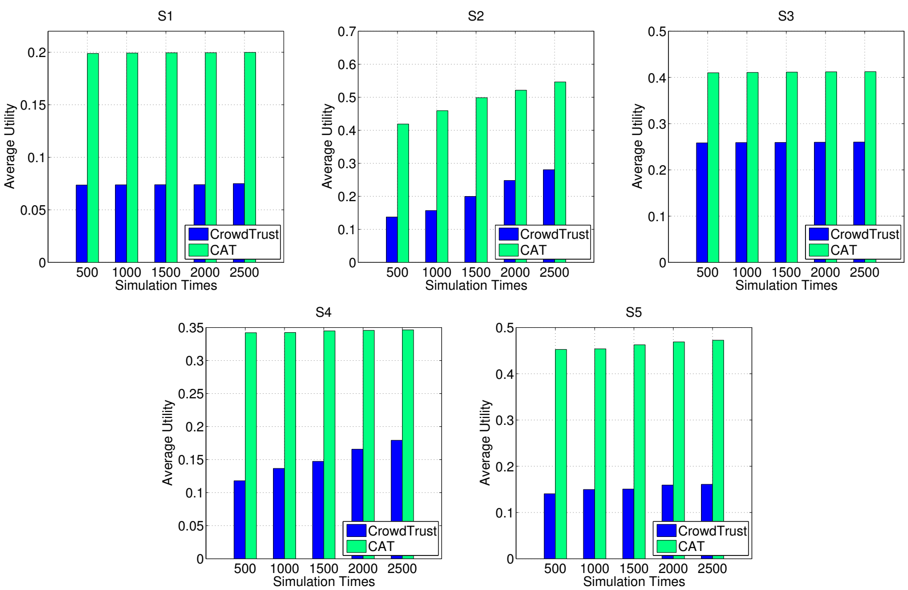

Figure 8. The comparison of average utility on Slashdot

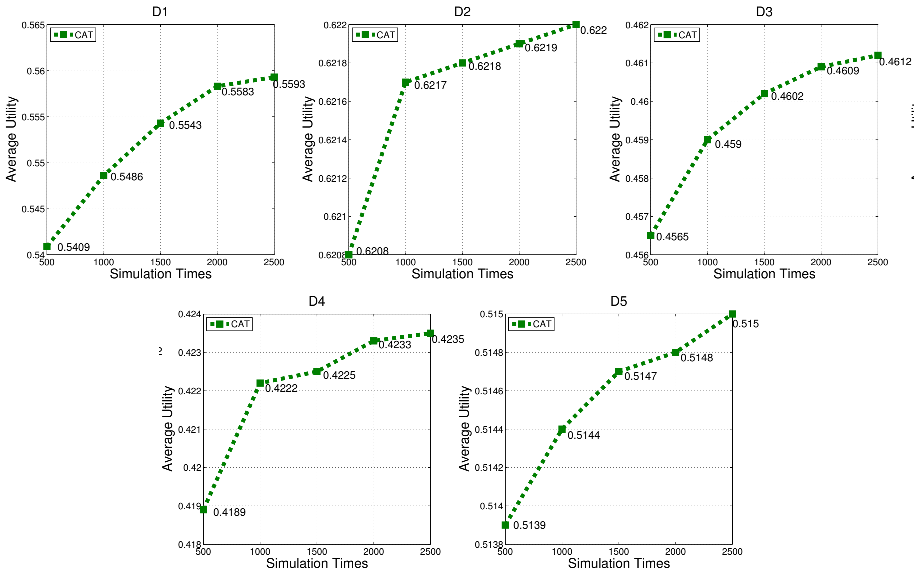

Figure 9. Average utility on DBLP with different simulation times

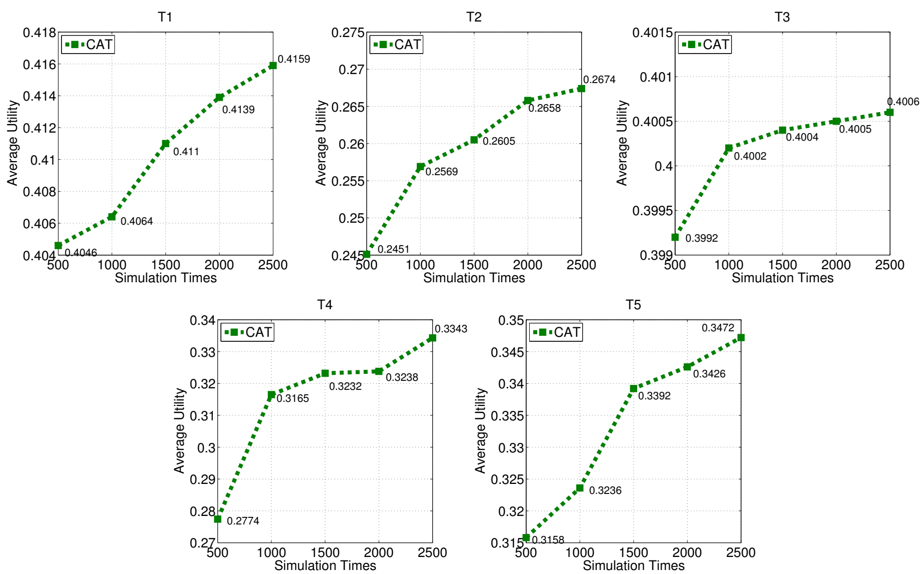

Figure 10. Average utility on Twitter with different simulation times

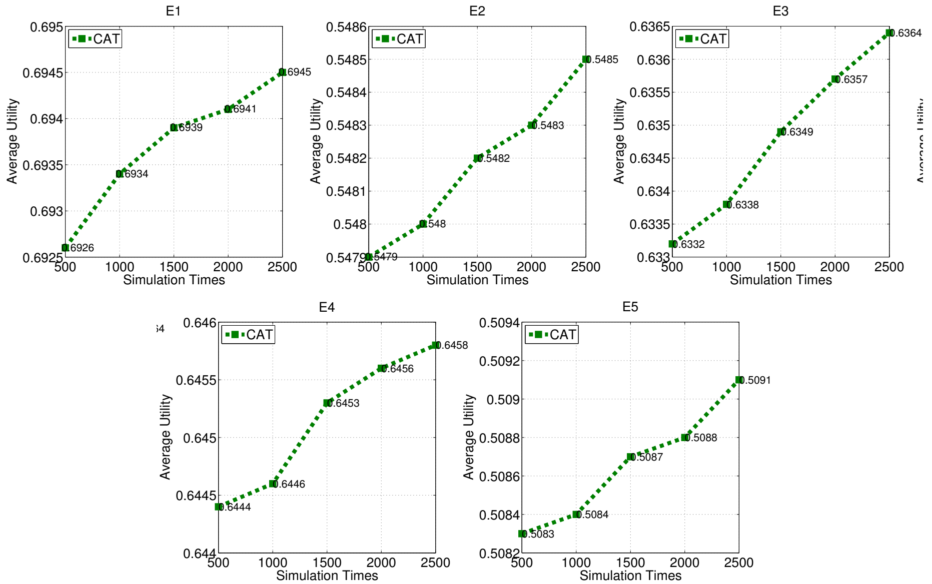

Figure 11. Average utility on Epinions with different simulation times

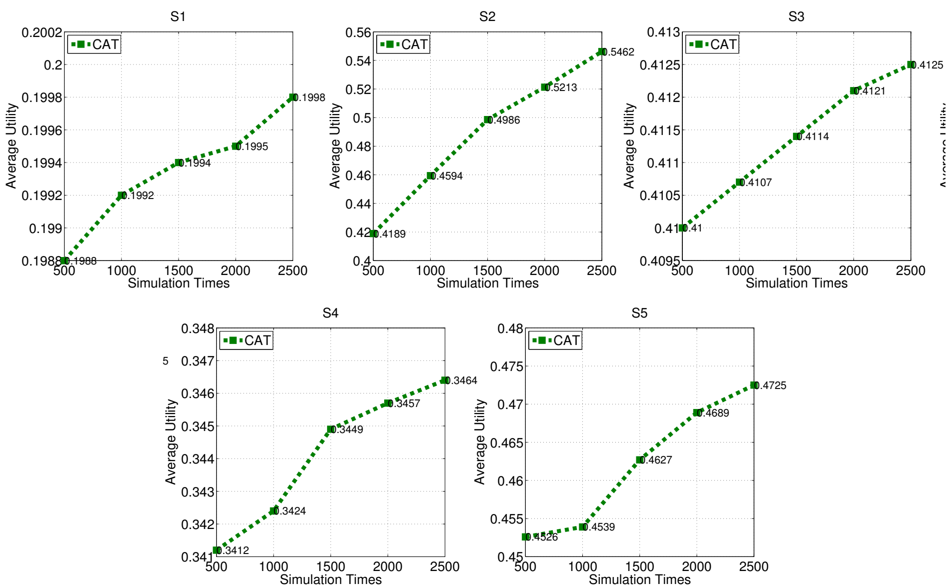

Figure 12. Average utility on Slashdot with different simulation times

i.e., only when the target worker satisfies both of the constraints of *social contexts* (i.e., T, R, ρ) and *task based contexts* (i.e., $TaTrust, RaTrust$), the worker can be selected as a trustworthy worker. However, *CrowdTrust* only takes the *task based contexts* into consideration, i.e., as long as the target worker satisfies the *task based contexts* constraints, the worker can be selected, and that leads to utility delivered by *CrowdTrust* can be remarkably lower than that of *CAT*. Based on the statistics, under the same simulation times, on average, the utility values delivered by *CrowdTrust* is 0.3056 and by *CAT* is 0.395. On average, *CAT* can deliver 22.6 % higher utility values than *CrowdTrust*.

Result 2. Figure 8 to Figure 11 plot the average utility of *CAT* with different simulation times. From the figures, we can see that with the increase of simulation times, the average utility values delivered by *CAT* increase. This is because the more the simulation times, the more the nodes are identified by *Backward Search* process, which can provide more information for the *Forward Search* process to find higher average utility of social path from a source requestor to a target worker. In addition, at each simulation, we store the context values of the social path with the minimal $\delta_{(p)}$ at the identified node during the *Backward Search* process and select the social path with maximum $U_{(p)}$ values at *Forward Search* process. So the more the simulation times, the higher possibility to select a social path with higher utility. Moreover, based on the property of our proposed algorithm, when the intermediate modes record the optimal aggregated social context values, the utility can converge to the optimal value. The simulation times of the convergence can be different with different scales and structures of social networks. Based on our statistic, on average, the curve will converge to be horizontal when the simulation times reach 3 000 times.

7 CONCLUSIONS

In this paper, we have proposed a context-aware trust-oriented worker selection method, which takes both the *social contexts* and the *task based contexts* into consideration. In order to solve the NP-Complete trustworthy worker selection problem, we have proposed a new approximation algorithm *CAT*. The results of experiments conducted on four real-world datasets has demonstrated that our trust model can effectively identify trustworthy workers.

Acknowledgments

This work was partially supported by Natural Science Foundation of China (Grant Nos. 61303019, 61572336, 61532018, 61402313, 61502324), Doctoral Fund of Ministry of Education of China (20133201120012), Postdoctoral Science Foundation of China (2015M571805, 2016T90492), Collaborative Innovation Center of Novel Software Technology and Industrialization, Jiangsu, China, and the Opening Project of Guangdong Province Key Laboratory of Big Data Analysis and Processing (2017-002).

REFERENCES

[1] Howe, J.: The Rise of Crowdsourcing. Wired Magazine, Vol. 14, 2006, No. 6, pp. 1–4.

[2] Yang, X. S.—Cheng, R.—Mo, L.—Kao, B.—Cheung, D. W.: On Incentive-Based Tagging. 2013 IEEE 29th International Conference on Data Engineering (ICDE), IEEE, 2013, pp. 685–696, doi: 10.1109/icde.2013.6544866.

[3] Wang, J.—Kraska, T.—Franklin, M. J.—Feng, J.: Crowder: Crowdsourcing Entity Resolution. Proceedings of the VLDB Endowment, Vol. 5, 2012, No. 11, pp. 1483–1494, doi: 10.14778/2350229.2350263.

[4] Hung, N. Q. V.—Tam, N. T.—Miklós, Z.—Aberer, K.: On Leveraging Crowdsourcing Techniques for Schema Matching Networks. Database Systems for Advanced Applications, Springer, 2013, pp. 139–154, doi: 10.1007/978-3-642-37450-0_10.

[5] Feng, A.—Franklin, M.—Kossmann, D.—Kraska, T.—Madden, S. R.—Ramesh, S.—Wang, A.—Xin, R.: CrowdDB: Query Processing with the VLDB Crowd. Proceedings of the VLDB Endowment, Vol. 4, 2011, No. 12.

[6] Liu, X.—Lu, M.—Ooi, B. C.—Shen, Y.—Wu, S.—Zhang, M.: CDAS: A Crowdsourcing Data Analytics System. Proceedings of the VLDB Endowment, Vol. 5, 2012, No. 10, pp. 1040–1051, doi: 10.14778/2336664.2336676.

[7] Marcus, A.—Wu, E.—Karger, D.—Madden, S.—Miller, R.: Human-Powered Sorts and Joins. Proceedings of the VLDB Endowment, Vol. 5, 2011, No. 1, pp. 13–24, doi: 10.14778/2047485.2047487.

[8] Park, H.—Garcia-Molina, H.—Pang, R.—Polyzotis, N.—Parameswaran, A.—Widom, J.: Deco: A System for Declarative Crowdsourcing. Proceedings of the VLDB Endowment, Vol. 5, 2012, No. 12, pp. 1990–1993.

[9] Whang, S. E.—Lofgren, P.—Garcia-Molina, H.: Question Selection for Crowd Entity Resolution. Proceedings of the VLDB Endowment, Vol. 6, 2013, No. 6, pp. 349–360, doi: 10.14778/2536336.2536337.

[10] Li, H.—Zhao, B.—Fuxman, A.: The Wisdom of Minority: Discovering and Targeting the Right Group of Workers for Crowdsourcing. Proceedings of the 23rd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2014, pp. 165–176, doi: 10.1145/2566486.2568033.

[11] Cao, C. C.—She, J.—Tong, Y.—Chen, L.: Whom to Ask?: Jury Selection for Decision Making Tasks on Micro-Blog Services. Proceedings of the VLDB Endowment, Vol. 5, 2012, No. 11, pp. 1495–1506, doi: 10.14778/2350229.2350264.

[12] Cao, C. C.—Tong, Y.—Chen, L.—Jagadish, H.: Wisemarket: A New Paradigm for Managing Wisdom of Online Social Users. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 455–463, doi: 10.1145/2487575.2487642.

[13] Forlines, C.—Miller, S.—Guelcher, L.—Bruzzi, R.: Crowdsourcing the Future: Predictions Made with a Social Network. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, ACM, 2014, pp. 3655–3664, doi: 10.1145/2556288.2556967.

[14] Brabham, D. C.: Crowdsourcing as a Model for Problem Solving an Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies, Vol. 14, 2008, No. 1, pp. 75–90.

[15] Parameswaran, A. G.—Garcia-Molina, H.—Park, H.—Polyzotis, N.—Ramesh, A.—Widom, J.: Crowdscreen: Algorithms for Filtering Data with Humans. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, ACM, 2012, pp. 361–372, doi: 10.1145/2213836.2213878.

[16] Zhao, Z.—Wei, F.—Zhou, M.—Chen, W.—Ng, W.: Crowd-Selection Query Processing in Crowdsourcing Databases: A Task-Driven Approach. International Conference on Extending Database Technology, Brussels, 2015.

[17] Adler, P. S.: Market, Hierarchy, and Trust: The Knowledge Economy and the Future of Capitalism. Organization Science, Vol. 12, 2001, No. 2, pp. 215–234.

[18] Miller, R.: Intimate Relationships. McGraw-Hill Higher Education, 2007.

[19] Guilford, J. P.: The Nature of Human Intelligence. 1967.

[20] Berger, P. L.—Luckmann, T.: The Social Construction of Reality: A Treatise in the Sociology of Knowledge. No. 10. Penguin UK, 1991.

[21] Gentle, J. E.—Härdle, W. K.—Mori, Y.: Handbook of Computational Statistics: Concepts and Methods. Springer Science and Business Media, 2012.

[22] Korkmaz, T.—Krunz, M.: Multi-Constrained Optimal Path Selection. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2001), Vol. 2, IEEE, 2001, pp. 834–843, doi: 10.1109/infcom.2001.916274.

[23] Liu, G.—Liu, A.—Wang, Y.—Li, L.: An Efficient Multiple Trust Paths Finding Algorithm for Trustworthy Service Provider Selection in Real-Time Online Social Network Environments. 2014 IEEE International Conference on Web Services (ICWS), IEEE, 2014, pp. 121–128, doi: 10.1109/ICWS.2014.29.

[24] Ye, B.—Wang, Y.—Liu, L.: Crowd Trust: A Context-Aware Trust Model for Worker Selection in Crowdsourcing Environments. 2015 IEEE International Conference on Web Services (ICWS), IEEE, 2015, pp. 121–128, doi: 10.1109/ICWS.2015.26.

[25] Mislove, A.—Marcon, M.—Gummadi, P. K.—Druschel, P.—Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. Internet Measurement Comference, 2007, pp. 29–42, doi: 10.1145/1298306.1298311.

[26] Wang, Y.—Li, L.—Liu, G.: Social Context-Aware Trust Inference for Trust Enhancement in Social Network Based Recommendations on Service Providers. World Wide Web, Vol. 18, 2015, No. 1, pp. 159–184.

[27] Liu, G.—Wang, Y.—Orgun, M. A.—Liu, H.: Discovering Trust Networks for the Selection of Trustworthy Service Providers in Complex Contextual Social Networks. 2012 IEEE 19$^{\mathrm{th}}$ International Conference on Web Services (ICWS), IEEE, 2012, pp. 384–391.

[28] Liu, G.—Wang, Y.—Orgun, M. A. et al.: Social Context-Aware Trust Network Discovery in Complex Contextual Social Networks. AAAI, Vol. 12, 2012, pp. 101–107.

[29] Liu, G.—Wang, Y.: Trust-Oriented Service Provider Selection in Complex Online Social Networks. Advanced Web Services, Springer, 2014, pp. 363–380, doi: 10.1007/978-1-4614-7535-4_15.

[30] Zheng, X.—Wang, Y.—Orgun, M. A.—Liu, G.—Zhang, H.: Social Context-Aware Trust Prediction in Social Networks. International Conference on Service-Oriented Computing, Springer, 2014, pp. 527–534, doi: 10.1007/978-3-662-45391-9_45.

[31] Golbeck, J.—Hendler, J.: Inferring Binary Trust Relationships in Web-Based Social Networks. ACM Transactions on Internet Technology (TOIT), Vol. 6, 2006, No. 4, pp. 497–529, doi: 10.1145/1183463.1183470.

[32] Tang, J.—Zhang, J.—Yan, L.—Li, J.—Zhang, L.—Su, Z.: Arnetminer: Extraction and Mining of Academic Social Networks. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2008, pp. 990–998, doi: 10.1145/1401890.1402008.

Yang Zhao is current Master Student at the Research Center on Advanced Data Analytics (ADA) in Soochow University, China. Her research interest is social influence evaluation.

Guanfeng Liu is Associate Professor of the Research Center on Advanced Data Analytics (ADA) in Soochow University, China. He received his Ph.D. degree in computer science from Macquarie University, Australia in 2013. His research interests include social networks and graph mining.

Kai Zheng is currently Professor in the School of Computer Science at Soochow University. He received his Ph.D. degree in computer science from The University of Queensland in 2012. His research focus is to find effective and efficient solutions for managing, integrating and analyzing big data for business, scientific and personal applications. He has been working in the area of social-media analysis, spatial-temporal databases, uncertain databases, data mining and bioinformatics.

An Liu is Associate Professor of the Research Center on Advanced Data Analytics (ADA) in Soochow University, China. He received the Ph.D. degree from the University of Science and Technology of China, Hefei, and the City University of Hong Kong, Hong Kong, both in computer science. His research interests are in the area of service-oriented computing, social network, information security, and spatial databases.

Zhixu Li is currently Associate Professor at the Research Center on Advanced Data Analytics (ADA) in Soochow University, China. He used to work as a research fellow at King Abdullah's University of Science and Technology (KAUST). Prior to that, he received his Ph.D. degree from the University of Queensland in 2013, and his B.Sc. and M.Sc. degrees from Renmin University of China in 2006 and 2009, respectively. His research interests are data cleaning, big data applications, information extraction and information retrieval.

Xiaofang Zhou is Professor of computer science with The University of Queensland and Adjunct Professor with the School of Computer Science and Technology, Soochow University, China. He received his B.Sc. and M.Sc. degrees in computer science from Nanjing University, China, in 1984 and 1987, respectively, and his Ph.D. degree in computer science from The University of Queensland, Australia, in 1994. His research interests include spatial and multimedia databases, high performance query processing, web information systems, data mining, bioinformatics, and e-research.