Information Diffusion Meets Invitation Mechanism

Shiqi Zhang*
National University of Singapore
Southern University of Science and Technology
PyroWis AI
shiqi@pyrowis.ai

Jiachen Sun
Wenqing Lin†
jiachensun@tencent.com
edwlin@tencent.com
Tencent

Xiaokui Xiao
National University of Singapore
xkxiao@nus.edu.sg

Yiqian Huang
Bo Tang†
huangyq2020@mail.sustech.edu.cn
tangb3@sustech.edu.cn
Southern University of Science and Technology

ABSTRACT

The dissemination of information is a complex process that plays a crucial role in real-world applications, especially when intertwined with friend invitations and their ensuing responses. Traditional diffusion models, however, often do not adequately capture this invitation-aware diffusion (IAD), rendering inferior results. These models typically focus on describing the social influence process, i.e., how a user is informed by friends, but tend to overlook the subsequent behavioral changes that invitations might precipitate. To this end, we present the Independent Cascade with Invitation (ICI) model, which incorporates both the social influence process and multi-stage behavior conversions in IAD. We validate our design through an empirical study on in-game IAD. Furthermore, we conduct extensive experiments to evaluate the effectiveness of our proposal against 6 state-of-the-art models on 6 real-world datasets. In particular, we demonstrate that our solution can outperform the best competitor by up to 5× in cascade estimation and 17.2% in diffusion prediction. We deploy our proposal in the seed selection and friend ranking scenarios of Tencent's online games, where it achieves improvements of up to 170% and 20.3%, respectively.

CCS CONCEPTS

• **Information systems** → **Social networks**; **Massively multiplayer online games**; **Social recommendation**.

KEYWORDS

diffusion model; information propagation; online gaming platform

ACM Reference Format:
Shiqi Zhang, Jiachen Sun, Wenqing Lin, Xiaokui Xiao, Yiqian Huang, and Bo Tang. 2024. Information Diffusion Meets Invitation Mechanism. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3589335.3648337

*This work was done while Shiqi Zhang was an intern at Tencent.
†Corresponding authors: Bo Tang and Wenqing Lin.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05.
https://doi.org/10.1145/3589335.3648337

1 INTRODUCTION

The invitation-aware diffusion (IAD) describes the process by which information spreads from one user to another via an *invitation mechanism*, characterized by the behaviors of sending and accepting invitations. IAD is ubiquitous in various real-world social platforms, e.g., WeChat [38], Yahoo Messenger [15], LinkedIn [1], and Tencent's online gaming platforms [48]. In contrast to actions such as liking or commenting on a stranger's tweet, the invitation behavior typically occurs between existing friends, thereby spreading through *established social relationships*. For instance, online gaming platforms often organize events to strengthen friendships [19, 28–30, 34, 35, 48]. In these events, users are encouraged to invite their friends to play together and the accepting friends can further invite their friends, hence creating a cascade of invitations. Understanding the mechanism of IAD is an important problem and underpins a variety of applications, such as influence maximization [24], rumor detection [33], diffusion prediction [14], network robustness verification [32] and influencer pricing [21, 51]. However, previous works about IAD [1, 15, 38] mainly focus on exploring the macroscopic properties. For example, [15, 38] analyze the size and depth of the diffusion tree starting from selected users (called seeds), while [1] highlights that user homophily plays an important role in IAD.

In this work, we aim to design an IAD model that captures the dissemination of invitation behaviors via social connections. Despite the numerous diffusion models [2–4, 9, 10, 16–18, 27, 31] proposed in recent decades, adapting them to encompass the invitation mechanism still poses challenges. The first lies in the unclear nature of the *social influence process* of IAD, i.e., how users are activated (or informed) by others. To explain, most existing model derive their influence processes from two traditional ones: Independent Cascade (IC) [16], which assumes that each individual is independently influenced by their active friends, and Linear Threshold (LT) [18], which suggests that a user is influenced only after a sufficient number of friends have been activated. Thus, a critical question arise: does the social influence process of IAD conform to the patterns of IC, LT, or neither? Secondly, the transition of invitation and acceptance behaviors further complicates the dynamics of IAD, rendering existing models inadequate for capturing these processes.

To address these challenges, this work presents an IAD model called *Independent Cascade with Invitation* (ICI), which categorizes active users into three roles: invitee, acceptor, and inviter, with transitions between these roles occurring progressively. Specifically, an uninformed user has the chance to become an invitee based on the independent inviting process. Then, the user may transition from invitee to acceptor and subsequently to inviter with specific transition probabilities. This contrasts with IC, which assumes that users will accept and invite others unconditionally once influenced. To validate the design of ICI, we conduct an empirical study focused on the in-game IAD scenario. It not only confirms that the influence process of IAD aligns with IC, but also reveals that deconstructing IC's influence process into an independent inviting procedure coupled with multi-stage behavior transitions is statistically close to the dynamics observed in IAD. Furthermore, we integrate ICI into four key applications namely cascade estimation [4, 26], target recommendation [48], diffusion prediction [7, 14], and influence maximization [24]. For each application, we offer detailed analyses regarding the correctness and computational complexity, highlighting ICI's applicability across a spectrum of scenarios.

We experimentally evaluate the proposed ICI model against 6 representative competitors across 6 real-world datasets. In particular, we demonstrate that the proposed model outperforms all competitors in terms of RMSE, AUC, and MAP scores while (i) estimating the macroscopic features of diffusion [26] and (ii) predicting the activation of each user [7] across all tested datasets. Furthermore, we deploy our solution to two real-world scenarios in online gaming platforms of Tencent. Here, we leverage the model's ability to estimate the number of acceptors for two critical tasks: seed selection and target friend recommendation, achieving improvements of up to 170% and 20.3% in their respective evaluation metrics.

To summarize, we make the following contributions in this work:

- We devise an IAD model ICI, which has been validated through an empirical study and applied in four applications.
- We conduct experiments to show the superiority of ICI over competitors on cascade estimation and diffusion prediction tasks.
- We deploy ICI to seed selection and friend recommendation in online games, achieving significant improvements.

2 PRELIMINARIES

This section introduces diffusion models and the in-game invitation-aware diffusion (IAD), and then highlights the goal of this work.

2.1 Diffusion Models

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a social network, where $\mathcal{V}$ is a set of nodes representing users and $\mathcal{E}$ is a set of edges representing relationships. We assume that each edge $(u, v) \in \mathcal{E}$ is directed, indicating that v is a follower of and can be influenced by u. We call u (resp. v) the in-neighbor (resp. out-neighbor) of v (resp. u) and use $\mathcal{N}_u^{in}$ (resp. $\mathcal{N}_u^{out}$) to denote the set of in-neighbors (resp. out-neighbors) of u. Given a $\mathcal{G}$ and a set $\mathcal{S}$ of chosen users (called seeds), a diffusion model assumes that each $u \in \mathcal{V}$ has two possible states, inactive or active, and captures the diffusion of a given item from $\mathcal{S}$ in a stochastic manner. Initially, seeds are in active states, and subsequently, active users attempt to influence their inactive out-neighbors through an influence process. Most of existing models [2–4, 9, 10, 16–18, 27, 31, 42] extend the Independent Cascade (IC) [16] and Linear Threshold (LT) [18] models, whose influence processes are as follows.

IC model. Given a $\mathcal{G}$ and an $\mathcal{S}$, IC introduces the influence probability $p_{u,v}$ for each edge $(u, v) \in \mathcal{E}$, which indicates the likelihood that v is successfully activated by the in-neighbor u. A diffusion instance of the IC model first tags each node in $\mathcal{S}$ to be active and leaves the rest inactive at step 0. At the following step $t > 0$, each user u activated at step $t - 1$ has *one chance* to independently activate the inactive out-neighbor v with probability $p_{u,v}$.

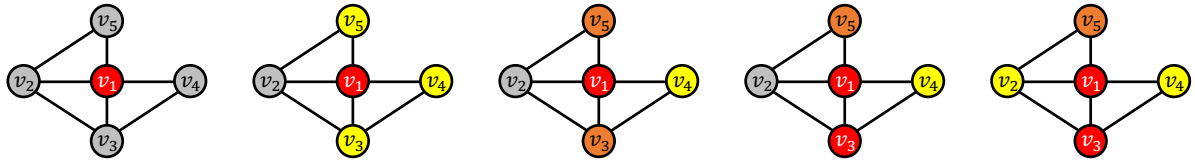
LT model. Unlike IC, the influence process of LT follows the intuition that an inactive user will switch to be active when a sufficient number of his/her in-neighbors have been activated. Formally, given a $\mathcal{G}$ and an $\mathcal{S}$, LT assumes that each edge $(u, v) \in \mathcal{E}$ is associated with an edge weight $w_{u,v}$ satisfying $\sum_{u \in \mathcal{N}_v^{in}} w_{u,v} \leq 1$. In LT, the threshold $\theta_v \in [0, 1]$ is uniformly sampled and assigned to each user $v \in \mathcal{V}$. For any step $t > 0$, an inactive user v is activated if $\phi(v, t) \geq \theta_v$, where $\phi(v, t)$ is v's threshold function and is the summation of $w_{u,v}$ w.r.t. v's in-neighbor u activated before step t.

During a diffusion instance of IC or LT, once a node is activated, it remains active in all subsequent steps. The instance terminates if no more users can be activated, and the influence spread $\sigma_{\mathcal{G}}(\mathcal{S})$ is defined as the expected number of active users from $\mathcal{S}$.

2.2 In-Game Invitation-Aware Diffusion

Event description. On Tencent's online gaming platforms, the service provider regularly conducts friendship-enhancing events to foster interactions among friends. Before an event, the service provider selects a set $\mathcal{V}_s$ of source users and a set $\mathcal{V}_t$ of target users based on historical activeness and the event requirement. For each source user $u \in \mathcal{V}_s$, a limited number of target friends are selected from $\mathcal{V}_t \cap \mathcal{N}_u^{out}$ in terms of specific recommendation strategies. As the event begins, each source user, upon their login, receives detailed information about the event and a list of recommended target friends they are encouraged to invite. Upon receiving an invitation from a source user u, a target user v is notified and decides whether to accept the invitation. If v accepts and interacts with u, both users are rewarded with the event's incentives. Due to the intersection between $\mathcal{V}_s$ and $\mathcal{V}_t$, if the target is also designated as a source user, it can invite its own target friends, thereby further propagating the event within $\mathcal{G}$. To summarize, the propagation of this event encompasses two primary elements: (i) a social influence process through *inviting relationships* and (ii) two stand-alone user behaviors, namely *source invitation* and *target acceptance*.

Dataset cleaning. The logs of a friendship-enhancing event consist of two parts: (i) an invitation dataset with tuples $(u, v, T_{u,v})$, representing that the source user $u \in \mathcal{V}$ invited the friend $v \in \mathcal{N}_u^{out}$ at timestamp $T_{u,v}$, and (ii) an acceptance dataset with tuples (v, T_v), representing that the target v accepted the invitation from one of the source friend and engaged in this event at timestamp T_v. For a better understanding of event dynamics, we clean the logs by retaining only the earliest timestamp for each distinct invitation relationship and each accepting invitee in the datasets. Additionally, we find that the invitation behavior is cascading on the social network, and hence construct the *diffusion trees* from the invitation dataset. In particular, each tree is initialized to the invitation relationships, starting from the seed inviter who spontaneously



(a) Step 0 (b) Step 1–(1) (c) Step 1–(2) (d) Step 1–(3) (e) Step 2–(1)

Figure 1: A running example of ICI (uninformed nodes: light grey, invitees: yellow, acceptors: orange, inviters: red).

engages in the event without receiving any friends' invitations. Subsequently, we add the directed edge (v, w) to the tree if (i) there exists an edge (u, v) on the tree satisfying $T_{u,v} < T_{v,w}$ and (ii) the tree is still acyclic after insertion.

2.3 Problem Formulation

We aim to devise an IAD model that captures the invitation dissemination via social connections of the given $\mathcal{G}$. Moreover, we assume that the IAD model is a progressive and single-item model. In particular, the progressive model means that an active (or informed) node will not be deactivated at the later step. The single-item model describes the propagation of a single item, as the friendship-enhancing event is unique on the platform in a specific period.

3 PROPOSED MODEL

As per the above-said description of IAD, we elaborate on the proposed model: *Independent Cascade with Invitation* (ICI), followed by conducting an empirical study to justify the design choices in ICI.

3.1 Formulation

User roles. We first define the following user roles to distinguish the components in the IAD. In particular, the *uninformed* is the user who has not received the invitation about the event (i.e., inactive user). Moreover, the active user is classified into three roles: (i) *invitee*, the user who has received the event invitation from friends; (ii) *acceptor*, the invitee who has accepted the invitation; (iii) *inviter*, the acceptor who has sent the invitation to friends.

Diffusion procedure. Given a social network $\mathcal{G}$, a seed set $\mathcal{S}$, and constants β and γ, a diffusion instance of the ICI model unfolds in the discrete steps, where the inviting process and further behavior conversions are independent. In particular, seed users are set to inviters, and other users remain uninformed at step 0. For any step $t > 0$, denote $\mathcal{V}_r^{(t-1)}$ as the set of users who become inviters at step $t-1$. For an uninformed user v at step $t-1$, the role transition at step t is illustrated as follows:

(1) Each friend inviter $u \in \mathcal{V}_r^{(t-1)} \cap \mathcal{N}_v^{in}$ has one chance to independently invite v with probability $p_{u,v}$. If there exists a friend inviter u successfully invites v, then v will become an invitee.
(2) If v successfully becomes the invitee, then v will become an acceptor with probability β, or remain an invitee.
(3) If v successfully becomes the acceptor, then v will turn into an inviter with probability γ, or still act as the acceptor.

A diffusion instance repeats these at each step until no new inviters exist. Notably, for each active user, the roles invitee, acceptor, and inviter are in ontological priority, e.g., an inviter is also an acceptor. In addition, ICI will degrade to IC when $\beta = \gamma = 1$.

Running example. Figure 1 illustrates a diffusion instance of ICI. Given an undirected graph with nodes $v_1, v_2, \ldots, v_5$, we first pick v_1 as the seed (inviter) at step 0 (Figure 1(a)). At step 1, v_1 successfully sends invitations to friends v_3, v_4, v_5 (Figure 1(b)). After being invited, all invitees flip the coin with a head probability of β. Among them, v_3 and v_5 achieve heads and become acceptors (Figure 1(c)). Subsequently, the same coin-flip operation occurs for each acceptor with a success probability of γ, which results in v_3 becoming a new inviter (Figure 1(d)) and inviting v_2 (Figure 1(e)).

Model outputs. We focus on the acceptor role in ICI as it can reflect the user engagement w.r.t. the information and is paramount to many real-world applications, such as signing up to join LinkedIn [1] or logging in to play with others [48]. In light of the live-edge graph in [24], we define the *invitation snapshot* of $\mathcal{G}$, reflecting the edge status and the user roles when an instance stops.

Definition 3.1 (Invitation Snapshot). Given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, an invitation snapshot $\mathcal{L} = (\mathcal{V}_r, \mathcal{V}_a, \mathcal{V}_l, \mathcal{E}_l)$ is a subgraph of $\mathcal{G}$, where $\mathcal{V}_l = \mathcal{V}$ is the node set associated with an acceptor set $\mathcal{V}_a \subseteq \mathcal{V}_l$ and an inviter set $\mathcal{V}_r \subseteq \mathcal{V}_a$, and $\mathcal{E}_l \subseteq \mathcal{E}$ is the set of inviting relationships. The sampling procedure of $\mathcal{L} \in \Omega$ is as follows:

(1) Include each edge $(u, v) \in \mathcal{E}$ into $\mathcal{E}_l$ with $p_{u,v}$ probability, i.e., $\Pr(\mathcal{E}_l) = \prod_{(u,v)\in\mathcal{E}_l} p_{u,v} \cdot \prod_{(u,v)\in\mathcal{E}\setminus\mathcal{E}_l} (1 - p_{u,v})$.
(2) Add each invitee $v \in \mathcal{V}_e = \{v : (u, v) \in \mathcal{E}_l\}$ to $\mathcal{V}_a$ with β probability, i.e., $\Pr(\mathcal{V}_a) = \beta^{|\mathcal{V}_a|} \cdot (1-\beta)^{|\mathcal{V}_e|-|\mathcal{V}_a|}$.
(3) Add each acceptor $v \in \mathcal{V}_a$ to $\mathcal{V}_r$ with γ probability, i.e., $\Pr(\mathcal{V}_r) = \gamma^{|\mathcal{V}_r|} \cdot (1-\gamma)^{|\mathcal{V}_a|-|\mathcal{V}_r|}$.

As each step is independent, the probability of sampling an $\mathcal{L}$ is

$$\Pr(\mathcal{L}) = \Pr(\mathcal{E}_l) \cdot \Pr(\mathcal{V}_a) \cdot \Pr(\mathcal{V}_r). \tag{1}$$

Additionally, we define a reachable set $\Gamma_{\mathcal{L}}(\mathcal{S})$ as the set of acceptors directly invited by $\mathcal{S}$ or reachable from $\mathcal{S}$ by a path of inviters in $\mathcal{L}$, Based on Definition 3.1 and the principle of deferred decisions [36], we introduce two outputs for applications in Section 4.

Definition 3.2 (Accepting Spread). Given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, a seed set $\mathcal{S}$, and ICI with constants β and γ, let $\mathcal{L} \in \Omega$ be any invitation snapshot in Definition 3.1. The accepting spread from $\mathcal{S}$ under ICI is defined as the expected number of acceptors in $\mathcal{L}$:

$$\sigma_{\mathcal{G}}(\mathcal{S}, \beta, \gamma) = \mathbb{E}_{\mathcal{L}\sim\Omega}\left[|\Gamma_{\mathcal{L}}(\mathcal{S})|\right] = \sum_{\mathcal{L}\in\Omega} \Pr(\mathcal{L}) \cdot |\Gamma_{\mathcal{L}}(\mathcal{S})|.$$

Definition 3.3 (Accepting Probability). Given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, a seed set $\mathcal{S}$, and ICI with β and γ, let $\mathcal{L} \in \Omega$ be any invitation snapshot in Definition 3.1 and $\mathbb{I}(\cdot)$ be an indicator function. The accepting probability of v is $ap_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma) = \mathbb{E}_{\mathcal{L}\sim\Omega}\left[\mathbb{I}\left(v \in \Gamma_{\mathcal{L}}(\mathcal{S})\right)\right]$.

3.2 Justifications

To justify the design of ICI, we collect the logs in a friendship-enhancing event *TXG-A* of Tencent's role-playing game, where users are from different isolated servers and only connected to others on the same server. Hence, IAD only happens among users on each stand-alone server. After preprocessing, *TXG-A* returns 20.4 thousand invitation tuples and 12.8 thousand acceptance tuples. In the sequel, we first show the existence of IAD and behavior conversions related to invitation and acceptance. Next, we elucidate the rationale for choosing IC as the base model. At last, we explain the reason for separating IC's influence process into the independent inviting procedure and two stand-alone states. Notably, similar outcomes can be observed in other gaming datasets of Section 6.

Figure 2(a) displays the distribution of diffusion tree depths originating from all seed inviters in *TXG-A*. It reveals that the depth of the diffusion tree adheres to an exponential distribution, with only about 12% of diffusion trees exceeding a depth of three. A similar trend has been observed in other real-world diffusion scenarios [8, 15, 38]. In Figure 2(b), we report the distributions of conversion rates for acceptance and invitation behaviors within servers. Specifically, the conversion rate for the acceptance (resp. invitation) behavior signifies the proportion of acceptors among all invitees (resp. inviters among all acceptors) within each server. As shown in Figure 2(b), the conversion rates that a user transitions into an acceptor and further into an inviter tend to concentrate around 0.9 and 0.6, respectively. This observation highlights the presence of multi-stage behavior conversion, indicating that once informed, a user is likely to accept an invitation and subsequently engage as an inviter with certain probabilities.

Figure 2(c) reports the distribution of the number of invitations invitees receive and the number of invitations they receive after accepting. Specifically, we find that a user can receive invitations from multiple distinct friends, with about 11% of invitees being invited by more than two distinct friends. Furthermore, users continue to receive invitations from other friends even after accepting the invitation from one friend. Notably, among users invited more than once, 76% experience repeated invitations twice or three times after acceptance. This phenomenon can be explained by the fact that, in the context of the friendship-enhancing event, two source users are permitted to invite a common friend from their recommendation lists. Consequently, a target user may receive multiple inviting notifications from different friends. This observation suggests that the invitation procedure resembles the influence process of IC. Specifically, each user can be independently invited by friends who have become inviters. It is worth noting that we provide a quantitative comparison between the ground-truth and model-predicted diffusions in Section 6, demonstrating that the IC-predicted diffusion aligns more closely with the ground-truth than LT.

Unlike the IC model, which consolidates the inviting procedure and acceptance behavior into the influence probability of each edge, the ICI model differentiates these processes by positing that the transition from invitee to acceptor does not depend on the number of invitations received. To justify this distinction, we introduce the notation $ar(s,t)$, representing the actual acceptance rate after being invited t times in server s on *TXG-A*, and compute $ar(s,t)$ for each server s and $t > 0$. Subsequently, we leverage $ar(s,1)$ and different

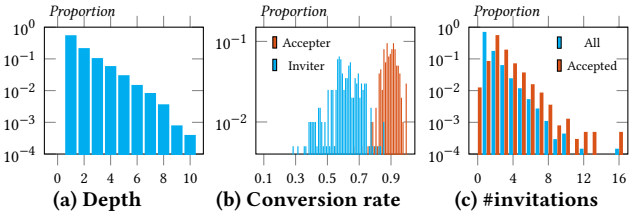


Figure 2: The histograms on *TXG-A*: (a) the depth of each diffusion tree; (b) the conversion rate from the acceptor to the inviter (cyan) and from the invitee to the acceptor (red); (c) the number of invitations invitees receive (cyan) and invitees receive after accepting (red).

assumptions to predict $ar(s,t)$ when $t > 1$. Specifically, we denote $\hat{ar}_i(s,t)$ as the acceptance rate predicted under the IC assumption, where the acceptance rate increases with the number of invitations received, i.e., $\hat{ar}_i(s,t) = 1 - (1 - ar(s,1))^t$. In contrast, under the ICI assumption, the acceptance behavior remains unaffected by the number of invitations, resulting in a predicted acceptance rate $\hat{ar}_o(s,t) = ar(s,1)$. We compare the absolute errors of IC-predicted rate $|ar(s,2) - \hat{ar}_i(s,2)|$ and ICI-predicted rate $|ar(s,2) - \hat{ar}_o(s,2)|$ in each server s, finding that the average absolute errors are 0.130 and 0.088, respectively. This indicates that the conversion assumption in ICI aligns more closely with the online gaming scenario. Furthermore, we conduct a two-sample t-test, yielding a one-sided p-value of 0.016, signifying that the smaller error of ICI is statistically significant. Additionally, we evaluate this assumption through a sensitivity analysis, as elaborated in Section 6. Similar findings have also been reported by a recent study [47].

4 APPLICATIONS

This section outlines the utilization of ICI in various applications. Section 4.1 and Section 4.2 introduce the utilization of accepting spread (Definition 3.2) and accepting probability (Definition 3.3) for macroscopic and microscopic tasks, respectively. At last, we aim to apply the ICI model to the well-studied influence maximization problem. Following IC, we define the graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, where the inviting probability $p_{u,v}$ is associated with each edge $(u,v) \in \mathcal{E}$. For ease of exposition, we defer all proofs to Appendix A.

4.1 Macroscopic Tasks

Problem 1 (Cascade Estimation [4, 26]). *Given a social network $\mathcal{G}$ and a seed set $\mathcal{S}$, the cascade estimation problem aims to predict the size and growth of the diffusion tree starting from $\mathcal{S}$.*

Problem 2 (Target Recommendation [48]). *Given a social network $\mathcal{G}$, a budget k, a source set $\mathcal{V}_s$ and a target set $\mathcal{V}_t$, the objective of target recommendation is to select at most k target neighbors from $\mathcal{V}_t \cap \mathcal{N}_u^{out}$ for each source user $u \in \mathcal{V}_s$, such that the likelihood that the user engagement among all returned pairs is maximized.*

Due to the #P-hardness [12] of evaluating $\sigma_{\mathcal{G}}(\cdot)$, we leverage the *Monte-Carlo (MC) simulation* to estimate the accepting spread, which simulates the discrete propagation step of ICI starting from $\mathcal{S}$ and takes the average number of acceptors in r trials as the estimation. Initially, the estimation $\hat{\sigma}_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)$ is set to 0. For each of r trials, at step $t = 0$, MC simulation adds the seeds in $\mathcal{S}$ to

the invitee set $\mathcal{V}_e$, acceptor set $\mathcal{V}_a$, and inviter set $\mathcal{V}_r^{(0)}$. At the following step t, each user $u \in \mathcal{V}_r^{(t-1)}$ that became an inviter at $t-1$ flips a coin with the head probability $p_{u,v}$ to invite each uninformed friend $v \in \mathcal{N}_u^{out} \backslash \mathcal{V}_e$ and v will be marked as an invitee if the result is head. The new invitee v will have the probability β to be an acceptor and be included in $\mathcal{V}_a$. After becoming an acceptor, v will further have the probability γ to become an inviter and be included into $\mathcal{V}_r^{(t)}$. This trial terminates at step t if $\mathcal{V}^{(t)} = \emptyset$, and the estimated spread $\hat{\sigma}_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)$ receives an increment of $|\mathcal{V}_a|/r$. The following theorem shows the correctness and time complexity.

Theorem 4.1. *Given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, a set $\mathcal{S}$, and* ICI *with β and γ, $\hat{\sigma}_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)$ derived by MC simulation is an unbiased estimator. For any error ϵ and failure probability p_f, by setting $r = \frac{3|\mathcal{V}| \cdot \ln(2/p_f)}{|\mathcal{S}| \cdot \epsilon^2}$,*

$$\Pr\left[|\hat{\sigma}_{\mathcal{G}}(\mathcal{S}, \beta, \gamma) - \sigma_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)| \leq \epsilon \cdot \sigma_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)\right] \geq 1 - p_f$$

and the worst-case time complexity of MC simulation is $O(r \cdot |\mathcal{E}|)$.

By MC simulation, the size and growth of the diffusion tree in Problem 1 can be estimated by the overall accepting spread and the number of acceptors at each step, respectively. Regarding Problem 2, inspired from the prior work [11], we take the estimated single accepting spread $\hat{\sigma}_{\mathcal{G}}(\{u\}, \beta, \gamma)$ as the influence centrality of each user u, and select the friends with top k largest single accepting spread to recommend.

4.2 Microscopic Task

Problem 3 (Diffusion Prediction [7, 14]). *Given a social network $\mathcal{G}$ and a seed set $\mathcal{S}$, the diffusion prediction problem aims to identify the users in $\mathcal{G}$ that are directly or indirectly influenced by $\mathcal{S}$.*

Solving Problem 3 by MC simulation resembles the above-said solution. However, MC simulation has two distinctions for diffusion prediction. In particular, we initialize the estimation $\hat{ap}_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma)$ as 1 for each seed $v \in \mathcal{S}$ and 0 for all other $v \in \mathcal{V} \backslash \mathcal{S}$. During each iteration of the r simulation trials, if a user v becomes a new acceptor, the estimated value $\hat{ap}_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma)$ increments by $1/r$. The following theorem shows the correctness and time complexity.

Theorem 4.2. *Given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$, a set $\mathcal{S}$, and* ICI *with β and γ, for any $v \in \mathcal{V}$, $\hat{ap}_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma)$ derived by MC simulation is an unbiased estimator. For any error ϵ, threshold δ, and failure probability p_f, by setting $r = \frac{3 \ln(2 \cdot |\mathcal{V}|/p_f)}{\delta \cdot \epsilon^2}$,*

$$\Pr\left[|\hat{ap}_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma) - ap_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma)| \leq \epsilon \cdot ap_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma)\right] \geq 1 - p_f$$

for each $v \in \mathcal{V}$ with $ap_{\mathcal{G}}(v, \mathcal{S}, \beta, \gamma) \geq \delta$ and the worst-case time complexity of MC simulation is $O(r \cdot |\mathcal{E}|)$.

It is worth noting that Problems 1–3 can be efficiently solved by MC simulation by setting $r = 1,000$–$10,000$ [7, 24], which is sufficient for these tasks. In addition, the whole MC simulation procedure is only invoked once for Problems 1 and 3. As for Problem 2, the simulation can be parallelly invoked from target users.

4.3 Influence Maximization

Problem 4 (Influence Maximization [24]). *Given a social network $\mathcal{G}$, an integer k, and a diffusion model* M*, the influence maximization problem asks for a seed set $\mathcal{S}$ with cardinality $|\mathcal{S}| = k$ to maximize the overall influence spread $\sigma_{\mathcal{G}}(\mathcal{S})$ under* M*.*

The crux of taking ICI (resp. accepting spread) as the input model M (resp. the objective function) of Problem 4 is showing that the accepting spread is monotone and submodular. Specifically, denote the set function on nodes of $\mathcal{G}$ as $f : 2^{|\mathcal{V}|} \to \mathbb{R}$, which is (i) *monotone* if $f(\mathcal{S}) \leq f(\mathcal{T})$ for any $\mathcal{S} \subseteq \mathcal{T}$; and is (ii) *submodular* if $f(\mathcal{S} \cup \{v\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{v\}) - f(\mathcal{T})$ for any $\mathcal{S} \subseteq \mathcal{T}$ and $v \in \mathcal{V} \backslash \mathcal{T}$. The following theorem shows that the accepting spread also satisfies monotonicity and submodularity.

Theorem 4.3. *Given a social network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$ and* ICI *with constants β and γ, the accepting spread function on any seed set $\mathcal{S} \subseteq \mathcal{V}$ is monotone and submodular.*

In light of Theorem 4.3 and the fact that ICI is orthogonal to influence maximization, a plethora of existing approximation solutions [25, 43, 44, 50] can be applied to select a size-k seed set $\mathcal{S}$ such that $\sigma_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)$ is $(1 - 1/e - \epsilon)$-approximate. Considering that the worst-case complexity by MC simulation reaches $O(r \cdot k \cdot |\mathcal{V}| \cdot |\mathcal{E}|)$, state-of-the-art solutions [5, 43, 44] estimate the spread based on random Reverse-Reachable (RR) sets. To construct an RR set $R_{\mathcal{L}}(v)$ under ICI, we first sample a snapshot $\mathcal{L}$ as per Definition 3.1, and then check whether the node v is an acceptor. If $v \in \mathcal{V}_a$, we will include v and all nodes directly pointing or reachable to v via a path of inviters into $R_{\mathcal{L}}(v)$, otherwise $R_{\mathcal{L}}(v) = \emptyset$. In practice, we do not need to materialize $\mathcal{L}$, and can employ a breadth-first search starting from v instead. By definition, for fixed v and $\mathcal{S}$, we can obtain that $R_{\mathcal{L}}(v) = \{u : v \in \Gamma_{\mathcal{L}}(\{u\})\}$ and $\Pr[v \in \Gamma_{\mathcal{L}}(\mathcal{S})] = \Pr[\mathcal{S} \cap R_{\mathcal{L}}(v) \neq \emptyset]$. We call $R_{\mathcal{L}}(v)$ a random RR set where v is randomly selected from $\mathcal{V}$ and denote $\mathcal{R}_{\mathcal{G}}$ as a set of random RR sets. Based on the proof in Borgs et al. [5], given a $\mathcal{G} = (\mathcal{V}, \mathcal{E}, p)$ and an $\mathcal{R}_{\mathcal{G}}$, we can derive that $|\mathcal{V}| \cdot \frac{\Lambda_{\mathcal{R}}(\mathcal{S})}{|\mathcal{R}_{\mathcal{G}}|}$ is an unbiased estimator of $\sigma_{\mathcal{G}}(\mathcal{S}, \beta, \gamma)$, where $\Lambda_{\mathcal{R}}(\mathcal{S})$ is the number of random RR set $R_{\mathcal{L}}(v) \in \mathcal{R}_{\mathcal{G}}$ satisfying $\mathcal{S} \cap R_{\mathcal{L}}(v) \neq \emptyset$. Motivated by this connection, we can leverage existing solutions based on OPIM-C [43] to return a seed set satisfying $(1 - 1/e - \epsilon)$-approximate with the probability at least $1 - p_f$ in the expected time of $O\left(k \ln |\mathcal{V}| + \frac{1}{\epsilon^2} \ln \frac{1}{p_f} \cdot (|\mathcal{V}| + |\mathcal{E}|)\right)$.

5 RELATED WORK

In this part, we illustrate how other typical diffusion models CT-IC [10, 17, 31], IC-N [9], LT-C [4] and F-TM [2] extending IC and LT, and justify their differences from ours. We skip other variants [3, 27, 42] as the required features are unavailable in our problem, rendering them degraded to IC or LT. Other related work focuses on learning-based models for specific downstream tasks, such as inferring influence probabilities from known diffusion trees [2, 3, 6, 7, 14, 17, 39] and predicting the next-activated user by sequential models [23, 40, 45, 46], which are outside the scope of this study.

CT-IC model. Continuous-Time IC (CT-IC) [10, 17, 31] extends IC by introducing time delays in information transmission. Once activated, users start to communicate with inactive neighbors using an independent meeting probability. If a meeting occurs at the step before the deadline τ, CT-IC follows the influence process of IC, providing one opportunity for user u to activate v. The spread in CT-IC is the expected number of active users before τ.

IC-N model. IC with Negative Opinions (IC-N) [9] introduces opinion diversity by distinguishing between positive and negative

Table 1: Dataset statistics ($K = 10^3$, $M = 10^6$).

Dataset	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	$\|\mathcal{S}\|$	Spread	Type
TXG-A	153.0K	2.3M	10.3K	12.8K	Invitation
TXG-B	155.5K	2.5M	4.9K	12.6K	Invitation
TXG-C	155.9K	2.5M	4.4K	11.0K	Invitation
TXG-D	133.9K	2.1M	12.2K	76.4K	Invitation
Diggs	279.6K	1.5M	0.6K	8.1K	Vote
Twitter	456.6K	12.5M	27.0K	38.7K	Retweet

active states. Each seed in $\mathcal{S}$ is initially activated and becomes positive with probability q or negative otherwise. Subsequently, when a user v is activated by an in-neighbor u, v is converted as follows: (i) v becomes negative if u is negative; (ii) otherwise, v becomes positive with probability q and negative otherwise. IC-N's spread is measured by the expected number of positive active users.

LT-C model. In LT with Colors (LT-C), user ratings for a product lead to three active states: adopted, promoted, and inhibited. After randomly assigning seeds to these states, LT-C follows LT's influence process, but the threshold function for an inactive user v combines active in-neighbor u's rating and edge weight $w_{u,v}$. Upon activation, a user either (i) becomes adopted with a probability or (ii) serves as a message bridge, with a chance of (ii-a) becoming promoted with a probability or (ii-b) becoming inhibited otherwise. LT-C defines the spread as the expected number of adopted users.

F-TM model. F-TM [2] also follows the influence process in LT but integrates more information into the threshold function, including the edge weight, user's positive feeling w.r.t. each feature of the given item, and internal resistance to being influenced. At last, the threshold function is wrapped into a logistic function.

Remarks. In contrast to ICI, none of the prior IC-based models considers inviter and adopter states. Among LT-based models, while LT-C distinguishes between awareness and adoption, it is still inadequate in capturing IAD due to the mismatch of influence process.

6 EXPERIMENTS

In this part, we first introduce the experimental settings and then evaluate the performance of the proposed model on cascade estimation and diffusion prediction tasks. All of our experiments are conducted on an in-house cluster consisting of hundreds of machines, each of which runs CentOS, and has 16GB memory and 12 Intel Xeon Processor E5-2670 CPU cores. For reproducibility, the source code is available at: https://github.com/jeremyzhangsq/ICI.

6.1 Experiments Setup

Datasets. We use four friendship-enhancing events from Tencent's online games and preprocess the logs of invitation relationships and user behaviors as explained in Section 2.2. Furthermore, we take the snapshot of $\mathcal{G}$ before the release time as the input graph, since $\mathcal{G}$ for a particular online game evolves when new users are registered, or friendships are modified. Notice that all datasets have been anonymized to avoid any leakage of private information. Besides the dataset about invitation diffusion, we also choose two other types of diffusion datasets *Diggs* [20] and *Twitter* [13]. In particular, the *Diggs* contains the diffusion of vote behaviors w.r.t a given story on the platform, and the *Twitter* dataset records the diffusion of *retweet* behaviors on Twitter about the announcement

Table 2: The RMSE of estimating overall spreads ($\times 10^3$).

Model	*TXG-A*	*TXG-B*	*TXG-C*	*TXG-D*	*Diggs*	*Twitter*
IC	40.6	32.7	32.7	39.7	40.9	13.2
CT-IC	*20.9*	*8.3*	*8.1*	*22.9*	30.8	42.0
IC-N	23.4	14.8	14.9	23.8	*22.0*	*76.7*
LT	97.1	100.0	101.7	88.6	59.6	227.4
LT-C	69.6	71.9	73.6	63.2	42.7	161.1
F-TM	103.1	112.0	113.4	92.2	120.6	241.6
ICI	**11.2**	**1.7**	**2.1**	**13.4**	**7.2**	**37.1**

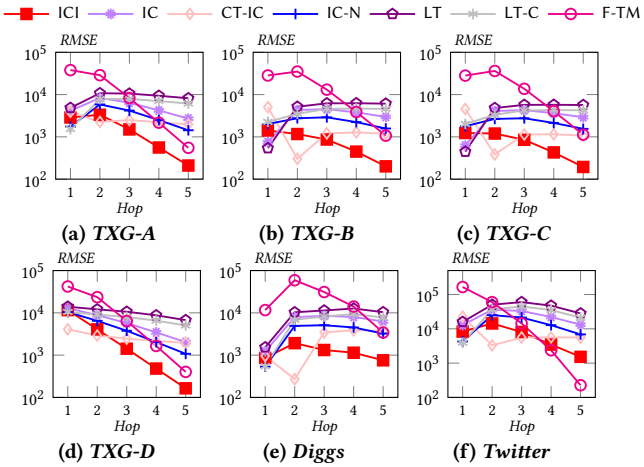


Figure 3: The RMSE of estimating spreads in each hop.

of the discovery of a new particle. For public datasets, we preserve the behaviors on the social connections and treat the user who first posted the story or tweet as the seed. The dataset statistics, including the graph, seeds, and actual spread, are shown in Table 1.

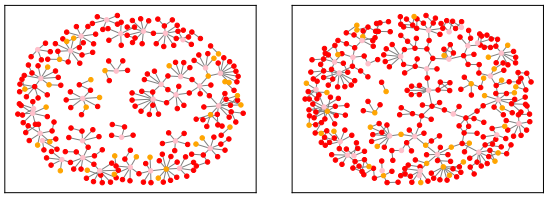
Models and parameter settings. We compare the proposed ICI model with six representative diffusion models as mentioned in Section 5: (i) IC-based models: IC [16], CT-IC [10, 17, 31], IC-N [9]; (ii) LT-based models: LT [18], LT-C [4] and F-TM [2]. Recall that LT-C and F-TM require extra information about user ratings and feelings for the product, respectively, which are not available in the provided datasets and are set to 1 for a fair comparison. Moreover, we follow the basic setting in [4, 24, 44, 50] and assign $1/|\mathcal{N}_v^{in}|$ to each edge (u, v) as the edge probability and weight for IC-based and LT-based models, respectively. For other parameters, we follow the default parameter settings for all competitors, and set the conversion constants in ICI to $\beta = 0.9$ and $\gamma = 0.6$ as explored in Section 3.2. Notice that this setting is derived from *TXG-A*, the results of which may overstate ICI's actual capabilities.

6.2 Cascade Estimation

In this section, we evaluate the performance of ICI and the competitors on the cascade estimation task (Problem 1). Given the actual seed set $\mathcal{S}$, we conduct 1,000 Monte-Carlo simulations [7] to estimate the spread for each model. After that, we follow the prior work [4] and compute the Root Mean Square Error (RMSE) between the model-predicted spread and the ground-truth spread. Regarding the model-predicted spread, ICI employs the accepting spread in Definition 3.2, and each competitor utilizes the specific spread function as explained in Section 5. We report the average RMSE after

Table 3: The AUC (%) and MAP (%) of different models in diffusion prediction (the best is bold and the second best is italic).

Model		IC	CT-IC	IC-N	LT	LT-C	F-TM	IC+	ICI
TXG-A	AUC	82.11±0.08	79.30±0.10	82.36±0.10	78.29±0.03	77.77±0.07	77.32±0.17	*82.58±0.12*	**83.36±0.06**
	MAP	20.07±0.13	18.35±0.08	20.34±0.12	16.51±0.23	16.15±0.19	18.99±0.19	*20.69±0.05*	**20.71±0.12**
TXG-B	AUC	81.96±0.05	80.76±0.05	83.06±0.11	74.17±0.04	73.98±0.10	75.95±0.17	*83.30±0.15*	**84.43±0.10**
	MAP	19.48±0.06	20.13±0.06	21.05±0.11	12.41±0.12	12.37±0.14	16.10±0.24	*21.54±0.18*	**22.05±0.15**
TXG-C	AUC	82.26±0.09	81.23±0.07	83.35±0.13	73.56±0.06	73.28±0.07	75.06±0.17	*83.56±0.13*	**84.90±0.08**
	MAP	18.82±0.12	19.42±0.08	20.43±0.16	11.10±0.21	10.89±0.09	13.83±0.20	*20.81±0.11*	**21.41±0.09**
TXG-D	AUC	78.20±0.04	74.30±0.11	78.47±0.08	78.12±0.04	77.11±0.08	75.57±0.21	*78.35±0.06*	**78.98±0.07**
	MAP	20.04±0.04	16.43±0.06	20.03±0.03	20.03±0.08	19.14±0.18	20.01±0.14	*20.08±0.04*	**20.11±0.02**
Diggs	AUC	86.65±0.03	82.03±0.04	87.58±0.06	87.82±0.02	87.83±0.03	**90.18±0.05**	88.06±0.03	*89.67±0.06*
	MAP	10.19±0.02	7.25±0.01	11.52±0.12	11.85±0.08	12.02±0.06	**26.21±0.14**	12.23±0.03	*15.95±0.22*
Twitter	AUC	70.39±0.04	72.37±0.04	72.88±0.03	69.91±0.03	69.29±0.05	68.80±0.06	*76.62±0.04*	**77.97±0.04**
	MAP	15.97±0.03	19.12±0.04	18.27±0.06	14.35±0.04	14.59±0.06	15.40±0.04	*21.17±0.03*	**22.40±0.05**



(a) ICI (b) IC-N

Figure 4: The diffusion visualization on a server of *TXG-D* (seed: pink, true positive: orange, false positive: red).

10 repeated trials and skip the standard derivation as it is always three orders of magnitude smaller than the average.

Spread estimation. We first report the RMSE of each model in predicting the overall spread starting from $\mathcal{S}$. As shown in Table 2, the proposed ICI model outperforms all competitors on the invitation and other types of diffusion datasets. In particular, the RMSE score of ICI is up to 5 × better than that of the best competitor CT-IC. Furthermore, we find that the IC-based models perform much better than the LT-based models in each invitation diffusion dataset, coherent to the observation illustrated in Section 3.2.

Growth estimation. We next evaluate the performance of each model in predicting the spread growth. Akin to the actual diffusions, the diffusion predicted by each model can also be preprocessed into a diffusion tree, as described in Section 2.2. We compare the RMSE between the number of predicted and true active users in each hop t, in which the active users have the same shortest distance t from the seed set $\mathcal{S}$. As reported in Figure 3, ICI has the best RMSE in most hops of each dataset. This is because the multi-stage role transition makes spread converge faster than competitors as the hop increases. Notice that the RMSE of CT-IC dramatically decreases from hop 1 to 2, as the activation process of CT-IC is postponed by the communication probability. We also find that F-TM has a similar trend of RMSE as ICI, but the RMSE in the smaller hop is always worse than ICI. To explain, the logistic format threshold function makes numerous users infected at the earlier step and leaves almost no reachable inactive users at the later step.

6.3 Diffusion Prediction

In this part, we evaluate the performance of different models in the micro-level diffusion prediction task (Problem 3). Given a social network $\mathcal{G}$ and an actual seed set $\mathcal{S}$, let $y_u \in \{0, 1\}$ be the label of $u \in \mathcal{V}$, where $y_u = 1$ if u is directly or indirectly infected by any user in $\mathcal{S}$ and $y_u = 0$ otherwise, and let $\hat{y}_u \in [0, 1]$ be the activation likelihood of u, which is the number of activitions of u over 1,000 simulations. Following previous works [3, 6, 7, 14], we repeat each approach 10 times and report the Area Under Curve (AUC) and Mean Average Precision (MAP).

Table 4: Performance of varying β on *TXG-D*.

	AUC (%)			MAP (%)		
β	0.3	0.6	0.9	0.3	0.6	0.9
IC+	75.86	78.09	78.35	17.96	19.60	20.09
ICI	**78.99**	**79.00**	**78.98**	**20.10**	**20.06**	**20.11**

Overall evaluation. We first report the average and the standard deviation of AUC and MAP scores for the proposal and all competitors. In Table 3, the proposed ICI outperforms all competitors on *TXG-A*, *TXG-B*, *TXG-C*, *TXG-D*, and *Twitter*, and is the second-best approach on *Diggs* in terms of both evaluation metrics. For example, ICI is 6.3% and 22.4% better than the best competitor CT-IC on *TXG-D* in terms of AUC and MAP, respectively. In addition, ICI improves the best competitor IC-N (resp. CT-IC) by 7.0% (resp. 17.2%) on *Twitter* in terms of AUC (resp. MAP), demonstrating that our proposal is effective in capturing various types of diffusion.

Case study. We next conduct a case study to compare the diffusion visualization induced by ICI and the leading competitor IC-N. We randomly select a server on *TXG-D* with 33 seeds as the test graph. We conducted 1,000 simulations, starting from the seeds, under both models. We use the state-of-the-art solution PPRviz [49] to visualize the diffusion trees generated by both models, where true (resp. false) positive infected nodes are marked in orange (resp. red). Figure 4 shows that both models produce a similar number of true positives. However, IC-N produced 30.3% more false positives than ICI, highlighting ICI's higher precision in diffusion prediction.

Ablation study. At last, we justify the effectiveness of the user roles defined in ICI. For the fair comparison, we extend the conventional IC model by involving the conversion rates into the influence probability, i.e., assigning $p_{u,v} = \beta \cdot \gamma / |\mathcal{N}_v^{in}|$ to each edge $(u, v) \in \mathcal{E}$, and call this variant the IC+ model. The only difference between ICI and IC+ is that the IC+ merges the operations in the role transitions of ICI into the social influence process. As reported in Table 3, we find that the new variant can outperform all competitors on all datasets except for *Diggs* due to utilizing the information of behavior conversion. However, the IC+ model is still beaten by ICI in all

Table 5: Performance of seed selection in Game X.

Model	ICI	IC	Degree
Spread	**2286**	*1923*	843
Invite Rate	**46.20%**	*39.64%*	32.44%

cases. For example, the MAP score of ICI is 30.5% better than that of the IC+ model on *Diggs*. In addition, to validate the design choice of the independent acceptor role, we evaluate the performance of ICI and IC+ by varying β. As reported in Table 4, as β decreases, ICI and IC+ become more distinguishable, and the effectiveness of ICI grows more significant. Specifically, when $\beta = 0.3$, ICI improves IC+ by 4.1% and 10.7% in terms of AUC and MAP, respectively, underscoring the necessity of having the stand-alone probability β.

7 DEPLOYMENT

We have deployed the proposed ICI model for the seed selection and target recommendation scenarios in several online games of Tencent, as illustrated in the sequel. The system setting for the online deployment follows that in Section 6. Due to the network effect, we follow [41] and partition all users into communities with high connectivity and profile homophily. We then conduct the online A/B testing that randomly assigns the live traffic in the same communities to a treatment group. Each approach is initially computed based on the graph instance ahead of the event and is then updated daily by using the latest graph snapshot.

7.1 Seed Selection

Tencent's online gaming platforms often organize viral marketing events, where a set of influential players (called seeds) are selected and treated as initial lucky users with a virtual incentive. Each lucky user u can invite its friend v following the invitation mechanism introduced in Section 2.2. v will also become a lucky user after accepting the invitation and playing with u. This leads to the spread of the luck privilege throughout the social network. Accordingly, the seed selection is paramount to the effectiveness of the event.

We deploy (i) the degree centrality, (ii) the proposed ICI model, and (iii) the competing IC model to separately select $k = 5000$ seeds for a viral marketing event on Tencent's battle royale game X with 227 million quarter-active users and 4 billion relationships. Specifically, the solution degree centrality is a well-adopted baseline for various Tencent's viral marketing events [22], by which the users with top-k largest degree centrality are selected as seeds. Regarding the diffusion model ICI (resp. IC), we follow the influence maximization (Problem 4) and greedily select k seeds by the state-of-the-art solution OPIM-C [43], such that the spreads of selected seeds under ICI (resp. IC) are maximized. We evaluate each approach by (i) spread, the number of lucky users excluding seeds, and (ii) invite rate, the fraction of lucky users who invite friends. The higher spread and invite rate indicates better quality. As reported in Figure 5, the approaches based on influence maximization are better than the degree centrality in both evaluation metrics, demonstrating the usefulness of the influence maximization problem in real-world viral marketing. Furthermore, ICI outperforms competitors in both metrics. Notably, ICI improves IC (resp. degree) by 15.6% (resp. 170%) in terms of spread, and improves IC (resp. degree) by 15% (resp. 37.5%) in terms of invite rate.

Table 6: Performance of target recommendation in Game Y.

Month	Measure	ICI	IC	Intimacy
Aug.	**Invite Rate**	**9.60%**	6.24%	*7.98%*
	Pay Rate	**35.15%**	*32.91%*	26.71%
Sep.	**Invite Rate**	**17.89%**	*16.85%*	16.15%
	Pay Rate	**30.91%**	24.53%	*29.80%*

7.2 Target Recommendation

Recall in Section 2.3 that a user can only invite the target friend from a recommendation list with a limited size of k during the friendship-enhancing event. Therefore, judiciously recommending k target friends for each user is pivotal to the event's performance, which motivates the target recommendation task (Problem 2). We deploy (i) Intimacy [34, 48], (ii) ICI and (iii) IC to a monthly friendship-enhancing event on a Tencent's role-playing game Y with 2.5 million monthly active users and 6.5 million relationships. Specifically, Intimacy is the well-adopted score in the target recommendation, which records the number of historical activities/interactions between friends, e.g., co-playing and gifting. Following the explanation in Section 4.1, we estimate the spread starting from v under a diffusion model (i.e., ICI or IC) for each target $v \in \mathcal{V}_t$, and take it as the influence centrality score of v. We sort each score in descending order and select the top-k target nodes to recommend. We evaluate the effectiveness of each treatment group by the click rate and pay rate. In particular, the click rate is the fraction of acceptors that invite friends, and the pay rate is the fraction of invitees that pay for this event. The higher rates indicate better quality. Table 6 reports the performance of each solution in August and September. Most notably, the treatment group ICI outperforms all competitors on both metrics and two monthly events. In August, ICI improves the best competitor Intimacy (resp. IC model) by 20.3% (resp. 6.8%) in terms of invite rate (resp. pay rate). In addition, ICI improves the best competitor IC (resp. Intimacy) by 6.2% (resp. 3.7%) on invite rate (resp. pay rate) in September.

8 CONCLUSIONS AND FUTURE WORK

In this work, we introduce a diffusion model, ICI, to capture the information dissemination process in the friend invitation scenario and evaluate its performance through extensive experiments on six different types of diffusion datasets. Our results show that ICI outperforms six state-of-the-art methods in terms of effectiveness in both cascade estimation and diffusion prediction. Additionally, the deployment of ICI in seed selection and friend ranking scenarios results in significant improvement. In future work, it would be interesting to learn personalized parameters β and γ for each user to enhance performance in other tasks, such as diffusion prediction.

ACKNOWLEDGMENTS

This work was partially supported by Singapore Ministry of Education Academic Research Fund Tier 3 (Grant No. A-8001561-00-00), Proxima Beta (Grant No. A-8000177-00-00), Shenzhen Fundamental Research Program (Grant No. 20220815112848002), Guangdong Provincial Key Laboratory (Grant No. 2020B121201001), and a research gift from Huawei Gauss department. Dr. Bo Tang is also affiliated with the Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, China.

REFERENCES

[1] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, Jure Leskovec, and Mitul Tiwari. 2015. Global diffusion via cascading invitations: Structure, growth, and homophily. In *WWW*.

[2] Nicola Barbieri and Francesco Bonchi. 2014. Influence maximization with viral product design. In *ICDM*.

[3] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2013. Topic-aware social influence propagation models. *KIS* (2013).

[4] Smriti Bhagat, Amit Goyal, and Laks VS Lakshmanan. 2012. Maximizing product adoption in social networks. In *WSDM*.

[5] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. 2014. Maximizing social influence in nearly optimal time. In *SODA*. 946–957.

[6] Simon Bourigault, Cedric Lagnier, Sylvain Lamprier, Ludovic Denoyer, and Patrick Gallinari. 2014. Learning social network embeddings for predicting information diffusion. In *WSDM*.

[7] Simon Bourigault, Sylvain Lamprier, and Patrick Gallinari. 2016. Representation learning for information diffusion through social networks: an embedded cascade model. In *WSDM*.

[8] Meeyoung Cha, Alan Mislove, and Krishna P Gummadi. 2009. A measurement-driven analysis of information propagation in the flickr social network. In *WWW*.

[9] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. 2011. Influence maximization in social networks when negative opinions may emerge and propagate. In *ICDM*.

[10] Wei Chen, Wei Lu, and Ning Zhang. 2012. Time-critical influence maximization in social networks with time-delayed diffusion process. In *AAAI*.

[11] Wei Chen and Shang-Hua Teng. 2017. Interplay between social influence and network centrality: a comparative study on shapley centrality and single-node-influence centrality. In *WWW*.

[12] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*. 1029–1038.

[13] Manlio De Domenico, Antonio Lima, Paul Mougel, and Mirco Musolesi. 2013. The anatomy of a scientific rumor. *Scientific reports* (2013).

[14] Shanshan Feng, Gao Cong, Arijit Khan, Xiucheng Li, Yong Liu, and Yeow Meng Chee. 2018. Inf2vec: Latent representation model for social influence embedding. In *ICDE*.

[15] Sharad Goel, Duncan J Watts, and Daniel G Goldstein. 2012. The structure of online diffusion networks. In *EC*.

[16] Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* (2001).

[17] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Schölkopf. 2011. Uncovering the temporal dynamics of diffusion networks. In *ICML*.

[18] Mark Granovetter. 1978. Threshold models of collective behavior. *American journal of sociology* (1978).

[19] Qintian Guo, Sibo Wang, Zhewei Wei, Wenqing Lin, and Jing Tang. 2022. Influence Maximization Revisited: Efficient Sampling with Bound Tightened. *TODS* 47, 3 (2022), 12:1–12:45.

[20] Tad Hogg and Kristina Lerman. 2012. Social dynamics of digg. *EPJ Data Science* (2012).

[21] Shixun Huang, Junhao Gan, Zhifeng Bao, and Wenqing Lin. 2023. Managing Conflicting Interests of Stakeholders in Influencer Marketing. *PACMMOD* 1, 1 (2023), 1–27.

[22] Shixun Huang, Wenqing Lin, Zhifeng Bao, and Jiachen Sun. 2022. Influence maximization in real-world closed social networks. *PVLDB* 16, 2 (2022), 180–192.

[23] Mohammad Raihanul Islam, Sathappan Muthiah, Bijaya Adhikari, B Aditya Prakash, and Naren Ramakrishnan. 2018. Deepdiffuse: Predicting the'who'and'when'in cascades. In *ICDM*.

[24] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *KDD*.

[25] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *KDD*.

[26] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deepcas: An end-to-end predictor of information cascades. In *WWW*.

[27] Wenxin Liang, Chengguang Shen, Xiao Li, Ryo Nishide, Ian Piumarta, and Hideyuki Takada. 2019. Influence maximization in signed social networks with opinion formation. *Access* (2019).

[28] Wenqing Lin. 2019. Distributed Algorithms for Fully Personalized PageRank on Large Graphs. In *WWW*. 1084–1094.

[29] Wenqing Lin. 2021. Large-Scale Network Embedding in Apache Spark. In *KDD*. 3271–3279.

[30] Wenqing Lin, Feng He, Faqiang Zhang, Xu Cheng, and Hongyun Cai. 2020. Initialization for Network Embedding: A Graph Partition Approach. In *WSDM*. 367–374.

[31] Bo Liu, Gao Cong, Dong Xu, and Yifeng Zeng. 2012. Time constrained influence maximization in social networks. In *ICDM*.

[32] Alvis Logins, Yuchen Li, and Panagiotis Karras. 2020. On the robustness of cascade diffusion under node attacks. In *WWW*.

[33] Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. 2015. Point process modelling of rumour dynamics in social media. In *ACL*.

[34] Siqiang Luo, Xiaokui Xiao, Wenqing Lin, and Ben Kao. 2019. Efficient batch one-hop personalized pageranks. In *ICDE*.

[35] Siqiang Luo, Xiaokui Xiao, Wenqing Lin, and Ben Kao. 2020. BATON: Batch One-Hop Personalized PageRanks with Efficiency and Accuracy. *TKDE* (2020).

[36] Michael Mitzenmacher and Eli Upfal. 2017. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press.

[37] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical programming* (1978).

[38] Jiezhong Qiu, Yixuan Li, Jie Tang, Zheng Lu, Hao Ye, Bo Chen, Qiang Yang, and John E Hopcroft. 2016. The lifecycle and cascade of wechat social messaging groups. In *WWW*.

[39] Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. 2008. Prediction of information diffusion probabilities for independent cascade model. In *International conference on knowledge-based and intelligent information and engineering systems*. Springer, 67–75.

[40] Aravind Sankar, Xinyang Zhang, Adit Krishnan, and Jiawei Han. 2020. Inf-vae: A variational autoencoder framework to integrate homophily and influence in diffusion prediction. In *WSDM*.

[41] Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. 2017. Detecting network effects: Randomizing over randomized experiments. In *KDD*.

[42] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social influence analysis in large-scale networks. In *KDD*.

[43] Jing Tang, Xueyan Tang, Xiaokui Xiao, and Junsong Yuan. 2018. Online processing algorithms for influence maximization. In *SIGMOD*.

[44] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence maximization in near-linear time: A martingale approach. In *SIGMOD*.

[45] Jia Wang, Vincent W Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. 2017. Topological recurrent neural network for diffusion prediction. In *ICDM*.

[46] Yongqing Wang, Huawei Shen, Shenghua Liu, Jinhua Gao, and Xueqi Cheng. 2017. Cascade Dynamics Modeling with Attention-based Recurrent Neural Network.. In *IJCAI*.

[47] Shiqi Zhang, Yiqian Huang, Jiachen Sun, Wenqing Lin, Xiaokui Xiao, and Bo Tang. 2023. Capacity Constrained Influence Maximization in Social Networks. In *KDD*. 3376–3385.

[48] Shiqi Zhang, Jiachen Sun, Wenqing Lin, Xiaokui Xiao, and Bo Tang. 2022. Measuring Friendship Closeness: A Perspective of Social Identity Theory. In *CIKM*. 3664–3673.

[49] Shiqi Zhang, Renchi Yang, Xiaokui Xiao, Xiao Yan, and Bo Tang. 2023. Effective and efficient pagerank-based positioning for graph visualization. *PACMMOD* 1, 1 (2023), 1–27.

[50] Shiqi Zhang, Xinxun Zeng, and Bo Tang. 2021. RCELF: A residual-based approach for Influence Maximization Problem. *Information Systems* (2021).

[51] Yuqing Zhu, Jing Tang, and Xueyan Tang. 2020. Pricing influential nodes in online social networks. *PVLDB* 13, 10 (2020), 1614–1627.

A PROOFS

Proof of Theorem 4.1. Let Z_i be the fraction of acceptors over all users in the i-th simulation, i.e., $|\mathcal{V}_a|/|\mathcal{V}|$, then we can derive that $Z_i \in [0,1]$, $\mathbb{E}[Z_i] = \sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)/|\mathcal{V}|$, and $\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma) = \frac{|\mathcal{V}|}{r}\sum_{i=1}^{r} Z_i$. Let $Z = \sum_{i=1}^{r} Z_i$, then

$$\mathbb{E}[\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)] = \mathbb{E}\left[\frac{|\mathcal{V}|}{r}\cdot Z\right] = \frac{|\mathcal{V}|}{r}\sum_{i=1}^{r}\mathbb{E}[Z_i] = \sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)$$

and $\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)$ is an unbiased estimator of $\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)$.

Lemma A.1 (Chernoff's Inequality [36]). *Let $Z_1, Z_2, \ldots, Z_r$ be independent random variables with $Z_i \in [0,1]$ ($\forall 1 \le i \le r$) and $Z = \sum_{i=1}^{r} Z_i$. For any $0 < \epsilon < 1$,*

$$\Pr\left[|Z - \mathbb{E}[Z]| \ge \epsilon \cdot \mathbb{E}[Z]\right] \le 2\exp\left(-\frac{\epsilon^2\cdot\mathbb{E}[Z]}{3}\right).$$

According to Lemma A.1, we can obtain that

$$\begin{aligned}&\Pr\left[|\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)-\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)|\geq\epsilon\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)\right]\\=&\Pr\left[|\tfrac{r}{|\mathcal{V}|}\cdot\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)-\tfrac{r}{|\mathcal{V}|}\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)|\geq\tfrac{r}{|\mathcal{V}|}\cdot\epsilon\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)\right]\\=&\Pr\left[|Z-\mathbb{E}[Z]|\geq\epsilon\cdot\mathbb{E}[Z]\right]\leq 2\exp\left(-\tfrac{\epsilon^2\cdot r\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)}{3\cdot|\mathcal{V}|}\right).\end{aligned}$$

Based on the above and setting $r=\frac{3|\mathcal{V}|\cdot\ln(2/p_f)}{|\mathcal{S}|\cdot\epsilon^2}$, we have

$$\begin{aligned}&\Pr\left[|\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)-\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)|\leq\epsilon\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)\right]\\=&1-\Pr\left[|\hat{\sigma}_{\mathcal{G}}(\mathcal{S},\beta,\gamma)-\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)|\geq\epsilon\cdot\sigma_{\mathcal{G}}(\mathcal{S},\beta,\gamma)\right]\geq 1-p_f.\end{aligned}$$

As each of r simulations costs $O(|\mathcal{E}|)$ time, the worst-case complexity of MC simulation is $O(r\cdot|\mathcal{E}|)$.

Proof of Theorem 4.2. For a fixed node $v\in\mathcal{V}$, let $Z_i\in\{0,1\}$, where $Z_i=1$ if v becomes an acceptor starting from $\mathcal{S}$ in the i-th simulation and $Z_i=0$ otherwise, then we can derive that $\mathbb{E}[Z_i]=ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)$, and $\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)=\sum_{i=1}^{r}Z_i/r$. Let $Z=\sum_{i=1}^{r}Z_i$, then

$$\mathbb{E}[\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)]=\mathbb{E}[Z/r]=ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)$$

and $\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)$ is an unbiased estimator of $ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)$. According to Chernoff's inequality (Lemma A.1), we can obtain that

$$\begin{aligned}&\Pr\left[|\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)-ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)|\geq\epsilon\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)\right]\\=&\Pr\left[|r\cdot\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)-r\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)|\geq\epsilon\cdot r\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)\right]\\=&\Pr\left[|Z-\mathbb{E}[Z]|\geq\epsilon\cdot\mathbb{E}[Z]\right]\leq 2\exp\left(-\tfrac{\epsilon^2\cdot r\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)}{3}\right).\end{aligned}$$

For each v with $ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)\geq\delta$, based on the above and setting $r=\frac{3\ln(2\cdot|\mathcal{V}|/p_f)}{\delta\cdot\epsilon^2}$, we have

$$\Pr\left[|\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)-ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)|>\epsilon\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)\right]<\frac{p_f}{|\mathcal{V}|}.$$

By union bound, $|\hat{ap}_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)-ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)|\leq\epsilon\cdot ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)$ holds for each $v\in\mathcal{V}$ of $ap_{\mathcal{G}}(v,\mathcal{S},\beta,\gamma)\geq\delta$ with probability at least $1-p_f$. Analogously, each of r simulation costs $O(|\mathcal{E}|)$ time, leading the worst-case complexity to $O(r\cdot|\mathcal{E}|)$.

Proof of Theorem 4.3. For a fixed snapshot $\mathcal{L}$, we first prove that $|\Gamma_{\mathcal{L}}(\mathcal{S})|$ is monotone and submodular. In particular, given any $\mathcal{S}\subseteq\mathcal{T}$, each $v\in\Gamma_{\mathcal{L}}(\mathcal{S})$ is also reachable from $\mathcal{S}$'s superset $\mathcal{T}$, i.e., $v\in\Gamma_{\mathcal{L}}(\mathcal{T})$, hence $|\Gamma_{\mathcal{L}}(\mathcal{S})|\leq|\Gamma_{\mathcal{L}}(\mathcal{T})|$ and the monotonicity holds. Given any $\mathcal{S}\subseteq\mathcal{T}$ and $v\in\mathcal{V}\backslash\mathcal{T}$, each $u\in\Gamma_{\mathcal{L}}(\mathcal{T}\cup\{v\})\backslash\Gamma_{\mathcal{L}}(\mathcal{T})$ is the node that is not reachable from $\mathcal{T}$ but is reachable from v. Since $\mathcal{S}$ is the subset of $\mathcal{T}$, $u\in\Gamma_{\mathcal{L}}(\mathcal{T}\cup\{v\})\backslash\Gamma_{\mathcal{L}}(\mathcal{T})$ is also not reachable from $\mathcal{S}$, i.e., $\Gamma_{\mathcal{L}}(\mathcal{T}\cup\{v\})\backslash\Gamma_{\mathcal{L}}(\mathcal{T})\subseteq\Gamma_{\mathcal{L}}(\mathcal{S}\cup\{v\})\backslash\Gamma_{\mathcal{L}}(\mathcal{S})$. In other words, $|\Gamma_{\mathcal{L}}(\mathcal{S}\cup\{v\})|-|\Gamma_{\mathcal{L}}(\mathcal{S})|\geq|\Gamma_{\mathcal{L}}(\mathcal{T}\cup\{v\})|-|\Gamma_{\mathcal{L}}(\mathcal{T})|$ and the submodularity holds. Based on Eq. (1) and the fact that any non-negative linear combination of monotone and submodular functions is also monotone and submodular [37], the proof completes.