Title	**Privacy exposure of online social search**
Author(s)	**Xu, K; Li, VOK**
Citation	**The IEEE Conference and Exhibition on Global Telecommunications Conference (GLOBECOM 2010), Miami, FL., 6-10 December 2010. In Proceedings of GLOBECOM 2010, 2010, p. 1-5**
Issued Date	**2010**
URL	**http://hdl.handle.net/10722/142823**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Privacy Exposure of Online Social Search

Kuang Xu and Victor O.K. Li
Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam, Hong Kong, China

***Abstract*—Online social search brings forth a new way to harness the Internet for answers. However, the personal and often sensitive information is unwittingly exposed to others when a person looks for an expert via the underlying social network. In this paper, we propose a model in which a node's behavior of looking for an expert is adjusted by his awareness of the potential expertise of his contacts. We derive the optimal distribution of nodes' awareness level that minimizes the system's privacy exposure, and prove that it corresponds to the unique Nash equilibrium. Our analysis shows that the optimal distribution over a posed question is inversely proportional to the *square root* of the corresponding expertise density.**

I. Introduction

Web search engine became popular since the last decade when there was a great deal of online content that could be indexed. Today, there are many humans who can be indexed, bringing forth a new way to harness the Internet for answers. Seeking answers in this context is based on the recent flourishing of online social networks (OSN). We refer to the OSN-based information search as online social search (OSS). Compared with searching for an answer on the Web which may yield scattershot results, asking a human expert from your *small world* [1] often gives you more subjective or tailored-made recommendations. Thus OSS utilizes the underlying network structure of OSNs to look for experts [2] via friends, and friends of friends. In turn, the answer emerges from people we trust, or people trusted by those we trust [3], whose collective expertise will likely be much higher than that of the relatively few people we happen to know personally. OSS has received attention in both research [4] and actual applications [5].

However, OSNs with millions of willing participants are no longer niche phenomena [6]. While a question is passed on along multi-step chains of acquaintances, the personal and often sensitive information is also unwittingly exposed to friends as well as strangers. In this paper, we investigate how social context affects the revelation of OSS users' personal information. In particular, we look at the amount of private information that is exposed in a referral session when a person looks for answers to a specific question. A referral session refers to the process from when a question is injected into the system until an answer is found. The privacy exposure here concerns the number of references a question is directed to in a referral session. Distinct from anonymous remailer [7] or anonymity network [8] which forwards messages concealing the senders' and the routers' identifying information, OSS applications require revealing a user's real identity. Therefore, when a question is passed along chains of acquaintances, the questioner's personal information accompanying his identity is also exposed to the intermediate nodes in the chains, and the more references receiving a question, the higher the questioner's privacy exposure.

How to properly forward a question to an appropriate expert is non-trivial, as one is confronted with the trade-off between forwarding the question to as many contacts as possible (e.g. flooding), and hence straining the willingness of possible responders [9], and forwarding them to a more compact set of contacts, thus missing an appropriate expert. Considering the above trade-off, in this paper, instead of having a referral agent flood questions to all its contacts or send them to a predetermined set of contacts, we utilize *random walk* to model a node's referral strategy (Section II). Since a node in a real OSS application maintains its local social network and directs a question only to those who are most likely to have an answer, we equip the nodes in our model with the intelligence of awareness, assuming every node is aware of the potential expertise of his contacts, and the capacity of a node's awareness towards another node is limited. Consequently, the behavior of looking for a reference is adjusted by this social context. Based on the model, we study how the distribution of nodes' awareness level $\mathcal{R}$ affects the exposure of nodes' privacy (Section III). In particular, to minimize the system's privacy exposure degree (PED), we derive the optimal distribution of nodes' awareness level in terms of the percentage of nodes that may have appropriate answers to the posed questions in the system (namely, expert density), and prove that the optimal distribution constitutes the unique Nash equilibrium. The analysis shows that the optimal distribution over a posed question is inversely proportional to the *square root* of the corresponding expert density. We also analyze the performance bounds of the system's minimal privacy exposure degree with heterogeneous settings of the number of a node's contacts. To empirically study the system, we apply the modeled referral strategy to the crawled data of a set of OSNs [10] with various settings (Section IV). The simulation result validates our analyses, and further shows that the underlying network connectivity has a negative relationship with the system's privacy exposure. Finally, we conclude this study with suggestions for future work (Section V).

II. Modeling

In this section, we model the nodes' behavior of looking for references. We first present the assumptions, definitions, and

* This research is supported in part by the University of Hong Kong Strategic Research Theme of Information Technology.

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

the question we study in the system, followed by the referral strategy of a node.

A. Assumption and definition

We consider an OSN as an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (OSN users) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (social ties) in the network. Each edge means one-hop question-forwarding is possible between the pair of nodes. Let $n = |\mathcal{V}|$ be the number of users in the system. We also denote by $\mathcal{N}_u \subseteq \mathcal{V}$ the set of neighbors of nodes u, and $d_u = |\mathcal{N}_u|$ the number of users in this set. Since a node maintains its local social network, we equip the nodes in our model with the intelligence of awareness, assuming every node is aware of the potential expertise of his neighbors. We denote by $S_u(v, i)$ a node u's awareness of its neighbor v's expertise, with respect to question i. $S_u(v, i)$ takes one of three possible values $\{-1, 0, 1\}$, such that

$$S_u(v,i) \triangleq \begin{cases} 1, & u \text{ knows } v \text{ is an expert on } i, \\ -1, & u \text{ knows } v \text{ is not an expert on } i, \\ 0, & \text{otherwise.} \end{cases}$$

The above assumption studies the simplest case that classifies nodes' awareness into three types. In other words, for every question i, we divide a node's neighbors into three possible sets. Nodes in the first set are regarded as experts on question i, and those in the second set are considered as not holding relevant answers to question i, and the rest of the neighbors are those that the node is uncertain about whether they are expert on i.

DEFINITION 1. *The **awareness level** of neighbors' expertise in question i is defined as*

$$r_i \triangleq \frac{\delta_i}{\sum_{u \in \mathcal{V}} d_u}, \quad \text{where}$$

$\delta_i = |\{e_{uv} | v \in \mathcal{N}_u \text{ and } S_u(v,i) \neq 0, \text{ for all } u \in \mathcal{V}\}|$.
e_{uv} in the above definition refers to the directed edge from node u to its neighbor v. We consider directed edge since, unlike a social tie between two people which represents the fact that the two people know each other, awareness here describes to what extent a person unilaterally feels whether he knows the expertise of another person on a particular question, and may not be symmetric. Thus r_i refers to the percentage of the directed edges among nodes that satisfy $S_u(v, i) \neq 0$. In other words, r_i is the probability that node u knows the potential expertise of node v, either "v is an expert on question i" or "v is a layman (not an expert) on question i", for each $v \in \mathcal{N}_u$.

In this paper, we study the steady state of OSS and assume that there are in total m posed questions in the system. We denote by $\mathcal{R} = [r_1, r_2, ..., r_m]$ the distribution of nodes' awareness level over m posed questions. Since in a practical system the capacity[1] of a node's awareness of another node's expertise is limited, we set the constraint

$$\sum_{i=1}^{m} r_i = c,$$

and the constant c represents the average capacity of a node's awareness towards another node.

DEFINITION 2. *The **expert density** on question i is defined as*

$$e_i \triangleq \frac{l_i}{n},$$

where $l_i > 0$ refers to the number of people that have relevant answers to question i in the system. We also denote by $\mathcal{E} = [e_1, e_2, ..., e_m]$ the expert density distribution over m posed questions.

DEFINITION 3. *The **privacy exposure degree** of a referral session for Question i, denoted by $PED(i)$, is defined as the expected number of nodes Question i has visited before it reaches an expert on it.*

We consider the above definition due to the fact that when a question is passed on along chains of friends, the questioner's personal and often sensitive information (i.e. at least the question itself) is also exposed to the intermediate nodes in the chains. We further denote the system's privacy exposure degree as PED, where $PED = \frac{1}{m} \cdot \sum_{i=1}^{m} PED(i)$, i.e., it is the average privacy exposure in the system.

Based on the above definitions, we attempt to answer the following questions: *How does the distribution of nodes' awareness level $\mathcal{R}$ affect the exposure of nodes' privacy? Given the expert density distribution $\mathcal{E}$ and the nodes' awareness capacity c, what is the optimal $\mathcal{R}$ which minimizes the system's privacy exposure degree PED?*

B. Referring strategy

Here, from a node's perspective, we introduce the referral strategy we utilize in this paper. If a node that receives Question i has expertise in i, it responds to the node that poses i with an answer, and this referral session is considered successful; otherwise, it checks to see whether some of its neighbors are potential experts on i and, if so, forwards i to a randomly selected expert neighbor (i.e. pick a member from that set uniformly with equal probability). If there are no expert neighbors on i, it forwards i randomly to one of its neighbors excluding those who are are not experts on i, and if there are no neighbors of this category, the referral session is considered failed.

In [11], we have analyzed the single step success rate of a referral session with the above strategy. The probability that a referral to an expert on Question i is satisfied at one step from node u is

$$q_i = 1 - (1 - e_i)(1 - r_i e_i)^{d_u} \tag{1}$$

III. Privacy Exposure Optimization

In this section, we study the effect of nodes' awareness distribution over different questions on the exposure of the questioners' privacy. In the system, heterogeneous setting of

[1] In reality, we also mean the social familiarity from a person to another is limited.

d (the number of neighbors of a node) of each node leads to heterogenous success rate at each step, which complicates the analysis. To obtain a clear vision of the privacy exposure problem of the modeled referral strategy of OSS, we first consider the homogeneous setting of d and then the heterogenous case.

A. Optimization

Our goal is to find the system's optimal awareness distribution $\mathcal{R}$, given $\mathcal{E}$, such that the system's minimal privacy exposure degree PED_{min} can be achieved with the capacity limit of nodes' awareness towards their neighbors. That is, to minimize

$$PED(\mathcal{E},\mathcal{R}) = \frac{1}{m}\cdot\sum_{i=1}^{m} PED(i) \tag{2}$$

subject to a constraint of $\mathcal{R}$:

$$\sum_{i=1}^{m} r_i = c \tag{3}$$

where c is a given constant.

THEOREM 1. *The optimal distribution of nodes' awareness level that minimizes the system's privacy exposure degree satisfies*

$$r_i \;\propto\; \frac{\sqrt{\frac{1}{e_i}}}{\sum_{j=1}^{m}\sqrt{\frac{1}{e_j}}}, \qquad i=1,...,m \tag{4}$$

and the minimal privacy exposure degree is

$$PED_{min} \;=\; \gamma\Big(\sum_{i=1}^{m}\sqrt{\frac{1}{e_i}}\Big)^2 \tag{5}$$

where γ is a system-dependent constant.

Proof: Statistical analogy between random walk and uniform sampling has been well studied [12]. One may approximate the consecutive random walk steps with independent sampling from a uniform distribution of the nodes in the system. The precision of approximation depends on the underlying network connectivity (i.e. the spectral gap of a graph). Since the typical OSN topologies exhibit properties that guarantee a large average degree [10], it is appropriate to approximate the question-forwarding strategy with uniform sampling in our context. Therefore, we approximate the distribution of the privacy exposure of a referral session with the Geometric distribution, and the expectation is $1/q_i$. That is,

$$\begin{aligned} PED(i) &= \sum_{j=1}^{\infty} j\cdot(1-q_i)^{j-1}q_i \\ &= \frac{1}{q_i} \end{aligned} \tag{6}$$

where q_i is the probability of finding an answer for question i at one particular step.

Thus, our problem is, given $\mathcal{E}$, to minimize

$$PED(\mathcal{E},\mathcal{R}) = \frac{1}{m}\cdot\sum_{i=1}^{m}\frac{1}{q_i} \tag{7}$$

with the capacity constraint of nodes' awareness towards their neighbors (Eqn. (3)).

By substituting (1) into (7), we obtain

$$PED(\mathcal{E},\mathcal{R}) = \frac{1}{m}\cdot\sum_{i=1}^{m}\frac{1}{1-(1-e_i)(1-r_ie_i)^d} \tag{8}$$

In this paper, we assume $e_i \ll 1$, which allows us to simplify the derivation by omitting e_i^2 and other higher order items in Eqn. (8). A similar assumption is also made in [12] [13]. The approximation lead to a suboptimal solution to the original optimization problem. To minimize the privacy exposure degree of the system, we need to solve the following constrained optimization problem

$$\textit{minimize}\quad PED(\mathcal{E},\mathcal{R}) = \frac{1}{m}\cdot\sum_{i=1}^{m}\frac{1}{(1+r_id)e_i} \tag{9}$$

$$\textit{s.t.}\quad \sum_{i=1}^{m} r_i = c \quad \textit{and}\quad 0<r_i,e_i<1 \tag{10}$$

where $i=1,...,m$. The inequality in the constraint (10) is due to the physical limitation of the awareness level and expert density.

We can solve this problem with the *Lagrangian multipliers* method [14]. The Lagrangian of the above problem is given by

$$\Lambda(\mathcal{R},\lambda) = \frac{1}{m}\cdot\sum_{i=1}^{m}\frac{1}{(1+r_id)e_i} + \lambda(\sum_{i=1}^{m} r_i - c)$$

Thus, we obtain

$$1+r_id = \sqrt{\frac{d}{\lambda m e_i}} \tag{11}$$

Combining (11) and the constraint (10), we can determine the nodes' optimal awareness distribution

$$r_i \;=\; (\frac{m}{d}+c)\cdot\frac{\sqrt{\frac{1}{e_i}}}{\sum_{j=1}^{m}\sqrt{\frac{1}{e_j}}} - \frac{1}{d}, \qquad i=1,...,m \tag{12}$$

and we obtain the distribution rule (4). Furthermore, we can derive the suboptimal value of $PED(\mathcal{E},\mathcal{R})$ with the above awareness distribution by combining Eqns. (1) and (7), and

$$PED_{min} \;=\; \frac{1}{m(m+dc)}\cdot\Big(\sum_{i=1}^{m}\sqrt{\frac{1}{e_i}}\Big)^2. \tag{13}$$

■

Eqns. (12) present the optimal distribution of nodes' awareness level for a practical situation where the expert density on any question is far less than 1. We can see that the nodes' optimal awareness distribution over a posed questions is inversely proportional to the square root of the corresponding expert density in the system. Consider two posed questions i and k, we have

$$\frac{r_i}{r_k} \;\propto\; \sqrt{\frac{e_k}{e_i}}.$$

This shows that the optimal awareness distribution favors

questions on which there are fewer experts in the system (e.g. if $e_i < e_k$, then $r_i > r_k$).

Nash equilibrium. In the above analysis, we regard the capacity of nodes' awareness towards their neighbors as a constant. In order to reduce the system's privacy exposure to its minimum, each node need to contribute as much of its awareness as possible. However, in a practical system, certain nodes may not be willing to contribute their awareness up to the capacity, and we are interested in the stable *social norm* in the system. Here, we maintain the objective function in the problem (9) and relax the constraint (10) slightly. That is,

$$s.t. \quad \sum_{i=1}^{m} r_i \leq c \quad and \quad 0 < r_i, e_i < 1 \tag{14}$$

LEMMA 1. *To minimize the system's privacy exposure, all nodes in the system are required to contribute up to the capacity of their awareness towards their neighbors.*

Proof: Suppose that certain nodes do not contribute their awareness up to the capacity, namely, $\sum_{i=1}^{m} r_i = c^* \leq c$. Denote by $\mathcal{R}^* = [r_1^*, r_2^*, ..., r_m^*]$ the optimal distribution of nodes' awareness level defined by Eqns. (12), and by combining Eqn. (13) we obtain

$$\begin{aligned} PED_{min}^* &= \frac{1}{m(m+dc^*)} \cdot \Big(\sum_{i=1}^{m} \sqrt{\frac{1}{e_i}}\Big)^2 \\ &\geq PED_{min}. \end{aligned}$$

We can see that the system with $\mathcal{R}^*$ incurs higher privacy exposure. ■

From Eqns. (12), we are informed of the uniqueness of the optimal distribution of node's awareness level. Together with Lemma 1, we can immediately arrive at the following conclusion.

THEOREM 2. *The optimal distribution of nodes' awareness level that minimizes the system's privacy exposure degree constitutes the unique Nash equilibrium.*

B. Performance bound

With the heterogenous setting of d in the system, the single step success rate of a referral session differs at each step. Therefore it is difficult to obtain the closed-form solution. In this section, we study the upper and lower bounds of the system's minimal privacy exposure degree PED_{min} with heterogeneous settings of d.

THEOREM 3. *For nodes with heterogeneous d, given the expert density distribution $\mathcal{E}$, they system's minimal privacy exposure degree PED_{min} is bounded by*

$$\begin{aligned} PED_{min} &\geq \frac{1}{m(m+cd_{max})} \cdot \Big(\sum_{i=1}^{m} \sqrt{\frac{1}{e_i}}\Big)^2, \\ PED_{min} &\leq \frac{1}{m(m+cd_{min})} \cdot \Big(\sum_{i=1}^{m} \sqrt{\frac{1}{e_i}}\Big)^2. \end{aligned} \tag{15}$$

where $d_{min} = min(d_u)$ and $d_{max} = max(d_u)$, $u \in \mathcal{N}$.

Proof: From Eqn. (1) we know q_i is an increasing function with respect to d_u. Thus, $\forall u \in \mathcal{N}$,

$$q_{min}(i) \leq q_i \leq q_{max}(i), \tag{16}$$

where $q_{min}(i)$ and $q_{max}(i)$ correspond to the single step success rate with $d_u = d_{min}$ and $d_u = d_{max}$, respectively. We compare the referral session of heterogeneous d with two other referral sessions that are characterized by homogeneous single step success rates of $q_{min}(i)$ and $q_{max}(i)$. Denote by $PED(i)_U$ and $PED(i)_L$ the privacy exposure degrees of two referral sessions for question i with homogeneous setting of d_{min} and d_{max}, respectively. From (16), we obtain $\forall i$, $1 \leq i \leq m$,

$$PED(i)_L \leq PED(i) \leq PED(i)_U \tag{17}$$

Consider the linear combination of the system's average privacy exposure degree for the m posed questions, we can obtain, for any $\mathcal{E}$ and $\mathcal{R}$,

$$\frac{1}{m} \cdot \sum_{i=1}^{m} PED(i)_L \leq \frac{1}{m} \cdot \sum_{i=1}^{m} PED(i) \leq \frac{1}{m} \cdot \sum_{i=1}^{m} PED(i)_U,$$

which can be written as

$$PED_L \leq PED \leq PED_U. \tag{18}$$

Since (18) holds for any distribution of node's awareness level $\mathcal{R}$, we can obtain

$$min\ PED_L \leq PED_{min} \leq min\ PED_U.$$

According to Eqn. (5), we obtain the performance bound (15). ■

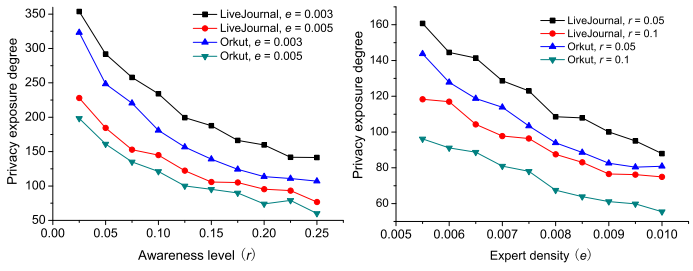
OSN	Orkut	LiveJournal
Number of nodes	3,072,441	5,284,457
Estimated crawled fraction	11.3%	95.4%
Number of links	223,534,301	77,402,652
Av. no. of friends per node	106.1	16.97
Fraction of symmetric links	100.0%	73.5%

TABLE I
STATISTICS OF THE OSN DATASETS [10].

IV. EVALUATION

In this section, we study empirically the privacy exposure of referral sessions in the system, based on our modeled referring strategy. We utilize the connectivity data of a set of OSNs, namely Orkut, LiveJournal, collected by Mislove et al. [10]. Orkut is a website of explicitly defined social network to help people make new friends and maintain existing relationships. LiveJournal is an online social network of bloggers. The major statistics of these datasets are summarized in Table I. We believe it is more realistic to evaluate the system on these real social network data[2]. Since the networks are too big for evaluation, we sample several different portions from each network with Snowball sampling [15], and the size of the sampled networks is set to 1×10^4.

[2]Utilizing datasets from actual OSS applications such as Aardvark (which is based on Facebook) would be ideal, but those datasets are unavailable since Facebook prohibits automated crawlers.

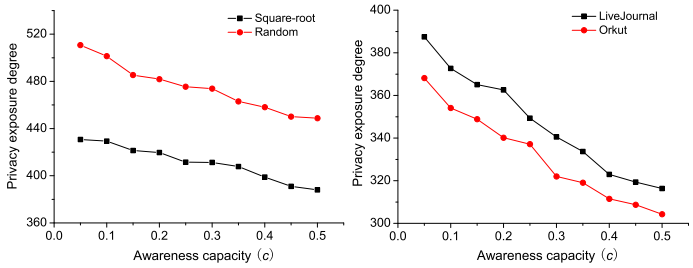


(a) Varying awareness level (b) Varying expert density

Fig. 1. Privacy exposure degree of a single referral session.

Figure 1(a) and Figure 1(b) show the privacy exposure degree of a single referral session with varying awareness level r and expert density e, respectively. The results are obtained as averages of 1000 simulation runs. They illustrate the performance improvement as r or e increases. We can see that the privacy exposure degree of a single referral session decreases as r or e increases, matching our previous analysis. We also observe that the resulting performance from Orkut is better than that from LiveJournal.

We further study the effect of node's awareness distribution on the system's overall privacy exposure. In the simulation, 20 questions are randomly posed, and we select 20 different pairs of (e, r) according to Eqn. (12) (referred to as Square-root distribution), with $e = 5 \times 10^{-4}, 1 \times 10^{-3}, ..., 1 \times 10^{-2}$ (20 values with increment of 5×10^{-4}). For comparison, we also generate 20 random values of r subject to the capacity constraint (3) (referred to as Random distribution). Figure 2(a) shows the results under the two awareness level distributions based on a 5-regular graph of 1×10^4 nodes. The x-axis represents the awareness capacity c. We can see that the system's privacy exposure degree with the Square-root distribution of r is smaller than that with the Random distribution of r. This verifies Theorem 1 that the Square-root distribution of r incurs the minimal privacy exposure. In addition, we observe that the performance improves as c increases, which matches our analysis. Figure 2(b) presents the simulation results under the two OSN datasets. Again, we observe that the system's privacy exposure degree from Orkut is smaller than that from Livejournal. Note that the Orkut dataset has higher average degree than that of LiveJournal. This leads us to an empirical conclusion: the network connectivity, characterized by the average degree, is negatively related to the system's privacy exposure degree (i.e. positively related to the performance).



(a) Different distributions of r (b) Different OSN datasets

Fig. 2. The system's privacy exposure degree ($m = 20$).

V. Conclusion

In this paper, we study the privacy exposure problem of OSS. We utilize random walk to model a node's referring behavior, adjusted by the node's awareness of the potential expertise of its neighbors. We derive the optimal distribution of nodes' awareness level, which minimizes the system's privacy exposure degree, and prove it constitutes the unique Nash equilibrium. The analysis shows that the optimal distribution over a posed question is inversely proportional to the square root of the corresponding expert density. In addition, we present the performance bounds of the system's minimal privacy exposure degree with heterogeneous settings on the number of a node's contacts. The evaluation based on crawled data of several OSNs validates our analysis. Our model is a first step to analytically study the OSS system. In the future, we would like to develop tools to capture the users' awareness of their friends' expertise. In addition, we currently assume the homogeneous setting that answers from multiple experts on a question are the same, and it would be interesting to study response filtering considering the system's trust [16] and reputation [17] mechanisms, which are closely related to the system's privacy exposure.

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun 1998.
[2] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks," *Science*, vol. 296, pp. 1302–1305, May 2002.
[3] T. Tassier and F. Menczer, "Emerging small-world referral networks in evolutionary labor markets," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 5, pp. 482–492, Oct 2001.
[4] J. Zhang and M. V. Alstyne, "SWIM: fostering social network based information search," in *Proceedings of ACM CHI '04*, 2004.
[5] Aardvark, "http://vark.com/," .
[6] A. Acquisti and R. Gross, "Imagined communities: awareness, information sharing, and privacy on the Facebook," in *Proceedings of the 6th Workshop on Privacy Enhancing Technologies*, 2006.
[7] G. F. du Pont, "The time has come for limited liability for operators of true anonymity remailers in cyberspace: an examination of the possibilities and perils," *Journal of Techonlogy Law and Policy*, 2001.
[8] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for Web transactions," *ACM Transactions on Information and System Security*, 1998.
[9] H. Kautz, B. Selman, and M. Shah, "Referral Web: combining social networks and collaborative filtering," *Communications of the ACM*, vol. 40, no. 3, pp. 63–65, 1997.
[10] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007.
[11] K. Xu, J. Xie, and V. O.K. Li, "Locating experts via online social networks," in *Proceedings of IEEE ICC '10*, 2010.
[12] C. Gkantsidis, M. Mihail, and A. Saberi, "Random walks in peer-to-peer networks," in *Proceedings of IEEE INFOCOM '04*, 2004.
[13] N. Bisnik and A. A. Abouzeid, "Optimizing random walk search algorithms in p2p networks," *Comput. Netw.*, vol. 51, no. 6, pp. 1499–1514, 2007.
[14] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
[15] S. K. Thompson, *Sampling*, John Wiley & Sons Inc., 2 edition, April 2002.
[16] F. E. Walter, S. Battiston, and F. Schweitzer, "A model of a trust-based recommendation system on a social network," *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 16, no. 1, pp. 57–74, Feb 2008.
[17] L. Mui, M. Mohtashemi, and A. Halberstadt, "A computational model of trust and reputation for e-businesses," in *Proceedings of HICSS '02*, 2002.