A Thorough Comparison Between Independent Cascade and Susceptible-Infected-Recovered Models*

Panfeng Liu[1], **Guoliang Qiu**[1], **Biaoshuai Tao**[2], **Kuan Yang**[2]

[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University
liupf22@sjtu.edu.cn, guoliang.qiu@sjtu.edu.cn

[2] John Hopcroft Center for Computer Science, Shanghai Jiao Tong University
bstao@sjtu.edu.cn, kuan.yang@sjtu.edu.cn

Abstract

We study cascades in social networks with the independent cascade (IC) model and the Susceptible-Infected-recovered (SIR) model. The well-studied IC model fails to capture the feature of *node recovery*, and the SIR model is a variant of the IC model with the node recovery feature. In the SIR model, by computing the probability that a node successfully infects another before its recovery and viewing this probability as the corresponding IC parameter, the SIR model becomes an "out-going-edge-correlated" version of the IC model: the events of the infections along different out-going edges of a node become dependent in the SIR model, whereas these events are independent in the IC model. In this paper, we thoroughly compare the two models and examine the effect of this extra dependency in the SIR model. By a carefully designed coupling argument, we show that the seeds in the IC model have a stronger influence spread than their counterparts in the SIR model, and sometimes it can be significantly stronger. Specifically, we prove that, given the same network, the same seed sets, and the parameters of the two models being set based on the above-mentioned way, the expected number of infected nodes at the end of the cascade for the IC model is weakly larger than that for the SIR model, and there are instances where this dominance is significant.

We also study the influence maximization problem (the optimization problem of selecting a set of nodes as initial seeds in a social network to maximize their influence) with the SIR model. We show that the above-mentioned difference in the two models yields different seed-selection strategies, which motivates the design of influence maximization algorithms specifically for the SIR model. We design efficient approximation algorithms with theoretical guarantees by adapting the reverse-reachable-set-based algorithms, commonly used for the IC model, to the SIR model.

Extended version — https://arxiv.org/abs/2408.11470

1 Introduction

The study of information diffusion in social networks, such as Facebook, Twitter, and WeChat, has garnered significant attention in the fields of communication media and social science (Kempe, Kleinberg, and Tardos 2015; Domingos and Richardson 2001; Brown and Reingen 1987; Richardson and Domingos 2002; Rigobon 2002; Pastor-Satorras and Vespignani 2001; Lerman and Ghosh 2010). Information diffusion is typically characterized by *cascading*—a fundamental social network process in which a number of nodes, called *seeds*, initially possess a certain attribute or piece of information and may spread to their neighbors.

Numerous diffusion models have been developed so far. Among them, the *independent cascade model* (Kempe, Kleinberg, and Tardos 2015) is well-known and extensively studied. In the independent cascade model (hereinafter denoted by IC), each edge (u, v) is assigned a probability $p_{u,v}$. An infected node u attempts to infect its neighbor v *only once*, with a success probability of $p_{u,v}$. The events of successful infections along different edges are independent. In the IC model, an infected node remains infected throughout the cascade.

However, in many real-world scenarios, an infected node may *recover*. This phenomenon can be observed in various situations. For example, consider when an individual subscribes to a magazine or signs up for a fitness membership card under the influence of friends; eventually, this subscription or membership may be terminated. This termination could result from factors such as the user losing interest in the service or the subscription/membership expiring. Another example arises when someone introduces a new product to a friend; initially, they may promote it, but eventually, they tire of advertising it after a few days. Similarly, when a user shares a post on Twitter, their friends may retweet it only for a short period before it becomes overshadowed by other posts. The recovery of nodes may signify the loss of a certain attribute or the end of spreading this attribute, both of which undoubtedly impact the cascading process.

The concept of node recovery is effectively captured by *epidemic models*. These models divide individuals into distinct states, including *susceptible*, *infected*, and *recovered*. Various models have been formulated based on feasible state transitions (Kermack and McKendrick 1927; Allen 1994; Greenwood and Gordillo 2009). The *Susceptible-Infected-Recovered model* (Kermack and McKendrick 1927) (hereinafter denoted by SIR) includes the process from susceptible to infected and eventually to recovery and permanent immunity. In the SIR model, each node u has a *recover rate* γ_u. In each round, an infected node u attempts to infect each

*Biaoshuai Tao is the corresponding author.

neighbor v, succeeding with probability $\beta_{u,v}$, and then the node u recovers with probability γ_u. If u does not recover, it remains infected and continues to attempt to infect its neighbors in the next round and subsequent rounds until it recovers. Once u recovers, it can no longer be infected. Obviously, after a sufficiently long time, all infected nodes will be recovered, and nodes can only be either susceptible or recovered. From the applicational perspective, we are interested in the number of the nodes *that have been infected* (e.g., in the examples in the previous paragraph, the advertiser immediately receives benefits when the users pay the subscription fees/membership fees). Equivalently, we are interested in the number of the *recovered* at the end of the cascade.

Plausible similarity between IC and SIR models. Previous work has observed the similarity between IC and SIR (see, e.g., Chapter 8.2 in Chen (2020)). In the SIR model, the probability that a node u successfully infects a neighbor v before u's recovery is given by

$$p_{u,v} = \sum_{t=1}^{\infty} \gamma_u (1-\gamma_u)^{t-1} \left(1 - (1-\beta_{u,v})^t\right), \qquad (1)$$

where $\gamma_u(1-\gamma_u)^{t-1}$ represents the probability of node u recovering in round t, and $(1-(1-\beta_{u,v})^t)$ is the probability of node u successfully infecting v in at least one of the t rounds. By considering $p_{u,v}$ in Equation (1) as the IC parameter, we can observe the similarity in the cascading processes between both models. In the SIR model, we can "aggregate" the infection attempts along each edge over multiple rounds, and the overall probability becomes the IC parameter.

However, this observation raises a question and thus leads to the distinction between the IC model and the SIR model. In the SIR model, due to this aggregation, infections along the incident edges to a node are no longer independent. Intuitively, the event of u successfully infecting one of its neighbors v_1 is positively correlated with the event of u infecting another neighbor v_2, as success on one edge increases the likelihood of u not recovering for a longer period, thus enhancing the probability of success on another edge. This dependency creates a difference between the two models. Our objective is to explore whether this additional dependency on outgoing edges enhances or hampers the spread of information. Consequently, we consider the SIR model as an outgoing-edges-dependent variant of the IC rather than following the original SIR definition.

From the above analysis, we can interpret the SIR model by the IC model if the above dependency can be ignored. Past literature has noticed the similarity and this seemingly "slight" difference between the two models. However, it is unclear how significant this difference is. For example, given the same graph and the same seed set, and assuming that the parameters setting in IC and SIR satisfy Equation (1), is the spread of the seeds in these two models substantially different?

Influence maximization with SIR model. The *influence maximization problem* (InfMax), proposed by Kempe, Kleinberg, and Tardos (2015); Richardson and Domingos (2002); Domingos and Richardson (2001), is the problem of selecting a set of "influencers" (also known as *seeds*) in social networks to maximize the expected number of influenced agents. The InfMax problem has attracted remarkable attention from researchers due to its wide range of applications, including social advertising (Camarero and San José 2011; Domingos and Richardson 2001; Richardson and Domingos 2002), product adoption (Brown and Reingen 1987; Bass 1976; Mahajan, Muller, and Bass 1990; Lappas, Liu, and Terzi 2011), disease analysis (Zhan et al. 2018), rumor control (Wu and Pan 2017) and social computing (Liu et al. 2023). More comprehensive literature reviews can be found in books (Chen 2020; Chen, Castillo, and Lakshmanan 2022) and survey (Li et al. 2018).

The InfMax problem with the IC model is well-studied. In the first paper where the model is proposed, Kempe, Kleinberg, and Tardos (2015) show that a simple greedy algorithm achieves a $(1-1/e)$-approximation, and no polynomial time $(1-1/e+\epsilon)$-approximation algorithm exists assuming P $\neq$ NP. After this, extensive work has focused on designing faster algorithms while keeping the theoretical approximation guarantee (Borgs et al. 2014; Chen, Peng, and Lee 2012; Goyal, Lu, and Lakshmanan 2011; Tang, Shi, and Xiao 2015; Leskovec et al. 2007; Kempe, Kleinberg, and Tardos 2015). One of the most successful types of such algorithms is based on *reverse-reachable sets* (Borgs et al. 2014; Tang, Xiao, and Shi 2014; Tang, Shi, and Xiao 2015; Chen 2018), which yields a nearly linear time algorithm with the $(1-1/e)$ approximation ratio.

However, for those epidemic models including the SIR model, most of the existing results have focused on the *dynamic* of these epidemic models and analyzed equilibrium states (Sene 2020; Kabir, Kuga, and Tanimoto 2019; Ehrhardt, Gašper, and Kilianová 2019; Long and Wang 2020; Kumar et al. 2020). However, few studies have explored them *in the context of influence maximization*. On the other hand, maximizing the number of nodes that *have been infected at least once* remains a natural problem in multiple real-world applications. For example, in viral marketing, advertisers receive benefits or payments based on nodes' infection, irrespective of subsequent recoveries. Nonetheless, nodes' recoveries can negatively affect the cascading process. Therefore, advertisers must consider this factor to attain a more comprehensive understanding of the cascade, enabling them to identify the optimal initial influencers.

We have previously seen the similarity and the subtle difference between the IC model and the SIR model. Regarding InfMax, are the seeding strategies considerably different for the two models even if the parameters satisfy Equation (1)? If the difference is insignificant, the InfMax algorithms for the IC model can be directed applied to the SIR model by reducing the SIR model to the IC model based on Equation (1). Otherwise, we should look for algorithms that are specifically designed for the SIR model.

1.1 Our Contributions

Stronger propagation effect in the IC model. As mentioned earlier, in the SIR model, through Equation (1), we can compute the probability that a vertex u "eventually" infects v before its recovery and obtain an "aggregated prob-

ability" $p_{u,v}$. This establishes an equivalence between the SIR model and the IC model, except that the events that the infections succeed along different edges become dependent in the SIR model. We intensively study how this seemingly minor difference in dependency affects the influence spread and the seeding strategy.

Firstly, in Section 3 and Section 4, we show that this dependency can only *harm* the influence spread. Specifically, we prove the following novel observation: when comparing the two models, given the same graph, the same seed set, and the cascade model parameters related by Equation (1), the expected number of infected vertices under the IC model is always weakly larger than that under the SIR model. This is proved by developing a novel coupling method.

Furthermore, we show that the above gap in the expected number of infections can be made arbitrarily large in some instances. In addition, we also observed that, in certain networks, the optimal seeding strategies for the SIR and IC models are different, leading to significant differences in propagation outcomes. This motivates the need for designing algorithms specifically for the SIR model.

Approximation algorithms for the SIR model and its variants. We remark that the influence spread functions for all the proposed epidemic models based on SIR are *submodular* functions. This implies that the simple greedy algorithm achieves a $(1-1/e)$-approximation. However, the conventional greedy algorithm based on Monte-Carlo sampling is known to be slow in practice. We adapt those reverse-reachable-set-based algorithms to the SIR setting, which yields a $(1-1/e-\varepsilon)$-approximation algorithm with probability at least $1-n^{-\ell}$ for graphs with n nodes and parameters ε, ℓ, where the algorithm's running time is *nearly linear* in the number of nodes n.

1.2 Structure of This Paper

In Section 2, we define the two diffusion models and the influence maximization problem. In Section 3, we discuss the live-edge graph formulation of the IC and SIR models, and adapt the reverse-reachable set technique to the SIR model. These will be used in all the later sections. Our results for the theoretical comparison of the IC and SIR models are in Section 4. In Section 5, we discuss the algorithm design from the aspect of the InfMax problem with the SIR model.

More related work is discussed in the full version of our paper (Liu et al. 2024).

2 Model and Preliminaries

A social network can be represented as a directed graph $G = (V, E)$ with n nodes (i.e., *users*) and m directed edges (i.e., *social connections* between *users*). For any directed edge $(u, v) \in E$, we say (u, v) is an incoming edge (resp. outgoing edge) of v (resp. u). We also call u an incoming neighbor of v and v an outgoing neighbor of u.

Definition 1 (Diffusion Model and Influence Spread (Kempe, Kleinberg, and Tardos 2015))**.** *Given a social network (directed graph) $G = (V, E)$, a **diffusion model** Γ is a (possibly random) function that maps from a vertex set S (the seeds that are initially infected) to a vertex set $\Gamma_G(S)$ (the set of influenced vertices at the end of the spreading). We omit the subscript G when there is no confusion. Denote by $\sigma(S) = \mathbf{E}[|\Gamma(S)|]$ the expected number of vertices influenced by S.*

The goal of the *influence maximization* problem (InfMax) is to select at most k nodes as *seeds* to maximize the number of *influenced* nodes on the social network G.

Definition 2 (Influence Maximization (Kempe, Kleinberg, and Tardos 2015))**.** *Given a social network (directed graph) $G = (V, E)$, a diffusion model Γ, and a positive integer k, the objective of influence maximization is to select a subset $S \subseteq V$ with $|S| \leq k$ that maximizes the expected influence spread $\sigma(S)$.*

2.1 Independent Cascade Model

In the Independent Cascade (IC) model (Kempe, Kleinberg, and Tardos 2015), the nodes could be *active* or *inactive* in a given directed graph $G = (V, E)$. A node $v \in V$ could be activated by each of its incoming active neighbors independently. More precisely, each directed edge $e = (u, v) \in E$ is associated with an activation probability $p_e = p_{u,v} \in [0, 1]$. The influence spread of an active seed set $S \subseteq V$ unfolds in discrete timestamps as follows.

1. At timestamp 0, only nodes in S are active.
2. At each timestamp $t = 1, 2, \ldots$, each newly activated node u from the previous timestamp gets one chance to activate its inactive outgoing neighbors; and for each inactive outgoing neighbor v, u tries to activate v with a probability $p_{u,v}$. The attempts to activate neighbors are independent of each other. If multiple incoming neighbors of an inactive node attempt to activate it, each attempt is considered separately with its own probability.
3. The diffusion process terminates when no inactive node gets activated in a timestamp.

In the above process, once a node becomes active, it remains active throughout. When considering the IC model, i.e., $\Gamma = \mathrm{IC}_{\boldsymbol{p}}$ where $\boldsymbol{p} = \{p_e\}_{e \in E}$ in Definition 1, $\mathrm{IC}_{\boldsymbol{p},G}(S)$ is the set of active nodes at the end of the above diffusion process.

In Kempe, Kleinberg, and Tardos (2015), it is shown that the influence spread of the IC model can be simulated by evaluating the number of *reachable* nodes from the seed set in the *live-edge graph*. To be specific, the random live-edge graph $\mathcal{G}_{\mathrm{IC}}(G, \boldsymbol{p})$ corresponding to the instance $(G, \boldsymbol{p})$ under the IC model is generated by including each directed edge $e \in E$ in G with probability p_e. Each node v is active if it is reachable from the seed set S, that is, there exists a directed path from some nodes $v \in S$ to u in the live-edge graph. More discussions about live-edge graphs can be found in Section 3.

2.2 Susceptible-Infected-Recovered Model

In the Susceptible-Infected-Recovered Model (SIR) model (Kermack and McKendrick 1927), the nodes could be *susceptible*, *infected*, or *recovered*. The SIR diffusion process is characterized by a directed graph $G = (V, E)$ together with two sets of parameters $\boldsymbol{\beta} = \{\beta_e\}_{e \in E}$ and $\boldsymbol{\gamma} = \{\gamma_v\}_{v \in V}$, where each node $v \in V$ is assigned a

recovery probability $\gamma_v \in (0, 1]$ and each directed edge $e = (u, v)$ is associated with an infection probability $\beta_e = \beta_{u,v} \in (0, 1]$. The influence spread of an infected seed set $S \subseteq V$ unfolds in discrete timestamps as follows:

1. At timestamp 0, all nodes in the seed set S are initially infected, while the remaining nodes are considered susceptible.
2. At each timestamp $t = 1, 2, \ldots$, each node u that is infected at timestamp t performs the following operations sequentially:
   - for each susceptible outgoing neighbors v, u infects v with probability $\beta_{u,v}$;
   - u gets recovered with a recovery probability γ_u and remains infected otherwise.
3. The diffusion process terminates when all nodes are either recovered or susceptible.

When we are considering the SIR model, i.e., $\Gamma = \text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ in Definition 1, $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma},G}(S)$ is the set of recovered nodes at the end of the above diffusion process.

Remark. *It is implied that once a node becomes recovered in the* SIR *model, it can never be infected again or infect other nodes. Note that all infected nodes will eventually be recovered after a sufficiently long time.*

We also consider the spread of influence within a specific time frame T in the SIR model, referred to as the *Truncated Susceptible-Infected-Recovered* (TSIR) Model. In this model, with a seed set $S \subseteq V$ infected at time 0, the diffusion process unfolds as described earlier but stops at time T. The influence spread $\text{TSIR}_{\boldsymbol{\beta},\boldsymbol{\gamma},T,G}(S)$ is defined as the number of nodes that have been infected by time T. Thus, when taking $\Gamma = \text{TSIR}_{\boldsymbol{\beta},\boldsymbol{\gamma},T}$, the "influenced vertices" in Definition 1 are those that are infected or recovered.

3 Live-Edge Graph & Reverse Reachable Set

Similar to the equivalent formulation proposed by Kempe, Kleinberg, and Tardos (2015) for the IC model, we demonstrate that the influence spread of the diffusion models we consider can be formulated using a model-specific live-edge graph. This live-edge graph formulation allows us to study the models using the unified characterization of the *reverse reachable set* (Borgs et al. 2014) in our later discussion.

3.1 Live-Edge Graph Formulation

The live-edge graph of a diffusion model is a random spanning sub-graph of G that reflects the diffusion behavior. Specifically, each edge in the original network is either "*live*" (active) or "*blocked*" (inactive), based on the model's spread probabilities, and the live-edge graph is the subgraph consisting of live edges. Roughly speaking, each edge (u, v) is live (included in the live-edge graph) with a probability, which represents the likelihood of u activating v during the influence spread period. Thus, a live-edge graph is an equivalent representation of a spreading process on G based on a diffusion model, and it serves as a sampling graph that captures one possible propagation of the model. By evaluating the number of *reachable* nodes from a given seed set S in the *live-edge graph*, we can simulate the spreading ability of S in the diffusion process.

Definition 3 (Live-edge Graph Formulation)**.** *Given a social network (directed graph) $G = (V, E)$ with the diffusion model Γ, we say Γ is a live-edge graph diffusion model if there exists a measure μ over the collection of (possibly edge-weighted) spanning sub-graphs of G, such that the influence spread $\sigma(S)$ equals the expected number of reachable nodes from S in the live-edge graph $\mathcal{G}$ sampled from μ for any seed set $S \subseteq V$.*

For convenience, we abuse $\mathcal{G}$ to represent the edge set $E(\mathcal{G})$ of the live-edge graph $\mathcal{G}$ and thus $e \in \mathcal{G}$ means that e is one of the edges in $\mathcal{G}$ for any $e \in E$ in subsequent discussion.

In the following, we introduce the live-edge graph formulation for the SIR model. For the TSIR model, the live-edge graph characterization can be found in the full version of our paper.

The live-edge graph for the SIR **model.** The influence spread of seed nodes in the SIR model on instance $(G = (V, E), \boldsymbol{\beta}, \boldsymbol{\gamma})$ can be characterized by the number of reachable nodes in the live-edge graph

$$\mathcal{G}_{\text{SIR}} = \mathcal{G}_{\text{SIR}}(G, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathcal{G}_{\text{SIR}}\left(G, \{\boldsymbol{R}_v\}_{v\in V}, \{\boldsymbol{I}_e\}_{e\in E}\right)$$

induced by a collection of independent variables

$$\boldsymbol{R}_v = (R_{v,1}, R_{v,2}, \ldots), \quad \text{and} \quad \boldsymbol{I}_e = (I_{e,1}, I_{e,2}, \ldots),$$

where $R_{v,t} \sim \texttt{Bern}(\gamma_v)$ for any $v \in V, t \in \mathbb{N}^+$ and $I_{e,t} \sim \texttt{Bern}(\beta_e)$ for any $e \in E, t \in \mathbb{N}^+$ are Bernoulli random variables. Here, $R_{u,t} = 1$ denotes that u is recovered at the t-round after u's infection, and $R_{v,t} = 0$ otherwise. Similarly, $I_{(u,v),t}$ is the indicator random variable for the event that u successfully infects v at the t-th round after u is infected. Specifically, each directed edge $e = (u, v) \in E$ is included in $\mathcal{G}_{\text{SIR}}$ if there exists some $t^* \geq 1$ such that $I_{e,t^*} = 1$ and $R_{u,[t^*-1]} = (R_{u,1}, R_{u,2}, ..., R_{u,t^*-1})$ is a sequence of zeros with length $t^* - 1$. In other words, the influence spread from node u to node v succeeds at timestamp t^* before the node u gets recovered, and the edge e represents a successful infection event between nodes u and v that could have occurred. Here, we say a node v is reachable from u in $\mathcal{G}_{\text{SIR}}$ if there exists a directed path from u to v in $\mathcal{G}_{\text{SIR}}$. An example is illustrated below.

Example 1. Consider a graph $G = (V, E)$ with $V = \{u, v, w\}$ and $E = \{(u, v), (u, w)\}$. Assume we have sampled indicator random variables $\boldsymbol{R}_u = (R_{u,1}, R_{u,2}, \ldots) = (0, 0, 0, 1, \ldots)$, $\boldsymbol{I}_{(u,v)} = (I_{(u,v),1}, I_{(u,v),2}, \ldots) = (0, 0, 1, \ldots)$, and $\boldsymbol{I}_{(u,w)} = (I_{(u,w),1}, I_{(u,w),2}, \ldots) = (0, 0, 0, 0, 0, 1, \ldots)$. In this example, u is recovered at round 4. The infection from u to v is successful at round 3, and the infection from u to w is successful at round 6. Therefore, u can successfully infect v before u's recovery, while u cannot infect w before its recovery. Correspondingly, in the live-edge graph, (u, v) is live and (u, w) is blocked, i.e., $(u, v) \in \mathcal{G}_{\text{SIR}}$ and $(u, w) \notin \mathcal{G}_{\text{SIR}}$.

The proof of the following proposition is straightforward: this is just a rephrasing of the same stochastic process.

Please refer to the full version of our paper for the details of the proof.

Proposition 4. *$\mathcal{G}_{\text{SIR}}$ defined above is a live-edge graph formulation of $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$, namely,*

$$\sigma_{\text{SIR}}(S) = \mathbf{E}\left[\textit{the number of reachable nodes from } S \textit{ in } \mathcal{G}_{\text{SIR}}\right].$$

As mentioned in Section 1, we are interested in comparing the two diffusion processes corresponding to IC and SIR respectively with the equal marginal probability for u successfully infecting v along each edge (u, v). Equation (1) ensures this, which is described in the proposition below (whose proof is straightforward).

Proposition 5. *Given any directed graph $G = (V, E)$ together with the diffusion models $\text{IC}_{\boldsymbol{p}}$ and $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ where the parameters satisfying Equation* (1)*, it holds that $\mathbf{Pr}[e \in \mathcal{G}_{\text{IC}}] = \mathbf{Pr}[e \in \mathcal{G}_{\text{SIR}}]$ for each $e \in E$.*

3.2 Reverse Reachable Set Characterization

The influence spread of the aforementioned models can be characterized by the notion of the *reverse reachable set* (Borgs et al. 2014) under the live-edge graph formulation. When considering a live-edge graph diffusion model, we sometimes use $\mathcal{G}$ to represent its diffusion process instead of Γ. The definition of the reverse reachable set can be stated as follows.

Definition 6 (Reverse Reachable Set)**.** *Given a directed graph $G = (V, E)$ with the live-edge graph $\mathcal{G}$, the reverse reachable set of a node $v \in V$, denoted by $\text{RR}_{\mathcal{G}}(v)$, is the set of all nodes in $\mathcal{G}$ that can reach v. Furthermore, let $\text{RR}_{\mathcal{G}}$ be the random set $\text{RR}_{\mathcal{G}}(v)$ with v selected uniformly at random from V.*

Proposition 7. *Given a directed graph $G = (V, E)$ with a live-edge graph diffusion model $\mathcal{G}$, the probability that the diffusion process from any seed set $S \subseteq V$ can influence a node v equals the probability that S overlaps with the set $\text{RR}_{\mathcal{G}}(v)$, i.e., $\mathbf{Pr}[S \cap \text{RR}_{\mathcal{G}}(v) \neq \varnothing]$. Furthermore, the influence spread satisfies:*

$$\begin{aligned}\sigma(S) &= \sum\nolimits_{v \in V} \mathbf{Pr}\left[S \cap \text{RR}_{\mathcal{G}}(v) \neq \varnothing\right] \\ &= |V| \cdot \mathbf{Pr}\left[S \cap \text{RR}_{\mathcal{G}} \neq \varnothing\right]\end{aligned}$$

In later discussion, we sometimes use $\text{RR}_{\text{IC}}(v)$ (RR_{IC}, resp.) to denote the reverse reachable set of a node v (random reverse reachable set, resp.) corresponding to the live-edge graph $\mathcal{G}_{\text{IC}}$ for simplicity. The notations $\text{RR}_{\text{SIR}}(v)$, $\text{RR}_{\text{TSIR}}(v)$, RR_{SIR}, and RR_{TSIR} are defined similarly.

4 Theoretical Comparison Between the IC and SIR Models

As discussed in Section 1, we can "aggregate" the infection attempts in multiple rounds along each edge in the SIR model so that the SIR model and the IC model can be related by Equation (1). However, if we view the SIR model in this way, infections across different outgoing edges of a node are correlated. In this section, we prove that this correlation *negatively* affects the cascade: given a graph G and a set of seeds S and setting the parameters of IC and SIR to satisfy Equation (1), the influence spread of IC dominants SIR. Moreover, we further show that the differences between IC and SIR lead to different seeding strategies.

We first show the positive correlation property with respect to the occurrence of the edges in the SIR model (Section 4.1). Together with the reverse reachable set characterization of influence spread, we demonstrate that the influence spread in the IC model dominates the one in the corresponding SIR model conditioned on the comparable spreading ability on each edge by a coupling between the reverse reachable set (Theorem 9 in Section 4.2). Furthermore, in certain scenarios, we find that the IC model can significantly dominate the SIR model. In our full version of the paper, We delve deeper into these scenarios to claim that the influence of IC could significantly dominate SIR, which further implies different seeding strategies.

4.1 Positive Correlation in the SIR Model

According to the live-edge graph formulation of the SIR model, the positive correlation can be described as follows: the probability of an edge being included in the live-edge graph decreases if other edges in the underlying graph G are not included.

Lemma 8. *Given a directed graph $G = (V, E)$ with the diffusion model $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$, we have*

$$\mathbf{Pr}\left[e \in \mathcal{G}_{\text{SIR}} \mid E' \cap \mathcal{G}_{\text{SIR}} = \varnothing\right] \leq \mathbf{Pr}\left[e \in \mathcal{G}_{\text{SIR}}\right]$$

for any $E' \subseteq E$ and $e \in E \setminus E'$.

A tiny example is helpful to understand Lemma 8. Given $G = (V, E)$ with $V = \{u, v, w\}$, $E = \{(u, v), (u, w)\}$, and letting $e = (u, v)$ and $E' = \{(u, w)\}$, Lemma 8 says that the event $(u, w) \notin \mathcal{G}_{\text{SIR}}$ makes the edge (u, v) less likely to be live. The proof of this lemma is listed in the full version of our paper. Intuitively, in the above example, if knowing (u, w) is not live, u is more likely to recover at earlier rounds, which decreases the chance that (u, v) is live. In general, knowing a set of edges fails to be live makes an edge less likely to be live.

Remark. *The dependency exists only on the outgoing edges from the same vertex. It is obvious from the definition of* SIR *that the events $e_1 \in \mathcal{G}_{\text{SIR}}$ and $e_2 \in \mathcal{G}_{\text{SIR}}$ are independent if e_1 and e_2 are not outgoing edges of the same vertex.*

To show that IC dominates SIR given the same seed set and with parameters satisfying Equation (1), a natural idea is to couple the two spreading processes such that the set of infected vertices in the SIR process is a subset of the set of infected vertices in the IC process. However, in the following example, we demonstrate that such a coupling is unlikely to exist.

Consider a seed set with a single seed s that has many outgoing neighbors. In the IC model, the events that s successfully infects its neighbors are independent. In the SIR model, due to Equation (1), s infects each neighbor with the same probability as in the IC model. Thus, the expected number of infected neighbors is the same in both models due to the linearity of expectation. Intuitively, due to the positive correlation, in the SIR model, the number of infected neighbors is

more likely to be either very small or very large. However, because of the same expectation, it is impossible to find a coupling such that the set of infected neighbors in SIR is a subset of the set of infected neighbors in IC. In fact, from this example, it is not even intuitively clear that IC is superior to SIR.

In the next section, we will use a coupling on sampling the reverse-reachable sets of an arbitrary vertex v. We will show that viewing the cascade in such a "backward" way helps us better understand the relationship between the two models.

4.2 IC **Dominates** SIR

Given a directed graph $G = (V, E)$, we rigorously prove that the influence spread of any seed set in the IC model dominates that of the SIR model, under the assumption that the parameters satisfy Equation (1).

Theorem 9. *Given any directed graph $G = (V, E)$ together with the diffusion models $\text{IC}_{\boldsymbol{p}}$ and $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ where the parameters satisfy Equation* (1)*, we have for any set $S \subseteq V$,*

$$\sigma_{\text{IC}}(S) \geq \sigma_{\text{SIR}}(S).$$

According to the linearity of the expectation and the reverse reachable set characterization of the influence spread Proposition 7, it suffices to show the following lemma.

Lemma 10. *Given any directed graph $G = (V, E)$ together with the diffusion models $\text{IC}_{\boldsymbol{p}}$ and $\text{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$, where the parameters satisfy Equation* (1)*, we have for any seed set $S \subseteq V$ and any node $v \in V$,*

$$\mathbf{Pr}\left[S \cap \text{RR}_{\text{IC}}(v) \neq \varnothing\right] \geq \mathbf{Pr}\left[S \cap \text{RR}_{\text{SIR}}(v) \neq \varnothing\right].$$

The lemma above straightforwardly implies Theorem 9.

Proof of Theorem 9. According to Proposition 7, we have

$$\begin{aligned}\sigma_{\text{IC}}(S) &= \sum\nolimits_{v \in V} \mathbf{Pr}\left[S \cap \text{RR}_{\text{IC}}(v) \neq \varnothing\right] \\ &\geq \sum\nolimits_{v \in V} \mathbf{Pr}\left[S \cap \text{RR}_{\text{SIR}}(v) \neq \varnothing\right] = \sigma_{\text{SIR}}(S)\end{aligned}$$

where the inequality follows from Lemma 10.

Some intuitions for the correctness of Lemma 10. To prove Lemma 10, considering an arbitrary fixed v, we define a coupling between $\mathcal{G}_{\text{IC}}$ and $\mathcal{G}_{\text{SIR}}$ such that $\text{RR}_{\text{SIR}}(v) \subseteq \text{RR}_{\text{IC}}(v)$. The existence of such a coupling immediately implies the lemma. It then remains to define such a coupling. We describe the high-level ideas here, and the full proof is available in the full version of our paper.

The edges in both $\mathcal{G}_{\text{IC}}$ and $\mathcal{G}_{\text{SIR}}$ are revealed on a need-to-know basis. Specifically, the edges are revealed in a reverse Breadth-First-Search process: let U denote the set of vertices that can reach vertex v, initialized as $U = \{v\}$; in each iteration, for every vertex x that has out-neighbors in U, we reveal the corresponding outgoing edges of x; if one of these outgoing edges is live, then x is included in U. In addition, the corresponding Breadth-First-Search tree forms an in-arborescence (a directed tree rooted at v where each edge is from the child to the parent): if the node x has at least one outgoing edge connecting to U that is live, we pick an arbitrary such live edge and include it in the in-arborescence. Notice that, after this process, unrevealed edges have no effect on $\text{RR}_{\text{IC}}(v)$ or $\text{RR}_{\text{SIR}}(v)$. We couple the two reverse Breadth-First-Search processes for $\text{RR}_{\text{IC}}(v)$ and $\text{RR}_{\text{SIR}}(v)$ respectively such that the in-arborescence for the former is a superset of the in-arborescence for the latter. The fact that each node in an in-arborescence has an out-degree at most 1 ensures that this is always possible. This is better illustrated by the following example.

Suppose at a certain stage of the Breadth-First-Search process where, in both IC and SIR processes, U is the set of vertices that are already in the in-arborescence, and suppose now we are revealing the outgoing edges of a vertex $x \notin U$ to see if x is in the in-arborescence. Suppose, for example, $y_1, y_2, y_3 \in U$ are all the outgoing neighbors of x that belong to U. If at least one of (x, y_1), (x, y_2), and (x, y_3) is live in the SIR process, then x is included in U in the next iteration. Our coupling ensures that x is also included in U in the IC process. To see this, taking an example where $(x, y_2), (x, y_3) \in \mathcal{G}_{\text{SIR}}$ and $(x, y_1) \notin \mathcal{G}_{\text{SIR}}$, which happens with probability $\mathbf{Pr}\left[(x, y_1) \notin \mathcal{G}_{\text{SIR}}\right] \cdot \mathbf{Pr}\left[(x, y_2) \in \mathcal{G}_{\text{SIR}} \mid (x, y_1) \notin \mathcal{G}_{\text{SIR}}\right] \cdot \mathbf{Pr}\left[(x, y_3) \in \mathcal{G}_{\text{SIR}} \mid (x, y_1) \notin \mathcal{G}_{\text{SIR}}, (x, y_2) \in \mathcal{G}_{\text{SIR}}\right]$. The live edge with the smallest index, which is (x, y_2) in this case, is included in the in-arborescence, and x is included in U. We consider these three probabilities in the IC model. For the first, we have $\mathbf{Pr}\left[(x, y_1) \in \mathcal{G}_{\text{SIR}}\right] = \mathbf{Pr}\left[(x, y_1) \in \mathcal{G}_{\text{IC}}\right]$ due to Equation (1). For the second, we have

$$\begin{aligned}&\mathbf{Pr}\left[(x, y_2) \in \mathcal{G}_{\text{SIR}} \mid (x, y_1) \notin \mathcal{G}_{\text{SIR}}\right] \\ \leq &\mathbf{Pr}\left[(x, y_2) \in \mathcal{G}_{\text{SIR}}\right] && \text{(Lemma 8)} \\ = &\mathbf{Pr}\left[(x, y_2) \in \mathcal{G}_{\text{IC}}\right] && \text{(Equation (1))} \\ = &\mathbf{Pr}\left[(x, y_2) \in \mathcal{G}_{\text{IC}} \mid (x, y_1) \notin \mathcal{G}_{\text{IC}}\right]. \\ & && \text{(Independence in IC)}\end{aligned}$$

By coupling the events for the first two edges (x, y_1) and (x, y_2), we already ensure that x is included in U in the IC process as well. Notice that, for the third probability, although the relationship between $\mathbf{Pr}\left[(x, y_3) \in \mathcal{G}_{\text{SIR}} \mid (x, y_1) \notin \mathcal{G}_{\text{SIR}}, (x, y_2) \in \mathcal{G}_{\text{SIR}}\right]$ and $\mathbf{Pr}\left[(x, y_3) \in \mathcal{G}_{\text{IC}} \mid (x, y_1) \notin \mathcal{G}_{\text{IC}}, (x, y_2) \in \mathcal{G}_{\text{IC}}\right]$ cannot be implied by Lemma 8, we have already included x in U regardless of the status of (x, y_3). Even if the conditional probability that (x, y_3) is live for SIR is higher than it is for IC, this does not give SIR any advantages.

More generally, if x has many out-neighbors $y_1, y_2, \ldots, y_k$ in U and i is the smallest index such that (x, y_i) is live in the SIR process, by Lemma 8, the event that $(x, y_1), \ldots, (x, y_{i-1})$ are blocked reduces the chance that (x, y_i) is live in the SIR process, while this event has no effect on the chance that (x, y_i) is live in the IC process. By coupling the event corresponding to the status of the first i edges, we can make x be included in U for the IC process whenever this happens in the SIR process. Although (x, y_i) being live may increase the chances of $(x, y_{i+1}), \ldots, (x, y_k)$ being live in the SIR process, this does not give SIR any advantages against IC as x is already included in U due to the inclusion of (x, y_i)

in the in-arborescence. The fact that we only need one live edge from x to U and the correlation property described in Lemma 8 make IC superior to SIR.

Notice that the main purpose of the example above is to give readers some intuitions. A rigorous proof can be found in our full version of the paper. In our paper's full version, we also show that, conditioned on the comparable spreading ability of each edge, we also demonstrate that IC could significantly dominate SIR in some special networks, and this dominance further yields significantly different seeding strategies under the two models.

5 InfMax in the SIR-based Models

In this section, we present algorithmic results for InfMax in the SIR model. It is worth noting that InfMax in various diffusion models can be formulated as a *submodular* optimization problem. In our case with the SIR model, the expected influence spread function $\sigma(\cdot)$ is also submodular. This follows from the fact that a diffusion model with a live-edge graph formulation is always submodular (Kempe, Kleinberg, and Tardos 2015). With submodularity, it is well-known that the simple greedy algorithm achieves a $(1-1/e)$-approximation (Nemhauser, Wolsey, and Fisher 1978). The same holds for the variant TSIR model (definition included in the paragraph right before Section 3).

However, computing the function $\sigma(\cdot)$ is a #P-hard problem (Kempe, Kleinberg, and Tardos 2015; Chen, Wang, and Wang 2010). To apply the greedy algorithm, the influence spread function $\sigma(\cdot)$ is approximated by the Monte-Carlo method, so we have a randomized algorithm that achieves a $(1 - 1/e - \varepsilon)$-approximation with high probability. Notice that this approximation ratio is optimal assuming P $\neq$ NP, as InfMax is NP-hard to approximate to within a factor of more than $(1 - 1/e)$ for the IC model (Kempe, Kleinberg, and Tardos 2015), and the special case of the SIR model with $\gamma_u = 1$ for every $u \in V$ is exactly the IC model.

Despite that the greedy algorithm runs in polynomial time and provides a constant theoretical approximation guarantee, greedy algorithms based on the Monte-Carlo are slow in practice. Researchers have then focused on designing algorithms that are fast in practice while keeping the theoretical approximation guarantee (Borgs et al. 2014; Chen, Peng, and Lee 2012; Goyal, Lu, and Lakshmanan 2011; Tang, Shi, and Xiao 2015; Leskovec et al. 2007; Kempe, Kleinberg, and Tardos 2015). One of the most successful types of such algorithms is based on *reverse-reachable sets* (Borgs et al. 2014; Tang, Xiao, and Shi 2014; Tang, Shi, and Xiao 2015), which yields a nearly linear time algorithm with $1 - 1/e$ approximation ratio. Inspired by the success of these reverse-reachable-set-based algorithms for the InfMax problem in IC model, we design an IMM-based efficient algorithm for the InfMax problem in SIR model.

The reverse-reachable-set-based algorithms are based on the observation in Proposition 7. Our objective is to find S that maximizes $\mathbf{Pr}[S \cap \mathrm{RR}_{\mathcal{G}} \neq \varnothing]$. This is done by sampling a sufficient number of copies of $\mathrm{RR}_{\mathcal{G}}$ and finding S that intersects as many copies as possible, which becomes a classical k-max coverage problem. This type of algorithms has been widely used for the IC model and proved successful in terms of scalability. Below, we highlight how we adapt them to the SIR model and its variants.

For the SIR model, one natural idea is to sample each $\mathrm{RR}_{\mathrm{SIR}}$ based on the live-edge graph definition in Section 3.1. However, both $\boldsymbol{R}_v$ and $\boldsymbol{I}_e$ are sequences of infinitely many random variables. To accurately and efficiently sample a reverse reachable set $\mathrm{RR}_{\mathrm{SIR}}$, we instead directly calculate the probability that an edge is live. For example, $\mathbf{Pr}[e \in \mathcal{G}_{\mathrm{SIR}}]$ can be computed by Equation (1), which is a geometric series that admits a compact formula. More generally, for a vertex u with a set E'_u of u's incident edges and a particular incident edge e, the conditional probability $\mathbf{Pr}[e \in \mathcal{G}_{\mathrm{SIR}} \mid E'_u \cap \mathcal{G}_{\mathrm{SIR}} = \varnothing]$ can also be expressed as a geometric series by applying the law of conditional probability. With these, the reverse reachable set $\mathrm{RR}_{\mathrm{SIR}}$ can be sampled in a way similar to the proof of Lemma 10. Details for our algorithms are available in the full version of our paper. As for the time complexity, since (as we have seen) the probability that an edge is live can be computed by the sum of a geometric series, which takes $O(1)$ time, the time complexity for sampling an RR set for the SIR model is the same as it is in the IC model. Therefore, the time complexity of `SIRIMM` is the same as the original IMM algorithm for the IC model.

Theorem 11. *There exists an algorithm that takes as inputs $k, \ell \in \mathbb{Z}^+$ and $\epsilon \in \mathbb{R}^+$, and a directed graph $G = (V, E)$ where $|V| = n, |E| = m$ with the diffusion model $\mathrm{SIR}_{\boldsymbol{\beta},\boldsymbol{\gamma}}$ and outputs a $(1-1/e-\epsilon)$-approximately optimal expected influence spread of k seeds with at least $1-1/n^{\ell}$ probability in $O\left((k+l)\cdot(n+m)\cdot\log n/\varepsilon^2\right)$ time.*

The case with the TSIR model is more involved due to the extra time-dependency. Intuitively, $\mathrm{RR}_{\mathrm{TSIR}}(v)$ is a truncated version of $\mathrm{RR}_{\mathrm{SIR}}(v)$. In addition, we need extra information in addition to the probability $\mathbf{Pr}[e = (u, v) \in \mathcal{G}_{\mathrm{SIR}}]$, as it now matters in which round the infection across (u, v) succeeds. Therefore, we have to apply the original definition in Section 3.1 to sample a reverse reachable set. Due to the time-dependency feature, we need to truncate those vertices in $\mathrm{RR}_{\mathrm{SIR}}(v)$ that are too far away from v. This is done by applying Dijkstra's algorithm to build a reverse shortest path tree rooted at v. Notice that we do not need to consider infinitely many random variables for $\boldsymbol{R}_v$ and $\boldsymbol{I}_e$: vertices that are not recovered or edges that are not active for long periods need not be considered due to the time-dependency of the TSIR model. In our full version of the paper, we formally define the live-edge graph formulation for the TSIR model and present the details for our algorithm. Sampling an RR set in the TSIR model is more time-consuming as we need to flip a coin for each vertex and each edge at most T times and we need to run Dijkstra's algorithm. The time complexity is increased by a factor of $O(T \log n)$.

Theorem 12. *There exists an algorithm that takes as inputs $k, \ell, T \in \mathbb{Z}^+$ and $\epsilon \in \mathbb{R}^+$, and a directed graph $G = (V, E)$ where $|V| = n, |E| = m$ with the diffusion model $\mathrm{TSIR}_{\boldsymbol{\beta},\boldsymbol{\gamma},T}$ and outputs a $(1-1/e-\epsilon)$-approximate optimal expected influence spread of k seeds with at least $1 - 1/n^{\ell}$ probability in $O\left(T\cdot(k+l)\cdot(n+m)\cdot\log^2 n/\varepsilon^2\right)$ time.*

Acknowledgments

The research of Biaoshuai Tao was supported by Key Laboratory of Interdisciplinary Research of Computation and Economics (Shanghai University of Finance and Economics), Ministry of Education, and the National Natural Science Foundation of China (No. 62102252). The research of Kuan Yang was supported by the NSFC grant No. 62102253.

References

Allen, L. J. 1994. Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences*, 124(1): 83–105.

Bass, F. M. 1976. *A New Product Growth Model for Consumer Durables*, 351–353. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-51565-1.

Borgs, C.; Brautbar, M.; Chayes, J.; and Lucier, B. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, 946–957. SIAM.

Brown, J. J.; and Reingen, P. H. 1987. Social Ties and Word-of-Mouth Referral Behavior*. *Journal of Consumer Research*, 14(3): 350–362.

Camarero, C.; and San José, R. 2011. Social and attitudinal determinants of viral marketing dynamics. *Computers in Human Behavior*, 27(6): 2292–2300.

Chen, W. 2018. An issue in the martingale analysis of the influence maximization algorithm imm. In *Computational Data and Social Networks: 7th International Conference, CSoNet 2018, Shanghai, China, December 18–20, 2018, Proceedings 7*, 286–297. Springer.

Chen, W. 2020. Network Diffusion Models and Algorithms for Big Data. *POSTS & TELECOM PRESS, Bei Jing*.

Chen, W.; Castillo, C.; and Lakshmanan, L. V. 2022. *Information and influence propagation in social networks*. Springer Nature.

Chen, W.; Wang, C.; and Wang, Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1029–1038.

Chen, Y.-C.; Peng, W.-C.; and Lee, S.-Y. 2012. Efficient algorithms for influence maximization in social networks. *Knowledge and Information Systems*, 33: 577–601.

Domingos, P.; and Richardson, M. 2001. Mining the Network Value of Customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, 57–66. New York, NY, USA: Association for Computing Machinery. ISBN 158113391X.

Ehrhardt, M.; Gašper, J.; and Kilianová, S. 2019. SIR-based mathematical modeling of infectious diseases with vaccination and waning immunity. *Journal of Computational Science*, 37: 101027.

Goyal, A.; Lu, W.; and Lakshmanan, L. V. 2011. Celf++ optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, 47–48.

Greenwood, P. E.; and Gordillo, L. F. 2009. Stochastic epidemic modeling. *Mathematical and statistical estimation approaches in epidemiology*, 31–52.

Kabir, K. A.; Kuga, K.; and Tanimoto, J. 2019. Analysis of SIR epidemic model with information spreading of awareness. *Chaos, Solitons & Fractals*, 119: 118–125.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2015. Maximizing the Spread of Influence through a Social Network. *Theory Comput.*, 11: 105–147.

Kermack, W. O.; and McKendrick, A. G. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772): 700–721.

Kumar, S.; Ahmadian, A.; Kumar, R.; Kumar, D.; Singh, J.; Baleanu, D.; and Salimi, M. 2020. An efficient numerical method for fractional SIR epidemic model of infectious disease by using Bernstein wavelets. *Mathematics*, 8(4): 558.

Lappas, T.; Liu, K.; and Terzi, E. 2011. A survey of algorithms and systems for expert location in social networks. *Social network data analytics*, 215–241.

Lerman, K.; and Ghosh, R. 2010. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1): 90–97.

Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; and Glance, N. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 420–429.

Li, Y.; Fan, J.; Wang, Y.; and Tan, K.-L. 2018. Influence maximization on social graphs: A survey. *IEEE Trans. on Knowledge and Data Engineering*, 30(10): 1852–1872.

Liu, P.; Li, L.; Wen, Y.; and Fang, S. 2023. Identifying Influential Nodes in Social Networks: Exploiting Self-Voting Mechanism. *Big data*, 11(4): 296—306.

Liu, P.; Qiu, G.; Tao, B.; and Yang, K. 2024. A Thorough Comparison Between Independent Cascade and Susceptible-Infected-Recovered Models. *arXiv preprint arXiv:2408.11470*.

Long, Y.; and Wang, L. 2020. Global dynamics of a delayed two-patch discrete SIR disease model. *Communications in Nonlinear Science and Numerical Simulation*, 83: 105117.

Mahajan, V.; Muller, E.; and Bass, F. M. 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing*, 54(1): 1–26.

Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical programming*, 14: 265–294.

Pastor-Satorras, R.; and Vespignani, A. 2001. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6): 066117.

Richardson, M.; and Domingos, P. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 61–70.

Rigobon, R. 2002. Contagion: How to Measure It? In *Preventing Currency Crises in Emerging Markets*, NBER Chapters, 269–334. National Bureau of Economic Research, Inc.

Sene, N. 2020. SIR epidemic model with Mittag–Leffler fractional derivative. *Chaos, Solitons & Fractals*, 137: 109833.

Tang, Y.; Shi, Y.; and Xiao, X. 2015. Influence Maximization in Near-Linear Time: A Martingale Approach. In *ACM SIGMOD Conference*, 1539–1554.

Tang, Y.; Xiao, X.; and Shi, Y. 2014. Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, 75–86. New York, NY, USA: Association for Computing Machinery. ISBN 9781450323765.

Wu, P.; and Pan, L. 2017. Scalable influence blocking maximization in social networks under competitive independent cascade models. *Computer Networks*, 123: 38–50.

Zhan, X.-X.; Liu, C.; Sun, G.-Q.; and Zhang, Z.-K. 2018. Epidemic dynamics on information-driven adaptive networks. *Chaos, Solitons & Fractals*, 108: 196–204.