

Mechanism Design for Finding Experts Using Locally Constructed Social Referral Web

Lan Zhang^{*}, Xiang-Yang Li^{*,‡}, Yunhao Liu^{*,†}, QiuYuan Huang[§], Shaojie Tang[‡]

^{*} MOE Key Lab for Information System Security, School of Software,

Tsinghua National Lab for Information Science and Technology, Tsinghua University

[†] Department of Computer Science and Engineering, HKUST

[‡] Department of Computer Science, Illinois Institute of Technology

[§] Department of Electrical Engineering, University of Florida

Abstract—In this work, we address the problem of finding experts 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 search performance, 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 method and only about 2% of 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, Referral Web, Localized Searching, Small World.

I. INTRODUCTION

Finding experts is common and useful in real life. So far, there are two kinds of expert finding in industry and literature. One is searching in a large database with the global information. The other strategy is 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. 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 [3]. E.g., many users of online social networks have made their profiles unsearchable to public or only visible to friends. 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 called initiator) to the expert.

Online social networks, like Facebook and LinkedIn, are powerful resources for finding experts and constructing referral chains due to the abundant personal information and the homophily principle [6]. A lot of existing work have proved that social networks are searchable through short pairwise

connections [15], [16]. For online social networks (OSN), our analysis of Facebook dataset shows about 99% users are within 6 hops in average.

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 [8]. It has been reported in [17] 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. So our research focuses on similarity calculation and truthful local search mechanism design, while considering users' profit.

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 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 95.2%, which is significantly better than the best reported success rate 20% [2], [12].

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

II. SYSTEM MODEL AND PRELIMINARIES

A. Problem Formulation Using Online Social Networks

A social network is modeled by a graph $G = (V, E)$. By profiling or data collection, each user $v_i \in V$ 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. User v_i is called a neighbor of v_j in the social network G . 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 $\mathcal{A}_t = \langle a_t^1, a_t^2, \dots, a_t^m \rangle$. The initiator will ask her neighbors to help her to find a matching expert. The process will be iterated till 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 [13]. However, in the practical social network setting, following challenges need to be addressed to solve this problem:

- 1) computing the profile similarity and acquaintance probability,
- 2) designing efficient referral strategy using only local information under the budget constraint, and
- 3) designing a payment mechanism to make participants tell truth without any global information for lies checking.

B. 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 acquaintant to each other or have a shorter network distance* [6]. 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 [16], or geographical distance. In this work, we use users' profiles to estimate the probability of acquaintance.

Distance and similarity of attributes: The attribute could be discrete categorical characteristics such as social group or gen-

der, 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 $s_{ij}^k = 1 - \frac{d_{ij}^k}{\max_{(i,j)} d_{ij}^k}$. The **similarity vector** S_{ij} between users v_i and v_j is defined as $S_{ij} = \langle s_{ij}^1, s_{ij}^2, \dots, s_{ij}^m \rangle$.

The similarity of different attributes contribute differently to acquaintance probability [9]. Some attributes are strong evidences of acquaintance and some are weak. We model the acquaintance probability of two users using the logistic function about their profile similarity vector. Specifically, if $E_{i,j}$ is the event that the user v_i and v_j are acquaint, define the probability

$$\Pr(E_{i,j} | S_{ij} = X) = \text{logit}^{-1}(\beta X) = \frac{1}{1 + e^{-\beta X}}. \quad (1)$$

Here the function $\text{logit}^{-1}(y) = 1/(1 + e^{-y})$ and 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

$$r(X) = \frac{\Pr(E_{i,j} | S_{ij} = X)}{1 - \Pr(E_{i,j} | S_{ij} = X)} \quad (2)$$

Then for a vector $X = (x^1, x^2, \dots, x^m)$, we have

$$\beta \cdot X = \beta^0 + \beta^1 x^1 + \dots + \beta^m x^m = \ln(r(X)) \quad (3)$$

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. 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, the higher the probability they are acquaintances. Note that here the profile similarity could be negative values.

A simple computation from Bayes theory 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 derive the parameter β by learning $r(X)$ from the known part of a graph. With the parameter β , we can predict the acquaintance probability in an unknown graph given the profile vectors of two users.

In this work 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, in which the node set V_i includes herself and all her direct friends, and the edge set E_i consists of all the edges between vertices in V_i . Then $r(X)$ can be learned from the ego network. We compare the acquaintance probability estimated by globally learned β parameter and β parameter learned from ego networks. The result shows that the ego networks provide sufficient evidence on acquaintance.

III. STRATEGYPROOF MECHANISM DESIGN

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

A. Search Procedure

Before we present our mechanism, we first give an overview of the five phases during our expert search procedure:

- 1) **Initialization phase:** The initiator v_0 will stop the search if the target is among her neighbors.
- 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 the social referral will declare a price d^k , via a **sealed-bid** within a bounded bidding time window.
- 3) **Winner decision phase:** User v_0 chooses a winner neighbor, say v_w , as the next-hop vertex based on a score function 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 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.

B. Algorithmic Mechanism Design

We refer every selfish rational participant as an agent, who intends to maximize her own utility only. For a searching task, each agent v_k has a **private** true cost c^k to participate in the referrals. 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.*, $\max s_{ij}$, 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

$$u^k = \mathbf{p}^k(d) + v^k(c^k, \mathbf{O}). \quad (4)$$

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 $s_{k,t}$ as the normalized value of the profile similarity between v_k and v_t . The **score function** for winner decision is defined as

$$\eta_k = f(d^k, s_{k,t}). \quad (5)$$

f could be any function that entails: (1) $\partial \eta_k / \partial d^k \geq 1$ and (2) $\partial \eta_k / \partial s_{k,t} < 0$. Any intermediate user could choose the neighbor with the minimum η as the next link in the decision phase.

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 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 $\eta_x \geq \eta_k$. Then the value of **bonus** b^k is defined as:

$$b^k = \eta_x - \eta_k \quad (6)$$

Once the target is found, the bonus will be paid to agents in the social referral path from the remainder of the budget. Since the budget is limited, there's a chance that B is not enough to cover the bonus. 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. 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.

Theorem 1: Our mechanism \mathbf{M} is truthful. 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.

Omitted.

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

IV. EXPERIMENTS AND MEASUREMENT

A. Dataset

Here we use the real facebook data in MIT for our analysis and experiments [10], [11]. In this dataset the isolated vertices are ignored. There are 6440 users and 502504 friendship edges. In this undirected graph, the mean node degree is 78.0286 and the median node degree is 56. The node degree follows a power-law distribution. 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.

B. Acquaintance Probability Prediction Using Profiles

We learn the β parameter for the 7 attributes in a dataset by logistic regression. The detailed method is discussed in Section II. With the β parameter, given two users' profile vector, we can calculate their profile similarity and estimate their acquaintance probability. Figure 1 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.

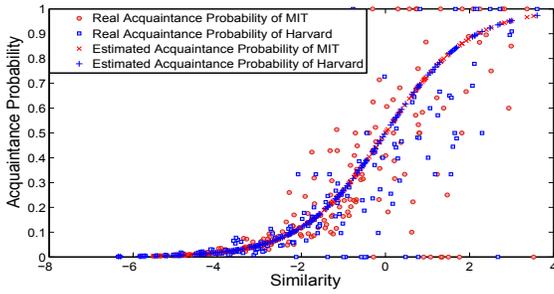


Fig. 1. Acquaintance probability change with similarity in MIT and Harvard Facebook datasets.

TABLE I

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%

C. Basic Max Similarity Search Strategy

The basic search strategy without paying intermediate users gives us a baseline search performance of the our expert finding mechanism.

If each user does not incur a cost for participating the social referral web, several search strategies could be used here: *random walk*, *high degree* [1], and *high similarity* (called **MaxSim**) strategies in this work.

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 strategies. Figure 2 presents the path length distribution of search result and Table I 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. If there is a path length constraint, e.g. 100, the success rate of the high degree strategy is reduced to 80.6%, while the success rate of our high similarity strategy remains 93.1%.

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.

D. Search with a Payment Mechanism

In this section, we study the search performance and budget requirement (*i.e.*, the minimum payment needed for having a successful search) 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

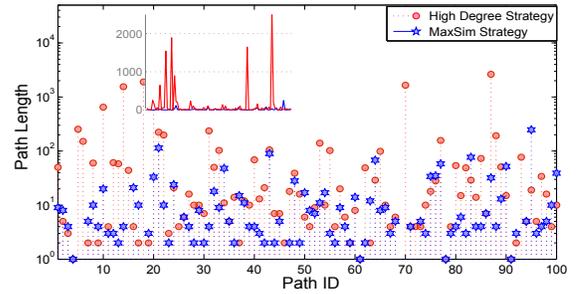


Fig. 2. Path lengths of different local search strategies for 100 pairs of randomly selected initiators and targets in MIT Facebook dataset. The small subfigure is the result with linear axis.

TABLE II

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 III

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

mean $\mu = 50$ and variance $\sigma = 10$. We compare the budgets with different cost distributions of one untruthful and two basic truthful search mechanisms:

- 1) **Mechanism MaxSim**: 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. We estimate the lowest budget of this untruthful mechanism by assuming that every user declares the true cost.
- 2) **Mechanism SumScore**: This is a truthful search mechanism as presented in Section III. The score function is $\eta^k = d^k + (1 - s_{k,t})$, which satisfies requirements for score function.
- 3) **Mechanism ProdScore**: This mechanism is similar to the mechanism SumScore, except that the score function is: $\eta^k = d^k \times (1 - s_{k,t})$. This mechanism is also truthful.

We randomly select 100 pairs of initiators and targets from the MIT dataset. Table II summarizes the search performance and Table III plots the average cost per hop in the social referral chain. Figure 3 and Figure 4 presents the required

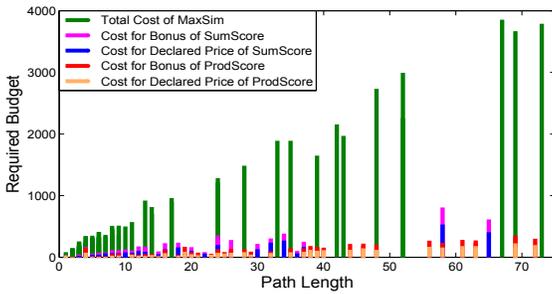


Fig. 3. The required budget v.s. different path length for three mechanisms with uniform cost distribution.

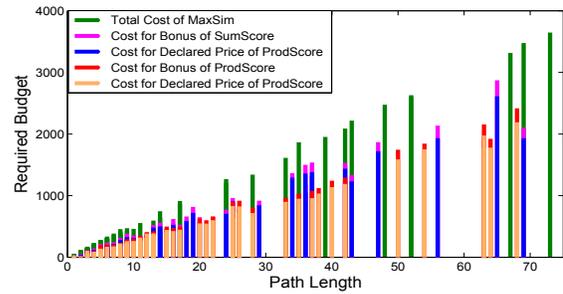


Fig. 4. The required budget v.s. path length for three mechanisms with normal cost distribution.

budget changing with the referral chain length.

Since Mechanism **MaxSim** is cost independent, it has the best search performance, but requires a much larger budget, which could be potentially unbounded when users lie about their cost. In fact, any other cost-independent/untruthful mechanisms, e.g. high degree mechanism, will suffer a large path cost. We found that 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. We also notice that the bonus will not incur significant extra pay, but it does motivate users to declare their true cost.

V. RELATED WORK

Given a description of a desired expertise, there are two kinds of expertise finding problem without using payment: (1) global expert finding: finding a person (e.g., [18]) or a group of ranked persons (e.g., [5]), similar enough to the desired one in a global database. (2) chain of social referrals: searching a desired person via a chain of acquaintance links using local acquaintance information in a social network, e.g. the *small world routing* [14]. Many research results were devoted to study various challenging small world routing problems, e.g., [4] [16], and ranking of closeness, e.g., [7]. 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 most paths were terminated because participants are *not sufficiently motivated* to relay messages [8]. Thus, we need design a mechanism that takes into account not only the participation cost of intermediate users, but also the selfish nature of these intermediate users.

VI. 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 need to design a truthful mechanism when the initiator wants to find a group of target experts.

ACKNOWLEDGMENT

The research of authors is partially supported by NSF CNS-0832120, NSF CNS-1035894, National Natural Science Foundation of China under Grant No. 61170216, National Basic Research Program of China (973 Program) under grant No. 2010CB328100, 2010CB334707, 2011CB302705, Tsinghua National Laboratory for Information Science and Technology (TNList), program for Zhejiang Provincial Key Innovative Research Team, and program for Zhejiang Provincial Overseas High-Level Talents (One-hundred Talents Program).

REFERENCES

- [1] ADAMIC, L. A., M. LUKOSE, R., PUNIYANI, A. R., AND HUBERMAN, B. A. Search in power-law networks. *Physical review E* 64, 4 (2001), 046135.
- [2] DODDS, P., MUHAMAD, R., AND WATTS, D. An experimental study of search in global social networks. *Science* 301, 5634 (2003), 827.
- [3] KAUTZ, H., SELMAN, B., AND SHAH, M. Referral web: combining social networks and collaborative filtering. *Communications of the ACM* 40, 3 (1997), 63–65.
- [4] KLEINBERG, J. Navigation in a small world. *Nature* 406, 6798 (2000), 845–845.
- [5] LAPPAS, T., LIU, K., AND TERZI, E. Finding a team of experts in social networks. In *ACM SIGKDD* (2009), pp. 467–476.
- [6] MILLER MCPHERSON, LYNN SMITH-LOVIN, J. M. C. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 2001.
- [7] OKAMOTO, K., CHEN, W., AND LI, X.-Y. Ranking of closeness centrality for large-scale social networks. In *The Second Int. Frontiers of Algorithmics Workshop (FAW)*, 2008.
- [8] SHARAD GOEL, ROBY MUHAMAD, D. W. Social search in "small-world" experiments. WWW, 2009, 701–710.
- [9] TANG, S.-J., YUAN, J., MAO, X., LI, X.-Y., CHEN, W., AND DAI, G. Relationship classification in large scale online social networks and its impact on information propagation. In *IEEE INFOCOM* (2011).
- [10] TRAUD, A. L., KELSIC, E. D., MUCHA, P. J., AND PORTER, M. A. Comparing community structure to characteristics in online collegiate social networks. *SIAM Review*, in press (arXiv:0809.0960), 2010.
- [11] TRAUD, A. L., MUCHA, P. J., AND PORTER, M. A. Social structure of facebook networks. arXiv:1102.2166, 2011.
- [12] TRAVERS, J., AND MILGRAM, S. An experimental study of the small world problem. *Sociometry* (1969), 425–443.
- [13] WANG, W., AND LI, X. Low-cost routing in selfish and rational wireless ad hoc networks. *IEEE Tran. on Mobile Computing* (2006), 596–607.
- [14] WATTS, D., AND STROGATZ, S. Collective dynamics of small-world networks. *Nature* 393 (1998), 440–442.
- [15] WATTS, D., AND STROGATZ, S. Small world. *Nature* 393 (1998), 440–442.
- [16] WATTS, D. J., DODDS, P. S., AND NEWMAN, M. E. J. Identity and search in social networks. *Science* 296 (2002), 1302–1305.
- [17] WHITE, H. Search parameters for the small world problem. *Social Forces* (1970), 259–264.
- [18] ZHANG, J., TANG, J., AND LI, J. Expert finding in a social network. *Advances in Databases: Concepts, Systems and Applications* (2010), 1066–1069.