

Mechanism Design for Finding Experts Using Locally Constructed Social Referral Web

Lan Zhang*, Xiang-Yang Li*[†], Jingsheng Lei[‡], Yunhao Liu*[§]

* Tsinghua University, [†] Illinois Institute of Technology

[‡] Shanghai University of Electric Power, [§] Hong Kong University of Science and Technology



Abstract—In this work, we address the problem of distributed expert finding using chains of social referrals and profile matching with only local information in online social networks. By assuming that users are selfish, rational, and have privately known cost of participating in the referrals, we design a novel truthful efficient mechanism in which an expert-finding query will be relayed by intermediate users. When receiving a referral request, a participant will locally choose among her neighbors some user to relay the request. In our mechanism, several closely coupled methods are carefully designed to improve the performance of distributed search, including, profile matching, social acquaintance prediction, score function for locally choosing relay neighbors, and budget estimation. We conduct extensive experiments on several datasets of online social networks. The extensive study of our mechanism shows that the success rate of our mechanism is about 90% in finding closely matched experts using only local search and limited budget, which significantly improves the previously best rate 20%. The overall cost of finding an expert by our truthful mechanism is about 20% of the untruthful methods, *e.g.* the method that always selects high-degree neighbors. The median length of social referral chains is 6 using our localized search decision, which surprisingly matches the well-known small-world phenomenon of global social structures.

Index Terms—Mechanism Design, Strategyproof, Social Networks, Referral Web, Distributed Search, Small World.

1 INTRODUCTION

Finding experts matching desired attributes is common and useful in real life, for example, to find a student studying math from a nearby collage to be a home tutor or to invite a computer engineer professor to give a talk. Online social networks, like Facebook and LinkedIn, are powerful resources for finding experts due to the abundant personal attribute information. So far, there are two methods of expert finding in industry and literature. One is searching in a large database with the global information. The other strategy is distributed search using a chain of social referrals through acquaintances. Searching a referral chain can help people find the experts who cannot be obtained by using search engines. Evans *et al.* [10] found that targeting questions to specific friends can receive in-depth answers. Asking a favor from or making friends with the target person are usually much easier with recommendations from his friends than just being a stranger [18]. Moreover, there

are limits as to the amount and the kinds of information that a user is able or willing to make available to the public at large [14]. *E.g.*, many users of online social networks have made their profiles unsearchable to public or only visible to friends or people from certain networks. Searching for some information or experts thus becomes a matter of searching the social network with a chain of personal referrals from the searcher (or initiator) to the expert with only local information.

The homophily principle [20] enables the efficient construction of referral chains in online social networks. A lot of existing work has proved that social networks are searchable through short pairwise connections [13], [31], [32]. For online social networks (OSN), our analysis (presented in Section 4) of a Facebook dataset shows about 99% users are within 6 hops on average. Compared to using a single social networking site, a cross-sites people search may be much more helpful with hundreds of social networking sites. For example, if you are looking for a trustworthy lawyer who works in your city and one of your good friends on Facebook knows a well reputed lawyer on LinkedIn who is unsearchable. In this case, an efficient cross-site user search/referral can be easily completed through your friend but not any single site search engine. Aggregation of information from multiple social networks for a person has been facilitated by projects such as OpenID, DataPortability, the social graph project [12], and various microformats. Existing aggregators include Friend-Feed, MyBlogLog, Jaiku, and Plaxo.

The social referral path is supposed to be done voluntarily in existing work. The results along this line typically emphasize that the completed paths tend to be short, thus ignoring the fact that a vast majority of paths never reach their ultimate targets [24]. It has been reported in [33] that the only parameter governing the success of a search is not related to the topology or search procedure, but the probability of termination at each step. Most paths are terminated for the reason that participants are *not sufficiently motivated* to relay messages. Thus, taking user's self-interest into consideration is necessary for a successful chain of social referrals. In this paper, we propose a new search mechanism for expert finding in online social networks. Our mechanism

exploits the homophily principle of social networks. Based on the similarity calculation, and acquaintance probability estimation, we design a truthful local search mechanism that works well even with possible selfish behavior.

Our Main Contributions: In this work, we design a truthful mechanism for expert finding by a chain of individuals from the initiator to the expert, where each intermediate user makes a decision using only local information. Our mechanism also takes the users' self-interest into account with a well-designed payment strategy. We assume that each intermediate user has a privately known cost of participating in the chain of social referrals. We theoretically prove that our mechanism is truthful, *i.e.*, each intermediate user will maximize her utility if she truthfully declared her cost and executed the search procedure. We also estimate the required budget for searching an expert with some profile. We conduct extensive experiments to study the performance of our mechanism. Our experimental results show that the social referral path found by our mechanism is significantly shorter than the one found by previous approaches. The total cost of intermediate agents participating in the chain is also much smaller than naive approaches such as using a high-degree neighbor. Moreover, the success rate of our localized search strategy is about 90%, which is significantly better than the best reported success rate 20% [9], [28]. Part of this work has been published in [44]. Compared with [44], in this paper, we complete the search method with incremental local learning, and add theoretical analysis and proof for our mechanism. We also greatly enrich our mechanism measurements with more dataset analysis and experimental analysis.

Paper Organization: The rest of the paper is organized as follows. In Section 2 we present the network model for expert finding and our similarity calculation method. In Section 3 we present our truthful mechanism for finding experts using chain of referrals. We report our evaluation results in Section 4, review the related work in Section 5 and conclude the paper in Section 6.

2 SYSTEM MODEL AND PRELIMINARIES

2.1 Problem Formulation Using Online Social Networks

A social network is modeled by a graph $G = (V, E)$. Every user $v_i \in V$ in G has a unique identity, called i for simplicity. By profiling or data collection, each user v_i is associated with an m -dimension **profile vector** $A_i = \langle a_i^1, a_i^2, \dots, a_i^m \rangle$, which represents her characteristics and social groups. Here the value a_i^j , represents a characterization of user i for the j -th attribute. The link $v_i v_j \in E$ between v_i and v_j is the acquaintance connection. For example, the friendship in Facebook, the co-authorship in co-author networks and communication in email networks. User v_i is called a neighbor of v_j in the social network G if the link $v_i v_j$ exists. We assume that each user only knows the profile of her neighbors.

In this work, we study finding experts in social networks via a chain of referrals by some intermediate users. Assume that there is an **initiator**, say v_0 , who wants to find an expert, characterized by a profile vector $\mathcal{A}_t = \langle a_t^1, a_t^2, \dots, a_t^m \rangle$. Due to the ambiguity of profile, we say that a user j has a matching profile with \mathcal{A}_t if the “**distance**” between her profile A_j and the target profile \mathcal{A}_t is within a small error bound ϵ . The initiator will ask her neighbors to help her to find a matching expert. The process will be iterated until the expert is found or some termination conditions are met (*e.g.*, the maximum number of referrals, or the total cost incurred for search). The output is a social referral path $\mathbf{P}(v_0, v_t)$ from the initiator v_0 to a target user v_t with the matching profile.

A major difference between the system model used in this study and previous studies for finding friends/experts is that here we assume that each user i has a cost c^i for querying her neighbors to get a target expert for some initiator. We assume that the cost c^i is privately known only to v_i . The initiator originally has a budget B for performing the task of finding experts in the social network. We say that finding experts using such a path $\mathbf{P}(v_0, v_t)$ is feasible if the total cost requested by users on this path is at most the budget B of the initiator. If we know the whole network and the cost vector, the problem becomes the simple shortest path problem. A truthful mechanism can also be designed in such centralized approach [29]. However, in the practical social network setting, several challenges need to be addressed to solve this problem. The challenges include the follows:

- 1) computing the profile similarity and acquaintance probability,
- 2) designing localized referral strategy since each intermediate user only knows her neighbors, and
- 3) designing a payment mechanism to make participants tell truth without any global information for lies checking.

As the number of users grows exponentially with the social distance from the initiator, a search strategy is needed to select the correct user among hundreds of acquaintances to form the next link of the referral path using only local information. There are two basic strategies in people/expert search: (1) choosing a high degree neighbor, or (2) “closest” neighbor by profile distance. In Section 4, we found by experiments that our profile similarity based strategy outperforms the high degree strategy greatly.

2.2 Similarity and Acquaintance Probability

Similarity breeds connection: It has been well observed that *the shorter the social distance between two users, the higher the probability that they are acquainted to each other or have a shorter network distance* [20]. A social distance is usually given as a metric of the similarity or relevance between two users. There are some existing social distance measurements, for example, organizational hierarchy distance [32], or geographical distance. In this work, we use the similarity of two users' attributes as

the social distance measurement to indicate the similarity between them. Then we use users' similarity to estimate their probability of acquaintance.

Distance and similarity of attributes: The attribute could be discrete categorical characteristics such as social group or gender, or numerical characteristics such as age or vertex degree. For each attribute a^k , an **attribute distance** $d_{ij}^k = d^k(a_i^k, a_j^k)$ is defined. For example, the distance of gender could be 0 for two users of the same sex, and 1 otherwise. The distance of age could be $d^k(x, y) = |x - y|$. Based on the attribute distance, the **attribute similarity** is given as

$$s_{ij}^k = 1 - \frac{d_{ij}^k}{\max_{(i,j)} d_{ij}^k}. \quad (1)$$

The **similarity** between users v_i and v_j is a vector S_{ij} (or denoted as $S(A_i, A_j)$) defined as $S_{ij} = \langle s_{ij}^1, s_{ij}^2, \dots, s_{ij}^m \rangle$.

We notice that the similarity of different attributes contribute differently to acquaintance probability. Some attributes are strong evidences of acquaintance and some are weak. Users who are close in the strong positive evidence have a high probability of acquaintance. We model the acquaintance probability of two users using the logistic function [37] about their profile similarity vector. Specifically, if $E_{i,j}$ is the event that the user v_i and v_j are acquainted and $\bar{E}_{i,j}$ is the event that they are not acquainted, define the probability $\Pr(E_{i,j} | S_{ij} = X) = \text{logit}^{-1}(\beta X) = \frac{1}{1 + e^{-\beta X}}$. Here the parameter β is an m -dimensional vector to be studied later. Let $r(X)$ be the odds ratio when that similarity equals X , *i.e.*, we define $r(X) = \frac{\Pr(E_{i,j} | S_{ij} = X)}{1 - \Pr(E_{i,j} | S_{ij} = X)} = \frac{\Pr(E_{i,j} | S_{ij} = X)}{\Pr(\bar{E}_{i,j} | S_{ij} = X)}$. Then for a vector $X = (x^1, x^2, \dots, x^m)$, we have

$$\beta \cdot X = \beta^0 + \beta^1 x^1 + \dots + \beta^m x^m = \ln(r(X)) \quad (2)$$

where β^0 is the intercept, and β^k describes the size of the contribution of the similarity of attribute a^k on the acquaintance relationship. A positive β^k means the similarity of a^k increases the probability of acquaintance, while a negative β^k means a decrease effect; a large β^k means that the attribute a^k strongly influences the probability. We define

$$\text{profile similarity:} \quad s_{ij} = \beta \cdot S_{ij},$$

which is the weighted summary of all the attributes similarities. For profile similarity, the larger the value, the closer the two profiles. In the Facebook dataset (introduced in Section 4.1), the largest profile similarity is about 3.5252 for the MIT dataset and 3.6063 for the Harvard dataset. Note that here the profile similarity could be negative values.

Let the probability that the similarity vector between two randomly selected profiles A_i and A_j equals X be $\Pr(S_{ij} = X)$. Then, the probability that v_i and v_j are acquaintances is $\Pr(E_{i,j} | S_{ij} = X) = \frac{\Pr(S_{ij} = X \wedge E_{i,j})}{\Pr(S_{ij} = X)}$ from Bayes theory [38]. Then a simple computation shows that $r(X) = \frac{\Pr(S_{ij} = X \wedge E_{i,j}) / \Pr(S_{ij} = X)}{\Pr(S_{ij} = X \wedge \bar{E}_{i,j}) / \Pr(S_{ij} = X)} = \frac{|\{(V_i, V_j) | S_{ij} = X \wedge E_{i,j}\}|}{|\{(V_i, V_j) | S_{ij} = X \wedge \bar{E}_{i,j}\}|}$. So we can use logistic regression to

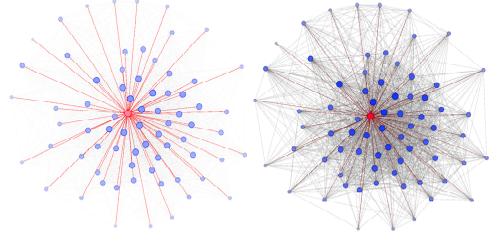


Fig. 1. The ego network of a random selected vertex in facebook social network graph, whose degree is 71.

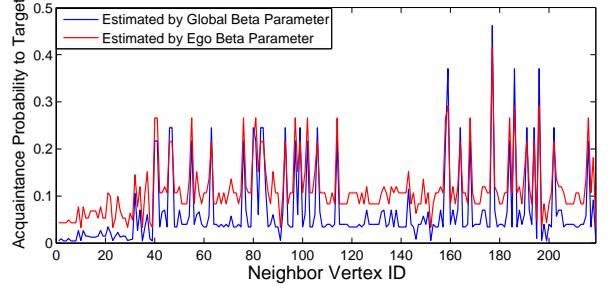


Fig. 2. The acquaintance probability between the neighbors of randomly selected ego network and target estimated by the local parameter β and global parameter β .

derive the parameter β by learning $r(X)$ from a known graph, such as a local graph like the ego network. With the parameter β , we can predict the acquaintance probability in an unknown graph given the profile vectors of two users. Note that the larger the profile similarity, the higher the probability they are acquaintances.

2.3 Ego Network and Incremental Local Learning

Recall that we assume that for finding experts in a social network, every person only knows the information of the target person and his current immediate one-hop neighbors in the social network. Here we present a method to learn the parameter β locally for acquaintance probability calculation.

Before a user starts a search, she can create an ego network (also called a personal network) centered on herself to learn all the information needed for acquaintance probability calculation. In the ego network, the node set V_i includes the vertex itself and all her direct friends in the social network G . The edge set E_i consists of all the edges between vertices in V_i . Figure 1 shows an example of the ego network of a randomly selected person from Facebook dataset. We can learn $r(X)$ from the ego network and β can be updated along the chain during the search procedure.

A natural question is then "Is ego network sufficient enough?". Newman [21] has shown that one's immediate neighbors in the acquaintance network are far from being a random sample, although they show some characteristics of the whole network. We compare the acquaintance probability estimated by globally learned β parameter and β

parameter learned from ego networks, as shown in Figure 2. Here the ego vertex and the target vertex are both chosen randomly. We find that the probabilities almost have the same relative magnitude among all neighbors in both estimations. The difference between the two lines are mainly caused by a higher density of the ego network than the whole graph. So the ego networks in a social network is a typical sample of the whole graph, and the result shows that the ego networks can give us some sufficient evidence on acquaintance.

3 STRATEGYPROOF MECHANISM DESIGN

In this section, we present our truthful mechanism for finding experts in a social network using social referrals.

3.1 Finding Experts Mechanism Definition

We suppose that all users are selfish and rational, which means that they will optimize their strategies to maximize their utilities and they will make consistent decisions in the same conditions. For a searching task, the initiator has a *privately known budget* B and each other user v_i has a *privately known cost* c^i when she is asked to participate in the referrals.

3.1.1 Search Procedure

Before we present our mechanism, we first give an overview of the search procedure for finding experts using a chain of social referrals. There are five phases during our expert search procedure:

- 1) **Initialization phase:** The initiator v_0 checks her neighbors to see whether the target is among them. If the target is among her neighbors, the search stops and it does not incur any additional cost. Otherwise, the initiator estimates the budget B based on the profile similarity s_{0t} between her profile and the target profile \mathcal{A}_t .
- 2) **Bidding phase:** The initiator v_0 announces the search task by giving the profile \mathcal{A}_t of the target expert. Any her neighbor v_k who intends to participate in the social referral will declare a price d^k , which is not necessarily her true cost, via a **sealed-bid** within a bidding time window.
- 3) **Winner decision phase:** The profile similarity between each neighbor v_k and the target is known to v_0 since v_0 can access the profile of v_k . User v_0 chooses a winner neighbor, say v_w , as the next-hop vertex based on a score function (which will be discussed in detail later) and pays a compensation d^k to the chosen neighbor v_w .
- 4) **Execution phase:** The selected neighbor v_w continues the search task as a new initiator whose budget is $B - d^k$ until an expert with a ‘‘closely matching’’ profile is found.
- 5) **Bonus payment phase:** Once the search is completed successfully, the initiator will pay a bonus to every agent in this social referral path.

3.1.2 Algorithmic Mechanism Design

In the expert finding algorithm, the **input** is the target profile \mathcal{A}_t , the budget B , and an acceptable lowest similarity ξ between the profile of the found expert and the sought target given by the initiator. The objective is to find the target by a chain of social referrals and pay intermediate neighbors with the limited budget.

We refer to every selfish rational participant as an agent in this game, who is also a vertex in the social graph. There are n agents. It will cost c^k for each agent v_k to perform a searching task. In this work, we assume that c^k is a **private** input, which is the true cost only known by v_k . When bidding for the task, an agent v_k can choose to declare a price d^k , which could be the true cost c^k or any other valid cost. Let $d = \langle d^1, d^2, \dots, d^n \rangle$. We define our expert finding **mechanism** as $\mathbf{M} = (\mathbf{O}, \mathbf{p})$, which is composed of an **output function** $\mathbf{O}(d)$ and an n -tuple **payment function** $\langle \mathbf{p}^1(d), \mathbf{p}^2(d), \dots, \mathbf{p}^n(d) \rangle$. An output \mathbf{O} is a social referral path consists of a sequence of vertices, say $\mathbf{O} = \{v_{j_1}, v_{j_2}, \dots, v_{j_l}\}$, where v_{j_1} is simply v_i and v_{j_l} has a matching profile with the sought target profile \mathcal{A}_t if the search is successful. The goal of the mechanism is to find a user, say v_{j_l} such that the profile similarity between the target profile \mathcal{A}_t and the profile \mathcal{A}_{j_l} of user v_{j_l} is maximized (at least a value ξ), while the total payment from the initiator is no more than her budget, *i.e.*, $\max S(\mathcal{A}_{j_l}, \mathcal{A}_t)$, while $\sum_{v_{j_k} \in \mathbf{O}} \mathbf{p}^{j_k} \leq B$. When the search fails, we denote the output as ϕ .

Each agent’s preferences are given by a **valuation function**: $v^k(c^i, \mathbf{O}) = -c^k$ if $v_k \in \mathbf{O}$, and 0 if $v_k \notin \mathbf{O}$. The **utility function** of the agent v_k is

$$u^k = \mathbf{p}^k(d) + v^k(c^k, \mathbf{O}). \quad (3)$$

The utility function is the objective function each agent aims to optimize in the bidding phase.

In the winner decision phase, the decider needs to select one from all participants as the next link. Without causing ambiguity and to simplify the expression, here we use d^k as the normalized value of the declared price of v_k and s_{kt} as the normalized value of the profile similarity between v_k and v_t . The **score function** for the winner decision is defined as

$$\eta_k = f(d^k, s_{k,t}). \quad (4)$$

f could be any function that entails: (1) $\partial \eta_k / \partial d^k \geq 1$ and (2) $\partial \eta_k / \partial s_{k,t} < 0$. After extensive testing, we use the following function in this work

$$\eta_k = d^k + (1 - s_{k,t}), \quad (5)$$

which has been proved to yield a good performance in our experiments. Any intermediate user could choose the neighbor with the minimum η as the next link.

The **payment function** is $\mathbf{p}^k(d) = d^k + b^k(d)$ if $v_k \in \mathbf{O}$ and 0 otherwise. Here b^k is the bonus paid to user v_k . When an agent v_k is chosen as a vertex of the social referral path, she will get d^k as her compensation no matter the search is successful or not. The bonus b^k will be paid in the payment phase *only* when a feasible output is found. In the decision

phase, an agent v_k is selected because it has the smallest score η_k among all the neighbors of an intermediate user. Let η_x be the second smallest score in that stage, and we have $\eta_x \geq d^k + (1 - s_{kt})$. Then the value of **bonus** b^k is defined as:

$$b^k = \eta_x - \eta_k = \eta_x - (1 - s_{kt}) - d^k. \quad (6)$$

Once the target is found, the bonus will be paid to agents in the social referral path from the remainder of the budget. In other words, we need

$$\sum_{v_k \in \mathbf{O}} b^k(d) \leq B - \sum_{v_k \in \mathbf{O}} d^k. \quad (7)$$

Since the budget is limited, there's a chance that B is not enough to cover the bonus, *i.e.*, inequality (7) is violated. In this case, the bonus will be paid to the agents in the descending order of the similarity of their successor agents until the budget is used, *i.e.*, the agent who has selected a neighbor with a higher profile similarity will get paid bonus first. Using the bonus strategy we provide incentives for agents to maximize their utilities by declaring the true cost as well as choosing the next vertex with a higher similarity to the target. So an agent could maximize her utility as well as optimize the objective function. This claim will be proved formally later.

With the definition of payment and valuation function, the **utility** of the agent v_k is

$$u^k = d^k - c^k + b^k. \quad (8)$$

3.2 Mechanism Analysis

We assume that agents are all rational and selfish and each agent intends to maximize her own utility only. Then we have that our mechanism is truthful. Recall that a mechanism is truthful if for all v_k and all d^k , each agent's strategy is to declare her true cost, *i.e.* $d^i = c^i$ and truth-telling maximizes her utility.

Theorem 1: Our mechanism **M** is truthful.

Proof: Please refer to Appendix A in the supplementary file. \square

It is obvious that our mechanism also satisfies the **participation constraints**, that is whenever an agent is truth-telling, her utility is non-negative.

4 EXPERIMENTS AND MEASUREMENT

4.1 Dataset

Facebook is the most popular online social network with more than 1 billion users, which makes it a powerful resource for finding experts. Here we use the real facebook data of MIT and Harvard for our analysis and experiments, which contains attributes of students and faculty from different departments. The dataset is provided by the work [26], [27]. In this dataset the isolated vertices are ignored. In the MIT dataset, there are 6440 users and 502504 friendship edges among them. In this undirected graph, the mean node degree is 78.0286 and the median node degree is 56. The graph diameter is 8 and the average path length between

two vertices is 2.72. 99% vertices are reachable to each other within 6 hops. For comparison, we also use the real facebook data from Harvard, which contains 15000 users. We present more detail about the dataset in the Appendix B.

4.2 Acquaintance Probability Prediction Using Profiles

In the dataset, there are seven attributes in each user's profile, which are "student/faculty", "gender", "major", "second major/minor", "dorm/house", "high school" and "year". We learn the β parameter for the 7 attributes in the dataset by logistic regression. The detailed method is discussed in Section 2. More detail about the attributes and β parameter are presented in the Appendix B in the supplementary file

With the β parameter, given two users' profile vector, we can calculate their profile similarity and estimate their acquaintance probability. Figure 3 presents the real acquaintance probability in MIT and Harvard via statistical analysis and the acquaintance probability estimated by our model via profile vector. It shows a good match between the real probability and our estimation. So our similarity calculation is an effective metric for acquaintance relationship.

4.3 Basic Max Similarity Search Strategy

Although shorter paths exist between vertices, it will be quite difficult to find the shortest path using only local acquaintance knowledge of immediate neighbors. In this section, we analyze the performance of our similarity based local search strategy and compare it with other local search strategies and real small world experiments. The basic search strategy without paying intermediate users gives us a baseline search performance of the our expert finding mechanism.

4.3.1 Performance of Different Search Strategies

If each user does not incur a cost for participating in the social referral web, several search strategies could be used here: *random walk*, *high degree* [8], and *high similarity* strategies. In the random walk strategy, an intermediate user will select a successor from her neighbors randomly and simply avoid the users already participated in the chain of social referral. It was proved that the path length by random walk is $O(\ln^2(N))$ for a random network of N nodes. In the high degree strategy, an intermediate user selects the neighbor with the highest degree, who is more likely to know the target by virtue of the fact that she knows so many people. The third strategy, high similarity strategy, called **MaxSim** in this work, will select the neighboring user with the largest profile similarity (higher acquaintance probability) to the target.

We select 1000 pairs of source and target from the MIT dataset randomly, and run these three strategies to *locally* find the chain of social referral with/without path length constraint. Our extensive experiments show that the random walk strategies performs the worst. Thus, we just compare the results between high degree and high similarity

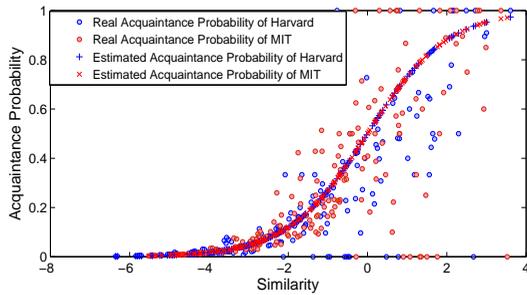


Fig. 3. Acquaintance probability change with similarity in MIT and Harvard Facebook datasets.

TABLE 1
Search results of high degree and similarity bias strategies.

	length constraint	mean	median	success rate
HighDegree	∞	126	12	99.8%
MaxSim	∞	22	6	95.2%
HighDegree	100	17	7	80.6%
MaxSim	100	11.6	6	93.1%

strategies. Figure 4 presents the path length distribution of paths and Table 1 presents the mean length, median length and successful rate of each strategy. Experiment results show that our high similarity local search strategy outperforms the high degree strategy in the mean length and median length greatly as we consider the attribute similarity to estimate the acquaintanceship. On the other hand, without path length constraint, the high degree strategy can achieve 99.8% success rate with some path that is more than 1000 hops long; our similarity bias local search strategy may have 4.8% failure, but has only one path with length exceeding 100 (length 114). We analyze the 48 failed searches (among 1000 search requests) and find that they all have targets with low node degrees, with a mean node degree 7.8. Meanwhile, they all have a very small profile similarity between the initiator and the target. If there is a path length constraint, e.g. 100, the successful rate of the high degree strategy is reduced to 80.6%, which is much lower than the successful rate of our high similarity strategy 93.1%.

Consider the result of the work in [7] of social network search. They used friendship network data from a community website, Club Nexus, with 2000 users, average node degree 8, and average shortest path 4. They compared 5 attributes simultaneously for the target and a user. In their experiments, the mean length is 135, and the median length is 28, which are much larger than our result, mostly because we learn the parameters locally to enable better acquaintanceship estimation.

4.3.2 Analysis of MaxSim Search Strategy

We compare the search result between MIT (6440 users) and Harvard (15000 users). The median path length with

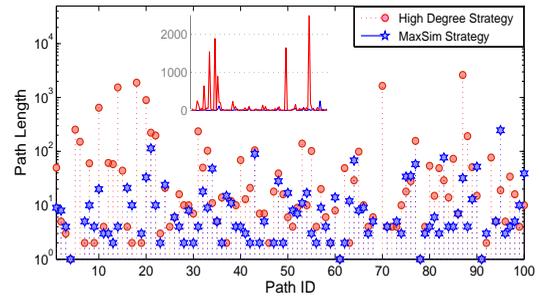


Fig. 4. Path lengths of different local search strategies for 100 pairs of randomly selected initiators and targets in MIT Facebook dataset. The small subfigure is the result with linear axis.

100 randomly selected pairs of initiator/targets in Harvard dataset is 5, the mean value is 46.7 without a length constraint. When there is a path length constraint (here we choose 100), the mean path length is 10.6 and the search success rate is 91%. As we have analyzed, the large mean path length is due to some extra long paths caused by some low node-degree targets and the low similarity between the initiator and target.

Figure 5 demonstrates the path length distribution of the 1000 found paths in the MIT dataset using our local search method. Since the average shortest path length between users in the network is 2.72, our experiment results show that our strategy is efficient to find almost shortest paths using local search with only local information.

Figure 6 presents the relationship between path length and acquaintance probability between two users. Generally, lower acquaintance probability will cause longer paths. However, according to Figure 5 and Figure 6, no matter how small the acquaintance probability is, short paths still exist and can be found using local search strategy.

We study the changes of the similarity with the target and the node degree along some random chosen social referral chains. We found that the node degree along the referral chain is irregular while the similarity increases. Please refer to Figure 17 (in Appendix C) for details. Further more, we analyze the similarity change along the true shortest path. We search the shortest paths of 10000 random selected pairs of source and target from the MIT dataset by Dijkstra algorithm. The result shows that 63% relays associate with increasing similarity to the target, only 18% relays associate with decreasing similarity. The analysis implies the reasonableness of our similarity bias strategy.

4.3.3 Performance with Real World Consideration

There are some existing shortest referral path search experiments in real world. We notice an interesting phenomenon here: the median path length of our search strategy is similar as the well-known small world experiments conducted by Travers and Milgram [28] and Dodds et. al. [9]. It implies that the online social network Facebook describes the acquaintance relationship of the real world

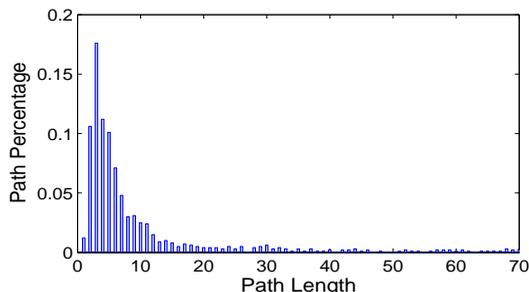


Fig. 5. The path length distribution for the 1000 paths experiment .

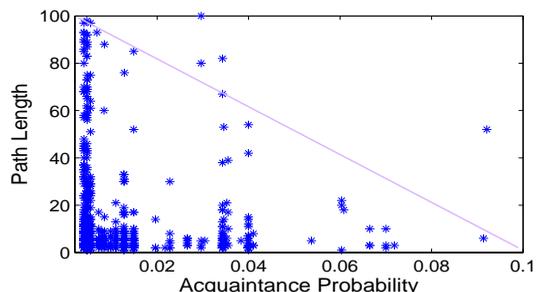


Fig. 6. The relationship between path length and acquaintance probability.

social network to some extent. However, the success rate of the small world experiments is very low.

In experiment conducted by Travers and Milgram [28], some people are able to construct short chains comprising an average of approximately five intermediaries, but 80% people in their experiments never reached the target. And in the most recently repeated small world experiment, conducted by Dodds et. al. [9] those completed chains were only 4 hops long, but only 0.4% of about 24,000 chains that started (a total of 384) reached their targets. Most paths were terminated because some participants drop the message instead of relaying messages. In Travers and Milgram’s experiment, chains were observed to terminate with probability 0.25. We therefore conduct a more realistic search experiment with an attrition rate at each step of the chain construction. The attrition rate indicates the probability a participant terminates the chain. We randomly select 1000 pairs of source and target. Figure 7 and Figure 8 illustrate the successful rate and chain length of the high degree search strategy and our high similarity search strategy. The increase of attrition rate courses great decrease of success rate. When the attrition rate is 0.25 which are consistent with Milgram’s experiment, the success rate is only 17% for high degree strategy, while it is 25% for our high similarity strategy which is 5% higher than the Milgram’s result. Moreover, a much shorter mean length is achieved with an attrition rate. The mean path length is 3.5 when attrition rate is 25% compared to 24.6 without attrition and the median length also decreases to 3 which approaches to the true average shortest path length 2.72. The existing work and our experiments show that, the main course of the failure is that the participants are not sufficiently motivated. As a result, a payment mechanism is quite necessary for a successful search.

To be more realistic, we consider the situation that an urgent request is demanded while some users are not online. The result shows that when the offline users are less than 70%, the success rate and path length don’t deteriorate significantly. Please refer to Figure 18 and Figure 19 (in Appendix D) for details.

4.4 Search with a Payment Mechanism

In this section, we study the performance of our truthful mechanisms with payment. We consider two distributions

of the real cost c^k : the uniform distribution in the range $[0, 100]$; the normal distribution with mean $\mu = 50$ and variance $\sigma = 10$. We compare the budgets with different cost distributions of three search mechanisms:

- 1) **Mechanism MaxSim**: In the decision phase of each intermediate user, a neighbor with the maximum similarity to the target will be selected without considering the declared price, and the mechanism will pay the user her declared price. Observe that this mechanism is not truthful: every user could declare a price as high as she likes. We estimate the lowest budget of this untruthful mechanism by assuming that every user declares the true cost.
- 2) **Mechanism SumScore**: This is our search mechanism presented in Section 3. The score function is $\eta^k = d^k + (1 - s_{k,t})$
- 3) **Mechanism ProdScore**: This mechanism is similar to the mechanism presented in Section 3, except that the score function is: $\eta^k = d^k \times (1 - s_{k,t})$. It is easy to show that this mechanism is also truthful, using techniques similar to that in Section 3.

4.4.1 Search Performance

We study the performance of these three different strategies by randomly selecting 100 pairs of initiators and targets from the MIT dataset. Table 2 summarizes the measurement of these three different mechanisms. Mechanism **MaxSim** is cost independent: it just tries to maximize the similarity. Mechanism **MaxSim** outperforms the other two in search. Mechanism **SumScore** performs a little worse than Mechanism **MaxSim**. Mechanism **ProdScore** performs the worst because its score function prefers nodes with smaller cost more than good similarity.

4.4.2 Budget Requirement

We investigate the cost and budget requirement with uniform distributed cost. Figure 9 and Figure 10 present the required budget of three mechanisms, *i.e.*, the minimum payment needed for having a successful search.

We then study the cost and budget requirement with normal distributed cost. Figure 11 and Figure 12 present the required budget.

Table 3 shows the average cost per hop in the social referral chain found by mechanisms **SumScore** and **ProdScore**.

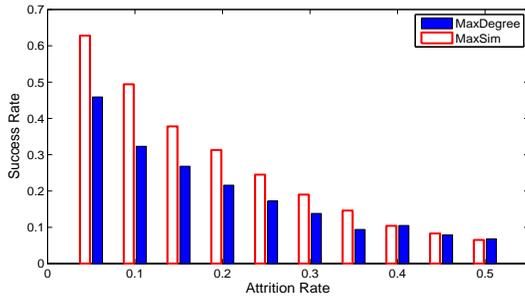


Fig. 7. The successful rate of MaxSim and High Degree search of 1000 random request with different attrition rate.

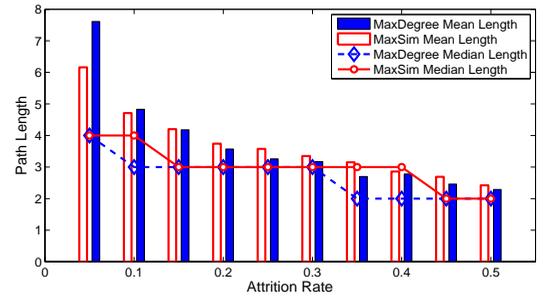


Fig. 8. The path length of MaxSim and High Degree search of 1000 random request with different attrition rate.

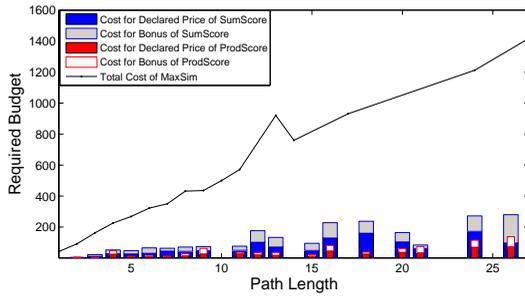


Fig. 9. The required budget v.s. different path length for three mechanisms with uniform cost distribution.

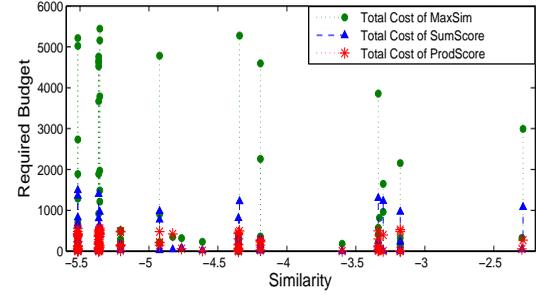


Fig. 10. The required budget v.s. different similarity for three mechanisms with uniform cost distribution.

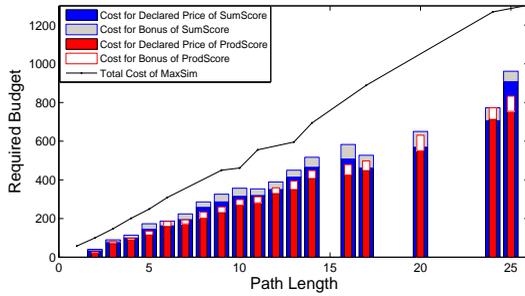


Fig. 11. The required budget v.s. path length for three mechanisms with normal cost distribution.

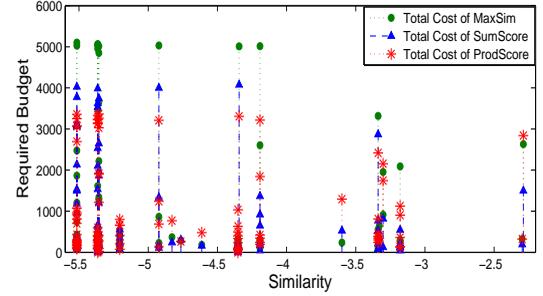


Fig. 12. The required budget v.s. different similarity for three mechanisms with normal cost distribution.

TABLE 2

Search performance of 3 mechanisms.

Uniform distribution of costs			
mechanism	mean length	median length	success rate
MaxSim	11.4	5	89%
SumScore	14.3	6	82%
ProdScore	26.5	19.5	68%
Normal distribution of costs			
mechanism	mean length	median length	success rate
MaxSim	11.4	5	89%
SumScore	15.1	6	87%
ProdScore	18.9	12	83%

TABLE 3

Average cost per-hop for mechanisms **SumScore** and **ProdScore**.

Uniform distribution of costs			
mechanism	declaration	bonus	total
SumScore	6.1	3.2	9.3
ProdScore	2.8	1.6	4.4
Normal distribution of costs			
mechanism	declaration	bonus	total
SumScore	32.3	3	35.3
ProdScore	28.9	2.5	31.4

We found that mechanism **ProdScore** generates a lower average cost at the cost of losing some search efficiency. Since Mechanism **MaxSim** is cost independent, the lower bound of its average per-hop cost is the mean value 50 with both uniform and normal cost distributions. In fact,

for any other cost-independent/untruthful mechanisms, e.g. high degree mechanism, the average per-hop cost is also at least 50.

The mechanism **MaxSim** has the best search performance, but requires a much larger budget, which could be potentially unbounded when users lie about their real cost.

Mechanism **SumScore** creates shorter paths with higher cost, while mechanism **ProdScore** achieves small average cost per-hop, but often found longer paths. Both mechanisms **SumScore** and **ProdScore** can find cheap and short paths to the target and produce similar overall path cost. Mechanism **SumScore** has a higher success rate (i.e. waste less pay) and satisfies our optimization objective better. By Figure 9, we notice that the budget isn't proportional to the product of mean price and path length, which implies that if only there exist cheap short paths, our mechanism can find them. We also notice that the bonus will not incur significant extra pay, especially when the users' costs follow normal distribution. Small bonus used in our mechanism does motivate users to declare their true cost.

5 RELATED WORK

Given a description of a desired expert, there are two kinds of expert finding methods without using payment: (1) global expert finding: finding a person or a group of ranked persons, similar enough to the desired one with global information; (2) distributed expert finding: searching a desired person via a chain of social referrals using local acquaintance information in a social network, e.g. the *small world routing*.

Global Expert Finding: This is usually based on a large database with global information and mainly focuses on linking humans to expertise areas (known as expertise retrieval). There are already some well-known large-scale online expert search systems like Spock and zoominfo. Some commercial expert search engines have been invested in companies like IBM and Microsoft. Some organization expert finding systems like [22] have studied knowledge management to utilize human knowledge within an organization as well as possible. Most research in this area focus on tracing, mining and organizing evidences of expertise, and linking the evidences to humans. Serdyukov *et al.* [23] designed schemes for expertise evidence finding outside a well organized database. They showed the importance of global web sources of expertise evidence for expert finding. Recently, social networks are also utilized to find professionals. Zhang *et al.* [34] proposed an approach to find the expertise by considering both personal information and relationship information with other experts. Lappas *et al.* [17] proposed to find a team of experts by considering the team formation problem in social networks. Balog *et al.* have done a lot of work about expertise retrieval [35], [36] and provided a good overview of expertise retrieval in [1]. All these schemes fall into the category of **global expert finding** by using some global databases or central server and focus on *expertise evidence mining and organization*. Social networks play a role of furnishing expertise evidence through providing the friendships between persons.

There are also some work [2]–[6] use social feature to improve information routing and dissemination in social networks.

In this work, we focus on searching people with matching attributes in online social network in a distributed manner

with only local information, which is also known as the small world routing problem. Finding evidence of expertise is not in the scope of this work.

Distributed Expert Finding (i.e., small world routing):

The small world problem has been popular since forty years ago [19], which conducted experiments to prove that everyone is connected to everyone else via six degrees of separation. In [30], the small world problem was divided into topological problem and algorithmic problem. The topological problem is that for a randomly chosen pair of individuals, there exists, with a high probability, a “short” chain of intermediaries that connect them. Here short is usually interpreted as proportional to the logarithm of the population size. The algorithmic small world problem is that in order to locate a service provider, a person must actively traverse some chains of referrals. An individual need effectively navigate these short chains themselves, with every intermediate individual having only local knowledge (e.g., her own friends) of the social network.

Many research results were devoted to study various challenging algorithmic small world problems, e.g., [7], [8], [15], [16], [25], [32]. The base of these algorithms is selecting a successor by some greedy methods that optimize some metrics (e.g., degree, geography, social distance and profile similarity). Adamic *et al.* [8] introduced several local search strategies which utilize high degree nodes in power-law graphs. Watts *et al.* [32] used the lowest common ancestor level in the social hierarchy as the similarity measurement between individuals. Adamic *et al.* [7] did some simulated experiments on the email network and verified the models proposed by Watts *et al.* [32] and Kleinberg [16]. The email experiments explain why individuals are able to successfully complete chains in small world experiments using only local information. They also pointed out that in an online social network, where the data is incomplete and hierarchical structures are not well defined, local search strategies are less effective. The difficulty of the local search is also mentioned by [24], which states that the search distance in social networks is fundamentally different from topological distance, where the mean and median topological distances tend to be similar. Pairs of individuals who are already “close”, in the sense of sharing social, geographical, and demographic attributes, can find each other, but distant pairs cannot.

These efforts focus on designing and analyzing local search algorithms to find short paths, with a hypothesis that people will participate voluntarily. It has been well documented that in the small world experiments a vast majority of paths never reach their ultimate targets [24]: most paths were terminated because participants are *not sufficiently motivated* to relay messages. Thus, we need design a truthful mechanism that takes into account not only the participation cost of intermediate users, but also the selfish nature of these intermediate users.

6 DISCUSSION AND CONCLUSION

In this work, we addressed the local social referral problem in a large scale social network by taking users' self-interests

into consideration and designed a truthful mechanism that reduces the length of social referral chain, reduces the cost of social referrals, and improves the success rate, compared with previous efforts. There are several interesting questions not fully studied in this work. For example, we are designing some privacy mechanisms to protect the desired expert of the initiator and the attributes of all the participants. *e.g.* [39]–[43]. We need to design a truthful mechanism when the initiator wants to find a target expert within certain hop distance, while tolerating the possibility of multiple experts were found by multiple social referral paths. We also need to design the mechanism to find a group of required experts who are well related. It is still challenging to estimate the path length and budget for the initiator.

REFERENCES

- [1] KRISZTIAN BALOG, YI FANG, MAARTEN DE RIJKE, PAVEL SERDYUKOV, AND LUO SI. Expertise Retrieval. In *Foundations and Trends in Information Retrieval*, 2012.
- [2] EYUPHAN BULUT, BOLESŁAW K. SZYMANSKI. Exploiting Friendship Relations for Efficient Routing in Mobile Social Networks. In *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [3] YUNSHENG WANG, WEI-SHIH YANG, JIE WU. Analysis of a Hypercube-based Social Feature Multi-Path Routing in Delay Tolerant Networks. In *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [4] SHENG WEN, WEI ZHOU, JUN ZHANG, YANG XIANG, WANLEI ZHOU, WEIJIA JIA. Modeling Propagation Dynamics of Social Network Worms. In *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [5] JIALU FAN, JIMING CHEN, YUAN DU, WEI GAO, JIE WU, YOUXIAN SUN. Geo-Community-Based Broadcasting for Data Dissemination in Mobile Social Networks. In *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [6] SHAO-JIE TANG, JIN YUAN, XUFEI MAO, XIANG-YANG LI, WEI CHEN, AND GUOJUN DAI. Relationship Classification in Large Scale Online Social Networks and Its Impact on Information. In *IEEE INFOCOM*, 2011.
- [7] ADAMIC, L., AND ADAR, E. How to search a social network. In *Social Networks*, 2005.
- [8] ADAMIC, L. A., M. LUKOSE, R., PUNIYANI, A. R., AND HUBERMAN, B. A. Search in power-law networks. In *Physical review*, 2001.
- [9] DODDS, P., MUHAMAD, R., AND WATTS, D. An experimental study of search in global social networks. In *Science*, 2003.
- [10] EVANS, B., KAIRAM, S., AND PIROLI, P. Do your friends make you smarter?: An analysis of social strategies in online information seeking. In *Information Processing & Management*, 2009.
- [11] FIELDS, L. 3 great databases for finding experts.
- [12] FITZPATRICK, B., AND RECORDON, D. Socialgraph, 2007. <http://bradfitz.com/social-graph-problem>.
- [13] GUARE, J. Six degrees of separation. In *Dramatists Play Service*, 1992.
- [14] KAUTZ, H., SELMAN, B., AND SHAH, M. Referral web: combining social networks and collaborative filtering. In *Communications of the ACM*, 1997.
- [15] KLEINBERG, J. Navigation in a small world. In *Nature*, 2000.
- [16] KLEINBERG, J. The small-world phenomenon: An algorithmic perspective. In *ACM symposium on theory of computing*, 2000.
- [17] LAPPAS, T., LIU, K., AND TERZI, E. Finding a team of experts in social networks. In *ACM SIGKDD*, 2009.
- [18] McDONALD, D. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003.
- [19] MILGRAM, S. The small world problem. In *Psychology today*, 1967.
- [20] MILLER MCPHERSON, LYNN SMITH-LOVIN, J. M. C. Birds of a feather: Homophily in social networks. In *Annual Review of Sociology*, 2001.
- [21] NEWMAN, M. E. J. Ego-centered networks and the ripple effect. In *Social Networks*, 2003.
- [22] REICHLING, T., VEITH, M., AND WULF, V. Expert recommender: Designing for a network organization. In *Computer Supported Cooperative Work*, 2007.
- [23] SERDYUKOV, P., AND HIEMSTRA, D. Being omnipresent to be almighty: The importance of the global web evidence for organizational expert finding. In *Proceedings of the SIGIR Workshop on Future Challenges in Expertise Retrieval*, 2008.
- [24] SHARAD GOEL, ROBY MUHAMAD, D. W. Social search in "small-world" experiments, In *WWW*, 2009.
- [25] SIMSEK, O., ET AL. Decentralized search in networks using homophily and degree disparity. In *Proceedings of the 19th international joint conference on Artificial intelligence*, 2005.
- [26] TRAUD, A. L., KELSIC, E. D., MUCHA, P. J., AND PORTER, M. A. Comparing community structure to characteristics in online collegiate social networks. In *SIAM Review*, 2010.
- [27] TRAUD, A. L., MUCHA, P. J., AND PORTER, M. A. Social structure of facebook networks. arXiv:1102.2166, 2011.
- [28] TRAVERS, J., AND MILGRAM, S. An experimental study of the small world problem. In *Sociometry*, 1969.
- [29] WANG, W., AND LI, X. Low-cost routing in selfish and rational wireless ad hoc networks. In *IEEE TMC*, 2006.
- [30] WATTS, D., AND STROGATZ, S. Collective dynamics of small-world networks. In *Nature*, 1998.
- [31] WATTS, D., AND STROGATZ, S. Small world. In *Nature*, 1998.
- [32] WATTS, D. J., DODDS, P. S., AND NEWMAN, M. E. J. Identity and search in social networks. In *Science*, 2002.
- [33] WHITE, H. Search parameters for the small world problem. In *Social Forces*, 1970.
- [34] ZHANG, J., TANG, J., AND LI, J. Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications*, 2010.
- [35] K. BALOG AND M. DE RIJKE. Combining candidate and document models for expert search. In *TREC*, 2008.
- [36] K. BALOG, M. DE RIJKE, AND W. WEERKAMP. Bloggers as experts: Feed distillation using expert retrieval models. In *ACM SIGIR*, 2008.
- [37] JORDAN, M.I. AND OTHERS. Why the logistic function? A tutorial discussion on probabilities and neural networks. In *Computational Cognitive Science Technical Report*, 1995.
- [38] BERNARDO, J.M. AND SMITH, A.F.M. Bayesian theory. In *Measurement Science and Technology*, 2001.
- [39] ZHANG, L. AND LI, X.Y. AND LIU, Y.H. AND JUNG, T. Verifiable Private Multi-party Computation: Ranging and Ranking. In *IEEE INFOCOM*, 2013.
- [40] JUNG, T. AND MAO, X.F. AND LI, X.Y. AND TANG, S.J. AND GONG, W. AND ZHANG, L. Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation. In *IEEE INFOCOM*, 2013.
- [41] LI, X.Y. AND JUNG, T. Search me if you can: privacy-preserving location query service. In *IEEE INFOCOM*, 2013.
- [42] JUNG, T. AND LI, X.Y. AND WAN, Z. AND WAN, M. Privacy preserving cloud data access with multi-authorities. In *IEEE INFOCOM*, 2013.
- [43] ZHANG, L. AND LI, X.Y. AND LIU, Y.H. Message in a Sealed Bottle: Privacy Preserving Friending in Social Networks. In *IEEE ICDCS*, 2013.
- [44] ZHANG, L. AND LI, X.Y. AND LIU, Y.H. AND HUANG, Q.Y. AND TANG, S.J. Mechanism Design for Finding Experts Using Locally Constructed Social Referral Web. In *IEEE INFOCOM*, 2012.



Lan Zhang is a Ph.D. student in the department of Computer Science and Technology, Tsinghua University, Beijing, China. She received the Bachelor degree (2007) in School of Software at Tsinghua University. Her research interests span social networks, privacy, secure multi-party computation and mobile computing, etc.



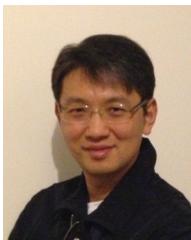
Xiang-Yang Li is a professor at the Illinois Institute of Technology. He is recipient of China NSF Outstanding Overseas Young Researcher (B). Dr. Li received MS (2000) and PhD (2001) degree at Department of Computer Science from University of Illinois at Urbana-Champaign, a Bachelor degree at Department of Computer Science and a Bachelor degree at Department of Business Management from Tsinghua University, P.R. China, both in 1995. He published a mono-

graph "Wireless Ad Hoc and Sensor Networks: Theory and Applications". He co-edited several books, including, "Encyclopedia of Algorithms". His research interests include mobile computing, cyber physical systems, wireless networks, security and privacy, and algorithms. Dr. Li is an editor of several journals, including IEEE Transaction on Parallel and Distributed Systems, IEEE Transaction on Mobile Computing. He is a senior member of IEEE and a member of ACM.



Jingsheng Lei is a professor and the Dean of the College of computer science and technology, Shanghai University of Electronic Power. He received his B.S. in Mathematics from Shanxi Normal University in 1987, and MS and Ph.D in Computer Science from Xinjiang University in 2000 and 2003 respectively. He has wide research interests, mainly including machine learning, data mining, pattern recognition and Cloud computing. In these areas he has published more than 80

papers in international journals or conferences. He serves as an Editor-in-Chief of the Journal of Computational Information Systems. He is the Members of the Artificial Intelligence and Pattern Recognition Technical Committee of the CCF (China Computer Federation), Members of the Machine Learning Technnical Committee of the CAAI (Chinese Association of Artificial Intelligence), Members of Academic Committee of ACM Shanghai Chapter.



Yunhao Liu received the BS degree in automation from Tsinghua University, China, in 1995, the MS and PhD degrees in computer science and engineering from Michigan State University, in 2003 and 2004, respectively. He is now Cheung Kong Professor at Tsinghua University, as well as a faculty member with the Hong Kong University of Science and Technology. His research interests include wireless sensor network, peer-to-peer computing, and pervasive computing.

APPENDIX A PROOF OF THEOREM 1

Proof: We will denote $\langle d^1, \dots, d^{k-1}, d^{k+1}, \dots, d^n \rangle$ by d^{-k} . We use c^k as the normalized value of the true cost of v_k and accordingly all the other currency values are normalized too. So the true score of v_k is $\eta_k = c^k + (1 - s_{kt})$.

When an agent declares her true cost $d^k = c^k$, her maximum utility is

$$\max u^k = \begin{cases} 0 & v_k \notin \mathbf{O}(c^k, d^{-k}), \\ b^k & v_k \in \mathbf{O}(c^k, d^{-k}), \end{cases}$$

$$\text{bonus: } b^k = \eta_x - \eta_k \geq 0. \quad (\text{A.1})$$

When η_k is the smallest score, η_x is the second smallest score among v_k 's competitors. So we have $\eta_x > \eta_k$, and given d^{-k} , the utility of v_k is determined as the true bonus b^k .

If the agent intends to lie by declaring a price $d^{k'} = c^k + \delta$, and the fake score is

$$\eta_k' = c^k + \delta + (1 - s_{kt}). \quad (\text{A.2})$$

Now we will prove that no matter $\delta > 0$ or $\delta < 0$, her utility can not be improved. Note that, an agent cannot lie about her similarity with the target profile because her profile is open to her friends in a social network. Thus the similarity score s_{kt} is always the true value.

First we analyze the case $\delta > 0$, i.e. the agent intends to declare a price higher than her true cost to increase her compensation. In this case, if $v_k \notin \mathbf{O}(c^k, d^{-k})$, i.e. η_k is not the smallest score, then η_k' is still not the smallest value, and v_k still will not be selected. When $v_k \in \mathbf{O}(c^k, d^{-k})$, i.e. η_k is the smallest score, v_k will be selected if she tells truth. But when she lies about the cost, if $\delta > \eta_x - \eta_k$, η_x will be the smallest score and v_k will not be selected and thus lose everything. If $\delta \leq \eta_x - \eta_k$, her utility is

$$u^{k'} = d^{k'} - c^k + b^{k'} \\ = c^k + \delta - c^k + \eta_x - (1 - s_{kt}) - (c^k + \delta), \quad (\text{A.3})$$

which is still b^k . So, in the case $\delta > 0$ her max utility is:

$$\max u^{k'} = \begin{cases} 0 & v_k \notin \mathbf{O}(c^k, d^{-k}), \\ b^k & v_k \in \mathbf{O}(c^k, d^{-k}) \text{ and } \delta \leq \eta_x - \eta_k, \\ 0 & v_k \in \mathbf{O}(c^k, d^{-k}) \text{ and } \delta > \eta_x - \eta_k. \end{cases}$$

Then we analyze the case $\delta < 0$, i.e. the agent intends to declare a price lower than her true cost to win the bidding. It's similar to the situation above, except that when η^k is not the smallest score and η_y is the true smallest score, if $\delta \leq \eta^y - \eta^k$, the utility of v_k is $u^{k'} = \eta^y - \eta^k < 0$, i.e. agent v_k will get a negative utility.

Consequently, in our mechanism no matter what price v_k declares, if η_k is already the smallest score, the maximum utility will be b^k . When she declares a price higher than c^k , there is a higher risk to fail the competition and get nothing; If η_k is not the smallest score, the maximum utility of v_k will be 0. Besides when she declares a price lower than c^k , there is a higher risk to get a negative utility. So, an agent's

best strategy to maximize her utility is to tell the truth, i.e. $d^k = c^k$. \square

APPENDIX B DATASET ANALYSIS

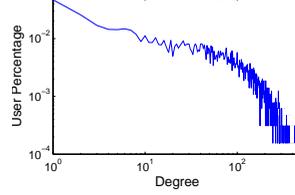


Fig. 13. The facebook users' degree distribution in MIT.

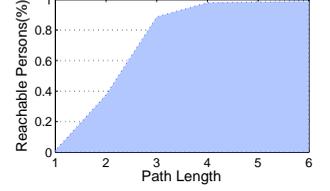


Fig. 14. Reachable users v.s. the path length in MIT dataset.

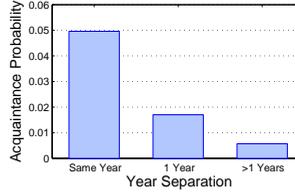


Fig. 15. Acquaintance probability with different year separation.

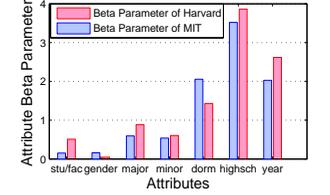


Fig. 16. β parameters for each attribute learned from MIT and Harvard Facebook Data.

In the MIT dataset, there are 6440 users and 502504 friendship edges among them. In this undirected graph, the mean node degree is 78.0286 and the median node degree is 56. The node degree distribution is illustrated in Figure 13 which follows a power-law distribution. A few nodes have very high degree, and many nodes have low degree. The graph diameter is 8 and the average path length between two vertices is 2.72. As shown in Figure 14, 90% vertices are reachable to each other within 3 hops and 99% vertices are reachable to each other within 6 hops.

In the dataset, there are seven attributes in each user's profile. In our experiment, for the categorical attributes "student/faculty", "gender", "major", "second major/minor", "dorm/house", "high school", the similarity is 1 if they have the same value, the similarity is 0 if they have different values. For the numerical attribute "year", the acquaintance probability is quite small when two users are more than one year apart as shown in Figure 15. Therefore we consider three conditions: the same year, a year apart, and more than a year apart. Accordingly the attribute similarities of "year" are 1, 0.5 and 0.

We learn the β parameter for the 7 attributes in a dataset by logistic regression. The detailed method is discussed in Section 2. Figure 16 illustrates the learned β parameters of MIT and Harvard Facebook networks. As we see, different attributes contribute unequally to the acquaintance relationship and the parameters depend on networks. The β parameters of MIT we used in the experiments is summarized in Table 4.

TABLE 4
 β parameters learned using global network.

β_0	-5.5157	status	0.1509
gender	0.1592	major	0.5927
minor	0.5368	dorm	2.0545
high school	3.5201	year	2.0267

APPENDIX C SIMILARITY AND DEGREE CHANGES ALONG REFERRAL CHAINS

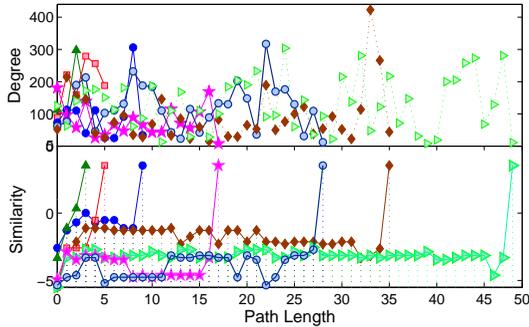


Fig. 17. The similarity and degree changes along some randomly selected paths.

We study the changes of the similarity with the target, and the node degree along 7 randomly chosen social referral chains from the search result of MaxSim strategy and high degree strategy. As shown in Figure 17, we found that the node degree along the chain is irregular while the similarity increases. It is interesting to find that except those very short paths, there are some typical plateaus along the similarity line. The plateau may be due to the procedure of searching the target will remain in some social community (with similar profiles), until the social referral finds someone who has a shortcut to the target. The number of the plateaus implies the number of different social communities crossed by the social referral path.

APPENDIX D PERFORMANCE WITH DIFFERENT ONLINE USER RATIOS

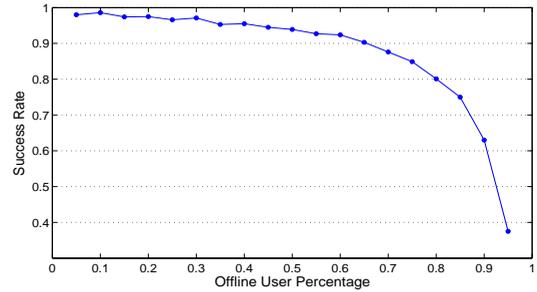


Fig. 18. The successful rate of MaxSim of 1000 random request with different online user percentage.

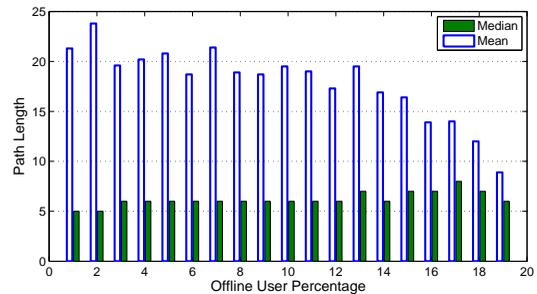


Fig. 19. The path length of MaxSim of 1000 random request with different online user percentage.

To evaluate the performance of our search mechanism in practice, we consider a more realistic situation that an urgent request is demanded while some users are not online. Figure 18 and Figure 19 shows that when the offline users are less than 70%, the success rate and path length don't deteriorate significantly.