Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 898761, 10 pages
http://dx.doi.org/10.1155/2014/898761

Research Article

Optimized Reputable Sensing Participants Extraction for Participatory Sensor Networks

Weiwei Yuan,[1] Donghai Guan,[1] and Yuanfeng Jin[2]

[1] *College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2] *College of Science, Yanbian University, Yanji 133002, China*

Correspondence should be addressed to Yuanfeng Jin; jinyuanfeng@gmail.com

Received 1 April 2014; Accepted 28 July 2014; Published 29 September 2014

Academic Editor: Lu Zhen

By collecting data via sensors embedded personal smart devices, sensing participants play a key role in participatory sensor networks. Using information provided by reputable sensing participants ensures the reliability of participatory sensing data. Setting a threshold for the reputation, and those whose reputations are bigger than this value are regarded as reputable. The bigger the threshold value is, the more reliable the extracted reputable sensing participant is. However, if the threshold value is too big, only very limited participatory sensing data can be involved. This may cause unexpected bias in information collection. Existing works did not consider the relationship between the reliability of extracted reputable sensing participants and the ratio of usable participatory sensing data. In this work, we propose a criterion for optimized reputable sensing participant extraction in participatory sensor networks. This is achieved based on the mathematical analysis on the ratio of available participatory sensing data and the reliability of extracted reputable sensing participants. Our suggested threshold value for reputable sensing participant extraction is only related to the power of sensing participant's reputation distribution. It is easy to be applied in real applications. Simulation results tested on real application data further verified the effectiveness of our proposed method.

1. Introduction

Combining the power of crowd, the ubiquitously available smart devices, and the high speed Internet, participatory sensor networks have raised more and more attention nowadays [1–3]. Participants use their personal smart devices, which have embedded sensors such as camera, microphone, GPS, accelerometer, digital compass, and gyroscope, to gather data and make them available for large-scale applications [4–6]. The performance and the efficiency of participatory sensing applications heavily depend on the quality of data contributed by the sensing participants [7, 8]. By providing data via their portable devices, any participant may join the participatory sensor networks at any time, ubiquitously. The data trustworthiness becomes more crucial than the traditional wireless sensor networks. It is vital to ensure the reliability of information in participatory sensor networks.

One way to ensure the reliability of participatory sensing data is to use the data from reputable sensing participants. The reputable sensing participants refer to those who have high reputation in participatory sensor networks. Reputation is what is generally said or believed about a person's or a thing's character or standing [9]. Reputation is closely related to trust, while they are two distinct concepts. Trust is the measure of willingness to believe in a user based on their competence and behavior within a specific time [10]. Trust reflects the user's personal view on others, while reputation measures the reliability from the global point of view; that is, it reflects the trustworthiness of the target user based on the opinions in some specific community.

The main challenge in extracting reputable sensing participants is to maximize the reliability of reputable sensing participants with maximum available participatory sensing data. The most straightforward way to extract reputable sensing participants is to set a threshold for the reputation. If a sensing participant's reputation is bigger than the threshold value, it is regarded as reputable. The bigger the threshold value is, the more reliable the reputable sensing participant

is. However, if the threshold value is set to be too big, only very limited sensing data can be involved. This may cause unexpected bias in the information collection of participatory sensing applications. So there should be some tradeoff between the reliability of the reputable sensing participants and the usable participatory sensing data.

To the best of our knowledge, no work has given clear criteria on how to extract reputable sensing participants in participatory sensor networks. Some existing works in other related areas [11] ambiguously identified the reputable users. For example, in case the reputation is represented by a number varying from 0 to 1, some works [12] choose those whose reputations are bigger than 0.7 or 0.8 as reputable users. They did not explain why the threshold value was chosen and whether this value is optimized to achieve maximum usable participatory sensing data.

To solve the problems of existing works, this paper contributes to propose a criterion for optimized reputable sensing participant extraction in participatory sensor networks. It is achieved via the mathematical analysis on the sensing participant's reputation distribution, the ratio of the available participatory sensing data, and the reliability of the extracted reputable sensing participants. Our suggested optimized threshold value for reputable sensing participant extraction is only related to the power of the sensing participant's reputation distribution. It is easy to be applied in the real applications. In addition, we give general guidance on optimized reputable sensing participant extraction for those whose sensing participant's reputation distribution is not known. It is suggested that, for the normalized reputation in the range varying from 0 to 1, the optimized reputation threshold value for reputable sensing participant extraction should be no more than 0.1. Otherwise, the ratio of available participatory sensing data is too low to be used. Simulation results tested on the real application data further verified that our proposed method can achieve maximum useable participatory sensor data with maximum reliability.

The following sections of this paper are organized as follows. Section 2 introduces the related works, Section 3 gives our proposed method on reputable sensing participant extraction, Section 4 gives the experimental results, and the last section concludes this paper and points out the future work.

2. Related Works

Reputation is closely related to trust, while there are clear differences between these two concepts. The differences between trust and reputation can be illustrated by the following distribution.

> "I trust you because of your good reputation" [13].
>
> "I trust you despite your bad reputation" [13].

Based on the above statements, we can see that the reputation describes the reliability of users from the global point of view. It can be regarded as a collective measure of trustworthiness based on the personal trust in a community [14]. If a user has high reputation in a community, it means general users in this community regard this user as the reliable user. If a user has low reputation in a community, it means significant number of users in this community do not trust this user. Reputation can relate to a group or to an individual [14]. A group's reputation be modeled as the average of all its members' individual reputation or as the average of how the group is perceived as a whole by external parties [13]. On the contrary, user trust describes the user reliability from the personal point of view. It can be derived from a combination of the objective reputation and personal experience [14]. Different users have their own tastes, so items liked by one user do not necessarily have to be liked by other users. Similarly, though some users are trusted by most users in the community, that is, these users have high reputations, they do not necessarily have to be regarded as reliable by all users: some users may not trust them for various reasons; for example, they have bad personal interaction experiences with these reputable users. Yet generally speaking, the users with higher reputations are more probable to be regarded trustworthy by others. This is because reputation reflects the general opinions, which are the opinions of most users in the community.

Reputation based mechanisms are widely used in the various applications, such as reputation based access control [15, 16], reputation based trust-aware recommender system [10], and reputation based risk management [17]. In these applications, reputation is used to measure the global trustworthiness of users or used together with other attributes to measure the personal trustworthiness of users for specific users. In these applications, information given by users with bigger reputation is regarded as more reliable and more acceptable by the target user, while the users with smaller reputation are regarded as less reliable and less acceptable.

Different research works have their own criteria to identify the reputable users [10]. A common method is to set a threshold value for the reputation to extract reputable users. If the reputation of a user is bigger than the threshold value, this user is regarded as a reputable user. Otherwise, this user is not regarded as a reputable user. To the best of our knowledge, no work has analyzed the ratio of available information by setting this threshold value. They just intuitively set some value for the threshold value [10]; for example, in [12], the authors mentioned that those whose reputations are bigger than 0.75 are regarded as reputable users, but it did not mention the ratio of useable information by setting this threshold value.

3. Proposed Method for Optimized Reputable Sensing Participant Extraction

Let $G = (V, E)$ be the participatory sensor network, in which V is the collection of participatory sensing participants and E is the collection of social relations between participatory sensing participants. For $v_i \in V$, let $I(v_i)$ represent v_i's indegree in the participatory sensor network, $I(v_i) = \sum_j e_{ji}$, $e_{ji} \in E$. The indegree of a participatory sensing participant in the participatory sensor network represents the number of trust this participatory sensing participant received from

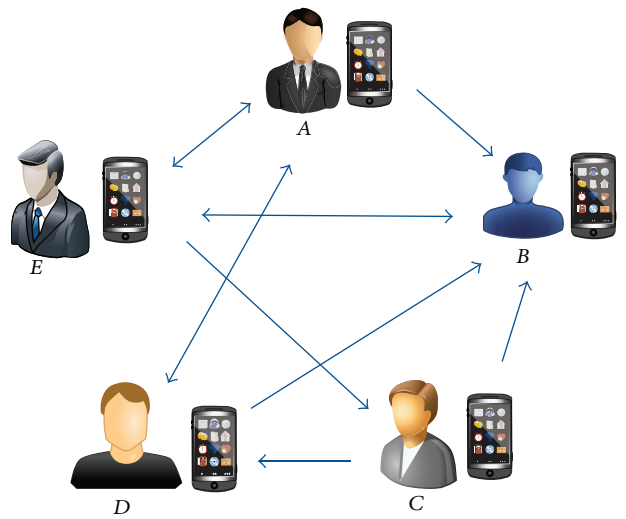


Figure 1: An example of the reputation calculation in the participatory sensor network.

other participatory sensing participants. The bigger indegree a participatory sensing participant has, the more participatory sensing participants trust it. So the indegree of participatory sensing participant represents its trustworthiness from the global point of view, that is, the general opinions of participatory sensing participants in the participatory sensor network. We therefore use the participatory sensing participant's indegree to represent its reputation in the participatory sensor network. For the convenience of comparison, we further normalize the reputation into the range varying from 0 to 1.

Let $R(v_i)$ be the reputation of v_i in G as follows:

$$R(v_i) = \frac{I(v_i)}{I_{\max}}, \tag{1}$$

in which $I_{\max}$ is the maximum indegree of participatory sensing participants in G, $R(v_i) \in [0, 1]$.

An example is given in Figure 1 to illustrate the calculation of reputation: there are five sensing participants involved in the participatory sensor network. The indegree of each sensing participant is equal to the edges pointing to them; that is, $I(A) = 2$, $I(B) = 4$, $I(C) = 1$, $I(D) = 2$, $I(E) = 2$. Using (1), the reputation of each sensing participant in the participatory sensor network is $R(A) = 0.5$, $R(B) = 1$, $R(C) = 0.25$, $R(D) = 0.5$, $R(E) = 0.5$.

We have verified in our previous work [10] that the trust network is the scale-free network. The scale-free network is the network whose degree distribution follows the power law; that is, $P(k) \sim k^{-\gamma}$, where $P(k)$ is the probability that a randomly selected node has k connections and γ is the power of the degree distribution. The most notable characteristic in the scale-free network is the existence of node whose degrees greatly exceed the average. These highest degree nodes are often called "hubs." Though the number of hubs is limited, they dominate the connectivity of the scale-free network. The comparison between the structure of the random network and the structure of scale-free network is given in Figure 2.

Based on the properties of the scale-free network, the sensing participant's indegree distribution can be represented as

$$f(I(v_i)) = aI(v_i)^{-\gamma}, \tag{2}$$

in which $f(\cdot)$ is the distribution of $I(v_i)$, γ is the power, and a is the coefficient.

Using $R(v_i)$ to substitute $I(v_i)$ in (2), we get

$$f(R(v_i)) = a'R(v_i)^{-\gamma}. \tag{3}$$

The power of the reputation distribution, that is, γ, is the same as that of the indegree distribution, and only the coefficient, that is, a', is different from that of the indegree distribution shown in (2).

For the convenience of representation, let $x = R(v_i)$. Let $g(\cdot)$ be the ratio of involved sensing participants, that is, the useable participatory sensing data in participatory sensor networks. It can be calculated as

$$g(x) = \int_x^1 f(x)\,dx = \int_x^1 a'x^{-\gamma}dx = \frac{a'}{1-\gamma}\left(1 - x^{1-\gamma}\right). \tag{4}$$

The goal of this work is to maximize the reliability of extracted reputable sensing participants with maximum usable participatory sensing data. Let x be the threshold value for reputable participatory sensing participant extraction. By increasing x, the reliability of extracted reputable participatory sensing participant is increasing, while the involved information given by extracted reputable participatory sensing participant is decreasing. So we further optimize this procedure as follows.

Let $h(x)$ be the reliability of extracted reputable participatory sensing participant as follows:

$$h(x) = cx, \quad c > 0, \tag{5}$$

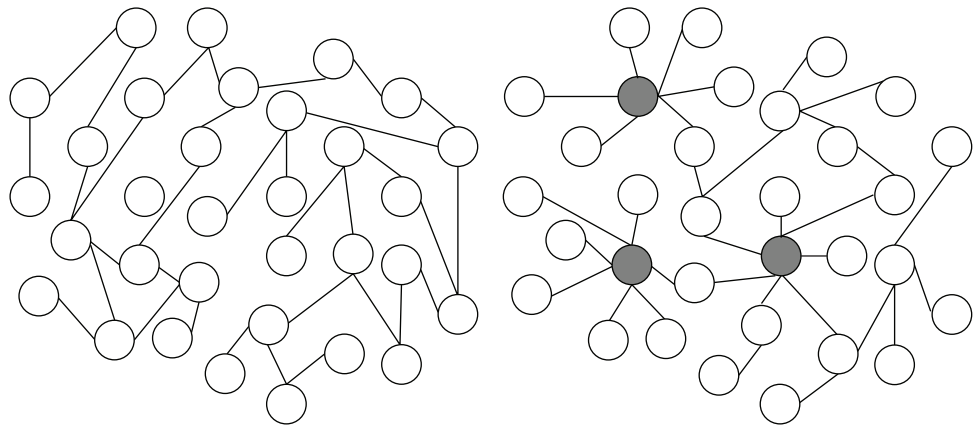
in which x is the threshold value for reputable participatory sensing participant extraction, and c is a constant. The bigger the value of x is, the more reliable the extracted reputable participatory sensing participants are.

We set a fitness function for the optimization as follows:

$$\text{fitness}(x) = h(x) * g(x). \tag{6}$$

So the goal is to find the maximum value of fitness(x).

There are various methods to calculate the maximum value of a function. In this work, we choose the most popular



(a) (b)

Figure 2: Comparison between the structure of (a) random network and (b) scale-free network. In the scale-free network, the hubs are highlighted with black nodes.

one, that is, via calculating the derivative of a function and finding the extrema. The calculation is as follows:

$$\begin{aligned}
&\text{fitness}(x)' \triangleq 0 \\
&\quad \Longrightarrow \left(h(x) * g(x)\right)' = 0 \\
&\quad \Longrightarrow \left(cx \int_x^1 a' x^{-\gamma} dx = \right)' = 0 \\
&\quad \Longrightarrow ca' \int_x^1 x^{-\gamma} dx - ca' x x^{-\gamma} = 0 \\
&\quad \Longrightarrow \frac{1}{1-\gamma}\left(1 - x^{1-\gamma}\right) - x^{1-\gamma} = 0
\end{aligned} \tag{7}$$

$$\begin{aligned}
&k \triangleq 1 - \gamma \\
&\frac{1}{k}\left(1 - x^k\right) - x^k = 0 \\
&\quad \Longrightarrow (k+1)\, x^k = 1 \\
&\quad \Longrightarrow x = \left(\frac{1}{k+1}\right)^{1/k} \\
&\quad \Longrightarrow x = \left(\frac{1}{2-\gamma}\right)^{1/(1-\gamma)}, \quad 1 < \gamma < 2.
\end{aligned}$$

Since the functions in (2) and (3) approximately follow the participatory sensing participant's indegree distribution and reputation distribution, we further adjusted the value in (7), letting it be closer to the value in real cases. The reputation threshold value for reputable sensing participant extraction is adjusted as

$$x = c\left(\frac{1}{2-\gamma}\right)^{1/(1-\gamma)}, \quad 1 < \gamma < 2, \tag{8}$$

in which c is a constant.

Table 1: The basic information of the simulation data.

ID	Network name	Number of nodes	Number of edges	Average degree
1	CA_AstroPh	18772	396089	21.10
2	CA_CondMat	23133	186915	8.08
3	CA_GrQc	5242	28988	5.53
4	CA_HepPh	12008	237038	19.74
5	Epinions	75879	509148	6.71
6	P2p_Gnutella05	8846	31846	3.60
7	P2p_Gnutella08	6301	20793	3.30
8	P2p_Gnutella09	8114	26046	3.21
9	Slashdot	77360	905886	11.71

In our proposed method, the computational complexity of reputable sensing participant extraction is $o(c)$, in which c is a constant. By using our proposed method, it is easy to estimate the optimized value for reputable sensing participant extraction. For real applications with different scales, the only parameter one needs to know is γ, that is, the power of the sensing participant's reputation distribution or the power of the sensing participant's indegree distribution in the participatory senor network.

4. Experimental Results

Nine networks extracted from the real applications are used to simulate the sensing participant network. Experiments are held on these networks to verify the effectiveness of our proposed idea. The basic information of these networks, including the number of nodes, the number of edges, and the average degree of these networks, is given in Table 1. Each node in the simulation data is used to simulate the sensing participant, and the edges in the simulation data are used to simulate the social ratios between the sensing participants in participatory sensor networks. CA_AstroPh dataset is

extracted from Arxiv ASTRO-PH (AstroPhysics) collaboration network. CA_CondMat dataset is extracted from Arxiv COND-MAT (Condense Matter Physics) collaboration network. CA_GrQc dataset is extracted from Arxiv GR-QC (General Relativity and Quantum Cosmology) collaboration network. CA_HepPh dataset is extracted from Arxiv HEP-PH (High Energy Physics-Phenomenology) collaboration network. These four datasets are from the e-print arXiv and cover scientific collaborations between authors' papers submitted to AstroPhysics category, Condense Matter Category, General Relativity and Quantum Cosmology category, and High Energy Physics-Phenomenology category. In these four applications, if an author i coauthored a paper with author j, the graph contains an undirected edge from i to j. If the paper is coauthored by k authors this generates a completely connected subgraph on k nodes. All these four datasets cover papers in 124 months, representing essentially the complete history of their specific section. Epinions dataset is a who-trust-whom online social network of a general consumer review site (Epinions.com). Members of the site can decide whether to "trust" each other. All the trust relationships interact and form the Web of Trust which is then combined with review ratings to determine which reviews are shown to the user. P2p_Gnutella05 dataset, P2p_Gnutella08 dataset, and P2p_Gnutella09 dataset are extracted from a sequence of snapshots of the Gnutella peer-to-peer file sharing network. There are a total of 9 snapshots of Gnutella network. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts. The difference between these three datasets is that they are collected in different time period. Slashdot dataset is extracted from a technology-related news website known for its specific user community. The website features user-submitted and editor-evaluated current primarily technology-oriented news. Slashdot has Slashdot Zoo feature which allows users to tag each other as friends or foes. More details of these datasets can be found in [18].

To use our proposed method on reputable sensing participant extraction, we first evaluate the reputation distribution of each simulated sensing participant network. The experimental results are shown in Figure 3: the sensing participant's reputation distribution follows the power law, that is, the distribution given in (3). The power law distribution which approximately follows the sensing participant's reputation distribution is given in each subfigure of Figure 3. The reputations of most sensing participants in participatory sensor networks are small. The number of sensing participants who have high reputations in participatory sensor networks is very limited. In addition, the simulation results show that not only the number of sensing participants with high reputation but also the number of sensing participants with median reputation is very limited. For all 9 simulation datasets, the ratio of sensing participants whose reputation is bigger than 0.2 is very low. For some networks, such as Epinions, CA_CondMat, CA_HepPh, and Slashdot, even the ratio of sensing participants whose reputation is bigger than 0.1 is very limited. This is very different from the intuitive thought in some reputation based applications [6, 12]: they usually use those whose reputation is big in their model. However, the sensing participant's reputation distribution clearly shows that this is not appropriate in the real applications. Otherwise, only very limited information is available for use.

Table 2: The power of reputation distribution in simulation networks.

ID	Network name	Power of reputation distribution
1	CA_AstroPh	1.843
2	CA_CondMat	1.949
3	CA_GrQc	1.842
4	CA_HepPh	1.44
5	Epinions	1.541
6	P2p_Gnutella05	1.914
7	P2p_Gnutella08	1.788
8	P2p_Gnutella09	1.854
9	Slashdot	1.671

Based on the simulation results shown in Figure 3, the powers of the reputation distributions in simulation networks are summarized in Table 2. As shown in (8), we only consider the networks whose power of reputation distribution is between 1 and 2. We have verified in our previous work [11, 12] that the indegree distribution of most available trust networks is within this range. Trust is a kind of social relations between users. It is analogous to the relation between sensing participants in participatory sensor networks. So we regard the power of reputation distributions of most participatory sensor networks as within the range from 1 to 2. In our 9 simulation datasets, CA_CondMat has the biggest power of reputation distribution, while CA_HepPh has the smallest power of reputation distribution. Specifically, the power of reputation distribution is mainly within the range varying from 1.5 to 2.

As shown in (4), the ratio of useable information is based on the threshold value set for reputable sensing participant extraction. The bigger the threshold value is, the less the information available for use is. The relationship between the threshold value set for reputable sensing participant extraction and the ratio of usable information is given in Figure 4 for our simulation datasets. It is shown that if this threshold value is big, information available for use is very limited. For example, if the threshold value is 0.7, which is usually regarded as appropriate when selecting a big value in the range varying from 0 to 1, the ratio of useable information is less than 0.1% in all simulation datasets. Even if the threshold value is 0.4, which is not a very big value in the institutive thought, the ratio of useable information is less than 1% in all simulation datasets. Based on simulation results shown in Figure 4, it is suggested that the threshold value for reputable sensing participant extraction should not be big or even should not be a median value. Otherwise, only very limited information is available. Adequate information helps to reflect the real reliability of sensing participants from different aspects of views. So it is not recommended to filter out too much information. As shown in Figure 4, even if the threshold value for reputable sensing participant extraction is 0.1, the maximum ratio of available information

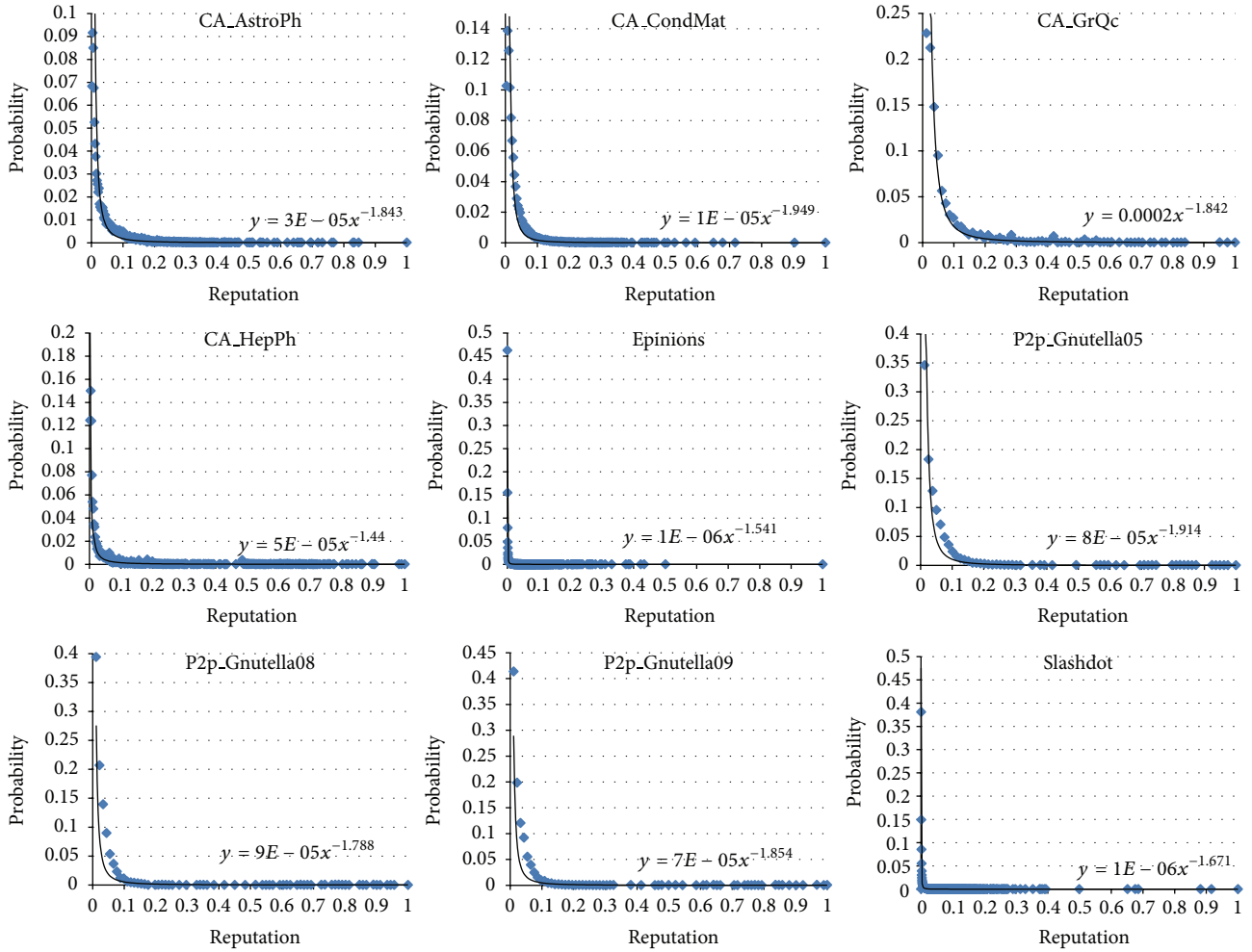


Figure 3: The reputation distribution of simulation networks.

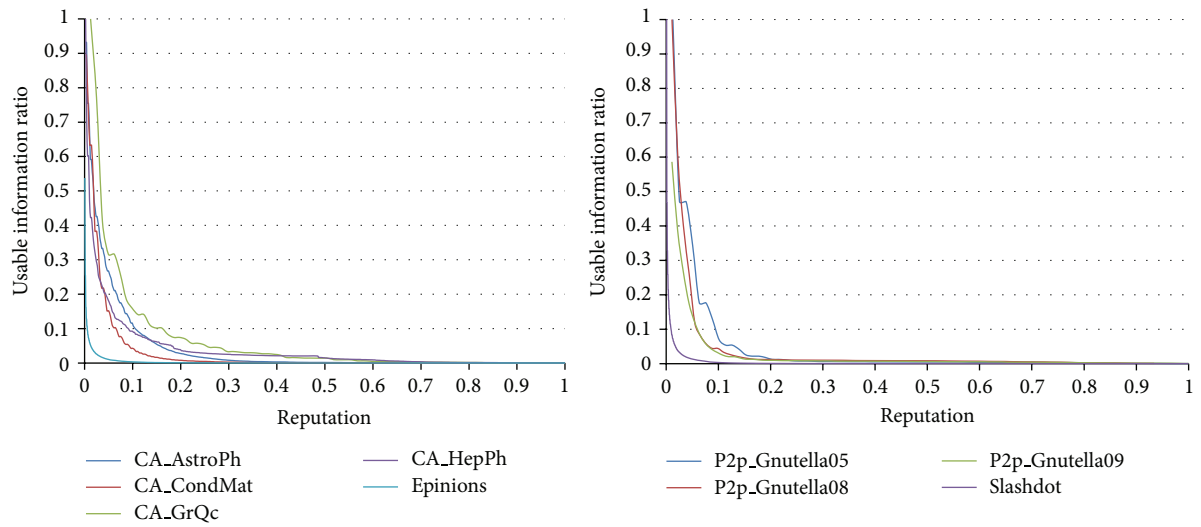


Figure 4: The relationship between reputation and the ratio of usable information.

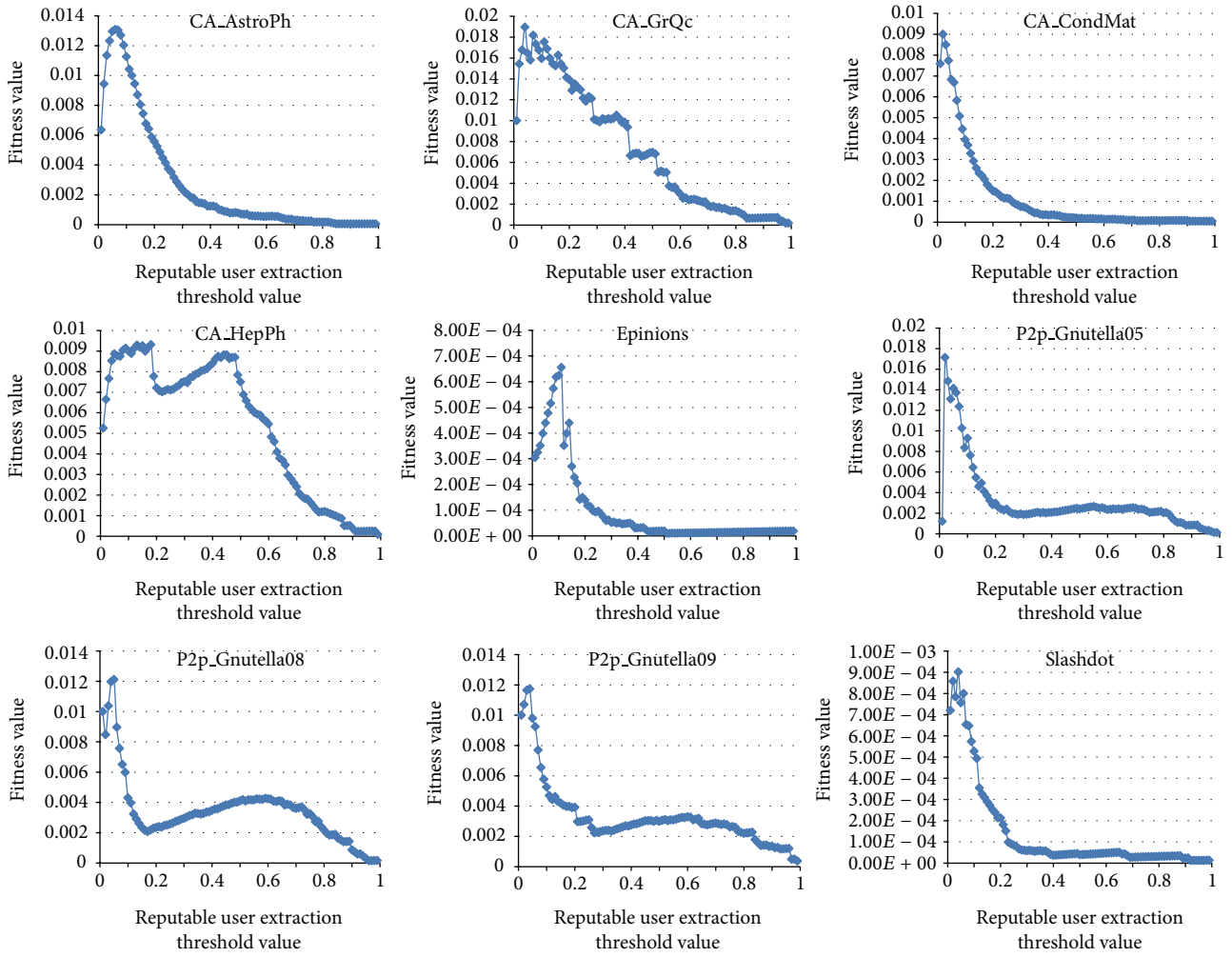


Figure 5: The relationship between reputation used for reputable sensing participant extraction and fitness value.

is less than 20% in simulation networks. If the threshold value for reputable sensing participant extraction is 0.2, available information is less than 10% in all 9 networks of our simulations. For example, for Epinions dataset, if the threshold value is set to be 0.1, available information is only around 1%.

As shown in (5), as the reputable sensing participant extraction threshold increases, the reliability of extracted reputable sensing participants increases. However, as shown in Figure 4, the information available for use decreases significantly with the increase of this threshold value. The fitness function in (6) is the product of user reliability and usable information. Figure 5 gives the fitness values of our simulation networks. The goal of our proposed idea is to find the proper threshold value which can achieve the maximum fitness values, that is, to achieve the maximum sensing participant reliability with maximum available information. As shown in Figure 5, for all 9 simulation networks, the maximum value of fitness function appears when the threshold value is small. In case the threshold value is big, for example, more than 0.6, all the fitness values decrease with the increasing of reputation values. In some simulation networks, for example, CA_HepPh and Epinions, in case the threshold value for reputable sensing participant extraction is a median value, though there exist some peak values for the fineness function, the fitness function only achieves local optimum. For all 9 simulation networks, the fitness functions achieve the global optimum when the reputation threshold value is small. This means it is possible to achieve the maximum usable information with the maximum reliability of reputable sensing participants if the reputation of sensing participants is not strictly controlled. This is very different from what was institutively thought in some existing works [11, 12], in which reputation of sensing participants was very strictly controlled; that is, they only regarded those who had extremely high reputation values as reputable. However, based on simulation results shown above, this is not suitable for real applications. This is because the reputation of most sensing participants in the participatory sensor network is very small, even if we set a small threshold value for reputable sensing participant extraction, significant number of sensing participants is filtered out. In this case, if we raise the threshold value for reputable sensing participant extraction, less and less information will be available.

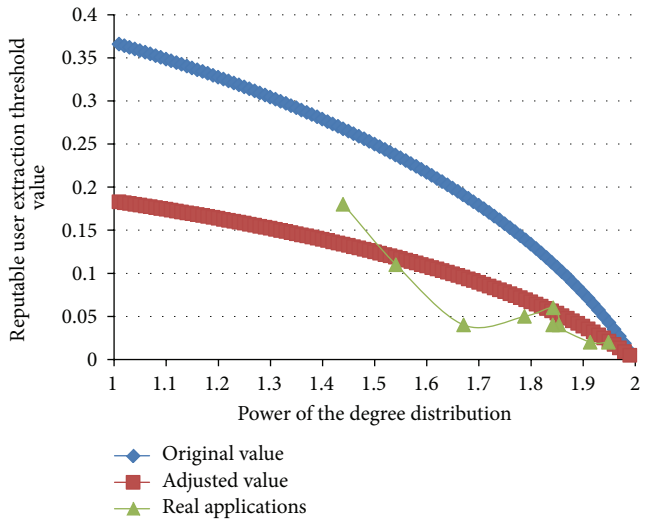


Figure 6: The optimized threshold value for reputable sensing participant extraction.

The statistics of the optimized value for reputable sensing participant extraction is given in Figure 6. The original value means the value calculated by (7). The adjusted value means the value calculated by (8), in which c is set to be 0.5 in this work. This constant value is chosen based on simulation results shown in Figure 4: if the threshold value for reputable sensing participant extraction is too big, very limited information can be involved; specifically, we regard 0.2 as an important value for this threshold value. If the threshold value is bigger than 0.2, no network in our simulation data can achieve more than 10% information, which is too limited in real applications. So we suggest that the threshold value for reputable sensing participant extraction should be no more than 0.2 to ensure the adequacy of usable information. Based on the distribution of the original value, setting c in (7) as 0.5 can ensure that the maximum threshold value for reputable sensing participant extraction is no more than 0.2. By the experiments held in Figure 5, we get the optimized threshold value for reputable sensing participant extraction, which is presented in Figure 6. As the power of sensing participant's reputation increases, the optimized threshold value for reputable sensing participant extraction tends to be decreasing. For our 9 simulation datasets, the optimized threshold value is concentrated in the range from 0 to 0.2. Specifically, the optimized threshold value for 7 out of the 9 simulation datasets is less than 0.05. This means to achieve the maximum reliability of extracted reputable sensing participants and the maximum usable information, we need to set a very small value for the reputation threshold value for reputable sensing participant extraction. The reliability of extracted reputable sensing participants may be far from our traditional expectation. For example, in CA_AstroPh network, the optimized threshold value for reputable sensing participant extraction is 0.06. Since reputation is in the range varying from 0 to 1, those whose reputations are bigger than 0.06 really cannot be thought to be reputable in traditional works. However, based on the sensing participant's reputation distribution, by setting the reputation threshold value for reputable sensing participant extraction to be 0.06, around 80% sensing participants cannot be involved in further operations. This small reputation threshold value can ensure those whose reputations are within top 20% are involved. In addition, this also ensures sufficient information is involved in participatory sensor networks. Based on simulation results shown in Figure 6, it is clear that our proposed method can approximately match the trend of the optimized reputation threshold value for reputable sensing participant extraction.

In some real applications, if the sensing participant's reputation distribution is not known, it is hard to know the parameter γ in (3). So we further estimate the general optimized reputation threshold value for reputable sensing participant extraction based on simulation results shown in this section. We suggest that the optimized reputation threshold value for reputable sensing participant extraction should be less than 0.1, at least no more than 0.2. This would ensure to involve the sensing participants whose reputations are within top 20%. Big reputation threshold value is not suggested to be set for reputable sensing participant extraction. Sensing participant's reputation does not follow liner distribution but follows the power law distribution. The ratio of available information decreases significantly with the increase of the reputation threshold value for reputable sensing participant extraction. Setting a big reputation threshold value would probably reduce the usable data, especially if the scale of participatory sensor network is not very large.

5. Conclusions and Future Works

The information reliability is the key to provide better services in participatory sensor network. One way to ensure the information reliability is to use information provided by reputable sensing participants. Reputation represents the reliability of sensing participants from general point of view, that is, general opinions of sensing participants in participatory sensor network. The sensing participants with bigger reputation are more reliable than those with small reputation value. So a threshold value can be set to extract reputable sensing participants, regarding all those whose reputation is bigger than this threshold value as reputable. Existing works just intuitively set this threshold value, letting it be a big value. For example, in case reputation is normalized to the range varying from 0 to 1, the threshold value is set to be 0.75 [12]. However, as the value of this reputation threshold value increases, the ratio of available participatory sensing data is decreasing significantly. Existing works did not consider the usable information by setting this threshold value. We argue that the reputation threshold value for reputable sensing participant extraction should not be a big value. Otherwise, very limited information will be available for use in participatory sensing networks. By increasing the reputation threshold value for reputable sensing participant extraction, the reliability of extracted sensing participants is increasing, while the ratio of available information is decreasing. So the reputation threshold value should be

optimized to achieve the maximum usable information with the maximum reputable sensing participant reliability.

In this work, we analyze sensing participant's reputation distribution and mathematically calculate the ratio of usable information by setting different reputation threshold value. To get the optimized threshold value on reputable sensing participant extraction, we set a fitness function, letting it be the product of sensing participant's reliability and the ratio of available information. Our goal is to get the threshold value which leads to the maximum fitness value. In this case, the reliability of sensing participant and the ratio of usable information are both maximized. By calculating the derivative of fitness function, we get the criteria on setting the reputation threshold value. The optimized reputable sensing participant extraction threshold value is only related to the power of sensing participant's reputation distribution (note that the power of sensing participant's reputation distribution should be in the range varying from 1 to 2 for our proposed method). Our proposed method is very efficient: its computational complexity is only $o(c)$. Nine networks extracted from real applications are used to analog the participatory sensor network and verify the effectiveness of our proposed method. The simulation results show that our proposed idea can approximately estimate the optimized threshold value on reputable sensing participant extraction. We also give general suggestions on setting the optimized reputable sensing participant threshold value based on the simulation results. Firstly, for the reputation value ranging from 0 to 1, the reputation threshold value for reputable sensing participant extraction should be no more than 0.2. Otherwise, the ratio of available information is too limited. Secondly, for reputation value ranging from 0 to 1, it is suggested that the optimized reputation threshold value for reputable sensing participant extraction should be around 0.1. In this case, the participatory sensor network can achieve enough information and the reputation of involved sensing participant is within top 20%.

We plan to focus on more details of reputable sensing participant extraction in the future work. In this work, the proposed method is for those whose power of sensing participant's reputation distribution is varying from 1 to 2. Though it has been shown [12] that most participatory sensing networks may satisfy this constraint, we will try some more methods to fulfill the requirement of other networks. We will use some optimization algorithms such as genetic algorithm and simulated annealing. By using these optimization algorithms, we can optimize reputable sensing participant extraction for more participatory sensor networks. However, the computational complexity by using these methods is much higher than that of our proposed method in this work. There should also be some tradeoff on the performance of reputable sensing participant extraction and the computational complexity.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 61100007, 61100081, and 11361066) and the collaborative research project under NSFC-NRF cooperative program (Grant no. 613111015).

References

[1] E. D. Cristofaro and C. Soriente, "Participatory privacy: enabling privacy in participatory sensing," *IEEE Network*, vol. 27, no. 1, pp. 32–36, 2013.

[2] H. Amintoosi and S. S. Kanhere, "A trust-based recruitment framework for multi-hop social participatory sensing," in *Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '13)*, pp. 266–273, IEEE, 2013.

[3] A. Manzoor, M. Asplund, and M. Bouroche, "Trust Evaluation for Participatory Sensing," in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pp. 176–187, Springer, Berlin, Germany, 2013.

[4] K. L. Huang, S. S. Kanhere, and W. Hu, "A privacy-preserving reputation system for participatory sensing," in *Proceedings of the 37th Annual IEEE Conference on Local Computer Networks (LCN '12)*, pp. 10–18, Clearwater, Fla, USA, October 2012.

[5] G. Sagl, T. Blaschke, E. Beinat, and B. Resch, "Ubiquitous geo-sensing for context-aware analysis: exploring relationships between environmental and human dynamics," *Sensors*, vol. 12, no. 7, pp. 9800–9822, 2012.

[6] I. Krontiris and N. Maisonneuve, "Participatory sensing: the tension between social translucence and privacy," in *Trustworthy Internet*, pp. 159–170, Springer, Milan, Italy, 2011.

[7] T. H. Silva, P. O. S. Vaz De Melo, J. M. D. Almeida, and A. A. F. Loureiro, "Uncovering properties in participatory sensor networks," in *Proceedings of the 4th ACM International Workshop on Hot Topics in Planet-Scale Measurement (HotPlanet '12)*, pp. 33–38, ACM, June 2012.

[8] S. S. Kanhere, "Participatory sensing: crowdsourcing data from mobile smartphones in urban spaces," in *Distributed Computing and Internet Technology*, pp. 19–26, Springer, Berlin, Germany, 2013.

[9] A. Jøsang, "Robustness of trust and reputation systems: does it matter?" in *Trust Management VI*, pp. 253–262, Springer, Berlin, Germany, 2012.

[10] W. Yuan, D. Guan, Y.-K. Lee, and S. Lee, "The small-world trust network," *Applied Intelligence*, vol. 35, no. 3, pp. 399–410, 2011.

[11] W. Yuan, D. Guan, Y.-K. Han, S. Lee, and Y.-K. Lee, "More reputable recommenders give more accurate recommendations?" in *Proceeding od the 7th International Conference on Ubiquitous Information Management and Communication (ICUIMC '13)*, New York, NY, USA, January 2013.

[12] W. Yuan, L. Shu, H.-C. Chao, D. Guan, Y.-K. Lee, and S. Lee, "iTARS: trust-aware recommender system using implicit trust networks," *IET Communications*, vol. 4, no. 14, pp. 1709–1721, 2010.

[13] T. Bhuiyan, A. Josang, and Y. Xu, "Trust and reputation management in web-based social network," in *Web Intelligence and Intelligent Agents*, pp. 207–232, InTech, Rijeka, Croatia, 2010.

[14] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[15] H. Hu, G.-J. Ahn, and J. Jorgensen, "Multiparty access control for online social networks: model and mechanisms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1614–1627, 2013.

[16] A. Ahmed and A. Alnajem, "Trust-aware access control: How recent is your transaction history?" in *Proceedings of the 2nd International Conference on Digital Information and Communication Technology and its Applications (DICTAP '12)*, pp. 208–213, May 2012.

[17] S. Zhao, G. Wu, G. Chen, and H. Chen, "Reputation-aware service selection based on QOS similarity," *Journal of Networks*, vol. 6, no. 7, pp. 950–957, 2011.

[18] http://snap.stanford.edu/data.