Worker Recruitment Strategy for Self-Organized Mobile Social Crowdsensing

En Wang*, Yongjian Yang*, Jie Wu, *IEEE Fellow*†, Dongming Luan*, Hengzhi Wang*
*Department of Computer Science and Technology, Jilin University, Changchun, China
†Department of Computer and Information Sciences, Temple University, Philadelphia, USA.

***Abstract*—Mobile crowdsensing recruits a massive group of mobile workers to cooperatively finish a sensing task through their smart devices (mobile phones, ipads, *etc*.). In this paper, the communication in social network for delivering the sensing data of mobile crowdsensing is considered, where some requesters publish the sensing tasks to all the Point of Interests (PoIs), and the workers are recruited to take the sensing data in the PoI until they could communicate with the requester through an offline and online connection. We first use the semi-Markov model to predict the offline encounter situation. Then, the worker's utility is decided by both the offline encounter and social connection probabilities. The Worker Recruitment for Self-organized MSC (WEO) is further presented through recruiting a set of workers, who have the maximum communication probability with the requesters. We prove that the optimal recruitment problem is NP-hard, and we introduce a practical greedy heuristic method for this problem, the performance of the greedy method is also tested. Two real-world traces, *roma/taxi* and *epfl* are tested in our simulations, where WEO always achieves the highest delivery ratio of sensing tasks among different recruitment strategies.**

***Index Terms*—Worker recruitment, Self-organized, Mobile social crowdsensing**

I. Introduction

Recently, a popular sensing scheme, *mobile crowdsensing* [1] has attracted the attention of researchers. This scheme recruits some mobile workers to coordinately perform a complex sensing task through their equipped devices. These devices are widely embedded of a variety of sensors (e.g., temperature sensing element, humidity sensing element, and acceleration transducer) as well as a high level computing ability. In addition to addressing the sensing data, some new-scheme services are born as traffic predictions, finding parking spots, air quality monitoring, etc [2, 3].

Recently, selecting suitable user set [4], [5], [6] and encouraging them to participate in MCS [7], [8], [9] are the two main researching parts. The most important challenge in terms of mobile crowdsensing is to decide a suitable set of workers who can contribute most (data quality, sensing time, *etc.*) to finish the sensing task. As a result, worker recruitment is most discussed and also important topic. However, previous works mainly focus on recruiting workers according to the offline encounter among devices, while ignoring communication through the online social network. This is not comprehensive because online communication is also an efficient way to exchange the stored data. Hence, we pay attention to a self-organized MSC, where the sensing tasks are published in each PoI, and the workers are recruited in the PoIs, in order to take

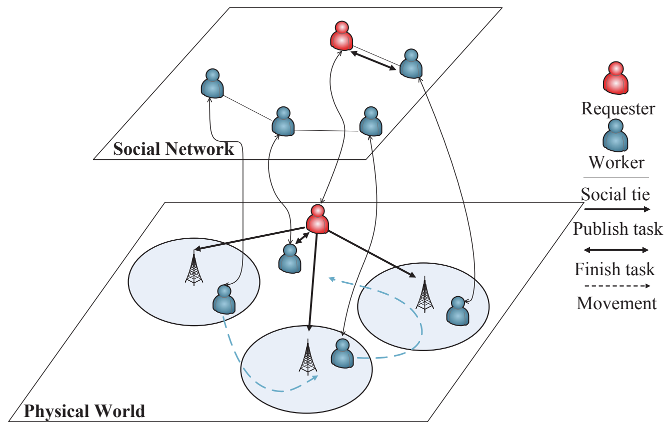


Fig. 1. Illustration of worker recruitment problem in mobile social crowdsensing.

the sensing data until they could communicate both online and offline [10] with the requester. We use the Semi-Markov model to calculate the encounter situation of the physical world and regard the connection in two hops as the online communication probability to measure the transmission ability. Then, we prove that the worker recruiting problem is NP-hard. Subsequently, a practical greedy heuristic solution is proposed in this paper for solving the recruiting problem.

Fig. 1 illustrates the worker recruitment strategy in MSC, in the double-layer network structure, a set of task requesters R jointly launch some tasks for sensing the data in each PoI (e.g., the task is to test the air quality in each PoI area). Supposing that a task is assigned to the PoIs, and some workers move among the PoIs, our purpose is to select some workers to sense the data (i.e, test air quality) and deliver the air quality data to one of the requesters. The delivery could be achieved through the following two ways: (1) the workers with the sensing data accidentally encounter one of the requesters in the physical world (e.g., Bluetooth, WiFi); or (2) the workers communicate with the requesters through the social network (e.g., Wechat, Twitter). The sensing task for a PoI is finished if and only if any worker recruited by this PoI could communicate (online or offline) with one of the requesters before the deadline of the sensing task.

For solving the above recruiting problem, the important thing is selecting a group of suitable workers for the task considering both the encounter probability in the offline physical

978-1-5386-5156-8/18/$31.00 ©2018 IEEE

world and connection probability in the online social network. This problem is challenging because,

1) When recruiting the workers in a PoI we could not know the detailed positions and friendships of the requesters; hence, it is difficult for us to predict the communication probabilities (online and offline) between the workers and requesters.
2) It is hard for us to decide which worker to be recruited taking both the online and offline communication probabilities into consideration.
3) The optimization problem of recruiting a set of best workers is NP-hard. We need to present a greedy algorithm for solving this problem.

In order to overcome the above challenges, we try to solve two problems: the first one is deciding a comprehensive probability, in which Semi-Markov is used to calculate the offline encounter probability between the recruiters and the requesters; then, we predict the social communication probability of them through the two-hop social relationship. The other problem is an NP-hard problem, where we attempt to recruit a set of best workers who have the highest communication probability with the requesters. Then, we introduce a practical greedy heuristic method for the problem, the performance of the greedy method is also tested. Finally, two reality traces, *roma/taxi* and *epfl* are tested in our simulations, where WEO always achieves the highest delivery ratio of sensing tasks among different recruitment strategies.

The main contributions are briefly summarized as follows:

- In an offline mobile network, we use a PoIs-based prediction method that predicts the inter-worker contact probability.
- We propose a calculation method for a worker's utility, which consists of online communication probability and offline contact probability.
- Based on the worker's utility, we propose a worker recruitment strategy for self-organized mobile social crowdsensing, which adopts a practical greedy heuristic for solving the NP-hard recruiting problem.
- Two real-world traces: *roma/taxi* and *epfl* are tested in our simulations, where WEO always achieves the highest delivery ratio of sensing tasks among different recruitment strategies.

The rest parts are organized as follows: The problem modeling and formulating are presented in Section II. The online social connecting probability and offline encounter probability are described in Section III. The main proposed strategy WEO is shown in Section IV, and the NP-hard problem is also solved in this section. In Section V, we evaluate the delivery ratio and delay of WEO. We review the related work in Section VI, while Section VII summarizes this paper.

II. PROBLEM MODELING AND FORMULATING

A. Problem Model

A self-organized MSC is taken into consideration, which includes a group of mobile workers $W = \{w_1, w_2, \cdots, w_n\}$. Some workers could jointly change to be requesters and publish some tasks, then the workers are denoted by the requester set $R = \{r_1, r_2, \cdots, r_m\}, m \leq n$. Workers could be regarded as the collectors of the sensing data. When they finish collecting the sensing data, they could jointly compromise the sensing data and upload them to the task requester. The upload process could be done through both online and offline networks, which are described in the previous section. In social networks, each worker has some social friends, which are also in W. In the physical world, some PoIs: $L = \{1, 2, \cdots, l\}$ exist in the map. If a worker meets a requester in a PoI, the worker can upload the stored data, then the sensing task is successfully finished. Here, the data size for publishing the sensing task is considered as an acceptable cost for a requester, so they publish the task to each PoI through 4G. However, the size for uploading sensing data is not an acceptable cost for a worker; hence, they upload the data to a requester through WiFi APs.

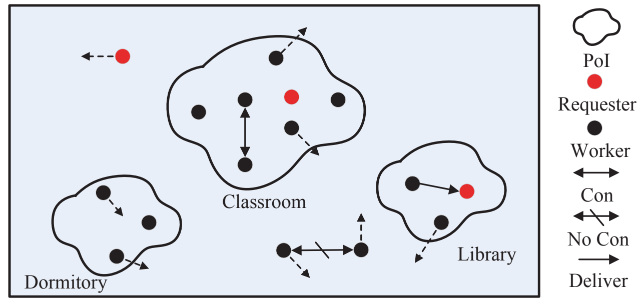


Fig. 2. The physical world of self-organized MSC.

In this model, when two workers are in the same PoI or they have a social relationship, they could be regarded as being in contact. Two workers cannot communicate with each other when one of them is not in a PoI. The self-organized physical world network is shown in Fig. 2. There are three PoIs in the map and when a worker and a requester are in the same PoI, the worker could finish the sensing task successfully. Due to the reason that we pay attention to the connection condition, the communication duration and bandwidth are assumed to be enough for the workers and requesters. Moreover, Table I shows the detailed explanations of symbolsr.

B. Problem Formulation

The above self-organized network model is considered in this subsection. The requesters would like to publish a task with the purpose to collect the following data: $D = d_1, d_2, \cdots, d_l$ at time t. For all the sensing data, they have an uploaded deadline T_d. In other words, all the sensing data should be uploaded before the deadline. That is to say, the sensing data must be received by one of the requesters before time $t + T_d$. Only the workers of PoIs could be recruited to choose one of them: (1) finish the sensing task through social connection or (2) upload the collected data to a requester through the encounter in the physical world (move into the same PoI).

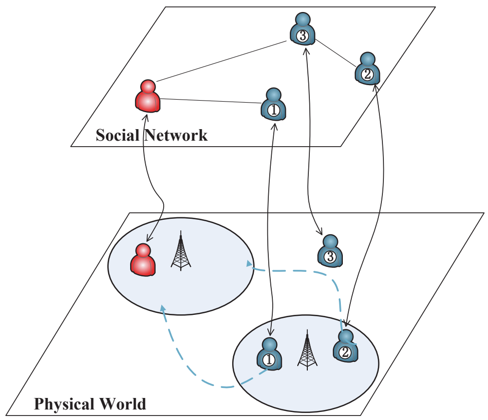


Fig. 3. The problem description, which shows that it is relatively easier for Worker 1 and Worker 2 to deliver the sensing data, which is in red color.

Assuming that the process of recruiting is done as the following steps: requester publishes a sensing task, then the workers accept the tasks, and try to finish them through both online and offline opportunities.

The first question: which worker is the most suitable one in the PoI. As shown in Fig. 3, workers 1 and 2 are both in PoI, and each worker has a contact probability with the requester in the physical world. However, at the same time, it also has a communication probability in social network. Since we want to recruit one worker, we have to decide which one is better between them.

The second question: we want to recruit a set of the best workers, (e.g., k workers), how can we design a suitable recruitment strategy.

III. Online and Offline Communication Probabilities

In this section, we predict offline contact probabilities by Semi-Markov model [11] and predict social connection probabilities between workers and requesters through online social relationship, respectively. Then, the worker's utility is decided through the total communication probability (online and offline) [12].

A. Predicting Offline Encounter Probability

The ever-been PoIs of worker k are stored as the set $H^k = \{1, 2, 3, \cdots, l\}$, if a worker is being in a PoI, then its status could be regarded as one item in H. There are totally l places in H, and the n_{th} ever-been PoI of worker k is H_n^k. The beginning time/entering time of worker k for the n_{th} place is T_n^k.

A semi-Markov [13] could be used to model the mobility pattern in this paper, a two-tuple (H_n^k,T_n^k) is the main part of the semi-Markov model. The reason of using this model is that the probability of a worker k changing from status H_n^k to status H_{n+1}^k is independent of status H_{n-1}^k. Random holding time $D_n^k = T_{n+1}^k - T_n^k$ is formulated.

The main part of proposed semi-Markov is in Eq. 1, $A_{ij}^k(t)$ is the probability of worker k changing from PoI i to PoI j.

TABLE I
Main Symbols

Symbol	Meaning
T_d	task deadline
l	PoI number
D_n^k	time for worker k to stay in n_{th} status
$W_{ij}^k(t)$	change probability of worker k from status i to status j, and also the time is t
P_{ij}^k	change probability from status i to status j
$V_{ij}^k(t)$	change probability of worker k from status i to status j, and the change time is less than t
$V_i^k(t)$	leave probability of worker k from status i, and the leave time is less than t
U_i	the utility of worker i

Obviously, H_{n+1}^k has a tight relationship with H_n^k, but has no concern with H_{n-1}^k.

$$\begin{aligned} A_{ij}^k(t) &= P(H_{n+1}^k = j, D_n^k \le t | H_0^k \cdots H_n^k, T_0^k \cdots T_n^k) \\ &= P(H_{n+1}^k = j, D_n^k \le t | H_n^k = i) \end{aligned} \tag{1}$$

Then, Eq. 2 shows the changing probability from PoI i to PoI j, num_i^k is the number of changes out of PoI i, no matter where is the destination, and sum_{ij}^k is used to record the number of changing times.

$$P_{ij}^k = P(H_{n+1}^k = j | H_n^k = i) = sum_{ij}^k / sum_i^k \tag{2}$$

Then, $V_{ij}^k(t)$ is the probability of worker k leaving i for j, and the entering time to j is less than t.

$$\begin{aligned} V_{ij}^k(t) &= P(D_n^k \le t | H_n^k = i, H_{n+1}^k = j) \\ &= \sum_{x=1}^{t} P(D_n^k = x | H_n^k = i, H_{n+1}^k = j)) \end{aligned} \tag{3}$$

And $V_i^k(t)$ is the probability of worker k leaving the PoI i, and the leaving time is less than t as follows:

$$V_i^k(t) = P(D_n^k \le t | H_n^k = i) = \sum_{j=1, j \ne i}^{l} V_{ij}^k(t). \tag{4}$$

For simultaneous Eqs. 1- 4, we could achieve the semi-Markov main part W_{ij}^k, which is shown as Eq. 5.

$$\begin{aligned} W_{ij}^k(t) &= P(H_{n+1}^k = j, D_n^k \le t | H_0^k \cdots H_n^k, T_0^k \cdots T_n^k) \\ &= P(D_n^k \le t | H_n^k{=}i, H_{n+1}^k{=}j) P(H_{n+1}^k{=}j | H_n^k{=}i) \\ &= V_{ij}^k(t) P_{ij}^k \end{aligned} \tag{5}$$

$R_{ij}^k(t)$ shows the probability that the worker k would be j after time t with the condition that now it is in PoI i. Obviously, if we know $R_{ij}^k(t)$, and the current place of worker, we could predict the future location of this worker after time t. In order to achieve $R_{ij}^k(t)$, we consider a special case: worker

k keeps staying in the beginning PoI i until time t. It does not change to any other PoI:

$$P(D_n^k > t | H_n^k = i) = 1 - V_i^k(t) \quad (6)$$

Then, another case is taken into consideration: worker k has more than one change before time t, and the change occurs at PoI r, time x:

$$P(H_t^k = j | H_0^k = i \text{ and at more than one change to } r) = \sum_{r=1}^{l} \sum_{x=1}^{t} (W_{ir}^k(x) - W_{ir}^k(x-1)) R_{rj}^k(t-x), \quad (7)$$

where $(W_{ir}^k(x) - W_{ir}^k(x-1))$ means that the first change occurs at PoI r, time x.

Finally, we obtain $R_{ij}^k(t)$ as follows:

$$R_{ij}^k(t) = \begin{cases} \sum_{r=1}^{l} \sum_{x=1}^{t} (W_{ir}^k(x) - W_{ir}^k(x-1)) R_{rj}^k(t-x), j \neq i \\ 1 - V_i^k(t) + \\ \sum_{r=1, r\neq i}^{l} \sum_{x=1}^{t} (W_{ir}^k(x) - W_{ir}^k(x-1)) R_{rj}^k(t-x), j = i \end{cases} \quad (8)$$

By now, P_{ij} can be calculated through Eq. 2. We assume that the PoI list of H_n^k is $2, 3, 2, 5, 2, 4, 5$, and $P_{23}^k = 1/3$. Obviously, $sum_{ij}^k \leq sum_i^k$ and $P_{ij}^k \leq 1$. By storing sum_{ij}^k and sum_i^k, each worker can keep track of its P matrix. Similarly, Eq. 3 could be used to calculate V_{ij}^k. For instance, the time list of changing from PoI i to PoI j is $1, 2, 4, 7, 7, 8, 9$, then $V_{ij}^k(4) = 3/7$.

Next, we attempt to achieve the encounter probability between a worker and a requester. $V_{ij}^k(t)$ gives the probability that, at time 0, the worker k in i, and it change to j at time t. In general, the nearest status of worker a is l_a at time h_a, and for requester b is status l_b, time h_b. Therefore, the encounter probability between worker a and requester b could be calculated as follows:

$$C_{ab}^i(h) = R_{l_a i}^a(h - h_a) R_{l_b i}^b(h - h_b), h > 0, \quad (9)$$

where $h > h_a > 0$, and $h > h_b > 0$. Then, the encounter probability between worker a and requester b is

$$C_{ab}(h) = \sum_{i=1}^{l} C_{ab}^i(h), h > 0. \quad (10)$$

Next, the probability of the first encounter occurs before h is defined as follows:

$$E_{ab}(h) = C_{ab}(h) \prod_{t=0}^{h-1} (1 - C_{ab}(t)), h > 0. \quad (11)$$

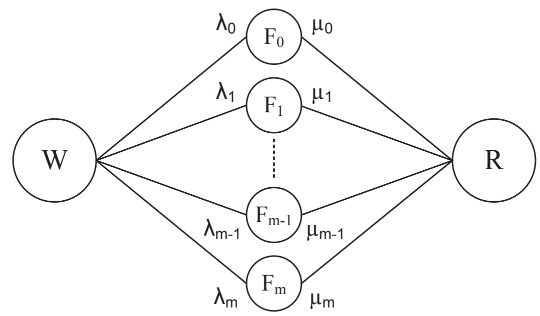


Fig. 4. An illustration of a two hop theory model (worker-friend-requester).

The first encounter probability between worker a and requester b before the deadline is:

$$F_{ab} = \sum_{h=1}^{T_d} E_{ab}(h) \quad (12)$$

Then, the probability that a worker i with the sensing data encounters a requester y is shown in Eq. 13.

$$F_{w_i r_y}^{off} = \sum_{h=1}^{T_d} E_{w_i r_m}(h) \quad (13)$$

B. Predicting Online Connection Probability

Due to limited social network information, we use the two-hop routing theory to predict the communication probability, which uses local social network information (i.e., friends and friends' friends). Moreover, [3] shows that two-hop routing achieves a high delivery ratio. In two-hop routing, each worker records its two-hop friend information to obtain Worker-Friend-Requester paths (denoted as W-F-R, in Fig. 4).

As shown in Fig. 4, λ and μ represent the social communication probability; then, for a worker i, its communication probability with the requester y is defined as Eq. 14:

$$F_{w_i r_y}^{on} = 1 - \prod_{z=1}^{m} (1 - \lambda_z \mu_z) \quad (14)$$

According to Eq. 13 and Eq. 14, for a worker w_i, the probability of communicating with a requester (r_y) is shown in Eq. 15.

$$F_{w_i r_y} = 1 - (1 - F_{w_i r_y}^{on})(1 - F_{w_i r_y}^{off}) \quad (15)$$

Therefore, we could obtain the utility of worker i, in the follow equation, m represents the number of requesters.

$$U_i = F_{w_i R_M} = \sum_{y=1}^{m} F_{w_i r_y} \quad (16)$$

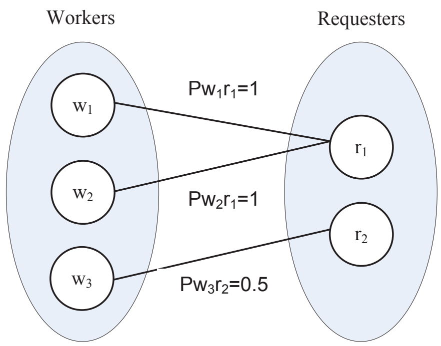


Fig. 5. There are many different recruitment strategies. In a Top 2 worker recruitment strategy, if we just recruit the worker with highest utility, then, workers 1 and 2 will be recruited. However, workers 1 and 3 are obviously a better worker set than workers 1 and 2.

IV. Worker Recruitment Strategy

Here, we attempt to present a Worker Recruitment in MSC (WEO) for solving the question: how can we recruit a set of best k workers for maximizing the total delivery ratio of the sensing data?

A. *Top k Worker Recruitment (WEO)*

Each worker has an utility (a comprehensive connecting probability in both online and offline networks) for a requester. We attempt to recruit the optimal k workers, who have the highest total-utility so that the problem above could be turned out to be NP-hard through the following theorem [14].

Theorem 1. *The k workers selection problem is NP-hard.*

Proof. Suppose that the utilities for all the workers are equal to 1, in this case, for every worker, it could surely communicate with at least one requester before the deadline. In other words, every worker could cover at least one requester. Obviously, this special situation has no difference with the k set cover problem. The worker is regarded as the set, and the requester is considered as an item. Then we try to recruit a best group of workers in order to cover as many as possible requesters. As we know, the k set cover problem is NP-hard; then, the special case of the general k worker recruitment must be NP-hard. Hence, the top k worker recruitment problem is more complex than an NP-hard problem so that the theorem holds. □

It is worth noting that a lot of greedy algorithms could be used to solve the above NP-hard problem. An easy solution is that the worker with the highest utility could be recruited in a higher priority. The same operation is repeated for k times, then the k workers recruiting problem could be solved. However, the recruited set is not optimal with the example shown in Fig. 5. If we just recruit the worker with the highest utility, then workers 1 and 2 will be recruited, while the utility of worker set 1 and 2 is 1 (1+0). However, a better strategy is to recruit workers 1 and 3, because the utility of worker set 1 and 3 is 1.5 (1+0.5).

Algorithm 1 Greedy heuristic for WEO worker recruitment

Input:
Total set of workers in this PoI: N
Set of recruiters: W
W' total utility: U_W
Output:
Top k worker set: W
$W \longleftarrow \varnothing$; $U_W = 0$
for i =1 to k **do**
  $w \longleftarrow \arg\max_{w \in N \setminus W} U_{W \cup \{w\}}$
  $W = W \cup \{w\}$; update U_W
return W

To optimize the delivery ratio, we should recruit a suitable set of k workers. Focusing on the NP-hard problem, we propose a greedy heuristic algorithm: WEO, rather than recruiting the highest worker's utility, we pay attention to maximizing the recruited worker set's utility U_W, which is defined as the communication probability between worker W and all the requesters. As shown in Fig. 5 in WEO, we will select the workers 1 and 3 (or 2 and 3), rather than workers 1 and 2. Algorithm 1 shows the detail process of WEO.

B. *Approximation Ratio*

Theorem 2. *U_W is a submodular function. For two worker sets W_1 and W_2, if $W_1 \subseteq W_2$, then $\forall w_k \in N \setminus W$, the submodular property holds, i.e., $U_{W_1 \cup \{w_k\}} - U_{W_1} \geq U_{W_2 \cup \{w_k\}} - U_{W_2}$.*

Proof. A special case: $|W_2| - |W_1| = 1$ is first considered, we try to prove that $U_{W_1 \cup \{w_k\}} - U_{W_1} \geq U_{W_2 \cup \{w_k\}} - U_{W_2}$. Then, a general case $|W_2| - |W_1| = \omega > 1$ is further proved.

First, we assume that $|W_2| - |W_1| = 1$ and $W_2 \setminus W_1 = \{w_h\}$ according to $W_1 \subseteq W_2$. To prove the submodular property of U_W, we focus on the relationship between w_k and one of requesters $\forall r_j \in R$, there are following three cases:

Case 1: w_k has no communication probability (no matter online or offline) with r_j. Then, $P_{w_k r_j} = 0$. So we have $U_{W_1 \cup \{w_k\}} = U_{W_1}$ and $U_{W_2 \cup \{w_k\}} = U_{W_2}$. Consequently, $U_{W_1 \cup \{w_k\}} - U_{W_1} = U_{W_2 \cup \{w_k\}} - U_{W_2} = 0$.

Case 2: w_k has a communication probability (online or offline) with r_j, but w_h has no communication probability with r_j. Then, $P_{w_h r_j} = 0$. According to Eq.2, $U_{W_2} = U_{W_1 \cup \{w_h\}} = U_{W_1}$, and $U_{W_2 \cup \{w_k\}} = U_{W_1 \cup \{w_k\} \cup \{w_h\}} = U_{W_1 \cup \{w_k\}}$. So we can get $U_{W_1 \cup \{w_k\}} - U_{W_1} = U_{W_2 \cup \{w_k\}} - U_{W_2}$.

Case 3: Both w_k and w_h have a communication probability with r_j. Then for all the workers w_i in W_1, the communication probability with r_j is defined as P_{1j}; similarly, for W_2, the communication probability with r_j is defined as P_{2j}. It is not difficult to find that, $P_{1j} \leq P_{2j}$, then $U_{W_1 \cup \{w_k\}} - U_{W_1} = 1 - (1 - P_{1j})(1 - P_{kj}) - P_{1j}$. Similarly, $U_{W_2 \cup \{w_k\}} - U_{W_2} = 1 - (1 - P_{2j})(1 - P_{kj}) - P_{2j}$. Therefore, we have

TABLE II
SIMULATION SETTINGS

Parameter	Traces	
	epfl	roma/taxi
Task Deadline	300,400,500,$\cdots$,900	
Probability of Social Communication	0.02,0.03,0.04,$\cdots$,0.08	
Time Unit (s)	30	15
PoI Number	13	10
PoI Range (m)	80	200
Worker Number	368	158
Requester Number	4,5,6,$\cdots$,10	

$$
\begin{aligned}
&(U_{W_2\cup\{w_k\}} - U_{W_2}) - (U_{W_1\cup\{w_k\}} - U_{W_1}) \\
&= (1-(1-P_{2j})(1-P_{kj})-P_{2j})-(1-(1-P_{1j})(1-P_{kj})-P_{1j}) \\
&= (P_{1j} - P_{2j})P_{kj} \leq 0
\end{aligned} \tag{17}
$$

Therefore, $U_{W_1\cup\{w_k\}} - U_{W_1} \geq U_{W_2\cup\{w_k\}} - U_{W_2}$.

Above all, $U_{W_1\cup\{w_k\}} - U_{W_1} \geq U_{W_2\cup\{w_k\}} - U_{W_2}$ holds for $\forall r_j \in R$ in all cases. So, the following case is taken into consideration: $|W_2| - |W_1| = \omega \geq 1$. We assume that $W_2 \backslash W_1 = \{w_h, w_{h+1}, \cdots, w_{h+\omega-1}\}$. Then, we have $U_{W_1\cup\{w_k\}} - U_{W_1} \geq U_{W_1\cup\{w_k\}\cup\{w_h\}} - U_{W_1\cup\{w_h\}} \geq U_{W_1\cup\{w_k\}\cup\{w_h\}\cup\{w_{h+1}\}} - U_{W_1\cup\{w_h\}\cup\{w_{h+1}\}} \geq \cdots \geq U_{W_1\cup\{w_k\}\cup\{w_h\}\cup\cdots\cup\{w_{h+\omega-1}\}} - U_{W_1\cup\{w_h\}\cup\cdots\cup\{w_{h+\omega-1}\}} = U_{W_2\cup\{w_k\}} - U_{W_2}$

Therefore, U_W is a submodular function. Theorem 2 is proved. □

Theorem 3. *For a positive submodular function f, we attempt to select a set of k items: W by selecting an item that provides the maximize value improvement for an continuous k times. Let W^* be optimal performance. Then $f(W) \geq (1-1/e) \cdot f(W*)$.*

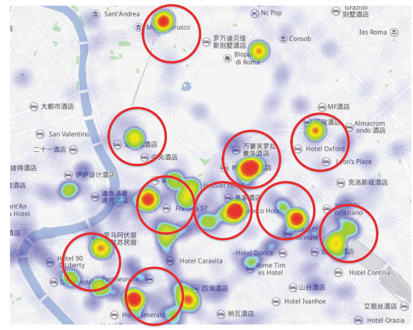
Proof. Submodular functions have a desirable attribute, which is shown as follows: if we have a nonnegative function f that is submodular, and it does not decrease when we add an element e to the set: $f(W \cup \{e\}) \geq f(W)$ for all elements e and sets W. The purpose is to select a k-item set S, which maximizes $f(W)$. This is obviously a well-known set-covering problem, and also an NP-hard optimization problem, [15] proves that a factor of $1 - 1/e$ exists for the above problem. Obviously, the top k user recruiting problem matches all the requirements of the above submudular property, hence, Theorem 3 is proved. □

Therefore, the proposed top k recruiting strategy in this paper could achieve a $(1-1/e)$- approximation of the optimal recruiting strategy.

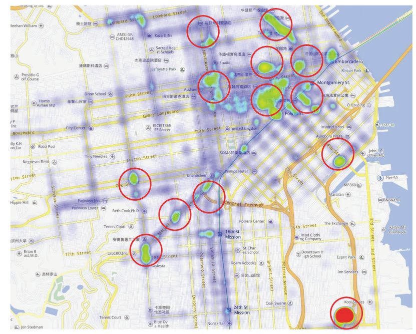
V. PERFORMANCE EVALUATION

A. The Traces Used and Settings

Two real-word datasets are used in this paper to test the performances of WEO: *roma/taxi trace set* [16] and *epfl trace set* [17]. There are 320 taxi drivers who work in the city area of Rome, Italy, in *roma/taxi trace set*. The locations of the taxis are recorded in a periodic way through uploading the GPS position to the server by the taxi drivers. In the *epfl trace set*, 500 taxis' locations are collected and recorded in about one month of San Francisco Bay Area [11], it also includes GPS coordinates of the drivers. The above two traces are both collected by taxi drivers, they could communicate with each other through short-distance communication protocol (Bluetooth or Wifi), and also they could disseminate messages to each other by online social network. Hence, the two datasets could be used to test our proposed recruitment strategies.



(a) roma/taxi



(b) epfl

Fig. 6. The PoI locations in Baidu map.

First of all, we filter out some discrete traces and also move out the worker trace, which is far away from the area most workers are. Then, by using Baidu map, we embed the traces into the actual map. With the help of JavaScript API in Baidu map, a thermodynamic chart is formulated. The red area represents that it is covered more than 400 times by the workers' traces, and we also find the associate PoIs in each data set (Fig. 6). Some workers are randomly selected as the requesters, and the other users as the workers. The relationship among all the users are generated randomly and will not change during the whole simulation. The simulation settings in this paper are shown in Table II, where time unit means the minimal collecting period; in other words, time unit is the interval time for reporting GPS data. For example, if the simulation time of *epfl* is 100 periods, then the actual time is 100 time units. This equals 3000s, this is because the time unit in *epfl* is 30s.

B. Algorithms Performances

For testing the performances of WEO, simulations are done in the above two reality traces. Two performances are mainly concerned in this paper: (1) the accuracy for the worker's utility and (2) the delivery performance of sensing data for WEO.

The purpose in terms of the first performance is to test whether the worker's utility is reasonable and accurate, we compare the four 1-worker recruitment strategies: WEO, WEO-on, WEO-off and RR. WEO is our strategy and recruits only one worker with the highest utility in every PoI. WEO-on recruits the worker with the highest online utility, while WEO-off recruits the worker with the highest offline utility. RR (Randomly Recruiting) just randomly selects a worker in the available worker set.

Considering the second performance, our purpose is to verify that whether WEO could get the highest delivery ratio when we need to recruit k workers in each POI, compared to the RL and RR recruitment strategies. WEO is proposed in this paper, which recruits a set of k workers, who have the highest U_W with the requesters. RL (Recruitment Largest) recruits k workers, whose utilities are the largest among the workers in the PoI. RR (Recruitment Randomly) also randomly recruits workers; however, in this group of simulations, RR recruits k workers, while not one worker in the available worker set.

A lot of performances are tested in the simulations, while we just take the most important performance metric into consideration the delivery ratio of the sensing data. This is the ratio of the sensing data number successfully uploaded to the requesters over the total number of the generated sensing data.

C. Simulation Results

1) Worker's Utility: To test the accuracy of the worker's utility, we simulate the delivery ratios of the four 1-worker recruitment strategies: WEO, WEO-on, WEO-off, and RR, in the two reality traces: *roma/taxi* and *epfl*. Fig. 7 and Fig. 8 show the simulation results in terms of worker's utility.

As described in Section V-B, we attempt to test the accuracy of the proposed worker's utility through recruiting only one worker in each PoI. We compare the performance to draw our conclusions. As shown in Fig. 7, in reality trace *roma/taxi*, we compare the four 1-worker recruitment strategies: WEO, WEO-on, WEO-off, and RR. The results show that WEO obtains the best performance compared with the other recruitment strategies, which means that the worker's utility leads to a right guidance in terms of worker's communication ability with the requesters. In other words, the proposed worker's utility in this paper could represent the worker's communication ability with the requesters.

It is worth noting that the ranking of the above four recruitment strategies is: WEO>WEO-on>WEO-off>RR, which means that the online social communication probability plays a more important role in the delivery of the sensing data, compared with the offline encounter probability. RR recruits

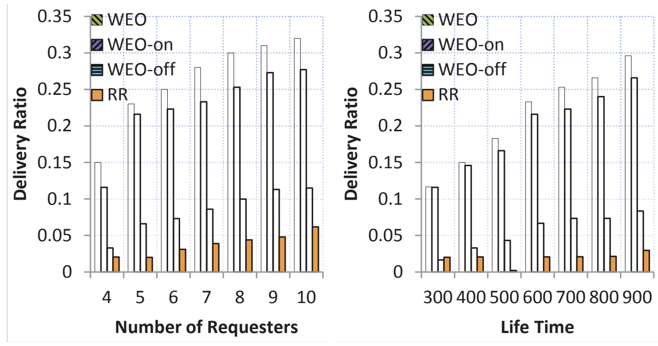


Fig. 7. The simulation results in terms of worker's utility in the roma/taxi data set.

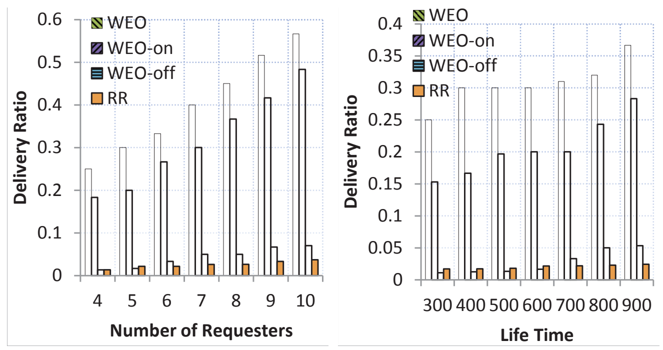


Fig. 8. The simulation results in terms of worker's utility in the epfl data set.

a random worker in each PoI, which is why its delivery performance is the lowest.

Moreover, we test the performances of the above four worker selecting methods along with the growth of the requester number and life time, respectively. As shown in Fig. 8, the performance of the sensing data appears to follow an upward trend along with the growth of the requester number, which is reasonable because more requesters results in more communication opportunities between the workers and the requesters, hence the delivery performance becomes higher. Similarly, the delivery ratio of the sensing data also appears an increasing trend with the increasing life time, because if we have enough time, then a higher communication probability between the recruited workers and the requesters will be obtained. The performance results of Fig. 8 are similar with that of Fig. 7.

2) Delivery Performance: In this subsection, to test the delivery performances of WEO, two sets of simulations are first done, they use two reality traces, *roma/taxi* and *epfl*, respectively. Because we try to test the situation that the sensing data are uploaded by WEO, we treat the workers in WEO as sources and treat the requesters as destinations. From sources to destinations, we attempt to recruit the best k workers in each PoI, who are most valuable to assist in the uploading process. For testing the performances of WEO, we test the successfully delivered data number in two real-world data sets. Fig. 9 and Fig. 10 give the detailed performances of WEO in terms of delivery ratio of sensing data.

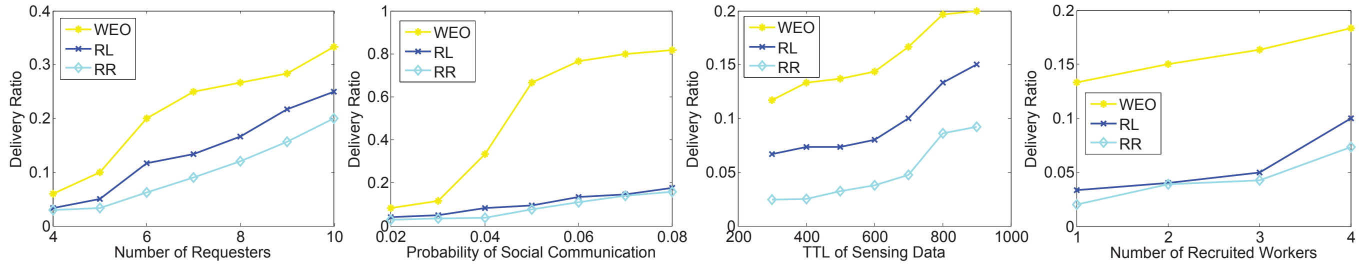


Fig. 9. Delivery ratio performance comparison on the *roma/taxi trace set*.

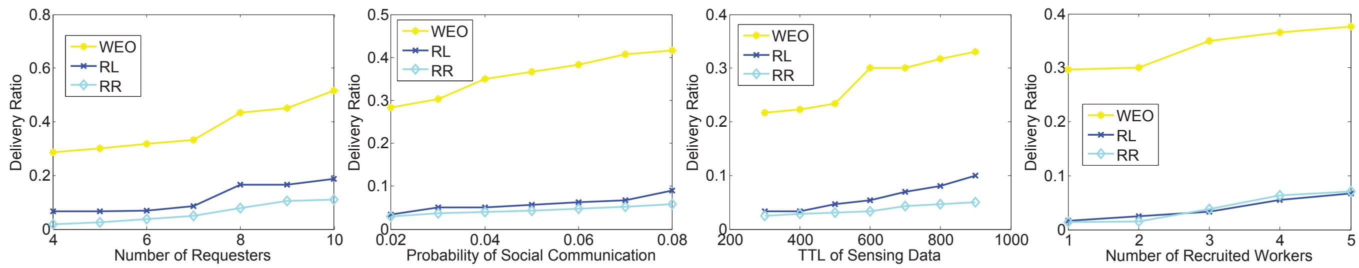


Fig. 10. Delivery ratio performance comparisons on the *epfl trace set*.

WEO recruits a set of k workers, who has the highest U_W (calculated in Section IV) with the requesters, while RL recruits the k workers with the highest utility, and RR randomly recruits k workers in each PoI. Simulation results in Fig. 9 show that WEO achieves the highest delivery ratio of sensing data. This is because WEO considers the workers as a set and measures the total utility of the set, rather than considering an individual worker.

Moreover, we compare the delivery performances of the three worker selection methods: WEO, RL and RR, along with the growth of the requester number, social probability, TTL and the number of the recruited workers, respectively. Figures show that WEO always obtains a higher deliver performance than that of RL, which proves theoretical results of Section IV-A. And we could notice that the delivery ratio of the sensing data appears a growing trend with increasing requester number, social probability, TTL and number of the recruited workers, which is not difficult for us to understand. The simulation results of Fig. 10 are almost same with that of Fig. 9, we do not repeat the description.

In conclusion, in the two reality traces *roma/taxi* and *epfl*, we first compare the accurate of utility calculation, and get the conclusion that we could obtain a well-done utility. Then, we test the deliver performance in terms of the three recruitment strategies. The simulation results show that WEO achieves the highest delivery performance.

VI. Related Work

A. Incentive Mechanisms

Zheng *et al.* [18] analyzed the issue of maximizing the weighted coverage in mobile crowdsensing and proposed an incentive mechanism which is proved to be budget-friendly and efficient for mobile crowdsensing. Yang *et al.* [19] proposed different corresponding incentive mechanisms for a crowdsourcer based and user based schemes. Chen *et al.* [20] recognized network effects as internal incentives, analyzed its effects on the composition of external incentives, then proposed two external mechanisms to motivate more users and increase the revenue of the crowdsourcer. Yang *et al.* [21] provided an unsupervised learning method to recognize the remaining process as a cooperative game, then proposed an approach using Shapley value to get users remuneration. Han *et al.* [22] focused on a Bayesian pricing problem to motivate participants efficiently with enough data qualities to get sensing robustness and changed the problem from a non-submodular one into submodular. Feng *et al.* [23] studied the situation of motivating users to join in crowdsensing tasks with smartphones and proposed a reverse auction scheme to address the interaction among users and requesters.

The above works encourage workers to take part in the crowdsensing task through some suitable incentive mechanisms. These are the element jobs of this paper and also an important part in MSC. However, even though we have an incentive mechanism to encourage workers for finishing the sensing task, we still need to decide which workers are optimal recruiters to finish the task most efficiently.

B. Recruitment Strategy

Xiao *et al.* [14] analyzed the Deadline based Recruitment problem in the probabilistic cooperation mobile crowdsensing and presented a user selection method considering induction time and a greedy algorithm called DUR. Li *et al.* [24] considered a user selection strategy with sundry tasks for minimizing the cost while maintaining high probability coverage. Then they proposed an offline and online algorithm to get the feasible solution. Yi *et al.* [25] provided an algorithm which is for solving the VPR problem in order to recruit vehicle users. Pu *et al.* [26] considered an online user selection scheme to achieve the optimization of the total serving level and proposed a method by using the dynamic planning. Karaliopoulos *et al.* [27] predicted user's location dynamically and minimized the total cost using the deterministic and stochastic mobility models, while achieving a good optimization goal in terms of finishing the sensing data.

The above works pay attention to user recruitment strategy. However, almost all the works do not consider utilizing the social influence to assist in finishing the sensing tasks, and also do not consider the propagation in online social network, hence, both online and offline deliverings will increase the task finishing propobility.

VII. CONCLUSION

We have looked into the problem of worker selection in double-layer MCS campaigns drawing on opportunistic networking and social networking methods. First, we formulate the sensing data transmission into two parts: online social connection and offline mobile contact. Next, by using the semi-Markov and the two-hop social network theory, we propose WEO, where we recruit a set of best workers with the highest communication probability (online and offline) with the requesters. Moreover, we prove that the optimal recruitment problem is NP-hard, and we introduce a practical greedy heuristic method for this problem, the performance of the greedy method is also tested. Two reality traces, *roma/taxi* and *epfl* are tested in our simulations, where WEO always achieves the highest delivery ratio of sensing tasks among different recruitment strategies.

REFERENCES

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] Y. Gao, W. Dong, K. Guo, X. Liu, Y. Chen, X. Liu, J. Bu, and C. Chen, "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," in *Proc. of IEEE INFOCOM 2016*.

[3] Y. Hu, G. Dai, J. Fan, Y. Wu, and H. Zhang, "Blueaer: A fine-grained urban pm2.5 3d monitoring system using mobile sensing," in *Proc. of IEEE INFOCOM 2016*.

[4] D. Zhang, H. Xiong, LeyeWang, and G. Chen, "Crowdrecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *Proc. of ACM UbiComp 2014*.

[5] H. Xiong, D. Zhang, G. Chen, L. Wang, V. Gauthier, and L. E. Barnes, "iCrowd: Near-Optimal Task Allocation for Piggyback Crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 2010–2022, 2016.

[6] B. Guo, C. Chen, D. Zhang, and Z. Yu, "Mobile crowd sensing and computing: when participatory sensing meets participatory social media," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 131–137, 2016.

[7] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proc. of ACM WWW 2013*.

[8] X. Gong, X. Chen, J. Zhang, and H. V. Poor, "Exploiting Social Trust Assisted Reciprocity (STAR) Toward Utility-Optimal Socially-Aware Crowdsensing," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 3, pp. 195–208, 2015.

[9] S. Reddy, D. Estrin, M. Hansen, and M. Srivastavai, "Examining micro-payments for participatory sensing data collections," in *Proc. of ACM UbiComp 2010*.

[10] J. Li, Z. Cai, M. Yan, and Y. Li, "Using crowdsourced data in location-based social networks to explore influence maximization," in *Proc. of IEEE INFOCOM 2016*.

[11] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, 2018.

[12] Y. Yang, Y. Xu, E. Wang, K. Lou, and D. Luan, "Exploring influence maximization in online and offline double-layer propagation scheme," *Information Sciences*, vol. 450, no. 2018, pp. 182–199, 2018.

[13] Q. Yuan, I. Cardei, and J. Wu, "An Efficient Prediction-Based Routing in Disruption-Tolerant Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 19–31, 2012.

[14] M. Xiao, J. Wu, H. Huang, L. Huang, and C. Hu, "Deadline-sensitive user recruitment for mobile crowdsensing with probabilistic collaboration," in *Proc. of IEEE ICDCS 2016*.

[15] G. Comuejols, M. Fisher, and G. Nemhauser, "Location of bank accounts to optimize float." *Management Science*, vol. 1977, no. 23.

[16] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD dataset roma/taxi (v. 2014-07-17)," Downloaded from http://crawdad.org/roma/taxi/20140717, 2014.

[17] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD dataset epfl/mobility (v. 2009-02-24)," Downloaded from http://crawdad.org/epfl/mobility/20090224, Feb. 2009.

[18] Z. Zheng, F. Wu, X. Gao, H. Zhu, G. Chen, and S. Tang, "A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–14, 2017.

[19] D. Yang, G. Xue, G. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Transactions on Networking*, pp. 1–13, 2015.

[20] Y. Chen, B. Li, and Q. Zhang, "Incentivizing crowdsourcing systems with network effects," in *Proc. of IEEE INFOCOM 2016*.

[21] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, and G. Chen, "On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 832–847, 2017.

[22] K. Han, H. Huang, and J. Luo, "Posted pricing for robust crowdsensing," in *Proc. of ACM MobiHoc 2016*, pp. 261–270.

[23] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *Proc. of IEEE INFOCOM 2014*, pp. 1231–1239.

[24] H. Li, T. Li, and Y. Wang, "Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks," in *Proc. of IEEE MASS 2015*.

[25] K. Yi, R. Dua, L. Liu, Q. Chend, and K. Gao, "Fast participant recruitment algorithm for large-scale vehicle-based mobile crowd sensing," *Pervasive and Mobile Computing*, vol. PP, no. 99, pp. 1–12, 2017.

[26] L. Pu, X. Chen, J. Xu, and X. Fu, "Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing," in *Proc. of IEEE INFOCOM 2016*.

[27] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, "User recruitment for mobile crowdsensing over opportunistic networks," in *Proc. of IEEE INFOCOM 2015*.