Non-Cooperative Information Diffusion in Online Social Networks under the Independent Cascade Model

Yile Yang, Zhiyi Lu, Victor O.K. Li, *IEEE Fellow*, and Kuang Xu
The University of Hong Kong

***Abstract*—In this paper, we present the first detailed analysis of influence maximization in *noncooperative* social networks under the Independent Cascade Model. We propose a new influence model based on the Independent Cascade Model and prove the approximation guarantees for influence maximization in noncooperative settings. We structure the influence diffusion into two stages, namely, seed node selection and influence diffusion. In the former, we introduce a modified hierarchy-based seed node selection strategy which can take node noncooperation into consideration. In the latter, we propose a game-theoretic model to characterize the behavior of noncooperative nodes and design a Vickrey-Clarke-Groves-like scheme to incentivise cooperation. Then we study the budget allocation problem between the two stages, and show that a marketer can utilize the two proposed strategies to tackle noncooperation intelligently. We evaluate our proposed schemes on large coauthorship networks, and the results show that our seed node selection scheme is very robust to noncooperation and the Vickrey-Clarke-Groves-like scheme can effectively stimulate a node to become cooperative.**

***Keywords*—influence maximization; cooperative; social network.**

I. Introduction

In order to maximize revenues, it is critical for online social viral marketers to identify pilot users who are "influential" to seed product adoption cascade. The influence maximization problem [1] is about selecting optimal initial seed nodes. However, intermediate (i.e., non-pilot) users cannot be ignored. It is shown in advertising research literature [2] that pilot users are more cooperative to recommend the product to their social neighbors because they feel obligated after accepting discounts, or even free samples from advertisers. However, non-pilot users may not be willing to pass on the influence, for the recommendation action may cost time, credibility, privacy, etc. We believe that it is important to study influence maximization problem in the context of noncooperative social networks.

It is natural to provide incentives to nodes in the social network for them to be cooperative, and incentive mechanism design is a hot topic in networking research. For example, in Peer-to-Peer (P2P) applications, [3] proposes to design a taxation scheme to incentivise user participation in P2P streaming. For file swarming applications, [4] introduces an auction-based model to improve incentives in BitTorrent. [5] also designs an incentive scheme named Networked Asynchronous Bilateral Trading (NABT) to deter free-riding in P2P applications. The Vickrey-Clarke-Groves (VCG) auction scheme is also a powerful tool for designing incentive schemes in networks such as mobile ad hoc networks and wireless networks made up of selfish nodes [6] [7].

[8] claims that there are two ways to enhance the efficiency of targeting influential users in large-scale online social networks (OSNs). One is to further optimize the greedy algorithm in [1], as in [9]. The other is to propose new heuristics to solve the problem. [8] tries both methods respectively, and suggests that the second way is more promising based on experimental results. However, the degree discount heuristic proposed in [8] is only applicable to the Independent Cascade Model (ICM) in which the propagation probability is very small. There are several follow-on papers focusing on the second way to enhance the efficiency as claimed in [8], like [10] [11], and [12]. We find the hierarchy-based algorithm [12] the only algorithm that can be readily adapted to consider node noncooperation.

Prior efforts on solving influence maximization problems are mainly focused on modeling the influence diffusion process [13] [14] and solving the corresponding optimization problems [15] [16]. [1] [17] first show that this optimization problem is NP-hard, and provide a greedy heuristic which can achieve near-optimal performance guarantee. However, existing influence maximization algorithms and newly proposed influence diffusion models (e.g., [10] [11] [14]) do not distinguish between seed nodes (or pilot users) and nonseed nodes and assume all nodes are cooperative in propagating influence.

The efforts that are similar to ours are [18] [19], which model social networks that allow negative opinions to emerge and propagate. [18] extends ICM to incorporate a parameter, namely, quality factor q such that a node in the network has probability q to be negatively activated and to propagate negative opinion. [19] uses a novel heat diffusion process to model social network marketing campaigns in which negative comments can spread.

To the best of our knowledge, the only paper that considers the budget in influence maximization problem is [20]. [20] generalizes the influence maximization problem with bud-

Y. Yang, Z. Lu, V.O.K. Li and K. Xu are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. Email: yyl.henry@gmail.com, {zylu, vli, xukuang}@eee.hku.hk.
This research is supported in part by the Research Grants Council of Hong Kong, under Grant No. 17264116.
Manuscript received Feb 02, 2015; revised June 20, 2017.

get constraint, namely, the budgeted influence maximization (BIM) problem. The BIM problem is similar to the budget allocation problem (BAP) discussed in this paper. However, we consider node noncooperation and BIM does not.

For a comprehensive survey on information diffusion in OSNs, readers are referred to [21]. [22] is a good tutorial on the cascading behavior in OSNs.

In our previous work [23], we design a VCG-like scheme to incentivise cooperation in noncooperative social networks. In this paper, we study the problem of influence maximization in noncooperative social networks more thoroughly by structuring the influence diffusion process into two stages. The two stages are seed node selection and influence diffusion. For the first stage, we design a hierarchy-based seed selection algorithm which considers node noncooperation. The VCG-like incentive scheme proposed in [23] is utilized in the second stage. A viral marketer under the two-stage framework can either choose one of the two strategies, or combine them together to achieve a satisfactory solution for the noncooperation problem. Simulation results on large social network dataset show that the proposed strategy can effectively encourage cooperation among the participating nodes. To the best of our knowledge, we are the first to investigate node noncooperation on influence diffusion in social networks under ICM. Other related work includes [24], but the paper is based on the Linear Threshold Model (LTM) while this work is based on the ICM model. Moreover, in this paper, we give a more detailed study than [24] on noncooperative influence maximization by discussing properly incentivising intermediate nodes. In addition, different from the flow-based seed selection algorithm designed for LTM, we develop a hierarchy-based seed selection scheme, which is more suitable for ICM.

The rest of this paper is structured as follows. In Section II, we briefly describe the problem formulation and diffusion models. Then we generalize the original ICM into a noncooperative one and prove some nice properties of the model. This section also proposes the two-stage view of the noncooperative influence maximization problem. We propose a hierarchy-based seed selection algorithm in noncooperative social networks in Section III. The VCG-like incentive scheme designed for the influence diffusion stage is discussed in Section IV. Then we discuss the BAP in Section V. The evaluation results are shown in Section VI. We conclude this study in Section VII.

II. System Model

We first formulate the influence maximization problem and introduce the noncooperative ICM. Following our prior work [24] which proved the submodularity of the noncooperative Linear Threshold Model (LTM), we further present several useful properties of the noncooperative ICM. Finally we describe our two-stage view of the noncooperative influence maximization problem.

A. Problem formulation

We consider an OSN as a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (OSN users) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (social ties) in the network. We also denote by $\mathcal{N}_u \subseteq \mathcal{V}$ the set of neighbors of node u. Each node in the system can either be active or inactive. As more neighbors of an inactive node become active, it is more likely to switch to being active. A node cannot return to the inactive state once it becomes active. All nodes are inactive at the beginning of the influence propagation process and marketing practitioners initially activate K nodes to seed the information cascade in the social network. The process ends when no more nodes can be activated. The influence maximization problem is defined as follows: Determine the K-node seed set to achieve the maximal expected active nodes at the end of the process.

B. Diffusion models

The Independent Cascade Model (ICM) [25] is a popular diffusion model. In ICM, a node i activated at time t has a probability $p_{i,j}$ to successfully activate its inactive neighbor j at time $t+1$. Node i does not have any further opportunities to activate j again whether it succeeds or not.

C. Noncooperative influence maximization under ICM

Traditional ICM implicitly assumes that nodes in the system will not reserve their influence capacities during the propagation process. To account for non-cooperativeness the standard ICM is generalized such that Node j is activated by Node i with probability $\alpha_{i,j} \cdot p_{i,j}$, where $\alpha_{i,j} \in [0,1]$ is the cooperativeness level of Node i on its neighbor Node j. We assume that node cooperativeness levels are static during the entire diffusion process.

D. Properties of the noncooperative ICM

We now discuss some nice properties of the noncooperative ICM. First we define a set function $\sigma(\cdot)$ to be submodular if $\sigma(S \cup \{v\}) - \sigma(S) \geq \sigma(T \cup \{v\}) - \sigma(T)$ for all $v \in \mathcal{V} \setminus T$ and $S \subseteq T$, i.e., $\sigma(\cdot)$ satisfies a "diminishing returns" requirement: the marginal gain from adding a node to a set T is at most the same as the marginal gain from adding the same node to a subset of T. In addition, we say that $\sigma(\cdot)$ is monotone if $\sigma(T) \geq \sigma(S)$ for all $S \subseteq T$, that is, $\sigma(\cdot)$ will at least stay the same after adding elements to the original set. We also define a greedy algorithm as follows: starting from an empty set, the algorithm iteratively selects a seed which achieves the highest incremental change of $\sigma(\cdot)$. The result of [26] shows that the optimum of a non-negative, monotone submodular objective function can be approximated to within a factor of $(1 - 1/e)$ (around 63%, here e is the base of the natural logarithm) using the greedy algorithm. [1] further proves that the final influence function $\sigma(\cdot)$, which is the expected number of active nodes in the network at the end of the diffusion process, is submodular. Thus the greedy algorithm can also achieve $(1-1/e)$ approximation for the influence maximization problem. Based on [1], we prove that the influence function under the proposed noncooperative ICM also satisfies the requirement of submodularity, so that a greedy algorithm can also achieve the same $(1 - 1/e)$ performance guarantee.

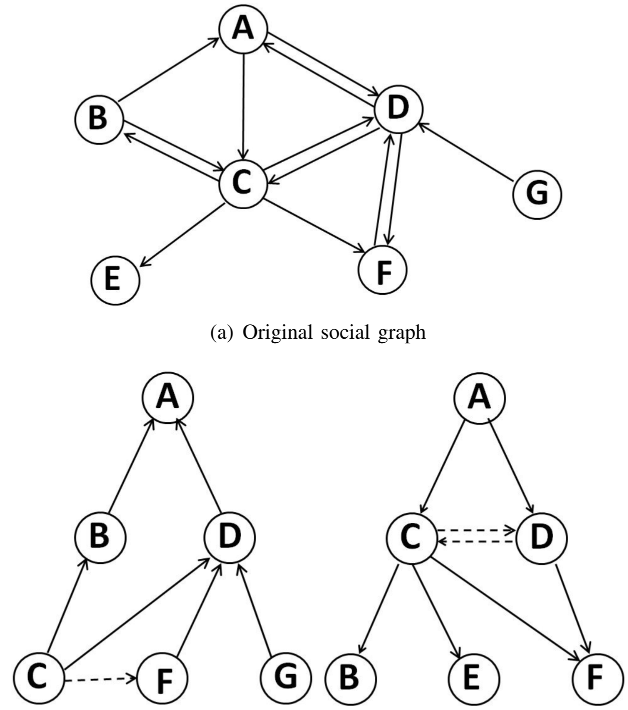
Lemma 1. *[1] The influence function $\sigma(\cdot)$ is submodular for an arbitrary instance of the ICM.*

Theorem 1. *The influence function $\sigma(\cdot)$ of noncooperative ICM is submodular.*

Proof. Since the cooperativeness parameters $\alpha_{i,j}$ are static, the noncooperative ICM is equivalent to a standard ICM in which $p'_{i,j} = \alpha_{i,j} \cdot p_{i,j}$. Thus, according to Lemma 1, the influence function of noncooperative ICM is also submodular. □

Proving that the influence function under noncooperative ICM also satisfies the requirement of submodularity not only shows that the model has a performance guarantee, but also implies that the incentive needed for the advertising campaign should show similar property, since the amount of incentive needed is closely related to the seed-node set size. It is also intuitively satisfying that incentive as a function of seed-node set size would show a "diminishing returns" property. The detailed study of the relationship between the amount of incentive and seed-node set size in noncooperative influence maximization problem will be our future work.

Since there is no explicit formula for the influence function, we have to simulate the influence diffusion process for R rounds. The average number of final active nodes is then the estimated value of $\sigma(\cdot)$. With a large enough R, this estimate may be made arbitrarily close to the real value of $\sigma(\cdot)$.



(a) Original social graph

(b) Input hierarchical network (c) Output hierarchical network

Fig. 1: Up-to-2-hop hierarchical network for Node A [12]

E. A two-stage solution to solve node noncooperation

To solve the node noncooperation problem, we examine the influence diffusion process from a two-stage perspective as follows.

1) Seed node selection stage: This stage corresponds to when a viral marketer selects seed nodes in order to start an influence cascade. Current influence maximization literature mostly focuses on this stage. We claim that the initiator can take the noncooperativeness of non-seed (intermediate) nodes into consideration when choosing initial users to activate. We propose a modified hierarchy-based seed node selection method which can take node noncooperation into account. The details can be found in Section III.

2) Influence diffusion stage: After the viral marketer has chosen and activated the seed node set according to various selection methods, the second stage starts. The influence will propagate through the social network at this stage. The final result of the marketing campaign (i.e. the size of the final activated nodes) can be improved if intermediate nodes are cooperative to forward the influence. We introduce a VCG-like incentive scheme to stimulate cooperation of non-seed nodes. A detailed description and discussion are provided in Section IV.

III. Stage I: Seed Node Selection

As indicated in [27] [28], human emotions such as happiness and loneliness can spread no more than three hops in a large social network. These findings inspire the authors in [12] to design a heuristic which approximates the influential power of a single node in the social network as the influence in its up-to-2-hop social circle. Extensive simulations show that the proposed heuristic performs well both in terms of accuracy and efficiency on calculating social influence.

The special property of only considering the up-to-2-hop neighbors in the heuristic proposed in [12] also makes it suitable for a modification to consider node noncooperation. As we have discussed in Section I, normally seed nodes are more cooperative in forwarding influence received compared to their one-hop, non-seed neighbors during influence diffusion. Thus it is straightforward to adapt the original algorithm to a noncooperative version.

In this section we first give a brief description of the algorithm introduced in [12], then we describe the modifications necessary to account for node noncooperation.

A. Up-to-2-hop hierarchical network

Before describing the original hierarchy-based algorithm in [12], we first introduce the concepts of up-to-2-hop hierarchical network. To estimate the influential power of a single node, we have to build an up-to-2-hop hierarchical network of this node. There are two types of such networks, namely, input and output hierarchical networks. Both networks contain three levels of nodes, namely, (i) head node (i.e., the node being considered), (ii) 1-hop neighbors of the head node, and (iii) 2-hop neighbors of the head node. Both networks can be built using Breadth-First Search (BFS). Figure 1 is the example of how to formulate the up-to-2-hop input and output hierarchical networks. Figure 1 (a) is the original social graph. In Figure 1 (b) and (c), we treat Node A as the head node and generate the corresponding input and output hierarchical networks, respectively. The construction of both input and output hierarchical networks follows two principles as follows:

1. Intra-level link removal principle: We ignore links between nodes of the same level in the hierarchical network in order to simplify calculations. Take Figure 1 (b) as example, Nodes C and F are both 2-hop neighbors of the head node. Under the Intra-level link removal principle, we delete the edge from Node C to F (the dotted line in Figure 1 (b)) in the constructed network.

2. Hop-first principle: When a node is a 1-hop neighbor as well as 2-hop neighbor of the head node, we treat it as a 1-hop neighbor. In this way we can ensure the uniqueness of the constructed hierarchical network. For example, in Figure 1 (a), Node C is both 1-hop neighbor (via path $A \rightarrow C$) and 2-hop neighbor (via path $A \rightarrow D \rightarrow C$) of the head node. Under the hop-first principle, we treat Node C as the 1-hop neighbor in Figure 1 (c).

We use input hierarchical network to calculate the influence from all active, up-to-2-hop neighbors of the head node on the head node itself. In other words, we evaluate the probability a node is activated by the seed set by approximating it as the probability this node is activated by its active, up-to-2-hop neighbors in its input hierarchical network. Consider the head node i and the active seed set S in the hierarchical network, we define the up-to-2-hop input hierarchical influence as $InInf_i(S)$. The calculation on $InInf_i(S)$ is as follows:

$$1-(1-p)^{|NH_i^{in}\cap S|}\cdot \prod_{\substack{j\in NH_i^{in}\\ j\notin S}} [1-p(1-(1-p)^{|NH_j^{in}\cap S|})] \tag{1}$$

where NH_i^{in} is the set of nodes which are the in-degree neighbors of Node i in the input hierarchical network H. In (1), $(1-p)^{|NH_i^{in}\cap S|}$ represents the probability that none of the active 1-hop neighbors succeed in activating Node i. Similarly, $\prod_{j\in NH_i^{in}, j\notin S} [1-p(1-(1-p)^{|NH_j^{in}\cap S|})]$ represents the likelihood that Node i fails to be activated by any active 2-hop neighbors via non-blocked[1] 2-hop paths in its input hierarchical network.

The output hierarchical network is utilized to estimate the influential power of the head node in terms of the expected number of nodes the head node can activate in its output hierarchical network. We define the up-to-2-hop output hierarchical influence as $OutInf_i(S)$. The calculation on $OutInf_i(S)$ is given as follows:

$$OutInf_i = p\cdot d_i^{out} + \sum_{j\in MH_i^{out}} [1-(1-p^2)^{|NH_j^{in}|}] \tag{2}$$

where d_i^{out} is the out-degree of Node i in the output hierarchical network H, and MH_i^{out} is the 2-hop out-degree neighbor set of Node i in H. Since NH_j^{in} can be considered as the set of 2-hop paths from Node i to Node j, the term $1-(1-p^2)^{|NH_j^{in}|}$ corresponds to the probability that Node j is eventually activated.

For the detailed descriptions about the construction method and properties of the hierarchical networks, and explanations about the definitions of $InInf_i(S)$ and $OutInf_i(S)$, readers are referred to [12].

B. Hierarchy-based algorithm for the influence maximization problem

After introducing hierarchical networks, [12] further borrows two important concepts, namely, intensification and diversification from meta-heuristic optimization theory to design the hierarchy-based influence maximization algorithm. In the hierarchy-based influence maximization algorithm, intensification means that the viral marketer should choose seed nodes which have common neighbors so that they can work collectively to increase the probability of activation. On the contrary, diversification indicates that the viral marketer should spread out the seed nodes in different neighborhoods in order to avoid influence overlapping, i.e., the influential ranges of seed nodes overlap, which is kind of wasteful. The authors further propose the marginal influence increment (MII) to strike a balance between these two important factors.

MII calculates the marginal increment of activations a single node can achieve given the existing seed set. Suppose S is the existing seed set and Node i is the candidate, then MII is defined as follows:

$$MII_i(S) = (1-p_i(S))(1+INT_i(S)+DIV_i(S)) \tag{3}$$

In (3), $1-p_i(S)$ stands for the likelihood that Node i fails to be activated by S. $p_i(S)$ can be approximated by $InInf_i(S)$ as mentioned in Section III-A. $INT_i(S)$ is the intensification part. The physical meaning of $INT_i(S)$ is the number of nodes which are common out-degree neighbors that Node i and S shares. $DIV_i(S)$ represents the diversification part, which means the number of nodes which are only out-degree neighbors of Node i, but not out-degree neighbors of any node in S. Under a greedy selection strategy, we want to choose a node with the largest MII. The equations for $INT_i(S)$ is:

$$\sum_{\substack{j\notin S, j\in N_i^{out}\\ j\in N_r^{out}, \exists r\in S}} [(1-p)^{|N_j^{in}\cap S|}-(1-p)^{|N_j^{in}\cap S|+1}]\cdot[1+ p\cdot(d_j^{out}-|N_j^{out}\cap\{S\cup\{i\}\}|)] \tag{4}$$

and $DIV_i(S)$ is as follows:

$$\sum_{\substack{j\notin S, j\in N_i^{out}\\ j\notin N_r^{out}, \forall r\in S}} p\cdot[1+p\cdot(d_j^{out}-|N_j^{out}\cap\{S\cup\{i\}\}|)] \tag{5}$$

where N_j^{in} is the set of in-degree neighbors of Node j in the social graph, and d_j^{out} is the out-degree of Node j in the social graph.

The hierarchy-based algorithm roughly works as follows: we first sort all nodes according to their $OutInf_i(S)$ values by constructing the up-to-2-hop output hierarchical network.

[1]In [12], a 2-hop path $r \rightarrow j \rightarrow i$ (Node i is the head node, Nodes r and j are 2-hop and 1-hop neighbors of the head node, respectively) is defined as blocked if both Nodes r and j belong to the active seed set S. The concept of being blocked is useful when we generalize the hierarchy-based influence maximization algorithm to consider node noncooperation in Section III-C.

Then we choose the node with the largest $OutInf_i(S)$ value as the first node in the seed node set. After that, for each node not in S, we calculate their $MII_i(S)$ values, and add the node with the largest value into S. We repeat until K nodes are selected for influence diffusion.

Finally, a formal statement of the hierarchy-based algorithm is given in Algorithm 1.

Algorithm 1 Hierarchy-based heuristic

```
Let 1, ..., N be nodes.
Input:
    Network G(V, E) with activation probability p.
    A given integer K;
Output:
    The final target set S_t;
S_t ← ∅
Calculate OutInf_i for each node i
Select u = argmax_i{OutInf_i}
S_t ← S_t ⋃ {u}
while |S_t| < K do
    for i = 1 to N do
        if i ∉ S_t then
            Calculate MII_i(S_t) for node i
            Select v = argmax_i{MII_i(S_t) | i ∉ S_t}
            S_t ← S_t ∪ {v}
        end if
    end for
end while
```

For a detailed description about the hierarchy-based seed-selection algorithm with examples, readers are referred to [12].

C. Hierarchy-based algorithm for the noncooperative influence maximization problem

For the noncooperative version of the hierarchy-based seed-selection algorithm, we just need to modify the equations of the metrics used in the algorithm. The modified equations are described in the following sections.

1) Input hierarchical network under the noncooperative case: To calculate the up-to-2-hop input hierarchical influence $InInf_i(S)$ considering node noncooperation, if the 1-hop neighbor of Node i is an active seed node, then it is cooperative to activate Node i (i.e., $\alpha_{j,i} = 1$, j neighbor of Node i, if Node j is a seed node). However, if a 2-hop neighbor of the head node is an active seed node and tries to influence the head node through a non-blocked path, even if the 1-hop neighbor on that path is successfully activated, it may not be cooperative to forward the influence it receives to the head node since it is not a seed node (i.e., $\alpha_{j,i} < 1$, j neighbor of Node i, if Node j is an intermediate node on a 2-hop path $r \to j \to i$, and Node r is a seed node).[2] Thus the equation for $InInf_i(S)$ in a noncooperative social network is as follows:

$$1-(1-p)^{|NH_i^{in} \bigcap S|} \cdot \prod_{\substack{j \in NH_i^{in} \\ j \notin S}} [1-\alpha_{j,i} \cdot p \cdot (1-(1-p)^{|NH_j^{in} \bigcap S|})] \tag{6}$$

2) Output hierarchical network under the noncooperative case: As discussed in Section III-A, the up-to-2-hop output hierarchical influence $OutInf_i$ is the expected number of nodes activated by Node i in the output hierarchical network H, including 1-hop and 2-hop neighbors. In a noncooperative social network, if the head node is a seed node, then it is cooperative to activate its 1-hop neighbors. However, the 1-hop neighbors activated by the head node may not be cooperative to forward the influence it receives to the 2-hop neighbors of the head node due to lack of incentives. Thus the equation for $OutInf_i(S)$ in a noncooperative social network becomes

$$OutInf_i(S) = p \cdot d_i^{out} + \sum_{j \in MH_i^{out}} [1 - \prod_{k \in NH_j^{in}} (1 - \alpha_{k,j} \cdot p^2)] \tag{7}$$

3) Marginal influence increment under the noncooperative case: The modifications regarding the calculation of the intensification $INT_i(S)$ and the diversification $DIV_i(S)$ in order to get the marginal influence increment $MII_i(S)$ in noncooperative social networks are similar to Section III-C2. The 1-hop neighbors activated either collectively by the active seed set S (i.e., the nodes activated because of intensification) or only by the head node (i.e., the nodes activated because of diversification) may not be cooperative to forward the influence received to the 2-hop neighbors of the head node due to lack of incentives. Thus the calculations are as follows:

$$INT_i(S) = \sum_{\substack{j \notin S,\, j \in N_i^{out} \\ j \in N_r^{out},\, \exists r \in S}} [(1-p)^{|N_j^{in} \bigcap S|} - (1-p)^{|N_j^{in} \bigcap S|+1}] \cdot (1 + \sum_{\substack{r \in N_j^{out} \\ r \notin S \bigcup \{i\}}} \alpha_{j,r} \cdot p) \tag{8}$$

$$DIV_i(S) = \sum_{\substack{j \notin S,\, j \in N_i^{out} \\ j \notin N_r^{out},\, \forall r \in S}} p \cdot (1 + p \cdot \sum_{\substack{r \in N_j^{out} \\ r \notin S \bigcup \{i\}}} \alpha_{j,r}) \tag{9}$$

The noncooperative version of hierarchy-based influence maximization algorithm works the same way as Algorithm 1. We just need to replace (1), (2), (4), (5) with (6), (7), (8), and (9).

IV. Stage II: Influence Diffusion

We introduce an incentive mechanism to solve the node noncooperation problem in this section. We first introduce a game-theoretic framework to model node noncooperation in influence propagation. Next, we describe the incentive method and derive some nice properties of the mechanism, namely, individual-rationality (IR) and incentive-compatibility (IC). Then we compare the proposed scheme to a fixed price incentive mechanism to show some of its other advantages. Finally, we discuss implementation details of the proposed mechanism.

[2] In noncooperative ICM, the cooperativeness levels are determined by individual nodes themselves. However, for simplicity, we adopt the two-tiered, static node cooperativeness when explaining algorithms in this paper. That is, we set $\alpha_{i,j} = 1$ if Node i belongs to the seed-node set and $\alpha_{i,j} < 1$, otherwise. The results can be generalized in a straightforward manner.

A. A VCG-like incentive mechanism to solve the noncooperation problem

We define $C(i)$ as the cost of an individual node i during the influence diffusion process. The utility of an individual node without payment should be

$$\begin{aligned} U_i &= -C(i) \\ &= -D \cdot \sum_{j\ neighbor\ of\ i} \alpha_{i,j} \cdot p_{i,j} \end{aligned} \tag{10}$$

In (10) we model $C(i)$ as the sum of the influence probabilities Node i imposes on all its neighbors mainly to reflect the fact that the more a single node can impact its friends, the more reward it will ask for from the initiator of the viral marketing campaign, because "influence" here is considered a scarce commodity. Also an influential node (e.g., a celebrity) in the social network may have already expended a large amount of resources (e.g., time, money, privacy, etc.) in order to cultivate its impact. $D \geq 0$ is the cost-of-influence parameter, which converts the amount of influence an individual node exerts into cost. It is assumed to be constant over the whole network and known to the operator in our model.

The action of Node i is denoted as $\alpha_i = (\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,|\mathcal{N}_i|})$, in which $0 \leq \alpha_{i,j} \leq 1$, $j = 1, 2, \ldots, |\mathcal{N}_i|$. We also assume that all nodes determine their actions (i.e., cooperativeness levels) at the beginning of the game simultaneously.

Theorem 2. *Without payment, the strategy $\hat{\alpha}_i$ which constitutes the Nash equilibrium should be $\hat{\alpha}_{i,j} = 0$, j neighbor of i.*

Proof. Suppose that Node i chooses an action α_i different from $\hat{\alpha}_i$, with $\alpha_i = (\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,|\mathcal{N}_i|})$, *s.t.* $\exists \alpha_{i,j} \neq 0$. From the utility function (10) we can see that Node i can obtain a better payoff by setting $\alpha_{i,j} = 0$. Thus α_i is a strictly dominated action and cannot be used in any Nash equilibrium. So the strategy which constitutes the Nash equilibrium should be in the form $\hat{\alpha}_{i,j} = 0$, j neighbor of i. □

Define VCG-like payment to Node i as

$$\begin{aligned} M_i &= \beta \cdot q \cdot (\sigma(A) - \sigma_{-i}(A)) + C(i) \\ &= \beta \cdot q \cdot (\sigma(A) - \sigma_{-i}(A)) + D \cdot \sum_{j\ neighbor\ of\ i} \alpha_{i,j} \cdot p_{i,j} \end{aligned} \tag{11}$$

where $\sigma(A) - \sigma_{-i}(A)$ is the difference of the expected final active node set size when Node i exists, and the expected size if Node i does not exist. $q \geq 0$ is the amount of reward the initiator is willing to pay for a successful activation. $\beta \geq 0$ is the premium control parameter, used by the viral marketer to ensure that the total payment is within the budget. The detailed usage of this parameter will be explained in Section V. In other words, (11) means that besides compensating the individual cost $C(i)$, the initiator will additionally pay Node i for its contribution during the influence diffusion stage.

Our proposed VCG-like incentive scheme is different from the standard scheme in the following two aspects. First, in traditional mechanism design theory, the goal of the VCG auction is to encourage each selfish agent in the game to disclose its private information ("types") to the auctioneer [29]. For example in [30], under the VCG payment scheme, each node may choose to report its true forwarding cost so that the least cost path can be found correctly. But the objective of our proposed incentive mechanism is to ensure that each selfish node is cooperative in the sense that they will exert all its influence capacity (i.e. $\alpha_{i,j} = 1$, $j = 1, 2, \ldots, |\mathcal{N}_i|$ for an arbitrary Node i). Second, in standard VCG, the cost the operator has to compensate is the reported value claimed by the selfish node, while in (11) the cost $C(i)$ is the actual cost Node i has incurred taking its cooperativeness level into consideration.

Although there are some differences, the proposed VCG-like payment (11) is similar to the standard VCG payment formula in structure in that they both consist of two parts: premium, and some kind of "cost" (reported value or true value). More importantly, the proposed VCG-like scheme shares some nice properties of the standard scheme, and such properties will be discussed next.

First we introduce Lemma 2 which is useful in showing that the VCG-like incentive scheme will encourage nodes in the network to be cooperative. In order to prove the lemma, an equivalent view of the ICM proposed in [1] needs to be described first.

The probability $p_{i,j}$ in ICM represents the likelihood Node i will activate Node j when Node i becomes active while at the same time Node j is inactive. The outcome of this random event can be viewed as the flipping of a coin of bias $p_{i,j}$. In fact we can flip the coin corresponding to each of the edges in the social network G at the beginning of the cascading process and the result will only be revealed when Node i is active while its neighbor Node j is inactive. After all the coins have been flipped in advance, we declare edges in G for which the coin flip result in heads as live and the remaining edges as blocked. In this graph, it is clear that a node will be active at the end of the cascading process if it is on a path consisting of only live edges from the target set A. Further we can see that the number of nodes that are active at the end of the cascading process will be the number of the nodes that are on paths consisting of only live edges from the target set A. This equivalent view also shows that the final activated set size under ICM is an order-independent outcome, that is, if a node has several newly activated neighbors, the order of their activating attempts will not affect the final result. For a detailed discussion on the equivalent view, readers are referred to [1].

Lemma 2. *The expression $\sigma(A) - \sigma_{-i}(A)$ is always non-negative, i.e., $\sigma(A) - \sigma_{-i}(A) \geq 0$.*

Proof. Based on the order-independent equivalent view of the ICM process [1], we can divide the diffusion process of one sample point X in a sample space S into two steps. The first step is to simulate the diffusion process in the whole graph, but assuming all the incoming edges of Node i to be "blocked" and Node i itself to be inactive. The active set size at the end of the first step is thus $\sigma_{X,-i}(A)$. In the second step, we keep the original states of incoming edges (blocked or live) of i, and activate Node i if it is in the original seed set A. The result at the end of the second step is thus $\sigma_X(A)$. If Node i is activated first at the beginning of step two (i.e. $i \in A$), then it is obvious that $\sigma_X(A) > \sigma_{X,-i}(A)$. Consider the case in which $i \notin A$. If there is a path from some node in A to i consisting entirely of live edges, then Node i will be active and in turn may possibly initiate a cascading process (i.e. $\sigma_X(A) > \sigma_{X,-i}(A)$); otherwise, Node i will end up inactive and the diffusion process ends (i.e. $\sigma_X(A) = \sigma_{X,-i}(A)$). In general, $\sigma_X(A) \geq \sigma_{X,-i}(A)$. Since

$$\sigma(A) = \sum_{X \in S} P[X]\sigma_X(A) \tag{12}$$

and

$$\sigma_{-i}(A) = \sum_{X \in S} P[X]\sigma_{X,-i}(A) \tag{13}$$

So $\sigma(A) - \sigma_{-i}(A) \geq 0$. □

Lemma 3. *[31] Let $p' \in [0,1]^{|\mathcal{E}|}$ be the true influence probabilities on each edge. Given a target set A, then*

$$\sigma(A) = \sum_{i=1}^{n} u_i(p', A) + |A| \tag{14}$$

where $u_i(p', A)$ is the expected number of neighbors activated by Node i, given the target set.

Theorem 3. *The strategy $\hat{\alpha}_i$ which constitutes the Nash equilibrium should be $\hat{\alpha}_{i,j} = 1$, j neighbors of i, under the VCG-like payment scheme.*

Proof. Consider an arbitrary Node i, and fix the cooperativeness levels of the other nodes. If Node i is cooperative (i.e. $\alpha_{i,j} = 1$, j neighbors of i), with the VCG-like payment, the utility function of Node i now becomes

$$\begin{aligned} U_i &= M_i - C(i) \\ &= \beta \cdot q \cdot (\sigma(A) - \sigma_{-i}(A)) \\ &+ D \cdot \sum_{j\, neighbor\, of\, i} p_{i,j} - D \cdot \sum_{j\, neighbor\, of\, i} p_{i,j} \\ &= \beta \cdot q \cdot (\sigma(A) - \sigma_{-i}(A)) \end{aligned} \tag{15}$$

In Lemma 2 we have proved that $\sigma(A) - \sigma_{-i}(A) \geq 0$. Since $\beta \geq 0$ and $q \geq 0$, so $U_i \geq 0$.

Let $\alpha'_i = (\alpha'_{i,1}, \alpha'_{i,2}, \ldots, \alpha'_{i,|\mathcal{N}_i|})$, *s.t.* $\exists \alpha'_{i,j} < 1$ be the cooperativeness level of a noncooperative Node i, then the utility becomes

$$\begin{aligned} U'_i &= M'_i - C'(i) \\ &= \beta \cdot q \cdot (\sigma'(A) - \sigma'_{-i}(A)) + D \cdot \sum_{j\, neighbor\, of\, i} \alpha'_{i,j} \cdot p_{i,j} \\ &- D \cdot \sum_{j\, neighbor\, of\, i} \alpha'_{i,j} \cdot p_{i,j} \\ &= \beta \cdot q \cdot (\sigma'(A) - \sigma'_{-i}(A)) \\ &= \beta \cdot q \cdot (\sigma'(A) - \sigma_{-i}(A)) \text{(since the cooperativeness levels of other nodes are fixed)} \end{aligned} \tag{16}$$

Actually the true influence probability vector p' in Lemma 3 can be represented as (p'_i, p'_{-i}), where $p'_i = (p'_{i,1}, p'_{i,2}, \ldots, p'_{i,|\mathcal{N}_i|})$ is the true influence probability Node i has on its neighbors while p'_{-i} is the true influence probability vector on all other edges in the graph. According to Lemma 3, the final expected active node set size contains the expected number of neighbors activated by each node, given target set A. For each i, Node i can influence more neighbors when it is cooperative, i.e. $u_i((p_i, p'_{-i}), A) \geq u_i((p'_i, p'_{-i}), A)$. Thus $\sigma(A) \geq \sigma'(A)$ and $U_i \geq U'_i$. The cooperative strategy always maximizes the node utility. □

Theorem 3 implies that the VCG-like incentive scheme satisfies two important properties. The first property is individual-rationality (IR), that is, for each player, it is always better (i.e. achieving at least no less utility) to join the game than not participating. Combining Lemma 2 and Theorem 3 we can see that the individual utility of Node i is always nonnegative (0 is the utility when not participating in the game) under the proposed incentive scheme, so our scheme is IR. The other property is incentive-compatibility (IC) — each player prefers to act in accordance with the objective of the mechanism. Theorem 3 has proved that the dominant strategy for a single node is to be cooperative to exert all its influence capacity under the VCG-like scheme, which is exactly the design objective of the proposed scheme, so the scheme is also IC. IR and IC are also two nice properties of the standard VCG auction [32].

B. Advantages of the VCG-like scheme over the fixed price incentive scheme

Another possible, also intuitive incentive scheme is as follows:

$$\begin{aligned} M_i &= \varepsilon + C(i) \\ &= \varepsilon + D \cdot \sum_{j\, neighbor\, of\, i} \alpha_{i,j} \cdot p_{i,j} \end{aligned} \tag{17}$$

where ε can be any arbitrary positive number. Under this scheme, besides compensating for the individual cost $C(i)$, the operator will also pay a fixed amount of incentive ε.

It can be easily shown that under the fixed price incentive scheme, being cooperative (i.e. $\alpha_{i,j} = 1$, $j = 1, 2, \ldots, |\mathcal{N}_i|$ for an arbitrary Node i) is the weakly dominant strategy for a selfish node. In other words, the utility of an individual node is the same (i.e. ε) whether it is cooperative or not. However, Theorem 3 has already shown that under the VCG-like scheme, being cooperative is the strongly dominant strategy. That is to say, the individual utility is maximized if a selfish node chooses to exert all its influence capacity. From this aspect the VCG-like scheme is superior to the fixed price scheme.

Another drawback of the fixed price scheme is that every node can get the same premium ε regardless of its ability to impact others. That means the fixed price scheme is not "fair" in the sense that the specific contribution of an individual node during the influence diffusion process is ignored. Some "influential" nodes in the social network may thus find this property discouraging. In contrast, in the VCG-like scheme (11), $\sigma(A) - \sigma_{-i}(A)$, which is the difference of the final performance when Node i does not exist, exactly quantifies the contribution of Node i during the diffusion process. To conclude, being more "fair" is another advantage of the proposed VCG-like incentive scheme.

C. Some technical discussions on the VCG-like incentive scheme

1) How do we get the cooperativeness levels of nodes?

Suppose the social network marketer has already determined the influence probability on each edge in the network through various methods (e.g. machine learning techniques [33]). Since the marketer has to pay the real cost (i.e. $C(i)$ in (11)), it is vital for the proposed incentive scheme to get the cooperativeness levels (i.e. $\alpha_{i,j}$) of nodes correctly. In this model we assume that both nodes on the edge (u, v) have various information about the properties of the edge. There is a similar assumption in [31]. The difference is that in our proposed incentive scheme, it is obvious that the influencer (i.e., the node u in edge (u, v)) has motivation to lie about its cooperativeness level in order to get a higher payment. So in the VCG-like incentive scheme, the influencee (i.e., the node v in edge (u, v)) will report the influence probability $p'_{u,v}$ the influencer node u has exerted on it to the marketer, thus

$$\alpha_{u,v} = \frac{p'_{u,v}}{p_{u,v}} \tag{18}$$

2) How do we calculate the premium given to Node i (i.e. $\sigma(A) - \sigma_{-i}(A)$)?

Another possible concern on the VCG-like incentive scheme is the calculation of $\sigma(A) - \sigma_{-i}(A)$ via simulation. Since both terms of the premium require the expected final active node set size, the efficiency would be greatly improved if we can reduce the Monte-Carlo simulation times needed while preserving the accuracy of the result.

To solve this problem, from the proof of Lemma 2, we see that the expression $\sigma(A) - \sigma_{-i}(A)$ at one sample point X (i.e. $\sigma_X(A) - \sigma_{X,-i}(A)$) is exactly the marginal increase in the active set size at the end of the present step compared to the previous step. So for each simulation run we can store the value of $\sigma_X(A) - \sigma_{X,-i}(A)$ and the average number over all simulation runs is the result desired. By using this method we can avoid running two simulations separately and the efficiency is hence improved.

V. The Budget Allocation Problem

In the previous two sections we have introduced two methods, namely, the hierarchy-based seed node selection method and VCG-like incentive scheme to solve the node noncooperation problem in the influence maximization process. These two methods work at different stages. It is obvious that the final influence diffusion in terms of the size of the active node set will improve if there are more seeds chosen at the seed node selection stage, or the network is more cooperative at the influence diffusion stage. In real life, however, the marketer only has a finite budget to promote his product. If one puts all of the budget in selecting as many seed nodes as possible, the final outcome might not be satisfactory because of the noncooperation of non-seed nodes. On the other hand, if one puts too much efforts on providing incentives to non-seed nodes so that they can be cooperative during the influence diffusion process, the cascading effect may be limited because too few nodes are activated at the beginning. Therefore, there exists a tradeoff between a larger seed set and a more cooperative network. To study the tradeoff more rigorously, we formulate the budget allocation problem as follows: Given r as the cost of activating a seed node and B the budget at the viral marketer's disposal, what is the optimal seed node set size so that the final active set size is maximized? Define $K_{max} = \lfloor \frac{B}{r} \rfloor$ as the largest possible number of seed nodes the viral marketer can activate. Since normally K_{max} is a small number compared to the total number of nodes in the social network, one way to solve the problem is to enumerate from 1 to K_{max} seed nodes with the proposed seed node selection heuristic and incentive scheme embedded, so that the optimal value can be found.

A. Decision criteria of the budget

Let $S_n = \mathcal{V} \setminus S_t$ be the non-seed node set. Because β in (11) can be arbitrarily small, we can use this parameter to control the premium given to non-seed nodes in order to control the total budget. If a viral marketer wants to activate K seed nodes as well as provide all non-seed nodes the required incentives to encourage their cooperativeness, it needs enough budget to cover the activation cost of seed nodes and incentive costs of non-seed nodes. Mathematically, the criteria for a viral marketer to decide whether he has enough budget or not should be

$$B \geq r \cdot K + D \cdot \sum_{i \in S_n} \sum_{j\, neighbor\, of\, i} p_{i,j} \tag{19}$$

We omit $\alpha_{i,j}$ in (19) since we have proved in Theorem 3 that nodes will be cooperative (i.e., $\alpha_{i,j} = 1$, j neighbor of i) under the VCG-like payment scheme.

However, when a viral marketer decides whether it has enough budget at the beginning of the campaign, it has not determined the seed node set yet. In other words, it cannot know S_t and therefore S_n in advance. It is also reasonable to assume that $|S_t| \ll |S|$. Thus a decision criteria not only reflecting the reality more accurately, but also a good approximation is as follows

$$B \geq r \cdot K + D \cdot \sum_{i \in S} \sum_{j\, neighbor\, of\, i} p_{i,j} \tag{20}$$

We will use (20) in the rest of this paper.

B. Seed node selection scheme when budget is not enough

Suppose the viral marketer only has enough budget to activate K initial nodes, but cannot afford to pay all other non-seed nodes incentives as required in the VCG-like scheme. In other words,

$$r \cdot K \leq B \leq r \cdot K + D \cdot \sum_{i \in S} \sum_{j\, neighbor\, of\, i} p_{i,j} \tag{21}$$

Then the viral marketer can only provide incentives to a portion of the total nodes in the network. In order to select a satisfactory seed node set considering node noncooperation, the marketer needs to estimate the percentage of cooperative nodes (i.e., seed nodes and non-seed nodes receiving VCG-like incentives) in the network beforehand. We define the incentive coverage estimation factor as the ratio between the remaining budget and the sum of individual costs, that is

$$\eta = \frac{B - r \cdot K}{D \cdot \sum_{i \in S} \sum_{j\, neighbor\, of\, i} p_{i,j}} \tag{22}$$

In (22) the denominator is the sum of the individual costs of all nodes in the network. The reason is the same as in Section V-A.

Since the physical meaning of η can also be interpreted as the probability for a non-seed node to be cooperative during the influence propagation, thus the cooperativeness level of a non-seed node now becomes a random variable λ with two possible values, α and 1.

$$\lambda = \begin{cases} 1, & with\, probability\, \eta \\ \alpha, & with\, probability\, 1 - \eta \end{cases}$$

We further define the equivalent cooperativeness level of a non-seed node as the expected value of λ, that is

$$\alpha' = \eta + (1 - \eta) \times \alpha \tag{23}$$

This α' will be used for choosing seed nodes under the noncooperative version of the hierarchy-based influence maximization algorithm.

C. A greedy heuristic to allocate the incentive

In order to maximize the advertising revenue, one has to choose a cost-effective way to allocate the remaining budget $B - r \cdot K$ among non-seed nodes. An intuitive allocation method is to provide incentives to non-seed nodes according to their "importance" until all the budget is used up. Various metrics can be utilized to measure the possible contribution of nodes to the influence propagation, e.g. betweenness-centrality, degree-centrality [1], etc. We distribute the remaining incentives to non-seed nodes in out-degree descending order. Algorithm 2 gives a formal statement of the greedy incentive allocation heuristic.

In Lines 10 - 12 of Algorithm 2, after the seed nodes are chosen, in order to ensure that the dominant strategy for an individual node is to be cooperative, the premium given to a cooperative non-seed node i $\beta \cdot q \cdot (\sigma(A) - \sigma_{-i}(A))$ needs to be greater than zero. Let B' be the remaining budget after activating K seed nodes and covering the individual costs of selected non-seed nodes. Also define $S_i \subseteq S_n$ as the set of non-seed nodes eligible for incentives and $m = |S_i|$. β can be calculated under the following procedure,

$$\beta \cdot q \cdot (m \cdot \sigma(A) - \sum_{i \in S_i} \sigma_{-i}(A)) = B'$$

$$\Rightarrow \beta = \frac{B'}{q \cdot (m \cdot \sigma(A) - \sum_{i \in S_i} \sigma_{-i}(A))} \tag{24}$$

Actually β also needs to be calculated even if the viral marketer has enough budget which satisfies (20). The procedure is exactly the same as (24). We only have to substitute S_n for S_i in (24).

Algorithm 2 Greedy degree-based incentive allocation heuristic

Let $u_1, ..., u_{n-K}$ be non-seed nodes

Input:

A given integer K, the cost-of-influence parameter D, the amount of budget B and the cost of activating a seed node r;

Out-degree-based centrality metric of non-seed nodes in descending order: $d^{out}_{g(1)}, \ldots, d^{out}_{g(n-K)}$. g is a permutation of $Y = \{1, 2, \ldots, n-K\}$ such that $\forall i \in Y \setminus \{n - K\}, d^{out}_{g(i)} \geq d^{out}_{g(i+1)}$.

Output:

The final non-seed node set receiving incentive S_i;

The premium control parameter β.

1: $B \leftarrow B - r \cdot K$
2: Start with $S_i = \emptyset$
3: **for** $i = 1$ to $n - K$ **do**
4: $v \leftarrow v_{g(i)}$
5: **if** $B - D \cdot \sum_{u\, neighbor\, of\, v} p_{v,u} \geq 0$ **then**
6: $S_i \leftarrow S_i \bigcup \{v\}$
7: $B \leftarrow B - D \cdot \sum_{u\, neighbor\, of\, v} p_{v,u}$
8: **end if**
9: **end for**
10: $m = |S_i|$
11: **if** $m > 0$ **then**
12: $\beta = \frac{B}{q \cdot (m \cdot \sigma(A) - \sum_{i \in S_i} \sigma_{-i}(A))}$
13: **end if**

Finally we give a flow chart of the complete budget allocation decision process in Figure 2.

VI. Evaluation

A. Dataset and influence model

Co-authorship networks are widely used in simulations since many key features of social networks can be captured [34]. Therefore, the dataset chosen here is Arxiv's co-authorship network under the General Relativity and Quantum Cosmology category [35], which was also used in [36]. In this collaboration graph, each node is an author, and an edge between two authors i and j means that they have co-authored a paper. We consider the co-authoring relationships between two authors only once in case they have co-authored more than one paper, and we only take the largest connected component of the graph into consideration, which contains 4158 nodes and 13422 edges.

To estimate the final active set size, $\sigma(A)$, we use the Monte-Carlo method under the noncooperative independent cascade model. We run simulations up to 10000 times and take the average value. Two-tiered, static node cooperativeness is adopted in the simulation. That is, we set $\alpha_{i,j} = 1$ if Node i belongs to the seed-node set, otherwise $\alpha_{i,j} = \alpha < 1$. The uniform activation probability p is another essential parameter in noncooperative ICM. However, its accurate value is nontrivial and cannot be learnt easily. Learning activation probability is a separate research area and here we set $p = 0.1$ according to the following observations:

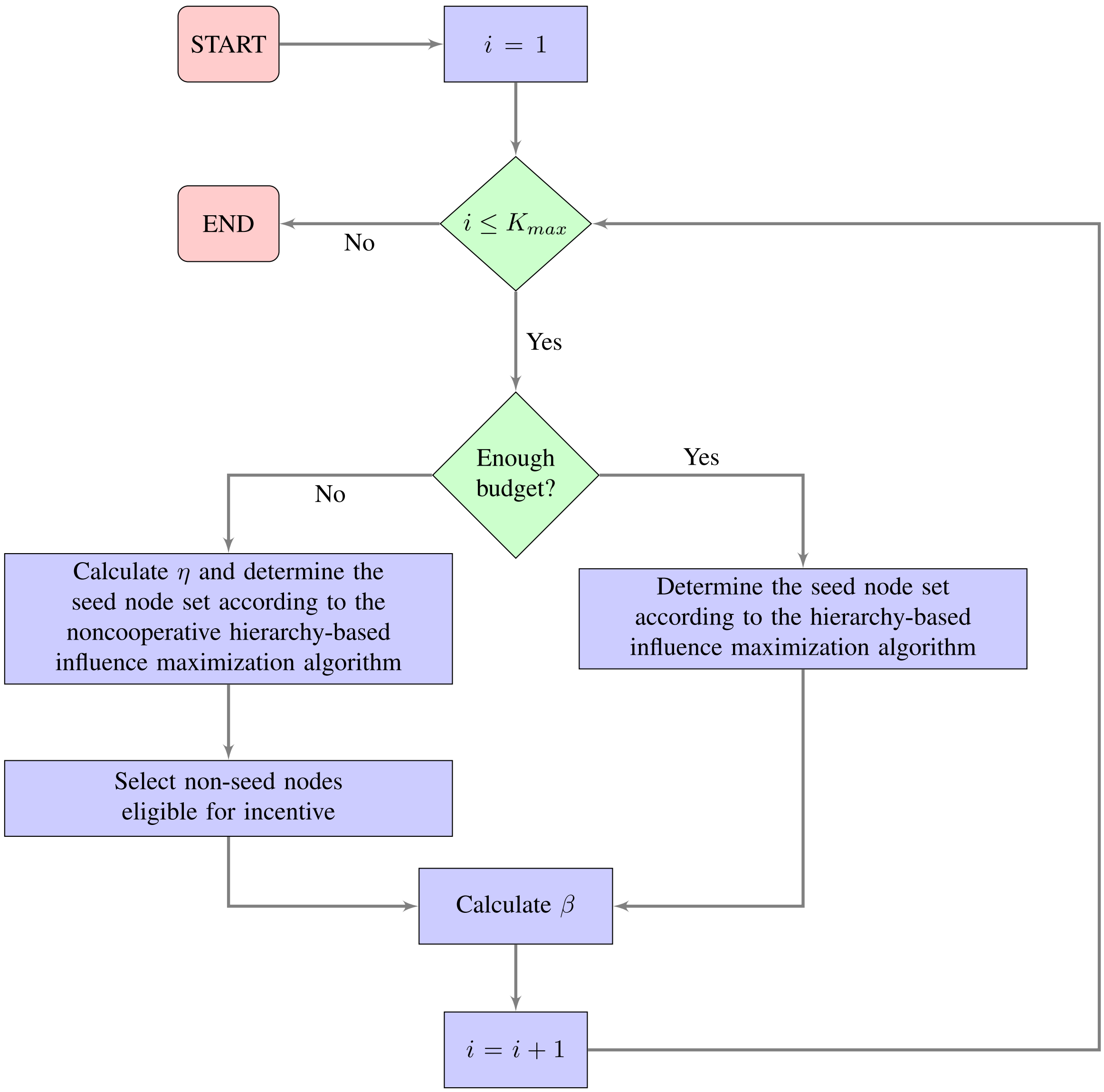


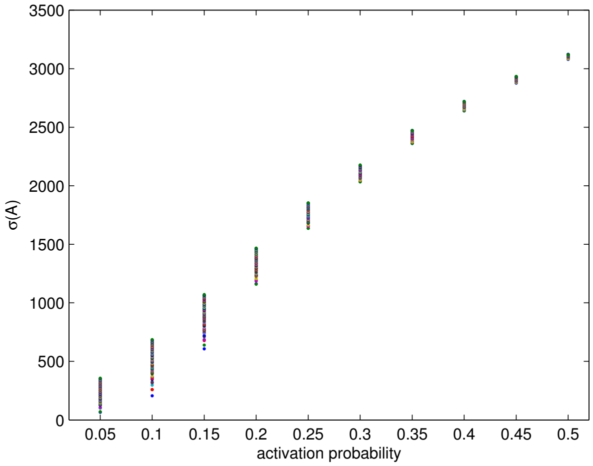
Fig. 2: Flow chart of the budget allocation decision process

In order to see the difference of information spreadings under different activation probabilities, we let p range from 0.05 to 0.5 with step 0.05, and $\alpha = 1$. For each specific p, we use the hierarchy heuristic to pick k seeds and test the influence spread, where $k = 1, 2, ..., 100$. The result is shown in Figure 3(a). Since we want to observe the effect of noncooperativeness on the influence spread, we would expect the difference of final active set size between 100 seeds (denoted as A_{100}) and just 1 seed (denoted as A_1) to be large, while their absolute values are not too small. Figure 3(b) shows $\sigma(A_{100})/\sigma(A_1)$ for different values of p. Though the ratio is large when $p = 0.05$, $\sigma(A_{100})$ is too small (less than 500). In the case of $p = 0.15$, the difference between $\sigma(A_{100})$ and $\sigma(A_1)$ is not obvious (less than 2 times). Therefore, $p = 0.1$ is the most appropriate choice.

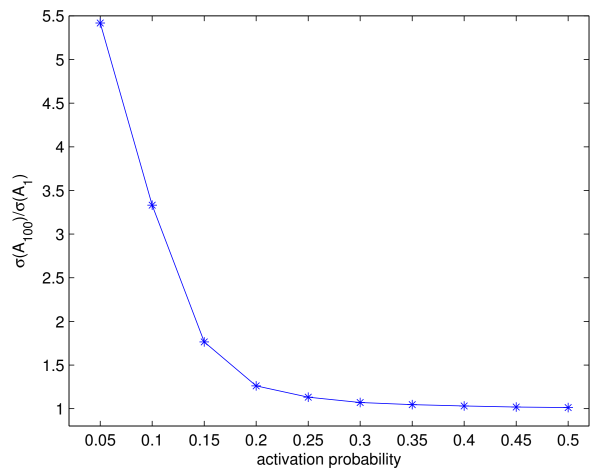
B. *Influence maximization under noncooperative ICM*

We compare our modified hierarchy-based algorithm in noncooperative ICM with three degree-based schemes.

- **Pure degree algorithm**. It simply selects seeds in an out-degree descending order [1]. It is commonly used in sociology to select influential nodes.
- **Single-discount heuristics**. The method is proposed in [8]. It is also based on degree centrality. Instead of simply using out-degree, it does not count the edge that links to a seed.
- **Degree-discount heuristics**. It is also proposed in [8]. Although it is degree-based as well, the method is not as simple as the former two. For each seed candidate v, if we denote d_v as its out degree, and t_v as the number of its out neighbors that are seeds already, then we calculate the metric $1+p\times[d_v - 2t_v - (d_v - t_v)t_v p]$ and select the node with the largest value as the new seed. The method recalculates the metric for the remaining candidates and



(a)



(b)

Fig. 3: Simulation results under ICM with different p

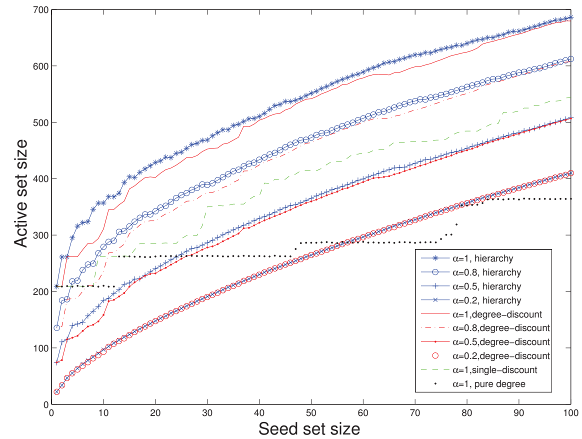


Fig. 4: Comparison of different seed selection schemes ($p = 0.1$)

iteratively run the algorithm until k seeds are picked.

Four cooperativeness levels $\alpha = 1, 0.8, 0.5, 0.2$ are used when comparing the four methods. However, none of the above three methods take node noncooperation into consideration. That is to say, for each method, once a seed set is decided, it will be used in all four cooperativeness levels. Figure 4 shows the performance comparison of these algorithms. We can see that the hierarchy heuristic clearly outperforms the other three under all α. Even in the case $\alpha = 0.8$, the hierarchy heuristic performs better than both single-discount and pure degree heuristics in a cooperative network, and thus we do not show the non-cooperative cases of both single-discount and pure degree heuristics here. In other words, the modified hierarchy heuristic is efficient and robust under non-cooperative ICM.

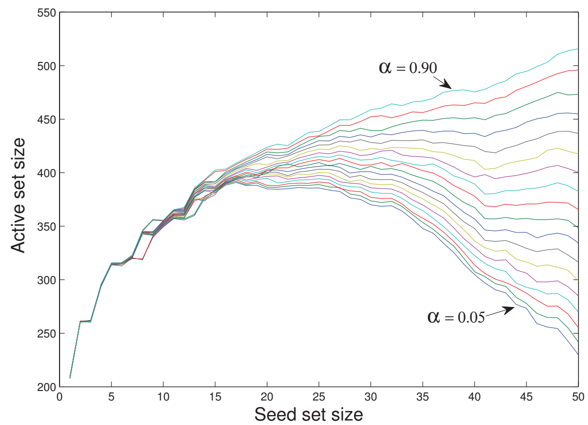
C. The budget allocation problem

As shown in Figure 4, the final active set will be larger if there are more seeds or the network is more cooperative. If budget is enough, we can use incentives to get all nodes to be cooperative. In real life, however, the marketer only has a finite budget to promote his product. Therefore, a tradeoff exists between a larger seed set and a more cooperative network. In this section, we conduct a simulation to illustrate the tradeoff. The Arxiv's co-authorship network is also used here and we make the following parameter settings:

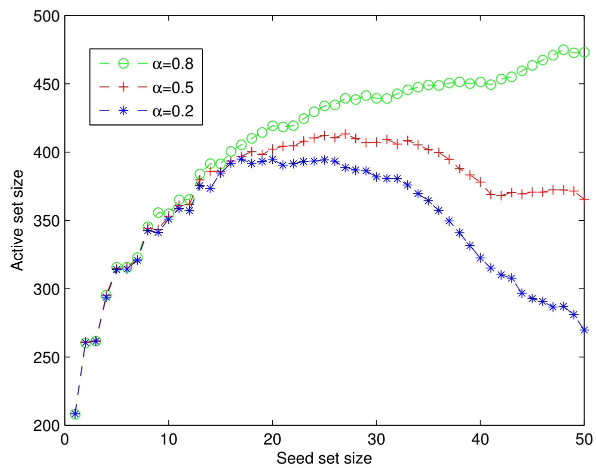
- The cost-of-influence parameter $D = 100$.
- The total budget $B = 268440$ and the cost of activating a seed node $r = 5368.8$. It means $K_{max} = 50$, i.e., a viral marketer can activate at most 50 seed nodes.

Under these settings, the simulation of the budget allocation process works like this: Firstly, decide the seed size $K = 1, 2, ..., K_{max}$. Secondly, use the method described in Section V-B to calculate α' and pick the K seeds using the hierarchy heuristic. Thirdly, use Algorithm 2 to distribute the remaining budget $B - Kr$ to the non-seed nodes, high out-degree first. Finally simulate the diffusion process 10000 times and get the average size of the final active set. We do simulations for $\alpha = 0.05, 0.10, ..., 0.90$, and Figure 5(a) shows the result. The first observation from the figure is that the performances of the hierarchy heuristic when the size of seed set is small are unsatisfactory under all cooperativeness levels. This indicates that no matter what the cooperativeness level is, it is always unwise to spend all the money to encourage non-seed nodes. Also we can conclude that the optimal policy varies according to different α. We select three cases: $\alpha = 0.2, 0.5, 0.8$ for a detailed illustration in Figure 5(b).

From the figure we can see that when $\alpha = 0.8$, the performance reaches the peak when 50 nodes are chosen as seed nodes, i.e., the viral marketer uses all his budget to activate seed nodes without providing incentives to non-seed nodes. An important implication for the viral marketer is that if the network is sufficiently cooperative, there is no need to spend any budget on encouraging cooperation. However when the network is not that cooperative, we should spare some efforts to get people to actively participate. To be more specific, if the network is cooperative (e.g. $\alpha = 0.8$), we should spend all budget on seeds, while if the network is non-



(a) $\alpha = 0.05, 0.10, ..., 0.90$, in steps of 0.05



(b) $\alpha = 0.2, 0.5, 0.8$

Fig. 5: Simulation results of budget allocation, $p = 0.1$

cooperative (e.g. $\alpha = 0.2$), the best result appears when 17 nodes are chosen as seed nodes.

D. BAP under LTM

We have shown the impact of node noncooperation on influence diffusion under LTM in [24]. Here we further study BAP under LTM through experiments. The process of budget allocation is similar to that described in VI-C, except we use the flow-based centrality metric proposed in [24] as the seed node selection method under LTM. We investigate the performance of BAP under LTM with different D, the cost-of-influence parameter. The result of the budget allocation simulation is shown in Figure 6. The x-axis represents the number of initially active nodes and the y-axis represents the final active set size. An immediate observation from the figure is that the performance of the budget allocation algorithm improves as D decreases on every seed node set size. This is because a smaller D means it is cheaper to incentivize non-seed nodes, and the performance improves due to more cooperative non-seed nodes. Moreover, we can see that for

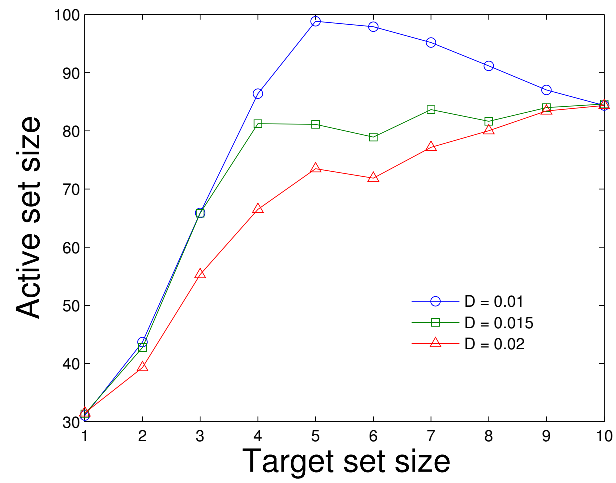


Fig. 6: Performance of the budget allocation problem with different D under LTM

$D = 0.01$, the performance reaches the peak when 5 nodes are chosen as seed nodes, while for $D = 0.015$ and 0.02, the performance reaches the peak when ten nodes are chosen as seed nodes, i.e. the viral marketer uses all his budget to activate seed nodes without providing incentives to non-seed nodes. An important implication for the viral marketer is that if it is too expensive to incentivize non-seed nodes, there is no need to spend any budget on encouraging cooperation.

VII. Conclusion

Influence maximization in noncooperative social networks is studied in this paper. Firstly, we generalize the original ICM into a noncooperative version and show that noncooperative ICM also possesses the nice property (i.e., submodularity) of the original ICM. Then, a two-stage solution is provided. For the seed node selection stage, we propose a variant of the hierarchy-based seed node selection strategy which takes node noncooperation into consideration. For the influence diffusion stage, a VCG-like incentive scheme is designed to encourage node cooperation. The proposed mechanism is IR and IC. Simulation results on a large co-authorship network show that node cooperation is very important to achieve a satisfactory advertising outcome. Evaluation results also indicate that the modified hierarchy-based influence maximization algorithm outperforms other seed node selection algorithms under various noncooperative scenarios. We also study the budget allocation problem between the two stages, and show that a marketer can utilize them to tackle noncooperation intelligently.

References

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

[2] J. E. Phelps, R. Lewis, L. Mobilio, D. Perry, and N. Raman, “Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email,” *Journal of Advertising Research*, vol. 44, no. 04, pp. 333–348, 2004.

[3] Y.-H. Chu, J. Chuang, and H. Zhang, "A case for taxation in peer-to-peer streaming broadcast," in *Proceedings of the ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems*, 2004, pp. 205–212.
[4] D. Levin, K. LaCurts, N. Spring, and B. Bhattacharjee, "Bittorrent is an auction: Analyzing and improving bittorrent's incentives," in *Proceedings of the ACM SIGCOMM*, 2008, pp. 243–254.
[5] Z. Liu, H. Hu, Y. Liu, K. W. Ross, Y. Wang, and M. Mobius, "P2P trading in social networks: The value of staying connected," in *Proceedings of the IEEE INFOCOM*, 2010.
[6] L. Anderegg and S. Eidenbenz, "Ad hoc-VCG: a truthful and cost-efficient routing protocol for mobile ad hoc networks with selfish agents," in *Proceedings of the MobiCom*, 2003, pp. 245–259.
[7] H. Zhou, K.-C. Leung, and V. O. K. Li, "Auction-based bandwidth allocation and scheduling in noncooperative wireless networks," in *Proceedings of the IEEE ICC*, 2014.
[8] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD*, 2009, pp. 199–208.
[9] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD*, 2007, pp. 420–429.
[10] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *IEEE 10th International Conference on Data Mining*, 2010, pp. 88–97.
[11] A. Goyal, W. Lu, and L. Lakshmanan, "SIMPATH: An efficient algorithm for influence maximization under the linear threshold model," in *Proceedings of the 11th IEEE International Conference on Data Mining*, 2011, pp. 211–220.
[12] X. Fan and V. O. K. Li, "Hierarchy-based algorithm for the influence maximization problem in social networks," Department of Electrical and Electronic Engineering, The University of Hong Kong, Tech. Rep. TR-2015-002, 2015.
[13] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Trans. Web*, vol. 1, no. 1, 2007.
[14] E. Even-Dar and A. Shapira, "A note on maximizing the spread of influence in social networks," in *Internet and Network Economics*, ser. Lecture Notes in Computer Science, X. Deng and F. Graham, Eds. Springer Berlin Heidelberg, 2007, vol. 4858, pp. 281–286.
[15] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD*, 2001, pp. 57–66.
[16] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the eighth ACM SIGKDD*, 2002, pp. 61–70.
[17] D. Kempe, J. Kleinberg, and E. Tardos, "Influential nodes in a diffusion model for social networks," in *Automata, Languages and Programming*, 2005, pp. 1127–1138.
[18] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate," in *SDM*, 2011, pp. 379–390.
[19] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, 2008, pp. 233–242.
[20] H. Nguyen and R. Zheng, "On budgeted influence maximization in social networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1084–1094, 2013.
[21] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: a survey," *SIGMOD Rec.*, vol. 42, no. 1, pp. 17–28, 2013.
[22] J. Kleinberg, *Cascading Behavior in Networks: Algorithmic and Economic Issues*. Cambridge University Press, 2007.
[23] Y. Yang, V. O. K. Li, and K. Xu, "An incentive scheme for non-cooperative social networks under the independent cascade model," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 2013.
[24] ——, "Influence maximization in noncooperative social networks," in *Proceedings of the IEEE GLOBECOM*, 2012.
[25] J. Goldenberg, B. Libai, and E. Muller, "Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata," *Academy of Marketing Science Review*, vol. 2001, no. 9, pp. 1–18, 2001.
[26] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
[27] J. T. Cacioppo, J. H. Fowler, and N. A. Christakis, "Alone in the crowd: the structure and spread of loneliness in a large social network." *Journal of personality and social psychology*, vol. 97, no. 6, pp. 977–991, 2009.
[28] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network," *BMJ: British medical journal*, vol. 337, 2008.
[29] Y. Narahari, D. Garg, R. Narayanam, and H. Prakash, *Game theoretic problems in network economics and mechanism design solutions*. Springer, 2009.
[30] H. Zhou, K.-C. Leung, and V. O. K. Li, "Auction-based schemes for multipath routing in selfish networks," in *Proceedings of the IEEE WCNC*, 2013.
[31] M. Mohite and Y. Narahari, "Incentive compatible influence maximization in social networks and application to viral marketing," in *The 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 1081–1082.
[32] V. Krishna, *Auction Theory*. Academic Press, 2002.
[33] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 241–250.
[34] M. E. J. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, 2001.
[35] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Trans. Knowl. Discov. Data*, vol. 1, 2007.
[36] X. Fan and V. O. K. Li, "The probabilistic maximum coverage problem in social networks," in *Proceedings of the IEEE GLOBECOM*, 2011.

Yile Yang received his B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2009, and his Ph.D. degree from the Department of Electrical & Electronic Engineering, The University of Hong Kong (HKU) in 2014. He now works as a staff researcher at Lenovo Machine Intelligence Lab in Hong Kong. His research interests are in the area of social networks and data mining.

Zhiyi Lu received the B.Eng. degree in Information Engineering from Shanghai Jiao Tong University in 2012 and the PhD degree in Electrical and Electronic Engineering from the University of Hong Kong in 2016. He is currently a postdoctoral fellow at the Department of Electrical and Electronic Engineering of the University of Hong Kong. His research interests include online social networks, optimization techniques and game theory.

Victor O.K. Li (S'80-M'81-F'92) received SB,SM, EE and ScD degrees in Electrical Engineering and Computer Science from MIT in 1977, 1979, 1980, and 1981, respectively. He is Chair Professor of Information Engineering, Cheng Yu-Tung Professor in Sustainable development and Head of the Department of Electrical and Electronic Engineering at the University of Hong Kong (HKU). He has also served as Assoc. Dean of Engineering and Managing Director of Versitech Ltd., the technology transfer and commercial arm of HKU. He served on the board of China.com Ltd., and now serves on the board of Sunevision Holdings Ltd. and Anxin-China Holdings Ltd., listed on the Hong Kong Stock Exchange. Previously, he was Professor of Electrical Engineering at the University of Southern California (USC), Los Angeles, California, USA, and Director of the USC Communication Sciences Institute. His research is in the technologies and applications of information technology, including clean energy and environment, social networks, wireless networks, and optimization techniques. Sought by government, industry, and academic organizations, he has lectured and consulted extensively around the world. He has received numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, the UK Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the Hong Kong Special Administrative Region, China. He is a Registered Professional Engineer and a Fellow of the Hong Kong Academy of Engineering Sciences, the IEEE, the IAE, and the HKIE.

Kuang Xu was a research fellow with the University of Hong Kong (HKU). He received his Ph.D degree from HKU in 2010.