Nonprogressive Diffusion on Social Networks: Approximation and Applications

Yunduan Lin
Civil and Environmental Engineering Department, University of California, Berkeley, yunduan_lin@berkeley.edu

Heng Zhang
W. P. Carey School of Business, Arizona State University, hengzhang24@asu.edu

Renyu Zhang
CUHK Business School, The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk

Zuo-Jun Max Shen
Industrial Engineering and Operations Research Department, University of California, Berkeley, maxshen@berkeley.edu

Nonprogressive diffusion describes the spread of behavior on a social network, where agents are allowed to reverse their decisions as time evolves. It has a wide variety of applications in service adoption, opinion formation, epidemiology, etc. To offer an efficient framework for evaluating and optimizing nonprogressive diffusion, we introduce a comprehensive model and a Fixed-Point Approximation (FPA) scheme. This approximation scheme admits both theoretical guarantee and computational efficiency. We establish that the approximation error is inherently related to the network structure, and derive order-optimal bounds for the error using two novel network metrics. We show that the FPA scheme is most accurate for dense and large networks that are generally prohibitive to analyze by simulation. Taking the widely studied influence maximization and optimal pricing problems on a social network as examples, we further illustrate the broad applications of our FPA scheme. Finally, we conduct comprehensive numerical studies with synthetic and real-world networks. In real networks, the FPA scheme shows 70-230 times more speed up in computation time than simulation while achieving a mean absolute percentage error of less than 3.48%. Moreover, our proposed two network metrics are reliable indicators of the FPA scheme's performance.

Key words: Nonprogressive network diffusion, Large-scale network approximation, Network centrality, Influence maximization, Pricing

1. Introduction

Social networks fundamentally shape our lives. People are more receptive to information shared by their friends and relatives (Lu et al. 2013) and more inclined to make a purchase when informed by their acquaintances (Ma et al. 2015, Bapna and Umyarov 2015). It is even more so in the digital era—globally, 4.62 billion people, approximately 58.4% of the population worldwide, used online social network platforms, such as Facebook, YouTube, and Tiktok, by January 2022 (Datareportal 2022). These platforms extend the reach and complexity of our social networks, as both friends and strangers online contribute to shaping our opinions and choices.

Within such networks, each agent both affects and is affected by others, setting the stage for the diffusion of information and behavior. In this work, we use the term diffusion to represent the general phenomenon of information or behavior spread when agents' beliefs or behaviors are influenced by their social connections. Platforms that harness this power of network diffusion can substantially boost their impact and profitability (Shriver et al. 2013). However, understanding diffusion within social networks is a complex undertaking. It involves not just individual behaviors, but also the intricate relationships that bind them. This complexity has made network analysis an enduring subject of study that has engaged generations of researchers (see books Jackson 2010).

Network diffusion analysis spans multiple domains, such as computer science (Kempe et al. 2003, Acemoğlu et al. 2013), economics (Sadler 2020, Acemoglu et al. 2011), operations management (Song and Zipkin 2009, Candogan et al. 2012, Shen et al. 2017, Wang and Wang 2017) and epidemiology (Kermack and McKendrick 1927, Drakopoulos and Zheng 2017). In the seminal paper (Kempe et al. 2003), diffusion processes are broadly categorized into progressive and nonprogressive types. While progressive diffusion deals with unidirectional changes in the state—such as adopting new technology or purchasing a product—our study focuses on *nonprogressive diffusion*, which allows for bidirectional state transitions. This framework is especially relevant in contexts where decisions can be reversed, such as signing up for a membership program, belief propagation influenced by social learning behaviors, being infected in a pandemic, etc.

Analyzing network diffusion primarily adheres to one of two approaches. The first is microfounded by capturing the concrete network topology and the stochastic evolution of agent states. Notable models of this include the independent cascade model (Goldenberg et al. 2001) and the linear threshold model (Granovetter 1978, Schelling 1978). While these models capture fine granularity, detailing the diffusion on an individual basis over time, their intricate nature often leads to computational challenges. In most cases, simulation happens to be the only viable tool to analyze such models, making the optimization, even for a sparse and moderate-sized network, time-consuming (Chen et al. 2009). Conversely, the second approach offers a macroscopic view, simplifying the diffusion process. Some models, (e.g., Bass models; Bass 1969), bypass the intricacies of network topology, focusing on the overall population. Others, like Candogan et al. (2012), Jackson et al. (2020), ignore the stochasticity and focus on the equilibrium outcome. This macro lens, while sacrificing detailed characterization of the diffusion, facilitates efficient analyses and generates sharper insights.

Our work bridges these two approaches in the context of nonprogressive diffusion by providing a simple, efficient, and accurate approximation scheme. We base on a general diffusion model that takes into account heterogeneous agents, local network effects, and network topology. While characterizing the long-run adoption rate for each agent in such a detailed model may seem technically intractable, we investigate a fixed-point approximation (FPA) scheme that estimates the adoption

rates through a set of easily solvable fixed-point equations. Notably, we show that the FPA scheme comes with provable guarantees. Its performance is associated with the network structure and improves for larger and denser networks. We also propose metrics at both node and network levels that can efficiently indicate the FPA's performance for different network structures. Moreover, the FPA scheme further paves the way for optimizing operational decisions, such as the influence maximization and pricing problems in the nonprogressive diffusion context. It allows for straightforward problem formulation and algorithm development, that are not just computationally efficient but also yield near-optimal solutions. In summary, through the FPA scheme, we show that the diffusion outcome characterized by the "micro-model" can be accurately approximated by an easy-to-analyze "macro-model", integrating the advantages of both modeling paradigms.

1.1. Contributions and Organization

Our research contributions are summarized as follows:

- **Provable approximation scheme for a general nonprogressive diffusion model.** We investigate nonprogressive diffusion through a micro-founded, dynamic and stochastic model, which captures local network effects and individual heterogeneity. Our model generalizes the well-known linear threshold (LT) model and adapts it to the nonprogressive diffusion (also see related discussions in the literature review). Within this model, we propose the FPA scheme to approximate the limiting adoption probability of each agent. To validate this approach, we develop a nontrivial "fixed-point sandwich" technique, establishing an order-optimal error bound. This bound indicates its superior performance for large and dense networks, which are otherwise challenging to simulate. These bounds naturally link to novel network structure metrics we propose to gauge the performance of the FPA scheme: the *inverse in-degree centrality* and the *inverse in-degree density*. These metrics provide valuable insights into both node-level and network-wide structures of the network, serving as reliable indicators for the performance of FPA in diverse network configurations. Under mild technical assumptions, we also show that our bound is tight. Our large-scale empirical studies highlight the FPA scheme's superior performance over a wide range of networks. For example, it achieves a mean absolute percentage error of less than 3.48% among all tested real-world networks while concurrently accelerating computation by factors ranging from 70 to 230, compared with traditional simulation methods.
- **Wide applicability in optimizing operational decisions.** The FPA scheme offers a powerful tool to reformulate and solve operational decision-making problems in the nonprogressive diffusion setting. By virtue of our approximation error bound, the reformulated problem will lead to efficient algorithms and provable high-quality decisions for problems that were previously only able to be solved by cumbersome simulation-based algorithms. We take the influence

maximization (IM) and pricing problems on a social network as examples. For the IM problem, we show that under technical conditions, the influence function is submodular with regard to the seed set in the reformulated problem. This extends the greedy algorithm to more general settings, with significant efficiency improvement. For the pricing problem, we can also provide near-optimal algorithms by the FPA reformulation. Specifically, under technical conditions, the pricing problem under perfect price discrimination can be reformulated as a convex program. In more general settings where perfect price discrimination is infeasible, we derive approximate gradient expressions for the direct price optimization problem, with which near-optimal solutions can be achieved efficiently.

The remainder of the paper is structured as follows: In the following of this section, we review the related literature. Section 2 introduces the nonprogressive diffusion model and characterizes the limiting adoption rate. In Section 3, we describe the FPA scheme and demonstrate our main theoretical results. Then in Section 4, we establish the order-optimal error bound followed by extensive numerical experiments in Section 5. We study the IM and pricing problem using our FPA scheme in Section 6. Section 7 concludes this paper. Throughout this paper, we use increasing and decreasing in the non-strict sense.

1.2. Literature Review

Our paper is broadly related to the literature on network diffusion. We first review diffusion models in different settings. Then, we discuss the optimization problems that involve network diffusion.

Diffusion Models. Various models have been proposed across disciplines to characterize diffusion for specific applications. However, a consistent trade-off can be observed: researchers often have to choose between the conciseness of the model and practical efficiency. For instance, the Bass model (Bass 1969) ignores most information on network structure and agents but enjoys the advantages of the analytical expressions on some critical values, allowing for easy optimization (Agrawal et al. 2021, Lin et al. 2021). In contrast, the LT model (Granovetter 1978, Schelling 1978) incorporates the network structure, but is computationally challenging, as evidenced by Chen et al. (2010).

Given our specific focus on nonprogressive diffusion, we will discuss some parallel streams of work related to our study. First, while the LT model is designed for a progressive case, a nonprogressive variant[1] has been introduced by Kempe et al. (2003). This model retains most features of the traditional LT model but selects the random threshold independently at each time period, unlike the fixed random threshold of the original model. Our model builds upon this nonprogressive LT model, introducing agent heterogeneity and accommodating a more arbitrary randomness distribution. Second, our work is related to the social learning literature (Jadbabaie et al. 2012, Chandrasekhar

[1] Hereafter, we will refer to this as the nonprogressive LT model to differentiate it from its progressive counterpart.

et al. 2020, Allon et al. 2019), where agents form beliefs towards a binary signal of the world based on their neighbors' beliefs. While this body of work predominantly examines the learning process and the final network-wide belief distribution, our emphasis is on characterizing individual agent adoption states for an arbitrary network diffusion instance. Third, a variety of engineering and economics applications describe the interactions across the network using network games (e.g., see Ballester et al. 2006, Candogan et al. 2012, Afèche et al. 2023, Baron et al. 2022, Feng et al. 2022). A central goal of this literature is to analyze various types of equilibria. Although our fixed-point approximation is reminiscent of the equilibrium in the network games, our focus diverges in its relation to a concrete micro-founded model. Fourth, a number of operations management studies incorporate network externality into consumer choice models. This type of work, serving for the subsequent assortment or pricing problem, generally simplifies the network structure. For example, some studies consider only global effects by looking at market-wide adoption averages (Du et al. 2016, Wang and Wang 2017), while others restrict their focus to myopic local proxy or specific types of networks (Gopalakrishnan et al. 2022, Xie and Wang 2020). Our work, instead, accounts for full network information and operates under a more general setting. Finally, our work also relates to works studying the mean-field approximation for stochastic processes (Benaïm and Weibull 2003, Van Mieghem et al. 2008). While these works typically offer a deterministic description at the population level, we go further by addressing the operational aspects at the individual level.

Optimization with Network Diffusion. The FPA scheme is applicable in a wide variety of applications. In this paper, we highlight its applications in two examples: influence maximization and pricing within social networks, and provide a concise review of the literature on these themes. Kempe et al. (2003) first consider the issue of choosing an influential set of seed agents to maximize the spread of diffusion influence as a discrete optimization problem. They show that the IM problem, under the LT model, is NP-hard for both progressive and nonprogressive cases. This NP-hard result can also be extended to various other diffusion models. Moreover, to evaluate the total influence under different influential sets, extensive simulations are required so that it is time-consuming to even achieve an approximate solution. We refer readers to the survey (Li et al. 2018) for a comprehensive review of the existing approaches. These approaches compromise either accuracy or efficiency and are not ideal for practical use. However, with the FPA scheme, we can effectively balance both. We also point out that a recent paper (Chan et al. 2020) specifically studies the IM problem with the nonprogressive LT model, closely aligning with our setting. For the pricing problem, there is a growing literature in the economics and operations management communities that considers the presence of network effects (Anari et al. 2010, Hu et al. 2020, Li 2020, Yang and Zhang 2022, Huang et al. 2022). Recent studies on the single-item pricing problem with the network effect can be found in Candogan et al. (2012), Du et al. (2018), and Nosrat et al. (2021). Compared with these three

papers, our framework as well as the proposed algorithms can be used to consider a more general and flexible setting, with theoretical guarantees rooted in our micro-founded model.

2. Nonprogressive Network Diffusion Model

In this section, we first introduce the network diffusion model and then characterize the limiting behavior of each agent within it. This model can be applied to various nonprogressive diffusion settings, among which we use service adoption on an online social network platform for illustration.

2.1. Preliminaries and Formulation

We model the social network platform (e.g., TikTok) as a graph $G = (V, E)$ with n agents, where $V := \{1, 2, ..., |V|\}$ is the set of agents and $E := \{1, 2..., |E|\}$ is the set of directed edges. A directed edge $(i, j) \in E$, where $i, j \in V$, implies that agent j is influenced by agent i, and we call i an *in-neighbor* of j. We interpret $(i, j) \in E$ as j following i on the platform. We use $\mathcal{N}_i$ to denote the set of all in-neighbors for agent i (i.e., $\mathcal{N}_i := \{j \in V : (j, i) \in E\}$) and $d_i := |\mathcal{N}_i|$ to denote the in-degree (i.e., the number of in-neighbors). Throughout, we use agent and node interchangeably.

We use t to denote the discrete time period, starting with $t = 0$ as the service launch time. Define $Y_i(t) \in \{0, 1\}$ as the state of agent i at period t, where $Y_i(t) = 1$ (resp. $Y_i(t) = 0$) means the adoption (resp. nonadoption) of the service in this period. The initial state $\mathbf{Y}(0)$ follows an arbitrary distribution on $\{0, 1\}^n$. For all $t \geq 1$, each agent's decision to adopt is governed by their realized utility $u_i(t)$ during that period, as given by

$$u_i(t) := v_i + \beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t). \tag{1}$$

Without loss of generality, we normalize the utility of nonadoption to 0, and thus $Y_i(t) = \mathbb{1}\{u_i(t) \geq 0\}$. As shown by (1), $u_i(t)$ consists of three parts: the idiosyncratic intrinsic value v_i, *the local network effect* $\beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i}$ and the random noise $\epsilon_i(t)$. The value v_i reveals the personalized preference and remains constant over time. From an analytical point of view, v_i can be estimated from agent features such as demographic information and behavioral data with the support of big data. It may also be affected by the platform's operational strategies. For example, the price of a paid service (e.g., YouTube Premium) will definitely affect whether and how, the agent likes it. The local network effect captures the peer influence on the agent from in-neighbors, with parameter β to quantify the network effect intensity. If agent i has no in-neighbors (i.e., $\mathcal{N}_i = \emptyset$), we set this term to 0. Finally, we assume the random noise $\epsilon_i(t)$ is independent and identically distributed (i.i.d.) across agents and time. We assume, without loss of generality, that $\mathbb{E}[\epsilon_i(t)] = 0$. For now, we impose no further constraints on its distribution, except for the following mild condition.

Assumption 1 **(Lipschitz Continuity)**. *The random noise $\epsilon_i(t)$ has an L-Lipschitz continuous cumulative distribution function (CDF): $|F_\epsilon(x) - F_\epsilon(y)| \leq L|x - y|$ for any $x, y \in \mathbb{R}$.*

We require that the noise distribution is sufficiently smooth. Assumption 1 is satisfied by any continuous distribution with a bounded probability density function (PDF), making common distributions like the uniform, logistic, or normal distribution compatible with our model. We also impose a bound for the network effect intensity β that facilitates characterizing the limit of diffusion.

ASSUMPTION 2 **(Bounded Network Effects)**. *The network effect satisfies* $|\beta| < 1/L$.

Parameter β quantifies the magnitude of network externality. Similar assumptions are commonly made in the network economics literature (e.g., see Horst and Scheinkman 2006, Wang and Wang 2017, Xu 2018, Jackson et al. 2020, Gopalakrishnan et al. 2022). In these settings, such assumptions are often introduced to ensure that the equilibrium of a network game uniquely exists. However, our model assigns additional significance to Assumption 2. It not only excludes divergent or periodic behavior in the long run in our diffusion model (Proposition 1) but also guarantees a valid fixed-point characterization of the limiting adoption probabilities (Proposition 2). In Section 5.1, we also extend our discussion through extensive numerical experiments to investigate the implications of Assumption 2 being violated. For the remainder of this analysis, we assume that the network effects are positive, operating under the assumption that $0 < \beta < \frac{1}{L}$. However, it should be noted that our results can be generalized to scenarios where $-\frac{1}{L} < \beta < 0$, thereby covering situations of negative network effects as well.

A natural goal in such a setting is to quantify the total diffusion in the network. In line with prior studies (Kempe et al. 2003), we focus on the limiting adoption probability. Provided it converges, it also represents accumulated reward (frequency of adoption) in the long run.

We remark on the notations. Hereafter, we use a bold math notation to denote the collection of a particular variable over all agents in vector form. It is important to note that the network structure and intrinsic values together identify a specific diffusion case. Meanwhile, the noise distribution and the network effect intensity make up the diffusion environment. Accordingly, a specific diffusion instance is represented by a quadruple $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$. Sequences of such instances are likewise represented by a series of these quadruples.

2.2. A Markov Chain Perspective

Notably, each diffusion instance can be characterized by a Markov chain (MC), of which the state space is the set of indicator vectors denoting all possible combinations of adoption decisions, represented by $\{0,1\}^n$. The transition probability from state $\mathbf{y}$ to $\mathbf{y}'$ can be computed as

$$P(\mathbf{y},\mathbf{y}') = \prod_{i\in V} \mathbb{P}(Y_i(t)=y_i'|\mathbf{Y}(t-1)=\mathbf{y}) = \prod_{i\in V} F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}\right)^{1-y_i'} \cdot \left[1-F_\epsilon\left(-v_i-\beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{d_i}\right)\right]^{y_i'}.$$

As our primary interest lies not in the individual MC states, but rather in the overall adoption probability for each agent. To that end, we define the adoption probability of agent i at time t as

$$q_i(t) \;:=\; \mathbb{P}(Y_i(t)=1) \;\equiv \sum_{\mathbf{y}\in\{0,1\}^n} \mathbb{1}\{y_i=1\}\cdot\mathbb{P}(\mathbf{Y}(t)=\mathbf{y}). \tag{2}$$

We have the following proposition on the limiting behavior of $\mathbf{q}(t)$ when t tends to infinity.

PROPOSITION 1 **(Limiting Adoption Probability)**. *Under Assumptions 1 and 2, for any initial state* $\mathbf{Y}(0)$*, the adoption probability of each agent* i *converges to*

$$\lim_{t\to\infty} q_i(t) = q_i^* := \sum_{\mathbf{y}\in\{0,1\}^n} \mathbb{1}\{y_i=1\}\cdot\pi(\mathbf{y}),$$

when t *increases, where* $\boldsymbol{\pi}$ *is the stationary distribution of the MC that satisfies* $\boldsymbol{\pi}=\boldsymbol{\pi}P$*.*

As made clear in the proof of Proposition 1, with assumptions on random noise and network effect, the MC only has a *single aperiodic recurrent class*. Thus, a limiting distribution $\boldsymbol{\pi}$ leads to the limiting adoption probabilities $\mathbf{q}^*$. By the standard MC theory, one can easily verify that

$$\lim_{t\to\infty}\frac{1}{t}\cdot\sum_{s=1}^{t} Y_i(s) = q_i^* \quad \text{a.s.} \quad \text{and} \quad \lim_{t\to\infty}\frac{1}{t}\cdot\sum_{s=1}^{t} q_i(s) = q_i^*, \quad \forall i\in V, \tag{3}$$

for any initial state $\mathbf{Y}(0)$. As a result, this enables us to leverage the limiting adoption probability $\mathbf{q}^*$ when formulating optimization problems related to nonprogressive diffusion. Specifically, various operational problems, such as the influence maximization problem (Section 6.1) and the pricing problem on a social network (Section 6.2), can be generally framed as:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \qquad g\Big(\mathbf{q}^*(G,\mathbf{v}(\mathbf{x}),F_\epsilon(\cdot),\beta),\mathbf{x}\Big). \tag{4}$$

Here, $\mathbf{x}$ represents platform decisions with $\mathcal{X}$ denoting the feasible set of such decisions. For simplicity, we only consider decisions influence diffusion outcomes by altering the intrinsic values. with a slight abuse of notation, $\mathbf{v}(\cdot)$ represents intrinsic values as a function of platform decisions, and $\mathbf{q}^*(\cdot)$ denotes the mapping from a diffusion instance to the limiting probability vector. Finally, $g(\cdot,\cdot)$ is the objective function that depends on the diffusion outcome $\mathbf{q}^*$ and decision variables $\mathbf{x}$. For example, the influence maximization problem can be formulated as (4) with the decision vector $\mathbf{x}$ defined as setting the intrinsic utility of a set of seed users to sufficiently high levels; and the objective $g(\cdot,\cdot)$ defined as the limiting total expected adoptions $\sum_{i\in V} q_i^*$. For the optimal pricing problem, the decision vector $\mathbf{x}$ is the price vector that affects the intrinsic value of each agent, and the objective $g(\cdot,\cdot)$ is the expected profit under the limiting adoption probability, i.e., $\sum_{i\in V} q_i^* x_i$. The specific formulations of these problems will be presented in Section 6.

Solving the optimization problem (4) is challenging, due to the absence of closed-form expressions for $\mathbf{q}^*$ and the exponential growth in the MC state space. It is intractable to construct the transition matrix even for a moderate-sized network, let alone to calculate $\mathbf{q}^*$. Therefore, problem (4) is generally intractable either analytically or computationally, which motivates us to develop our approximation scheme for $\mathbf{q}^*$ presented in Section 3.

Before presenting our approximation scheme, it is worth situating our diffusion model within the broader network diffusion literature. Our model is most closely related to the LT model (Granovetter 1978), which is one of the most widely studied diffusion models. We extend the LT model for nonprogressive settings (see Kempe et al. 2003), by introducing the heterogeneity of agents' intrinsic values and incorporating different random noise distributions. One advantage of the LT model is its ability to closely represent rational decision-making by agents, thereby characterizing the evolution of diffusion processes. While the LT model offers a solid micro-foundation for diffusion, it comes with the drawback of computational intractability. Our subsequent approximation technique offers a practical solution to this limitation. Our model also bears similarities to the game-theoretic discrete choice models with network effects (Du et al. 2016, Wang and Wang 2017). There are two noteworthy distinctions. First, our model accommodates a broader range of choice models than what the standard logistic distribution covers. Second, we include the stochasticity of network effects by assuming that the network effects come from realized average adoptions, as opposed to the more commonly assumed expected adoption rates. Interestingly, our result shows that, in the long term, the expected adoption rates that arise naturally from a game-theoretic perspective are also well-justified in a dynamic and micro-founded setting. This convergence validates the use of expected adoption rates, effectively bridging the LT and game-theoretic choice models.

3. Fixed-Point Approximation (FPA)

In this section, we introduce the FPA scheme. We present a comprehensive analysis of the scheme's performance, offering theoretical upper bounds for approximation error. Furthermore, we put forth two critical metrics designed to evaluate the difficulty of approximation and the performance of the FPA scheme; see (6). In this section, we assume that Assumptions 1 and 2 hold.

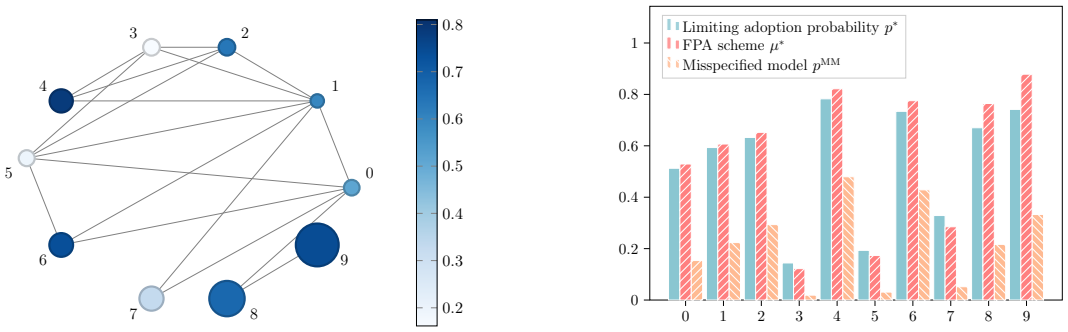
3.1. Overview and Motivating Example

For a given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, we will show that the limiting adoption probability $\mathbf{q}^*$ can be reasonably approximated by the solution $\boldsymbol{\mu}^*$ of the following simple system of equations:

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i}\right), \text{ for all } i \in V. \tag{5}$$

We begin with a motivating example to demonstrate the values of $\mathbf{q}^*$ and $\boldsymbol{\mu}^*$. This particular instance features heterogeneity in both network connectivity and intrinsic values among agents. To

facilitate an intuitive understanding of the impact of network effects, we also introduce a misspecified model as a benchmark scenario. In this misspecified model, the adoption probability for each agent i, q_i^{MM}, is calculated as $\mathbb{E}[\mathbb{1}\{v_i+\epsilon_i \geq 0\}]$ so the network effects are ignored. For detailed information about this example instance, including numerical results, please refer to Appendix C.1.



(a) Approximation results on the network structure

(b) Adoption probability values

Figure 1 **The 10-node example to illustrate the FPA scheme. In the left subfigure, the color denotes the true value of $\mathbf{q}^*$ and the size denotes the absolute error between $\mathbf{q}^*$ and $\boldsymbol{\mu}^*$.**

Figure 1a presents the network structure and the approximation results. Clearly, nodes with fewer neighbors exhibit higher errors, whereas well-connected nodes yield smaller errors. Further insights can be gained from Figure 1b, which enumerates the values of $\mathbf{q}^*$, $\boldsymbol{\mu}^*$ and $\mathbf{q}^{\mathrm{MM}}$. The strong impact of network effects is underscored by the large discrepancy between $\mathbf{q}^*$ and $\mathbf{q}^{\mathrm{MM}}$. Against $\mathbf{q}^*$ as a baseline, the mean absolute error values for $\boldsymbol{\mu}^*$ and $\mathbf{q}^{\mathrm{MM}}$ are 0.045 and 0.310, respectively. These observations confirm the high quality of the FPA solution and suggest that the approximation tends to be more accurate for agents with central positions in the network.

In the following, we quantitatively measure the deviation between $\mathbf{q}^*$ and $\boldsymbol{\mu}^*$, with a particular focus on its dependence on network structures. The key technical challenge arises from the fact that adoption decisions exhibit temporal and spatial correlations, compounded by the non-linearity introduced by a general distribution of noise. To overcome this challenge, we employ a novel "fixed-point sandwich" technique to provide the theoretical guarantee for general diffusion instances.

3.2. The Approximation Error

We first remark on notations before formally presenting our analysis on error bound. Given an network $G=(V,E)$ with n nodes, we define the matrix $\tilde{\mathbf{A}} \in \mathbb{R}^{n\times n}$ such that $\tilde{A}_{ij}=\frac{1}{d_i}$ if an edge is directed from j to i, and $\tilde{A}_{ij}=0$ otherwise. This matrix can be viewed as a transformation of the network's adjacency matrix $\mathbf{A}$, where $A_{ij}=1$ if there is an edge directing from i to j and $A_{ij}=0$ otherwise. Specifically, one obtains $\tilde{\mathbf{A}}$ by scaling row i of $\mathbf{A}^\top$ by $\frac{1}{d_i}$. It is worth noting that $\tilde{\mathbf{A}}$ is a row scholastic matrix; that is, $\tilde{\mathbf{A}}\mathbf{e}=\mathbf{e}$ where $\mathbf{e}$ is a vector of ones. Additionally, we introduce the vector

$\mathbf{b} = \left(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_n}\right)^\top$, consisting of the reciprocal of each node's in-degree. Lastly, we define $\rho := L\beta$ as the discount parameter that characterizes the diffusion environment ($\rho < 1$ by Assumption 2). Let $d_{\min}$ be the minimum in-degree of the network with $d_{\min} > 0$.

We introduce two centrality metrics critical in our analysis, which we term as the *inverse in-degree centrality* $\mathcal{C}(G;\rho)$ and the *inverse in-degree density* $\mathcal{D}(G)$. They are defined as follows:

$$\mathcal{C}(G;\rho) := (1-\rho)\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b} = (1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} \quad \text{and} \quad \mathcal{D}(G) := \frac{\mathbf{e}^\top \mathbf{b}}{n}. \tag{6}$$

The *inverse in-degree centrality* $\mathcal{C}(G;\rho)$ is an n-dimensional vector that captures the centrality of each agent, with its i-th entry denoted by $\mathcal{C}_i(G;\rho)$. It bears similarities to the classical Bonacich centrality (Bonacich 1987), taking the form of summed discounted matrix powers. On the other hand, it introduces distinct features regarding the inverse in-degree properties, making it specifically tailored to evaluate the FPA scheme. Importantly, this centrality measure is well-defined given that $\tilde{\mathbf{A}}$ is row stochastic and $\rho < 1$ by Assumption 2. The *inverse in-degree density* $\mathcal{D}(G)$ is a scalar that represents the average inverse in-degree of all agents in the network, serving as an aggregate measure of the FPA scheme's performance. Together, these two centrality metrics offer comprehensive evaluations, affording both individualized and holistic views of the network. Such a dual perspective not only enhances our understanding of the FPA scheme but also provides actionable insights into its application across different network configurations.

For clarity in our subsequent theoretical analysis, we adopt the notations $|\cdot|$ and $(\cdot)^{\frac{1}{2}}$ to represent entry-wise operations on vectors. For instance, for vector $\mathbf{q}$, we define $|\mathbf{q}| := (|q_1|, |q_2|, \ldots, |q_n|)^\top$ and $\mathbf{q}^{\frac{1}{2}} := (\sqrt{q_1}, \sqrt{q_2}, \ldots, \sqrt{q_n})^\top$. We also define the constant $C_\rho := \frac{\rho}{(1-\rho)\sqrt{(1-\rho/2)}}$. We now present our technical key result. In Theorem 1, our primary focus is to characterize the entry-wise error between the limiting probability $\mathbf{q}^*$ and our FPA solution $\boldsymbol{\mu}^*$.

THEOREM 1 **(Entry-wise Error Bound of the FPA Scheme)**. *Under Assumptions 1 and 2, for any diffusion instance* $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$*, the absolute difference between the limiting adoption probability* $\mathbf{q}^*$ *and the fixed-point solution* $\boldsymbol{\mu}^*$ *can be upper bounded by*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\le\; C_\rho \cdot \left[\mathcal{C}(G;\rho)\right]^{\frac{1}{2}}.$$

Theorem 1 shows that the entry-wise error of FPA is closely related to the network structure through the inverse in-degree centrality $\mathcal{C}(G;\rho) := (1-\rho)\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b} = (1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}$. Note that $\sum_{\ell=0}^{\infty} \rho^\ell = 1/(1-\rho)$ and $\tilde{\mathbf{A}}$ is a row stochastic matrix, so each component of $\mathcal{C}(G;\rho)$ is always within the range $[0,1]$. Similar to the Bonacich centrality, $\mathcal{C}(G;\rho)$ can also be interpreted as the weighted sum of different entries of $\mathbf{b}$, the vector of inverse in-degrees. These

weights capture the connectedness between node pairs. To illustrate, consider $\mathcal{P}_{(j,i)}$ as the set of directed paths from agent j to agent i. The weight allocated to the inverse in-degree b_j in the i-th term of $\mathcal{C}(G,\rho)$ is $(1-\rho)\sum_{P\in\mathcal{P}_{(j,i)}}\rho^{|P|}\prod_{k\in P}\frac{1}{d_k}$. Remarkably, this weight decays exponentially fast with respect to the length of the path that connects nodes j and i. As a consequence, the inverse in-degree centrality of each node is predominantly affected by the inverse in-degrees of its nearby nodes. This result rationalizes the observations in Figure 1.

To further understand the intuition of $\mathcal{C}(G,\rho)$, we let $d_{\min,i}(\ell)$ be the minimum in-degree of any node j that is connected to node i via a path of length ℓ. It is straightforward that $d_{\min,i}(\ell)\geq d_{\min}$. By Theorem 1, it then holds that

$$\mathcal{C}_i(G;\rho)\leq(1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)}\ \leq\ \frac{1}{d_{\min}},\ \text{for all } i\in V, \tag{7}$$

where the first inequality holds because $\tilde{\mathbf{A}}$ is a row stochastic matrix. Consequently, we arrive at:

$$|q_i^*-\mu_i^*|\ \leq\ C_\rho\cdot\sqrt{(1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)}}, \tag{8}$$

which suggests that the error is small for nodes with large in-degrees and only distantly connected to nodes with low in-degrees. A subsequent corollary is then derived directly from (7).

Corollary 1 **(ℓ_∞-Norm Bound)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G,\mathbf{v},F_\epsilon(\cdot),\beta\big)$*, the* ℓ_∞*-norm of the difference between* $\mathbf{q}^*$ *and* $\boldsymbol{\mu}^*$ *can be upper bounded by*

$$\|\mathbf{q}^*-\boldsymbol{\mu}^*\|_\infty\ \leq\ C_\rho\cdot\sqrt{\frac{1}{d_{\min}}}. \tag{9}$$

Corollary 1 removes the dependence on the specific network structure from the bound to highlight a worst-case convergence rate as the network expands. Specifically, for a sequence of diffusion instances characterized by an increasing minimum in-degree $d_{\min}$, the maximal deviation shrinks at a rate of $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. As $d_{\min}$ approaches infinity, $\boldsymbol{\mu}^*$ is asymptotically equal to $\mathbf{q}^*$. This simplified bound clearly indicates that the FPA scheme can perform better in larger and denser networks. We remark that although (8) and (9) are both intuitively appealing, these bounds are looser than the one based on the inverse in-degree centrality presented in Theorem 1.

The ℓ_∞-norm, while easy to understand, has its limitations. It relies entirely on the minimal in-degree of the network, rendering it overly conservative and vulnerable to the isolated outliers. For most real-world networks, the minimal in-degree is often quite small even if its size n is large, limiting the applicability of this bound. Corollary 2 below counters this limitation by introducing an upper bound characterized by the scaled ℓ_1-norm. This bound takes into account the inverse in-degree density across the entire network $\mathcal{D}(G)$, providing a more reliable and effective metric

compared to the minimal in-degree $d_{\min}$. Define $r(G) := \max_{i\in V}(\sum_{j=1}^n A_{ij}/\sum_{j=1}^n A_{ji})$ as the largest out-degree to in-degree ratio among all nodes in the network. It always holds that $r(G) \geq 1$, with a smaller value indicating a more evenly distributed degrees in the network.

COROLLARY 2 **(Scaled ℓ_1-Norm Bound)**. *Under Assumptions 1 and 2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$*, the scaled* ℓ_1*-norm of the difference can be upper bounded by*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\leq\; C_\rho \cdot \sqrt{\frac{\|\mathcal{C}(G,\rho)\|_1}{n}}. \tag{10}$$

If $r(G) < 1/\rho$*, the bound can be further simplified as*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\leq\; C_\rho \cdot \sqrt{\frac{1-\rho}{1-\rho r(G)} \cdot \mathcal{D}(G)}. \tag{11}$$

In light of (7), the worst-case network-structure-free bound on $\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1$ is also of order $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. However, Corollary 2 provides more meaningful bounds. Particularly, (10) bounds the scaled ℓ_1-norm of the error by that of the inverse in-degree centrality $\mathcal{C}(G,\rho)$. In poorly conditioned networks with a large $r(G)$, large weights may apply to nodes with a small in-degree, pushing this bound closer to the worst-case scenario of $\mathcal{O}\left(\sqrt{1/d_{\min}}\right)$. Yet, for most networks, we believe that (10) gives an accurate characterization of the ℓ_1-norm that takes into account the whole network structure. When the network is appropriately conditioned (i.e., $r(G) < 1/\rho$), we obtain a more transparent bound in (11) characterized by $\mathcal{D}(G)$, which is an aggregate measure of the network structure. Unlike the minimal in-degree which offers a perspective of extreme nodes, $\mathcal{D}(G)$ provides a holistic view. It emphasizes the average inverse in-degree, thereby giving an encompassing depiction of network connectivity and underscoring the relationship between network structure and the FPA error. Moreover, the inverse in-degree density $\mathcal{D}(G)$ is also computationally more efficient than the aforementioned inverse in-degree centrality $\mathcal{C}(G,\rho)$, which requires inverting an n-by-n matrix. As a consequence, $\mathcal{D}(G)$ is a more practical performance indicator for the FPA scheme across different networks.

Another key observation pertains to the largest out-to-in-degree ratio $r(G)$. Interestingly, the derived upper bound is tighter as $r(G)$ decreases, indicating a better performance of the FPA scheme for more balanced networks. We also highlight that the assumption $r(G) < 1/\rho$ for the second part of Corollary 2 is not restrictive in general. Notably, all undirected graphs and balanced directed graphs satisfy that $r(G) = 1 < 1/\rho$. Studies such as Mislove (2009) also validate the balanced nature of social networks in practice. In particular, active agents (i.e., those who create many links) also tend to be popular (i.e., they are the target of many links). This high correlation is generally attributed to the prevalence of reciprocal links in social networks.

We remark on two facts. First, all the aforementioned bounds apply to networks where $d_{\min} > 0$. For any standalone node i with no in-neighbors, the local network effect term in (1) is zero. Thus, for these nodes, $\mu_i^* = q_i^*$ holds trivially. Therefore, the inclusion of such nodes would only tighten the derived bounds. Further discussions of this aspect are deferred to the numerical experiments in Section 5. Second, it is easy to show that the constant C_ρ is strictly decreasing in ρ and converges to zero as ρ converges to zero. Therefore, our bounds suggest that FPA works better when ρ, which represents the compound effect of network externality and noise magnitude, is small.

Operationalizing the FPA solution. The significance of the FPA solution $\boldsymbol{\mu}^*$ lies in the fact that it allows us to reformulate and simplify problem (4). Instead of solving (4) directly, we can replace $\mathbf{q}^*$ with $\boldsymbol{\mu}^*$ and approximate (4) as follows:

$$\underset{\mathbf{x}\in\mathcal{X},\boldsymbol{\mu}}{\text{maximize}} \qquad g(\boldsymbol{\mu},\mathbf{x}) \quad \text{s.t. } \boldsymbol{\mu} = \mathbf{h}\Big(\boldsymbol{\mu}\,;\, G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta\Big), \tag{12}$$

where $\mathbf{h}(\cdot\,; G, \mathbf{v}, F_\epsilon(\cdot), \beta)$ is the adoption evolution operator (AEO) induced by the diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, which we will formally define using (13) in Section 3.3. The approximate problem (12) offers an explicit formulation by incorporating the FPA scheme as a constraint. This stands in contrast to the implicit variable $\mathbf{q}^*$ in problem (4), which emerges from a complex stochastic process. The formulation (12) thus simplifies the optimization problem and facilitates its solution.

We advocate for the use of this approximate optimization problem (12). The benefits are as follows: (i) *Theoretical Guarantee.* The FPA scheme is especially appealing due to its superb performance, particularly for large and dense networks. Stronger results can be obtained for specific network structures, such as regular networks and random graphs (see Sections 4.2 and 5.2). From a practical standpoint, many real-world networks are large and continuously expanding, making the FPA scheme a promising tool (see Section 5.3). (ii) *Computational Efficiency.* The FPA scheme offers significant computational advantages over traditional Markov Chain Monte Carlo (MCMC) simulations. According to Rheinboldt (1998), a fixed-point iteration converges to the FPA solution in linear time. In contrast, MCMC simulations require considerable computational resources and become more cumbersome as network sizes grow. (iii) *Insights on Network Structure.* Our proposed metrics, namely inverse in-degree centrality and inverse in-degree density, serve as accurate indicators of the FPA scheme's performance. These metrics offer actionable insights on whether to employ the FPA scheme, depending on specific accuracy and efficiency goals. (iv) *Closed-Form Expression.* The approximate problem (12) is considerably more tractable than (4), which paves the way for developing more efficient algorithms tailored for specific problems (see Section 6).

3.3. Proof Sketch of Theorem 1

In this section, we sketch the proof of Theorem 1, which is our main methodological contribution. The key idea is to construct an approximate process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ for a given instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$. We show that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ closely aligns with the adoption probability process $\{\mathbf{q}(t)\}_{t=0}^{\infty}$. As a result, $\mathbf{q}^*$, as the limit of $\mathbf{q}(t)$, is expected to be closely approximated by $\boldsymbol{\mu}^*$, the limit of $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$.

Specifically, we define $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ as a deterministic dynamic system throughout the time horizon:

$$\mu_i(t) = \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j(t-1)}{d_i}\right) & t>0 \end{cases}, \quad \text{for all } i \in V. \tag{13}$$

By Proposition 1, $\mathbf{q}^*$ has a unique value regardless of the initial distribution of the adoption state $\mathbf{Y}(0)$. Without loss of generality, we assume $Y_i(0) = 0$ for all $i \in V$. Indeed, by Proposition 1, the limiting distribution $\mathbf{q}^*$ is irrelevant to the starting adoption status so any bound on $\mathbf{q}^* - \boldsymbol{\mu}^*$ derived under $\mathbf{Y}(0) = 0$ applies to an arbitrary initial distribution of $\mathbf{Y}(0)$. We introduce $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$ as the mapping function that allows us to express $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ in the form $\boldsymbol{\mu}(t) = \mathbf{h}\big(\boldsymbol{\mu}(t-1)\big)$ for $t \geq 1$. We refer to $\mathbf{h}(\cdot)$ as the adoption evolution operator (AEO) and define a family of auxiliary AEOs $\mathcal{H} := \big\{\mathbf{h}_{\boldsymbol{\zeta}}(\cdot) = \mathbf{h}(\cdot) + \boldsymbol{\zeta} : \boldsymbol{\zeta} \in \mathbb{R}^n\big\}$. We proceed by discussing the properties of any AEO $\mathbf{h} \in \mathcal{H}$ and its role in shaping the approximate diffusion process $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$.

PROPOSITION 2 **(Partial Order Preserving, Existence, and Uniqueness)**. *Any AEO* $\mathbf{h} \in \mathcal{H}$ *satisfies the following properties* (*i*) *and* (*ii*)*, and the induced dynamic system* $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ *defined by fixed-point iteration* $\boldsymbol{\nu}(t) = \mathbf{h}\big(\boldsymbol{\nu}(t-1)\big)$ *satisfies the following property* (*iii*)*:*

(*i*) $\mathbf{h}(\mathbf{a}) \leq \mathbf{h}(\mathbf{b})$ *if* $\mathbf{a} \leq \mathbf{b}$.

(*ii*) *There exists a unique fixed-point solution* $\boldsymbol{\nu}^* \in \mathbb{R}^n$ *with* $\mathbf{h}(\boldsymbol{\nu}^*) = \boldsymbol{\nu}^*$.

(*iii*) *For any initial state* $\boldsymbol{\nu}(0)$*, the dynamic system* $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ *satisfies* $\lim_{t\to\infty} \boldsymbol{\nu}(t) = \boldsymbol{\nu}^*$.

The proof of Proposition 2(i) follows from the definition of $\mathbf{h}(\cdot)$. Proposition 2(ii) and (iii) are consequences of the fact that $\mathbf{h}(\cdot)$ is a contraction mapping. Note that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ is the special case of the induced dynamic system $\{\boldsymbol{\nu}(t)\}_{t=0}^{\infty}$ when $\boldsymbol{\nu}(0) = \mathbf{q}(0)$. Given these properties of $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$, for any diffusion instances under Assumptions 1 and 2, we can always find the well-defined FPA solution $\boldsymbol{\mu}^*$ for limiting adoption probability $\mathbf{q}^*$ by solving the system of equations $\mathbf{h}(\boldsymbol{\mu}^*) = \boldsymbol{\mu}^*$.

To show that $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ closely approximates $\{\mathbf{q}(t)\}_{t=0}^{\infty}$, we face two challenges. The first stems from the temporal and spatial dependencies among (un)adoptions. Specifically, an agent's adoption utility is directly shaped by the behavior of their immediate in-neighbors. Over time, these localized correlations not only accumulate but also spread across the network. The second challenge arises from the non-linearity of the CDF F_ϵ of a general distribution. This non-linearity makes it difficult to analytically track the transition of adoption states over time and particularly complicates the

characterization of adoption correlations. To address these challenges, our subsequent efforts focus on bounding the spatio-temporal variances and the nonlinear dynamics in a sequential manner.

In our first analytical phase, we focus on bounding the variance of agents' adoptions, particularly addressing the local network effect term presented in (1). This term is essentially an average over a set of mutually dependent random variables. To quantify its variability, we introduce $\kappa_i(t) := \mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t)\right)$ as the variance of the in-neighbor adoption fraction for agent i. Lemma 1 provides an upper bound of this variance for each period over the entire time horizon.

LEMMA 1 **(In-Neighbor Variance Bound)**. *Under Assumptions 1 and 2, for any diffusion instance* $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ *and* $t \geq 0$*, the in-neighbor adoption variance can be upper bounded by*

$$\boldsymbol{\kappa}(t) \;\leq\; \frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{t-1}\left(\frac{\rho^2}{2}\right)^{\tau}\tilde{\mathbf{A}}^{\tau}\right]\mathbf{b}.$$

Since $\tilde{\mathbf{A}}$ is row stochastic, Lemma 1 bounds $\boldsymbol{\kappa}$ by (approximately) the weighted sum of inverse in-degrees $\mathbf{b}$. This essentially implies that as the number of in-neighbors increases, the variance decreases. In other words, having more in-neighbors reduces the impact of any single neighbor, thereby reducing the mutual dependence among the adoptions of different agents. As time progresses, this upper bound gradually increases, which can be interpreted as a discounted contribution from the neighbors that are connected through a path of length $t-1$. As we can see, the network structure plays a critical role here. The expression $\mathbf{I} + \sum_{\tau=1}^{t-1}\left(\frac{\rho^2}{2}\right)^{\tau}\tilde{\mathbf{A}}^{\tau}$ offers insights into how stochasticity is disseminated, including both spatial and temporal aspects. We can then connect this bound to the centrality measure that resembles our inverse in-degree centrality as follows:

$$\boldsymbol{\kappa}(t) \;\leq\; \lim_{t\to\infty}\frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{t-1}\left(\frac{\rho^2}{2}\right)^{\tau}\tilde{\mathbf{A}}^{\tau}\right]\mathbf{b} \;=\; \frac{1}{4}\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}, \text{ for all } t \geq 0. \tag{14}$$

With the variance bound (14), we move to our second analytical phase of bounding the nonlinear dynamics. Although the adoption probability process $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ lacks a closed-form expression, we expect that its transition between consecutive time periods akin to the AEO $\mathbf{h}(\cdot)$.

LEMMA 2 **(Fixed-Point Deviation of Adoption Probability)**. *Under Assumptions 1 and 2, for any diffusion instance* $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$ *and* $t \geq 1$*, we have*

$$\left|\mathbf{h}\big(\mathbf{q}(t-1)\big) - \mathbf{q}(t)\right| \;\leq\; \boldsymbol{\delta},$$

where $\boldsymbol{\delta} = \left[\left(\frac{\rho}{2}\right)^2\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]^{\frac{1}{2}}$.

Building on Lemma 1, Lemma 2 establishes a connection between the transitions of $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$, providing a one-step guarantee for their similarity. Based on this fact, we use a

"fixed-point sandwich" technique to prove the final results in Theorem 1. Specifically, we define a lower bound system $\{\underline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ and an upper bound system $\{\overline{\boldsymbol{\mu}}(t)\}_{t=0}^{\infty}$ as follows:

$$\underline{\mu}_i(t) = \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \underline{\mu}_j(t-1)}{d_i}\right) - \delta_i & t>0 \end{cases}, \quad \text{for all } i \in V,$$

$$\overline{\mu}_i(t) = \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \overline{\mu}_j(t-1)}{d_i}\right) + \delta_i & t>0 \end{cases}, \quad \text{for all } i \in V.$$

Employing our auxiliary AEOs, these two systems can be expressed as two fixed-point iterations: $\underline{\boldsymbol{\mu}}(t) = \mathbf{h}_{-\boldsymbol{\delta}}(\underline{\boldsymbol{\mu}}(t-1))$ and $\overline{\boldsymbol{\mu}}(t) = \mathbf{h}_{\boldsymbol{\delta}}(\overline{\boldsymbol{\mu}}(t-1))$, with $\underline{\boldsymbol{\mu}}^*$ and $\overline{\boldsymbol{\mu}}^*$ being the fixed-point solutions to these two systems, respectively. The proof of Theorem 1 uses these two fixed-point iterations to sandwich both $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$. When t goes to infinity, $|\mathbf{q}^* - \boldsymbol{\mu}^*|$ can be bounded by $\left|\overline{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^*\right|$, which we can quantify by the lemmas above. This allows us to show that the entire trajectories of $\{\mathbf{q}(t)\}_{t=0}^{\infty}$ and $\{\boldsymbol{\mu}(t)\}_{t=0}^{\infty}$ are close to each other. Therefore, the FPA solution $\boldsymbol{\mu}^*$ serves as a good approximation for the limiting adoption probability $\mathbf{q}^*$.

4. Improved Error Bounds for the FPA Scheme

In this section, we delve deeper into the error bound of the FPA scheme. We demonstrate that, by introducing a mild additional assumption on the noise distribution $F_\epsilon(\cdot)$, a tighter upper bound can be derived. Moreover, we also present a matching lower bound of the same order with this refined upper bound, closing the gap in our analysis. For the subsequent analysis in this section, we will proceed under Assumption 3.

ASSUMPTION 3 **(Smoothness Condition)**. *The random noise $\epsilon_i(t)$ has a differentiable probability density function (PDF) $f_\epsilon(\cdot)$ with its derivative upper bounded by $|f'_\epsilon(\cdot)| \le L_f$.*

This assumption mainly requires the smoothness of the PDF f_ϵ. It is worth noting that this assumption is fairly mild, given that many commonly used distributions inherently exhibit high degrees of differentiability, including but not limited to the normal and logistic distributions.

4.1. Improved Upper Bounds

Recall that Theorem 1 establishes an upper bound for the error of FPA at the order of $\left[\mathcal{C}(G;\rho)\right]^{\frac{1}{2}}$. Under Assumption 3, Theorem 2 below refines this upper bound to a lower order of $\mathcal{C}(G;\rho) \le \left[\mathcal{C}(G;\rho)\right]^{\frac{1}{2}}$, where the inequality follows immediately from $\mathcal{C}(G;\rho) \le \mathbf{e}$. Define the constant $\tilde{C} := \frac{L_f \beta^2}{4(1-\rho)(1-\rho/2)}$, which is increasing in both ρ and L_f.

THEOREM 2 **(Improved Entry-Wise Error Bound of the FPA Scheme)**. *Under Assumptions 1, 2 and 3, for any diffusion instance $\left(G, \mathbf{v}, F_\epsilon(\cdot), \beta\right)$, we have*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\le\; \tilde{C} \cdot \mathcal{C}(G;\rho) \quad \textit{and} \quad \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\le\; \tilde{C} \cdot \frac{1}{d_{\min}}. \tag{15}$$

Our inverse indegree centrality $\mathcal{C}(G;\rho)$ remains essential in the improved bounds. In light of the refined bound (15), we also sharpen the scaled ℓ_1-norm of the approximation error.

COROLLARY 3 **(Improved Bound for Scaled ℓ_1-Norm)**. *Under Assumptions 1, 2 and 3, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *with* $r(G) < 1/\rho$, *we have*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; \frac{(1-\rho)\tilde{C}}{1-\rho r(G)} \cdot \mathcal{D}(G). \tag{16}$$

Finally, we remark that the proof of Theorem 2 largely parallels that of Theorem 1, with the primary difference in that we employ a second-order Taylor expansion of $F_\epsilon(\cdot)$, enabled by Assumption 3 to bound the deviation $|\mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t)|$. Lemma 3 refines Lemma 2 by removing the square root in the upper bound.

LEMMA 3 **(Improved Fixed-Point Deviation of Adoption Probability)**. *Under Assumptions 1, 2 and 3, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *and* $t \ge 1$, *we have*

$$\Big|\mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t)\Big| \;\le\; \frac{L_f \beta^2}{8}\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}.$$

4.2. A Matching Lower Bound

Our results in Section 4.1 highlight a linear dependence of FPA's error upper bounds on the network measures such as $\mathcal{C}(G;\rho)$, $1/d_{\min}$, and $\mathcal{D}(G)$. We next establish matching lower bounds of the same orders, suggesting that the order-optimality of the upper bounds.

THEOREM 3 **(A Lower Bound with Regular Graphs)**. *For any* $(n,d) \in \mathbb{Z}_+^2$ *with* $d \le n$, *there exists a diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *satisfying Assumptions 1, 2 and 3 such that* $|V| = n$, G *is* d*-regular and*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\ge\; c \cdot \frac{1}{d}\mathbf{e}, \tag{17}$$

where $c \approx 0.003 > 0$. *Consequently, it holds that*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \ge c \cdot \mathcal{C}(G;\rho), \quad \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \ge c \cdot \frac{1}{d_{\min}}, \quad \textit{and} \quad \frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \ge c \cdot \mathcal{D}(G). \tag{18}$$

For a regular graph G, it follows that $\mathcal{C}(G;\rho) = \mathbf{e}/d$ and $1/d_{\min} = \mathcal{D}(G) = 1/d$, so (18) follows from (17). Note that, for any minimal in-degree $d_{\min}$ and inverse in-degree density $\mathcal{D}(G)$, we can construct instances such that the bounds (15) and (16) are order-tight with respect to network-structure-based measures. We relegate the analysis regarding the (sub-)optimality of the constants C_ρ and $\tilde{C}$ to future research.

To show the lower bound (17), we construct a diffusion instance such that (i) the network structure is a d-regular (directed) graph $G = (V, E)$, (ii) each agent $i \in V$ has the same intrinsic value $v_i =$

-1.5, (iii) the strength of the network effect is $\beta = 1$, and (iv) the random noise follows $\epsilon_i(t) \sim$ Logistic$(0,1)$[2]. A key observation that enables the analysis is that for any node $i \in V$,

$$v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i} \in [v, v+\beta] = [-1.5, -0.5].$$

Then, we exploit the boundedness of $F_\epsilon(\cdot)$, $f_\epsilon(\cdot)$, and $f'_\epsilon(\cdot)$, as well as the convexity of $F_\epsilon(\cdot)$ in $[-1.5, -0.5]$ to derive sharp bounds that allow us to reverse the chain of inequalities in the analysis of the upper bounds, thus eventually establishing (17).

5. Numerical Experiments

In this section, we conduct a series of numerical studies, spanning across the 10-node example (i.e., Figure 1a), large-scale random networks, and real-world networks, to validate our FPA scheme in different scenarios. We underscore a few key insights from our numerics. Firstly, our FPA scheme consistently achieves superior approximation performance in every scenario, even for small and sparse networks or some instances where the assumptions necessary for the theoretical analysis, such as Assumption 2, are not met. Secondly, both the inverse in-degree centrality, $\mathcal{C}(G,\rho)$, and the inverse in-degree intensity, $\mathcal{D}(G)$, emerge as highly indicative metrics that offer precise evaluations of the FPA scheme's efficacy for a wide range of networks. Lastly, in terms of computational efficiency, FPA significantly outperforms alternative approaches, notably the MCMC method.

We note a fundamental challenge in measuring the accuracy of the FPA scheme that the ground-truth limiting adoption probabilities $\mathbf{q}^*$ are generally unknown. As outlined in Section 2, deriving $\mathbf{q}^*$ would require solving the stationary distribution of a large-scale MC, which is generally prohibitive both analytically and computationally.[3] Thus, as a workaround, we resort to the MCMC simulation with a long running time horizon to estimate the ground-truth $\mathbf{q}^*$. See Appendix C.3 for details.

Unless otherwise specified, we adhere to the following settings for all subsequent experiments. The limiting adoption probability $\mathbf{q}^*$ is computed by MCMC simulations, while the FPA solution $\boldsymbol{\mu}^*$ is obtained through fixed point iteration, with an initial value $\boldsymbol{\mu}(0) = \mathbf{0}$ and a convergence criterion of 10^{-5}. The noise distribution follows $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim}$ Logistic$(0,1)$ and the network effect intensity is set to be $\beta = 3.5$, so $\rho = 0.875$. By setting a relatively high value for ρ, our experiments aim to provide insights into the near-worst-case scenarios, offering a robust assessment of the FPA scheme. Further, to quantify performance for a specific diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, we use a self-normalized measure, the mean absolute percentage error (MAPE) across all agents, defined as MAPE $= \frac{1}{n}\sum_{i\in V} \frac{|\mu_i^* - q_i^*|}{q_i^*} \times 100\%$, rather than absolute errors as in the theoretical discussion in previous sections. Doing so allows to focus on the relative errors, which is more interpretable and facilitates the comparisons across different scales of adoption probabilities.

[2] The CDF of Logistic(μ,σ) is $F_\epsilon(x) = 1/(1+\exp\{-(x-\mu)/\sigma\})$.

[3] For some highly structured symmetric networks (such as star networks and complete networks), solving the stationary distribution is tractable. See Appendix C.2 for details.

5.1. Revisiting the Motivating Example

In this subsection, we revisit the 10-node motivating example introduced in Section 3.1 to numerically validate our theoretical results. We focus on two aspects: the role of our centrality measure as a node-level metric for the performance of FPA, and a thorough examination of Assumption 2, which serves as a sufficient condition for all our theoretical results. For the experiments in this subsection, the limiting adoption probability $\mathbf{q}^*$ is achieved by solving the exact MC stationary distribution. We use the same set of parameters and still compute the FPA solution initialized at $\boldsymbol{\mu}(0) = \mathbf{0}$.

5.1.1. Discussions on the inverse in-degree centrality. As highlighted in Theorem 1, the upper bound of the approximation error is intrinsically linked to the inverse in-degree centrality, $\mathcal{C}(G, \rho)$. Each component of $\mathcal{C}(G, \rho)$ is computed as a weighted average of the inverse in-degrees across all nodes, emphasizing the connectivity of each node to those with different in-degrees. In Figure 2, we juxtapose the absolute approximation error $|q_i^* - \mu_i^*|$ with the inverse in-degree centrality $\mathcal{C}_i(G, \rho)$ for this illustrative set of 10 nodes, which show a discernible positive correlation between these two metrics. This underscores the significance of inverse in-degree centrality as a sharp node-level indicator in evaluating the performance of the FPA scheme.

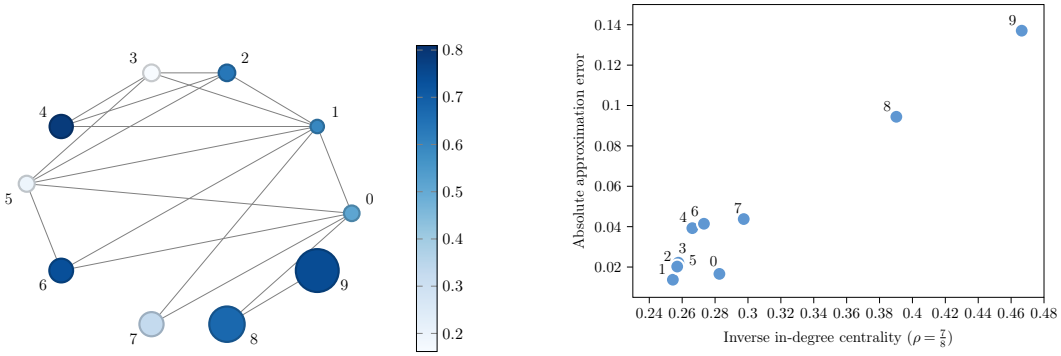
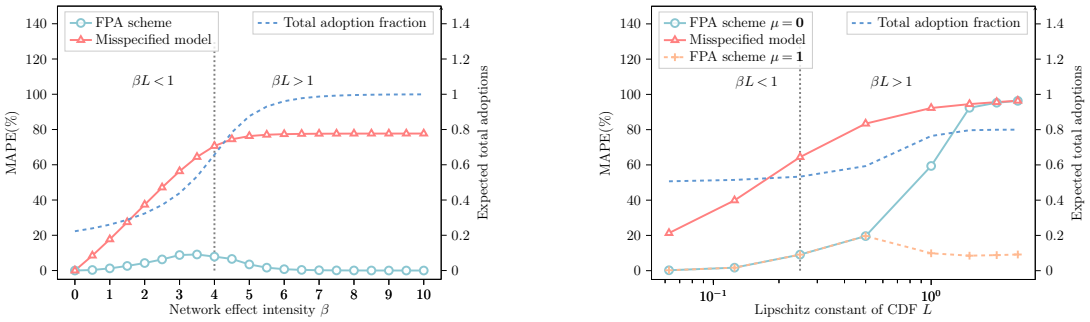


Figure 2 **Analysis of the 10-node example instance. Left: Reproduction of the network diagram from Figure 1a; Right: Illustration of the relationship between the approximation error and inverse in-degree centrality.**

5.1.2. Discussions on Assumption 2. Although similar assumptions as $\rho = \beta L < 1$ are commonly made in the network literature (e.g., see Huang et al. 2022), its implications on the FPA scheme warrant further exploration—especially when this assumption is not satisfied. The parameter ρ has two key elements, namely the network effect intensity β and the Lipschitz constant L of $F_\epsilon(\cdot)$. We experiment by varying each of these two components, benchmarking against the misspecified model where the network diffusion is not incorporated (similar to the setup in Section 3.1). Figure 3 plots the MAPE of the FPA scheme for instances with different values of β and L.

(i) *Sensitivity analysis regarding the network effect intensity*. We vary β from 0 to 10 and keep all other parameters fixed. Given that the noise distribution follows $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ with $L =$



(a) Sensitivity regarding network effect intensity

(b) Sensitivity regarding noise distribution

Figure 3 **Sensitivity of the FPA error against the discount parameter ρ. In the plot on the right, FPA scheme 0 (1, resp.) represents the FPA solution initialized with $\boldsymbol{\mu}(0) = \mathbf{0}$ ($\boldsymbol{\mu}(0) = \mathbf{1}$, resp.).**

$f_\epsilon(0) = 0.25$, our experiments encompass both scenarios when Assumption 2 is satisfied or violated. As illustrated in Figure 3a, the MAPE first increases with β, but at a notably slower rate compared to the misspecified model. Subsequently, the MAPE gradually declines to 0 when $\rho = \beta L > 1$.

For $\rho < 1$, it is not surprising that the FPA scheme performs exceptionally well when β is close to 0 (i.e., the network effect is weak). When β increases, the network effect becomes increasingly influential on user adoption behavior, leading to a slight degradation in the FPA's quality. This dependency on β is primarily reflected in the constant C_ρ (and $\tilde{C}$) in our theoretical results. Nevertheless, even when $\beta = 3.5$ where the MAPE peaks at 9.11%, the performance remains commendable, substantially lower than what is observed in the misspecified model (64.42%). This underscores the superb quality of the FPA scheme even with strong network effects in such a small network.

When $\rho > 1$, we observe an intriguing trend: the MAPE decreases in β. This demonstrates the robust and resilient performance of the FPA. The underlying reason can be attributed to the fact that L is not uniformly tight for $F_\epsilon(\cdot)$. Specifically, with large values of β, network effects heavily influence user behavior, driving the adoption probabilities close to 1 for agents. Thus, the nominal utility in the FPA scheme, represented by $v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} \mu_j(t-1)}{d_i}$, gravitates towards the flat areas of the CDF where the Lipschitz constant is significantly smaller than L, ensuring a unique limit of the FPA. Thus, even if the nominal utility for adoption, given by $v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i}$, occasionally lands in CDF regions with larger Lipschitz constants, thereby violating Assumption 2, the resilience of the FPA scheme remains evident.

(ii) *Sensitivity analysis regarding the noise distribution*. The experiment above may lead the readers to conceive that Assumption 2 is conservative, but we demonstrate that it is not the case by varying L. Particularly, we assess the performance of the FPA scheme across a family of Logistic distributions defined by $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, s)$, where s ranges from 0.0625 to 2.5. The associated Lipschitz constant can be derived as $L = \frac{1}{4s}$. All other parameters are fixed as in Section 3.1.

In Figure 3b, we observe a continuous increase of MAPE as L increases. Indeed, when $\rho > 1$, the FPA scheme's performance deteriorates drastically, approaching that of the misspecified model. This deterioration can be attributed to the violation of Proposition 2, i.e., there may be multiple solutions to the FPA scheme. In particular, the FPA solution initialized at $\boldsymbol{\mu}(0) = \mathbf{0}$ diverges from $\mathbf{q}^*$. To offer a more comprehensive view, we also present an additional FPA solution initialized at $\boldsymbol{\mu}(0) = \mathbf{1}$ in Figure 3b. Notably, these two solutions exhibit divergent performance, with the latter significantly outperforming the former when $\rho > 1$. We remark that other FPA solutions might exist with different initial values and it is difficult to know which FPA solution provides a better performance *a priori*.

Upon closer examination, an important difference emerges between the two scenarios. In the first, the intrinsic utility $\mathbf{v}$ is in the same location within the noise distribution. In contrast, the second scenario can exhibit extreme placements of $\mathbf{v}$, either to the far left or far right relative to the noise distribution, especially when s is small. Consequently, in the latter scenario, certain agents are highly inclined to adopt, while others lean towards non-adoption. When Assumption 2 is breached in this context, the FPA scheme has multiple solutions, complicating its application. In short, the effects of violating Assumption 2 on the FPA's quality are multifaceted, contingent on the specific circumstances. We leave further explorations of these phenomena to future studies.

5.2. Random Networks

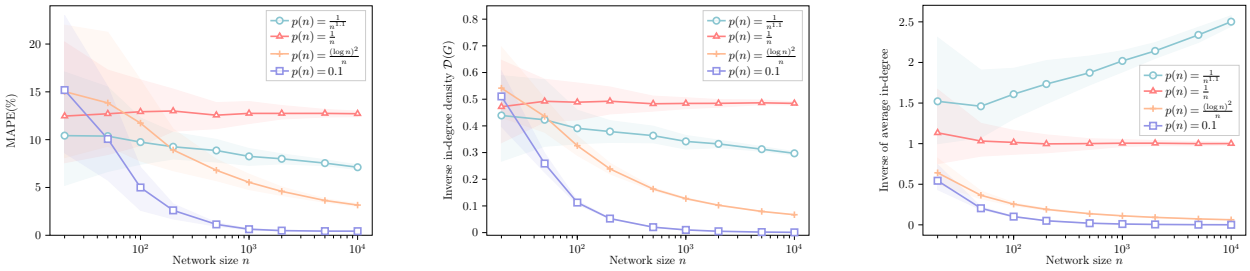
In this section, we evaluate the performance of the FPA scheme across a range of well-studied random networks, with a focus on Erdős-Rényi networks of different sizes and densities. Additionally, we explore the FPA scheme's performance in power-law networks, varying the exponent and degree correlation to gain insights into its sensitivity to network structure. These supplementary results are provided in Appendix C.4. To ensure the robustness of our empirical findings, we run 50 repetitions for each combination of random network parameters.

In the following, we consider a sequence of directed Erdős-Rényi networks, each denoted by $G(n, p(n))$, where n represents the number of nodes and $p(n)$ represents the network density—the probability that any given edge connecting two nodes exists. In an Erdős-Rényi network, edges are present independent of each other. To thoroughly understand the FPA scheme, we conduct experiments focusing on two aspects: the sensitivity of FPA's performance with respect to network structures and the computational efficiency of the FPA scheme.

5.2.1. The FPA scheme's accuracy with regard to network structure. We assess the performance of the FPA scheme by varying the size and density of Erdős-Rényi networks.

(i) *Network size*. We vary the network size n from 20 to 10,000, and select densities $p(n)$ from the set $\{\frac{1}{n^{1.1}}, \frac{1}{n}, \frac{(\log n)^2}{n}, 0.1\}$. The choice of these values is motivated by the critical ranges of $p(n)$

identified in the random graph literature. Further discussions on the properties of different Erdős-Rényi networks can be found in Appendix C.5 (see, also, Huang et al. 2022).



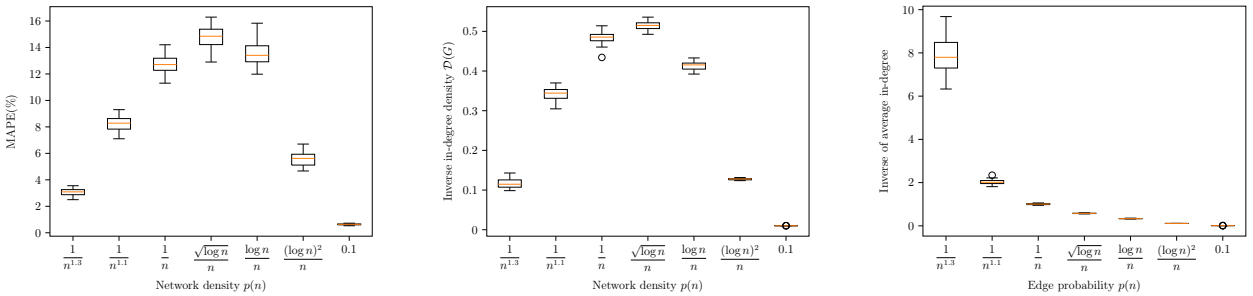
(a) MAPE against n (b) $\mathcal{D}(G)$ against n (c) Inverse of average in-degree against n

Figure 4 Performance of the FPA scheme on Erdős-Rényi networks of different network sizes. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval.

In Figure 4a, we examine how MAPE varies with the network size n. The MAPE either decreases or remains stable as n grows, but the rate of decrease differs depending on $p(n)$. Interestingly, the decrease rate is faster when $p(n)$ is relatively small or large and slower when $p(n)$ is moderate. At first glance, this may seem contradictory to our theoretical insights, which suggest that the FPA scheme performs better in larger, denser networks. However, our theoretical results, as discussed in Sections 3 and 4, are developed under the assumption $d_{\min} > 0$, thereby excluding standalone nodes. In the FPA solution, these standalone nodes, which do not receive any influence but can exert influence (i.e., their out-degree can be positive), are perfectly approximated, thereby leading to a decline in MAPE once included. Consequently, we conclude that for dense networks (i.e., $p(n) = \frac{(\log n)^2}{n}$ and $p(n) = 0.1$), the fraction of standalone nodes remains minimal, but the in-degrees increases and converges to its mean value $np(n)$ as n increases, resulting in a rapid decrease in MAPE—consistent with our theoretical predictions. Conversely, for very sparse networks (i.e., $p(n) = \frac{1}{n^{1.1}}$), the increasing fraction of standalone nodes with growing n also drives a reduction in MAPE.

The observation in Figure 4a reveals that some traditional measures of network density, such as $p(n)$ and average degree, may not effectively evaluate the FPA scheme's performance, because critical information regarding network configuration is absent in these measures. To confirm this intuition, Figures 4b and 4c present two metrics—the inverse in-degree density $\mathcal{D}(G)$ and the inverse of average in-degree, which is the metric widely used to measure network densities. We compute $\mathcal{D}(G)$ as $\frac{1}{n}\sum_{i\in V_i:d_i>0}\frac{1}{d_i}$ to reflect the impact of standalone nodes. As n increases, we find that $\mathcal{D}(G)$ follows a trend that matches that of MAPE, suggesting that it is a sharp indicator of the scheme's performance. Meanwhile, the average in-degree, which captures the traditional view of network density, cannot reflect the trend.

(ii) *Network density*: We fix a medium network size of $n = 1,000$ and focus on various densities $p(n) \in \left\{\frac{1}{n^{1.3}}, \frac{1}{n^{1.1}}, \frac{1}{n}, \frac{\sqrt{\log n}}{n}, \frac{\log n}{n}, \frac{(\log n)^2}{n}, 0.1\right\}$. In Figure 5a, the MAPE first increases and then decreases as the network gets denser, reflecting a trend that aligns with Figure 4a. The peak MAPE



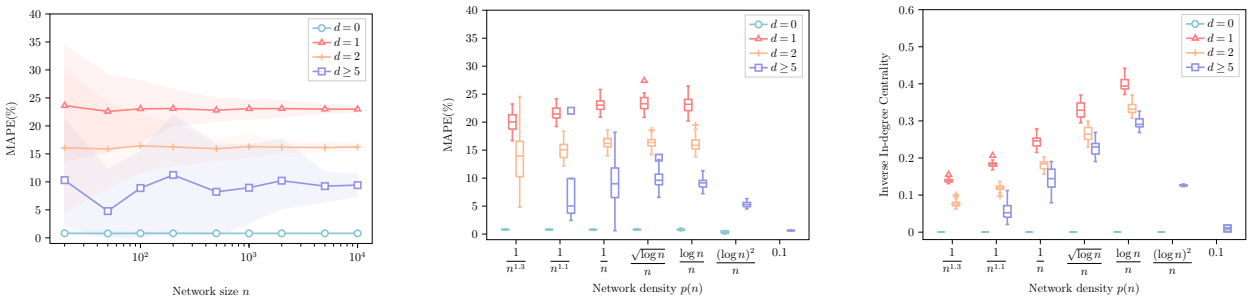
(a) MAPE against $p(n)$ (b) $\mathcal{D}(G)$ against $p(n)$ (c) Inverse of average in-degree against $p(n)$

Figure 5 Performance of the FPA scheme on Erdős-Rényi networks of different network densities.

is achieved at $p(n) = \frac{\sqrt{\log n}}{n}$. When $p(n)$ is very small (i.e., $p(n) = \frac{1}{n^{1.3}}$) or very large (i.e., $p(n) = 0.1$), the MAPE is less than 5%. Recall that we set ρ close to 1 to demonstrate the near-worst-case performance. Even under such a setting, the FPA scheme performs exceptionally well. Figures 5b and 5c further show the inverse in-degree density $\mathcal{D}(G)$ and the inverse of average in-degree, respectively, for different random graphs. The trends largely mirror those seen in Figures 4b and 4c, again confirming that $\mathcal{D}(G)$ is a robust indicator of PFA's performance.

We observe that the enhanced performance of dense networks can also be partly attributed to the largest out-in-degree ratio $r(G)$. As highlighted in Corollary 2, we show that the upper bound for approximation error increases as $r(G)$ increases. In dense Erdős-Rényi networks, both the in-degrees and the out-degrees of nodes tend to cluster around the mean value. This contributes to the density of the network while also making it more balanced. To explore the role of network imbalance level further, we conduct an extensive analysis with power-law networks, constructing in-degree and out-degree sequences with different correlations (see Appendix C.4).

Finally, we complement our analysis by singling out the agents with low in-degrees, who, as illustrated by both numerical and theoretical analyses, significantly affect the performance of FPA. We analyze the MAPE for the nodes with $d = 0, 1, 2$, and also those with $d \geq 5$ for comparison, visualized in Figure 6, across various network sizes and densities. Key takeaways from our observations include the following: Standalone agents, who have an in-degree of 0 always show zero error. However, agents with an in-degree of 1 exhibit larger MAPE compared to those with larger in-degrees, far exceeding the network-wise MAPE as shown in Figures 4a and 5a. These findings reaffirm our node-level theoretical predictions (Theorem 1 and 2), suggesting that the FPA scheme is more accurate for nodes with higher in-degrees. Additionally, for both agents with in-degrees of 1 and 2, we find



(a) MAPE against n (b) MAPE against $p(n)$ (c) $\mathcal{C}(G;\rho)$ against $p(n)$

Figure 6 Degree-level MAPE of the FPA scheme. Left: MAPE with regard to n and $p(n)$ is fixed to be $\frac{1}{n}$. The x-axis is in the log scale. Shaded areas represent the 95% confidence interval. Middle: MAPE with regard to $p(n)$ and n is fixed to be $1,000$. Right: Inverse in-degree centrality with regard to $p(n)$.

that their MAPEs remain relatively stable as n increases. However, there is a slight upward trend with growing $p(n)$. This can be attributed to the influence of more distant neighbors, highlighting the importance of capturing network's overall structure and connectivity beyond mere in-degrees. We observe that such information is comprehensively represented by our inverse indegree centrality measure, as depicted in Figure 6c. The patterns observed in the centrality align closely with those of MAPE. Indeed, this alignment would be nearly impeccable if the mean absolute error was used instead. However, we have omitted the corresponding results due to space constraints.

5.2.2. The computational efficiency of the FPA scheme. We evaluate the computational efficiency by comparing the CPU time required to calculate $\boldsymbol{\mu}^*$ and that for approximating $\mathbf{q}^*$ with MCMC. To ensure a fair comparison, we report the computational time of the MCMC process once its real-time MAPE falls below that achieved by the FPA scheme. Keeping the network density constant at $n(p) = 0.1$, we vary n from 20 to 10,000. As shown in Table 1, the runtime for both methods increases with n. However, the FPA scheme consistently outperforms MCMC by a substantial margin. The gap is even larger for large and dense networks. For instance, when $n = 10,000$, approximately 40 minutes are required for MCMC to match the performance of the FPA scheme, which completes the task in just 2.3 seconds.

Table 1 The CPU time required for the MCMC simulation and the FPA scheme.

Network size n	20	50	100	200	500	1,000	2,000	5,000	10,000
MCMC time (s)	0.1680	0.2576	0.5296	2.4152	18.9790	97.4521	286.7179	1315.3010	2366.8788
FPA time (s)	0.0015	0.0043	0.0074	0.0183	0.0438	0.1015	0.2298	1.0270	2.3044

In conclusion, the FPA scheme offers considerable advantages in computational efficiency across all the tested scenarios without a significant compromise in accuracy, which implies its potential to effectively characterize the diffusion process for a large variety of social networks.

5.3. Real-world Networks

Our numerical experiments in Section 5.2 focus on the random networks, which may not capture real-world phenomena (e.g., see Jackson 2010). As such, to evaluate the FPA scheme in more realistic settings, we test this scheme on real-world networks from the Network Repository (Rossi and Ahmed 2015). More specifically, we select five social friendship networks extracted from Facebook. These networks consist of people as nodes, with the edges representing friendship ties. An overview and the results for these networks are summarized in Table 2. For the raw data and additional summary statistics, readers can refer to the Network Repository website.[4]

Table 2 Experiment results for real-world networks.

Instance	n	$d_{\min}$	$d_{\max}$	Avg. in-degree	$\mathcal{D}(G)$	MAPE(%)	MCMC time (s)	FPA time (s)
Caltech36	770	1	248	43.2623	0.1108	3.48	4.5335	0.0636
Reed98	963	1	313	39.0696	0.0962	3.14	5.6228	0.0623
Haverford76	1,447	1	375	82.3621	0.0427	1.59	23.2347	0.1009
Simmons81	1,519	1	300	43.4338	0.0857	2.85	11.9185	0.1426
Amherst41	2,236	1	467	81.3542	0.0488	1.71	35.6749	0.1846

We emphasize three key observations from these experiments on real-world networks. First, the FPA scheme performs exceptionally well, achieving a maximum MAPE of just 3.48% across all evaluated networks. This suggests that the FPA scheme is not only accurate but also reliable for real-world applications. Second, the computational time required by the FPA scheme is significantly less than that of the MCMC method, with a factor ranging from 70 to 230. This underscores the computational efficiency of the FPA scheme, making it particularly well-suited for applications where scalability matters. Third, among various metrics such as network size n, minimal in-degree $d_{\min}$, maximal in-degree $d_{\max}$, and average in-degree, the inverse in-degree density $\mathcal{D}(G)$ stands out as the most reliable indicator of the FPA scheme's performance measured by MAPE.

Figure 7 presents a comprehensive overview of the relationship between the performance of FPA and the inverse in-degree density across different families of networks. The figure clearly plots a positive correlation between MAPE and $\mathcal{D}(G)$. This reaffirms that $\mathcal{D}(G)$ is not merely an upper bound of performance, but also a dependable and easy-to-compute metric to gauge the FPA scheme's efficacy. Notably, real-world networks typically demonstrate lower $\mathcal{D}(G)$ values and MAPE for FPA than several of the random networks we analyzed. This observation highlights the FPA scheme's practical relevance and resilience in real-world scenarios.

[4] See https://networkrepository.com/networks.php.

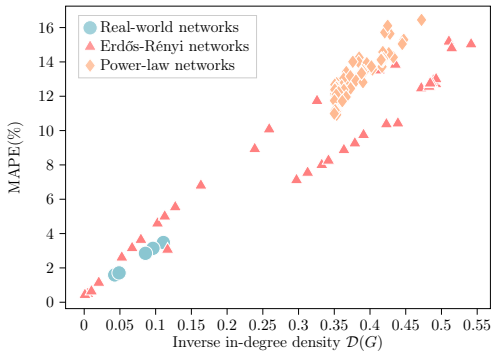


Figure 7 **Performance of the FPA scheme against the inverse in-degree density for different networks. Each node of the random networks represents the average values of the same pair of parameters across all repetitions.**

6. Applications of the Fixed-point Approximation Scheme

The FPA scheme can be applied to many classical operational decision problems that involve network diffusion. In this section, we consider two noteworthy applications: the influence maximization (IM) problem in network analysis, and optimal pricing on a social network in revenue management. Hereafter, we assume that the $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$ is known to the platform and confine our analysis of the optimization problems to a given instance.

We first formulate how the approximation error of FPA translates into the optimality gap of the original optimization problem. Consider the generally defined original problem (4) and its approximate formulation (12), we define the regret for a platform decision $\mathbf{x}$ as the difference between the optimal objective value and the objective value under $\mathbf{x}$. Formally, the regret for platform decision $\mathbf{x}$ is given by:

$$\text{Regret}(\mathbf{x}) = g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}^*), F_\epsilon, \beta), \mathbf{x}^*\Big) - g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}), F_\epsilon, \beta), \mathbf{x}\Big), \tag{19}$$

where $\mathbf{x}^*$ is the optimal solution derived from original problem (4).

6.1. Influence Maximization

In the IM problem, we aim to select a set of up to K seed users to adopt the service at the beginning, with the goal of maximizing the long-term expected total adoptions of the entire network. For example, the service provider may select the key influencers on social media as the initial adopters to promote the service to broader audiences. Additionally, we assume that the adoptions of seed users are irreversible, contrasting the standard nonprogressive diffusion setting. This choice serves two purposes. First, from an application perspective, it assumes that the influence of seed users is long-lasting, as is often the case in practice. Second, as demonstrated in Proposition 1, merely changing the initial states would not affect the long-run limit. Instead, by requiring the seed users' adoption irreversible, we effectively change the limiting behavior. In terms of problem formulation,

one can interpret it as increasing the intrinsic values sufficiently large for seed users so they will always adopt.

Given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, the original IM problem can be formulated as

$$\underset{S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} q_i^* = \lim_{t\to\infty} \sum_{i\in V} \mathbb{E}[Y_i(t)] \tag{20a}$$

$$\text{subject to} \quad Y_i(t) = \begin{cases} 1 & \forall i \in S, t \geq 1, \\ \mathbb{1}\left\{v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\} & \forall i \in V \setminus S, t \geq 1. \end{cases} \tag{20b}$$

The objective (20a) is the limiting total expected adoptions. Constraint (20b) describes the stochastic process that determines $\mathbf{q}^*$ with $Y_i(t) = 1$ for all $i \in S$ and $t \geq 0$ and initialization $\mathbf{Y}(0) = \mathbf{1}$.

Employing the FPA scheme, the approximate IM problem can be formalized as follows:

$$\underset{\boldsymbol{\mu}, S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} \mu_i \tag{21a}$$

$$\text{subject to} \quad \mu_i = 1, \qquad \forall i \in S, \tag{21b}$$

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i}\right), \quad \forall i \in V \setminus S. \tag{21c}$$

For ease of formulation, we use $\boldsymbol{\mu}$ as an explicit decision variable and a set of equality constraints specifies the FPA scheme which uniquely determines $\boldsymbol{\mu}^*(S)$ for any given $S \subseteq V$. We next derive a regret bound for the optimal solution S^{FPA} to (21) compared with the optimal solution to (20), S^*.

PROPOSITION 3 **(IM Regret Bound)**. *Under Assumptions 1 and 2, for any IM instance* $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$,

$$\text{Regret}(S^{\text{FPA}}) \;\leq\; 2C_\rho \sqrt{n \|\mathcal{C}(G, \rho)\|_1}.$$

The approximation error of the FPA scheme directly translates into the decision error. All our previous findings, including the refined bounds discussed in Section 4.1, can be extended to the approximate IM problem as well. For example, following (7), the worst-case regret bound in Proposition 3 can also be adjusted to one with order $\mathcal{O}(n/\sqrt{d_{\min}})$, which is sublinear in n when $d_{\min}$ increases with rate $\Omega(1)$.

While (21) provides an excellent approximation to the IM problem, solving it remains challenging. Under a mild technical condition, $g(\boldsymbol{\mu}^*(S), S)$ is also submodular in the seed set S as in the original optimal seeding problem. Specifically, we impose the following assumption.

ASSUMPTION 4 **(Restricted Convexity of the CDF)**. *The random noise CDF* $F_\epsilon(\cdot)$ *is convex on the range* $[-v_{\max} - \beta, -v_{\min}]$, *where* $v_{\max} := \max_{i\in V} v_i$ *and* $v_{\min} := \min_{i\in V} v_i$.

Assumption 4 covers a wide range of commonly studied cases. For example, the nonprogressive LT model, a special case of our model, naturally satisfies this assumption. Detailed discussions and additional examples supporting this assumption can be found in Appendix D.1. Under Assumption 4, the following theorem shows the submodularity of $g(\boldsymbol{\mu}^*(S), S)$.

THEOREM 4 **(Submodularity of Approximate IM Problem)**. *Under Assumptions 1, 2 and 4, $g(\boldsymbol{\mu}^*(S), S)$ is a submodular set function of seed set S.*

As an immediate corollary of Theorem 4, the well-known greedy algorithm (e.g., see Nemhauser et al. 1978) that recursively adds nodes with the largest marginal increase of total approximate adoption (i.e., adding node i that maximizes $\boldsymbol{\mu}^*(S \cup \{i\}) - \boldsymbol{\mu}^*(S)$) is applicable in our setting for solving the approximate IM problem. For instances that align with Assumption 4, the greedy algorithm provides a $(1-1/e)$-approximation solution to the approximate IM problem (21). Together with Proposition 3, the simple greedy approach also provides a high-quality solution to the original IM problem (20).

In summary, our approximate IM formulation presents several notable advantages. It facilitates establishing clear-cut conditions, such as Assumption 4, which allow us to affirm submodularity, thereby paving the way for efficient solution strategies. In contrast, verifying comparable conditions for the original IM problem can be intractable. Additionally, solution techniques like the greedy algorithm become markedly simpler to implement for the approximate IM problem. This simplification is especially pertinent for the greedy algorithm, with $\mathcal{O}(nK)$ calculations of limiting adoptions. To validate our findings, we conduct extensive numerical experiments on the approximate IM problem, with details provided in Appendix D.2. Importantly, our results indicate that the greedy algorithm for the approximate IM problem achieves near-optimal solutions irrespective of whether Assumption 4 holds or not. Moreover, this algorithm's performance is nearly on par with the simulation-based greedy algorithm for the original IM problem, surpassing many other heuristics. However, our greedy algorithm is dramatically faster than the simulation-based counterpart.

6.2. Optimal Pricing

Network effects often play an important role in determining customers' preferences for products or services, motivating an emerging literature takes into account network effects in revenue management problems (Du et al. 2016, 2018, Wang and Wang 2017, Chen and Shi 2023, Chen and Chen 2021, Gopalakrishnan et al. 2022). Our model is also naturally connected to this literature which often employs the axiomatic or game-theoretic models. We highlight that these models also arise naturally as approximations to the limiting customer purchasing behaviors in a dynamic and stochastic environment.

We assume that the firm uses pricing as the operational lever to steer consumers' adoption or purchase decisions, so we express the adoption utility for user i at time t as $u_i(t) = v_i - \gamma p_i + \beta \cdot \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t)$, where p_i is the price offered to user i and γ represents the price sensitivity. In general, we allow offering different prices to different consumers. Indeed, many platforms have the power to implement targeted price discrimination. Suppose the platform can set a maximum of m distinct prices, represented by $\mathbf{p} \in \mathbb{R}^m$. We define a transformation matrix $\mathbf{W} \in \mathbb{R}^{n\times m}$, where $W_{ik} = 1$ if consumer i is assigned with the k-th price, and $W_i = 0$ otherwise. When $m = n$ and $\mathbf{W} = \mathbf{I}_n$, customers face idiosyncratic prices. When $m = 1$ and $\mathbf{W} = \mathbf{e}_n$, customers face a homogeneous price. Various forms of price-discrimination in between, such as price discrimination based on high or low network connectivity, is also possible. We assume that the transformation matrix $\mathbf{W}$ is known by the platform; that is, the platforms have pre-determined m customer segments for pricing purposes. The ultimate objective is to identify an optimal price vector that maximizes profit.

Given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, the original pricing problem can be formulated as

$$\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{i\in V}\left(\sum_{k=1}^m W_{ik}p_k\right)\cdot q_i^* = \sum_{i\in V}\left(\sum_{k=1}^m W_{ik}p_k\right)\cdot \lim_{t\to\infty}\mathbb{E}[Y_i(t)] \tag{22a}$$

$$\text{subject to} \quad Y_i(t) = \mathbb{1}\left\{v_i - \gamma\sum_{k=1}^m W_{ik}p_k + \beta\frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\} \quad \forall i \in V, t \geq 1. \tag{22b}$$

The objective (22a) represents the total profit and constraint (22b) describes the stochastic process that determines $\mathbf{q}^*$ with given price vector $\mathbf{p}$.

Employing the FPA scheme, the approximate pricing problem can be formally stated as:

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \mathbf{W}\mathbf{p} \tag{23a}$$

$$\text{subject to} \quad \mu_i = 1 - F_\epsilon\left(-v_i + \gamma\sum_{k=1}^m W_{ik}p_k - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j}{d_i}\right),\ \forall i \in V. \tag{23b}$$

We use $\boldsymbol{\mu}$ as an explicit decision variable and use constraint (23b) to link $\mathbf{p}$ and $\boldsymbol{\mu}$. The approximate problem (23) is generally nonconvex due to the constraint (23b) and challenging to solve. Similar to the IM problem, we first establish the regret bound for the optimal solution to (23), denoted by $\mathbf{p}^{\text{FPA}}$, before discussing how to solve (23).

PROPOSITION 4 **(Regret Bound for Approximate Pricing Problem)**. *Under Assumptions 1 and 2, for any pricing instance* $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$,

$$\text{Regret}(\mathbf{p}^{\text{FPA}}) \ \leq\ p_{\max} C_\rho \sqrt{n\|\mathcal{C}(G,\rho)\|_1},$$

where $p_{\max} := \max\left\{\left\|\mathbf{p}^*\right\|_\infty, \left\|\mathbf{p}^{\text{FPA}}\right\|_\infty\right\}$.

Proposition 4 establishes a similar regret bound to Proposition 3, except that the bound being dependent on a derivative $p_{\max}$. In practice, the platform usually has a natural upper bound for the prices (e.g., the price under which no agent will make a purchase, regardless of his/her neighbors' adoption decisions), so $p_{\max}$ can be bounded by some constant. Hence, like Proposition 3, our bound in Proposition 3 also highlights a sublinear dependency on n and the network structure.

The distribution of random noise ϵ impacts the formulation and hardness of the problem. For some common utility distributions, such as the normal distribution, the optimal pricing problem is complex. Hereafter, we focus on the case with the logistic random noise where $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$. In this case, the formulation naturally relates to the existing revenue management literature (Li and Huh 2011, Gallego and Wang 2014, Golrezaei et al. 2020, Chen and Shi 2023), in which a proven-useful technique to analyze such a pricing problem is to transform it into an optimization problem in the demand space. Motivated by this technique, we consider the problem in both the adoption probability and the price spaces.

6.2.1. Profit maximization in the adoption probability space. When considering the adoption probability space, the pricing problem becomes less challenging when certain technical conditions hold. In a perfect price discrimination environment ($m=n$, $\mathbf{W}=\mathbf{I}_n$) where the platform can provide an idiosyncratic price/subsidy to each consumer and there are no price constraints, one can reformulate the problem as follows.[5]

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \mathbf{p} \tag{24a}$$

$$\text{subject to} \quad \mu_i = 1 - \frac{1}{1+\exp\{v_i + \beta \sum_{j\in\mathcal{N}_i} \mu_j/d_i - \gamma p_i\}}, \ \forall i \in V. \tag{24b}$$

Cancelling out $\mathbf{p}$, we can reformulate the problem in the adoption probability space as:

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{i\in V} \frac{1}{\gamma}\left(v_i + \beta \sum_{j\in\mathcal{N}_i} \frac{\mu_j}{d_i} + \ln\frac{1-\mu_i}{\mu_i}\right)\mu_i \tag{25a}$$

$$\text{subject to} \quad 0 \le \mu_i \le 1, \ \forall i \in V. \tag{25b}$$

When $\beta = 0$, the local network effect term does not play a role, so the problem is reduced to the classical pricing problem with a concave objective. We show that this property is preserved when β is sufficiently small.

THEOREM 5 **(Concavity of Price Optimization)**. *The objective of the pricing problem* (25) *is concave in* $\boldsymbol{\mu}$ *if and only if* $0 < \beta \le 3.375$.

[5] Possible negative prices mean that the platform can subsidize some users, in particular those who might have a large influence on the network. The platform incurs losses for these customers to promote a larger overall profit, as commonly found in practice.

Theorem 5 states that when the network diffusion intensity satisfies $0 < \beta \leq 3.375$, problem (25) is a convex optimization problem and the optimal adoption probability $\boldsymbol{\mu}^*$ can be solved by standard optimization techniques (i.e.. gradient methods). Given $\boldsymbol{\mu}^*$, we can then recover the optimal prices. Furthermore, we remark that both Theorem 5 and Assumption 2 require the network effect parameter to be relatively small, a condition consistently made in the related literature (e.g., Candogan et al. 2012).

6.2.2. Profit maximization in the price space. In a more general setting where perfect price discrimination is not feasible, the pricing problem cannot be reformulated in the adoption probability space. Thus, we need to study profit maximization directly in the price space. Particularly, we represent the adoption probability as an implicit function of price, $\boldsymbol{\mu}(\mathbf{p})$, and write the profit function as $\Pi(\mathbf{p})$. We can then derive the gradient of $\Pi(\mathbf{p})$ as follows:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}), \tag{26}$$

where the gradient of $\boldsymbol{\mu}(\mathbf{p})$ is not explicitly given. To obtain this gradient, we apply the implicit function theorem to (24b) (i.e., $\boldsymbol{\mu}(\mathbf{p}) = \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))$; see Appendix D.3 for details) and rewrite (26) as

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{-1} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}). \tag{27}$$

We can then apply the standard gradient descent techniques for nonlinear optimization problems to find the near-optimal solution. As a final remark, the profit maximization in the price space is valid under any noise distribution whereas the gradient-based approach can be easily applied to cases with more sophisticated price constraints (e.g., the box constraints).

Finally, we conduct numerical experiments to implement the gradient-based algorithms for the optimal pricing problem. We study two extreme scenarios, the perfect price discrimination case, where each consumer is offered a personalized price, and the uniform price case, where all consumers receive the same price. We show that, in both scenarios, near-optimal solutions can be achieved with the FPA scheme. Interested readers are referred to Appendix D.3 and Appendix D.4 for more detailed discussions of the implementation and results of the experiments, respectively.

7. Conclusion

In this study, we focus on nonprogressive diffusion on a social network, where agents can withdraw their previous decisions in accordance with a change in the social environment. We tide over the issues of the lack of a general modeling framework and efficient algorithms in the previous studies. Specifically, we base on a general nonprogressive diffusion model that is agent-based, considers the local network effect, and can be adapted to many utility models. We propose, with a

provable performance guarantee, a fixed-point approximation scheme that can accurately and efficiently approximate the limiting adoption probability for all agents and validate the results through extensive experiments. We provide order-optimal bounds for the approximation error and conduct a thorough analysis of its dependency with network structure. Finally, we investigate the conventional optimization problems based on the fixed-point approximation.

We also view one of our contributions as proposing a novel approach to studying the long-run behavior of the agents in networks in stochastic settings. In particular, there are several directions for future research, in which our method seems readily extendable. First, the adoptions may not change in each period but last for several periods in practice (e.g., a user needs to subscribe to Netflix for at least one month). It would be interesting to investigate how we can represent the limiting behavior in this scenario. Second, this work only considers a binary-choice case where each agent only decides to adopt or not. It is worth investigating whether similar results can be extended to a multiple-choice case (e.g., not to subscribe, to subscribe to a normal membership, or to subscribe to a premium membership). Finally, the local network effect is captured by the average adoption of the in-neighbors in our model. It is promising to consider the weighted average of in-neighbor adoptions where the network effect is asymmetric.

References

Acemoğlu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research, 38(1):1–27.

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. The Review of Economic Studies, 78(4):1201–1236.

Afèche, P., Liu, Z., and Maglaras, C. (2023). Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. Manufacturing & Service Operations Management, 25(5):1890–1908.

Agrawal, S., Yin, S., and Zeevi, A. (2021). Dynamic pricing and learning under the bass model. In Proceedings of the 22nd ACM Conference on Economics and Computation, pages 2–3.

Allon, G., Drakopoulos, K., and Manshadi, V. (2019). Information inundation on platforms and implications. In Proceedings of the 2019 ACM Conference on Economics and Computation, pages 555–556.

Anari, N., Ehsani, S., Ghodsi, M., Haghpanah, N., Immorlica, N., Mahini, H., and Mirrokni, V. S. (2010). Equilibrium pricing with positive externalities. In International Workshop on Internet and Network Economics, pages 424–431. Springer.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. Econometrica, 74(5):1403–1417.

Bapna, R. and Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. Management Science, 61(8):1902–1920.

Baron, O., Hu, M., and Malekian, A. (2022). Revenue volatility under uncertain network effects. Operations Research.

Bass, F. M. (1969). A new product growth for model consumer durables. Management science, 15(5):215–227.

Benaïm, M. and Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. Econometrica, 71(3):873–903.

Bonacich, P. (1987). Power and centrality: A family of measures. American journal of sociology, 92(5):1170–1182.

Candogan, O., Bimpikis, K., and Ozdaglar, A. (2012). Optimal pricing in networks with externalities. Operations Research, 60(4):883–905.

Chan, T.-H. H., Ning, L., and Zhang, Y. (2020). Influence maximization under the non-progressive linear threshold model. In International Workshop on Frontiers in Algorithmics, pages 37–48. Springer.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. Econometrica, 88(1):1–32.

Chen, N. and Chen, Y.-J. (2021). Duopoly competition with network effects in discrete choice models. Operations Research, 69(2):545–559.

Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 199–208.

Chen, W., Yuan, Y., and Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In 2010 IEEE international conference on data mining, pages 88–97. IEEE.

Chen, Y. and Shi, C. (2023). Network revenue management with online inverse batch gradient descent method. Production and Operations Management.

Datareportal (2022). Digital 2022: Global overview report. https://datareportal.com/reports/digital-2022-global-overview-report. Accessed: 2022-06-27.

Drakopoulos, K. and Zheng, F. (2017). Network effects in contagion processes: Identification and control. Columbia Business School Research Paper, (18-8).

Du, C., Cooper, W. L., and Wang, Z. (2016). Optimal pricing for a multinomial logit choice model with network effects. Operations Research, 64(2):441–455.

Du, C., Cooper, W. L., and Wang, Z. (2018). Optimal worst-case pricing for a logit demand model with network effects. Operations Research Letters, 46(3):345–351.

Erdős, P., Rényi, A., et al. (1960). On the evolution of random graphs. Publ. math. inst. hung. acad. sci, 5(1):17–60.

Feng, Q., Li, C., Lu, M., and Shanthikumar, J. G. (2022). Implementing environmental and social responsibility programs in supply networks through multiunit bilateral negotiation. Management Science, 68(4):2579–2599.

Gallego, G. and Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Operations Research, 62(2):450–461.

Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing letters, 12(3):211–223.

Golrezaei, N., Jaillet, P., and Liang, J. C. N. (2020). No-regret learning in price competitions under consumer reference effects. Advances in Neural Information Processing Systems, 33:21416–21427.

Gopalakrishnan, M., Zhang, H., and Zhang, Z. (2022). Multiproduct pricing under the multinomial logit model with local network effects. Decision Sciences.

Granovetter, M. (1978). Threshold models of collective behavior. American journal of sociology, 83(6):1420–1443.

Horst, U. and Scheinkman, J. A. (2006). Equilibria in systems of social interactions. Journal of Economic Theory, 130(1):44–77.

Hu, M., Wang, Z., and Feng, Y. (2020). Information disclosure and pricing policies for sales of network goods. Operations Research, 68(4):1162–1177.

Huang, J., Mani, A., and Wang, Z. (2022). The value of price discrimination in large social networks. Management Science, 68(6):4454–4477.

Jackson, M. O. (2010). Social and economic networks. Princeton university press.

Jackson, M. O., Lin, Z., and Yu, N. N. (2020). Adjusting for peer-influence in propensity scoring when estimating treatment effects. Available at SSRN 3522256.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-bayesian social learning. Games and Economic Behavior, 76(1):210–225.

Janson, S., Knuth, D. E., Łuczak, T., and Pittel, B. (1993). The birth of the giant component. Random Structures & Algorithms, 4(3):233–358.

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 137–146.

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.

Li, H. (2020). Optimal pricing under diffusion-choice models. Operations Research, 68(1):115–133.

Li, H. and Huh, W. T. (2011). Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. Manufacturing & Service Operations Management, 13(4):549–563.

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

Lin, Y., Wang, M., Shen, Z.-J. M., Zhang, H., and Zhang, R. P. (2021). Content promotion for online content platforms with network diffusion effect. Available at SSRN 3863104.

Lu, Y., Jerath, K., and Singh, P. V. (2013). The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. Management Science, 59(8):1783–1799.

Ma, L., Krishnan, R., and Montgomery, A. L. (2015). Latent homophily or social influence? an empirical analysis of purchase within a social network. Management Science, 61(2):454–473.

Mislove, A. E. (2009). Online social networks: measurement, analysis, and applications to distributed information systems. Rice University.

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. Random structures & algorithms, 6(2-3):161–180.

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—i. Mathematical programming, 14(1):265–294.

Newman, M. E., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. Physical review E, 64(2):026118.

Nosrat, F., Cooper, W. L., and Wang, Z. (2021). Pricing for a product with network effects and mixed logit demand. Naval Research Logistics (NRL), 68(2):159–182.

Rheinboldt, W. C. (1998). Methods for solving systems of nonlinear equations. SIAM.

Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Sadler, E. (2020). Diffusion games. American Economic Review, 110(1):225–70.

Schelling, T. C. (1978). Micromotives and macrobehavior. Technical report.

Shen, S., You, M., and Ma, Y. (2017). Single-commodity stochastic network design under demand and topological uncertainties with insufficient data. Naval Research Logistics (NRL), 64(2):154–173.

Shriver, S. K., Nair, H. S., and Hofstetter, R. (2013). Social ties and user-generated content: Evidence from an online social network. Management Science, 59(6):1425–1443.

Song, J.-S. and Zipkin, P. (2009). Inventories with multiple supply sources and networks of queues with overflow bypasses. Management Science, 55(3):362–372.

Van Mieghem, P., Omic, J., and Kooij, R. (2008). Virus spread in networks. IEEE/ACM Transactions On Networking, 17(1):1–14.

Wang, R. and Wang, Z. (2017). Consumer choice models with endogenous network effects. Management Science, 63(11):3944–3960.

Xie, T. and Wang, Z. (2020). Personalized assortment optimization under consumer choice models with local network effects. Available at SSRN 3788880.

Xu, H. (2018). Social interactions in large networks: A game theoretic approach. International Economic Review, 59(1):257–284.

Yang, N. and Zhang, R. P. (2022). Dynamic pricing and inventory management in the presence of online reviews. Production and Operations Management, 31(8):3180–3197.

Zeng, Z., Dai, H., Zhang, D. J., Zhang, H., Zhang, R., Xu, Z., and Shen, Z.-J. M. (2023). The impact of social nudges on user-generated content for social network platforms. Management Science, 69(9):5189–5208.

Online Appendices

Appendix A: Supporting Arguments for Section 3

In this Appendix, we start by showing the existence and uniqueness of the FPA solution.

Proof of Proposition 2: We first show the property (i) and then proof property (ii) and (iii) by showing that $\mathbf{h}(\cdot)$ is a contraction mapping.

Proof of (i): When $\boldsymbol{a} \leq \mathbf{b}$, we have $\sum_{j\in\mathcal{N}_i} a_j \leq \sum_{j\in\mathcal{N}_i} b_j$ for all $i \in V$. Since CDF $F_\epsilon(\cdot)$ is monotonically increasing, if $\boldsymbol{a} \leq \mathbf{b}$,

$$1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} a_j}{d_i}\right) \leq 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} b_j}{d_i}\right),$$

for all $i \in V$, which implies $\mathbf{h}(\boldsymbol{a}) \leq \mathbf{h}(\mathbf{b})$.

Proof of (ii) and (iii): It is trivial that $\mathbf{h}(\cdot)$ maps $\mathbb{R}^n$ to itself. Consider the Jacobian matrix of $\mathbf{h}(\boldsymbol{\mu})$, for all $\boldsymbol{\mu} \in \mathbb{R}^n$,

$$\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} = \begin{cases} 0, & j \notin \mathcal{N}_i \\ \dfrac{\beta}{d_i}\dfrac{\partial F_\epsilon\left(-v_i - \beta\frac{\sum_{j'\in\mathcal{N}_i}\mu_{j'}}{d_i}\right)}{\partial\left(-v_i - \beta\frac{\sum_{j'\in\mathcal{N}_i}\mu_{j'}}{d_i}\right)}, & j \in \mathcal{N}_i. \end{cases}$$

By Assumption 1, we can have $\left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \frac{\beta L}{d_i}$ for all $j \in \mathcal{N}_i$. Therefore, the ∞-norm of $\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}$ can be upper bounded as

$$\left\|\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}\right\|_\infty = \max_{i\in V}\sum_{j\in V}\left|\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j}\right| \leq \max_{i\in V} d_i\frac{\beta L}{d_i} = \beta L < 1$$

where the last inequality follows from Assumption 2.

Thus, for all $\boldsymbol{\mu} \in \mathbb{R}^n$, we have $\left\|\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}\right\|_\infty < 1$. It then implies that $\mathbf{h}(\boldsymbol{\mu})$ is a contraction mapping. By contraction mapping theorem, we conclude the proof. □

Then, we include the proof for the first analytical phase, namely, bounding the spatiotemporal variance. Before that, we show an important gradient for the proof in Lemma 4.

LEMMA 4 **(Variance of Lipschitz functions)**. *Let X be a random variable with a well-defined second moment. If $g(\cdot)$ is a L-Lipschitz continuous function, the following inequality holds:*

$$\mathrm{Var}(g(X)) \leq L^2\mathrm{Var}(X).$$

Proof of Lemma 4:

$$\mathrm{Var}(g(X)) = \mathrm{Var}(g(X) - g(\mathbb{E}[X])) \leq \mathbb{E}[(g(X) - g(\mathbb{E}[X]))^2] \leq \mathbb{E}[L^2(X - \mathbb{E}[X])^2] = L^2\mathrm{Var}(X)$$

where the first inequality follows since for any random variable Y, $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 \leq \mathbb{E}[Z^2]$, and the last inequality follows from the L-Lipschitz continuity. □

Proof of Lemma 1: Using the law of total covariance, we can decompose $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ into two parts:

$$\begin{aligned}\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \;=\; & \mathbb{E}_{\mathbf{Y}(t-1)}\Big[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(Y_i(t), Y_{i'}(t) \,\middle|\, \mathbf{Y}(t-1)\right)\Big] \\ & + \mathrm{Cov}_{\mathbf{Y}(t-1)}\Big(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_i(t) \mid \mathbf{Y}(t-1)\right], \mathbb{E}_{\boldsymbol{\epsilon}(t)}\left[Y_{i'}(t) \mid \mathbf{Y}(t-1)\right]\Big).\end{aligned}$$

The first term $\mathbb{E}_{\mathbf{Y}(t-1)}\left[\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}(Y_i(t), Y_{i'}(t)|\mathbf{Y}(t-1))\right]$ is always 0. The reason is as follows: by applying the law of total conditional covariance, we have

$$\begin{aligned}
&\mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\left(Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1)\right) \\
=\ &\mathbb{E}_{\boldsymbol{\epsilon}(t)}\Big[\mathrm{Cov}\left(Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\right)\Big| \mathbf{Y}(t-1)\Big] \\
&\qquad + \mathrm{Cov}_{\boldsymbol{\epsilon}(t)}\Big(\mathbb{E}\big[Y_i(t) \mid \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\big], \mathbb{E}\big[Y_{i'}(t) \mid \mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)\big] \;\Big|\; \mathbf{Y}(t-1)\Big).
\end{aligned}$$

The former term vanishes because $Y_i(t)$ and $Y_{i'}(t)$ are deterministic when given $\mathbf{Y}(t-1), \boldsymbol{\epsilon}_t$. The latter term is also zero since $\epsilon_i(t)$ and $\epsilon_{i'}(t)$ are independent of each other.

We then show that the second term $\mathrm{Cov}_{\mathbf{Y}(t-1)}\Big(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_i(t) \mid \mathbf{Y}(t-1)\big],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_{i'}(t) \mid \mathbf{Y}(t-1)\big]\Big)$ can be bounded recursively as

$$\begin{aligned}
&\mathrm{Cov}_{\mathbf{Y}(t-1)}\Big(\mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_i(t) \mid \mathbf{Y}(t-1)\big],\ \mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_{i'}(t) \mid \mathbf{Y}(t-1)\big]\Big) \\
=\ &\mathrm{Cov}\left(1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right),\ 1 - F_\epsilon\left(-v_{i'} - \beta\frac{\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}}\right)\right) \\
\le\ &\sqrt{\mathrm{Var}\left(F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right)\mathrm{Var}\left(F_\epsilon\left(-v_{i'} - \beta\frac{\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}}\right)\right)} \qquad (28a) \\
\le\ &\sqrt{(L\beta)^2\,\mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right)(L\beta)^2\,\mathrm{Var}\left(\frac{1}{d_{i'}}\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)} \qquad (28b) \\
\le\ &\frac{\rho^2}{2}\left[\mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) + \mathrm{Var}\left(\frac{1}{d_{i'}}\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)\right], \qquad (28c)
\end{aligned}$$

where (28a) follows from Cauchy–Schwarz inequality, (28b) follows from Lemma 4 and (28c) follows from the arithmetic-mean geometric-mean (AM-GM) inequality and the fact that $\rho = L\beta$. In summary, for any time $t \ge 1$, we can upper bound the covariance between any pair of nodes by

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \ \le\ \frac{\rho^2}{2}\left[\mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) + \mathrm{Var}\left(\frac{1}{d_{i'}}\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)\right]. \qquad (29)$$

Incorporating the trivial fact that the variance of a binary random variable can be no larger than $1/4$, we get $\mathrm{Var}(Y_i(t)) \le 1/4$ for all $i \in V$ and $t \ge 0$. As a consequence, we can show that

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t)\right) \ &= \ \frac{1}{d_i^2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(Y_j(t)\right) + \frac{1}{d_i^2}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_i, j'\neq j}\mathrm{Cov}\left(Y_j(t), Y_{j'}(t)\right) \\
&\le\ \frac{1}{4d_i} + \frac{d_i - 1}{d_i^2}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(\frac{1}{d_j}\sum_{k\in\mathcal{N}_j} Y_k(t-1)\right) \\
&\le \frac{1}{4d_i} + \frac{1}{d_i}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(\frac{1}{d_j}\sum_{k\in\mathcal{N}_j} Y_k(t-1)\right),
\end{aligned}$$

where the first inequality comes from (28) and the last inequality is trivial, given that $d_i > 0$ by definition.

Recalling the definition of $\tilde{\mathbf{A}}$ and $\mathbf{b}$, we can write this inequality in matrix form as

$$\boldsymbol{\kappa}(t) \ \le\ \frac{1}{4}\mathbf{b} + \frac{\rho^2}{2}\tilde{\mathbf{A}}\boldsymbol{\kappa}(t-1).$$

As a consequence, we can derive the upper bound for the variance recursively in the next. When $t=1$, agents indeed act independently given the initial adoption states $\mathbf{Y}(0)$. In other words, the diffusion process is initialized with $\mathrm{Cov}(Y_i(0), Y_{i'}(0)) = 0$ for all $i, i' \in V$. Therefore, we can express the upper bound as

$$\boldsymbol{\kappa}(1) \;\le\; \frac{1}{4}\mathbf{b}.$$

By induction, we can show that

$$\boldsymbol{\kappa}(t) \;\le\; \frac{1}{4}\mathbf{b} + \frac{\rho^2}{2}\tilde{\mathbf{A}}\boldsymbol{\kappa}(t-1) \;\le\; \frac{1}{4}\mathbf{b} + \frac{\rho^2}{2}\tilde{\mathbf{A}}\left(\frac{1}{4}\mathbf{b} + \frac{\rho^2}{2}\tilde{\mathbf{A}}\boldsymbol{\kappa}(t-2)\right) \;\le\; \cdots \;\le\; \frac{1}{4}\left(\mathbf{I} + \sum_{\tau=1}^{t-1}\frac{\rho^{2\tau}}{2^\tau}\tilde{\mathbf{A}}^\tau\right)\mathbf{b}.$$

The conclusion follows. □

We then move to the proof of the second analytical phase: bounding nonlinear dynamics.

Proof of Lemma 2: Let $\Delta_i(t) = \frac{\beta}{d_i}\big(\sum_{j\in\mathcal{N}_i} q_j(t-1) - \sum_{j\in\mathcal{N}_i} Y_j(t-1)\big)$. For any $i \in V$ and $t \ge 0$, the adoption probability of agent i at t can be written as

$$\begin{aligned} q_i(t) \;&=\; \mathbb{E}_{\mathbf{Y}(t-1)}\Big[\mathbb{E}_{\boldsymbol{\epsilon}(t)}\big[Y_i(t)\;\big|\mathbf{Y}(t-1)\big]\Big] \;=\; \mathbb{E}_{\mathbf{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right] \\ &=\; 1 - \mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right)\right]. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\left|\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right]\right| \\ =\;& \sqrt{\left(\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right]\right)^2} \\ \le\;& \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[\left(F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right)^2\right]} \\ \le\;& \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[L^2\big|\Delta_i(t-1)\big|^2\right]} \;=\; \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[(L\beta)^2\left(\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} - \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)^2\right]} \\ =\;& \sqrt{\rho^2\,\mathrm{Var}\left[\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right]}, \end{aligned}$$

where the first inequality follows by Jensen's inequality and the second inequality follows by Assumption 1.

Let $\boldsymbol{\delta} = \left[\left(\frac{\rho}{2}\right)^2\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}\right]^{\frac{1}{2}}$. Applying (14), we can obtain

$$\left|\mathbb{E}_{\mathbf{Y}(t-1)}\left[F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1)\right) - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right)\right]\right| \;\le\; \delta_i,$$

which further leads to

$$1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right) - \delta_i \;\le\; q_i(t) \;\le\; 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i}\right) + \delta_i.$$

In summary, we have

$$\Big|\mathbf{h}\big(\mathbf{q}(t-1)\big) - \mathbf{q}(t)\Big| \;\le\; \boldsymbol{\delta}.$$

and this concludes the proof. □

Finally, we prove our main result, i.e., Theorem 1.

Proof of Theorem 1: We first show by induction that, $\underline{\boldsymbol{\mu}}(t) \le \mathbf{q}(t) \le \overline{\boldsymbol{\mu}}(t)$ for each $t \ge 0$.

Base case: $t = 0$. By definition, we have $\underline{\boldsymbol{\mu}}(0) = \mathbf{q}(0) = \overline{\boldsymbol{\mu}}(0)$.

To show $t = s+1$: Assume that $\underline{\boldsymbol{\mu}}(t) \le \mathbf{q}(t) \le \overline{\boldsymbol{\mu}}(t)$. Then we have

$$\underline{\boldsymbol{\mu}}(s+1) \;=\; \mathbf{h}_{-\boldsymbol{\delta}}\big(\underline{\boldsymbol{\mu}}(s)\big) \;\le\; \mathbf{h}_{-\boldsymbol{\delta}}\big(\mathbf{q}(s)\big) \;\le\; \mathbf{q}(s+1) \;\le\; \mathbf{h}_{\boldsymbol{\delta}}\big(\mathbf{q}(s)\big) \;\le\; \mathbf{h}_{\boldsymbol{\delta}}\big(\overline{\boldsymbol{\mu}}(s)\big) \;=\; \overline{\boldsymbol{\mu}}(s+1),$$

where the first and last inequalities follow from Proposition 2(i) while the other two follow from Lemma 2.

By the contraction mapping theorem, we know that $\underline{\boldsymbol{\mu}}(t)$ (resp. $\overline{\boldsymbol{\mu}}(t)$) converges to $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) where $\underline{\boldsymbol{\mu}}^*$ (resp. $\overline{\boldsymbol{\mu}}^*$) is the fixed-point solution for $\mathbf{h}_{-\boldsymbol{\delta}}(\underline{\boldsymbol{\mu}}^*) = \underline{\boldsymbol{\mu}}^*$ (resp. $\mathbf{h}_{\boldsymbol{\delta}}(\overline{\boldsymbol{\mu}}^*) = \overline{\boldsymbol{\mu}}^*$). Thus, the following result holds,

$$\underline{\boldsymbol{\mu}}^* \le \mathbf{q}^* \le \overline{\boldsymbol{\mu}}^* \text{ and } \underline{\boldsymbol{\mu}}^* \le \boldsymbol{\mu}^* \le \overline{\boldsymbol{\mu}}^*. \tag{30}$$

By the definition of AEOs, the difference between $\underline{\boldsymbol{\mu}}^*$ and $\overline{\boldsymbol{\mu}}^*$ can be written as

$$\overline{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^* \;=\; \mathbf{h}(\overline{\boldsymbol{\mu}}^*) - \mathbf{h}(\underline{\boldsymbol{\mu}}^*) + 2\boldsymbol{\delta}.$$

Let $\Delta\boldsymbol{\mu} = \overline{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^*$, for all $i \in V$,

$$|\Delta\mu_i| \;\le\; \rho \left| \frac{\sum_{j\in\mathcal{N}_i} \overline{\mu}_j^*}{d_i} - \frac{\sum_{j\in\mathcal{N}_i} \underline{\mu}_j^*}{d_i} \right| + 2\delta_i \;=\; \rho \left| \frac{\sum_{j\in\mathcal{N}_i} \Delta\mu_i}{d_i} \right| + 2\delta_i,$$

where the inequality comes from Assumption 1.

In matrix form, we can write it as $|\Delta\boldsymbol{\mu}| \le \rho\tilde{\mathbf{A}}|\Delta\boldsymbol{\mu}| + 2\boldsymbol{\delta}$ or equivalently

$$(\mathbf{I} - \rho\tilde{\mathbf{A}})|\Delta\boldsymbol{\mu}| \le 2\boldsymbol{\delta}. \tag{31}$$

Recall that the inverse matrix $(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}$ can be expanded into the sum of matrix powers $\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell$. Given that all elements of $\tilde{\mathbf{A}}$ are non-negative, it follows that all elements of $(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}$ are also non-negative. Therefore, when we pre-multiply both sides of (31) by $(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}$, we obtain the inequality

$$|\Delta\boldsymbol{\mu}| \;\le\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\delta}. \tag{32}$$

Combining (30) and (32), we finally have the following chain of inequalities:

$$\begin{aligned}
|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\le\; |\Delta\boldsymbol{\mu}| \;\le\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\delta} \;&=\; 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1} \left[\left(\frac{\rho}{2}\right)^2 \left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1} \mathbf{b} \right]^{\frac{1}{2}} \\
&=\; \rho\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1} \left[\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1} \mathbf{b} \right]^{\frac{1}{2}} \;=\; \frac{\rho}{1-\rho} \left[(1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1} \right] \cdot \left[\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1} \mathbf{b} \right]^{\frac{1}{2}} \\
&\le\; \frac{\rho}{1-\rho} \left[(1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1} \left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1} \mathbf{b} \right]^{\frac{1}{2}} \;=\; \frac{\rho}{\sqrt{1-\rho}} \left[\left(\sum_{s=0}^{\infty} \rho^s \tilde{\mathbf{A}}^s \right) \left(\sum_{t=0}^{\infty} \left(\frac{\rho^2}{2}\right)^t \tilde{\mathbf{A}}^t \right) \mathbf{b} \right]^{\frac{1}{2}} \\
&=\; \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \left(\sum_{s,t\in\mathbb{Z}_+:\; s+t=\ell} \rho^s \cdot \left(\frac{\rho^2}{2}\right)^t \right) \tilde{\mathbf{A}}^\ell \mathbf{b} \right]^{\frac{1}{2}} \;=\; \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \frac{\rho^{\ell+1} - \left(\frac{\rho^2}{2}\right)^{\ell+1}}{\rho - \frac{\rho^2}{2}} \tilde{\mathbf{A}}^\ell \mathbf{b} \right]^{\frac{1}{2}} \\
&\le\; \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \frac{\rho^{\ell+1}}{\rho - \frac{\rho^2}{2}} \tilde{\mathbf{A}}^\ell \mathbf{b} \right]^{\frac{1}{2}} \;=\; \frac{\rho}{\sqrt{(1-\rho)(1-\rho/2)}} \left[\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} \right]^{\frac{1}{2}}.
\end{aligned}$$

where the third inequality follows from the Jensen's inequality provided $(1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}$ is a row-stochastic matrix. This concludes the proof. □

In the following, we prove the corollary for Theorem 1.

Proof of Corollary 2: From Theorem 1, it holds that

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; \frac{C_\rho}{n}\mathbf{e}^\top \mathcal{C}^{\frac{1}{2}}(G,\rho) \overset{(a)}{\le} \frac{C_\rho}{\sqrt{n}}\sqrt{\left\|\mathcal{C}^{\frac{1}{2}}(G,\rho)\right\|_1} \;=\; \frac{C_\rho\sqrt{1-\rho}}{\sqrt{n}}\sqrt{\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b}}$$

where (a) follows due to Cauchy-Schwarz inequality. Further, (a) proves (10) in the corollary.

In the following, we will bound $\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b}$. Let us define $\mathbf{D}=\mathrm{diag}(\mathbf{b})$, so that $\tilde{\mathbf{A}}=\mathbf{D}\mathbf{A}^\top$ holds where let us recall that $\mathbf{A}$ is the adjacency matrix. Further, we define

$$\mathbf{Q}(s) \;:=\; \mathbf{A}^\top\tilde{\mathbf{A}}^{s-1} \;=\; \mathbf{A}^\top(\mathbf{D}\mathbf{A}^\top)^{s-1}, \quad \forall\, s\ge 0.$$

Then, it holds that

$$\begin{aligned}\left\|\tilde{\mathbf{A}}^s\mathbf{b}\right\|_1 &= \mathbf{e}^\top\mathbf{D}\mathbf{Q}(s)\mathbf{D}\mathbf{e}\\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{d_i d_j}Q_{ij}(s) \le \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{2}\left(\frac{1}{d_i^2}+\frac{1}{d_j^2}\right)Q_{ij}(s) = \frac{1}{2}\left\|\mathbf{D}^2\mathbf{Q}(s)\mathbf{e}\right\|_1 + \frac{1}{2}\left\|\mathbf{Q}(s)\mathbf{D}^2\mathbf{e}\right\|_1, \end{aligned} \tag{33}$$

where the inequality follows from the AM-GM inequality.

We then bound the two terms in (33) as follows:

$$\frac{1}{2}\left\|\mathbf{D}^2\mathbf{Q}(s)\mathbf{e}\right\|_1 \;=\; \frac{1}{2}\left\|\mathbf{D}^2\mathbf{A}^\top\tilde{\mathbf{A}}^{s-1}\mathbf{e}\right\|_1 \;=\; \frac{1}{2}\left\|\mathbf{D}\mathbf{D}\mathbf{A}^\top\mathbf{e}\right\|_1 \;=\; \frac{1}{2}\left\|\mathbf{D}\mathbf{e}\right\|_1 \;=\; \frac{1}{2}\left\|\mathbf{b}\right\|_1,$$

where the second and the third inequalities follow because $\tilde{\mathbf{A}}$ is row-stochastic and

$$\frac{1}{2}\left\|\mathbf{Q}(s)\mathbf{D}^{-2}\mathbf{e}\right\|_1 \;=\; \frac{1}{2}\left\|(\mathbf{A}^\top\mathbf{D})^s\mathbf{D}\mathbf{e}\right\|_1 \;\le\; \frac{1}{2}\left\|\mathbf{A}^\top\mathbf{D}\right\|_1^s\left\|\mathbf{D}\mathbf{e}\right\|_1 \overset{(b)}{=} \frac{1}{2}r^s(G)\left\|\mathbf{D}\mathbf{e}\right\|_1 \;=\; \frac{1}{2}r^s(G)\left\|\mathbf{b}\right\|_1,$$

where the inequality follows from the definition of matrix-ℓ_1-norm as an operator norm and (b) follows since by inspection $r(G)=\mathbf{A}^\top\mathbf{D}$. Combined with (33), we get

$$\left\|\tilde{\mathbf{A}}^s\mathbf{b}\right\|_1 \le \frac{1}{2}\left(1+r^s(G)\right)\left\|\mathbf{b}\right\|_1 \le r^s(G)\left\|\mathbf{b}\right\|_1, \tag{34}$$

where the last inequality follows since by definition the largest out-in-degree ratio $r(G)\ge 1$. Therefore, as long as $\rho r(G)<1$,

$$\mathbf{e}^\top\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b} \;=\; \left\|\left(\mathbf{I}+\sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b}\right\|_1 \;\le\; \left\|\mathbf{b}\right\|_1 + \sum_{\ell=1}^{\infty}\rho^\ell\left\|\tilde{\mathbf{A}}^\ell\mathbf{b}\right\|_1 \;\le\; \frac{1}{1-\rho r(G)}\left\|\mathbf{b}\right\|_1,$$

where the first inequality follows from the subadditivity of norms and the last inequality follows from (34). This concludes the proof. □

Appendix B: Supporting Arguments for Section 4

For a refined upper bound on the approximation error, we first show the proof for Lemma 3, which is a refined version of Lemma 2.

Proof of Lemma 3: Let $X_i = -v_i - \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} Y_j(t-1)$ and $\nu_i = -v_i - \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} q_j(t-1)$ for all $i \in V$. For any $i \in V$ and $t \geq 1$, the adoption probability of agent i at t can be written as

$$q_i(t) \;=\; \mathbb{E}_{\mathbf{Y}(t-1)}\Big[\mathbb{E}\big[y_i(t) \mid \mathbf{Y}(t-1)\big]\Big] \;=\; \mathbb{E}_{\mathbf{Y}(t-1)}\left[1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i}\right)\right] \;=\; 1 - \mathbb{E}_{X_i}\big[F_\epsilon\left(X_i\right)\big].$$

With Assumption 3, we can apply Taylor expansion to $F_\epsilon\left(X_i\right)$ and get

$$\begin{aligned} \Big|\mathbb{E}_{X_i}\big[F_\epsilon\left(X_i\right) - F_\epsilon\left(\nu_i\right)\big]\Big| \;&=\; \left|\mathbb{E}_{X_i}\left[F_\epsilon\left(\nu_i\right) + f_\epsilon(\nu_i)(X_i - \nu_i) + \frac{1}{2} f'_\epsilon(C_i)(X_i - \nu_i)^2 - F_\epsilon\left(\nu_i\right)\right]\right| \\ &=\; \frac{1}{2}\left|\mathbb{E}_{X_i}\left[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2\right]\right|, \end{aligned} \tag{35}$$

where $\tilde{X}_i$ is a random variable such that $\tilde{X}_i$ lies in between the random variable X_i and ν_i.

Consequently, we can upper bound (35) by

$$\begin{aligned} \left|\mathbb{E}_{X_i}\left[F_\epsilon\left(X_i\right) - F_\epsilon\left(\nu_i\right)\right]\right| \;&=\; \frac{1}{2}\left|\mathbb{E}_{X_i}\left[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2\right]\right| \;\leq\; \frac{1}{2}\mathbb{E}_{X_i}\left[|f'_\epsilon(\tilde{X}_i)|(X_i - \nu_i)^2\right] \;\leq\; \frac{L_f}{2}\operatorname{Var}\left(X_i\right) \\ &=\; \frac{L_f \beta^2}{2}\operatorname{Var}\left(\frac{1}{d_i}\sum_{j \in \mathcal{N}_i} Y_j(t-1)\right), \end{aligned}$$

where the first inequality comes from Jensen's inequality and the second inequality is from Assumption 3. Let $\boldsymbol{\eta} = \frac{L_f \beta^2}{8} \cdot \left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1} \mathbf{b}$. By applying Lemma 1, we can finally get

$$\Big|\mathbb{E}\big[F_\epsilon(X_i) - F_\epsilon\left(\nu_i\right)\big]\Big| \;\leq\; \eta_i,$$

which further leads to

$$1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i}\right) - \eta_i \;\leq\; q_i(t) \;\leq\; 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i}\right) + \eta_i.$$

In conclusion, we have

$$\mathbf{h}(\mathbf{q}(t-1)) - \boldsymbol{\eta} \;\leq\; \mathbf{q}(t) \;\leq\; \mathbf{h}(\mathbf{q}(t-1)) + \boldsymbol{\eta}.$$

□

Based on Lemma 3, we then show the proof for the refined Theorem 2 and Corollary 3.

Proof of Theorem 2: Following the same steps leading to (32), with Lemma 3, we obtain

$$|\Delta \boldsymbol{\mu}| \;\leq\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\eta}. \tag{36}$$

Therefore, following the same line of analysis in the proof of Theorem 1, it holds that

$$\begin{aligned} |\mathbf{q}^* - \boldsymbol{\mu}^*| \;\leq\; |\Delta\boldsymbol{\mu}| \;\leq\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\eta} \;&=\; 2\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1} \cdot \frac{L_f\beta^2}{8} \cdot \left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} \\ &=\; \frac{L_f\beta^2}{4}\left[\left(\sum_{s=0}^{\infty}\rho^s\tilde{\mathbf{A}}^s\right)\left(\sum_{t=0}^{\infty}\left(\frac{\rho^2}{2}\right)^t\tilde{\mathbf{A}}^t\right)\mathbf{b}\right] \;\leq\; \frac{L_f\beta^2}{4(1-\rho/2)} \cdot \left(\mathbf{I} + \sum_{\ell=1}^{\infty}\rho^\ell\tilde{\mathbf{A}}^\ell\right)\mathbf{b} \\ &=\; \frac{L_f\beta^2}{4(1-\rho)(1-\rho/2)} \cdot \mathcal{C}(G;\rho). \end{aligned}$$

We conclude the proof. □

Proof of Corollary 3: By Theorem 2, we can upper bound the scaled ℓ_1-norm as

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\le\; \frac{\tilde{C}}{n}\mathbf{e}^\top \mathcal{C}(G,\rho) \;=\; \frac{(1-\rho)\tilde{C}}{n}\cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b}$$

Following the proof of Corollary 2, the last term can be bounded by

$$\frac{(1-\rho)\tilde{C}}{n}\cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell\right)\mathbf{b} \;\le\; \frac{(1-\rho)\tilde{C}}{n(1-\rho r(G))}\,\|\mathbf{b}\|_1\,,$$

and we conclude the proof. □

Proof of Theorem 3: We use the diffusion instance given in the main text to show the lower bound. We first remark on the following facts that will be used in the next. For this specific instance, the CDF F_ϵ, PDF f_ϵ and the derivative of PDF f'_ϵ are given by

$$F_\epsilon(x) = \frac{1}{1+e^{-x}}, \quad f_\epsilon(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad \text{and} \quad f'_\epsilon(x) = \frac{e^{-x}(e^{-x}-1)}{(1+e^{-x})^3}.$$

It is also convenient to define two constants that are crucial in showing the bounds as

$$u = \max_{0.5\le x\le 1.5}\{|f_\epsilon(x)|, |f'_\epsilon(x)|\} = f_\epsilon(0.5) \approx 0.235 \quad \text{and} \quad l = \min_{0.5\le x\le 1.5}\{|f_\epsilon(x)|, |f'_\epsilon(x)|\} = f'_\epsilon(0.5) \approx 0.058.$$

Specifically, we have for all $x\in[0.5,1.5]$, $l\le f_\epsilon(x)\le u$ and $-u\le f'_\epsilon(x)\le -l$. Recall that $-v_i - \beta\frac{1}{d_i}\sum_{j\in\mathcal{N}_i} Y_j(t-1) \in [-v-\beta, -v] = [0.5, 1.5]$. As a consequence, we remark that for the proof of Lemma 1 to be valid, it suffices to use the Lipschitz constant $L = u$, see (28b).

To lower bound the variance, we first provide a lower bound of the variance of the adoption indicator $Y_i(t)$ for each agent $i\in V$ and $t\ge 0$. Since $\mathrm{var}(Y_i(t)) = q_i(t)(1-q_i(t))$ where $q_i(t) = \mathbb{E}[1 - F_\epsilon(-v-\beta\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t-1))]$, we can derive that

$$\mathrm{var}\left(Y_i(t)\right) \;\ge\; \left(1-F_\epsilon(-v)\right)\cdot F_\epsilon(-v-\beta) \;=\; \left(1-F_\epsilon(1.5)\right)F_\epsilon(0.5) \;\approx\; 0.114. \tag{37}$$

We define constant $c_1 = \left(1-F_\epsilon(1.5)\right)F_\epsilon(0.5) \approx 0.114$.

By Lemma 1, we have

$$\boldsymbol{\kappa}(t) \;\le\; \frac{1}{4d}\left[\mathbf{I} + \sum_{\tau=1}^{t}\left(\frac{\rho^2}{2}\right)^\tau \tilde{\mathbf{A}}^\tau\right]\mathbf{e} \;=\; \frac{1}{4d}\left[1 + \sum_{\tau=1}^{t}\left(\frac{\rho^2}{2}\right)^\tau\right]\mathbf{e} \;\le\; \frac{1}{4\left(1-\frac{\rho^2}{2}\right)d}\boldsymbol{e}, \tag{38}$$

where $\rho = L\beta = u\beta \approx 0.235$. Furthermore, we are able to provide a lower bound on the pair-wise covariance, which is similar to (29) as

$$\mathrm{Cov}(Y_i(t), Y_{i'}(t)) \;\ge\; -\frac{\rho^2}{2}\left[\mathrm{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) + \mathrm{Var}\left(\frac{1}{d}\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)\right]. \tag{39}$$

Therefore, we can derive a lower bound for the in-neighbor variance as

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t)\right) \;&=\; \frac{1}{d^2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(Y_j(t)\right) + \frac{1}{d^2}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_i, j'\neq j}\mathrm{Cov}(Y_j(t), Y_{j'}(t)) \\
&\ge\; c_1\frac{1}{d} - \frac{d-1}{d^2}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(\frac{1}{d}\sum_{k\in\mathcal{N}_j} Y_k(t-1)\right) && \text{(40a)} \\
&\ge\; c_1\frac{1}{d} - \frac{1}{d}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(\frac{1}{d}\sum_{k\in\mathcal{N}_j} Y_k(t-1)\right) && \text{(40b)} \\
&\ge\; c_1\frac{1}{d} - \frac{\frac{\rho^2}{2}}{4\left(1-\frac{\rho^2}{2}\right)d}, && \text{(40c)}
\end{aligned}$$

where (40a) follows from (39) and (37), (40b) follows from the trivial fact that the variance is nonnegative, and (40c) follows from (38).

Define $c_2 = c_1 - \frac{\frac{\rho^2}{2}}{4\left(1-\frac{\rho^2}{2}\right)} \approx 0.106$. We then obtain that for all $i \in V$ and $t \geq 0$,

$$\mathrm{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t)\right) \;\geq\; c_2\frac{1}{d} \;=\; 0.106\frac{1}{d}.$$

We then bound the difference between $\mathbf{q}(t)$ and $\mathbf{h}\big(\mathbf{q}(t-1)\big)$. Similar to proof of Lemma 3, we obtain

$$\mathbb{E}\big[F_\epsilon(X_i) - F_\epsilon(\nu_i)\big] \;=\; \frac{1}{2}\mathbb{E}\left[f'_\epsilon(\tilde{X}_i)(X_i-\nu_i)^2\right] \;\leq\; -\frac{l}{2}\,\mathrm{Var}(X_i) \;=\; -\frac{l\beta^2}{2}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) \;\leq\; -\frac{l\beta^2 c_2}{2d}$$

where the first inequality follows because $\tilde{X}_i \in [0.5, 1.5]$ and $f'_\epsilon(x) < -l$ for x in this range. Therefore,

$$q_i(t) = 1 - \mathbb{E}[F_\epsilon(X_i)] \;\geq\; 1 - F_\epsilon(\nu_i) + \frac{l\beta^2 c_2}{2d} \;=\; h(\mathbf{q}(t-1))_i + \frac{l\beta^2 c_2}{2d}.$$

Letting $c_3 = l\beta^2 c_2/2 \approx 0.003$, we have

$$\mathbf{q}(t) \;\geq\; \mathbf{h}(\mathbf{q}(t-1)) + c_3\frac{1}{d}\boldsymbol{e}. \tag{41}$$

Finally, we lower bound the approximation error in a way analogous to Theorem 1. Let $\zeta = c_3/n$. We show $\mathbf{q}(t) \geq \overline{\overline{\boldsymbol{\mu}}}(t) \geq \boldsymbol{\mu}(t)$ by induction, where

$$\overline{\overline{\mu}}_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i}\overline{\overline{\mu}}_j(t-1)}{d_i}\right) + \zeta & t > 0 \end{cases}, \quad \text{for all } i \in V.$$

Base case $t = 0$: By definition, we have $\mathbf{q}(0) = \overline{\overline{\boldsymbol{\mu}}}(0) = \boldsymbol{\mu}(0)$.

To Show $t = s+1$: Assume that $\mathbf{q}(s) \geq \overline{\overline{\boldsymbol{\mu}}}(s) \geq \boldsymbol{\mu}(s)$. We have

$$\mathbf{q}(s+1) \;\geq\; \mathbf{h}_{\zeta\boldsymbol{e}}(\mathbf{q}(s)) \;\geq\; \mathbf{h}_{\zeta\boldsymbol{e}}(\overline{\overline{\boldsymbol{\mu}}}(s)) \;=\; \overline{\overline{\boldsymbol{\mu}}}(s+1) \;\geq\; \mathbf{h}_{\zeta\boldsymbol{e}}(\boldsymbol{\mu}(s)) \;\geq\; \mathbf{h}(\boldsymbol{\mu}(s))$$

where the first inequality follows from (41), the second and third inequalities follow Proposition 2 and the induction hypothesis, and the last inequality is trivial because $\zeta > 0$.

Thus, the following result holds,

$$\mathbf{q}^* \;\geq\; \overline{\overline{\boldsymbol{\mu}}}^* \;\geq\; \boldsymbol{\mu}^*, \tag{42}$$

where $\overline{\overline{\boldsymbol{\mu}}}^*$ is the limit of $\overline{\overline{\boldsymbol{\mu}}}(t)$. By the contraction mapping theorem, we have

$$\mathbf{q}^* - \boldsymbol{\mu}^* \;\geq\; \overline{\overline{\boldsymbol{\mu}}}^* - \boldsymbol{\mu}^* \;=\; \mathbf{h}(\overline{\overline{\boldsymbol{\mu}}}^*) - \mathbf{h}(\boldsymbol{\mu}^*) + \zeta\boldsymbol{e} \;\geq\; \zeta\boldsymbol{e} \;=\; 0.003\frac{1}{d}\mathbf{e},$$

where both inequalities follow from (42). We conclude the proof. □

Appendix C: Supplementary Numerical Experiments on the FPA Scheme

C.1. Illustration of the 10-Node Example Instance

To offer a clear illustration of the instance, we construct an undirected network comprising 10 nodes. The network structure is visualized in Fig. 1a, while the intrinsic values assigned to each agent are detailed in Table 3. We set the network effect intensity at $\beta = 3.5$ and assume that the random noise distribution is $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ for all $i \in V$ and $t \geq 0$. The characteristics of this example network, along with numerical results obtained from different models, are presented in Table 3.

Table 3 Characteristics and results of the 10-node example instance

Node	Degree	Intrinsic value v	q^*	μ^*	FPA error	p^{MM}	MM error
0	5	-1.7064	0.5126	0.5292	0.0166	0.1536	-0.3590
1	7	-1.2453	0.5932	0.6069	0.0137	0.2235	-0.3697
2	4	-0.8789	0.6325	0.6524	0.0199	0.2934	-0.3391
3	4	-3.9454	0.1442	0.1221	-0.0222	0.0190	-0.1253
4	3	-0.0822	0.7827	0.8219	0.0393	0.4795	-0.3032
5	5	-3.4441	0.1933	0.1731	-0.0202	0.0309	-0.1624
6	3	-0.2877	0.7341	0.7755	0.0414	0.4286	-0.3055
7	2	-2.9084	0.3287	0.2849	-0.0438	0.0517	-0.2770
8	2	-1.2859	0.6702	0.7646	0.0944	0.2166	-0.4536
9	1	-0.6963	0.7416	0.8786	0.1371	0.3326	-0.4090

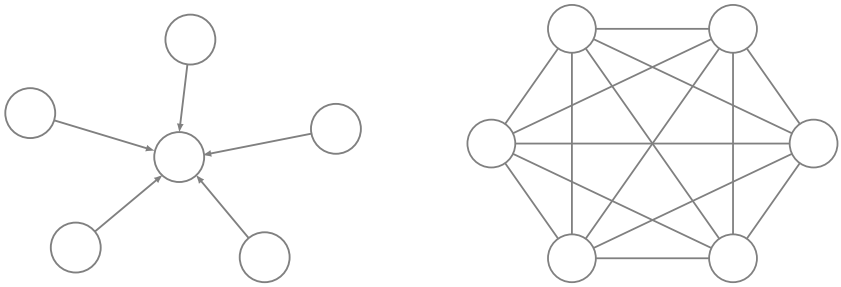
Note: p^* is calculated by first constructing a 1,024-state MC according to Section 2.2 and calculating the stationary distribution. μ^* is calculated by conducting fixed-point iteration according to (5), and FPA error equals $(\mu_i^* - q_i^*)$. p^{MM} is calculated as $\mathbb{E}[\mathbb{1}\{v_i + \epsilon_i \geq 0\}]$, and MM error equals $(\mu_i^{\text{MM}} - q_i^*)$.

C.2. Numerical Experiments on Highly-Structured Symmetric Networks

To illustrate the exact performance of the FPA scheme, we focus on two kinds of highly-structured symmetric networks, namely directed star network and complete network. These simple and symmetric structures make it easier to calculate the limiting adoption probability. We further simplify the diffusion instance by setting the intrinsic value of all agents in the network to be the same as v. This allows us either to directly compute the limiting adoption probability or to construct an MC with a much smaller state space.

Network instances:

- *Directed star network*. A star network consists of a central node and several surrounding nodes. We consider the directed version where the edges only point from surrounding nodes to the central node, not vice versa. Figure 8a shows an example of network size $n = 6$.
- *Complete network*. A complete network is the network where each node is directly connected to every other node. Figure 8b shows an example of network size $n = 6$.



(a) Directed star network (b) Complete network

Figure 8 Illustration of highly structured symmetric network structure.

For directed star networks, the adoption decisions of surrounding nodes are independent of each other. Therefore, we can directly calculate the limiting adoption probability of the central node as

$$q = \sum_{i=0}^{n-1} \binom{n-1}{i} \left(1 - F_\epsilon(-v)\right)^i F_\epsilon(-v)^{n-1-i} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)\right].$$

For complete networks, we can construct a more efficient MC by using the number of adopted agents as the MC states, rather than considering the combination of all agents' adoption states. The transition probability of this MC can be defined as

$$P(i,j) = \sum_{k=0}^{\min\{i,j\}} \binom{i}{k} \left[1 - F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right)\right]^k F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right)^{i-k} \binom{n-i}{j-k} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)\right]^{j-k} F_\epsilon\left(-v - \beta \frac{i}{n-1}\right)^{n-i-j+k}.$$

Hence, the limiting adoption probability for this MC can be easily calculated.

We measure the performance of the FPA scheme by the percentage error (PE) of the representative node, given in the following equation:

$$\text{PE} = \frac{\mu_i^* - q_i^*}{q_i^*} \cdot 100\%.$$

In directed star networks, we focus solely on the central node because the surrounding nodes have zero in-degree and can thus be perfectly approximated by the FPA scheme. In complete networks, the PE is identical for all nodes. Therefore, the PE for any arbitrary node in a complete network is equivalent to the mean average percentage error.

To assess the FPA scheme's performance, we investigate two scenarios for both types of network structures: (i) a sequence of diffusion instances with different intrinsic values, and (ii) a sequence of diffusion instances with different network sizes. For these experiments, we set the network effect intensity to be $\beta = 1$ and generate the random noise $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$.

(i) *The accuracy with regard to intrinsic values.* We choose the intrinsic value v from -5 to 5 in increments of 0.1. These instances are tested on networks of size $n \in \{10, 20, 30\}$. Figure 9 shows the PE of both network structures at different intrinsic values. Overall, all instances have a small absolute percentage error (less than

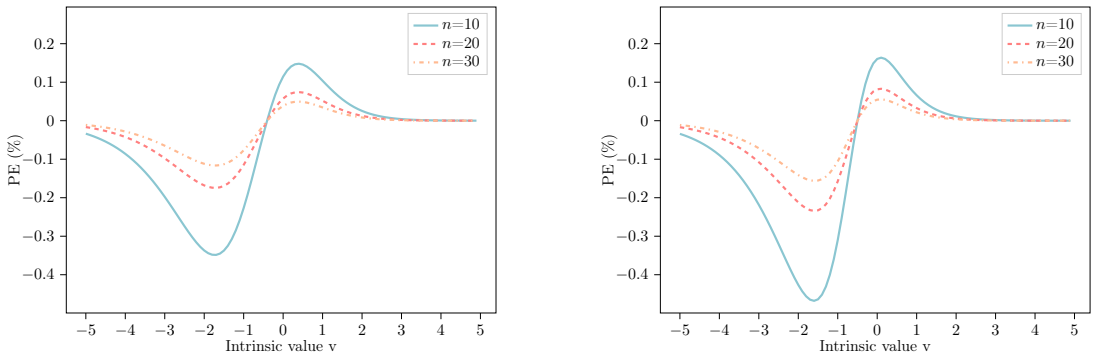


Figure 9 PE versus intrinsic value. Left: directed star network; Right: complete network.

0.5%), illustrating the high accuracy of the FPA solution. We notice that the PE curves of different network

structures possess similar shapes, however, the exact values are slightly different. In general, the FPA scheme tends to underestimate the adoption probability when the intrinsic values are small and overestimate it when they are large. There exist two critical points at around $v = -1.7$ and $v = 0.4$ where the PE reaches extremes. These points exhibit the worst cases and align with regions where the CDF F_ϵ has the highest curvature.

(ii) *The accuracy with regard to network size.* We then focus on instances with intrinsic values at the two previously mentioned critical points $v \in \{-1.7, 0.4\}$. We choose the network size n from 2 to 50. Figure 10 shows the PE across these different network sizes. Regardless of the network structure and the intrinsic

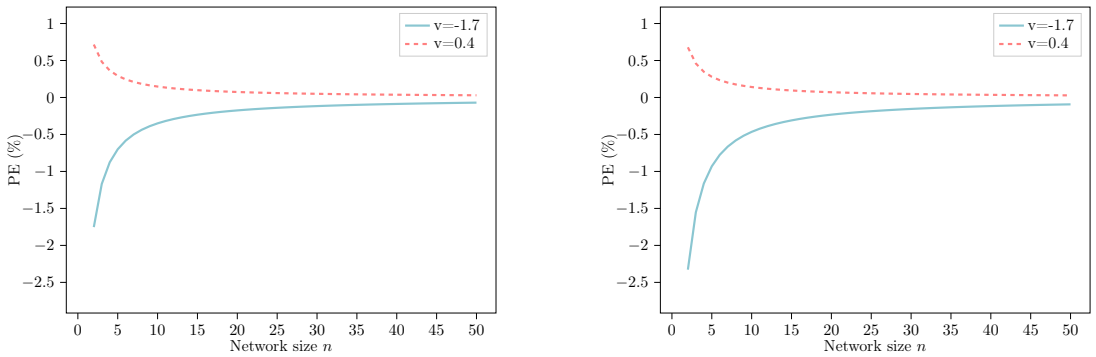


Figure 10 **PE versus network size. Left: directed star network; Right: complete network.**

values, PE converges to 0 rapidly when the network size increases. This can be theoretically confirmed, and we have explicitly demonstrated it in Corollary 1 or a refined version in Theorem 2. Our findings show that, for highly-structured networks, the FPA scheme offers excellent approximation quality and exhibits asymptotic convergence as the network size grows.

C.3. MCMC Simulation Settings

To estimate the limiting adoption probability for general instances where direct computation is impractical, we employ the MCMC simulation technique. The simulation initiates with all agents in a non-adopted state. We designate the first 1,000 time steps as the warm-up period to allow the system to reach a steady state, which we will elaborate on shortly. These initial steps are discarded from our analysis to avoid transient bias. In line with (3), we set the run length for each simulation replication to be 100,000 steps beyond the warm-up period. The adoption frequency of each agent throughout this period then serves as the ground truth for the limiting adoption probability.

In the following, we conduct additional experiments to empirically show when the MC enters a steady state so that the data samples can be gathered to calculate limiting adoption probability. Instead of focusing on the probability of each state of the MC, we use the average cumulative adoption proportion among the population as an indicator. This is represented by the following equation:

$$\frac{1}{t}\sum_{\tau=1}^{t}\frac{1}{n}\sum_{i\in V}Y_i(\tau).$$

In Figure 11, we show how the average cumulative adoption proportion changes with time. We test on 4 different diffusion instances, each represented by a randomly sampled Erdős-Rényi network $G(n, p(n))$.

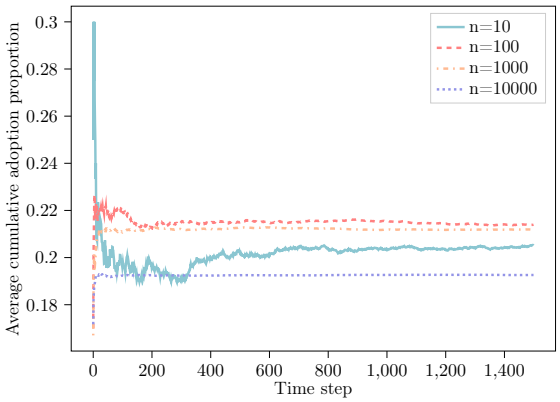


Figure 11 Average cumulative adoption proportion versus MCMC time steps

We choose the network size from $n \in \{10, 100, 1000, 10000\}$ and keep the probability of edge existence to be $p(n) = 0.1$. We set the network effect coefficient β to be 1. Our observations reveal that after 1,000 time steps, all tested trajectories have reached a steady state. Furthermore, larger networks appear to reach a steady state more rapidly. Additional tests on diffusion instances with varying parameters yielded similar results. Based on these results, we conclude that a warm-up period of 1,000 time steps is adequate for our problem context.

C.4. Numerical Experiments for Power-Law Networks.

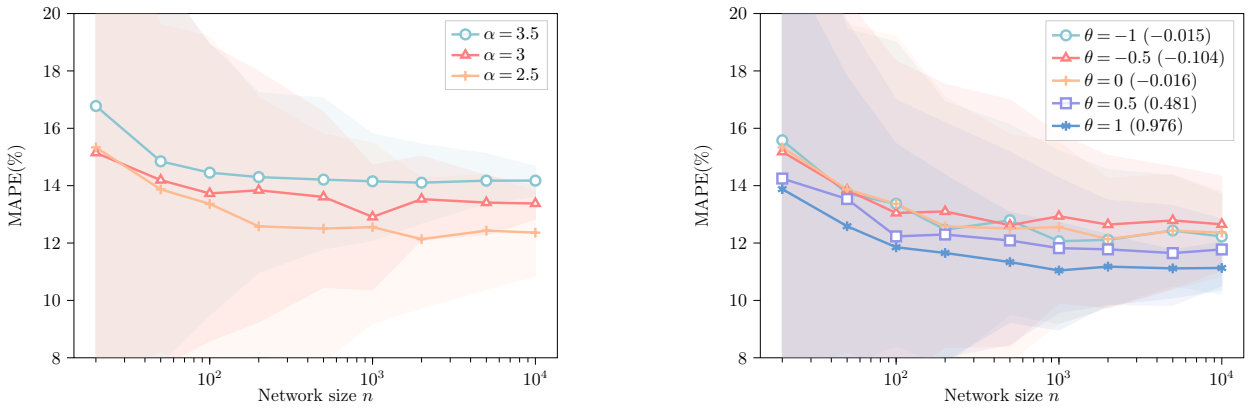
We extend our examination of the FPA scheme another important class of random networks i.e., power-law networks. These networks exhibit a degree distribution that follows a power-law pattern. We consider a sequence of directed power-law networks with n nodes and define the associated CDF of the degree distribution as $F_d(\cdot\,; n)$. The network in- and out-degrees are generated using the following CDF:

$$F_d(x; n) \;=\; \mathbb{P}(d \le x) \;=\; \frac{1 - \left(\frac{x}{d_{\min}}\right)^{1-\alpha}}{1 - \left(\frac{d_{\max}}{d_{\min}}\right)^{1-\alpha}} \quad \text{for} \;\; d_{\min} \le x \le d_{\max} \;=\; n,$$

where the α is the exponent of power-law distribution. We set $d_{\min}$ to be 2 and $d_{\max}$ to be n. Correspondingly, the probability mass function satisfies $f_d(x) \propto x^{-\alpha}$, which aligns with the conventional definition of a power-law distribution. Power-law networks often pose significant challenges for the analysis and optimization on networks due to the prevalence of low-degree nodes in such networks. Focusing on power-law networks with $d_{\min} = 2$ and $\rho = 0.875$ (see Section 5) allows us to test the limit of the FPA scheme.

We generate power-law network based on $F_d(\cdot\,; n)$ following the approach proposed by Huang et al. (2022). Detailed information on the generation process is included in Appendix C.5 for completeness. In this generation process we use an auxiliary parameter θ to account for the pairwise correlation between the in-degree and out-degree sequences. Although θ is not the exact correlation between these two sequences, it approximates the actual correlation between in- and out-degrees, particularly for large values of n.

We conduct two sets of experiments to test the FPA scheme across different power-law exponents α and pairwise correlations θ. For each parameter combination, we conduct 100 repetitions to ensure stable performance metrics. The results are presented in Figure 12. In general, The FPA scheme still performs



(a) MAPE against α (b) MAPE against θ

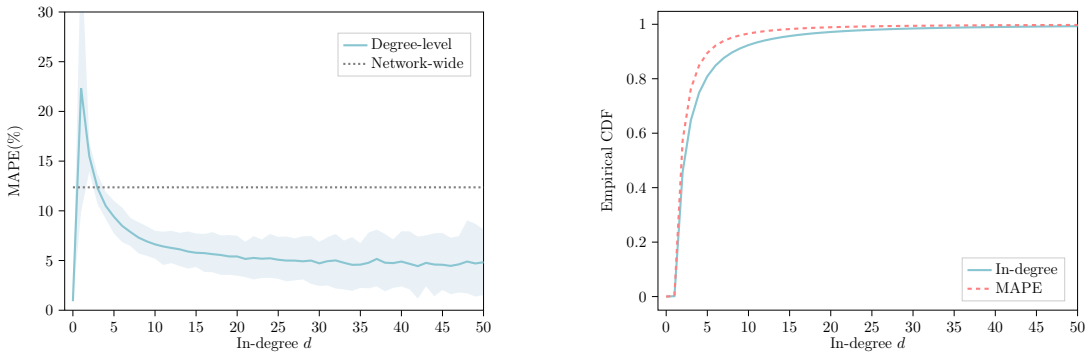
Figure 12 Performance of the FPA scheme on power-law networks of different α and θ values. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval. In the legend of the right subfigure, numbers within parentheses represent empirical correlations between in- and out-degrees.

reasonably well. Also, we observe a consistent decrease in the MAPE as n increases across all tested α and θ values, albeit at a relatively modest pace in comparison with Erdős-Rényi networks. Additionally, power-law networks exhibit both a higher mean and greater variance in MAPE. The increased mean MAPE in power-law networks is largely attributable to a higher proportion of nodes with low in-degrees. The increased variance, on the other hand, is primarily due to the more intricate structural variations inherent to power-law networks when specified parameters are used.

We note that the power-law exponent α has a crucial impact on the degree distribution. Typically, power-law networks feature an $\alpha > 2$ to avoid divergence in the expected degree. In the case where $\alpha = 3$, the network adheres to a model generated through the preferential attachment process. Accordingly, we select α from the set $\{2.5, 3, 3.5\}$ and set $\theta = 0$ to construct Figure 12a. Note that it is easy to see that when α increases, the proportion of low degree nodes increases, and consistent with our theoretical analysis, we find that the MAPE tends to increase as α increases.

In the second experiment, we generate the in-degree and out-degree sequences with $\theta \in \{-1, -0.5, 0, 0.5, 1\}$. Corollary 2 shows that the FPA scheme's performance is related to the imbalance level of the network, which can be captured by this pairwise correlation coefficient θ. Specifically, a large positive θ indicates a strong positive correlation between in-degree and out-degree sequences, resulting in a more balanced network. Conversely, a negative θ, suggests a more imbalanced network. From Figure 12b, the MAPE remains relatively stable when θ ranges between -1 and 0. However, it substantially diminishes as θ becomes positive, which aligns with our theoretical findings that the FPA scheme performs better on balanced networks.

Finally, we focus on the FPA scheme's performance for nodes with different in-degrees d. For illustration, we choose instances with $\alpha = 2.5$ and $\theta = 0$. In Figure 13a, we illustrate how the MAPE varies with respect to the in-degree d. Aside from standalone nodes—which display zero error—the MAPE consistently decreases for nodes with $d \geq 1$ as d increases. For nodes with more than 10 in-neighbors, the MAPE diminishes to less than 6.5%. Furthermore, the network-wide MAPE stands at approximately 12.36%; notably, only nodes with an in-degree of less than 3 exhibit errors above this level. In Figure 13b, we extend our analysis by



(a) MAPE against d (b) Empirical CDF for in-degree and MAPE

Figure 13 Performance of the FPA scheme on power-law networks with regard to the in-degree values. The dotted line in the left subfigure is the average network-wide MAPE. Shaded area represents 95% confidence interval.

displaying the empirical CDF for the in-degree distribution and the MAPE. Given the nature of power-law networks, a substantial number of nodes exhibit low in-degree. Moreover, these low in-degree nodes are also associated with larger errors. Specifically, 84.89% of the total error is attributable to agents with fewer than 5 in-neighbors, and 95.97% of the error can be attributed to agents with fewer than 10 in-neighbors.

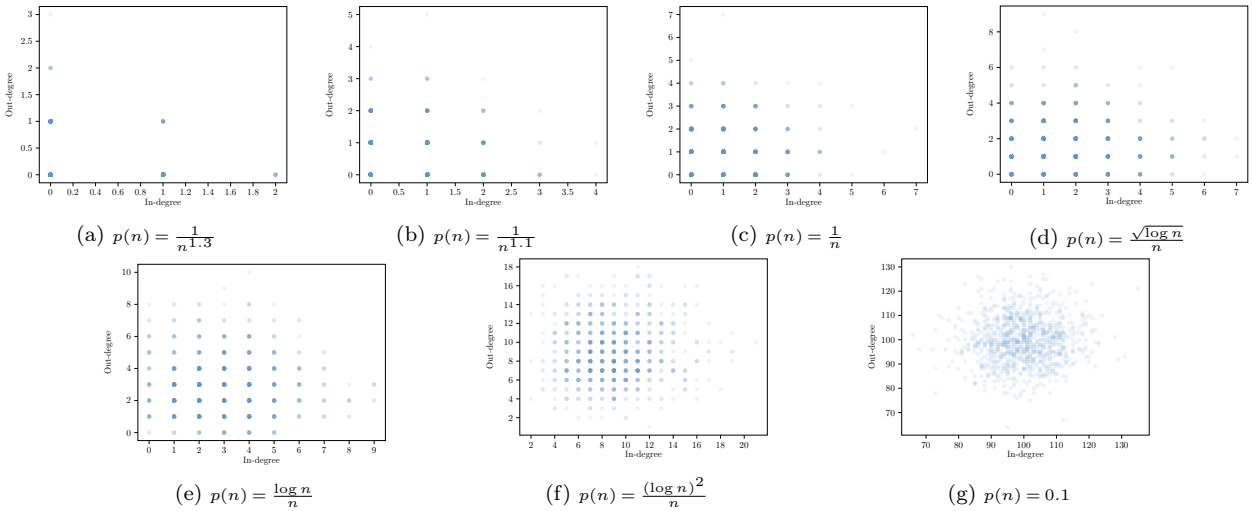
C.5. Supplementary Discussions on Random Networks.

In our numerical experiments of random networks, we generate our data following the setup outlined in Huang et al. (2022), which also offers an excellent discussion on the key properties of these networks. In the following, we revisit some of the discussions on parameter selection and instance construction for both Erdős-Rényi and power-law networks, supplementing them with additional numerical illustrations for more robust empirical support. For more details, please refer directly to this paper.

(i) *Erdős-Rényi networks*. In the asymptotic analysis of Erdős-Rényi networks, the density $p(n)$ plays a pivotal role in shaping the structural attributes of the network. Some critical cases are outlined as follows:

- When $p(n) = o(n^{-2})$, the Erdős-Rényi networks are empty almost surely (Erdős et al. 1960).
- When $p(n) = \mathcal{O}(n^{-(1+\epsilon)})$ for some $\epsilon > 0$, the expected in-degree and out-degree vanishes asymptotically. Such networks are called *very sparse* networks. They are probabilistically acyclic and fragmented.
- When $p(n) = \Theta(n^{-1})$, the expected in-degree and out-degree remain asymptotically bounded and positive. Such networks are called *critically sparse* networks. At this point, a phase transition occurs: as $p(n)$ increases from $\frac{1}{n} - \mathcal{O}(n^{-\frac{4}{3}})$ to $\frac{1}{n} + O(n^{-\frac{4}{3}})$, smaller components merge into a giant component comprising a positive fraction of nodes, and cycles begin to form (Janson et al. 1993).
- When $p(n) = \omega(\frac{\log n}{n})$, networks are called *dense* networks. These networks are highly likely to be connected, contain many cycles, and are asymptotically regular and balanced. Both in-degree and out-degree distributions concentrate around the mean value and converge asymptotically to a normal distribution.

It is important to note that dense networks are asymptotically regular and balanced. This underlying property aligns our numerical findings with the theoretical implications with regard to the imbalance level



(a) $p(n) = \frac{1}{n^{1.3}}$ (b) $p(n) = \frac{1}{n^{1.1}}$ (c) $p(n) = \frac{1}{n}$ (d) $p(n) = \frac{\sqrt{\log n}}{n}$

(e) $p(n) = \frac{\log n}{n}$ (f) $p(n) = \frac{(\log n)^2}{n}$ (g) $p(n) = 0.1$

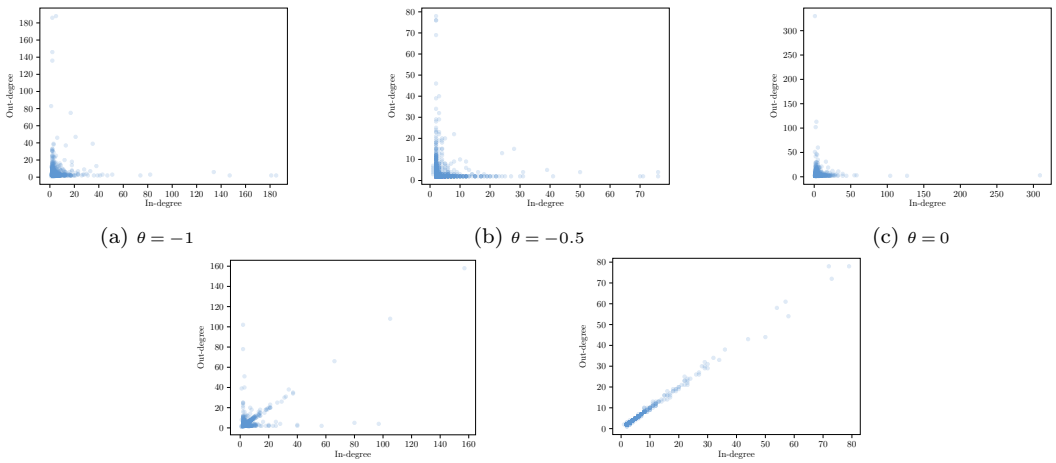
Figure 14 **In-degree and out-degree distributions of Erdős-Rényi networks with different densities** $p(n)$

of networks. In Figure 14, we illustrate the distribution of in-degree and out-degree pairs for each node in the Erdős-Rényi networks with different densities. As evidenced by Figure 14, the distribution of in-degrees and out-degrees approaches a normal distribution as $p(n)$ increases. More importantly, this trend also shows that the in-degree and the out-degree of a specific node become close to each other, contributing to a more balanced network structure.

(ii) *Power-law networks*. To avoid notation confusion, in this part, we let d_i^{in} and d_i^{out} denote the in-degree and out-degree of node i. To construct a power-law network, we require both the in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \ldots, d_n^{\text{in}}$ and the out-degrees $d_1^{\text{out}}, d_2^{\text{out}}, \ldots, d_n^{\text{out}}$ are i.i.d. sampled from distribution $F_d(\cdot, n)$. One distinctive aspect of our experiments with power-law networks is the introduction of a pairwise correlation parameter to capture the imbalance level of the network. A valid correlated in-degree sequence and out-degree sequence can be generated using the following procedure:

- Sample i.i.d. in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \ldots, d_n^{\text{in}}$ from the power-law distribution. Without loss of generality, assume this sequence is sorted in descending order.
- Sample i.i.d. random variables $Z_1, Z_2, \ldots, Z_n$ as follows: for each $i \in V$, $Z_i = 1$ with probability $|\theta|$, and $Z_i = 0$ with probability $1 - |\theta|$, where $\theta \in [-1, 1]$ is the parameter used to control the correlation. This parameter θ is not necessarily the correlation $\text{Cov}(d_i^{\text{in}}, d_i^{\text{out}})$.
- Define sets of nodes $I_0 : \{i : Z_i = 0, 1 \le i \le n\}$ and $I_1 : \{i : Z_i = 1, 1 \le i \le n\}$.
- If $\theta \ge 0$, set $d_i^{\text{out}} = d_i^{\text{in}}$ for $i \in I_1$ and set $\{d_i^{\text{out}} : i \in I_0\}$ by a random permutation of $\{d_i^{\text{in}} : i \in I_0\}$; If $\theta < 0$, set $d_i^{\text{out}} = d_{n-i+1}^{\text{in}}$ for $i \in I_0$ and set $\{d_i^{\text{out}} : i \in I_1\}$ by a random permutation of $\{d_{n-i+1}^{\text{in}} : i \in I_1\}$.
- Use a configuration model (Molloy and Reed 1995, Newman et al. 2001) to construct the directed random network with given in- and out-degree sequences.

Under this construction, when $\theta \ge 0$, the correlation is $\theta + O(n^{-1})$, so it asymptotically equals θ. When $\theta < 0$, the generated pairwise correlation may deviate from θ, and different values of θ yield similar degree sequences, as evidenced by Figure 15.



(a) $\theta = -1$ (b) $\theta = -0.5$ (c) $\theta = 0$

(d) $\theta = 0.5$ (e) $\theta = 1$

Figure 15 **In-degree and out-degree distributions of power-law networks with different pairwise correlations θ**

Appendix D: Supporting Arguments for Section 6

D.1. Supporting Arguments for Section 6.1

We first provide the proof to characterize the optimality gap of the approximate IM problem.

Proof of Proposition 3: The proof largely follows Corollary 2, where we bound the scaled ℓ_1-norm of the FPA error. Therefore, the regret can be bounded by

$$\begin{aligned}
\text{Regret}(S^{\text{FPA}}) &= \sum_{i\in V} q_i^*(S^*) - \sum_{i\in V} q_i^*\left(S^{\text{FPA}}\right) \\
&= \sum_{i\in V} q_i^*(S^*) - \sum_{i\in V} \mu_i^*(S^*) + \sum_{i\in V} \mu_i^*(S^*) - \sum_{i\in V} \mu_i^*\left(S^{\text{FPA}}\right) + \sum_{i\in V} \mu_i^*\left(S^{\text{FPA}}\right) - \sum_{i\in V} q_i^*\left(S^{\text{FPA}}\right) \\
&\le \left\|\mathbf{q}^*(S^*) - \boldsymbol{\mu}^*(S^*)\right\|_1 + \left(\sum_{i\in V} \mu_i^*(S^*) - \sum_{i\in V} \mu_i^*\left(S^{\text{FPA}}\right)\right) + \left\|\mathbf{q}^*\left(S^{\text{FPA}}\right) - \boldsymbol{\mu}^*\left(S^{\text{FPA}}\right)\right\|_1 \\
&\le \left\|\mathbf{q}^*(S^*) - \boldsymbol{\mu}^*(S^*)\right\|_1 + \left\|\mathbf{q}^*\left(S^{\text{FPA}}\right) - \boldsymbol{\mu}^*\left(S^{\text{FPA}}\right)\right\|_1 \\
&\le 2C_\rho\sqrt{n\|\mathcal{C}(G,\rho)\|_1},
\end{aligned}$$

where the first inequality holds trivially, the second inequality follows from the optimality of S^{FPA} for approximate IM problem (21), and the third inequality follows from Corollary 2. □

In the next, we discuss the applicability of Assumption 4.

Instances that satisfy Assumption 4. As we mentioned immediately after the assumption, the classical LT model (for nonprogressive diffusion) is a specific instance that meets this assumption. Recall we can recover the LT model by setting $v_i = -0.5$ and $\epsilon_i(t) \sim \mathcal{U}(-0.5, 0.5)$ for all $i \in V$ and $t \ge 1$. Therefore, for any $\beta > 0$, CDF F_ϵ can be expressed as $F_\epsilon(x) = \mathbb{1}\{x \ge -0.5\} \cdot (x + 0.5)$ on range $[0.5 - \beta, 0.5]$, which is convex. Additionally, some other diffusion instances related to common utility models can also meet Assumption 4. Some examples are: (i) Linear probability model: $v_i \ge -c$ and $\epsilon_i(t) \sim \mathcal{U}(-c, c)$ for all $i \in V$, $t \ge 0$. (ii) Logit model: $v_i \ge 0$ and $\epsilon_i(t) \sim \text{Logistic}(0, s)$ and for all $i \in V$, $t \ge 0$. (iii) Probit model: $v_i \ge 0$ and $\epsilon_i(t) \sim \mathcal{N}(0, s)$ and for all $i \in V$, $t \ge 0$. For many general distributions, the convexity assumption essentially requires the intrinsic values to be appropriately lower bounded.

Finally, we show the proof of the submodularity for the approximate IM objective.

Proof of Theorem 4: Consider two seed set $S_1 \subseteq S_2 \subseteq V$ and an additional user $w \in V \setminus S_2$, it is sufficient to show that $\boldsymbol{\mu}^*(S_2+\{w\}) - \boldsymbol{\mu}^*(S_2) \leq \boldsymbol{\mu}^*(S_1+\{w\}) - \boldsymbol{\mu}^*(S_1)$.

We consider constraints (21b) and (21c) as the dynamic system, that is, $\boldsymbol{\mu}(t) = \mathbf{h}(\boldsymbol{\mu}(t-1))$. We can notice that, for different seed sets, the transition function $\mathbf{h}$ is not the same. However, for all the users that are not selected as seed users, the transition function for the corresponding element is the same. With a little abuse of notation, in the following proof, we use $\mathbf{h}$ to denote the transition function for all users in $V \setminus (S_2+\{w\})$.

We want to show that at all time steps $t \geq 1$, the inequality $\boldsymbol{\mu}(S_2+\{w\},t) - \boldsymbol{\mu}(S_2,t) \leq \boldsymbol{\mu}(S_1+\{w\},t) - \boldsymbol{\mu}(S_1,t)$ always holds. For user $i \in S_2$, $\mu_i(S_2+\{w\},t) - \mu_i(S_2,t) = 0 \leq \mu_i(S_1+\{w\},t) - \mu_i(S_1,t)$. For user w, $\mu_w(S_2+\{w\},t) - \mu_w(S_2,t) = 1 - \mu_w(S_2,t) \leq 1 - \mu_w(S_1,t) \leq \mu_w(S_1+\{w\},t) - \mu_w(S_1,t)$. The above two inequalities hold because of Proposition 2(i). For all the other users in $V \setminus (S_2 \cup \{w\})$, we show by induction.

$\boldsymbol{t=0}$: First of all, $\boldsymbol{\mu}(S,0) = \mathbf{0}$ for all $S \subseteq S_2 \cup \{w\}$ by definition. Therefore,

$$\boldsymbol{\mu}(S_2+\{w\},0) - \boldsymbol{\mu}(S_2,0) = \boldsymbol{\mu}(S_1+\{w\},0) - \boldsymbol{\mu}(S_1,0).$$

Assume $\boldsymbol{t=s}$: The induction hypothesis holds such that

$$\boldsymbol{\mu}(S_2+\{w\},s) - \boldsymbol{\mu}(S_2,s) \leq \boldsymbol{\mu}(S_1+\{w\},s) - \boldsymbol{\mu}(S_1,s).$$

Show $\boldsymbol{t=s+1}$: We have

$$\begin{aligned}
\boldsymbol{\mu}(S_2+\{w\},s+1) - \boldsymbol{\mu}(S_2,s+1) &= \mathbf{h}(\boldsymbol{\mu}(S_2+\{w\},s)) - \mathbf{h}(\boldsymbol{\mu}(S_2,s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_2,s) + \boldsymbol{\mu}(S_1+\{w\},s) - \boldsymbol{\mu}(S_1,s)) - \mathbf{h}(\boldsymbol{\mu}(S_2,s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_1+\{w\},s)) - \mathbf{h}(\boldsymbol{\mu}(S_1,s)) \\
&= \boldsymbol{\mu}(S_1+\{w\},s+1) - \boldsymbol{\mu}(S_1,s+1),
\end{aligned}$$

where the first inequality comes from Proposition 2(i) and the second inequality comes from Assumption 4.

When t tends to infinity, we get the fixed-point solution $\boldsymbol{\mu}^*(S_2+\{w\}) - \boldsymbol{\mu}^*(S_2) \leq \boldsymbol{\mu}^*(S_1+\{w\}) - \boldsymbol{\mu}^*(S_1)$, and hence the submodularity is proved. □

D.2. Experiments on IM Problem

In the experiments, we consider two scenarios, one satisfies Assumption 4 and thus leads to a submodular influence function, while the other does not. For both scenarios, we assume that the intrinsic value $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-4,0)$ and $\beta = 3.5$. In addition, we assume the random noise to be $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-4,4)$ in the submodular case, while $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ in the nonsubmodular case.

The well-known greedy framework selects one user at each iteration which leads to the largest total adoptions. We refer to the algorithm that embeds the FPA solution into this greedy framework for the total influence evaluation as the greedy-FP algorithm. We randomly generate some small network instances to illustrate that greedy-FP can find a near-optimal solution. Although there is no theoretical guarantee for the nonsubmodular case, it is interesting to observe from the results that the FPA solutions are still of good quality. For either scenario, we generate 100 diffusion instances with random graph $G(15,0.5)$ and set the number of seed users to 5. We enumerate all the subsets to find the optimal seed set and evaluate the diffusion influence using MCMC. In Table 4, we show the numerical results of the greedy-FP algorithm.

Table 4 Numerical results of greedy-FP algorithm for IM problem.

Scenario	Percentage of instances where the optimal seeding is recovered	Optimality Gap (%) Mean	Max
Submodular	91	0.0194	0.4704
Non-submodular	84	0.0685	1.7444

We notice that in both the submodular and nonsubmodular scenarios, the greedy-FP algorithm can generate a near-optimal IM solution and even uncover the exact optimal solution for a large portion of instances. Meanwhile, the greedy-FP algorithm has a slightly better performance in the submodular case than in the nonsubmodular case but even in the nonsubmodular problem instances, it remains quite practical.

Furthermore, we choose a real-world network—*Caltech36* as introduced in Section 5.3 and compare the performance of greedy-FP with the traditional IM heuristics. Recall that the instance includes 765 agents with an average number of neighbors of 43. We define several benchmark strategies as follows. The DEG and EIG schemes are motivated by the important role of the centrality measures in diffusion discussed in the network economics literature (e.g., Ballester et al. 2006, Jackson 2010). We include them for completeness, but as substantiated in the numerical experiments, by overlooking the idiosyncratic features of the agents, these schemes are dominated by the FPA-based heuristic.

Benchmarks:

- *Greedy and MCMC* (greedy-MCMC): This is the classical algorithm used for the IM problem. The MCMC is embedded into a greedy framework for influence evaluation. The length of the MCMC run is set to 100,000 after the warm-up period.
- *Greedy and the FPA solution* (greedy-FP): This is our proposed algorithm. We embed the FPA solution into a greedy framework for an influence evaluation.
- *Greedy and low-resolution MCMC* (greedy-l-MCMC): The MCMC is embedded into a greedy framework for an influence evaluation. The length of the MCMC run is set to 50 so that the runtime is at the same scale as that of the FPA scheme.
- *Degree centrality* (DEG): Set K users with the largest degree to be seed users.
- *Eigenvector centrality* (EIG): Set K users with the largest eigenvector centrality to be seed users.
- *Model misspecification without network effect* (MM): This benchmark considers the misspecified model that ignores the network effect in the IM problem. This is the same as setting K users with the smallest intrinsic value to be seed users.
- *Random* (RAN): Randomly select K users to be seed users.

Figure 16 demonstrates the relative loss of the expected limiting adoptions compared with greedy-MCMC against the number of seed users. Similarly, we also consider both the submodular and nonsubmodular cases.

When the number of seed users increases from 0 to 30, the difficulty of the IM problem increases since the number of feasible solutions also increases. We observe that regardless of the number of seed users, the

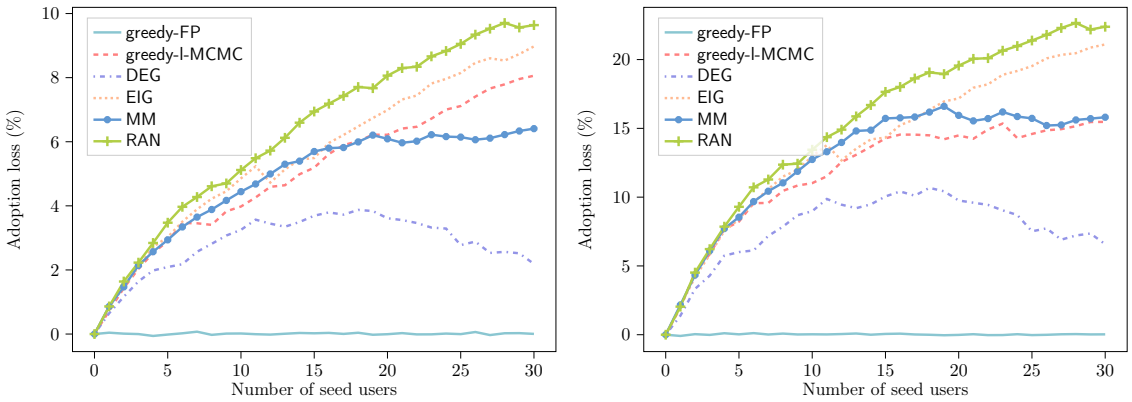


Figure 16 **Performance of different IM algorithms. Left: submodular case; Right: non-submodular case.**

performance of greedy-FP matches that of greedy-MCMC nearly perfectly. It significantly outperforms all the other benchmarks. This is no surprise to us; this is driven by the high accuracy of the FPA scheme. In particular, we also notice that the performance of the greedy framework with the MCMC method degrades drastically when the simulation length of the MCMC procedure is small. Compared with greedy-I-MCMC, greedy-FP achieves an improvement of 8.90% and 18.42% when $K = 30$ in the submodular and nonsubmodular cases. In short, we conclude that, by offering a significant efficiency gain, greedy-FP outperforms greedy-MCMC in solving the IM problem.

D.3. Supporting Arguments for Section 6.2

We first provide the proof to characterize the optimality gap of the approximate pricing problem.

Proof of Proposition 4: The proof aligns with Corollary 3. Therefore, the regret can be bounded by

$$\begin{aligned}
\text{Regret}(\mathbf{p}^{\text{FPA}}) &= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* + \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&\qquad + \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \\
&\le \left\| \left(\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*)\right)^\top \mathbf{W}\mathbf{p}^* \right\|_1 + \left(\boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W}\mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W}\mathbf{p}^{\text{FPA}}\right) \\
&\qquad + \left\| \left(\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})\right)^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \right\|_1 \\
&\le \left\| \left(\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*)\right)^\top \mathbf{W}\mathbf{p}^* \right\|_1 + \left\| \left(\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})\right)^\top \mathbf{W}\mathbf{p}^{\text{FPA}} \right\|_1 \\
&\le \left\| \mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*) \right\|_1 \left\| \mathbf{W}\mathbf{p}^* \right\|_\infty + \left\| \mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}}) \right\|_1 \left\| \mathbf{W}\mathbf{p}^{\text{FPA}} \right\|_\infty \\
&\le 2C_\rho \max\left\{ \left\| \mathbf{p}^* \right\|_\infty, \left\| \mathbf{p}^{\text{FPA}} \right\|_\infty \right\} \sqrt{n \|\mathcal{C}(G,\rho)\|_1},
\end{aligned}$$

where the first inequality follows trivially, the second inequality follows since the optimality of $\mathbf{p}^{\text{FPA}}$ for approximate pricing problem (23), the third inequality follows from Corollary 2, and the last one follows the properties of matrix operator norms. □

We then focus on the proof of the pricing problem in the adoption probability space.

Proof of Theorem 5: Let $\pi(\boldsymbol{\mu}) = \sum_{i\in V}\left(v_i + \beta\sum_{j\in\mathcal{N}_i}\frac{\mu_j}{d_i} + \ln\frac{1-\mu_i}{\mu_i}\right)\mu_i$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2\pi}{\partial\mu_i^2} = -\frac{1}{\gamma}\frac{1}{\mu_i(\mu_i-1)^2} \quad \text{and} \quad \frac{\partial^2\pi}{\partial\mu_i\partial\mu_j} = \frac{1}{\gamma}\mathbb{1}\{j\in\mathcal{N}_i\}\frac{\beta}{d_i} + \frac{1}{\gamma}\mathbb{1}\{i\in\mathcal{N}_j\}\frac{\beta}{d_j}.$$

For the diagonal elements of the Hessian matrix H_π, we can have $-1/[\mu(\mu-1)^2] \le -6.75$ holds for any $x\in[0,1]$. The inequality is tight when $\mu = 1/3$. For the nondiagonal elements of the Hessian matrix $\mathbf{H}_\pi$, we can find them related to the structure of network $G(V,E)$.

Therefore, we can have the Hessian matrix to be

$$\mathbf{H}_\pi = \operatorname{diag}\left(\left\{-\frac{1}{\mu_i(\mu_i-1)^2}\right\}_{i\in V}\right) + \beta\left(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top\right) \preceq -6.75I + \beta\left(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top\right)$$

By Gershgorin circle theorem, we can bound the eigenvalues of $\tilde{\mathbf{A}}$ by $-1 \le \lambda(\tilde{\mathbf{A}}) \le 1$. Since 1 is one of the eigenvalues of $\tilde{\mathbf{A}}$, we can have $\lambda_{\max}(\tilde{\mathbf{A}}) = 1$. Therefore, when $\beta \le 3.375$,

$$\lambda_{\max}\left(-6.75I + \beta\left(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top\right)\right) \le 0$$

which implies that $\mathbf{H}_\pi$ is negative semi-definite. The equality holds when $\beta = 3.375$.

As a result, $\mathbf{H}_\pi \preceq 0$ if and only if $\beta \le 3.375$. □

In the following, we illustrate the procedure of gradient descent for the pricing problem in the price space.

Gradient descent and approximate gradient descent. By taking the derivative on both sides of the fixed-point equation, we get

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} = \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\mathbf{p}} + \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}}\cdot\frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}.$$

By rearranging the terms, we obtain

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}}\cdot\left(I - \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}\right) = \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\mathbf{p}}.$$

Matrix $(\mathbf{I} - \partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))/\partial\boldsymbol{\mu}(\mathbf{p}))$ is guaranteed to be invertible. The reason is that, by Proposition 2, we know that $\mathbf{h}$ is a contraction mapping and $\|\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))/\partial\boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$.

With an eye toward implementation, we also notice that (27) involves the derivation of the gradient, which requires computing the inverse of an $n\times n$ matrix. When the network is large and dense, this calculation becomes intimidating. However, we notice that $\|\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))/\partial\boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$, and therefore, the spectral radius of $\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))/\partial\boldsymbol{\mu}(\mathbf{p})$ is smaller than 1, so we can expand the inverse as the sum of discounted matrix powers, which is similar to the centrality measure,

$$\left(\mathbf{I} - \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}\right)^{-1} = \lim_{n\to\infty}\sum_{\ell=0}^{k}\left(\frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}\right)^{\ell}.$$

This leads to the following k-th order approximate gradient:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} \approx \widetilde{\mathbf{G}}_k(\mathbf{p}) = \frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\mathbf{p}}\cdot\left(I + \sum_{\ell=1}^{k}\left(\frac{\partial\mathbf{h}(\mathbf{p},\boldsymbol{\mu}(\mathbf{p}))}{\partial\boldsymbol{\mu}(\mathbf{p})}\right)^{\ell}\right)\cdot\mathbf{W}\cdot\mathbf{p} + \mathbf{W}^\top\cdot\boldsymbol{\mu}(\mathbf{p}) \tag{43}$$

for the pricing problem. We expect such an easy-to-compute approximate gradient to lead to a significant efficiency gain, as is usually the case in the literature regarding approximate gradient descent (Ruder 2016). Previous works have applied similar low-order approximations for network effects for different purposes (e.g., see Candogan et al. 2012, Zeng et al. 2023). In subsequent numerical experiments, we find that $k = 2$ works very well in practice, leading to near-optimal solutions very quickly.

D.4. Experiments on Pricing Problems

In the experiments for pricing problems on a social network, the first issue is that the optimal pricing problem under the original diffusion model seems impossible to derive. We test over different randomly generated instances and find that the profits calculated via MCMC and the FPA scheme are quite close, with a percentage error almost uniformly bounded by 0.5% in our experiment.

In order to check the performance of the FPA scheme with regard to the total profit when price is considered, we test over three groups of instances. By fixing the expected number of neighbors to be 10, we generate diffusion instances with random graphs $G(20, 0.5)$, $G(100, 0.1)$, $G(1000, 0.01)$. For each instance, the agent is associated with an intrinsic value i.i.d. sampled from $\mathcal{U}(0, 4)$ and an offered price i.i.d. sampled from $\mathcal{U}(0, 4)$. We set $\beta = 3$ and $\gamma = 1$. In Figure 17, we show the distribution of profit difference among all diffusion instances. We notice that the absolute profit difference is small. Furthermore, as the network becomes larger,

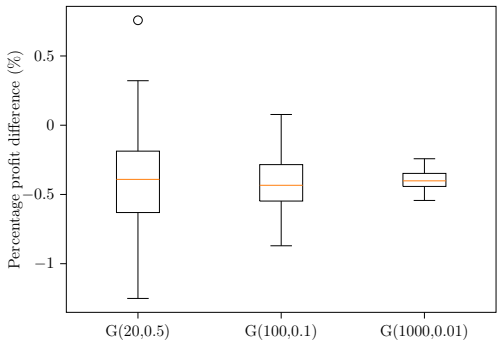


Figure 17 **Profit difference between MCMC and the FPA solution.**

the performance gap becomes more concentrated. In conclusion, we consider the FPA scheme can achieve almost the same performance as the simulation methods in the pricing problem.

Hereafter, we compare the pricing scheme under the FPA scheme as default. We assume $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$, which follows the theoretical analysis in Section 6.2. We study two extreme scenarios, the *perfect price discrimination* case, where each consumer is offered a personal price, and the *public price* case, where all consumers receive the same price.

In the perfect price discrimination scenario, we test three different algorithms. The first algorithm is the gradient descent method in the adoption probability space (grad-PROB). With a network effect parameter that satisfies Theorem 5, grad-PROB can find the global optimal solution. The second algorithm is the gradient method in the price space (grad-PRICE). The third algorithm considers the pricing problem without network diffusion, that is, the price is determined according to the standard logit model. We still refer to it as the model misspecification (MM) scheme.

For the public price case, we also test three different algorithms. However, in this case, the pricing problem cannot be considered in the adoption probability space. Instead, we use a grid search (GS) to find an upper-bound solution for the problem. Specifically, we divide the price into grids of tolerance ξ. For each price p, we upper bound the profit with $(p - \xi) \cdot \sum_{i \in V} \mu_i(p)$. The other two algorithms, grad-PRICE and MM, as discussed above, are applied here.

For both scenarios, we test on a real-world network—*Amherst41* as introduced in Section 5.3. For each diffusion instance, we set the price sensitivity as $\gamma = 0.1$, and the intrinsic value $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,4)$. In Figures 18 and 19, we plot the realized profit of three algorithms and the relative profit loss against different values of the network effect intensity β. The relative profit loss is compared with the optimal (upper bound) results from grad-PROB and GS, respectively. Furthermore, as remarked before, algorithm grad-PRICE involves the derivation of the gradient as (27), which requires calculating the inverse of a $n \times n$ matrix. We resort to the second-order approximate gradient, $\widetilde{G}_2(\mathbf{p})$, as given in (43).

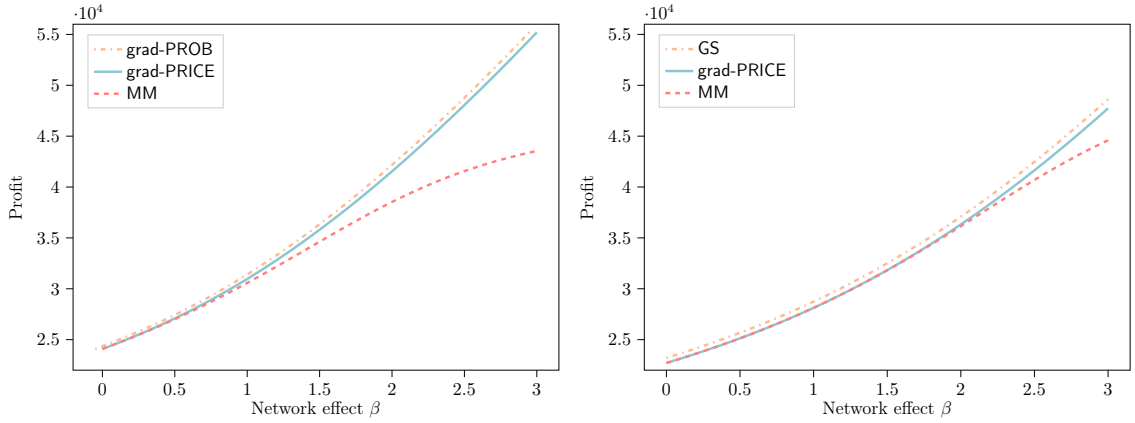


Figure 18 **Realized profit versus network effect. Left: with price discrimination; Right: without price discrimination (The curve of grad-PROB coincides with grad-PRICE in the left figure. In order to make them identifiable in the figures, we shift the grad-PROB to the left by 0.05.)**

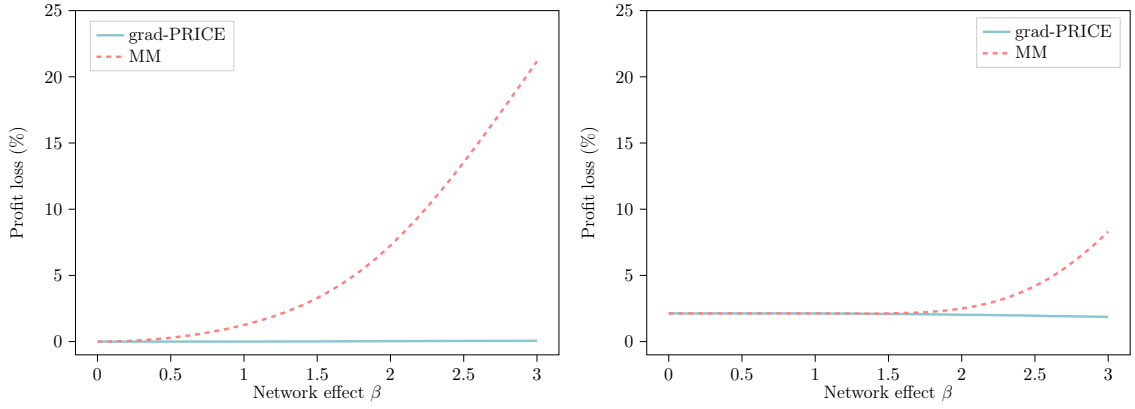


Figure 19 **Profit loss compared with grad-PROB/GS versus network effect. Left: with price discrimination; Right: without price discrimination**

We offer several observations from these two figures. First, grad-PRICE obtains a near-optimal solution in the case of price discrimination. This hints that we can use grad-PRICE to gain high-quality results in

the general pricing setting when grad-PROB is not applicable. Second, there is a significant performance degradation of MM when the network effect is large. When $\beta = 3$, the relative profit loss reaches 21.16% and 8.30% if the network effect is ignored, respectively. Third, comparing these two scenarios, we find that pricing discrimination can significantly increase the total profit, especially when the network effect is large.

Furthermore, we compare the performance of these algorithms on more instances in terms of their execution time and the quality of solutions. We assume parameters $(n, p(n), v_{\max}) \in \{100, 1000\} \times \{0.1, 0.9\} \times \{4, 10\}$, where $v_{\max}$ is a parameter representing the range of intrinsic value v_i. Specifically, we assume v_i is i.i.d. sampled from $\mathcal{U}(0, v_{\max})$. The numerical results for the two scenarios are shown in Tables 5 and 6. The grad-PRICE approach derives high-quality solutions in both scenarios. We notice that in the perfect price discrimination case, the run time for grad-PROB is less than grad-PRICE, although the margin is not too large and the run times of the two algorithms are on a similar scale. The profit difference between these two approaches is quite small, uniformly smaller than 0.2%. For the public price case, we set the tolerance of the grid search to be 0.5 within the range $[0, 100]$. grad-PRICE runs much faster than the grid search with a performance loss of up to 2%. The performance of MM remains poor across the two scenarios in this experiment, suggesting that the loss from ignoring network effects can be detrimental. In summary, our main message through the numerical experiments is twofold. First, it is important to incorporate the network effect into operational problems. The gain from doing so can be significant. Second, we advocate grad-PRICE as a practical method for price optimization. With our approximate gradient expression tailored to the network setting as in (43), grad-PRICE becomes a competitive price optimization technique. It can be efficiently implemented in various practical scenarios to find high-quality price solutions.

Table 5 Numerical results of pricing problem for randomly generated instances (perfect price discrimination)

Parameters $(n, p(n), v_{\max})$	grad-PROB	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,0.1,4)	0.035	0.115	0.064	0.079	0.100	0.007	19.0340	20.126	20.971
(100,0.1,10)	0.095	0.095	0.005	0.009	0.013	0.006	18.475	19.208	19.766
(100,0.9,4)	0.025	0.131	0.066	0.089	0.120	0.006	18.300	19.724	20.874
(100,0.9,10)	0.083	0.093	0.006	0.009	0.013	0.006	18.241	19.067	19.200
(10,000,0.1,4)	11.041	81.223	0.009	0.009	0.009	7.753	19.008	19.082	19.138
(10,000,0.1,10)	11.062	81.191	0.009	0.009	0.009	7.727	19.019	19.074	19.140
(10,000,0.9,4)	76.184	151.138	0.085	0.087	0.090	8.363	19.654	19.790	19.901
(10,000,0.9,10)	78.501	160.918	0.009	0.009	0.009	7.967	19.011	19.070	19.148

Table 6 Numerical results of pricing problem for randomly generated instances (public price)

Parameters $(n, p(n), v_{\max})$	GS	grad-PRICE				MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,0.1,4)	0.652	0.059	1.728	1.871	2.068	0.005	4.102	7.837	10.115
(100,0.1,10)	0.687	0.093	0.832	0.964	1.134	0.007	0.891	1.333	2.505
(100,0.9,4)	0.685	0.062	1.698	1.857	2.071	0.005	5.762	8.137	11.547
(100,0.9,10)	0.720	0.096	0.833	0.957	1.159	0.007	0.924	1.395	3.040
(10,000,0.1,4)	137.674	27.147	1.827	1.860	1.866	0.014	7.523	7.880	8.259
(10,000,0.1,10)	151.969	60.072	0.952	0.968	0.984	0.047	0.973	1.024	1.120
(10,000,0.9,4)	1070.717	91.743	1.835	1.860	1.866	0.024	7.634	7.879	8.294
(10,000,0.9,10)	1203.320	170.557	0.951	0.966	0.984	0.042	0.984	1.028	1.106