Incentive Mechanisms for Large-scale Crowdsourcing Task Diffusion based on Social Influence

Jia Xu, *Member, IEEE,* Gongyu Chen, Yuanhang Zhou, Zhengqiang Rao, Dejun Yang, *Senior Member, IEEE,* Cuihua Xie

***Abstract*—Crowdsourcing has become an effective tool to utilize human intelligence to perform tasks that are challenging for machines. Many incentive mechanisms for crowdsourcing systems have been proposed. However, most of existing mechanisms assume that there are enough participants to perform the crowdsourcing tasks. This assumption may not be true in large-scale crowdsourcing scenarios. To address this issue, we diffuse the crowdsourcing tasks via the social network. We study two task diffusion models, and formulate the problem of minimizing the total cost such that all tasks can be completed in expectation for each model. The topology based influence estimation and history based influence estimation based on the limited knowledge of social network are presented in this paper. Further, we present the global influence estimation method to measure the influence over the whole community with the full knowledge of social network. We design two sealed reverse auction based truthful incentive mechanisms, *MTD-L* and *MTD-IC*, for both diffusion models. Through both rigorous theoretical analysis and extensive simulations, we demonstrate that the proposed mechanisms achieve computational efficiency, individual rationality, truthfulness, and guaranteed approximation. Moreover, the global influence estimation based mechanisms always output the least social cost and overpayment ratio, and the history influence estimation based mechanisms show significant superiority in terms of task completion rate.**

***Keywords*—crowdsourcing; incentive mechanism design; reverse auction; social network**

I. Introduction

Crowdsourcing takes advantage of the wisdom of individuals, teams, and communities to complete tasks. One famous example is Wikipedia [1]. Recently, crowdsourcing has been widely used in many fields, including video analysis [2], knowledge discovery from web tables [3], conducting human robot interaction studies (e.g., RMS [4]), image quality assessment [5], and online marketplace (e.g., Amazon Mechanical Turk, AMT [6]). Moreover, with the rapid proliferation of smartphones, mobile crowdsourcing has become an efficient approach for large-scale sensing applications. Due to the high mobility of vehicles or the controllability of the unmanned vehicles, e.g., drones and driverless cars, vehicular crowdsourcing shows great superiority for large-scale tasks. There have been many vehicular crowdsourcing paradigms, such as energy crowdsourcing from AEVs (Autonomous Electric Vehicles) during peak time periods [7], vehicular crowdsourcing (VC) campaign for long-term and hash tasks in remote or dangerous areas [8], vehicular fog computing for real-time analytics of crowdsourced dash camera video [9].

Incentive mechanism design is important to many network computing paradigms [10], such as edge computing [11], transparent computing [12] and mobile crowdsourcing computing [13]. Particularly, incentive mechanisms are crucial to crowdsourcing while crowdsourcing workers may have associated costs for performing tasks. A lot of research efforts [13, 14, 15] have been focused on developing incentive mechanisms to entice users to participate in crowdsourcing. However, most of existing studies assume that there are enough participants to perform the crowdsourcing tasks. This assumption may not be true in real situations. First, at the early beginning, the crowdsourcing platform faces the cold-start problem and cannot provide sufficient workers for completing tasks. Moreover, the crowdsourcing platform also benefits the developed platform when it cannot find enough workers interested in some specific tasks. The platform can resort to our worker recruitment process to ask help from users in social networks. In addition, the bases of participants of crowdsourcing applications are still not big enough. According to [16], mobile crowdsensing applications have rarely scaled up to more than 1000 participants. Our statistical results show that there are 618.65 uncompleted HITs (Human Intelligence Tasks) in Amazon Mechanical Turk [6] per day on average from 2016-05-01 to 2016-05-20. Thus how to diffuse crowdsourcing tasks efficiently is a nontrivial issue. However, most of existing incentive mechanisms aim to stimulate the users to perform the crowdsourcing tasks rather than diffuse the tasks. As far as we known, there is no off-the-shelf incentive mechanism designed in the literature for the crowdsourcing task diffusion.

In reality, the celebrities have shown powerful influence in crowdsourcing campaigns. As an example of the "Digital Death" campaign [17] launched December 1, 2010 for World

J. Xu, G. Chen, Y. Zhou, and Z. Rao are with the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China. (e-mail: xujia, 1218043216, Q17010120, 1216043122@njupt.edu.cn).

D. Yang. is with Department of Computer Science, Colorado School of Mines, Golden, CO 80401, USA. (e-mail: djyang@mines.edu).

C. Xie. is with Oujiang College, Wenzhou University, Wenzhou, Zhejiang 325035, China. (e-mail: xiecuihua@wzu.edu.cn).

Corresponding author: Cuihua Xie. (e-mail: xiecuihua@wzu.edu.cn).

This research was supported in part by NSFC grants 61872193, NSF grants 1717315 and Zhejiang Philosophy and Social Sciences Foundation grants 19NDJC145YB.

TABLE I
SOCIAL NETWORK OWNED CROWDSOURCING SYSTEMS

Crowdsourcing system	**Service**	**Operator**	**Social Network**
Stepes [21]	global shared interpreting services on Mobile	Facebook	Facebook
Google Image Labeler [22]	label images as a side-effect of playing a game	Google	Google+
Translate Community [23]	improve google's translation quality	Google	Google+
Amazon Mechanical Turk [6]	online marketplace	Amazon	Amazon Spark
QQ-Crowd [24]	enhance the quality of products	Tencent	QQ

AIDS Day, the stars sacrifice their digital lives on Twitter and Facebook until their fans donate one million dollars to buy their lives back. Among the other celebrities featured in the campaign were Alicia Keys, Lady Gaga, Justin Timberlake, Usher and Serena Williams. The million dollar donation goal was reached in six days. It is intuitive to select influential users to diffuse large-scale tasks in the social network for improving the crowdsourcing participation level. Here, "large-scale tasks" represents that the workload of crowdsourcing tasks is huge for the current user base of the platform. For example, collecting sensing data from a large-scale area, complicated computation or data processing and collecting multiple sensing data/answers of crowdsourcing task to improve the variety and robustness. Thus, it is necessary to recruit participants from the social network through the task diffusion.

Online social network has become one of the most effective and efficient solutions for marketing and advertising. The extreme boom of online social network sites like Facebook, Twitter, Google+, Microblog and Wechat has been witnessed in the past decade. In social networks, one of the most significant events is called diffusion, such as the diffusion of news, innovations and product adoptions. Kempe *et al.* [18] proposed the two most popular influence diffusion models: independent cascade model and linear threshold model. A lot of research efforts have been focused on maximizing the spread of influence [18] or selecting minimum initial user set to diffuse influence [19, 20] in online social network. However, the existing studies mainly consider the scenario, where there is only one diffusion object. Moreover, the cost of diffusion and the corresponding incentive to diffusers are largely neglected in current research literature.

Many social network service sites have developed their own crowdsourcing systems. There are some realistic examples of existing crowdsourcing systems as shown in TABLE 1. These crowdsourcing systems can utilize the knowledge of the social network. It is intuitive to select the influential users to diffuse large-scale tasks in the social network for improving the crowdsourcing participation level.

In this paper, we consider the crowdsourcing with multiple large-scale tasks is launched in the platform, which is operated by an online social network site. Each of large-scale tasks requires a specific number of users to perform in order to ensure the variety and robustness of crowdsourced data. We term such tasks cooperative tasks. We consider the number of registered users is insufficient to complete the multiple cooperative tasks. The objective is designing truthful incentive mechanisms to minimize the total diffusion cost such that each of large-scale tasks can be completed in expectation through the task diffusion in online social network. To address the insufficient participation problem, we model the crowdsourcing task diffusion process as a sealed reverse auction. In our system model, each registered user bids for the tasks he/she can diffuse. The platform selects a subset of registered users and notifies the selected winners. The winners diffuse the tasks to their social neighbors. Then the influenced social neighbors perform the tasks. Finally, each winner obtains the payment, which is determined by the platform. The whole process is illustrated by Fig.1.

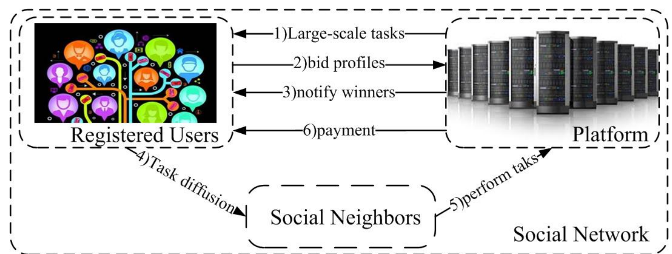


Fig. 1. Crowdsourcing task diffusion process through social network

The incentive mechanisms for the workers (influenced social neighbors in this paper) to perform the multiple cooperative tasks in crowdsourcing was proposed in [25]. Thus, we focus on addressing the insufficient participation problem in crowdsourcing, and only consider the incentive to the registered users for diffusing tasks in this work.

The problem of designing truthful incentive mechanisms to minimize the total cost for diffusing multiple large-scale crowdsourcing tasks is very challenging. First, the character of crowdsourcing task diffusion is that the influence of users not only depends on the network structure, but also the tasks to be diffused. This is because the uses usually have different influence for different tasks. However, influence evaluation methods based on network structure metrics [26, 27, 28] cannot capture this character of crowdsourcing task diffusion. In our crowdsourcing scenario, the diffusion models should enable the registered users diffuse multiple large-scale tasks simultaneously in order to meet the demands of crowdsourcing scenarios. Second, the traditional influence evaluation methods identify influential nodes based on the global knowledge of network structure metrics, such as degree centrality [28], betweenness centrality [26] and closeness centrality [27]. However, it is hard to obtain the global knowledge of network structure in crowdsourcing system. In addition, some studies [29, 30] need to calculate the influence between the social uses based on the predefined base influence, which is difficult to obtain in practice. Although many social network service sites have developed their own crowdsourcing systems, the limited knowledge can be obtained in most practical situations. Moreover, the registered user may take a strategic behavior by submitting dishonest bid price to maximize its utility.

The main contributions of this paper are as follows:

- To the best of our knowledge, this is the first work to design the auction-based truthful incentive mechanisms for the task diffusers in crowdsourcing systems through the influence diffusion models in social networks.
- We present two crowdsourcing task diffusion models: linear task diffusion model and independent cascade task diffusion model, and formulate the *Social Optimization Task Diffusion* (*SOTD*) problem for both models.
- We present two local influence estimation methods based on different knowledge of social network. Further, we extend the local influence estimation method to global influence estimation, which can estimate the influence over the whole social network.
- We design an incentive mechanism for each of two diffusion models to solve the *SOTD* problem. We show that the designed mechanisms satisfy desirable properties of computational efficiency, individual rationality, truthfulness, and guaranteed approximation. Moreover, our incentive mechanisms can largely reduce the total cost of task diffusion in crowdsourcing.

The rest of the paper is organized as follows. We review the state-of-art research in Section II. Section III formulates the diffusion models, and lists some desirable properties. Two influence estimation methods are proposed in Section IV. Section V and Section VI present the detailed design and analysis of our incentive mechanisms for both two diffusion models, respectively. We present the global influence estimation method in Section VII. Performance evaluation is presented in Section VIII. We conclude this paper in Section IX.

II. Related Work

A. Incentive Mechanism Design in Crowdsourcing

Many incentive mechanisms for crowdsourcing have been proposed thus far. Singer proposed a truthful budget feasible mechanism [31] based on the proportional share allocation rule. However, the designed mechanism was valid only for submodular functions. In [32], Singla *et al.* exploited a link between procurement auctions and multi-armed bandits to design mechanisms that are budget feasible, achieving near-optimal utility for the requester. The incentive mechanisms for the crowdsourcing system with biased requesters were proposed in [33]. Pricing mechanisms were also developed in [34] for the budget feasible maximizing task problem and the budget feasible minimizing payment problem based on the method in [31]. Yang *et al.* proposed two different models for smartphone crowd sensing [35]: the platform-centric model where the platform provides a reward shared by participating users, and the user-centric model where users have more control over the payment they will receive. In [13], Feng *et al.* formulated the location-aware collaborative sensing problem as the *winning bids determination problem*, and presented a truthful auction using the proportional share allocation rule. Koutsopoulos designed an optimal reverse auction [36], considering the data quality as *user participation level*. However, the *quality indicator*, which essentially measures the relevance or usefulness of information, is empirical and relies on user's historical information. In [37], Zhao *et al.* investigated the online crowdsourcing scenario where the users submit their profiles to the crowdsourcer when they arrive. The objective is selecting a subset of users for maximizing the value of the crowdsourcer under the budget constraint. They designed two online mechanisms, *OMZ*, *OMG* for different user models. Zhou *et al.* designed the incentive mechanisms for mobile crowdsensing systems under the platform's budget constraint and users' capacity constraint, which can ensure approximately maximized value of services for the crowdsourcer [38]. Zhang [14] *et al.* proposed *IMC*, which consider the competition among the requesters in crowdsourcing. Wang *et al.* divided the life cycle of each crowdsensing task in mobile crowdsensing into four phases: task allocation, incentive, data collection, and data publishing, and designed a privacy-preserving framework for mobile crowdsensing to protect users' privacy in the whole life cycle of mobile crowdsensing [39]. However, all above studies assume that there are enough participants who can perform the crowdsourcing tasks.

B. Incentive Mechanism Design in Other Applications

Many incentive mechanisms have been proposed in other applications, such as Device-to-Device (D2D) based content distribution [40, 41] and content distribution using Peer-to-Peer (P2P) protocols [42]. However, the incentive mechanisms designed for content distribution cannot be used for crowdsourcing task diffusion. First, either in D2D communication or P2P network, the content distribution is a one-hop transmission, thus there is no propagation process. Second, the content requester can always receive the content successfully once the content provider sends the content to it. Thus, the concepts of influence and diffusion model are not needed. However, in our crowdsourcing setting, the goal of task diffusion is to recruit the users to perform the tasks, and the users perform the task only if they have received sufficient influence. Thus, an appropriate diffusion model is needed in crowdsourcing scenario.

C. Influence Evaluation in Social Network

In [18], Kempe *et al.* proposed the two most popular and basic influence diffusion models: independent cascade model (IC) model and linear threshold (LT) model. Cheng *et al.* minimized the number of initial users while given quantity of users are influenced [20]. Zhu *et al.* studied how influence may use the initial users with minimum cost to spread its information to a certain threshold under competitors' hinder [19]. However, these classic influence diffusion models only consider the influence diffusion for single item. In our crowdsourcing scenario, we need to diffusion a set of tasks, and the extension of classic influence diffusion model is necessary.

The influence evaluation methods in social network have been widely studied. The traditional influence evaluation methods identify influential nodes based on the global knowledge of network structure metrics, such as degree centrality [28], betweenness centrality [26] and closeness centrality [27]. However, the computation complexity of calculating the betweenness and closeness centrality for all node pairs in the large-scale social networks, which usually contain millions of users, is very high [43, 44]. Moreover, it is hard to obtain the global knowledge of network structure in crowdsourcing system. Although many social network service sites have developed their own crowdsourcing systems, the limited knowledge, such as local network structure, can be obtained in most practical situations. Recently, Qiu *et al.* designed an end-to-end framework, *DeepInf*, to learn users' latent feature representation for predicting social influence [45]. However, *DeepInf* takes the shortest paths between all users, which is also hard to obtain, as the input of graph neural network. The character of crowdsourcing task diffusion is that the influence of users not only depends on the network structure, but also the tasks to be diffused. This is because the uses usually have different influence for different tasks. However, influence evaluation methods based on network structure metrics [26, 27, 28] cannot capture this character of crowdsourcing task diffusion.

D. Crowdsourcing based on Social Network

Some research works have studied the impact of the social relationship on the participation of workers by introduction of positive externality. Nie *et al.* obtained the *Nash Equilibrium* using *Stackelberg* game in complete and incomplete information scenarios, respectively [46]. The basic idea of [47] is to leverage the social ties among participants to promote global cooperation. A three-stage *Stackelberg* game is proposed to determine the strategy (contribution to the system) of each participant. These works aim to study the impact of the social relationship on the crowdsourcing system, rather than the problem of social user recruitment.

Recently, the method of task diffusion via social network are proposed to solve the insufficient participation problem in crowdsourcing systems. Xu *et al.* proposed a two-tiered social crowdsourcing architecture, where a set of agents are selected to diffuse the tasks to the social neighbors [48]. The objective is maximizing value from the winners' services under the budget constraint. Han *et al.* investigated a participant recruitment problem in which an initial set of recruited nodes (called seeds) need to make an optimal decision on what other nodes to recruit to perform the crowdsourcing task with the objective of maximizing the utility of seeds [49]. However, the cost of diffusion and the corresponding incentive to diffusers are neglected in the aforementioned studies.

Some incentive mechanisms have been proposed to stimulate the task diffusers. Wang *et al.* [29] proposed a social-network-assisted incentive mechanism to maximize the sensing coverage under the monetary budget constraint. In detail, they selected influential initial seeds from the MCS platform to recruit their social friends as new workers. The IC model and LT model with random base influence are extended by integrating mobile crowdsensing specific factors. Wang *et al.* [30] focused on the insufficient participation problem of MCS systems with limited number of workers, and propose to leverage social network to recruit workers for task completion as well as expanding the worker pool. However, the aforementioned two research provide a predefined reward to the initial users and cannot guarantee the individual rationality of initial users. In addition, both [29] and [30] need to calculate the influence between the social uses based on the predefined base influence, which is difficult to obtain in practice. Based on a three-layer network model, Chen *et al.* designed incentive mechanisms for both intermediaries and the crowdsensing platform and provided a solution to cope with the problem of user overlapping among intermediaries [50]. However, they did not specify the task diffusion model and influence computation. Different from the research above mentioned, this paper aims to propose an auction-based incentive mechanism to achieve the desirable economic properties and the guaranteed performance of task diffusion.

III. System Model and Desirable Properties

Although the traditional linear threshold model and independent cascade model for influence diffusion [18] have been applied in many different contexts, they cannot be directly applied in crowdsourcing systems. In general, multiple large-scale tasks need to be diffused simultaneously in crowdsourcing systems. Moreover, the crowdsourcing systems may have some specific demands for performing tasks, e.g., requiring multiple copies for each task from different users in order to improve the diversity or quality.

In this section, we model the crowdsourcing task diffusion as a reverse auction problem and present two different task diffusion models. Our task diffusion models derive from *linear threshold model*, which is proposed for modeling the collective behavior from the perspective of sociology [51], and *independent cascade model*, which is based on work in interacting particle systems [52] from probability theory. Both of them are widely applied in various areas, such as diffusion of innovations, rumors and diseases, strikes, voting, migration, and experimental social psychology.

In the research community of social network, *Linear threshold model* and *independent cascade model* are two mainstream models to characterize influence diffusion. Many studies of influence propagation conducted a comparative study by using both *models* [18, 20]. The recent work for crowdsourcing task diffusion also used the extensions of these two models [29]. In this paper, we also model the task diffusion process based on these two mainstream diffusion models to verify the feasibility of crowdsourcing task diffusion and conduct a comparative study for both models.

In our linear task diffusion model, a user can be influenced by each neighboring registered user according to the cumulative probability. While in the independent cascade task diffusion model, the registered user only has a single chance to influence each currently social neighbor. Moreover, we present some desirable properties of incentive mechanisms.

A. Auction Model

We consider a crowdsourcing platform operated by an online social network site. The platform publicizes a set $T = \{t_1, t_2, ..., t_m\}$ of m cooperative tasks. Each task $t_j \in T$ is associated with a *diffusion requirement* $r_j \in Z^+$, which is the least desirable number of unregistered users for performing

t_j through the task diffusion. Let $\mathbf{r} = (r_1, r_2, ..., r_m)$ be the *diffusion requirement profile* for all tasks. Each task $t_j \in T$ has a type a_j, such as target language and industry involved in Stepes [21], and image type in Google Image Labeler [22]. The task types are predefined by the platform, and different tasks can have the same task type. Let $\mathbf{a} = (a_1, a_2, ..., a_m)$ be the task type profile.

Assume that a set of registered users $U_R = \{1, 2, ..., n\}$ are interested in diffusing crowdsourcing tasks. Each registered user $i \in U_R$ submits a bid $B_i = (T_i, b_i)$, $T_i \subseteq T$. The task set T_i is associated with the cost c_i, which is the private information and known only to i. b_i is the claimed cost, which is the bid price that i wants to charge for diffusing T_i. If any registered user i is selected as a winner for task diffusion, he/she would diffuse T_i to his/her social neighbors excluding users in U_R. We consider that the registered users can always perform the tasks. Thus, we only diffuse the tasks to the unregistered users. To select the winners, the set of influence ϑ of every registered user when he/she diffuses any task to every social neighbor is estimated.

Given the task set T, the bid profile $\mathbf{B} = (B_1, B_2, ..., B_n)$, and the influence set ϑ, the incentive mechanism $\mathcal{M}(T, \mathbf{B}, \vartheta)$ outputs a winner set $S \subseteq U_R$ and a payment profile $\mathbf{p} = (p_1, p_2, ..., p_n)$ for each registered user. We define the utility of winning registered user i as the difference between the payment and its real cost:

$$u_i = p_i - c_i \tag{1}$$

Specifically, the utility of the losers would be zero because they are paid nothing in our designed mechanisms and there is no cost for task diffusion.

Since we consider the registered users are selfish and rational individuals, thus each registered user can behave strategically by submitting a dishonest bid price to maximize its utility.

In order to prevent the monopoly, we assume that all tasks still can be completed if any single registered user does not participate in the auction. This assumption is reasonable for crowdsourcing systems as made in [13, 53]. If a task can only be completed through the diffusion via a specific registered user, the platform can simply remove it from T.

B. Linear Task Diffusion Model

Let U_S be the set of social neighbors of all registered users excluding themselves, i.e., $U_R \cap U_S = \emptyset$. The size of U_S is q. We denote $\varphi_{i,v}(t_j)$ as the influence of when i diffuses task t_j to $v \in U_S$. Let ϑ be the set of $\varphi_{i,v}(t_j)$ for $\forall i \in U_R$, $\forall v \in U_S$. The influence to user v represents the probability that user v performs t_j. In the linear task diffusion model, the influence is cumulative. Thus, the probability of that user v performs t_j would increase when multiple registered users diffuse t_j to user v.

The objective of the incentive mechanism is minimizing the social cost such that each of tasks in T can be completed in expectation through the diffusion. We refer this problem as the *Social Optimization Task Diffusion* (*SOTD*) problem, which can be formulated as follows:

$$\min \sum\nolimits_{i \in U_R} x_i b_i \tag{2}$$

$$s.t. \quad \sum\nolimits_{i \in U_R} (x_i \sum\nolimits_{v \in U_S} \varphi_{i,v}(t_j)) \geq r_j, \ \forall t_j \in T \tag{3}$$

$$x_i \in \{0, 1\}, \ \forall i \in U_R \tag{4}$$

where x_i is the binary variable for each registered user $i \in U_R$. Let $x_i = 1$ if i is a winner; otherwise, $x_i = 0$.

Remark: Although the real cost c_i is only known by registered user i, we will prove that claiming a false cost cannot help to increase the utility of registered user i in our designed mechanisms. Thus we still use b_i when we attempt to minimize the social cost in the mechanisms designed below.

C. Independent Cascade Task Diffusion Model

In this model, the reverse auction process is the same as those in Section II-*A*. However, the influence is calculated differently. In this model, the registered user only has a single chance to influence the current social neighbors. After that, he/she cannot make any further attempts to influence the same social neighbors. Thus the probability that user v performs t_j is not cumulative, but only depends on the registered user, who is diffusing task to v currently. Essentially, the independent cascade task diffusion model considers that multiple registered users diffuse the tasks serially, thus the winners are selected sequentially. The *Social Optimization Task Diffusion* (*SOTD*) problem in the independent cascade task diffusion model can be formulated as follows:

$$\min \sum\nolimits_{i \in U_R} x_i b_i \tag{5}$$

$$s.t. \ \sum\nolimits_{i \in U_R} x_i (f_j(S_0 \cup \{i\}) - f_j(S_0)) \geq r_j, \ \forall t_j \in T \tag{6}$$

$$x_i \in \{0, 1\}, \ \forall i \in U_R \tag{7}$$

where x_i is the binary variable for each registered user $i \in U_R$. Let $x_i = 1$ if i is a winner; otherwise, $x_i = 0$. S_0 is the winner set when considering i. $f_j(S_0)$ is the influence of S_0 when diffusing t_j.

We need define $f_j(S_0)$ for any registered user subset S_0 according to the constraint of *SOTD* problem: each of tasks in T can be completed in expectation through the diffusion. Therefore, $f_j(S_0)$ should exactly represent the number of users in expectation in U_S to perform t_j when the registered users in S_0 diffuse t_j to U_S. We define $f_j(S_0)$ as the summation of joint probability of each user in U_S to perform t_j:

$$f_j(S_0) = \sum\nolimits_{v \in U_S} (1 - \prod\nolimits_{i \in S_0, t_j \in T_i} (1 - \varphi_{i,v}(t_j))) \tag{8}$$

D. Desirable Properties

Our objective is to design the incentive mechanisms satisfying the following four desirable properties:

- **Computational Efficiency:** An incentive mechanism $\mathcal{M}$ is computationally efficient if the outcome can be computed in polynomial time.
- **Individual Rationality:** Each registered user will have a non-negative utility when bidding its true cost, i.e., $u_i \geq 0, \forall i \in U_R$.
- **Truthfulness:** An incentive mechanism is truthful if reporting the true cost is a weakly dominant strategy for all registered users. In other words, no registered user can

improve its utility by submitting a false cost, no matter what others submit.

- **Social Optimization:** The objective is minimizing the social cost. We attempt to find optimal solution or approximation algorithm with low approximation ratio when there is no optimal solution terminated in polynomial time. For the latter, the approximation ratio is the ratio between approximation solution and the optimal solution.

The importance of the first two properties is obvious, because they together assure the feasibility of the incentive mechanism. The last two properties are indispensable for guaranteeing the compatibility and high performance. Being truthful, the incentive mechanisms can eliminate the fear of market manipulation and the overhead of strategizing over others for the registered users.

We list the frequently used notations in Table. II.

TABLE II
FREQUENTLY USED NOTATIONS

Symbol	Description
$T,\ m,\ t_j$	task set, number of tasks, task j
$\mathbf{r},\ r_j$	*diffusion requirement* profile, *diffusion requirement* of task t_j
$\mathbf{a},\ a_j$	task type profile, type of task t_j
$U_R,\ n$	set of registered users, number of registered users
T_i	task set of registered user i
$b_i,\ c_i$	bidding price of registered user i, cost of registered user i
$\mathbf{B},\ B_i$	bid profile, bid of registered user i
$U_S,\ q$	set of social neighbors, number of social neighbors
$\varphi_{i,v}\left(t_j\right)$	influence of registered user i to social neighbor v when diffusing task t_j
$v_i(t_j)$	influence of user i when diffusing t_j
$\mathbf{p},\ p_i$	payment profile, payment of registered user i
u_i	utility of registered user i
S	winner set
$f_j(S_0)$	influence of S_0 when diffusing t_j
$Jac(i,v)$	*Jaccard Similarity Coefficient* of i and v
$N_i,\ N_v$	set of social neighbors of i and v
U_G	set of all social users
$Shell(i)$	$\mathcal{K}-$shell index of registered user i
$Path(i,v)$	shortest path from i to v

IV. Influence Estimation

Before introducing our incentive mechanisms, we present the methods for estimating influence $\varphi_{i,v}(t_j)$, which is a key parameter of our incentive mechanisms for every social neighbor pair (i,v), $i \in U_R, v \in U_S, t_j \in T$. The influence should reflect the real-life relationships from the registered users to their social neighbors. Since we consider the crowdsourcing platform is operated by a social network, the crowdsourcing platform can obtain the social information with a certain level.

In this section, we propose two influence estimation methods: *Topology based Influence Estimation* (*TIE*) and *History based Influence Estimation (HIE)*. The *TIE* estimates the influence solely based on the topology of the social network. Therefore, it outputs the identical influence for all tasks. While the *HIE* estimates the influence according to the diffusion history between registered users and their social neighbors. The *HIE* can provide different influence values for different tasks. In general, the influence estimated by *TIE* is larger than that by *HIE*, and the social cost of *TIE* based mechanisms is smaller that of *HIE* based mechanisms. However, *HIE* based mechanisms can complete more tasks than *TIE* based mechanisms. Thus, the choose of influence estimation methods depends on the preference of crowdsourcing platform over social cost and task completion rate. We will give the detailed analysis in simulation section. In this paper, we present both *TIE* and *HIE* to fulfill the different requirement of crowdsourcing platform, improving the flexibility of our incentive mechanisms.

A. Topology based Influence Estimation

The *TIE* method relies on the observation that the relationship between two social neighbors is closer if they share more common social neighbors [54]. We use the *Jaccard Similarity Coefficient* to measure the relationship between two social users:

$$Jac(i,v) = \frac{|N_i \cap N_v|}{|N_i \cup N_v|},\ i \in U_R \cap N_v,\ v \in U_S \tag{9}$$

where N_i and N_v represent the set of social neighbors of i and v, respectively.

Then the influence of user i to user v when user i diffuses task t_j to user v can be calculated as:

$$\varphi_{i,v}(t_j) = \frac{Jac(i,v)}{\sum_{d \in U_R \cap N_v} Jac(d,v)},\ i \in U_R \cap N_v,\ v \in U_S \tag{10}$$

Note that the influence estimated by *TIE* is irrelevant to the tasks since *TIE* only utilizes the topology information of the social network.

B. History based Influence Estimation

In the *HIE*, the influence $\varphi_{i,v}(t_j)$ is considered as the proportion contribution of user i when user v is influenced to perform the tasks with type a_j in the history.

We introduce a binary variable $\eta_{i,v}(a_j)$. $\eta_{i,v}(a_j) = 1$ iff registered user i diffused tasks with type a_j to social neighbor v historically, and user v performed the task after the diffusion. Let $\Theta_v(a_j)$ be the set of registered users satisfying $\eta_{i,v}(a_j) = 1$:

$$\Theta_v(a_j) = \{\, i \,|\, \eta_{i,v}(a_j) = 1 \,\},\ i \in U_R,\ v \in U_S \tag{11}$$

Then we define the influence $\varphi_{i,v}(t_j)$ as follows:

$$\varphi_{i,v}(t_j) = \begin{cases} \dfrac{1}{\sum_{d \in \Theta_v(a_j)} \eta_{d,v}(a_j)}, & \text{if } i \in \Theta_v(a_j) \\ 0, & \text{else} \end{cases} \tag{12}$$

V. Incentive Mechanism for the Linear Task Diffusion Model

In this section, we present an *incentive Mechanism for Task Diffusion in the Linear Model* (*MTD-L*).

A. Mechanism Design

First of all, we attempt to find an optimal algorithm for the *SOTD* problem presented in (2)~(4). Unfortunately, as the following theorem shows, the *SOTD* problem is NP-hard.

Theorem 1. *The SOTD problem in the linear task diffusion model is NP-hard.*

Proof: We consider a special case of *SOTD* problem in the linear task diffusion model, where the *diffusion requirements* for all tasks in T are the same. Let $r_j = \sigma$ for $\forall t_j \in T, \sigma > 0$, where σ is sufficiently close to 0. This means that, in this special case, any task $t_j \in T$ can be completed upon there is any registered user $i \in U_R$ with $\sum_{v\in U_S} \varphi_{i,v}(t_j) > 0$. In this way, the problem can be simplified as selecting a subset $S \subseteq U_R$ with minimum total cost such that the registered users in S can diffuse each of tasks in T with positive influence. Since each registered user can bid for a subset of T with a cost, this special problem is actually an instance of the <u>*W*</u>*eighted* <u>*S*</u>*et* <u>*C*</u>*over* (*WSC*) problem, which can be formulated as follows:

$$\min \sum\nolimits_{i\in S} b_i \tag{13}$$

$$s.t. \quad \sum\nolimits_{i\in S} \sum\nolimits_{v\in U_S} \varphi_{i,v}(t_j) > 0, \ \forall t_j \in T \tag{14}$$

Since the *WSC* problem is a well-known NP-hard problem, the *SOTD* problem in the linear task diffusion model is NP-hard. ■

Algorithm 1: MTD-L

Input: task set T, bid profile $\mathbf{B}$, influence set ϑ, registered user set U_R, social neighbor set U_S, *diffusion requirements* $\mathbf{r}$

//Winner Selection Phase

1: $S \leftarrow \emptyset, \ r_j' \leftarrow r_j$;
2: **while** $\sum_{t_j\in T} r_j' \neq 0$ **do**
3:     $i \leftarrow \arg\min_{k\in U_R\setminus S} \dfrac{b_k}{\sum_{t_j\in T_k} \min\{r_j', v_k(t_j)\}}$;
4:     $S \leftarrow S \cup \{i\}$;
5:     **foreach** $t_j \in T_i$ **do**
6:         $r_j' \leftarrow r_j' - \min\{r_j', v_i(t_j)\}$;
7:     **end for**
8: **end while**

//Payment Determination Phase

9: **foreach** $i \in U_R$ **do** $p_i \leftarrow 0$;
10: **foreach** $i \in S$ **do**
11:     $U_R' \leftarrow U_R\setminus\{i\}, \ S' \leftarrow \emptyset, \ r_j'' \leftarrow r_j$;
12:     **while** $\sum_{t_j\in T} r_j'' \neq 0$ **do**
13:         $i_k \leftarrow \arg\min_{k\in U_R'\setminus S'} \dfrac{b_k}{\sum_{t_j\in T_k} \min\{r_j'', v_k(t_j)\}}$;
14:         $S' \leftarrow S' \cup \{i_k\}$;
15:         $p_i \leftarrow \max\{p_i, \dfrac{\sum_{t_j\in T_i} \min\{r_j'', v_i(t_j)\}}{\sum_{t_j\in T_{i_k}} \min\{r_j'', v_{i_k}(t_j)\}} b_{i_k}\}$;
16:         **foreach** $t_j \in T_{i_k}$ **do**
17:             $r_j'' \leftarrow r_j'' - \min\{r_j'', v_{i_k}(t_j)\}$;
18:         **end for**
19:     **end while**
20: **end for**
21: **return** $(S, \mathbf{p})$;

Since the *SOTD* problem in the linear task diffusion model is NP-hard, it is impossible to compute the winner set with minimum social cost in polynomial time unless P=NP. In fact, there is no $(1-\varepsilon)\ln n$ approximate polynomial time algorithm for *WSC* problem [55]. In addition, we cannot use the off-the-shelf VCG mechanism [56] since the truthfulness of VCG mechanism requires that the social cost is exactly minimized. We design our incentive mechanism in the linear task diffusion model, *MTD-L*, which follows a greedy approach. Illustrated in Algorithm 1, *MTD-L* consists of winner selection phase and payment determination phase.

In the winner selection phase, the registered users are sorted according to the effective influence unit cost. Given any task $t_j \in T$, the influence of registered user i is $v_i(t_j) = \sum_{v\in U_S} \varphi_{i,v}(t_j)$. The effective influence unit cost of registered user i is defined as $\dfrac{b_i}{\sum_{t_j\in T_i} \min\{r_j', v_i(t_j)\}}$. In each iteration of the winner selection phase, we select the registered user with minimum effective influence unit cost over the unselected registered user set $U\setminus S$ as the winner until the winners' influence can meet the diffusion requirement for each task in T.

In payment determination phase, for each winner $i \in S$, we execute the winner selection phase over $U\setminus\{i\}$, and the winner set is denoted as S'. We compute the maximum price that registered user i can be selected instead of each registered user in S'. We will prove that this price is a critical payment for registered user i later.

B. Mechanism Analysis

In the following, we present the theoretical analysis, demonstrating that *MTD-L* can achieve the desired properties of computational efficiency, individual rationality, truthfulness, and low approximation ratio.

Lemma 1. *MTD-L is computationally efficient.*

Proof: Finding the registered user with minimum effective influence unit cost takes $O(nmq)$, where computing the value of $\sum_{t_j\in T_k} \min\{r_j', v_k(t_j)\}$ takes $O(mq)$. Hence, the while-loop (line2-8) takes $O(n^2mq)$. In each iteration of the for-loop (line10-20), a process similar to line2-8 is executed. Hence the time complexity of the whole auction is dominated by this for-loop, which is bounded by $O(n^3mq)$.

Note that the influence estimated by *TIE* is irrelevant to the tasks, i.e., $\varphi_{i,v}(t_j) = \varphi_{i,v}(t_k)$ for all $t_j \in T_i, t_k \in T_i, t_j \neq t_k$. Hence computing the value of $\sum_{t_j\in T_k} \min\{r_j', v_k(t_j)\}$ can only take $O(q)$. Therefore, the time complexity of the auction is $O(n^2 \cdot \max\{nq, m\})$ when using *TIE*. ■

Lemma 2. *MTD-L is individually rational.*

Proof: Let i_k be registered user i's replacement which appears in the ith place in the sorting over $U_R\setminus\{i\}$. Since registered user i_k would not be at ith place if i is considered, we have $\dfrac{b_i}{\sum_{t_j\in T_i} \min\{r_j', v_i(t_j)\}} \leq \dfrac{b_{i_k}}{\sum_{t_j\in T_{i_k}} \min\{r_j', v_{i_k}(t_j)\}}$. Hence $b_i \leq \dfrac{\sum_{t_j\in T_i} \min\{r_j', v_i(t_j)\}}{\sum_{t_j\in T_{i_k}} \min\{r_j', v_{i_k}(t_j)\}} b_{i_k} = \dfrac{\sum_{t_j\in T_i} \min\{r_j'', v_i(t_j)\}}{\sum_{t_j\in T_{i_k}} \min\{r_j'', v_{i_k}(t_j)\}} b_{i_k}$, where the equality relies on the observation that $r_j' = r_j''$ for every $k \leq i$, which is due to the fact that $S = S'$

for every $k \leq i$. This is sufficient to guarantee $b_i \leq \max_{k\in U_R\setminus S'} \frac{\sum_{t_j\in T_i}\min\{r_j'', v_i(t_j)\}}{\sum_{t_j\in T_{i_k}}\min\{r_j'', v_{i_k}(t_j)\}} b_{i_k} = p_i$. ∎

Before analyzing the truthfulness of *MTD-L*, we first introduce the Myerson's Theorem [57].

Theorem 2. *([32, Theorem 2.1]) An auction mechanism is truthful if and only if:*

- *The selection rule is monotone: If user i wins the auction by bidding bi, it also wins by bidding $b_i' < b_i$;*
- *Each winner is paid the critical value: User i would not win the auction if it bids higher than this value.*

Lemma 3. *MLT-L is truthful.*

Proof: Based on Theorem 2, it suffices to prove that the selection rule of *MLT-L* is monotone and the payment p_i for each i is the critical value. The monotonicity of the selection rule is obvious as bidding a smaller value cannot push registered user i backwards in the sorting.

We next show that p_i is the critical value for registered user i in the sense that bidding higher p_i could prevent registered user i from winning the auction. Note that $p_i = \max_{k\in\{1,\ldots,e\}} \frac{\sum_{t_j\in T_i}\min\{r_j'', v_i(t_j)\}}{\sum_{t_j\in T_{i_k}}\min\{r_j'', v_{i_k}(t_j)\}} b_{i_k}$. If registered user i bids $b_i \geq p_i$, it will be placed after e since $b_i \geq \frac{\sum_{t_j\in T_i}\min\{r_j'', v_i(t_j)\}}{\sum_{t_j\in T_{i_e}}\min\{r_j'', v_{i_e}(t_j)\}} b_{i_e}$ implies $\frac{b_i}{\sum_{t_j\in T_i}\min\{r_j'', v_i(t_j)\}} \geq \frac{b_{i_e}}{\sum_{t_j\in T_{i_e}}\min\{r_j'', v_{i_e}(t_j)\}}$. Hence, registered user i would not win the auction because the first e registered users have met the diffusion requirement for each task in T. ∎

Then, we provide our analysis about the approximation ratio of the *MTD-L* auction using the dual fitting method [58]. We formulate the linear program relaxation of the *SOTD* problem defined in (2)~(4) as the normalized primal linear program **P**. The dual program is formulated in program **D**.

$$\mathbf{P}: \min \sum\nolimits_{i\in U_R} b_i x_i \tag{15}$$

$$s.t. \quad \sum\nolimits_{i:i\in U_R, t_j\in T_i} (v_i(t_j)x_i) \geq r_j, \ \forall t_j \in T \tag{16}$$

$$0 \leq x_i \leq 1, \ \forall i \in U_R \tag{17}$$

$$\mathbf{D}: \max \sum\nolimits_{t_j\in T} r_j y_j - \sum\nolimits_{i\in U_R} z_i \tag{18}$$

$$s.t. \quad \sum\nolimits_{t_j\in T_i} (v_i(t_j)y_j) - z_i \leq b_i, \ \forall i \in U_R \tag{19}$$

$$y_j \geq 0, \ \forall t_j \in T \tag{20}$$

$$z_i \geq 0, \ \forall i \in U_R \tag{21}$$

We define any task $t_j \in T$ as alive at any iteration in winner selection phase if its diffusion requirement is not fully satisfied. We define that task t_j is covered by T_i if $t_j \in T_i$ and t_j is alive when registered user i is selected. The coverage relationship is represented as $t_j \prec T_i$. Moreover, we define the minimum influence as Δv. Suppose when registered user i is selected, the residual diffusion requirement profile is $\{r_1^*, r_2^*, ..., r_m^*\}$ and T_i is the i_jth set that covers t_j, the corresponding normalized effective influence unit cost in terms of unit influence can be represented in (22):

$$w(t_j, i_j) = \frac{b_i \Delta v}{\sum_{t_j\in T_i}\min\{r_j^*, v_i(t_j)\}} \tag{22}$$

We assume that t_j is covered by h_j sets. Then we have $w(t_j, 1) \leq ... \leq w(t_j, h_j)$. We then define two constants $\Omega = \frac{1}{\Delta v}\sum_{t_j\in T} r_j$ and $\varepsilon = \max v_i(t_j)\,|T_i|\, b_i, i \in U_R, t_j \in T$.

Lemma 4. *The following pairs $(y_j, z_i), j \in T, i \in U_R$ are feasible to the dual program* **D**.

$$y_j = \frac{w(t_j, h_j)}{2\varepsilon H_n \Delta v}, \ \forall t_j \in T,$$

$$z_i = \begin{cases} \dfrac{\sum_{t_j\prec T_i}\left(\min\{r_j^*, v_i(t_j)\}(w(t_j,h_j) - w(t_j,i_j))\right)}{2\varepsilon H_\Omega \Delta v}, & i \in S \\ 0, & i \notin S \end{cases}$$

where $H_n = 1 + \frac{1}{2} + ... + \frac{1}{n}$, $H_\Omega = 1 + \frac{1}{2} + ... + \frac{1}{\Omega}$.

Proof: Suppose for any registered user $i \in U_R$, there are s_i tasks in T_i. We reorder these tasks in the order in which they are fully covered.

If $i \notin S$, then we have $z_i = 0$. Suppose when the last unit *diffusion requirement* of t_j is covered, the residual *diffusion requirement* profile is $\{r_1^+, r_2^+, ..., r_m^+\}$, then the total residual *diffusion requirements* of alive tasks contained by T_i are represented as $\sum_{k=j}^{s_i}\min\{r_k^+, v_i(t_k)\}$. We have

$$w(t_j, h_j) \leq \frac{b_i \Delta v}{\sum_{k=j}^{s_i}\min\{r_k^+, v_i(t_k)\}}$$

Therefore, we have

$$\begin{aligned}\sum\nolimits_{j=1}^{S_i}(v_i(t_j)y_j) - z_i &\leq \sum\nolimits_{j=1}^{S_i}\frac{v_i(t_j)b_i}{2\varepsilon H_\Omega \sum_{k=j}^{s_i}\min\{r_k^+, v_i(t_k)\}} - 0 \\ &\leq \frac{b_i}{H_\Omega}\left(1 + \frac{1}{2} + ... + \frac{1}{\Omega}\right) \leq b_i\end{aligned}$$

If registered user $i \in S$, then we assume that when registered user i is selected as a winner, s_i' tasks in T_i already been fully covered. We have

$$\begin{aligned}&\sum\nolimits_{j=1}^{S_i}(v_i(t_j)y_j) - z_i \\ &= \frac{\sum_{j=1}^{s_i}(w(t_j,h_j)v_i(t_j))}{2\varepsilon H_\Omega \Delta v} \\ &\quad - \frac{\sum_{j=s_i'+1}^{s_i}\min\{r_j^*, v_i(t_j)\}\left(w(t_j,h_j) - w(t_j,i_j)\right)}{2\varepsilon H_\Omega \Delta v} \\ &= \frac{\sum_{j=1}^{s_i'}(w(t_j,h_j)v_i(t_j))}{2\varepsilon H_\Omega \Delta v} + \frac{\sum_{j=s_i'+1}^{s_i}\min\{r_j^*, v_i(t_j)\}w(t_j,i_j)}{2\varepsilon H_\Omega \Delta v} \\ &\quad + \frac{\sum_{j=s_i'+1}^{s_i}(v_i(t_j)) - \min\{r_j^*, v_i(t_j)\}w(t_j,h_j)}{2\varepsilon H_\Omega \Delta v} \\ &\leq \frac{\sum_{j=1}^{s_i'}(w(t_j,h_j)v_i(t_j))}{2\varepsilon H_\Omega \Delta v} + \frac{\sum_{j=s_i'+1}^{s_i}\min\{r_j^*, v_i(t_j)\}w(t_j,i_j)}{2\varepsilon H_\Omega \Delta v} \\ &= \sum\nolimits_{j=1}^{s_i'}\frac{v_i(t_j)b_i}{2\varepsilon H_\Omega \sum_{k=j}^{s_i}\min\{r_k^*, v_i(t_j)\}} + \frac{b_i}{2\varepsilon H_\Omega} \\ &\leq \frac{b_i}{H_\Omega}\left(\frac{1}{s_i} + ... + \frac{1}{s_i - s_i' + 1} + 1\right) \leq b_i\end{aligned}$$

Hence, the pairs $(y_j, z_i), j \in T, i \in U_R$ are feasible to the dual program **D**. ■

Lemma 5. *MTD-L can approximate the optimal solution within a factor of* $2\varepsilon H_\Omega$.

Proof: By substituting the dual solution given in Lemma 4 into (18), we have

$$\begin{aligned}
&\sum\nolimits_{t_j \in T} r_j y_j - \sum\nolimits_{i \in U_R} z_i \\
&= \frac{\sum_{i \in S} \sum_{t_j < T_i} \left(\min\{r_j{}^*, v_i(t_j)\} \left(w(t_j, h_j) - w(t_j, i_j)\right)\right)}{2\varepsilon H_\Omega \Delta v} \\
&+ \frac{\sum_{t_j \in T} r_j w(t_j, h_j)}{2\varepsilon H_\Omega \Delta v} \\
&= \frac{\sum_{i \in S} \sum_{t_j < T_i} \min\{r_j{}^*, v_i(t_j)\} \frac{b_i \Delta v}{\sum_{t_j \in T_i} \min\{r_j{}^*, v_i(t_j)\}}}{2\varepsilon H_\Omega \Delta v} \\
&= \frac{\sum_{i \in S} b_i}{2\varepsilon H_\Omega} \le OPT
\end{aligned}$$

■

The above lemmas together prove the following theorem.

Theorem 3. *MTD-L is computationally efficient, individually rational, truthful and* $2\varepsilon H_\Omega$ *approximate for the linear task diffusion model.*

VI. Incentive Mechanism for the Independent Cascade Task Diffusion Model

In this section, we present an *incentive Mechanism for Task Diffusion in the Independent Cascade Model* (*MTD-IC*).

A. Mechanism Design

We first analyze the hardness of the *SOTD* problem in the independent cascade task diffusion model. The following theorem shows that it is hopeless to find the optimal solution in polynomial time.

Theorem 4. *The SOTD problem in the independent cascade task diffusion model is NP-hard.*

Proof: The proof is similar with that of Theorem 1. We consider a special case of *SOTD* problem in the independent cascade task diffusion model, where any task $t_j \in T$ can be completed upon there is any registered user $i \in U_R$ with $f_j(S_0 \cup \{i\}) - f_j(S_0) > 0$. This problem can be formulated as follows:

$$\min \sum\nolimits_{i \in S} b_i \tag{23}$$

$$s.t. \quad f_j(S) > 0, \ \forall t_j \in T \tag{24}$$

This is actually the *WSC* problem, which is a well-known NP-hard problem. Hence the *SOTD* problem in the independent cascade task diffusion model is NP-hard. ■

Since the *SOTD* problem in the independent cascade task diffusion model is also NP-hard, we use the similar algorithm framework in Algorithm 1 to design *MTD-IC*, which also follows a greedy approach and consists of winner selection phase and payment determination phase.

Different from *MTD-L*, the registered users are sorted according to the marginal influence unit cost. Given arbitrary task $t_j \in T$ and any register user subset S_0, the marginal influence of registered user i is $f_j(S_0 \cup \{i\}) - f_j(S_0)$. The marginal influence unit cost of registered user i is defined as $\frac{b_i}{\sum_{t_j \in T_i} \min\{r_j{}', f_j(S_0 \cup \{i\}) - f_j(S_0)\}}$. The payment rule is similar with that of *MTD-L*.

B. A Walk-Through Example

We use an example in Fig.2 to show how *MTD-IC* works. In this example, there are two tasks with $r_1 = r_2 = 2$ and three registered users with $B_1 = (\{1, 2\}, 6)$, $B_2 = (\{1, 2\}, 5)$, $B_3 = (\{1, 2\}, 4)$. $U_S = \{1, 2, 3, 4, 5\}$.

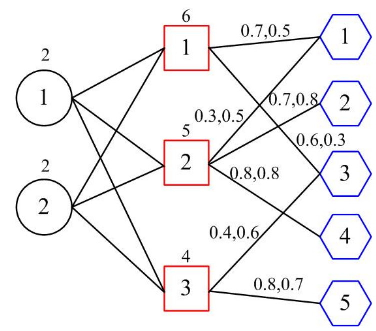


Fig. 2. Illustration for *MTD-IC*. The disks represent tasks, the squares represent registered users and the hexagons represent social neighbors. The numbers above the disks represent the *diffusion requirements*. The numbers above the squares represent the claimed prices. The first number above the lines connecting any registered user i and its social neighbor v represents the influence for task 1, i.e., $\varphi_{i,v}(t_1)$, and the second number represents the influence for task 2, i.e., $\varphi_{i,v}(t_2)$.

<u>Winner Selection:</u>

- First round: $S = \emptyset$, $r_1' = 2, r_2{}' = 2$.

$$\frac{b_1}{\min\{r_1{}', f_1(\emptyset \cup \{1\}) - f_1(\emptyset)\} + \min\{r_2{}', f_2(\emptyset \cup \{1\}) - f_2(\emptyset)\}}$$
$= 6/(\min\{2, f_1(\{1\})\} + \min\{2, f_2(\{1\})\}) = 6/2.1$;
$b_2/(\min\{2, f_1(\{2\})\} + \min\{2, f_2(\{2\})\}) = 5/3.8$;
$b_3/\min\{2, f_1(\{3\})\} + \min\{2, f_2(\{3\})\} = 4/2.5$. User 2 wins.

- Second round: $S = \{2\}$, $r_1{}' = 0.2$, $r_2{}' = 0$.

$$\frac{b_1}{\min\{0.2, f_1(\{1, 2\}) - f_1(\{2\})\} + \min\{0, f_2(\{1, 2\}) - f_2(\{2\})\}}$$
$= 6/\min\{0.2, 2.89 - 1.8\} = 6/0.2$;
$$\frac{b_3}{\min\{0.2, f_1(\{3, 2\}) - f_1(\{2\})\} + \min\{0, f_2(\{3, 2\}) - f_2(\{2\})\}}$$
$= 4/\min\{0.2, 2.6 - 1.8\} = 4/0.2$.

User 3 wins. $r_1{}' = 0$, $r_2{}' = 0$. $S = \{2, 3\}$.

<u>Payment Determination:</u>

- Payment for user 2: the winners are 3, 1 orderly.

$r_1{}'' = 2$, $r_2{}'' = 2$.
$$\frac{\min\{2, f_1(\{2\}) - f_1(\emptyset)\} + \min\{2, f_2(\{2\}) - f_2(\emptyset)\}}{\min\{2, f_1(\{3\}) - f_1(\emptyset)\} + \min\{2, f_2(\{3\}) - f_2(\emptyset)\}} b_3 = \frac{3.8}{2.5} \times 4.$$
$r_1{}'' = 0.8$, $r_2{}'' = 0.7$.

$$\frac{\min\{0.8, f_1(\{2,3\}) - f_1(\{3\})\} + \min\{0.7, f_2(\{2,3\}) - f_2(\{3\})\}}{\min\{0.8, f_1(\{1,3\}) - f_1(\{3\})\} + \min\{0.7, f_2(\{1,3\}) - f_2(\{3\})\}} b_1$$
$= 1 \times 6 = 6.$

$r_1'' = 0,\ r_2'' = 0.\ p_2 = \max\{\frac{3.8}{2.5} \times 4, 6\} \approx 6.08.$

- Payment for user 3: the winners are 2, 1 orderly.

$r_1'' = 2,\ r_2'' = 2.$
$$\frac{\min\{2, f_1(\{3\}) - f_1(\emptyset)\} + \min\{2, f_2(\{3\}) - f_2(\emptyset)\}}{\min\{2, f_1(\{2\}) - f_1(\emptyset)\} + \min\{2, f_2(\{2\}) - f_2(\emptyset)\}} b_2$$
$= \frac{2.5}{3.8} \times 5.$
$r_1'' = 0.2, r_2'' = 0.$
$$\frac{\min\{0.2, f_1(\{2,3\}) - f_1(\{2\})\} + \min\{0, f_2(\{2,3\}) - f_2(\{2\})\}}{\min\{0.2, f_1(\{1,2\}) - f_1(\{2\})\} + \min\{0, f_2(\{1,2\}) - f_2(\{2\})\}} b_1$$
$= 1 \times 6 = 6.$

$r_1'' = 0,\ r_2'' = 0.\ p_3 = \max\{\frac{2.5}{3.8} \times 5, 6\} = 6.$

C. Mechanism Analysis

Theorem 5. *MTD-IC is computationally efficient, individually rational, truthful and $2\varepsilon H_\Omega$ approximate for the independent cascade task diffusion model, where $\varepsilon = \max |T_i| b_i f_j(\{i\}), i \in U_R, j \in T$.*

Proof: Finding the registered user with minimum marginal influence unit cost takes $O(n^2 mq)$, where computing the value of $\sum_{t_j \in T_k} \min\{r_j', f_j(S_0 \cup \{k\}) - f_j(S_0)\}$ takes $O(mqn)$. Updating r_j' takes $O(m)$ time since each registered user can bid at most m tasks. Hence, the while-loop (line2-8) takes $O(n^3 mq)$. In each iteration of the for-loop (line10-20), a process similar to line2-8 is executed. Hence the time complexity of the whole auction is dominated by this for-loop, which is bounded by $O(n^4 mq)$.

Based on the similar analysis in Lemma 1, the time complexity of the auction is $O(n^2 \cdot \max\{n^2 q, m\})$ when using the influence estimated by *TIE*.

The proof of the individually rational, truthful and approximation ratio of *MTD-IC* is similar to that of *MTD-L*. ■

VII. Global Influence Estimation

So far, we have proposed two incentive mechanisms, which stimulate the registered users to diffuse the crowdsourcing tasks to their social neighbors. In practice, the diffusion behavior of a social user not only influences its social neighbors but also other users in the whole social community. In the scenario of crowdsourcing task diffusion, the social neighbors of winning registered users can continue the task diffusion due to the high influence of winning registered users. In the view of incentive mechanism design, we need a global influence estimation method for the registered users before executing the reverse auction. In this section, we propose the <u>G</u>lobal <u>I</u>nfluence <u>E</u>stimation (*GIE*).

In section III, we have proposed *TIE*, which uses the *Jaccard Similarity Coefficient* to measure the relationship between two social neighbors. For the global influence estimation, it is straightforward to calculate the multiplicative *Jaccard Similarity Coefficient* of all social neighbors in the path from any registered user i to any user $v \in U_G$, where U_G is the set of all social users excluding the registered users. However, in the context of global influence estimation, the influence of a user not only depends on the affinity with other users but also the social influence in the community.

To address this problem, we propose our *GIE* by integrating *Jaccard Similarity Coefficient* and *$\mathcal{K}-$shell decomposition* [59]. *$\mathcal{K}-$shell decomposition* has been widely used as a tool to analyze the structural properties of large graph.

We first give two definitions about $\mathcal{K}-$shell decomposition:

Definition 1 ($\mathcal{K}-$core). *Given a graph $G = (V, E)$, the subgraph G' is a $k-core$ of G if and only if it is the maximal subgraph of G such that the degree of every node in G' is at least $\mathcal{K}$.*

Definiton 2 ($\mathcal{K}-$shell). *A node is said to belong to the $\mathcal{K}$ shell if and only if it belongs to the $\mathcal{K}-core$ subgraph but not to the $\mathcal{K}+1-core$ subgraph.*

Given the graph $G = (V, E)$, where $V = U_R \cup U_G$. The $\mathcal{K}-$shell decomposition can be performed iteratively as follows. We prune all nodes with degree one and the arcs incident on them from the graph. We repeat pruning nodes with degree one on the graph iteratively until there is no node with degree one on the graph. The $\mathcal{K}-$shell index of the deleted nodes is 1. The same is done for $\mathcal{K} = 2$ ulteriorly and so on, until all nods are pruned from the graph.

Let $Shell(i)$ be the $\mathcal{K}-$shell index of any registered user $i \in U_R$. Then, for all $i \in U_R, v \in U_G$, we find the shortest path $Path(i, v)$ from i to v, and the influence of any registered user i to any social user v when i diffuses task t_j to user v can be calculated as:

$$\varphi_{i,v}(t_j) = \frac{Shell(i) \times \prod_{k,k' \in Path(i,v),(k,k') \in E} Jac(k,k')}{\sum_{d \in U_R} (Shell(d) \times \prod_{k,k' \in Path(d,v),(k,k') \in E} Jac(k,k'))} \quad (25)$$

Then, we can still use *MTD-L* and *MTD-IC* given in Section IV and Section V, respectively, to design the incentive mechanisms based on the global influence.

Lemma 6. *The time complexity of* GIE *is* $O(|V|^3)$.

Proof: Since finding the $\mathcal{K}-$shell decomposition can be computed in $O(|V| + |E|)$ [60], the time complexity of *GIE* is dominated by finding the shortest path for all nodes in graph G, which is bounded by $O(|V|^3)$.

Remark: Although *GIE* can measure the influence of any registered user i to any social user v, the time complexity is quite high. Moreover, *GIE* requires global social information of the community. Thus, *GIE* is suitable for small community.

VIII. Performance Evaluation

We have conducted thorough simulations to investigate the performance of proposed incentive mechanisms. We use the naming rule of *MTD-L/IC-T/H/G* to represent the *incentive <u>M</u>echanisms for <u>T</u>ask <u>D</u>iffusion in the <u>L</u>inear/<u>I</u>ndependent <u>C</u>ascade model using <u>TIE</u>/<u>HIE</u>/<u>GIE</u>* in our simulations. We evaluate the performance of our incentive mechanisms against *Fast-Selector* [29], which utilizes greedy-based seed selection to maximize the coverage of tasks. We replace the constraints of winner number and worker number with *diffusion requirement* proposed in our *SOTD* problem, i.e., *Fast-Selector* will keep selecting winners until all requirements are met. *Fast-Selector* sets a fixed reward for each task. We compare our mechanisms with extended *IC*-based *Fast-Selector* since the thresholds in *LT*-based mechanism are difficult to determine and have little connection with our mechanisms. Note that *Fast-Selector* is not a truthful seed selection method.

All the simulations were run on a Windows machine with Intel Xeon Platinum 8269CY and 8 GB memory. Each measurement is averaged over 100 instances.

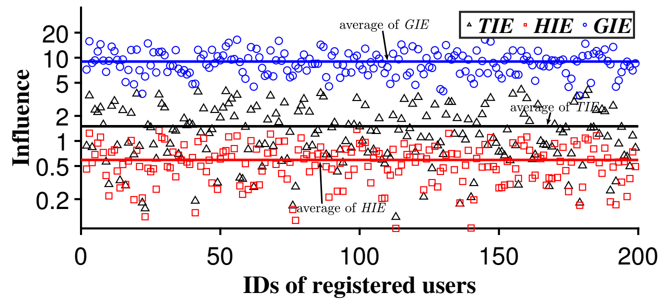


Fig. 3. Estimated influence of *TIE*, *HIE* and *GIE*

A. Simulation Setup

The simulations are based on Twitter Dataset [61], which has been built after monitoring the spreading processes of retweets on Twitter. We allocate the retweets to m crowdsourcing tasks with $\lceil m/2 \rceil$ task types. We set the default value of parameters as follows: The bidding price of users is randomly selected from the auction set [62], which contains 5017 bid prices for Palm Pilot M515 PDA from eBay. The *diffusion requirement* and the number of bidding tasks of each registered user are uniformly distributed in [2, 5] and [5, 10], respectively. Let $n = 200$, $m = 20$ be the default settings. Since the time complexity of *GIE* is quite high, we fix $|V| = 2000$, which is the number of all users in social network. However, we will vary the value of key parameters to explore the impacts of these parameters respectively. For *HIE*, we randomly choose 30% of retweets in the dataset as the diffusion history. For *Fast-Selector*, we randomly select the reward of each task from the auction set to incentivize users. We transform the two-dimensional task (temporal-spatial task) into multiple single-dimensional tasks without negative effect on calculating coverage in order to simulate the multiple cooperative tasks in our system model. Other parameters in *Fast-Selector* are set according to the original work.

We define overpayment ratio and completion rate to quantitatively measure the frugality and task diffusion performance, respectively.

Overpayment Ratio: The overpayment ratio is defined as $\left(\sum_{i\in S} p_i - \sum_{i\in S} b_i\right) / \sum_{i\in S} b_i$.

Completion Rate: let $R(S)$ be the total number of retweets from all users except the registered users in the social network for all crowdsourcing tasks based on the dataset. The completion rate is defined as $R(S) / \sum_{t_j\in T} r_j$.

We first measure the influence estimated through our influence estimation methods. Then we measure the number of winners, social cost, overpayment ratio, and completion rate, and reveal the impacts of the key parameters, including the number of registered users (n), the number of tasks (m) and diffusion requirement (r). Finally, we measure the running time of proposed influence estimation methods.

B. Influence Estimation

We first investigate the influence of 200 registered users, who are selected randomly from the Twitter network, estimated through *TIE*, *HIE* and *GIE*, respectively. The *TIE* and *HIE* influence of any registered user i is calculated as $\sum_{t_j\in T}\sum_{v\in U_S} \varphi_{i,v}(t_j)/m$. The *TIE* and *HIE* influence of any registered user i is calculated as $\sum_{t_j\in T}\sum_{v\in U_G} \varphi_{i,v}(t_j)/m$.

As shown in Fig.3, the average *TIE* influence (1.50) is larger than that of *HIE* (0.59). This is because the influence of *TIE* is estimated only using the topology, and the registered user will obtain significant influence if he/she has many common neighbors with his/her social neighbors. While in *HIE*, the registered user obtains the influence only if he/she retweets. As an example, the registered user 52 has the *TIE* influence of 3.95, however, the *HIE* influence is only 0.76. The average *GIE* influence (9.00) is quite high since it is the summation of all influence to all users in the social network, not only the social neighbors. In fact, the *GIE* influence towards a single social user is smaller than those of other two influence estimation methods.

C. Impact of n

To investigate the scalability of designed mechanisms, we vary the number of registered users from 200 to 600. As shown in Fig.4, the number of winners of our six incentive mechanisms increases with increasing registered users, while number of winners of *Fast-Selector* decreases. For our mechanisms, When the number of registered users becomes larger, the estimated influence becomes smaller, and each task needs more users to diffuse. However, the total influence of all registered users stays still. For *Fast-Selector*, the influence estimation method used is different from ours. The estimated influence (called possibility of task performance in [29]) does not change with the variation of number of registered users. The number of winners of *Fast-Selector* decreases with increasing number of registered users because *Fast-Selector* can select the seeds with more influence from a large set of registered users. There are two observations from Fig.4(a). First, the winners of *IC* model are more than that of *L* model with the same influence estimation method since *IC* model does not consider the cumulative effect of influence, and selects the registered user with minimum unit cost of joint probability difference iteratively. This implies that the *IC* model has to select more winners to satisfy the *diffusion requirements*. Second, the winners of *HIE* based mechanisms are more than that of *TIE* based mechanisms when using same diffusion model. This is because the average influence of *HIE* is less than that of *TIE*, which has been shown in Fig.3. Thus more winners are needed to meet the *diffusion requirements*. The winners of *GIE* based mechanisms are the least since

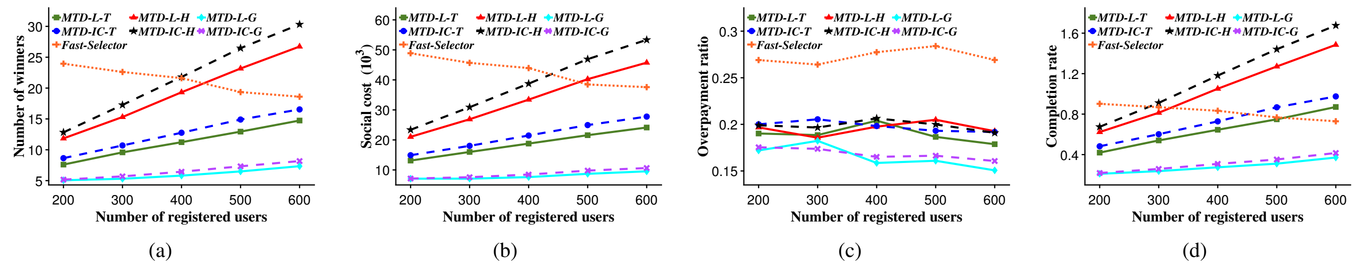


Fig. 4. Impact of the number of registered users (n): (a) Number of winners. (b) Social cost. (c) Overpayment ratio. (d) Completion rate

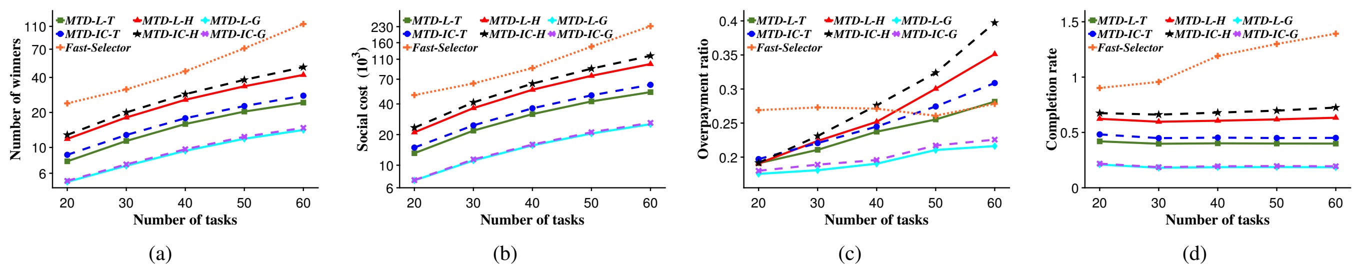


Fig. 5. Impact of the number of tasks (m): (a) Number of winners. (b) Social cost. (c) Overpayment ratio. (d) Completion rate

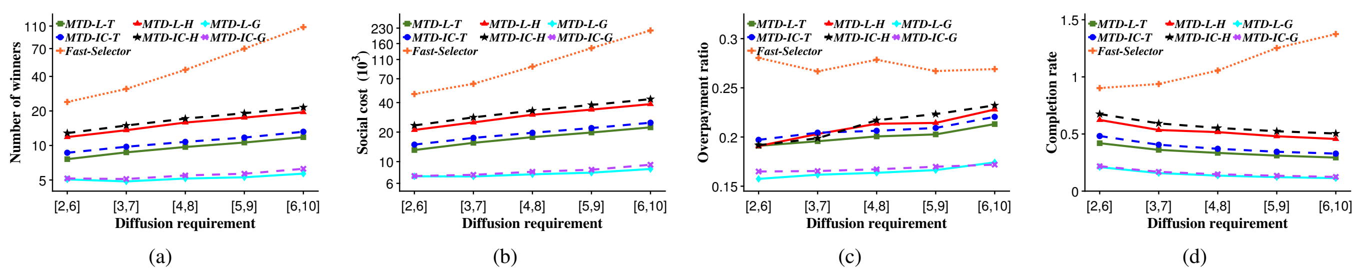


Fig. 6. Impact of diffusion requirement (r): (a) Number of winners. (b) Social cost. (c) Overpayment ratio. (d) Completion rate

the average influence of *GIE* is the largest among all three influence estimation methods.

As observed from Fig.4(b), the social cost of our mechanisms increases with the increasing number of registered users since there are more winners. The social cost largely depends on the number of winners. Thus, *GIE* based mechanisms have the smallest social cost, and the social cost of *TIE* based mechanisms is smaller than that of *HIE* based mechanisms. On the contrary, the social cost of *Fast-Selector* decreases with the decreasing number of winners.

We can see from Fig.4(c) that the overpayment ratio fluctuates with increasing registered users. In most cases, the overpayment ratio of our 6 mechanisms is lower than 0.2, achieving good frugality. The overpayment ratio is also used to represent the degree of competition among bidders. The overpayment ratio tends to be smaller with increasing number of registered users. This is because there are more choices of winners for platform to make, thus the competition among registered users is more intense. The payment of *Fast-Selector* is randomly chosen and is much higher than our mechanisms. Besides, the payment of *Fast-Selector* has little relation with variation of number of registered users, number of tasks and *diffusion requirement*.

We can see from Fig.4(d) that the increasing number of registered users helps much to complete the tasks in our mechanisms. This means that the diffusion requirement can reduce when there are sufficient registered users. Moreover, *HIE* based mechanisms show significant superiority in terms of completion rate. *MTD-IC-H* shows the best diffusion performance in most cases. The average completion rate of *MTD-IC-H* is 0.674, which is 1.4 times as high as *MTD-IC-T* (0.482) and 3.09 times as high as *MTD-IC-G* (0.218) under the default settings.The completion rate of *Fast-Selector* decreases with the decrease of winner number. An observation is that our mechanisms can achieve higher completion rate by selecting a fixed number of winners compared with *Fast-Selector*. In other words, the completion rate of *Fast-Selector* does not benefit from the increase of number of registered users.

D. Impact of m

The number of crowdsourcing tasks can depict the workload of diffusion. As shown in Fig.5, the number of winners and the social cost increase severely in our six incentive mechanisms with increasing m since the platform needs more registered users to diffuse the tasks. The overpayment ratio also increases with increasing m since the platform needs to recruit more registered users to diffuse the tasks, which mitigates the competition among registered users accordingly. The number of winners, social cost and completion rate of *Fast-Selector* increase with the increasing number of tasks.

Fast-Selector has the highest completion rate, but the social cost is 86.7%, 234.8%, 621.1% higher than *HIE*, *TIE* and *GIE* based mechanisms on average, respectively. Again, *HIE* based mechanisms show significant superiority in terms of completion rate. The impact of number of tasks is small on the average completion rate for all mechanisms since we consider that the *diffusion requirements* can be always satisfied. The average completion rate of *HIE* based mechanisms, *TIE* based mechanisms, and *GIE* mechanisms are 0.65, 0.45, and 0.21 under the default settings.

An interesting observation is that the performance difference of *MTD-L-G* and *MTD-IC-G* is very small. In fact, the two mechanisms choose almost the same winners. This is because that the influence estimated by *GIE* to single social user is too small such that the previously selected winners has small impact on the following selection in the winner selection phase. Thus the diffusion model used does not affect the selection of winners. In our simulations, only 7% outputs of the two mechanisms have different winners.

E. Impact of r

To investigate the performance for the tasks associated with different *diffusion requirement*, we vary the distribution interval of *cooperative index* from [2, 5] to [6, 10]. As can be seen from Fig.6, the number of winners and social cost increase slightly in all mechanisms with increasing *diffusion requirement* since the platform needs more registered users to diffuse each task on average, satisfying desirable least number of users for performing each task. The completion rate decreases with increasing r. For *Fast-Selector*, the change of all four metrics when the diffusion requirement increases is almost same as that when the number of tasks increases.

F. Running time

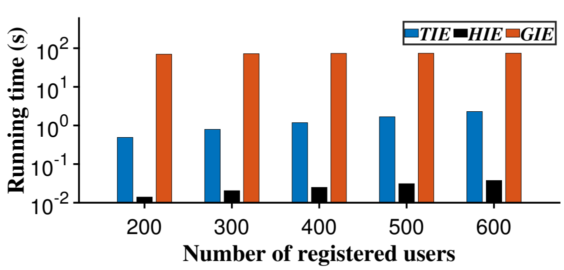


Fig. 7. Running time of *TIE*, *HIE*, and *GIE*

Finally, we test the running time of *TIE*, *HIE* and *GIE* for calculating $\varphi_{i,v}\left(t_j\right)$ for all i, v and t_j. It can be seen from Fig.7 that the running time of *TIE* and *HIE* increase with increasing number of registered users. Specifically, for *TIE*, the running time of calculating all influence for registered users is $O(n^2q + q^2n)$ since for each register user i and each social neighbor v, a loop of $n + q$ users have to be searched for common users. For *HIE*, the time complexity is dominated by calculating $\Theta_v(a_j)$, which is bounded by $O(nqm)$. Since m is much smaller compared to n and q, the running time of *HIE* is the smallest. The running time of *GIE* is $O(|V|^3)$, which is given in Lemma 6.

Overall, *HIE* based mechanisms usually leads more social cost and payment than those of *TIE* and *GIE* based mechanisms. However, *HIE* based mechanisms show significant superiority in terms of completion rate. *GIE* based mechanisms always output the least social cost and overpayment ratio, but the completion rate is much lower than the other two models. Moreover, the running time of *GIE* is quite high, thus *GIE* is only suitable for small community. In most cases, the completion rate of *Fast-Selector* is higher than our mechanisms, However, given the number of registered users, the social cost of our *HIE*, *TIE* and *GIE* based mechanisms is only 40.9%, 23.3%, 10.4% of this untruthful comparison mechanism on average, respectively.

IX. Conclusion

In this paper, we have designed the incentive mechanisms for large-scale crowdsourcing task diffusion through the social network. We have presented two task diffusion models, and formulated the *SOTD* problem for each of them. We have presented two influence estimation methods, *TIE* and *HIE*, based on the limited knowledge of social network. Furthermore, we have designed two incentive mechanisms: *MTD-L* and *MTD-IC* to solve the *SOTD* problem for the two task diffusion models, respectively. Further, we have proposed *GIE* to estimate the influence of registered users to the global social users. Through both rigorous theoretical analyses and extensive simulations, we have demonstrated that the proposed incentive mechanisms achieve computational efficiency, individual rationality, truthfulness, and guaranteed approximation. Our simulations also show that *GIE* based mechanisms always output the least social cost and overpayment ratio, and *HIE* based mechanisms show significant superiority in terms of completion rate.

References

[1] http://www.wikipedia.org

[2] K. Tahboub, N. Gadgil, J. Ribera, B. Delgado, and E. J. Delp, "An intelligent crowdsourcing system for forensic analysis of surveillance video," in *Proc. of SPIE The International Society for Optics and Photonics*, vol. 9407, p. 94070I, 2015.

[3] J. Fan, M. Lu, B. C. Ooi, W. Tan, and M. Zhang, "A hybrid machine-crowdsourcing system for matching web tables," in *Proc. of IEEE International Conference on Data Engineering*, 2014, pp. 976-987.

[4] R. Toris, D. Kent, and S. Chernova, "The robot management System: a framework for conducting human-robot interaction studies through crowdsourcing," *Journal of Human-Robot Interaction*, vol. 3, no.2, pp. 25-49, 2014.

[5] D. Ghadiyaram and A. C. Bovik, "Crowdsourced study of subjective image quality," in *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, 2014, pp. 84-88.

[6] https://www.mturk.com/mturk/welcome

[7] A. Yassine, M. S. Hossain, G. Muhammad, and M. Guizani, "Cloudlet-based intelligent auctioning agents for truthful autonomous electric vehicles energy crowdsourcing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5457-5466, 2020.

[8] C. Liu, Y. Zhao, Z. Dai, Y. Yuan, and K.K. Leung, "Curiosity-driven energy-efficient worker scheduling in vehicular crowdsourcing: a deep reinforcement learning approach," in *Proc. of IEEE 36th International Conference on Data Engineering*, 2020, pp. 25-36.

[9] C. Zhu, G. Pastor, Y. Xiao, and A. Ylajaaski, “Vehicular fog computing for video crowdsourcing: Applications, feasibility, and challenges,” *IEEE Communications Magazine*, vol. 56, no. 10, pp. 58-63, 2018.

[10] J. Ren, D. Zhang, S. He, Y. Zhang and T. Li, “A Survey on End-Edge-Cloud Orchestrated Network Computing Paradigms: Transparent Computing, Mobile Edge Computing, Fog Computing, and Cloudlet,” *ACM Computing Surveys*, vol. 52, no. 6, pp. 1-36, 2019.

[11] J. Ren, H. Guo, C. Xu and Y. Zhang, “Serving at the Edge: A Scalable IoT Architecture based on Transparent Computing,” *IEEE Network*, vol. 31, no. 5, pp. 96-105, 2017.

[12] L. Li, T. Quek, J. Ren, H. Yang, Z. Chen and Y. Zhang, “An Incentive-Aware Job Offloading Control Framework for Multi-Access Edge Computing,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, pp. 63-75, 2021.

[13] 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.

[14] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, “Truthful incentive mechanisms for crowdsourcing,” in *Proc. of IEEE INFOCOM*, 2015, pp. 2830-2838.

[15] L. Jiang, X. Niu, J. Xu, D. Yang, and L. Xu, “Incentivizing the workers for truth discovery in crowdsourcing with copiers,” in *Proc. of IEEE ICDCS*, 2019, pp. 1286-1295.

[16] Y Xiao, P. Simoens, P. Pillai, K. Ha, and M. Satyanarayanan, “Lowering the barriers to large-scale mobile crowdsensing,” in *Proc. of the 14th Workshop on Mobile Computing Systems and Applications*, 2013, pp. 1-6.

[17] https://en.wikipedia.org/wiki/Keep_a_Child_Alive

[18] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in *Proc. of ACM SIGKDD*, 2003, pp. 137-146.

[19] Y. Zhu, D. Li, and Z. Zhang, “Minimum cost seed set for competitive social influence,” in *Proc. of IEEE INFOCOM*, 2016, pp. 1-9.

[20] C. Long and R. C. Wong, “Minimizing seed set for viral marketing,” in *Proc. of IEEE International Conference on Data Mining*, 2011, pp. 427-436.

[21] https://www.stepes.com

[22] http://crowdsource.google.com/imagelabeler

[23] https://translate.google.com/community

[24] http://task.qq.com

[25] J. Xu, Z. Rao, L. Xu, D. Yang, and T. Li, “Incentive mechanism for multiple cooperative tasks with compatible users in mobile crowd sensing via online communities,” in *IEEE Transactions on Mobile Computing*, 2020, pp. 1618-1633.

[26] J. Park and B. Suh, “The impact of influential’s betweenness centrality on the wom effect under the online social networking service environment,” in *Pacific Asia Conference on Information Systems*, 2013, pp. 60-75.

[27] M. Li, Q. Zhang, and Y. Deng, “Evidential identification of influential nodes in network of networks,” *Chaos, Solitons & Fractals*, vol. 117, pp. 283-296, 2018.

[28] I. Gaye, G. Mendy, S. Ouya, I. Diop, and D. Seck, “Multi-diffusion degree centrality measure to maximize the influence spread in the multilayer social networks,” in *Proc. of International Conference on e-Infrastructure and e-Services for Developing Countries*, 2016, pp. 53-56.

[29] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, “Social-network-assisted worker recruitment in mobile crowd sensing,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1661-1673, 2018.

[30] Z. Wang, Y. Huang, X. Wang, J. Ren, Q. Wang, and L. Wu, “Social recruiter: dynamic incentive mechanism for mobile crowdsourcing worker recruitment with social networks,” *IEEE Transactions on Mobile Computing*, DOI: 10.1109/TMC.2020.2973958, 2020.

[31] Y. Singer, “Budget feasible mechanisms,” in *Proc. of IEEE Annual Symposium on Foundations of Computer Science*, 2010, pp. 765-774.

[32] A. Singla and A. Krause, “Truthful incentives in crowdsourcing tasks using regret minimization mechanisms,” in *Proc. of international conference on World Wide Web*, 2013, pp. 1167-1178.

[33] J. Xu, H. Li, Y. Li, D. Yang, and T. Li, “Incentivizing the biased requesters: Truthful task assignment mechanisms in crowdsourcing,” in *Proc. of IEEE SECON*, 2017, pp. 1-9.

[34] Y. Singer and M. Mittal, “Pricing mechanisms for crowdsourcing markets,” in *Proc. of international conference on World Wide Web*, 2013, pp. 1157-1166.

[35] D. Yang, G. Xue, X. Fang and J. Tang, “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing,” in *Proc. of ACM Mobicom*, 2012, pp. 173-184.

[36] I. Koutsopoulos, “Optimal incentive-driven design of participatory sensing systems,” in *Proc. of IEEE INFOCOM*, 2013, pp. 1402-1410.

[37] D. Zhao, X. Li, and H. Ma, “How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint,” in *Proc. of IEEE INFOCOM*, 2014, pp. 1213-1221.

[38] Y. Zhou, Y. Zhang, and S. Zhong, “Incentive mechanism design in mobile crowd sensing systems with budget restriction and capacity limit,” in *Proc. of IEEE ICCCN*, 2017, pp. 1-9.

[39] Z. Wang, X. Pang, J. Hu, W. Liu, Q. Wan, Y. Li, and H. Chen, “When mobile crowdsensing meets privacy,” *IEEE Communications Magazine*, vol. 57, no. 9, pp. 72-78, 2019.

[40] D. Wu, J. Yan, H. Wang, D. Wu, and R. Wang, “Social attribute aware incentive mechanism for device-to-device video distribution,” *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1908-1920, 2017.

[41] J. He, H. Wang, X. Chu, and T. Zhang, “Incentive mechanism and content provider selection for device-to-device-based content sharing,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2946-2957, 2019.

[42] M. Sirivianos, J. Park, H. J, X. Yang, and S. Jarecki, “Dandelion: Cooperative content distribution with robust incentives,” in *USENIX Annual Technical Conference*, 2007, pp. 157-170.

[43] D.B Johnson, “Efficient algorithms for shortest paths in sparse networks,” *Journal of the ACM (JACM)*, vol. 24, no. 1, pp. 1-13, 1977.

[44] U. Brandes, “A faster algorithm for betweenness centrality,” *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163-177, 2017.

[45] J. Qiu, J. Tang, H. Ma, Y. Dong, and K. Wang, “Deepinf: Social influence prediction with deep learning,” in *Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2110-2119.

[46] J. Nie, Z. Xiong, D. Niyato, P. Wang, and J. Luo, “A socially-aware incentive mechanism for mobile crowdsensing service market,” in *Proc. of IEEE Global Communications Conference*, 2018, pp. 1-7.

[47] G. Yang, S. He, Z. Shi, and J. Chen, “Promoting cooperation by the social incentive mechanism in mobile crowdsensing,” *IEEE Communications Magazine*, vol. 55, no. 3, pp. 86-92, 2017.

[48] J. Xu, C. Guan, H. Wu, D. Yang, L. Xu, and T. Li, “Online incentive mechanism for mobile crowdsourcing based on two-tiered social crowdsourcing architecture,” in *Proc. of IEEE SECON*, 2018, pp. 1-9.

[49] Y. Han, T. Luo, D. Li, and H. Wu, “Competition-based participant recruitment for delay-sensitive crowdsourcing applications in D2D networks,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, pp. 2987-2999, 2016.

[50] Y. Chen, H. Chen, S. Yang, X. Gao, Y. Guo, and F. Wu, “Designing incentive mechanisms for mobile crowdsensing with intermediaries,” *Wireless Communications and Mobile Computing*, vol. 2019, pp. 1-14, 2019.

[51] M. Granovetter, “Threshold models of collective behavior,” *American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.

[52] T. M. Liggett, “Interacting particle systems,” *Springer Science & Business Media*, 2012.

[53] J. Xu, J. Xiang, and D. Yang, “Incentive mechanisms for time window dependent tasks in mobile crowdsensing,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6353-6364, 2015.

[54] M. S. Granovetter, “The strength of weak ties,” *American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380, 1973.

[55] U. Feige, “A threshold of ln n for approximating set cover,” *Journal of the ACM (JACM)*, vol. 45, no. 4, pp. 634-652, 1998.

[56] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, “Algorithmic Game Theory,” *Cambridge University Press*, 2007.

[57] R. Myerson, “Optimal auction design,” *Mathematics of operations research*, vol. 6, no. 1, pp. 58-73, 1981.

[58] S. Rajagopalan and V. V. Vazirani, “Primal-dual RNC approximation algorithms for (multi)-set (multi)-cover and covering integer programs,” in *Proc. of IEEE FOCS*, 1993, pp. 322-331.

[59] D. Miorandi and F. D. Pellegrini, “K-shell decomposition for dynamic complex networks,” in *Proc. of IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2010, pp. 488-496.

[60] V. Batagelj and M. Zaversnik, “An O(m) algorithm for cores decomposition of networks,” *arXiv preprint cs/0310049*, 2003.

[61] M. D. Domenico, A. Lima, P. Mougel and M. Musolesi, “The anatomy of a scientific rumor,” *Scientific Reports 3*: 2980, 2013.

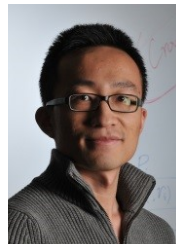
[62] http://www.modelingonlineauctions.com/datasets

Jia Xu (*M'15*) received the M.S. degree in School of Information and Engineering from Yangzhou University, Jiangsu, China, in 2006 and the PhD. Degree in School of Computer Science and Engineering from Nanjing University of Science and Technology, Jiangsu, China, in 2010. He is currently a professor in the School of Computer Science at Nanjing University of Posts and Telecommunications. He was a visiting Scholar in the Department of Electrical Engineering & Computer Science at Colorado School of Mines from Nov. 2014 to May. 2015. His main research interests include crowdsourcing, edge computing and wireless sensor networks. Prof. Xu has served as the PC Co-Chair of SciSec 2019, Organizing Chair of ISKE 2017, TPC member of Globecom, ICC, MASS, ICNC, EDGE. He currently serves as the Publicity Co-Chair of SciSec 2021.

Gongyu Chen received the bachelor's degree from the School of Mathematics and Statistics at Yancheng Normal University, Yancheng, China in 2017. She is currently studying for the M.S. degree in Nanjing University of Posts and Telecommunications in Nanjing, China. Her research interests are mainly in the areas of the crowdsouring, game theory and social network.

Yuanhang Zhou is now pursing the bachelor degree in Bell Honors School, Nanjing University of Posts and Telecommunications, Jiangsu, China. His research interests are mainly in the areas of the mobile crowdsensing, machine learning, and game theory.

Zhengqiang Rao received the bachelor's degree in School of communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2015. He is currently working toward the master's degree in the same university. His research interests are mainly in the areas of the mobile crowd sensing, wireless sensor networks, and machine learning.



Dejun Yang (*SM'19*) received the B.S. degree in computer science from Peking University, Beijing, China, in 2007 and the Ph.D. degree in computer science from Arizona State University, Tempe, AZ, USA, in 2013. Currently, he is an Associate Professor of computer science with Colorado School of Mines, Golden, CO, USA. His research interests include Internet of things, networking, and mobile sensing and computing with a focus on the application of game theory, optimization, algorithm design, and machine learning to resource allocation, security and privacy problems. Prof. Yang has served as the TPC Vice-Chair for Information Systems for IEEE International Conference on Computer Communications (INFOCOM) and currently serves an Associate Editor for the IEEE Internet of Things Journal (IoT-J). He has received the IEEE Communications Society William R. Bennett Prize in 2019 (best paper award for IEEE/ACM Transactions on Networking (TON) and IEEE Transactions on Network and Service Management in the previous three years), Best Paper Awards at IEEE Global Communications Conference (GLOBECOM) (2015), IEEE International Conference on Mobile Ad hoc and Sensor Systems (2011), and IEEE International Conference on Communications (ICC) (2011 and 2012), as well as a Best Paper Award Runner-up at IEEE International Conference on Network Protocols (ICNP) (2010).

Cuihua Xie received the M.S. degree in School of Information and Engineering from Yangzhou University, Jiangsu, China, in 2006 and the PhD. Degree in School of Economics and Management from Southeast University, Jiangsu, China, in 2011. He is a lecturer in e-commerce at Wenzhou University Oujiang College, China. His research interests include technological innovation in co-operatives, co-operatives in agricultural supply chain and e-commerce. Dr Xie has been leading a number of projects foundered by the Zhejiang Philosophy and Social Sciences Foundation, the Ministry of Education Industry-University Cooperation Collaborative Education Project, and the Basic Soft Science Research Projects of Wenzhou Science and Technology Bureau etc.