# Personalized Location Privacy in Mobile Networks: A Social Group Utility Approach

Xiaowen Gong*, Xu Chen*, Kai Xing†, Dong-Hoon Shin*, Mengyuan Zhang*, and Junshan Zhang*
*School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, AZ 85287
Email: {xgong9, xchen179, donghon.shin.2, mzhang72, junshan.zhang}@asu.edu
†School of Computer Science and Technology
University of Science and Technology of China, Hefei, Anhui 230027, China
Email: kxing@ustc.edu.cn

***Abstract*—With increasing popularity of location-based services (LBSs), there have been growing concerns for location privacy. To protect location privacy in a LBS, mobile users in physical proximity can work in concert to collectively change their pseudonyms, in order to hide spatial-temporal correlation in their location traces. In this study, we leverage the social tie structure among mobile users to motivate them to participate in pseudonym change. Drawing on a social group utility maximization (SGUM) framework, we cast users' decision making of whether to change pseudonyms as a socially-aware pseudonym change game (PCG). The PCG is further based on a general anonymity model that allows a user to have its specific anonymity set for personalized location privacy. For the SGUM-based PCG, we show that there exists a socially-aware Nash equilibrium (SNE), and quantify the system efficiency of the SNE with respect to the optimal social welfare. Then we develop an algorithm that greedily determines users' strategies, based on the social group utility derived from only the users whose strategies have been determined. We show that this algorithm can efficiently find a Pareto-optimal SNE with social welfare higher than that for the socially-oblivious PCG. We further show that the Pareto-optimal SNE can be achieved in a distributed manner. Numerical results corroborate that social welfare can be significantly improved by exploiting social ties.**

## I. Introduction

The proliferation of mobile devices is predicted to continue in the foreseeable future. In 2014, mobile phone shipments are projected to reach 1.9 billion units, about 7 times that of desktop and laptop combined [1]. With rapid growth of mobile networks, location-based services (LBS) have become increasingly popular recently (e.g., location-based navigation and recommendation). However, the providers of LBSs are often considered not trustworthy, due to the risk of leaking users' location information to other parties (e.g., sell users' location data). As a result, mobile users are exposed to potential privacy threats when using a LBS. Although a user can use a pseudonym for the LBS, an adversary can infer the user's real identity from its location traces (e.g., from the user's home and work addresses). To protect location privacy, an effective approach is to "confuse" the adversary using the notion of *anonymity* [2]: mobile users in physical proximity can change their pseudonyms simultaneously to form an anonymity set, so that the adversary cannot distinguish any of them from the others.

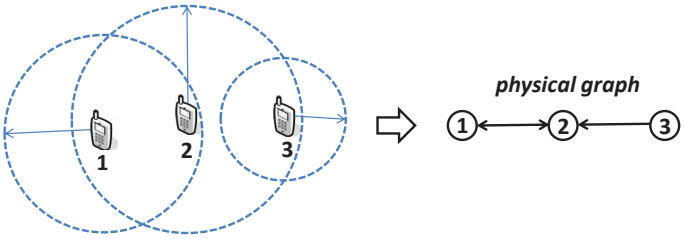


Fig. 1. Illustration of the general anonymity model: each user specifies its anonymity set for personalized location privacy by defining an anonymity range, e.g., a disk centered at the user's location. User 1 and 2 are out of user 3's anonymity range and thus are not in user 3's anonymity set (represented by no direct edge from user 1 or 2 to user 3); user 1 and 3 are within user 2's anonymity range and thus are in user 2's anonymity set (represented by directed edges from user 1 and 3 to user 2).

Clearly, as mobile devices are carried and operated by human beings, pseudonym change hinges heavily on human behavior. In particular, altruistic behaviors are widely observed among people with social ties[1]. It is then natural to ask "Is it possible to leverage social ties for pseudonym change to enhance location privacy?" The past few years have witnessed explosive growth of online social networks. In 2013, the number of online social network users worldwide has crossed 1.73 billion, nearly one quarter of the world's population [3]. As a result, social relationships influence people's interactions with each other in an unprecedented manner. This motivates us to exploit social ties among users for pseudonym change to improve their location privacy. Since pseudonym change typically incurs considerable overhead (e.g., service interruption, resource consumption [4]), users need strong incentives (e.g., adequate privacy gain) to participate in pseudonym change. We caution that secure protocols are needed to hide users' real identities when social information is used (which will be elaborated further in Section IV-E).

A basic assumption commonly used in existing studies [4]–[6] is that all users participating in pseudonym change have the *same* anonymity set. However, from an individual user's perspective, the set of users that can obfuscate its pseudonym

This research was supported in part by the U.S. NSF grants CNS-1422277, ECCS-1408409, CNS-1117462, DTRA grant HDTRA1-13-1-0029, and NSFC grants 61332004, 61170267.

[1]In this paper, "social tie" refers to "positive social tie" for brevity.

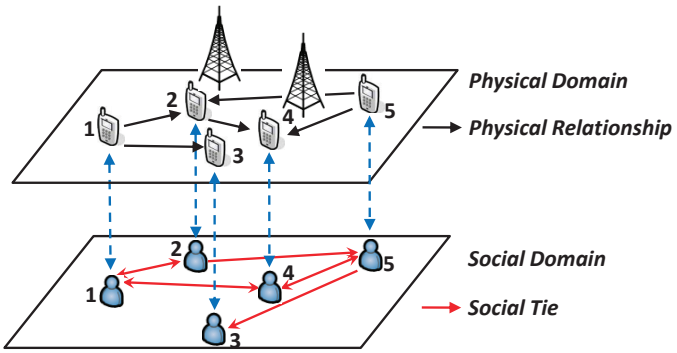


Fig. 2. Illustration of the social group utility maximization (SGUM) framework: a mobile network can be viewed as a virtual social network underlying a physical communication network. Mobile users have diverse social ties in the social domain, while mobile devices have diverse physical relationships in the physical domain.

(i.e., its anonymity set) can be different from that of another user, depending on users' physical locations. For example, a user with a higher level of privacy sensitivity can have a smaller anonymity set than others. It is thus desirable to meet users' needs for *personalized location privacy*. To this end, we consider a general anonymity model where a user can define its specific anonymity set different from others' (as illustrated in Fig. 1).

In this paper, we leverage the social tie structure among mobile users to motivate them to participate in pseudonym change. Drawing on a social group utility maximization (SGUM) framework recently developed in [7], we cast users' decision making of whether to participate in pseudonym change as a socially-aware pseudonym change game (PCG). The SGUM framework captures the impact of users' diverse social ties on the interactions of their mobile devices subject to diverse physical relationships (as illustrated in Fig. 2). The PCG is based on a general anonymity model that allows each user to have its specific anonymity set. In the SGUM-based PCG, each user aims to maximize its social group utility, defined as the sum of its individual utility and the weighted sum of its social friends' individual utilities. For the SGUM-based PCG, we are interested in answering the following important questions: Does the game admit a socially-aware Nash equilibrium (SNE)? What is the system efficiency of the SNE? How can we efficiently find a SNE with desirable system efficiency?

The main contributions of this paper can be summarized as follows.

- We propose a framework where the social tie structure among mobile users is leveraged to motivate them to participate in pseudonym change, based on a general anonymity model that allows each user to have its specific anonymity set for personalized location privacy. Taking a social group utility approach, we cast users' decision making of whether to change pseudonyms as a pseudonym change game.
- For the SGUM-based PCG, we first show that there always exists a socially-aware Nash equilibrium. Then we quantify the system efficiency of the SNE by bounding the gap between the optimal social welfare and the social welfare of the "best" SNE, which is the SNE that achieves the maximum social welfare among all SNEs. We observe that the best SNE is difficult to compute in general, and the often used best response updates would converge to a SNE with lower social welfare. We then develop an algorithm that greedily determines users' strategies, based on the social group utility derived from only the users whose strategies have been determined. We show that this algorithm can efficiently find a Pareto-optimal SNE with social welfare no less than that of the best SNE for the socially-oblivious PCG. We also show that the social welfare of the Pareto-optimal SNE increases as social ties increase, and its performance gap with respect to the optimal social welfare is bounded. We further show that the Pareto-optimal SNE can be achieved in a distributed manner.
- We evaluate the performance of the Pareto-optimal SNE by numerical results, which demonstrate that social welfare can be significantly improved by exploiting social ties.

The rest of this paper is organized as follows. Related work are reviewed in Section II. In Section III, we formulate the socially-aware pseudonym change game based on a general anonymity model for personalized location privacy, under the social group utility maximization framework. Section IV focuses on the analysis of the SGUM-based PCG. Numerical results are presented in Section V. Section VI concludes this paper.

## II. Related Work

With growing concerns for location privacy arising from pervasive mobile communication and computing, a great deal of research have been done to protect mobile users' location privacy. This work falls in the category of anonymity-based approaches [2], [4]–[6], [8]. Earlier studies [2], [9] show that an adversary can infer the real identity of a mobile user by analyzing the spatial-temporal correlation of its location traces. To overcome this vulnerability, pseudonyms should not only be changed over time but also be obfuscated across space to prevent inference attacks. Inspired by the notion of $k$-anonymity, Beresford and Stajano [2] introduced the notion of mix zone. By changing pseudonyms within a mix zone, users can make their new pseudonyms undistinguishable to the adversary. While the mix zone model assumes that all users have the same anonymity set, the general anonymity model proposed in this paper allows each user to define its specific anonymity set different from others'. A few work have studied users' interactions in pseudonym change based on game-theoretic models. The mix zone based pseudonym change has been studied in [4] as a non-cooperative game with complete or incomplete information. Auction-based mechanisms have been designed in [6] to incentive users to participate in pseudonym change. To our best knowledge, this paper is the first to exploit social relationships to improve location privacy by pseudonym change.

The social aspect of mobile networking is an emerging paradigm for network design and optimization [7], [10]–[15]. Recently, a social group utility maximization (SGUM)

framework is developed in [7], [14], which captures the impact of users' diverse social ties on the interactions of their devices subject to diverse physical coupling. A primary merit of this framework is that it provides rich modeling flexibility and spans the continuum between non-cooperative game and network utility maximization, two traditionally disjoint paradigms for network optimization.

## III. Model and Problem Formulation

### A. Privacy Threat in Location-based Services

We consider a mobile network where users obtain their locations via mobile devices that are capable of localization (e.g., by GPS or wireless access points based localization). Users send their locations to a LBS provider for a certain LBS (e.g., location-based navigation or recommendation), and the LBS provider feedbacks the desired results to the users based on their reported locations. To protect privacy, each user uses a pseudonym as its identity for the LBS.

As in [2], [6], [8], we assume that the LBS provider is untrusted, i.e., it may leak users' location traces to an adversary. For example, the adversary may steal the location data by hacking into the LBS system. The adversary aims to learn the real identity of a user by linking and analyzing the locations visited by the user's pseudonym. We also assume that users are honest-but-curious such that each user honestly follows the protocols with others (which will be discussed in Section IV-E), but is curious about others' private information. We further assume that the adversary may collude with a limited number of users to gain useful information for inferring a user's real identity.

The use of pseudonym allows short-term reference to a user (e.g., one pseudonym can be used for the navigation of an entire trip between two locations), which is useful for many LBSs and does not disclose private information. However, long-term linking among a user's locations should be prevented, as it may reveal sufficient information for inferring the user's real identity [9], [16], [17]. Although a user may hide explicit linking among its locations by changing its pseudonym, the adversary can still link different pseudonyms of the user by exploiting the spatial-temporal correlation of its locations. For example, consider a user that visits location $l_1$ with pseudonym Alice at time $t_1$, and then visits location $l_2$ which is close to location $l_1$ with pseudonym Bob at time $t_2$. If the adversary observes from the location traces that no other user changes its pseudonym between time $t_1$ and $t_2$, or there exists such a user but it does not visit any location close to location $l_1$ or $l_2$, then the adversary can infer that pseudonym Alice and Bob must refer to the same user, since only the same user can visit both location $l_1$ and $l_2$ within the limited period between time $t_1$ and $t_2$.

### B. Pseudonym Change for Personalized Location Privacy

To protect location privacy from inference attacks, an effective approach is based on the notion of anonymity: users in physical proximity can coordinate their pseudonym changes to happen simultaneously [2], so that the adversary cannot link their pseudonyms before the changes to their respective pseudonyms after the changes. Existing studies [4]–[6] assume that all users participating in pseudonym change have the *same* anonymity set. However, based on an individual user's belief of the adversary's power against its location privacy (e.g., the adversary's side information about that user), the set of users that it believes can obfuscate its pseudonym (i.e., its anonymity set) can be different from that of another user. Thus motivated, we consider a general anonymity model that can meet users' needs for personalized location privacy, depending on users' physical locations. In particular, each user specifies an *anonymity range* (a physical area) such that the set of users within the anonymity range constitute that user's *potential anonymity set*. For example, a user's anonymity range can be a disk centered at the user's location, with a large radius indicating a low level of privacy sensitivity (as illustrated in Fig. 1). Note that for two users at different locations, their anonymity ranges are different even when they have the same shape (e.g., two disks with the same radius but different centers), and thus their potential anonymity sets can be different.

Formally, consider a set of users $\mathcal{N} \triangleq \{1, \cdots, N\}$ where each user $i$ makes a decision $a_i$ on whether or not to participate in pseudonym change, denoted by $a_i = 1$ and $a_i = 0$, respectively. Based on users' physical locations, the privacy gain perceived by a user participating in pseudonym change depends on which users also participate. Each user $i$ incurs a cost of $c_i > 0$ to participate in pseudonym change. This cost is due to a number of factors, e.g., the participating users should stop using the LBS for a period of time. Based on the general anonymity model, the physical coupling among users can be captured by a physical graph $(\mathcal{N}, \mathcal{E}^P)$, where user $j$ is connected by a directed edge $e^P_{ji} \in \mathcal{E}^P$ to user $i$ if user $j$ is in user $i$'s potential anonymity set, denoted by $\mathcal{N}^P_{i-}$ (i.e., $j \in \mathcal{N}^P_{i-}$). Note that the physical coupling between two users can be asymmetric. The privacy gain perceived by a participating user $i$ is defined as its *anonymity set size*, i.e., the number of participating users in $\mathcal{N}^P_{i-}$. Note that the anonymity set size is a widely adopted privacy metric[2] for anonymity-based approaches. For example, $k$-anonymity is used as the privacy metric in [6], [8], where a user achieves location privacy if its pseudonym cannot be distinguished among $k$ users. Then the individual utility of user $i$, denoted by $u_i$, is given by

$$u_i(a_i, \boldsymbol{a}_{-i}) \triangleq a_i \left( \sum_{j \in \mathcal{N}^P_{i-}} a_j - c_i \right) \tag{1}$$

where $\boldsymbol{a}_{-i}$ denotes the vector of the strategies of all users except user $i$. If a user participates, its individual utility is its privacy gain minus its participation cost; otherwise, it is zero. Note that $c_i$ is a relative cost compared to privacy gain.

### C. Social Group Utility Maximization (SGUM)

Social relationships play an increasingly important role in people's interactions with each other. One important attribute

[2]Another privacy metric is the entropy of the adversary's uncertainty of a user's pseudonym. However, it is usually difficult to compute since it requires probability distribution which is difficult to attain.

of social relationship is that people are altruistic to their social "friends" (including friends, family, colleagues, etc.), as they care about their social friends' welfare. As a result, a user would take into account the effect of its behavior on its social friends. Recently, a social group utility maximization framework has been developed in [7], which captures the impact of mobile users' diverse social ties on the interactions of their mobile devices subject to diverse physical relationships.

Appealing to the SGUM framework [7], we model the social tie structure among the users in $\mathcal{N}$ by a social graph $(\mathcal{N}, \mathcal{E}^S)$, where user $i$ is connected by a directed edge $e_{ij}^S \in \mathcal{E}^S$ to user $j$ if user $i$ has a social tie with user $j$. We denote the *social tie* (level) from user $i$ to user $j$ as $s_{ij}$. We assume that each user $i$'s social tie to itself is $s_{ii} = 1$, and we normalize user $i$'s social tie to user $j \neq i$ as $s_{ij} \in (0, 1]$, which quantifies the extent to which user $i$ cares about user $j$ relative to user $i$ cares about itself. We also assume that $s_{ij} = 0$ if no social tie exists from user $i$ to user $j$. We define user $i$'s *social group* $\mathcal{N}_{i+}^S$ as the set of users that user $i$ has social ties with (i.e., $\mathcal{N}_{i+}^S \triangleq \{j \in \mathcal{N} | e_{ij}^S \in \mathcal{E}^S\}$).

To take into account the social ties among users, each user $i$ aims to maximize its social group utility, defined as

$$f_i(a_i, \boldsymbol{a}_{-i}) \triangleq u_i(a_i, \boldsymbol{a}_{-i}) + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} u_j(a_j, \boldsymbol{a}_{-j}). \quad (2)$$

Note that a user's social group utility consists of its own utility and the weighted sum of the individual utilities of the users in its social group. Therefore, the social group utility captures both physical coupling and social coupling among users in a unified way. Also note that a user does not need to know the individual utilities of its social friends (which may be their private information) to make the decision (as will be shown in equation (3)). In Section IV-E, we will discuss how social information can be used while preserving the privacy of users' real identities with each other.

### D. SGUM based Pseudonym Change Game

Based on the SGUM framework, users' socially-aware decision making for pseudonym change boils down to a social group utility maximization game. Specifically, each user $i \in \mathcal{N}$ is a *player* and its *strategy*[3] is $a_i \in \{0, 1\}$. Let $\boldsymbol{a} = (a_1, \cdots, a_n)$ denote the *strategy profile* consisting of all users' strategies. The *payoff* function of a user is defined as its social group utility function. Given the strategies of other users, each user $i$ aims to choose the *best response* strategy that maximizes its social group utility:

$$\underset{a_i}{\text{maximize}} \quad f_i(a_i, \boldsymbol{a}_{-i}), \ \forall i \in \mathcal{N}.$$

Similar in spirit to the Nash equilibrium of a standard non-cooperative game, the notion of *socially-aware Nash equilibrium* (SNE) applies to the SGUM game.

*Definition 1 (Socially-aware Nash Equilibrium [7]):* A strategy profile $\boldsymbol{a}^{sne} = (a_1^{sne}, \cdots, a_n^{sne})$ is a socially-aware Nash equilibrium of the SGUM-based PCG if no user can improve its social group utility by unilaterally changing its strategy, i.e.,

$$a_i^{sne} = \arg\max_{a_i} \quad f_i(a_i, \boldsymbol{a}_{-i}), \ \forall i \in \mathcal{N}.$$

[3]As we focus on pure strategies in this work, we use "strategy" and "action" interchangeably.

Due to the rational and autonomous nature of users, a SNE is a stable outcome that is acceptable by all users.

For the sake of system efficiency, a natural objective is to maximize the *social welfare* of the system, which is the total individual utility of all users denoted by $v(\boldsymbol{a}) \triangleq \sum_{i \in \mathcal{N}} u_i(\boldsymbol{a})$. A strategy profile $\boldsymbol{a}^* = (a_1^*, \cdots, a_n^*)$ is *social optimal* [18] if it achieves the maximum social welfare among all profiles, i.e., $v(\boldsymbol{a}^*) \geq v(\boldsymbol{a})$, $\forall \boldsymbol{a}$. Although the social optimal profile is the best outcome in terms of system efficiency, it is often not acceptable by all users. Then, it is desirable to achieve the "best" SNE, i.e., the SNE that achieves the maximum social welfare among all SNEs (referred to as "the best SNE").

Another desirable property for system efficiency is *Pareto-optimality*. A strategy profile $\boldsymbol{a}^{po} = (a_1^{po}, \cdots, a_n^{po})$ is Pareto-optimal [18] if there does not exist a Pareto-superior profile $\boldsymbol{a}' = (a_1', \cdots, a_n')$ such that no user achieves a worse individual utility while at least one user achieves a better individual utility, i.e.,

$$u_i(a_i', \boldsymbol{a}_{-i}') \geq u_i(a_i^{po}, \boldsymbol{a}_{-i}^{po}), \ \forall i \in \mathcal{N}$$

with at least one strict inequality.

For the SGUM-based PCG, we are interested in answering the following important questions:

- Does the game admit any SNE? If yes, what is the system efficiency of the best SNE?
- How can we efficiently find a SNE? What is the system efficiency of this SNE? Is it Pareto-optimal?

To answer these questions, in Section IV we will focus on the analysis of the SGUM-based PCG.

## IV. Social Group Utility Maximization based Pseudonym Change Game

### A. Benchmark: Socially-oblivious Pseudonym Change Game

As the benchmark, we start with a basic case of the PCG: the PCG for socially-oblivious users (SO-PCG), i.e., $s_{ij} = 0$, $\forall e_{ij}^S \in \mathcal{N}^S$. In this case, each user is selfish and the social group utility degenerates to the individual utility.

For SO-PCG, there can exist multiple SNEs[4] with different values of social welfare (as illustrated in Fig. 3). For system efficiency, it is desirable to achieve the best SNE, which is the SNE that achieves the maximum social welfare among all SNEs. To find this SNE, we can use best response updates as follows: with all users' actions initially set to 1, each user asynchronously updates its action as its best response to other users' actions (no two users update at the same time). We illustrate how it works by an example in Fig. 3. We use $\mathcal{N}_1(\boldsymbol{a}) \triangleq \{i \in \mathcal{N} | a_i = 1\}$ to denote the set of participating users. The above result is formally stated below.

*Proposition 1:* For SO-PCG, best response updates can find the SNE that achieves the maximum social welfare among all SNEs.

Due to space limitation, the proof is given in our online technical report [19]. As the best SNE achieves the maximum system efficiency among all SNEs, we will use the best SNE for SO-PCG as the benchmark for the general case of the PCG: the PCG for socially-aware users (SA-PCG).

[4]For SO-PCG, an SNE is equivalent to a NE for a standard non-cooperative game. For consistency of terminology, we still call it "SNE" in this case.

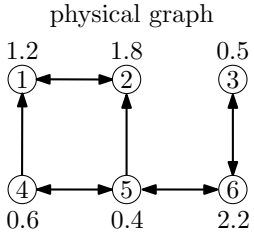


Fig. 3. An example of SO-PCG. The number beside a user is its cost. Using best response updates, we have $u_6 = 2 - 2.2 < 0 \rightarrow a_6 = 0 \rightarrow u_3 = 0 - 0.5 < 0 \rightarrow a_3 = 0 \rightarrow \mathcal{N}_1 = \{1, 2, 4, 5\}$, which is a SNE. It is also Pareto-superior to the other two SNEs with $\mathcal{N}_1 = \{4, 5\}$ and $\mathcal{N}_1 = \varnothing$, and hence is the best SNE.

### B. Existence and Efficiency of SNE

We next study SA-PCG. We first establish the existence of SNE. Using (1) and (2), we have

$$\begin{aligned} & f_i(1, \boldsymbol{a}_{-i}) - f_i(0, \boldsymbol{a}_{-i}) \\ &= u_i(1, \boldsymbol{a}_{-i}) - u_i(0, \boldsymbol{a}_{-i}) + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} \left(u_j(1, \boldsymbol{a}_{-i}) - u_j(0, \boldsymbol{a}_{-i})\right) \\ &= \sum_{j \in \mathcal{N}_{i-}^P} a_j - c_i + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} a_j. \end{aligned} \tag{3}$$

It is clear from (3) that no user participating is always a SNE. We thus conclude that at least one SNE exists.

Then we show an important property of the social group utility function. It follows from (3) that

$$\begin{aligned} & f_i(1, \boldsymbol{a}_{-i}) - f_i(0, \boldsymbol{a}_{-i}) - \left(f_i(1, \boldsymbol{a}'_{-i}) - f_i(0, \boldsymbol{a}'_{-i})\right) \\ &= \sum_{j \in \mathcal{N}_{i-}^P} a_j - c_i + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} a_j - \left( \sum_{j \in \mathcal{N}_{i-}^P} a'_j - c_i + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} a'_j \right) \\ &= \sum_{j \in \mathcal{N}_{i-}^P} (a_j - a'_j) + \sum_{j \in \mathcal{N}_{i+}^S} s_{ij} (a_j - a'_j). \end{aligned} \tag{4}$$

Let $\boldsymbol{a} \leq \boldsymbol{a}'$ denote entry-wise inequality (i.e., $a_i \leq a'_i$, $\forall i \in \mathcal{N}$). The property below follows from (4).

*Property 1 (Supermodularity):* If $\boldsymbol{a}_{-i} \leq \boldsymbol{a}'_{-i}$, then $f_i(1, \boldsymbol{a}_{-i}) - f_i(0, \boldsymbol{a}_{-i}) \leq f_i(1, \boldsymbol{a}'_{-i}) - f_i(0, \boldsymbol{a}'_{-i})$.

Property 1 implies that if a user's best response strategy is to participate, then it remains the best response strategy if more users participate; if a user's best response strategy is to not participate, then it remains the best response strategy if less users participate.

To quantify the system efficiency of the SNE, we provide a bound of the gap between the social welfare of the best SNE and the optimal social welfare in the following result.

*Theorem 1:* The performance gap between the maximum social welfare among all SNEs and the optimal social welfare is upper bounded by $\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_{i+}^S} (1 - s_{ij})$.

The proof is given in our online technical report [19]. Theorem 1 shows that the performance gap decreases as social ties increase. In particular, when users are socially-oblivious (i.e., $s_{ij} = 0$, $\forall e_{ij}^S \in \mathcal{N}^S$), the performance gap reaches the maximum, and the best SNE for SA-PCG degenerates to the best SNE for SO-PCG; when users are fully altruistic (i.e., $s_{ij} = 1$, $\forall e_{ij}^S \in \mathcal{N}^S$), the performance gap becomes 0, and the best SNE degenerates to the social optimal profile. This demonstrates that the SNE spans the continuum between a NE for a standard non-cooperative game and the optimal solution for network utility maximization, two traditionally disjoint paradigms for network optimization.

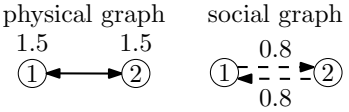


Fig. 4. Using best response updates, we have $f_1(1,1) - f_1(0,1) = 1 - 1.5 + 0.8 > 0$, $f_2(1,1) - f_2(1,0) = 1 - 1.5 + 0.8 > 0$, and hence $\mathcal{N}_1 = \{1, 2\}$ is a SNE. However, it is not Pareto-optimal, since it is Pareto inferior to $\mathcal{N}_1 = \varnothing$ as $u_1(0,0) = u_2(0,0) = 0 > 1 - 1.5 = u_1(1,1) = u_2(1,1)$. Furthermore, its social welfare is less than that with $\mathcal{N}_1 = \varnothing$ as $v(1,1) = -1 < 0 = v(0,0)$, where $\mathcal{N}_1 = \varnothing$ is also a SNE for SO-PCG.

### C. Computing Pareto-Optimal SNE

In this subsection, we turn our attention to finding a SNE with desirable system efficiency.

For the PCG for fully altruistic users (i.e., $s_{ij} = 1$, $\forall e_{ij}^S \in \mathcal{N}^S$), it is clear that the social optimal profile $\boldsymbol{a}^*$ is a SNE, and is the solution to the following problem:

$$\begin{aligned} \underset{\boldsymbol{a}}{\text{maximize}} \quad & \sum_{i \in \mathcal{N}} a_i \left( \sum_{j \in \mathcal{N}_{i-}^P} a_j - c_i \right) \\ \text{subject to} \quad & a_i \in \{0, 1\}, \ \forall i \in \mathcal{N}. \end{aligned} \tag{5}$$

Observe that problem (5) is an integer quadratic programming, which is difficult to solve in general[5]. Since the PCG for fully altruistic users is a special case of SA-PCG, it is also difficult to compute the best SNE for SA-PCG. Based on this observation, our objective below is to efficiently compute a SNE with desirable system efficiency.

To compute a SNE of SA-PCG, a naive approach is to use best response updates in a similar way as with SO-PCG: with all users' actions initially set to 1, each user asynchronously updates its action as its best response to other users' actions. Due to Property 1, a user who changes its strategy from 1 to 0 will never change it back to 1, and thus the best response updates always converges to a SNE. However, it has drawbacks as illustrated by an example in Fig. 4: the SNE may not be Pareto-optimal and its social welfare may be worse than that of a SNE for SO-PCG. Thus motivated, our objective below is to efficiently find a SNE such that 1) it is Pareto-optimal and 2) its social welfare is no less than that of the best SNE for SO-PCG, which is the benchmark.

*1) Algorithm Design:* To this end, we design an algorithm as described in Algorithm 1. The main idea of the algorithm is to greedily determine users' strategies, depending on the social group utility derived from *the users whose strategies have been determined* (referred to as "determined users"), denoted by

$$f'_i(a_i, \boldsymbol{a}_{-i}) \triangleq u_i(a_i, \boldsymbol{a}_{-i}) + \sum_{j \in \mathcal{N}_{i+}^S \setminus \overline{\mathcal{N}}} s_{ij} u_j(a_i, \boldsymbol{a}_{-i})$$

where $\overline{\mathcal{N}}$ denotes the set of users whose strategies have not been determined (referred to as "undetermined users"). An undetermined user's action is fixed once it becomes determined.

Specifically, the algorithm proceeds in rounds and each round consists of phase I and phase II. In phase I, with all undetermined users' actions initially set to 1, an undetermined

[5]We conjecture that problem (5) is an NP-hard problem.

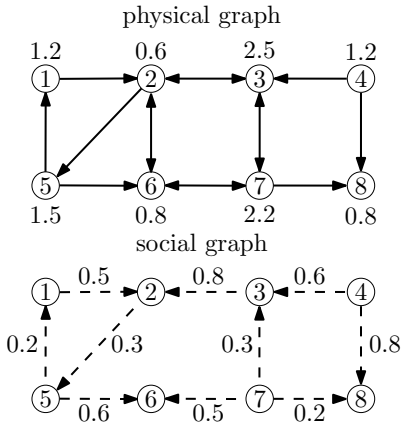


Fig. 5. An example that illustrates how Algorithm 1 works: The number beside a user is its cost; the number beside a social edge is its social tie.

user's action is changed from 1 to 0 if that improves its social group utility derived from the determined users, i.e.,

$$f_i'(1,\boldsymbol{a}_{-i}) - f_i'(0,\boldsymbol{a}_{-i}) = u_i(1,\boldsymbol{a}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a_j < 0$$

until no such user exists. Then the undetermined users whose actions remain 1 become determined and their actions are fixed to 1. In phase II, with all undetermined users' actions initially set to 0, an undetermined user becomes determined and its action is fixed to 1 if that improves its social group utility derived from the determined users, until no such user exists. The algorithm terminates when no undetermined user becomes determined during either phase I or phase II of a round.

**Algorithm 1:** Compute the Pareto-optimal SNE for SA-PCG

$\overline{\mathcal{N}} \leftarrow \mathcal{N}$;
**repeat**
  // Phase I;
  $\boldsymbol{a} \leftarrow (1,\cdots,1)$, $\mathcal{N}_I \leftarrow \overline{\mathcal{N}}$;
  **while** $\exists i \in \overline{\mathcal{N}}$ *such that* $u_i(1,\boldsymbol{a}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a_j < 0$ **do**
    $a_i \leftarrow 0$, $\mathcal{N}_I \leftarrow \mathcal{N}_I \setminus \{i\}$;
  **end**
  // Phase II;
  $\overline{\mathcal{N}} \leftarrow \overline{\mathcal{N}} \setminus \mathcal{N}_I$, $\mathcal{N}_{II} \leftarrow \varnothing$;
  **while** $\exists i \in \overline{\mathcal{N}}$ *such that* $u_i(1,\boldsymbol{a}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a_j \geq 0$ **do**
    $a_i \leftarrow 1$, $\overline{\mathcal{N}} \leftarrow \overline{\mathcal{N}} \setminus \{i\}$, $\mathcal{N}_{II} \leftarrow \mathcal{N}_{II} \cup \{i\}$;
  **end**
**until** $\mathcal{N}_I \cup \mathcal{N}_{II} = \varnothing$;
**return** $\boldsymbol{a}^e \leftarrow \boldsymbol{a}$;

We use an example in Fig. 5 to illustrate how Algorithm 1 works and outline the steps as follows.

- Phase I of 1st round: $u_1 = 1 - 1.2 < 0 \rightarrow a_1 = 0$; $u_5 = 1 - 1.5 < 0 \rightarrow a_5 = 0$; $u_4 = 0 - 1.2 < 0 \rightarrow a_4 = 0 \rightarrow u_3 = 2 - 2.5 < 0 \rightarrow a_3 = 0$; $u_7 = 2 - 2.2 < 0 \rightarrow a_7 = 0 \rightarrow u_8 = 0 - 0.8 < 0 \rightarrow a_8 = 0$; $u_2 = 1 - 0.6 > 0$; $u_6 = 1 - 0.8 > 0$; $\mathcal{N}_I = \{2, 6\}$.
- Phase II of 1st round: $u_5 + s_{56} = 1 - 1.5 + 0.6 > 0 \rightarrow a_5 = 1 \rightarrow u_1 + s_{12} = 1 - 1.2 + 0.5 > 0 \rightarrow a_1 = 1$; $u_3 + s_{32} = 1 - 2.5 + 0.8 < 0$; $u_7 + s_{76} = 1 - 2.2 + 0.5 < 0$; $u_4 = 0 - 1.2 < 0$; $u_8 = 0 - 0.8 < 0$; $\mathcal{N}_{II} = \{1, 5\}$.
- Phase I of 2nd round: $u_4 = 0 - 1.2 < 0 \rightarrow a_4 = 0 \rightarrow u_3 + s_{32} = 2 - 2.5 + 0.8 > 0$; $u_7 + s_{76} = 2 - 2.2 + 0.5 > 0$; $u_8 = 1 - 0.8 > 0$; $\mathcal{N}_I = \{3, 7, 8\}$.
- Phase II of 2nd round: $u_4 + s_{43} + s_{48} = 0 - 1.2 + 0.6 + 0.8 > 0 \rightarrow a_4 = 1$; $\mathcal{N}_{II} = \{4\}$.

Since the size of the set of undetermined users $\overline{\mathcal{N}}$ is upper bounded by $N$, the computational complexity of phase I and phase II of a round is bounded by $O(N^2)$. Since at least one user is determined during a round, the algorithm must terminate within $N$ rounds. Therefore, the running time of the algorithm is bounded by $O(N^3)$. In Section V, numerical results will demonstrate that the computational complexity of Algorithm 1 is nearly a quadratic function of the number of users. In Section IV-D, we will discuss a distributed version of Algorithm 1.

*Theorem 2:* For SA-PCG, the strategy profile $\boldsymbol{a}^e = (a_1^e, \cdots, a_n^e)$ computed by Algorithm 1 is a Pareto-optimal SNE.

The proof is given in the Appendix. As the SNE computed by Algorithm 1 is Pareto-optimal, it is desirable for system efficiency. Next we show that its social welfare is no less than that of the best SNE for SO-PCG. To this end, we first show that the Pareto-optimal SNE is monotonically "increasing" with respect to social ties.

*Theorem 3:* For SA-PCG, when social ties increase (i.e., $s_{ij}' \geq s_{ij}$, $\forall i, j \in \mathcal{N}$), the corresponding Pareto-optimal SNE $\boldsymbol{a}^{e'}$ satisfies that $\boldsymbol{a}^{e'} \geq \boldsymbol{a}^e$ and $v(\boldsymbol{a}^{e'}) \geq v(\boldsymbol{a}^e)$.

The proof is given in our online technical report [19]. Intuitively, with larger social ties to other users, a user is more likely to participate in favor of its social group utility, even at the cost of reducing its individual utility. Theorem 3 confirms this intuition: as social ties become larger, more users participate at the Pareto-optimal SNE. Furthermore, the social welfare of the Pareto-optimal SNE also increases.

When Algorithm 1 is used for SO-PCG, we can see that it works exactly the same as the best response updates used to find the best SNE for SO-PCG, and thus they find the same profile. Based on this observation, using Theorem 3, we have the following result.

*Corollary 1:* The social welfare of the Pareto-optimal SNE for SA-PCG is no less than that of the best SNE for SO-PCG.

Corollary 1 guarantees that the social welfare of the Pareto-optimal SNE is no less than the benchmark SNE for SO-PCG. In Section V, numerical results will demonstrate that the Pareto-optimal SNE is efficient, with a performance gain up to 20% over the benchmark.

*2) Efficiency of Pareto-Optimal SNE:* Next we investigate the social welfare of the Pareto-optimal SNE compared to the optimal social welfare. To this end, we first show that the set of participating users at the Pareto-optimal SNE is a subset of that at the social optimal profile.

*Lemma 1:* For SA-PCG, the Pareto-optimal SNE $\boldsymbol{a}^e$ satisfies that $\boldsymbol{a}^e \leq \boldsymbol{a}^*$.

The proof is given in our online technical report [19]. Let $\mathcal{N}_\Delta \triangleq \mathcal{N}_1(\boldsymbol{a}^*) \setminus \mathcal{N}_1(\boldsymbol{a}^e)$ denote the set of users that participate

under the social optimal profile but not under the Pareto-optimal SNE. Using Lemma 1, we have the following result.

*Theorem 4:* The performance gap between the social welfare of the Pareto-optimal SNE and the optimal social welfare is upper bounded by $\sum_{i\in\mathcal{N}_\Delta}\sum_{j\in\mathcal{N}_{i-}^P\cap\mathcal{N}_\Delta}1+\sum_{i\in\mathcal{N}_\Delta}\sum_{j\in\mathcal{N}_{i+}^S\setminus\mathcal{N}_\Delta}(1-s_{ij})$. Furthermore, this bound decreases as social ties increase.

The proof is given in our online technical report [19]. Theorem 4 shows that the performance gap decreases as social ties increase. The first part of the performance bound is due to the fact that we trade the social welfare optimality of the computed SNE (among all SNEs) for computational efficiency. As a result, when users are fully altruistic (i.e., $s_{ij}=1$, $\forall i,j\in\mathcal{N}$), the second part becomes 0, while the first part could be greater than 0. In Section V, numerical results will demonstrate that the Pareto-optimal SNE is efficient, with a performance gap less than 10% on average compared to the optimal social welfare for real social dataset.

### D. Distributed Computation of Pareto-Optimal SNE

The Pareto-optimal SNE computed by Algorithm 1 can be achieved in a distributed manner. To this end, each user first obtains its potential anonymity set and its social ties with others (which will be discussed in Section IV-E). Following Algorithm 1, each user checks if it should change its strategy according to the condition in line 5 or 10 based on other users' strategies, and if yes, announces the change to all users. With time divided into slots, a random backoff mechanism can be used so that at most one user announces a change of strategy in a time slot. If no user announces a change, it indicates the end of phase I or phase II of a round. Therefore, all users keep track of the current state of the algorithm as it proceeds, and thus can act correctly according to the algorithm. The computational complexity of the distributed version of Algorithm 1 is almost the same as the centralized version, and is upper bounded by $O(N^3)$. Note that each user only knows the strategies of other users during the execution the algorithm, and thus users' privacy is preserved. After reaching the Pareto-optimal SNE, the users who decide to change their pseudonyms implement their pseudonym changes.

### E. Further Discussions

We assume that there is a third party platform where users interact with each other to make pseudonym change decisions and coordinate their pseudonym changes. The platform only serves to allow users to exchange information. We assume that the platform is honest-but-curious such that it honestly delivers messages among users, but is curious about users' private information. To protect privacy, each user also uses a pseudonym as its identity on the platform (which can be different from that used for the LBS). To make a socially-aware pseudonym change decision, each users needs to know its potential anonymity set and its social ties with others. This can be achieved in a privacy-preserving manner using secure protocols as discussed below. Note that the platform is not involved in the computing tasks of these protocols.

A user can learn whether another user is within its anonymity range using a certain private proximity detection protocol [20], [21]. For example, the protocol proposed in [20] can be used if the anonymity range is a disk. Specifically, the protocol involves several message exchanges between the two users, including one message that contains encrypted values that are functions of a user' location or the radius of the anonymity range. The protocol guarantees that both users can only learn the binary result of whether or not one is in another's anonymity range, and neither user can learn the other's location or anonymity range. In addition, since location information is encrypted in the messages, the platform cannot learn any user's location information. Similarly, the protocol in [21] can be used if the anonymity range is a convex polygon. Therefore, each user can learn its potential anonymity set without revealing its location information.

A user can also learn its social tie with another user without disclosing one's real identity to the other. To this end, each user keeps a social profile consisting of the social communities that it belongs to (e.g., a community of colleagues at the same workplace), and sets a *single* social tie level for each community based on its social relationships with those in the community. Each community is identified by a predefined key that is only known to the community's members. Using a certain private matching protocol such as [22], [23], two users can learn whether they have a community in common, and if yes, which community[6] it is. In particular, the protocol involves several message exchanges between the two users, including one message that contains encrypted values that are functions of the keys of a user's social communities. The protocol ensures that both users can only know the community they have in common (if it exists), and neither user can learn any additional social information of the other, or pretend to have a community in common with the other. Since a community typically has many members, neither user can know the other's real identity even when they know the community they both belong to. In addition, since social information is encrypted in the messages, the platform cannot learn any user's social information. Therefore, each user can learn its social ties with those in its potential anonymity set while keeping their real identities private. Note that although the adversary might collude with multiple users, it is almost infeasible for the adversary to find a sufficient number of colluding users who have social ties with a specific user, in order to infer the user's real identity.

## V. Simulation Results

In this section, we provide numerical results to illustrate the system efficiency of the Pareto-optimal SNE computed by Algorithm 1. We compare the Pareto-optimal SNE with the best SNE for SO-PCG, and the social optimal solution which is found by exhaustive search.

We consider $N$ mobile users randomly located in a square area with side length 500 m. We assume that the anonymity range of each user is a disk centered at the user's location

[6]To protect privacy, only one community in common is revealed if they have multiple communities in common.

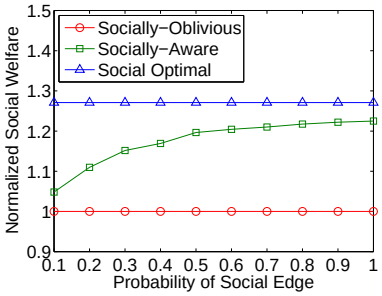


Fig. 6. Impact of $P_S$.

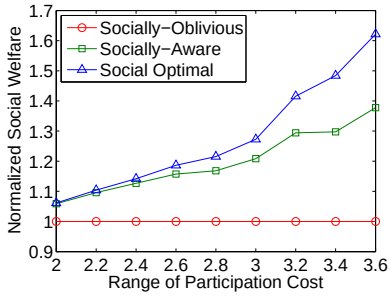


Fig. 7. Impact of $\overline{C}$.

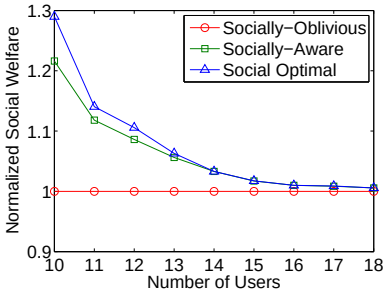


Fig. 8. Impact of $N$.

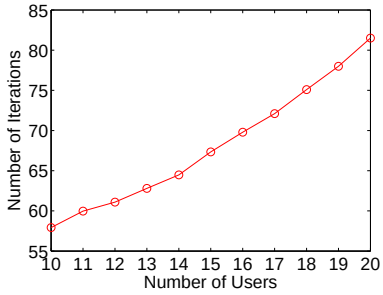


Fig. 9. Computational complexity versus $N$.

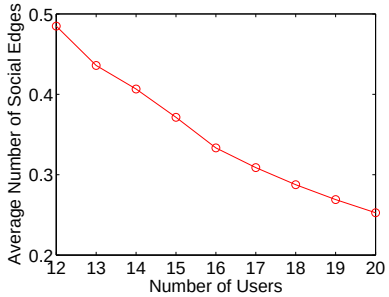


Fig. 10. Average number of social edges per user versus $N$ for the real dataset based social network.

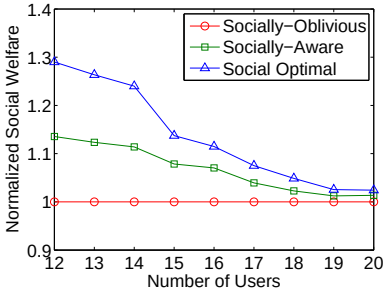


Fig. 11. Impact of $N$ for the real dataset based social network.

with radius randomly chosen from $\{200\text{ m}, 300\text{ m}, 400\text{ m}\}$. Based on users' physical locations and anonymity ranges, there exists a physical edge from user $i$ to user $j$ if user $i$ is in the anonymity range of user $j$. We assume that a user's cost of changing pseudonym is uniformly distributed in $[0, \overline{C}]$. We simulate the social graph using two methods as follows.

### A. Erdos-Renyi Model Based Social Graph

We simulate the social graph based on the Erdos-Renyi (ER) graph model [24], where a social edge exists between any two users with probability $P_S$. We assume that the social tie is 1 if it exists. We set the default values of parameters as $N = 10$, $P_S = 0.5$, $\overline{C} = 3$. For each parameter setting, we average the results over 1000 runs.

We illustrate the impact of $P_S$, $\overline{C}$, and $N$ on the normalized social welfare in Fig. 6, 7, and 8, respectively. We observe from these figures that socially-aware users significantly outperforms socially-oblivious users, especially when $P_S$ or $\overline{C}$ is large, or $N$ is small. This is because more users participate in pseudonym change when they are socially-aware, which improves the social welfare. On the other hand, the social welfare of socially-aware users is close to the optimal social welfare. Fig. 6 shows that the performance of socially-aware users improves when $P_S$ increases, with a performance gain up to 20% over socially-oblivious users and a performance gap about 10% on average from the optimal social welfare. This is because larger social ties further encourage users to participate. Fig. 7 and 8 show that the performance gaps from socially-oblivious users to socially-aware users and the social optimal solution decrease as $\overline{C}$ and $N$ increase, respectively. This is because with a lower participation cost, or a higher privacy gain of participation due to the increased total number of users, more users would participate even when they are socially-oblivious. Therefore, the performance gap decreases as it depends on the users that participate only when they are socially-aware. We plot the number of iterations for running Algorithm 1 versus $N$ in Fig. 9. We observe that the computational complexity increases nearly quadratically as the number of users increases. This shows that the algorithm is scalable for a large number of users.

### B. Real Data Trace Based Social Graph

We simulate the social graph according to the real dataset from Brightkite [25], which is a online social network based on mobile phones. We plot the average number of social edges of a user versus the total number of users in Fig. 10. We illustrate the impact of $N$ on the social welfare in Fig. 11, where we set the range of participation cost as $\overline{C} = 3$. We can see that the performance gain of socially-aware users can achieve up to 15% over socially-aware users, and its performance gap from the optimal social welfare is less than 10% on average. This verifies the effectiveness of exploiting social ties for improving location privacy based on real social dataset.

## VI. Conclusion

In this paper, we have studied a socially-aware pseudonym change game for personalized location privacy, based on a general anonymity model with user-specific anonymity sets. The game is based on a social group utility maximization framework that captures diverse social ties and diverse physical relationships among mobile users. For the SGUM-based PCG, we show that there exists a socially-aware Nash equilibrium, and quantify the performance gap of the SNE

with respect to the optimal social welfare. Then we develop an algorithm that can efficiently find a Pareto-optimal SNE, with social welfare no less than the maximum social welfare among all SNEs for the socially-oblivious PCG. We further show that the social welfare of the Pareto-optimal SNE increases as social ties increase, and its performance gap with respect to the optimal social welfare is upper bounded. Numerical results demonstrate that social welfare can be significantly improved by exploiting users' social ties.

## REFERENCES


[1] "Gartner: Worldwide PC, tablet and mobile phone shipments to grow 4.5 percent in 2013." [Online]. Available: http://www.gartner.com/newsroom/id/2610015

[2] A. Beresford and F. Stajano, "Location privacy in pervasive computing," in *IEEE Pervasive Computing 2003*.

[3] "eMarketer: Social networking reaches nearly one in four around the world." [Online]. Available: http://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976

[4] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes, "On non-cooperative location privacy: A game-theoretic analysis," in *ACM CCS 2009*.

[5] X. Liu, H. Zhao, M. Pan, H. Yue, X. Li, and Y. Fang, "Traffic-aware multiple mix zone placement for protecting location privacy," in *IEEE INFOCOM 2012*.

[6] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for $k$-anonymity location privacy," in *IEEE INFOCOM 2013*.

[7] X. Chen, X. Gong, L. Yang, and J. Zhang, "A social group utility maximization framework with applications in database assisted spectrum access," in *IEEE INFOCOM 2014*.

[8] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving $k$-anonymity in location based services," in *IEEE INFOCOM 2013*.

[9] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *SECURECOMM 2005*.

[10] K.-C. Chen, M. Chiang, and H. Poor, "From technological networks to social networks," *IEEE J. Sel. Area. Comm.*, vol. 31, no. 9, pp. 548–572, Sept. 2013.

[11] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Area. Comm.*, vol. 26, no. 5, pp. 748–760, June 2008.

[12] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *ACM MOBIHOC 2009*.

[13] X. Chen, B. Proulx, X. Gong, and J. Zhang, "Social trust and social reciprocity based cooperative D2D communications," in *ACM MOBIHOC 2013*.

[14] X. Gong, X. Chen, and J. Zhang, "Social group utility maximization game with applications in mobile social networks," in *IEEE Allerton 2013*.

[15] ——, "Social group utility maximization in mobile networks: From altruistic to malicious behavior," in *IEEE CISS 2014*.

[16] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," in *IEEE Pervasive Computing 2006*.

[17] J. Krumm, "Inference attacks on location tracks," in *IEEE Pervasive Computing 2007*.

[18] M. Osborne and A. Rubinstein, *A course in game theory*. MIT Press, 1994.

[19] "Personalized location privacy in mobile networks: A social group utility approach," Technical Report. [Online]. Available: http://informationnet.asu.edu/pub/SGUM-infocom15-TR.pdf

[20] G. Zhong, I. Goldberg, and U. Hengartner, "Louis, Lester and Pierre: Three protocols for location privacy," in *ACM PETS 2007*.

[21] B. Mu and S. Bakiras, "Private proximity detection via computational geometric approaches," in *ACM MobiDE 2013*.

[22] R. Zhang, Y. Zhang, J. Sun, and G. Yan, "Fine-grained private matching for proximity-based mobile social networking," in *IEEE INFOCOM 2012*.

[23] L. Zhang, X.-Y. Li, and Y. Liu, "Message in a sealed bottle: Privacy preserving friending in social networks," in *IEEE ICDCS 2013*.

[24] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, pp. 17–61, 1960.

[25] "SNAP: Network datasets: Brightkite." [Online]. Available: http://snap.stanford.edu/data/loc-brightkite.html


## APPENDIX

For convenience, let $\mathcal{N}_{I,k}$ and $\mathcal{N}_{II,k}$ denote the set of users that become determined in phase I and phase II of round $k$ in Algorithm 1, respectively. Let $\mathcal{N}_0(\boldsymbol{a}) \triangleq \{i \in \mathcal{N} | a_i = 0\}$ denote the set of users with action 0 under profile $\boldsymbol{a}$.

### *Proof of Proposition 1*

Since SO-PCG is a special case of SA-PCG, Property 1 also applies to the individual utility function $u_i$. Therefore, due to Property 1, a user who changes its strategy from 1 to 0 will not change it back to 1. As a result, best response updates always terminates and results in a profile $\boldsymbol{a}^o$ that is a SNE.

Next we show that $\boldsymbol{a}^o$ achieves the maximum social welfare among all SNEs. It suffices to show that $\boldsymbol{a}^o$ is Pareto-superior to any other SNE. To this end, we first show that a profile $\boldsymbol{a}'$ is not a SNE if $\mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^o) \neq \varnothing$. Suppose such $\boldsymbol{a}'$ is a SNE. Let $i \in \mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^o)$ be the first user among $\mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^o)$ whose action is changed to 0, and $\bar{\boldsymbol{a}}$ be the profile right before that change. Since $\boldsymbol{a}' \leq \bar{\boldsymbol{a}}$, we have $u_i(\boldsymbol{a}') \leq u_i(\bar{\boldsymbol{a}}) < 0 = u_i(0, \boldsymbol{a}'_{i-})$ due to that 0 is the best response strategy. This shows that $\boldsymbol{a}'$ is not a SNE. Therefore, for any SNE $\boldsymbol{a}'$ other than $\boldsymbol{a}^o$, we must have $\boldsymbol{a}' < \boldsymbol{a}^o$. Then for each $i \in \mathcal{N}_1(\boldsymbol{a}')$, we have $u_i(\boldsymbol{a}') \leq u_i(\boldsymbol{a}^o)$. For each $i \in \mathcal{N}_0(\boldsymbol{a}')$, since $\boldsymbol{a}^o$ is a SNE, we have $u_i(\boldsymbol{a}') = 0 = u_i(0, \boldsymbol{a}^o_{-i}) \leq u_i(\boldsymbol{a}^o)$. Therefore $\boldsymbol{a}^o$ is Pareto-superior to $\boldsymbol{a}'$. Thus we show that $\boldsymbol{a}^o$ is the best SNE.

### *Proof of Theorem 1*

We first show that we can construct a SNE from the social optimal profile $\boldsymbol{a}^*$ using best response updates: with all users' actions initially set according to the social optimal profile, a user's action is changed from 1 to 0 if that can improve its social group utility. To this end, we first show that this algorithm can terminate. Without loss of generality, we assume that there does not exist $\boldsymbol{a}'$ with $\boldsymbol{a}' > \boldsymbol{a}^*$ such that $v(\boldsymbol{a}') \geq v(\boldsymbol{a}^*)$. Suppose there exists $i \in \mathcal{N}_0(\boldsymbol{a}^*)$ such that $f_i(1, \boldsymbol{a}^*_{-i}) \geq f_i(0, \boldsymbol{a}^*_{-i})$. Then we have

$$v(1, \boldsymbol{a}^*_{-i}) - v(0, \boldsymbol{a}^*_{-i}) = u_i(1, \boldsymbol{a}^*_{-i}) + \sum_{j \in \mathcal{N}^S_{i+}} a_j$$

$$\geq u_i(1, \boldsymbol{a}^*_{-i}) + \sum_{j \in \mathcal{N}^S_{i+}} s_{ij} a_j = f_i(1, \boldsymbol{a}^*_{-i}) - f_i(0, \boldsymbol{a}^*_{-i}) \geq 0$$

where the first equality follows a similar manipulation as in (3), and the second equality follows from (3). This contradicts the previous assumption. Therefore we must have $f_i(1, \boldsymbol{a}^*_{-i}) < f_i(0, \boldsymbol{a}^*_{-i})$ for each $i \in \mathcal{N}_0(\boldsymbol{a}^*)$. Then, due to Property 1, the algorithm must terminate and results in a profile $\boldsymbol{a}^b$, which is a SNE and satisfies that $\boldsymbol{a}^* \geq \boldsymbol{a}^b$.

Next we show an upper bound on $v(\boldsymbol{a}^*) - v(\boldsymbol{a}^b)$. For any $i \in \mathcal{N}_1(\boldsymbol{a}^*) \setminus \mathcal{N}_1(\boldsymbol{a}^b)$, let $\bar{\boldsymbol{a}}$ be the profile right before $a_i$ is

changed to 0 in the algorithm. Then we have

$$\begin{aligned} v_i^{\Delta} &\triangleq v(1,\bar{\boldsymbol{a}}_{-i}) - v(0,\bar{\boldsymbol{a}}_{-i}) = u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S} \bar{a}_j \\ &= u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S} s_{ij}\bar{a}_j + \sum_{j\in\mathcal{N}_{i+}^S} (1-s_{ij})\bar{a}_j \\ &= f_i(1,\bar{\boldsymbol{a}}_{-i}) - f_i(0,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S} (1-s_{ij})\bar{a}_j \\ &< \sum_{j\in\mathcal{N}_{i+}^S} (1-s_{ij})\bar{a}_j \le \sum_{j\in\mathcal{N}_{i+}^S} (1-s_{ij}) \end{aligned}$$

where the last equality follows from (3), and the first inequality is due to that 0 is the best response strategy. Therefore we have

$$v(\boldsymbol{a}^*) - v(\boldsymbol{a}^b) = \sum_{i\in\mathcal{N}_1(\boldsymbol{a}^*)\setminus\mathcal{N}_1(\boldsymbol{a}^b)} v_i^{\Delta} \le \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}_{i+}^S} (1-s_{ij}).$$

*Proof of Theorem 2*

We first show that $\boldsymbol{a}^e$ is a SNE. We consider three cases of each user $i$ as follows.

*Case 1*: $i \in \mathcal{N}_1(\boldsymbol{a}^e)$ and $i \in \mathcal{N}_{I,k}$

Let $\boldsymbol{a}'$ be the profile right after phase I during which $i$ remains in $\mathcal{N}_I$. Since $\boldsymbol{a}^e \ge \boldsymbol{a}'$, using (3) we have

$$f_i(1,\boldsymbol{a}^e_{-i}) - f_i(0,\boldsymbol{a}^e_{-i}) \ge u_i(1,\boldsymbol{a}'_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a'_j \ge 0$$

where the second inequality is due to the condition in line 5.

*Case 2:* $i \in \mathcal{N}_1(\boldsymbol{a}^e)$ and $i \in \mathcal{N}_{II,k}$

Let $\boldsymbol{a}'$ be the profile right after $i$ becomes determined in phase II. Since $\boldsymbol{a}^e \ge \boldsymbol{a}'$, using (3) we have

$$f_i(1,\boldsymbol{a}^e_{-i}) - f_i(0,\boldsymbol{a}^e_{-i}) \ge u_i(1,\boldsymbol{a}'_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a'_j \ge 0$$

where the second inequality is due to the condition in line 10.

*Case 3:* $i \in \mathcal{N}_0(\boldsymbol{a}^e)$

Since $i$ is not included in $\mathcal{N}_{II}$ in phase II of the last round, using (3) we have

$$f_i(1,\boldsymbol{a}^e_{-i}) - f_i(0,\boldsymbol{a}^e_{-i}) = u_i(1,\boldsymbol{a}'_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}a'_j < 0$$

where the inequality is due to the condition in line 10.

Next we show that $\boldsymbol{a}^e$ is Pareto-optimal. Suppose there exists $\boldsymbol{a}'$ that is Pareto-superior to $\boldsymbol{a}^e$. It suffices to show that i) $\mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^e) = \varnothing$ and ii) $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}') = \varnothing$. We first show part i). Suppose $\mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^e) \neq \varnothing$. Then for each $i \in \mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^e)$, we have $u_i(\boldsymbol{a}') \ge u_i(\boldsymbol{a}^e) = 0$. Let $i$ be the first user among $\mathcal{N}_1(\boldsymbol{a}') \setminus \mathcal{N}_1(\boldsymbol{a}^e)$ whose action is set to 0 during phase I of the last round, and $\bar{\boldsymbol{a}}$ be the profile right before $a_i = 0$ is performed. Since $\bar{\boldsymbol{a}}_{-i} \ge \boldsymbol{a}'_{-i}$, we have $u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}} s_{ij}\bar{a}_j \ge u_i(1,\bar{\boldsymbol{a}}_{-i}) \ge u_i(1,\boldsymbol{a}'_{-i}) \ge 0$, which contradicts to the condition in line 5.

Next we show part ii). Suppose $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}') \neq \varnothing$. Since we have shown part i), we must have $\boldsymbol{a}' < \boldsymbol{a}^e$. Then for each $i \in \mathcal{N}_1(\boldsymbol{a}') \subset \mathcal{N}_1(\boldsymbol{a}^e)$, we have $u_i(\boldsymbol{a}') \le u_i(\boldsymbol{a}^e)$. Since $u_i(\boldsymbol{a}') = 0 = u_i(\boldsymbol{a}^e)$ for each $i \in \mathcal{N}_0(\boldsymbol{a}') \cap \mathcal{N}_0(\boldsymbol{a}^e)$, there must exist $i \in \mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}')$ such that $u_i(\boldsymbol{a}^e) < u_i(\boldsymbol{a}') = 0$. Suppose $i$ is included in $\mathcal{N}_{II}$ during phase II of some round. Let $\bar{\boldsymbol{a}}$ be the profile right before $a_i = 1$ is performed. Since $\bar{\boldsymbol{a}} \le \boldsymbol{a}^e$, we have $u_i(\bar{\boldsymbol{a}}) \le u_i(\boldsymbol{a}^e) < 0$. Then it follows from $0 \le f_i(1,\bar{\boldsymbol{a}}_{-i}) - f_i(0,\bar{\boldsymbol{a}}_{-i}) = u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S} s_{ij}\bar{a}_j$ that there must exist $j \in \mathcal{N}_{i+}^S$ such that $\bar{a}_j = 1$, and therefore $a^e_j = 1$. If $j \in \mathcal{N}_1(\boldsymbol{a}')$, we have $u_j(\boldsymbol{a}^e) - u_j(\boldsymbol{a}') \ge a^e_i - a'_i = 1 > 0$, which is a contradiction. Therefore we must have $j \in \mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}')$ and $u_j(\boldsymbol{a}^e) \le u_j(\boldsymbol{a}') = 0$. Let $\hat{\boldsymbol{a}}$ be the profile right before $a_j = 1$ is performed. Since $j$ is included before $i$, we have $u_j(\hat{\boldsymbol{a}}) < u_j(\boldsymbol{a}^e) \le 0$. Then we can use the above argument sequentially, until we find some $k$ that leads to contradiction.

*Proof of Theorem 3*

Let $\mathcal{N}'_{I,k}$ be the set of users in $\mathcal{N}_{I,k}$ during the execution that computes $\boldsymbol{a}^{e'}$. For each $i \in \mathcal{N}_{I,1}$, we have

$$u_i(1,\boldsymbol{a}'_{-i}) + \sum_{j\in\mathcal{N}_{i+}^{S'}\setminus\overline{\mathcal{N}'}} s'_{ij}a_j \ge u_i(1,\boldsymbol{a}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^{S}\setminus\overline{\mathcal{N}}} s_{ij}a_j \ge 0.$$

Therefore we must have $\mathcal{N}_{I,1} \subseteq \mathcal{N}'_{I,1}$. Similarly, we can show that for any $i \in \mathcal{N}_{II,1} \setminus \mathcal{N}'_{I,1}$, we must have $i \in \mathcal{N}'_{II,1}$. Using this argument sequentially, we can show that $\cup_{i=1}^k (\mathcal{N}_{I,i} \cup \mathcal{N}_{II,i}) \subseteq \cup_{i=1}^k \left(\mathcal{N}'_{I,i} \cup \mathcal{N}'_{II,i}\right)$ for any $k$, and therefore $\boldsymbol{a}^e \le \boldsymbol{a}^{e'}$. When a user becomes determined with action 1, the increment of social welfare of determined users by changing its action from 0 to 1 is no less than the increment of its social group utility involving only determined users, which is non-negative. Therefore we can see that $v(\boldsymbol{a}^e) \le v(\boldsymbol{a}^{e'})$.

*Proof of Lemma 1*

Suppose $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}^*) \neq \varnothing$. Without loss of generality, we assume that $\boldsymbol{a}^*$ is unique. Suppose the first participating user among $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}^*)$ is in $\widehat{\mathcal{N}} \triangleq \mathcal{N}_{I,k} \setminus \mathcal{N}_1(\boldsymbol{a}^*)$. Let $\overline{\mathcal{N}}_{I,k}$ be the set of users in $\overline{\mathcal{N}}$ right before phase I of round $k$ and $\bar{\boldsymbol{a}}$ be the profile right after that phase. Define $\boldsymbol{a}' \triangleq \boldsymbol{a}^* \vee \bar{\boldsymbol{a}}$ where $\vee$ denotes element-wise "or" operation such that $\mathcal{N}_1(\boldsymbol{a}') = \mathcal{N}_1(\boldsymbol{a}^*) \cup \widehat{\mathcal{N}}$. Then we have

$$\begin{aligned} v(\boldsymbol{a}') - v(\boldsymbol{a}^*) &= \sum_{i\in\widehat{\mathcal{N}}} \left( u_i(1,\boldsymbol{a}'_{-i}) + \sum_{j\in\mathcal{N}_{i+}^P\setminus\widehat{\mathcal{N}}} a'_j \right) \\ &\ge \sum_{i\in\widehat{\mathcal{N}}} \left( u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^P\setminus\widehat{\mathcal{N}}} \bar{a}_j \right) \\ &\ge \sum_{i\in\widehat{\mathcal{N}}} \left( u_i(1,\bar{\boldsymbol{a}}_{-i}) + \sum_{j\in\mathcal{N}_{i+}^S\setminus\overline{\mathcal{N}}_{I,k}} s_{ij}\bar{a}_j \right) \ge 0 \end{aligned}$$

where the first inequality follows from $\bar{\boldsymbol{a}} \le \boldsymbol{a}'$, and the last inequality follows from $\widehat{\mathcal{N}} \subseteq \mathcal{N}_{I,k} \subseteq \overline{\mathcal{N}}_{I,k}$ and the condition in line 5. This contradicts that $\boldsymbol{a}^*$ is unique. Similarly, if the first participating user among $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}^*)$ participates in phase II of some round, we can also show a contradiction. Therefore we must have $\mathcal{N}_1(\boldsymbol{a}^e) \setminus \mathcal{N}_1(\boldsymbol{a}^*) = \varnothing$.

*Proof of Theorem 4*

From Lemma 1 we have $\boldsymbol{a}^* \geq \boldsymbol{a}^e$. Then we have

$$v(\boldsymbol{a}^*) - v(\boldsymbol{a}^e) = \sum_{i\in\mathcal{N}_\Delta}\left(u_i(1,\boldsymbol{a}^*_{-i}) + \sum_{j\in\mathcal{N}^P_{i+}\setminus\mathcal{N}_\Delta} a^*_j\right)$$

$$= \sum_{i\in\mathcal{N}_\Delta}\left(u_i(1,\boldsymbol{a}^e_{-i}) + \sum_{j\in\mathcal{N}^P_{i-}\cap\mathcal{N}_\Delta} a^*_j + \sum_{j\in\mathcal{N}^P_{i+}\setminus\mathcal{N}_\Delta} a^e_j\right)$$

$$\leq \sum_{i\in\mathcal{N}_\Delta}\left(u_i(1,\boldsymbol{a}^e_{-i}) + \sum_{j\in\mathcal{N}^S_{i+}} s_{ij}a^e_j + \sum_{j\in\mathcal{N}^P_{i-}\cap\mathcal{N}_\Delta} a^*_j + \sum_{j\in\mathcal{N}^S_{i+}\setminus\mathcal{N}_\Delta}(1-s_{ij})a^e_j\right)$$

$$= \sum_{i\in\mathcal{N}_\Delta}\left(f_i(1,\boldsymbol{a}^e_{-i}) - f_i(0,\boldsymbol{a}^e_{-i}) + \sum_{j\in\mathcal{N}^P_{i-}\cap\mathcal{N}_\Delta} a^*_j + \sum_{j\in\mathcal{N}^S_{i+}\setminus\mathcal{N}_\Delta}(1-s_{ij})a^e_j\right)$$

$$\leq \sum_{i\in\mathcal{N}_\Delta}\left(\sum_{j\in\mathcal{N}^P_{i-}\cap\mathcal{N}_\Delta} a^*_j + \sum_{j\in\mathcal{N}^S_{i+}\setminus\mathcal{N}_\Delta}(1-s_{ij})a^e_j\right)$$

$$\leq \sum_{i\in\mathcal{N}_\Delta}\sum_{j\in\mathcal{N}^P_{i-}\cap\mathcal{N}_\Delta} 1 + \sum_{i\in\mathcal{N}_\Delta}\sum_{j\in\mathcal{N}^S_{i+}\setminus\mathcal{N}_\Delta}(1-s_{ij}).$$

Define $\mathcal{N}'_\Delta \triangleq \mathcal{N}_1(\boldsymbol{a}^*) \setminus \mathcal{N}_1(\boldsymbol{a}^{e'})$ where $\boldsymbol{a}^{e'}$ is the Pareto-optimal SNE when the social ties increase (i.e., $s'_{ij} \geq s_{ij}$, $\forall i,j \in \mathcal{N}$). By Theorem 3, we have $\mathcal{N}'_\Delta \subseteq \mathcal{N}_\Delta$. Then we can observe that the bound decreases when the social ties increase.