Integrating Social Network Analytics into Operations Management

By

Yunduan Lin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Associate Professor Scott Moura
Assistant Professor Zeyu Zheng

Spring 2024

Integrating Social Network Analytics into Operations Management

Abstract

Integrating Social Network Analytics into Operations Management

by

Yunduan Lin

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

The societal system is an intricate composition of individuals, each contributing their distinct demographics and experiences. It is not just a collection of people; it is an interconnected network that goes beyond the sum of its separate parts. Within this network, even a marginal action can have ripple effects, leading to the diffusion of behaviors, information, or, as painfully evidenced, infectious diseases. These intricate connections introduce significant challenges to the realm of operations management, including increased decision-making complexity, data overload in analysis, and a lack of theoretical guidance. All these challenges come together to form the central question that runs through my research: *How can we leverage the vast wealth of data and information available to navigate this intricate societal system for more effective operational decision-making?*

In response to this growing need, my dissertation contributes to the intersection of social network analytics and operations management. The objective is to create a more precise reflection of our interconnected societal systems, which, in turn, enables improved decision-making across a broad spectrum of platforms. To this end, I have employed a diverse tool set. These include optimization for high-quality problem-solving, data analytics to uncover actionable insights, machine learning to enable data-driven decision-making, network and graph theory to better understand the interconnected systems, and stochastic simulation for informed evaluation, etc.

The dissertation comprises three papers that each examine a different facet of integrating social network analytics with operations management. In Chapter 2, we explore the promotion optimization strategy with the consideration of the diffusion effects, drawing on extensive data from a large-scale online platform. In Chapter 3, we propose a general approximation framework to evaluate the impact of nonprogressive diffusion, delving into both its theoretical underpinnings and practical applications. In Chapter 4, we highlight the significant findings and set the stage for future research, particularly focusing on the challenges of learning user behavior within a social network with limited data availability.

To my parents, Keping Tang and Hui Lin.

Contents

List of Figures

List of Tables

Acknowledgments

I am immensely grateful to have been advised by Professor Zuo-Jun Max Shen. Before meeting him, I had not envisioned pursuing a path in academia or earning a Ph.D. degree. His exceptional support has led me through this struggling but rewarding journey. His vision and guidance will have a long-lasting impact on my future career. I would like to thank Professor Scott Moura for serving as the chair of my preliminary and qualifying exams and dissertation committee member. His academic guidance as the graduate student advisor in systems engineering has been invaluable. I would like to thank Professor Zeyu Zheng for being my dissertation committee member. His constructive feedback and fruitful discussions have been crucial in shaping both my work and my career plans. I would also like to thank Professor Joan Walker for being my preliminary and qualifying exam committee member.

It is my great fortune to collaborate with various other scholars. I am grateful to Professor Heng Zhang from ASU. I thank him for his patient academic guidance and boundless tolerance. Without his encouragement, I fear I would have given up on pursuing an academic path long ago. I am grateful to Professor Renyu Zhang from CUHK. I appreciate his invaluable information and his relentless efforts in helping me with my job search. I thank my undergraduate advisor, Professor Ruimin Li, from THU, who guided me in writing my first research paper. I extend my thanks to my coauthors, Mengxin Wang, Haoting Zhang, Donglin Zhan, Jinghai He, and Yifan Zheng, for their collaborative efforts and contributions to our joint work.

I am profoundly grateful for the companionship of so many friends throughout my PhD journey. My heartfelt thanks go to each member of Professor Shen's group, including Junyu Cao, Kexin Cao, Ying Cao, Jingchuan Chen, Ilgin Dogan, Amy Guo, Yiduo Huang, Hansheng Jiang, Shunan Jiang, Shaochong Lin, Mo Liu, Sheng Liu, Meng Qi, Shuo Sun, Mengxin Wang, Xin Weng, Lingfei Zhong, Pengxiang Zhou. I cherish the gatherings and leisure activities shared with my friends from the CEE department, especially during the early years of my PhD, including Di Chen, Sili Kong, Junren Liao, Qianhua Luo, Haobin Tang, Yuan Tian, Mingyuan Yang, Liang Zhang. I am equally thankful for the acceptance and warm companionship of my friends from the IEOR department in the later stages of my studies, including Ziyang Liu, Runhan Xie, Donghao Ying. Special thanks to Kaijing Ding, my roommate of three years, for the countless enjoyable moments we have shared. I am deeply appreciative of Zhe Fu, whose inspiration and consideration have enriched my life significantly, providing support in every small detail of daily living and personal growth. I am indebted to Hansheng Jiang for her hospitality, delicious meals, and valuable information. I thank Haoting Zhang for his encouragement and thoughtfulness, which have been instrumental in helping me navigate academic challenges, which have helped me navigate academic challenges as we struggled and strived together. My gratitude extends to Lingfei Zhong for her daily support and assistance with various personal matters. I also value the stimulating conversations and gatherings with Amy Guo, Tianyi Lin, Shuo Sun, and Jingxu Xu. Moreover, I am grateful for the enduring support and companionship of my friends from earlier life stages. Despite the

physical distance, special thanks to Xin Bao, Tonghui Liu, Lin Ma, Qinglin Ouyang, Lijun Wang, Linyu Xu, and Shuwen Zhang for their continued friendship and encouragement.

Finally, I would like to express my deepest gratitude to my family. I thank my parents for always being my safe harbor and providing me with their unwavering support. I am grateful to my grandparents for their constant concern, pride in my achievements, and understanding despite our long separations. I also thank my cousin, cousin-in-law, and their two cats for their care and companionship during the times we were away from our family.

Chapter 1

Introduction

Individuals interact in real-world to build relationships, share information, and create impacts. These interactions are effectively encapsulated within the framework of social networks, which serve as a fundamental basis for understanding and analyzing human behavior. In this dissertation, I will cover a series of studies that lie at the intersection of social network analytics and operations management, with a keen focus on resolving the inherent complexities that arise from this integration.

The primary challenge encountered in this integration is the intricate nature of network behaviors, which tend to conflict with the need for simplicity and tractability in operational decisions. Typically, the network effect, which propagates across the network structure and evolves over time, has a complicated form or, even worse, lacks closed-form expression. This complexity poses significant hurdles in accurately assessing the impact of social networks on individual and collective behavior, and subsequently integrating these insights into practical optimization frameworks. In the digital age, where online social networks dominate and are in a constant state of expansion, these challenges become even more pronounced, necessitating innovative approaches to manage and leverage the vast amount of data these networks generate.

This dissertation aims to bridge these gaps by developing unifying frameworks to provide simple, efficient, and high-quality solutions to operational problems tied to social networks. At its core, the aim is to contribute to our intellectual understanding of the following questions:

- How can we more efficiently evaluate the impact of network diffusion?
- How can network diffusion be integrated into operational decisions?
- How can we identify and learn about the diffusion effect using the limited data available?

In Chapter 2, we consider an in-depth analysis of the candidate generation and promotion optimization (CGPO) problem for online content platforms, emphasizing the crucial

role of the diffusion effect in shaping content promotion strategies. Content promotion policies are crucial for online content platforms to improve content consumption and user engagement. Unlike conventional methods that often sideline the complex dynamics of user interactions, this study introduces a sophisticated diffusion model that seamlessly integrates these elements with platform promotion strategies. Based on this diffusion model, the CGPO problem can be formulated as a mixed-integer program with nonconvex and nonlinear constraints, which is proved to be NP-hard. Despite the complexity of the CGPO problem, we prove the submodularity of its objective function, which enables us to find an efficient $(1 - 1/e)$-approximation greedy solution. Additionally, we investigate methods for estimating the diffusion model parameters using available online platform data and introduce novel double ordinary least squares (D-OLS) estimators. We demonstrate that the D-OLS estimators are consistent and have smaller asymptotic variances than traditional OLS estimators. By utilizing real data from a large-scale video-sharing platform, we show that our diffusion model effectively characterizes the adoption process of online content. Compared to the policy implemented on the platform, our proposed promotion policy increases total adoptions by 49.90%. This work highlights the essential role of diffusion in online content and provides actionable insights for online content platforms to optimize their content promotion policies by leveraging our diffusion model.

In Chapter 3, we delves into the phenomenon of nonprogressive diffusion, a critical type of social network behavior 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 network metrics are reliable indicators of the FPA scheme's performance.

In Chapter 4, I conclude the dissertation and chart out the prospective avenues for future research. A key focus is on understanding user behavior within social networks, especially in the face of limited data. This ongoing work navigates the complexities of learning user behavior in social networks, where the interdependence of user actions renders traditional machine learning approaches inadequate. We highlight the potential of simulation techniques to generate counterfactual data from network diffusion models, offering a viable strategy for overcoming the limitations of scarce real-world data, and setting the stage for future explorations in network behavior analysis.

1.1 Main Contribution

The primary goal of this dissertation is to present a perspective on the integration of social network analytics with operations management. The details are elaborated throughout this dissertation, drawing principally on the following publications authored by myself and my colleagues (Lin et al. 2024, 2022). The key contributions are outlined as follows:

Content Promotion for Online Content Platforms with the Diffusion Effect (Chapter 2)

- **A diffusion model for online content.** Our key contribution is the P-BDM for depicting online content diffusion, which takes into account the promotion policy and timeliness of content diffusion. Theoretically, the P-BDM characterizes the relationship between the platform's promotion decisions and the diffusion process, providing a concise way to optimize promotion policy. Practically, the P-BDM demonstrates effective alignment with real adoption patterns derived from an online video dataset.

- **Formulation and algorithmic design for the CGPO problem.** Under the P-BDM, we represent the CGPO problem as a challenging mixed-integer optimization problem that involves complex dynamics of content adoption. Despite the presence of nonconvex and nonlinear constraints, as well as its proven NP-hardness, we identify a crucial property known as the "monotonic property with nested sets". This leads to the establishment of the submodularity of the problem objective. Leveraging this property, we propose the accelerated greedy algorithm (AGA), building upon the well-known greedy algorithm for submodular maximization problems (Nemhauser et al. 1978) with a $(1 - 1/e)$-approximation ratio.

- **New estimation approach for the P-BDM.** We introduce the double ordinary least squares (D-OLS) method for estimating P-BDM parameters, taking advantage of the online platform's ability to distinguish different types of adopters. The D-OLS estimators are straightforward to compute and possess desired statistical properties. We show they yield smaller asymptotic variances compared to the OLS estimators and demonstrate their robustness when the promotion policy is endogenous with diffusion dynamics theoretically and numerically.

- **Extensive numerical experiments with real data.** We validate our models and algorithms using a large-scale real-world data set from an online video-sharing platform. Our observations are threefold. First, the promotion and diffusion coefficients for online content are negatively correlated, highlighting the complexity of the CGPO problem. Second, the policy generated by the AGA effectively strikes a balance between incorporating the diffusion effect and updating the promotion policy. The success of the AGA provides invaluable insights, such as the emphasis on the promotion effect over the diffusion effect and the distinction of promotion strategies for various content

based on their respective promotion and diffusion coefficients as well as content lifetime. Third, the AGA policy significantly outperforms the benchmark policy that disregards the diffusion effect, with an improvement of at least 49.90%.

Nonprogressive Diffusion on Social Networks: Approximation and Applications (Chapter 3)

- **Provable approximation scheme for a general 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.

Chapter 2

Content Promotion for Online Content Platforms with the Diffusion Effect

2.1 Introduction

In recent years, online content platforms such as TikTok and Instagram have achieved considerable success in parallel with the proliferation of social media. These platforms offer various forms of online content, including reviews, blogs, and videos, with the content serving as virtual products to attract users. However, several unique features of online content platforms set them apart from traditional online retailers. (i) *Platform objective*: While retailers aim to maximize the revenue obtained from selling products, content platforms aim to maximize the engagement of users and the impact of their content. For example, the total number of content clicks, which we adopt as the key metric in our work, is widely recognized as a vital metric for platform operations (Su and Khoshgoftaar 2009) to measure content consumption. (ii) *Scale*: The amount of content is orders of magnitude larger than the number of products on an online retail platform. New content is generated significantly faster than new products introduced on a retail platform. For instance, Amazon sells 12 million products in total (AMZScout 2021), while YouTube has more than 500 hours of videos (YouTube 2021) uploaded per minute, with an average video length of 11.7 minutes (Statista 2021). A rough estimate implies millions of new videos are uploaded every day. (iii) *User consumption behavior*: Unlike in an e-commerce setting where users directly search for a product of interest, online content platform users rely heavily on platform promotions and/or friends sharing content on a social network as ways to overcome information overload (Anandhan et al. 2018).

Therefore, online platforms actively promote content to users and foster an environment where users are encouraged to share interesting content. This leads to the phenomenon of content diffusion, wherein the content spreads to a larger audience beyond the scope of

direct platform promotion. Consequently, a greater number of users have the opportunity to discover and consume the content, thereby significantly amplifying user engagement. Two examples of diffusion are provided below.

Example 2.1.1 (The diffusion of one piece of content) *Some articles that announce breaking news or tell good stories might go viral on the platform when many users cascadingly re-post them. For instance, in 2020, a local news story about two missing children in Florida netted almost 3.5 million shares on Facebook (FOX32 2020).*

Example 2.1.2 (The diffusion of a trend) *Some content that is of the same category, similar in nature, or with homogeneous topics is usually associated with trending hashtags and/or headlines. As more users are aware of and engaged in this trend, the related content also becomes popular. For instance, the hashtag #squidgame has garnered 72.4 billion views on TikTok (TikTok 2021). Numerous TikTok users became part of the trend, and an enormous amount of content, including reenactments of the game, makeup looks, and Halloween costumes inspired by the TV show "Squid Game", was produced and viewed on TikTok.*

The online content platform's business model prompts the following research question:

> *How can a promotion policy be designed that selects a small subset of content from the enormous corpus to display to users in a way that maximizes the total content clicks?*

The existing literature often prioritizes maximizing the number of clicks through direct promotion, while neglecting the diffusion effect. It advocates promoting content with a high historical click-through rate in hopes of attracting more direct clicks (Feng et al. 2007, Liu et al. 2010). However, this type of promotion may overemphasize content that is already popular, creating a scenario where a limited set of content is continuously promoted, reducing overall content diversity. This "rich get richer" phenomenon can negatively impact user engagement and satisfaction (Vahabi et al. 2015), as users might be unable to discover new content. The challenge of the promotion policy lies in balancing between promoting trending content for immediate gain and promoting diverse content to stimulate user engagement for long-term platform sustainability. To the best of our knowledge, these trade-offs remain largely unexplored in the literature. The diffusion effect serves as one of the major sources of indirect gain. In this study, we aim to fill this gap by developing a diffusion-based promotion policy for online content platforms.

Machine learning-based promotion strategies in practice typically involve two stages: candidate generation and promotion optimization (Davidson et al. 2010, Covington et al. 2016). The candidate generation stage selects a small subset of content from a large corpus, while the promotion optimization stage allocates a limited promotion budget to each candidate content piece. This two-stage procedure balances computational efficiency and focuses the platform's attention on a small portion of content that can potentially generate high rewards. We follow this framework and introduce two distinct features that differ from

previous machine learning-based strategies. First, we incorporate the diffusion effect into our promotion policy. Second, we recognize that the candidate set selection can impact the optimal allocation of user attention to the content. We therefore carefully consider these two stages together to maximize the total number of clicks, rather than treating them as separate machine learning tasks as in previous literature. This leads to the candidate generation and promotion optimization (CGPO) problem that we focus on in this paper.

A central piece of the CGPO problem is the promotion Bass diffusion model (P-BDM) that we propose to characterize the diffusion process of online content. The P-BDM is adapted from the well-known Bass diffusion model (BDM, Bass 1969) and inherits its innovative and imitative effects, which we interpret as two sources of user clicks on content platforms: platform promotion and diffusion through user sharing, respectively. We show that the BDM is not suitable for modeling the diffusion process of online content because it fails to account for platform promotion policy and the timeliness of content diffusion. In contrast, the P-BDM explicitly captures both and provides an excellent fit for a real-world online content platform. Based on the P-BDM, we offer a set of complete solution techniques for the online content promotion problem. Firstly, we integrate the candidate generation and promotion optimization problems into a succinct mixed-integer program, allowing us to obtain high-quality approximate solutions with performance guarantees. Secondly, leveraging the high-granularity data commonly available on online platforms, we design a novel estimation method for the parameters in the P-BDM. Lastly, our modeling framework, optimization algorithm, and estimation method are demonstrated to be effective through counterfactual analyses based on real online content data.

The remainder of this chapter is structured as follows: In Section 2.2, we review the related literature. In Section 2.3, we discuss the formulation of the P-BDM. In Section 2.4, we formulate the CGPO problem and propose the AGA for solving it. In Section 2.5, we discuss the estimation issues for the P-BDM and propose the D-OLS method. Section 2.6 presents our numerical studies based on real-world data, followed by concluding statements in Section 2.7.

2.2 Literature Review

As we discussed earlier, promotion and diffusion are two primary drivers of rewards for online content platforms. Therefore, we focus our review on promotion policies and diffusion effect studies. An active stream of literature is about recommender systems (RSs), which focuses on investigating the connections between users and content. Various recommendation algorithms (Kitts et al. 2000, Covington et al. 2016) have been proposed to evaluate the probability of users clicking on a particular item which characterizes the immediate interactions between the platform and users. Some recent work (Lu et al. 2014, Besbes et al. 2016) demonstrates that maximizing the immediate item relevancy does not align with utility maximization. The reasons are various, and one of the most important issues is the consequent diffusion within the social network. This implies that the adoptions are not only

from directly targeted users but also from those who are influenced by them. Few studies incorporate this effect. Vahabi et al. (2015) are the first to mention that the social network can empower the utility maximization of RSs. They propose a social-diffusion-aware RS that can efficiently use recommendation slots and enhance the overall performance. Our work substantially differs from Vahabi et al. (2015). While they utilize a personalized recommendation scheme with a hard constraint that prevents neighbors from receiving identical content—we instead aim to characterize the diffusion trend across the population and find an optimal promotion policy.

We remark that our work does not emphasize understanding the relationship between users and content from a machine-learning perspective, as in the aforementioned literature. Typically, these works consider content recommendations for each user individually. Rather, we take a holistic approach and study the problem from an operations perspective. Our objective is to maximize the total clicks by modeling the whole problem as a diffusion process within the population and generating high-quality solutions using combinatorial optimization techniques.

To understand how user interactions influence adoption, many diffusion models have been proposed in the literature. Pioneered by Bass (1969), the BDM has become the most widely used model for new products, capturing the adoption trend with parsimonious differential equations. A sequence of work (Easingwood et al. 1983, Norton and Bass 1987) extends the BDM by incorporating different dynamics, such as nonuniform influence and multiple generations. The BDM has achieved tremendous success in predicting the adoption of a variety of products, including consumer durable goods, medical innovations (Sultan et al. 1990), and information technology innovations (Teng et al. 2002). Despite its long history, the BDM is also frequently applied in novel contexts (Jiang and Jain 2012, Agrawal et al. 2021). Our diffusion model extends the BDM by integrating promotion policy and the timeliness of online content diffusion.

The BDM, along with all the other models previously referenced, focuses on the global diffusion effect. In particular, each user is influenced by a universal diffusion effect, namely the overall adoption level within the entire market. In contrast, some research has also explored diffusion in the context of social networks, where each user is influenced only by their local neighborhood. This type of diffusion model can provide a more granular representation of the unique diffusion effect experienced by each user. The independent cascade model (Goldenberg et al. 2001) and linear threshold model (Granovetter 1978) are two such fundamental models that incorporate network structures. Kempe et al. (2003) then models the influence maximization problem as an algorithmic problem, aiming to identify the optimal subset of seed users that could trigger the maximum adoptions. For other applications, we refer readers to review papers (Kiesling et al. 2012, Zhang and Vorobeychik 2019). However, the diffusion process within a network is not easy to quantify through a simple formula; hence, the market characterization relies heavily on simulation techniques, making the subsequent optimization problem time-consuming to solve. Given the limitations, our work focuses on the global diffusion effect to maintain simplicity and efficiency in optimization problems.

Another branch of study relevant to our work is revenue management for online retailers. While users exhibit different behaviors, the process of candidate selection and promotion optimization shares some similarities with assortment and pricing problems. Golrezaei et al. (2020b) and Chen and Shi (2019) present the inventory and pricing strategies for strategic customers, who exhibit similar user behavior as online content platforms. Moreover, recent works also consider the network effect in operations management problems. Hu et al. (2016) consider the case where purchase decisions can be influenced by earlier purchases. Du et al. (2016), Wang and Wang (2017) propose a variant of the multinomial logit model incorporating the network effect in an assortment optimization problem. Follow-up works (Nosrat et al. 2021, Chen and Chen 2021) also involve different choice models. This line of research inspires us to consider similar problems on online content platforms.

2.3 Promotion Bass Diffusion Model

In this section, we introduce the promotion Bass diffusion model (P-BDM) to capture the adoption process of online content. We begin by highlighting a common issue of the Bass diffusion model to model the real-world online content adoption process. It motivates us to develop a new model, P-BDM, that incorporates the platform's promotion decisions, which serves as the foundation for our subsequent optimization and analysis.

2.3.1 Background and Motivation from a Large-Scale Video-Sharing Platform

We begin with a brief overview of the Bass diffusion model (BDM), which is a widely used model for describing the adoption process of new products in a population over time. The basic premise of the BDM is that adopters can be classified into two types: *innovators* and *imitators*. Innovators are individuals who independently decide to adopt a product, while imitators are influenced by those who have already adopted it. In the context of online content, we view the clicks on a content piece as adoptions. For this reason, we use the terms "click" and "adoption" interchangeably throughout.

The discrete-time BDM models the adoption process of a product over a discrete finite time horizon $t = 1, 2, \ldots, T$ in a market of population size m. The initial number of adopters is denoted by A_0, and the number of new adopters at each time period t is given by:

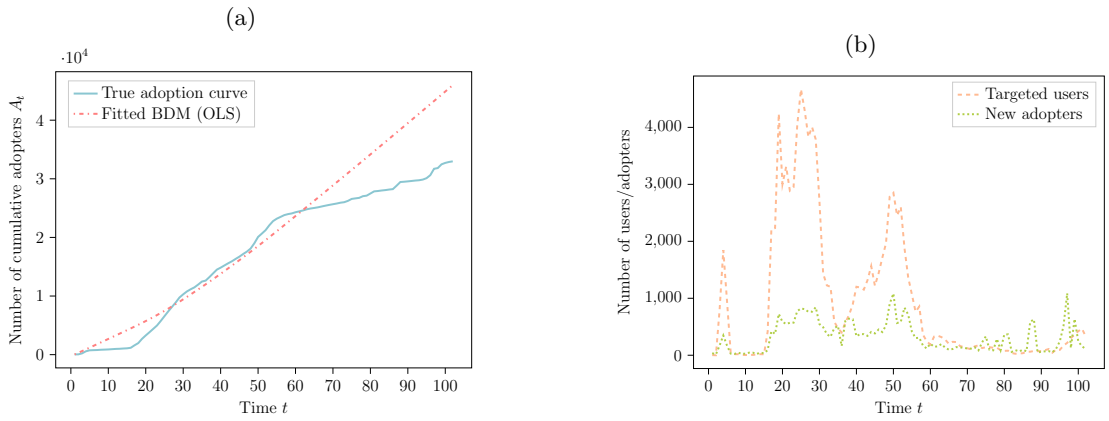
$$a_t = \left(p + q\frac{A_{t-1}}{m}\right)(m - A_{t-1}), \quad \forall\, t = 1, \ldots, T, \tag{2.1}$$

where $A_{t-1} = A_0 + \sum_{\tau=1}^{t-1} a_\tau$ represents the cumulative number of adopters up to time period $t-1$, and p and q are the innovative and imitative coefficients, respectively. In particular, $(p+qA_{t-1}/m)$ corresponds to the adoption rate of the non-adopters at period t. This indicates that the adoption behavior at time t is jointly influenced by two forces: the innovative effect

p and the imitative effect qA_{t-1}/m, which is proportional to the cumulative number of adopters. To ensure the discrete-time BDM is well-defined, it is commonly assumed that $p \geq 0$, $q \geq 0$, and $p + q \leq 1$.

Whereas it seems intuitive to apply the BDM to model the adoption of online content, empirical evidence may suggest otherwise. Particularly, we analyze the clickstream data from a large-scale video-sharing platform. For a detailed introduction about the platform and the data, please refer to Section 2.6.1. In Figure 2.1, we use a single video to showcase the typical pattern of the content diffusion process.

Figure 2.1: Illustration of Adoption Curves and the Corresponding Fitted BDM Curve for an Example Video



Notes. To ensure data anonymity, we have scaled the y-axis using a randomly selected number. (a) Cumulative adopters. (b) New adopters and targeted users.

In Figure 2.1a, we present both the actual cumulative adoption data over 102 periods and the fitted BDM curve. For clarity, we limit the range of these curves to the time frame within our observation. A comprehensive discussion on the fit of the BDM to online content adoption data, including the full trajectory of the fitted BDM curve, is provided in Appendix A.2.1.1.

A detailed explanation of our fitting method can be found in Section 2.5.1. Specifically, we first estimate the parameters of the BDM using data from the initial 60 periods, then use these parameters to generate a fitted curve for the full 102 periods. We evaluate the fitness and predictive power of this curve in two parts: from time period 0 to 60 and from 60 to 102. In the first part, although the fitted curve largely reflects the overall trend, it fails to capture the subtle swings of the curve from time to time. In the second part, the fitted BDM continues to predict a steady growth of new adopters, whereas the actual data shows a much slower rate of adoption. As a result, the predicted adoption number deviates

significantly from the actual value at the end. These observations suggest that the BDM may not be able to provide an appropriate description for the diffusion of online content. One of the significant factors contributing to this inconsistency is the BDM's assumption that all non-adopters will be impacted by the innovation effect, as outlined in (2.1). However, this is unlikely to hold for online content due to limited user time and targeted promotion strategies. Figure 2.1b further supports this claim by showing a strong correlation between targeted users and new adopters at each time period (the Pearson correlation coefficient is $\rho = 0.748$).

The discrepancy between the BDM and actual adoption data is not unique to this particular example; it is commonly observed in the data. This suggests that platform-controlled promotion plays a vital role in driving adoption and motivates us to develop a new diffusion model tailored to online content that incorporates platform promotion policy.

2.3.2 The Promotion Bass Diffusion Model

The key message conveyed in the previous analysis is that one must take into account the platform's *promotion policy* to capture the adoption patterns of online content. This motivates our promotion Bass diffusion model (P-BDM). It builds on the notation of the BDM and adapts it to the specifics of online content. It aims to model the adoption process of online content over a finite time horizon $t = 1, 2, \ldots, T$ in a market of size m. The model posits that the adoption of a content piece is driven by two forces: (i) *promotion effect*, which reflects the intrinsic preference of users towards a content piece, specifically how likely a user is to adopt a content piece as an individual when it is promoted by the platform, and (ii) *diffusion effect*, which represents the influence of the adopted population on others, i.e., the likelihood that a user will adopt a content piece that is shared by other adopters. The promotion and diffusion effects are the counterparts of innovative and imitative effects in the BDM; we, therefore, use these terms interchangeably in the following discussions. In a similar vein, we define the promotion coefficient p and diffusion coefficient q to characterize these two effects. Consistent with the BDM setting, we assume that $p \geq 0$, $q \geq 0$, and $p + q \leq 1$.

The P-BDM incorporates the platform's promotion policy as a new variable, denoted by $\boldsymbol{x} = (x_t : t = 1, 2, \ldots, T)$, which represents the fraction of users in the market that receive the promotion at each time period. For mathematical convenience in the subsequent formulation, we define this promotion fraction over the entire market size rather than over the remaining non-adopters, although these two definitions can be converted to each other as needed. Specifically, there are in total mx_t users receiving the promotion at time t. The platform does not promote any content piece to users who have already adopted it. Thus, the promotion fraction x_t is upper bounded by the fraction of the remaining non-adopters in the market at time $t - 1$ (i.e., $x_t \leq 1 - A_{t-1}/m$). Denoting A_0 as the initial number of

adopters, the P-BDM assumes that the number of new adopters at time t is

$$\begin{aligned} a_t = & \underbrace{\left(p + q\frac{A_{t-1}}{m}\right) m x_t}_{\text{Direct adopters}} + \underbrace{q\frac{A_{t-1}}{m}(m - A_{t-1} - m x_t)}_{\text{Indirect adopters}} \\ = & \underbrace{p m x_t}_{\text{Promotion effect}} + \underbrace{q\frac{A_{t-1}}{m}(m - A_{t-1})}_{\text{Diffusion effect}} . \end{aligned} \tag{2.2}$$

In the P-BDM dynamics (2.2), we categorize adopters based on whether they receive the promotion or not. For those not exposed to promotion, categorized as "indirect adopter", their adoption rate reflects only the diffusion effect and is expressed as qA_{t-1}/m. For those who receive the promotion, categorized as "direct adopters", their adoption rate is increased by the promotion effect, making it $(p + qA_{t-1}/m)$. It is worth noting that the BDM is a special case of the P-BDM where all non-adopters receive the promotion at each time period (i.e., $x_t = 1 - A_{t-1}/m, \forall t = 1, 2, \ldots, T$).

In practice, we also observe a *time decay effect* in online content diffusion, which can be attributed to the limited lifespan of content and the diminishing incentives for adopters to share it over time. To account for this timeliness feature, we incorporate a time-decay multiplicative factor γ for better alignment of the diffusion model with real-world data. We explain this approach in more detail in Appendix A.2.1.2. It is important to note that we consider the time-decay factor as an external influence, which means it does not complicate the theoretical analysis of the promotion optimization problem. Therefore, we assume $\gamma = 1$ until the discussion of numerical experiments (i.e., Section 2.6), where we will explore the effect of time-decay factor γ further.

2.4 Optimizing Content Adoptions

In practice, online platforms frequently undertake the mission of efficiently selecting and spotlighting featured content. This content, a distinct subset that the platform deliberately highlights or promotes, is usually selected due to its high quality and potential for stimulating trends. For instance, on the platform with which we collaborate, algorithms select high-quality content regularly for inclusion in the "trending video pool". With the intent of stimulating diffusion and creating buzz, this content is then blended with other material—content selected based on user interests by machine learning algorithms, advertisements, and more—and displayed to users. The content display process on this platform is representative, including two stages (Davidson et al. 2010, Covington et al. 2016): candidate generation and promotion optimization. The former involves the selection of a promising subset of content from the overall corpus, while the latter necessitates the platform's allocation of its limited promotional resources among the selected candidates. This process, driven by machine learning and reliant on decentralized algorithms, doesn't account for overall diffusion—an

objective of the trending video pool. In contrast, we formulate the content generation and promotion optimization (CGPO) problem as an optimization task to incorporate the diffusion effect into content promotion, using the P-BDM as a basis.

2.4.1 The Content Generation and Promotion Optimization Problem

We consider a platform with a fixed content corpus $\mathcal{V}$, operating within a market of unchanging size m. The platform can select up to K candidate content pieces. Its objective is to maximize total content adoptions over a fixed planning interval of length L given a promotion budget C. To achieve this, the platform needs to determine not only the promotion fraction for each content piece but also coordinate the timing of the promotion.

In line with the two-stage process, the platform first selects a subset $U \subseteq \mathcal{V}$ with no more than K candidates. This cardinality constraint reflects the natural upper bound on the size of the trending or featured video pool. Then, the platform determines the promotion policy $\boldsymbol{x} = (x_{v,t} : v \in U, t = 1, 2, \ldots, L)$ for the candidate set U. Here, each candidate $v \in U$ is promoted to a fraction $x_{v,t}$ of users at each time $t = 1, \ldots, L$. Subsequently, the CGPO problem is formulated as

$$\max_{U \subseteq \mathcal{V}:\ |U| \le K} \quad R(U; C) + R(\mathcal{V} \setminus U; 0)\,, \tag{2.3}$$

where $R(U; C)$ denotes the maximum total adoptions for the candidate set U, achievable by optimizing the promotion policy $\boldsymbol{x}$ within the promotion budget C. Similarly, $R(\mathcal{V} \setminus U; 0)$ denotes the total adoptions of the remaining set $\mathcal{V} \setminus U$ with budget 0. This is equivalent to none of the content in set $\mathcal{V} \setminus U$ being promoted. Notice that, given any $U \subseteq \mathcal{V}$ and promotion budget C, $R(U; C)$ can be defined as the optimal value of the following promotion optimization (PO) subproblem:

$$\max_{\boldsymbol{x} \ge \boldsymbol{0}, \boldsymbol{A}_{U,1:L}} \quad \sum_{v \in U} A_{v,L} \tag{2.4a}$$

$$\text{s.t.} \quad A_{v,t} = A_{v,t-1} + p_v m x_{v,t} + \frac{q_v}{m} A_{v,t-1}(m - A_{v,t-1}), \quad \forall v \in U \ \ \forall t = 1, \ldots, L, \tag{2.4b}$$

$$x_{v,t} \le 1 - \frac{A_{v,t-1}}{m}, \qquad \forall v \in U \ \ \forall t = 1, \ldots, L, \tag{2.4c}$$

$$m \sum_{t=1}^{L} \sum_{v \in U} x_{v,t} \le C. \tag{2.4d}$$

In this subproblem, $\boldsymbol{A}_{v,:} = (A_{v,t} : t = 0, 1, \ldots, L)$ denotes the cumulative adopters for $v \in U$ at each time and initial adoption $\boldsymbol{A}_{U,0} = (A_{v,0} : v \in U)$ is given. The objective (2.4a) represents the total adoptions of all candidates in set U at the end of the L-period planning interval; (2.4b) mandates the cumulative adopters follow the P-BDM diffusion dynamics, as defined in (2.2); (2.4c) ensures that the number of users receiving the promotion does

not exceed the remaining non-adopters; (2.4d) ensures that across the L periods a total promotion budget C on the number of impressions used for promoting these videos in the featured video pool. The rest of the platform capacity is reserved for other purposes such as displaying videos selected based on user interests and advertisements. In addition, we use $\bar{C} := C/mL$ to indicate the average promotion budget per user per time period, which can be an arbitrary given constant.

In a similar vein, we can define $R(\mathcal{V} \setminus U; 0)$. Specifically, it can be calculated according to the P-BDM diffusion dynamics (2.2) when $x_{v,t}$ is set to 0 for all $v \in \mathcal{V} \setminus U$ and $t = 1, 2, \dots L$.

We remark on the following before solving the CGPO problem. First, on the notation side, we use bold notation to denote collections of specific variables for a set of content pieces within a certain time period, in vector or matrix form. For example, $\boldsymbol{x} = (x_{v,t} : v \in \mathcal{V}, t = 1, \dots, L)$, $\boldsymbol{x}_{v,:} = (x_{v,t} : t = 1, \dots, L)$, and $\boldsymbol{x}_{U,t} = (x_{v,t} : v \in U)$. We use $i : j$ to denote a slice of the vector or matrix ranging from index i to j, where $i, j \in \mathbb{Z}_+$. For example, $\boldsymbol{x}_{v,2:L} = (x_{v,t} : t = 2, \dots, L)$. Second, one ultimate goal of content platforms is to maximize total adoptions over a long time horizon of T periods ($T \gg L$). Crafting a "true" optimal promotion policy for this entire period is challenging due to the dynamic platform environment, including regular updates to the content corpus and market size variations. However, content diffusion on platforms typically outpaces these environmental changes (Graffius 2022). As such, it is reasonable to design promotion policies periodically for a short period, in which the environment is relatively stable. The CGPO problem, therefore, seeks to identify such a dynamic policy within a L-period window. We can re-optimize it periodically to account for the environmental changes over the extended time horizon.

2.4.2 Promotion Optimization Given the Content Set

The CGPO problem inherently comprises two stages, namely, candidate generation (CG) and promotion optimization (PO). The CG stage, as represented in problem (2.3), is a combinatorial optimization problem that embeds the PO stage, as shown in problem (2.4). The primary challenge in solving the CGPO problem stems from the implicit interaction among content pieces, which is a consequence of the budget constraint (2.4d). This constraint not only leads to different selections of content but also necessitates corresponding adjustments in how the promotion budget is allocated to the selected content. These adjustments lead to a variety of diffusion outcomes. The complexity of this problem is formally captured in the following theorem:

Theorem 2.4.1 (NP-Hardness) *The CGPO problem* (2.3) *is NP-hard.*

For a detailed proof of Theorem 2.4.1, please refer to Appendix A.1.1.1. In this section, we first focus on the PO stage for a given candidate content set $U \subseteq \mathcal{V}$. We show how to solve the PO subproblem optimally and identify the key property that helps solve the entire CGPO problem.

Given set U, the PO problem (2.4) remains difficult to solve due to its non-convex nature. In the following, we first perform convex relaxation and show that the PO problem

is equivalent to its relaxed problem. Then, we highlight a critical ingredient in solving the relaxed problem, which is also essential to solving the entire CGPO problem.

2.4.2.1 Convex Relaxation

The nonconvexity of problem (2.4) originates from the set of equality constraints (2.4b), which include a quadratic term of $\boldsymbol{A}$ on the right-hand side. To transform this nonconvex feasible region into a convex one, we relax (2.4b) as inequalities as follows:

$$A_{v,t} \leq A_{v,t-1} + p_v m x_{v,t} + \frac{q_v}{m} A_{v,t-1}(m - A_{v,t-1}), \quad \forall v \in U \;\; \forall t = 1, \ldots, L\,. \tag{2.5}$$

We denote the relaxed problem as PO-CR, which uses (2.4a) as the objective and includes (2.5), (2.4c), and (2.4d) as the constraints. The PO-CR problem is a convex optimization problem and thus can be handled by commercial solvers. Any optimal solutions to the PO-CR problem serve as upper-bound solutions to the PO problem (2.4). Moreover, as we illustrate in Theorem 2.4.2 below, the PO-CR problem is in fact equivalent to the original PO problem (2.4).

Theorem 2.4.2 (Relaxation) *The PO problem* (2.4) *and relaxed problem PO-CR are equivalent.*

We remark that the equivalence is non-trivial because the decision variables $\boldsymbol{A}$ and $\boldsymbol{x}$ have opposing relationships in constraints (2.4b) and (2.4c). Specifically, increasing $A_{v,t}$ seems to increase the objective value due to (2.4b) but it lowers $x_{v,s}$ for $s \geq t+1$ due to (2.4c). To establish the equivalence, we show that the optimal solutions of the PO-CR problem are feasible solutions to the PO problem (2.4). The key intuition is that, under P-BDM dynamics, there is no benefit in "holding back" realized adoptions as a larger unadopted population for future promotion. In other words, achieving equality in constraint (2.4b) is more beneficial than maintaining the constraint as a strict inequality for a larger upper bound of $\boldsymbol{x}$ in constraint (2.4c). Complete proofs are in the appendix.

With the PO-CR problem at hand, we can directly find the optimal promotion policy for any given candidate set $U \subseteq \mathcal{V}$ using commercial solvers. However, to tackle the CGPO problem (2.3) as a whole, we need to utilize the optimality condition of the PO-CR problem, as detailed in Section 2.4.2.2, and then establish its link to the outer CG problem. This is accomplished by a reformulation that solely uses the promotion fraction $\boldsymbol{x}$ as the decision variable.

2.4.2.2 Monotonic Property with Nested Sets of PO Problems

Given that the adoption number $\boldsymbol{A}_{v,1:L}$ intrinsically depends on $\boldsymbol{x}$ and $\boldsymbol{A}_{v,0}$, we logically reformulate the PO-CR problem into a convex program that solely involves the promotion fraction $\boldsymbol{x}$:

$$\max_{\boldsymbol{x} \geq \boldsymbol{0}} \quad \sum_{v \in U} f_v(\boldsymbol{x}_{v,:}) \tag{2.6a}$$

$$\text{s.t.} \quad m \sum_{t=1}^{L} \sum_{v \in U} x_{v,t} \leq C, \tag{2.6b}$$

$$x_{v,t} \leq 1 - \frac{A_{v,0}}{m}, \forall\, v \in U \quad \forall\, t = 1, \ldots, L, \tag{2.6c}$$

where for all $v \in U$, function $f_v(\boldsymbol{x}_{v,:})$ is defined as

$$f_v(\boldsymbol{x}_{v,:}) := \max_{\boldsymbol{A}_{v,1:L}} \quad A_{v,L} \quad \text{s.t.} \quad (2.5), (2.4c). \tag{2.7}$$

This reformulation utilizes a series of black-box functions, f_v for each $v \in U$, to evaluate the adoptions of content v under a given promotion policy. To ensure $f_v(\boldsymbol{x}_{v,:})$ is well-defined, we include a set of redundant constraints (2.6c) which ensures that problem (2.7) always has a feasible solution. We elaborate on the rationale behind these constraints and the process of constructing a feasible solution to problem (2.7) for any given policy in Appendix A.1.1.3. This reformulation creates a crucial link between the PO problem for a specific candidate set and the CG problem, which encompasses a set of PO problems for any possible candidate sets. It naturally divides the PO problem into two steps: evaluation and optimization. Evaluation is through functions f_v, separable for each content piece $v \in \mathcal{V}$ and independent of the chosen set U. For example, if we consider two different candidate sets U_1 and U_2 where $v \in \mathcal{V}$ is included in both sets, function f_v will be consistently defined across both PO problems. Moreover, the reformulation preserves the convexity of the PO-CR problem, as we can demonstrate that f_v is a concave function for each $v \in \mathcal{V}$ in Lemma 2.4.1.

Lemma 2.4.1 (Concavity) *For any $v \in \mathcal{V}$, $f_v(\boldsymbol{x}_{v,:})$ is a concave function for $\boldsymbol{x}_{v,:} \in [0, 1 - A_{v,0}/m]^L$.*

Lemma 2.4.1 derives from the fact that problem (2.7) is a convex program. Based on this, we illustrate the optimality condition of the PO problem using the Lagrangian multiplier, which serves as a stepping stone to solving the entire CGPO problem. Specifically, we dualize the reformulation (2.6) as dual problem (2.8), with θ being the Lagrangian multiplier for constraint (2.6b):

$$\min_{\theta \geq 0} \quad \sum_{v \in U} h_v(\theta) + \theta C\,. \tag{2.8}$$

Here, $h_v(\theta)$ is defined as the optimal value function of the following maximization problem:

$$h_v(\theta) := \max_{\boldsymbol{x}_{v,:} \in [0, 1 - A_{v,0}/m]^L} f_v(\boldsymbol{x}_{v,:}) - \theta m \sum_{t=1}^{L} x_{v,t}\,. \tag{2.9}$$

For any candidate set $U \subseteq \mathcal{V}$, let $\theta^*(U)$ denote the optimal dual variables of dual problem (2.8). In Lemma 2.4.2, we provide a comparison of optimal dual variables for nested candidate sets $U_1 \subseteq U_2 \subseteq \mathcal{V}$.

Lemma 2.4.2 (Monotonic property with nested sets) *For any nested candidate sets $U_1 \subseteq U_2 \subseteq \mathcal{V}$, the optimal dual variables satisfy $\theta^*(U_1) \leq \theta^*(U_2)$.*

Lemma 2.4.2 implies that for nested candidate sets, the optimal dual variables of the larger set will always be greater or equal. This conclusion is grounded in the consistent definition of function f_v across PO problems with different sets. Lemma 2.4.2 not only enables us to efficiently search for the optimal dual solution θ^* without requiring a closed-form expression but also plays a crucial role in proving the submodularity of the GG problem. In particular, we employ this property to show that the marginal gain of the CG problem decreases monotonically as the content set U enlarges.

2.4.3 Candidate Generation

In this section, we address the candidate generation (CG) stage, which aims to select a subset of content pieces that yield the maximum total adoptions. Leveraging the "monotonic property with nested sets" for PO problems derived in the previous section, we approach this combinatorial optimization from another perspective. Instead of directly identifying the optimal candidate set, we focus on the comparison of total adoptions between two nested candidate sets. By comparing the marginal gains of incorporating an additional content piece into nested candidate sets, we show that the objective of the CG problem (2.3) is a monotone submodular set function. This finding enables us to apply the greedy algorithm for submodular maximization to solve the entire CGPO problem, thereby achieving an $(1-1/e)$-approximation. Moreover, we can further accelerate the greedy algorithm by leveraging the "monotonic property with nested sets" again.

2.4.3.1 Submodularity of the CGPO Objective

To verify that the CGPO objective (i.e., $R(U;C)+R(\mathcal{V}\setminus U;0)$) is a submodular set function of $U \subseteq \mathcal{V}$, we need to show that $R(U\cup\{w\};C)+R(\mathcal{V}\setminus(U\cup\{w\});0)-R(U;C)-R(\mathcal{V}\setminus U;0)$ is decreasing in U, for all $w \in \mathcal{V}\setminus U$. With simple algebra, this is equivalent to showing that for any given nested sets $U_1 \subseteq U_2 \subseteq \mathcal{V}$ and $w \in \mathcal{V} \setminus U_2$,

$$R(U_1 \cup \{w\};C) - R(U_1;C) - R(\{w\};0) \geq R(U_2 \cup \{w\};C) - R(U_2;C) - R(\{w\};0)\,. \tag{2.10}$$

The left and right sides of (2.10) represent the marginal gain of adding an additional content piece w to sets U_1 and U_2, respectively. The marginal gain is characterized by the difference between the optimal values of two different PO problems. Direct comparison of two marginal gains is intractable, as the optimal value of PO problem does not have a closed-form expression. To overcome this challenge, in the following, we express the marginal gain as the difference between the optimal values of the same PO problem under different promotion budgets instead.

At a higher level, $R(U;C)$ denotes the optimal adoptions when the promotion budget C is entirely allocated to the candidate set U, whereas $R(U\cup\{w\};C)$ denotes the optimal

adoptions when part of promotion budget $c \in [0, C]$ is allocated to U and the remaining $(C - c)$ is allocated to w. Hence, we can reformulate $R(U \cup \{w\}; C)$ as the optimal value of the following problem:

$$R(U \cup \{w\}; C) := \max_{0 \le c \le C} \left[R(U; c) + R(\{w\}; C - c) \right] , \tag{2.11}$$

where $R(U; c)$ is the maximum total adoptions of set U given a promotion budget c and $R(\{w\}; C - c)$ is the maximal total adoptions of content piece w with a promotion budget $(C - c)$.

Hence, if $c^*(U)$ denotes the optimal promotion budget allocated to set U in problem (2.11), the marginal gain from including content piece w given content set U can be expressed as

$$\begin{aligned} & R(U \cup \{w\}; C) - R(U; C) - R(\{w\}; 0) \\ = & \left[R(U; c^*(U)) - R(U; C) \right] + \left[R(\{w\}; C - c^*(U)) - R(\{w\}; 0) \right] . \end{aligned} \tag{2.12}$$

The marginal gain is decomposed into two parts: the adoption loss of set U due to the cannibalization of new content piece w, and the adoption gain resulting from w. Both parts can be depicted as the difference between optimal values of the same PO problem, with the promotion budget varied. This further enables us to use the Lagrangian multiplier to represent the marginal gain. Analogous to (2.8), we formulate the PO dual problem for set U and promotion budget c as

$$R(U; c) = \min_{\theta \ge 0} \sum_{v \in U} h_v(\theta) + \theta c \, , \tag{2.13}$$

where $h_v(\theta)$ adheres to the same definition in (2.9). By the envelope theorem, we can express the difference between two optimal values as an integral of the optimal dual variable, such as

$$R(U; c) - R(U; 0) = \int_{z=0}^{c} \theta^*(U; z) dz \, , \ \forall \ \theta^*(U; z) \in \Theta^*(U; z) \, ,$$

where $\Theta^*(U; z)$ is the set of optimal dual variables to problem (2.13) when the budget is z.

Consequently, the first term in (2.12) can be represented as

$$R(U; c^*(U)) - R(U; C) = \int_{z=0}^{c^*(U)} \theta^*(U; z) dz - \int_{z=0}^{C} \theta^*(U; z) dz = - \int_{z=c^*(U)}^{C} \theta^*(U; z) dz \, .$$

In a similar manner, we can express the second term in (2.12). Therefore, the marginal gain of adding piece w to the candidate set U can be represented by the optimal dual variables as

$$(2.12) = - \int_{z=c^*(U)}^{C} \theta^*(U; z) dz + \int_{z=0}^{C-c^*(U)} \theta^*(\{w\}; z) dz \, .$$

Hence, we transform the proof of submodularity, which essentially involves comparing the marginal gain of piece w over two nested sets U_1 and U_2, into a comparison between the optimal dual variables of PO problems with two nested sets. This leads us directly to Theorem 2.4.3.

Theorem 2.4.3 (Submodularity) *The CGPO objective, $R(U;C)+R(\mathcal{V}\backslash U;0)$, is a monotone submodular set function with respect to content set $U \subseteq \mathcal{V}$.*

The proof of Theorem 2.4.3 relies on transforming marginal gain and utilizing the "monotonic property with nested sets". The complete proof is included in Appendix A.1.1.4. As a result, the CGPO problem (2.3) can be viewed as a monotone submodular maximization problem with a cardinality constraint.

2.4.3.2 Accelerated Greedy Algorithm

The greedy algorithm (Nemhauser et al. 1978) provides an $(1-1/e)$-approximation for the monotone submodular maximization problem with a cardinality constraint. The algorithm iterates K times, selecting a content piece with the highest marginal gain in each iteration. The greedy algorithm is presented as Algorithm 1.

Algorithm 1: Greedy Algorithm for the CGPO Probem.

1 $U_0 := \emptyset$.
2 **for** $k \in [1, \ldots, K]$ **do**
3 **for** $v \in \mathcal{V} \setminus U_{k-1}$ **do**
4 Solve the PO problem (through its convex relaxation) for set $U_{k-1} \cup \{v\}$.
5 Let $R(U_{k-1} \cup \{v\}; C)$ be the optimal value.
6 **end**
7 $v^* := \arg\max_{w \in \mathcal{V} \setminus U_{k-1}} R(U_{k-1} \cup \{w\}; C) + R(\mathcal{V} \setminus (U_{k-1} \cup \{w\}); 0)$.
8 $U_k := U_{k-1} \cup \{v^*\}$.
9 **end**

Subsequently, we aim to demonstrate that an acceleration of the greedy algorithm can be achieved by exploiting the "monotonic property with nested sets". In each iteration, the greedy algorithm solves PO problems by adding an extra content piece to the selected set, which means it repeatedly solves PO problems for nested sets. Acceleration can be achieved by combining the Lagrangian relaxation technique with the greedy approach. The core idea is to utilize the optimal dual variable values from previously solved PO problems to create a more compact feasible region for subsequent iterations, which deals with expanded candidate sets. We formalize this idea as the accelerated greedy algorithm (AGA), which is detailed as follows.

Accelerated Greedy Algorithm:
For Line 4 in Algorithm 1,

(i) At iteration k, record the optimal Lagrangian dual variable when solving the PO problem with set $U_{k-1} \cup \{v\}$ as $\theta^*(U_{k-1} \cup \{v\})$.

(ii) At iteration $k+1$, when solving the PO problem for set $U_k \cup \{v\}$, set the lower bound of Lagrangian dual variable as $\max\{\theta^*(U_k), \theta^*(U_{k-1} \cup \{v\})\}$.

As indicated by Lemma 2.4.2, the optimal dual variable monotonically increases with each greedy iteration. By implementing the AGA, we do not treat the PO problems as separate convex programming problems but rather utilize knowledge from previous iterations to speed up the solving process. In the AGA, the search region of the dual variable is adaptively shrunk at each greedy iteration by updating the lower bound to match the optimal dual variable from previous iterations. Consequently, the AGA can significantly reduce the execution time of optimizing the CGPO problem by exploiting the problem structure in conjunction with the greedy algorithm.

2.5 Parameter Estimation

In this section, we discuss how to estimate the parameters of the P-BDM by adapting the classical methods for the BDM. We show that despite the challenges of estimating parameters for diffusion models, the data available on online platforms allows us to achieve high-quality estimates even when the promotion policy is endogenously determined throughout the diffusion process.

Although the BDM describes a deterministic diffusion dynamic, several probabilistic methods have been proposed to estimate its parameters. Bass (1969) first estimates the parameters using the ordinary least square (OLS) method. Schmittlein and Mahajan (1982), Srinivasan and Mason (1986) apply maximum likelihood estimation (MLE) and nonlinear least square to obtain better estimates. However, these methods and their analysis are complicated by the diffusion nature, such as autocorrelation, which exists among observations because diffusion happens as a dynamic process. The new adoptions at each time step depend on the previous cumulative adoption number and affect future adoptions. The intricate relationship of parameters, such as the cumulative adopters as a direct function of p, q, and m, further complicates the problem (a commonly used expression can be founded in Schmittlein and Mahajan 1982). These issues degrade the performance of the estimators. Estimating the parameters of the P-BDM introduces additional challenges due to promotion. The inclusion of promotion decisions makes the closed-form expression of cumulative adopters no longer exist, limiting our choice of tools. Additionally, as the platform determines promotion policy based on real-time adoption levels, the promotion fraction is correlated with the cumulative adopters and thus endogenous in the diffusion dynamics.

We revisit the OLS and MLE methods for the BDM and adapt them to the P-BDM, leading to new estimation methods, namely the D-OLS and D-MLE methods. We highlight that, while there are inherent deficiencies in estimating diffusion models as mentioned above,

we can largely alleviate these issues and improve the estimation results on online platforms. In fact, compared to traditional markets, we can extract additional information from online platforms, particularly by identifying adopters who have received promotions. We use a fixed design framework to underscore the theoretical benefits of this extra information. Although this analysis is stylized, the benefits we demonstrate are not merely fortuitous; they are also consistently observed in numerical experiments with both OLS-based and MLE-based estimators. In the following discussions, we focus on a fixed $v \in \mathcal{V}$ and omit the subscript v, and treat the market size m as fixed.

2.5.1 OLS Estimators

In this part, we discuss the OLS-based methods for estimating parameters in the P-BDM. We base our approach on the OLS method for the BDM as presented in Bass (1969), summarized in Appendix A.1.2.1. To estimate parameters in the P-BDM, we observe a sequence of observations $\{(a_t, x_t, A_t)\}_{t=1}^T$, which includes both the realization of promotion decisions and adoption numbers. The OLS method for the P-BDM relies on the following relationship:

$$a_t = p \cdot mx_t + q \cdot \frac{A_{t-1}}{m}(m - A_{t-1}) + \epsilon_t \,,$$

where p and q are the two parameters to estimate and ϵ_t is independent random noise with mean 0, as defined in the OLS estimation for the BDM. We obtain OLS estimators for p and q by considering mx_t and $(A_{t-1} - A_{t-1}^2/m)$ as two observed covariates. However, since the promotion fraction often correlates with adoption numbers, there can be certain colinearity between these two covariates, resulting in OLS estimators possibly yielding large variances.

To reduce the variances, we can leverage information about adopter types on online platforms. Specifically, out of the total new adopters (a_t), we can observe the number of direct adopters who receive the promotion (a_t^{d}) and the number of indirect adopters who do not receive the promotion (a_t^{i}). This yields a sequence of adoption data $\left\{\left(a_t^{\mathrm{d}}, a_t^{\mathrm{i}}, A_t, x_t\right)\right\}_{t=1}^T$. We propose a straightforward double-OLS (D-OLS) method based on the following relationships:

$$a_t^{\mathrm{d}} = p \cdot mx_t + q \cdot A_{t-1}x_t + \epsilon_t^{\mathrm{d}} \quad \text{and} \quad a_t^{\mathrm{i}} = q \cdot \frac{A_{t-1}}{m}(m - A_{t-1} - mx_t) + \epsilon_t^{\mathrm{i}}, \tag{2.14}$$

where the first equation in (2.14) focuses on the direct adopters targeted by promotion while the second focuses on the others; ϵ_t^{d} and ϵ_t^{i} are independent random noises such that $\epsilon_t = \epsilon_t^{\mathrm{d}} + \epsilon_t^{\mathrm{i}}$.

Our D-OLS method yields estimators, $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$, through the following steps:

(i) We use the OLS method to estimate $\hat{q}^{\text{D-OLS}}$ from the second equation in (2.14), resulting in:

$$\hat{q}^{\text{D-OLS}} = \frac{\sum_{t=1}^T \left[A_{t-1}(1 - x_t - \frac{A_{t-1}}{m})a_t^{\mathrm{i}}\right]}{\sum_{t=1}^T \left[A_{t-1}(1 - x_t - \frac{A_{t-1}}{m})\right]^2};$$

(ii) We use the OLS method again, but this time we substitute q in the first equation in (2.14) with the D-OLS estimator $\hat{q}^{\text{D-OLS}}$, to compute $\hat{p}^{\text{D-OLS}}$, which is given by:

$$\hat{p}^{\text{D-OLS}} = \frac{\sum_{t=1}^{T} \left[m x_t (a_t^{\text{d}} - \hat{q}^{\text{D-OLS}} A_{t-1} x_t) \right]}{\sum_{t=1}^{T} (m x_t)^2}.$$

By separating the estimation of two coefficients, the D-OLS method also alleviates the issue of correlation between the promotion fraction and adoption number. This method reduces the variance of estimators and enhances prediction accuracy. In the next section, we illustrate this improvement by analyzing the asymptotic properties of these estimators.

2.5.1.1 Asymptotic Properties

We now examine the asymptotic properties of the estimators. Our analysis reveals that D-OLS estimators are $\sqrt{n}$-consistent and possess smaller asymptotic variances than OLS estimators. Moreover, the reduction in variance becomes more pronounced when the promotion policy is endogenous with the diffusion dynamics.

In the traditional BDM literature, rigorous asymptotic analysis of estimation has been a challenging task due to the lack of an asymptotic framework for diffusion processes. To flesh out the comparison between OLS and D-OLS estimators, we consider a fixed design framework with a triangular sequence of infinite diffusion processes. Specifically, we consider a sequence of diffusion processes with an increasing market size $m_{(n)}$ for $n = 1, 2, \ldots$. We assume the observations come from a fixed-design triangular array, wherein the n-th row includes n observations from the diffusion process with market size $m_{(n)}$. We treat the covariates as fixed rather than random variables. This creates a framework amenable to theoretical analysis. For the n-th diffusion process, let $\left\{A_{i,(n)}\right\}_{i=1}^{n}$ denote the adopters at n different time steps and $\left\{x_{i,(n)}\right\}_{i=1}^{n}$ denote the consequent promotion fractions. We then define the empirical second-moment matrices of the OLS method, as well as the empirical second moments of the two estimation steps in the D-OLS method as follows:

$$Q_{(n)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i,(n)}^2 & \frac{1}{n}\sum_{i=1}^{n} x_{i,(n)} \bar{A}_{i,(n)}(1 - \bar{A}_{i,(n)}) \\ \frac{1}{n}\sum_{i=1}^{n} x_{i,(n)} \bar{A}_{i,(n)}(1 - \bar{A}_{i,(n)}) & \frac{1}{n}\sum_{i=1}^{n} \left[\bar{A}_{i,(n)}(1 - \bar{A}_{i,(n)})\right]^2 \end{pmatrix},$$

$$\tilde{Q}_{11,(n)} = \frac{1}{n}\sum_{i=1}^{n} x_{i,(n)}^2 \bar{A}_{i,(n)}, \text{ and } \tilde{Q}_{22,(n)} = \frac{1}{n}\sum_{i=1}^{n} \bar{A}_{i,(n)}^2 (1 - x_{i,(n)} - \bar{A}_{i,(n)})^2,$$

where $\bar{A}_{i,(n)} = A_{i,(n)}/m_{(n)}$ is the normalized adopter number (i.e., the fraction of adopters).

Our analysis is based on the following assumption, common for regression in fixed-design settings and reasonable in practice. With Q defined in the assumption, we let Q_{11} be the component in row one and column one of Q. Other components can be defined in a similar fashion.

Assumption 2.5.1 (Positive Definiteness) *We assume that the following limits exist:*

$$\lim_{n\to\infty} Q_{(n)} = Q, \lim_{n\to\infty} \tilde{Q}_{11,(n)} = \tilde{Q}_{11}, \text{ and } \lim_{n\to\infty} \tilde{Q}_{22,(n)} = \tilde{Q}_{22},$$

where Q is positive definite and $\tilde{Q}_{11}, \tilde{Q}_{22} > 0$.

We further suppose the scaled random noise for the n-th diffusion process $\bar{\epsilon} := \epsilon/m_{(n)}$ has variance σ^2. The following theorems, Theorem 2.5.1 and Theorem 2.5.2, show the asymptotic properties of D-OLS estimators. The detailed proof is given in Appendix A.1.2.2.

Theorem 2.5.1 (Consistency) *Suppose that the scaled random noise $\bar{\epsilon}_i^i := \epsilon_i^i/m_{(n)}$ and $\bar{\epsilon}_i^d := \epsilon_i^d/m_{(n)}$ are independently and identically distributed with mean zero and finite variance for all $i = 1, \ldots, n$, then D-OLS estimators $\hat{p}^{D\text{-}OLS}$ and $\hat{q}^{D\text{-}OLS}$ converge to the true parameters p and q in probability as n scales to infinity. That is,*

$$\hat{p}_{(n)}^{D\text{-}OLS} \xrightarrow{p} p \text{ and } \hat{q}_{(n)}^{D\text{-}OLS} \xrightarrow{p} q.$$

Theorem 2.5.1 implies that with sufficient observations, the true values of p and q can be uncovered.

Theorem 2.5.2 (Asymptotic Normality) *Suppose that the scaled random noise $\bar{\epsilon}_i^i$ and $\bar{\epsilon}_i^d$ are independently and identically distributed with mean zero and variance $(1-\eta)\sigma^2$ and $\eta\sigma^2$ for all $i = 1, \ldots, n$ for some $\eta \in (0,1)$, then when n scales to infinity,*

(i) D-OLS estimators $\hat{p}_{(n)}^{D\text{-}OLS}$ and $\hat{q}_{(n)}^{D\text{-}OLS}$ are asymptotically normal. Specifically,

$$\sqrt{n}\left(\hat{p}_{(n)}^{D\text{-}OLS} - p\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{11}}(1+\xi_1)\sigma^2\right) \text{ and } \sqrt{n}\left(\hat{q}_{(n)}^{D\text{-}OLS} - q\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{22}}(1+\xi_2)\sigma^2\right),$$

where $\xi_1 = \eta(\tilde{Q}_{11}^2/\tilde{Q}_{22}Q_{11} - 1)$ and $\xi_2 = \eta Q_{22}/\tilde{Q}_{22} - 1$.

(ii) OLS estimators $\hat{p}_{(n)}^{OLS}$ and $\hat{q}_{(n)}^{OLS}$ are asymptotically normal. Specifically,

$$\sqrt{n}\left(\hat{p}_{(n)}^{OLS} - p\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{11}}(1+\kappa)\sigma^2\right) \text{ and } \sqrt{n}\left(\hat{q}_{(n)}^{OLS} - q\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{22}}(1+\kappa)\sigma^2\right),$$

where $\kappa = Q_{12}^2/|Q|$.

We draw two insights based on Theorem 2.5.2. First, the ratio κ is not negligible, especially when the promotion policy is endogenous with diffusion dynamics. We observe that κ increases as the determinant of Q decreases. When x is highly colinear to $\bar{A}(1-\bar{A})$, κ approaches infinity while ξ_1 and ξ_2 remain bounded. Therefore, D-OLS estimators are more robust against correlations than OLS estimators. Second, when $\eta \leq \tilde{Q}_{22}/Q_{22}$, D-OLS estimators have smaller asymptotic variances than OLS estimators (see Proposition A.1.1, Appendix A.1.2). We note that, according to our real-world data set, the average of promotion fraction x_t is 0.00062 per hour, placing $\tilde{Q}_{22}/Q_{22}$ in close proximity to one. Consequently, we expect $\eta \leq \tilde{Q}_{22}/Q_{22}$ to be readily fulfilled in our setting, suggesting that D-OLS estimators present smaller asymptotic variances than OLS estimators. The advantage of D-OLS is clearly observed in subsequent numerical experiments.

2.5.2 MLE Estimators

While the OLS method is straightforward and computationally efficient, it lacks a rigorous probabilistic interpretation in a diffusion setting. On the other hand, the MLE method in Schmittlein and Mahajan (1982) for estimating the BDM is based on a rigorous probabilistic model. However, it requires an explicit expression of the cumulative adopter number A_t, which is not applicable in the P-BDM. Nonetheless, we show that MLE-based estimators can still be used in our setting.

When the platform cannot distinguish adopter types, the probabilistic counterpart is established as follows: At time t, there are $(m - A_{t-1})$ non-adopters, each of which has the same adoption probability as $(px_t/(1 - A_{t-1}/m) + qA_{t-1}/m)$. The log-likelihood function is formulated as

$$\begin{aligned}&\mathcal{LL}^{\text{MLE}}(p, q)\\ &= \sum_{t=1}^{T} a_t \log\left(\frac{mx_t}{m - A_{t-1}}p + \frac{A_{t-1}}{m}q\right) + (m - A_{t-1} - a_t)\log\left(1 - \frac{mx_t}{m - A_{t-1}}p - \frac{A_{t-1}}{m}q\right).\end{aligned}$$

When the platform can distinguish adopter types, the probabilistic counterpart is established as follows: At time t, there are $(m - A_{t-1})$ non-adopters. Each non-adopter has a probability of $mx_t/(m - A_{t-1})$ to be promoted by the platform. Given being promoted, the non-adopters adopt independently with probability $(p + qA_{t-1}/m)$. Otherwise, the non-adopters adopt independently with probability qA_{t-1}/m when not being promoted. The log-likelihood function is formulated as

$$\begin{aligned}\mathcal{LL}^{\text{D-MLE}}(p, q) = &\sum_{t=1}^{T}\left[a_t^{\text{i}}\log\left(\frac{A_{t-1}}{m}q\right) + (m - A_{t-1} - mx_t - a_t^{\text{i}})\log\left(1 - \frac{A_{t-1}}{m}q\right)\right]\\ &+ \sum_{t=1}^{T}\left[a_t^{\text{d}}\log\left(p + \frac{A_{t-1}}{m}q\right) + (mx_t - a_t^{\text{d}})\log\left(1 - p - \frac{A_{t-1}}{m}q\right)\right],\end{aligned}$$

and the derived estimators are named D-MLE estimators. In Appendix A.1.2.3, we show that both log-likelihood functions are concave, allowing us to use the gradient method for estimation.

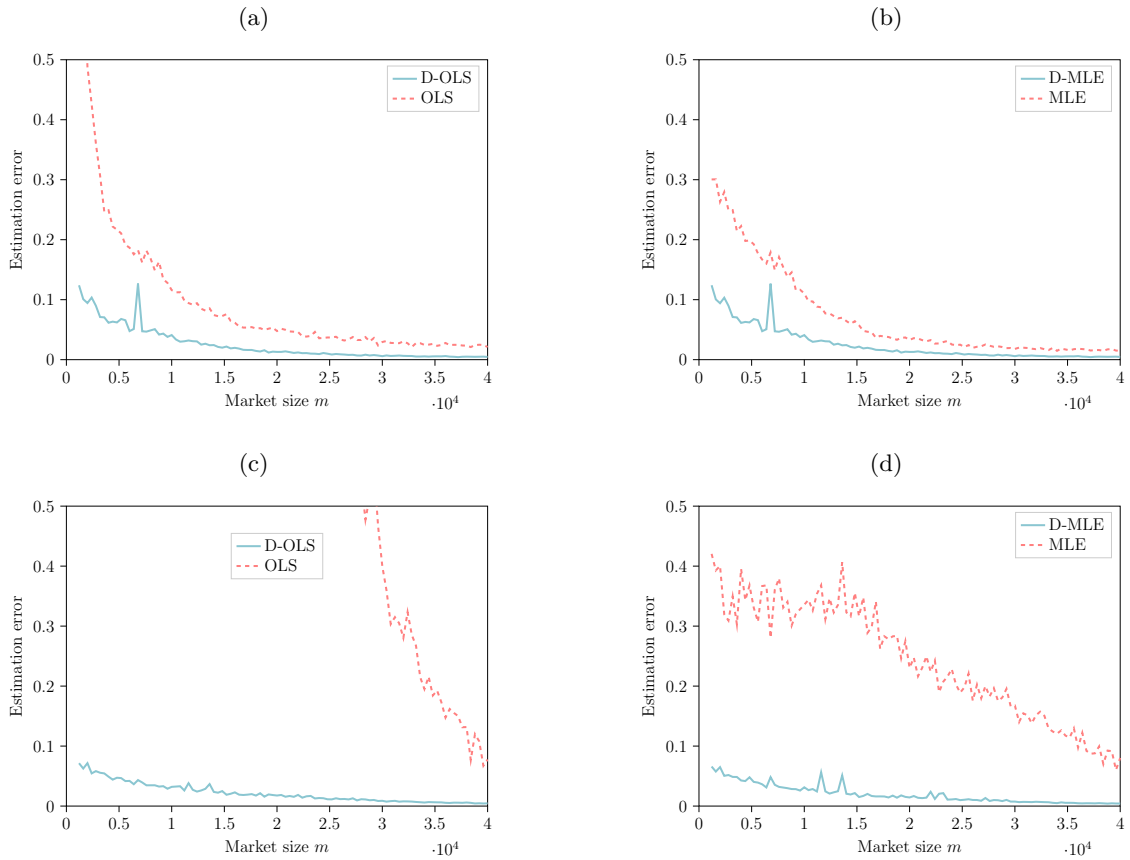
2.5.3 Comparing OLS-Based and MLE-Based Estimators with Simulation

We create a synthetic dataset by bootstrapping the diffusion processes of a content piece according to P-BDM dynamics. The diffusion processes are simulated based on the D-MLE probabilistic counterpart defined in Section 2.5.2 when adopter types can be distinguished. We assess the estimators under two promotion schemes: (i) Const: promotion fraction x_t remains constant; (ii) Linear: promotion fraction x_t has a positive linear relationship with adopter number A_{t-1}.

The true values of coefficients are set at $p = 0.523$ and $q = 0.062$. We run experiments with market sizes ranging from $m = 1,200$ to $m = 40,000$, and scale observation numbers with market size, as mentioned in Section 2.5.1.1. Performance is measured using estimation error of the parameters, which is the Euclidean distance between the estimators and true values $\sqrt{(p-\hat{p})^2 + (q-\hat{q})^2}$.

Figure 2.2 displays the results for Const and Linear schemes, respectively. Overall, we observe a significant improvement when adopter types can be distinguished, emphasizing the benefits of using additional data for estimating diffusion models.

Figure 2.2: Estimation Errors for Different Methods Against Market Size



Notes. Scale with the number of observations. (a) Const: OLS-based methods. (b) Const: MLE-based methods. (c) Linear: OLS-based methods. (d) Linear: MLE-based methods.

We offer two more observations. First, comparing Figures 2.2a with 2.2c, and Figures 2.2b with 2.2d, we see larger relative improvements under the Linear scheme compared to the Const scheme, particularly for OLS-based estimators. This not only indicates robustness in

MLE-based methods but also verifies the theoretical results in Theorem 2.5.2 and highlights the effectiveness of our proposed estimators. Second, comparing Figures 2.2a and 2.2b, or Figures 2.2c and 2.2d, we notice that D-OLS and D-MLE perform similarly when adopter types can be distinguished, while the MLE method outperforms the OLS method when they cannot be differentiated. In this case, the correlation among covariates creates difficulty for OLS estimators, but additional information about adoption types helps to greatly narrow the gap.

In summary, both the D-OLS and D-MLE methods perform well when working with data available on online platforms. While the D-MLE method is supported by a rigorous probabilistic framework, it is less computationally efficient. Given the similar performance of D-OLS and D-MLE, we opt to use the D-OLS method for other computational experiments with real data.

2.6 Numerical Results

In this section, we conduct a comprehensive analysis using data from a large-scale video-sharing platform. At a high level, we test the performance of the AGA promotion policy with this data set. To help readers better understand our numerical results, we provide the code for our analyses in a GitHub repository[1].

2.6.1 Platform and Data Overview

We obtain the data set from one of the most popular Chinese video-sharing platforms, similar to TikTok. The platform is fueled by user-generated content and has become a social phenomenon, with a massive user base sharing their daily lives. As of 2023, it has over 360 million daily active users and over 20 billion videos. Effective content promotion plays an important role in platform operations. While machine learning-based algorithms offer personalized recommendations curated based on user interests, promoting content that has the potential to go viral is challenging due to the difficulty of optimizing diffusion. As such, the issue addressed in this study is essential for the platform to maximize its impact and foster an engaged user community.

The data set consists of user behavior logs for 46,444 short videos, sampled from 518,646 users over 20 days (7/1/2020-7/20/2020). The logs contain timestamped records of video promotions and user behavior in terms of clicks. For each video, we identify two distinct user sets: $\mathcal{L}_P$, which comprises users who receive the promotion, and $\mathcal{L}_C$, which comprises users who click on it. Due to the presence of diffusion effects, some users click on videos without receiving promotions (i.e., $\mathcal{L}_C \setminus \mathcal{L}_P \neq \emptyset$). For ease of analysis, we aggregate the timestamped data hourly. Then, we calculate the promotion fraction as the ratio between the promoted users ($\mathcal{L}_P$) and market size m. We further identify adopter types: direct adopters ($\mathcal{L}_P \cap \mathcal{L}_C$) and indirect adopters ($\mathcal{L}_C \setminus \mathcal{L}_P$). In addition, each video is categorized by the platform

[1]See https://github.com/YunduanLin/Content_Promotion

according to its topic. The dataset includes videos from 61 category labels provided by the platform, ranging from 155 to 2,759 videos per category.

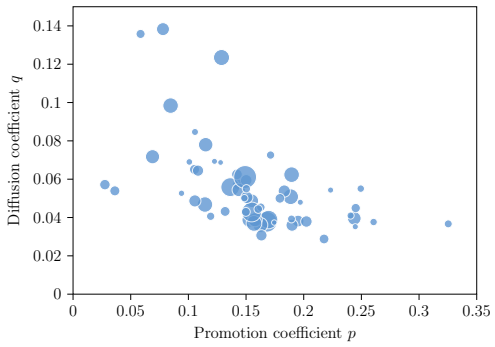
2.6.2 Model Calibration

In this section, we estimate the promotion and diffusion coefficients under the P-BDM specification with the real-world video data, comparing results with the BDM benchmark.

We use the D-OLS method to estimate p and q. During this process, we consider the following two key aspects. (i) *Time-decay factor*: We include a time-decay factor γ as a hyperparameter to reflect users' decreasing tendency to share content over time. See Appendix A.2.1.2 for more details. (ii) *Group estimation*: We estimate the same p and q values for each video category. We highlight that promotion decisions are often made at the early stages of a video's life cycle when limited data is available for estimation. Consequently, group-wise estimation is typically utilized to guarantee generalizability. In principle, we can adopt a contextual approach given the availability of the featurized information of each content piece. The group-based estimation can be seen as a special case of this approach, where the sole feature variable is the category information. For the sake of simplicity in this study, we use the category labels provided by the platform to determine groups. See Appendix A.2.1.3 for more details. For further details about our calibration process, including data splitting, hyperparameter selection, and the effect of time-decay factor γ, please refer to Appendix A.2.1.4. Next, we present the calibration results under the best time-decay factor $\gamma = 0.983$.

Distribution of p and q: Figure 2.3 depicts the distribution of estimated coefficients across 61 different categories. Notably, a negative correlation between p and q is observed,

Figure 2.3: Distribution of Estimated Promotion Coefficient p and Diffusion Coefficient q



Notes. Each point in the scatter plot represents a video category. The size of points represents the number of videos in each category.

with a Pearson correlation coefficient of $\rho = -0.5335$. A one-tailed t-test further supports the observation, with a t-statistic of -4.845 rejecting the null hypothesis at a significance level

of 0.05 (critical t-value of -1.671). These findings suggest that videos with a large promotion effect may not have a larger diffusion effect, highlighting the need for a promotion policy that accounts for the diffusion effect.

Performance of estimation: We evaluate the performance using the weighted mean absolute percent error (WMAPE), which can be calculated as WMAPE $= \sum_{t=1}^{T_v} |a_{v,t} - \hat{a}_{v,t}| / \sum_{t=1}^{T_v} a_{v,t}$ for video v, where $\hat{a}_{v,t}$ is the predicted number for new adopters. Overall, the P-BDM estimated with the D-OLS method achieves an average out-of-sample WMAPE of 38.96%. We assess the D-OLS approach with the P-BDM against two benchmarks. Our first point of comparison is the traditional OLS method, illustrating the advantages of the D-OLS method in the context of online platforms. Secondly, we contrast the P-BDM with the BDM, demonstrating the P-BDM's superior aptitude in managing online content adoptions. The out-of-sample WMAPEs for these two benchmarks register at 39.66% and 81.25%, respectively. The P-BDM shows a considerable improvement over the BDM and a moderate yet noticeable enhancement compared to the OLS method. In contrast to the simulation in Section 2.5.3, the colinearity issue brought up in Section 2.5.1.1 is not severe in this dataset. Through the use of a paired t-test, we ascertain that the improvement is statistically significant with a t-statistic of -35.48, smaller than the t-value corresponding to a 0.05 significance level (i.e., -1.645). Further, when we perform hypothesis tests for each category, we find that 53 out of the 61 categories display improvement at the 0.05 significance level. Two categories indicate deterioration, while the remaining six categories do not show significant changes.

To further illustrate the effectiveness of the P-BDM, we present two examples in Figure 2.4. Figure 2.4a uses the same video as the example in Section 2.3.1. To delineate the issue, we estimate the coefficients from a single video, rather than the entire category. That is, for each video, we use the first 60% data samples to estimate coefficients and generate the fitted curves for the entire time horizon using the estimated coefficients.

Compared to the BDM, the P-BDM fits not only the overall adoption trend but also the curve shape. While the BDM provides reasonable fit in early periods, a common issue observed is the underestimation of the diffusion coefficient. In some cases, as shown in Figure 2.4b, the estimated coefficient can even be negative, which lacks a valid real-world interpretation. These observations underscore the effectiveness of the P-BDM.

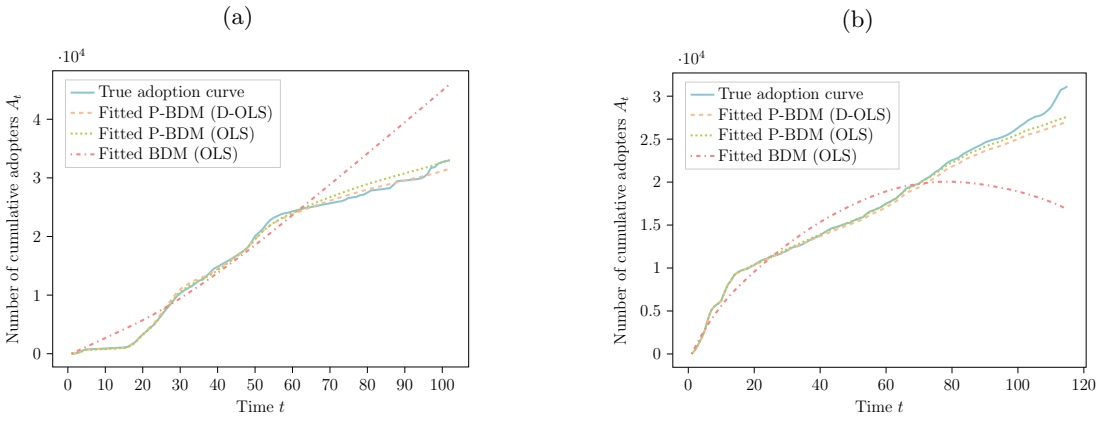
2.6.3 Experiments on the Accelerated Greedy Algorithm

In this section, we simulate the platform environment with estimated parameters to evaluate promotion policies. We name the policy decided by the AGA under the P-BDM as the AGA policy.

2.6.3.1 Long-term Performance with Different Planning Intervals

In practice, platforms are concerned with the long-term efficacy of promotion policies. Accordingly, we solve the CGPO problem every L periods using the AGA policy in the most recent platform environment.

Figure 2.4: Illustration of Adoption Curves and the Corresponding Fitted BDM/P-BDM Curves for Example Videos



Notes. To ensure data anonymity, we have scaled the y-axis using a randomly selected number. (a) Motivating example in Section 2.3.1. (b) Example of negative fitted BDM coefficient.

We simulate a 120-period time horizon with a market size of $m = 10,000$ and assess the AGA policy by varying the planning interval L from 1 to 20. More details are described as follows.

- *Video corpus*: The video corpus is initialized at $t = 0$ with 50 videos, all with zero adoptions. We assume that at each time step, five new videos are added with no initial adoptions, to be consistent with the practical operations of the platform. Each video is associated with parameters (p, q) randomly sampled from the empirical distribution estimated in Section 2.6.2.

- *User behavior*: At each time, users act according to the D-MLE stochastic counterpart of the P-BDM, as described in Section 2.5.2.

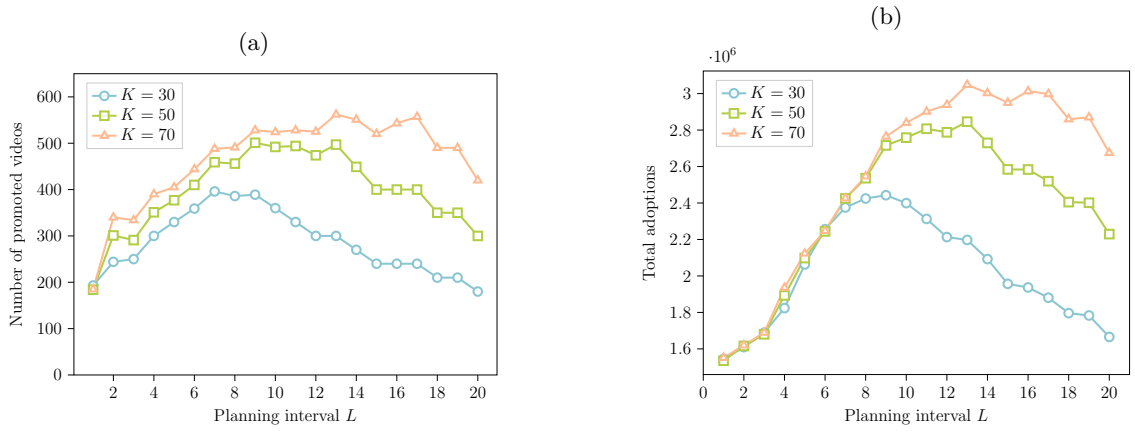
- *AGA implementation.* We assume that the platform employs the AGA with a planning interval of L. We solve the CGPO problem every L periods and implement the policy recommended by the AGA for these L periods. To keep the policy up-to-date with the platform environment, new videos added during the past L periods are included when solving the new CGPO problem instance, with initial adoption numbers set to match those at the end of the past L periods.

We first remark that the selection of L is crucial in striking a balance between the frequency of policy updates and the consideration of diffusion effects. A smaller L permits more frequent policy updates yet is myopic and ignores diffusion in the long term. Conversely, a larger L

considers more extended diffusion effects but may delay the promotion of new videos due to less frequent policy updates. When $L = 1$, the AGA policy is equivalent to ignoring all diffusion effects. We observe that this trade-off is influenced by the choice of candidate set size K and the promotion budget C. To elucidate this, we present results varying these two parameters separately.

In Figure 2.5, we fix the average promotion budget per user per period $\bar{C}$ at 6, and vary the size of the candidate set size to be $K \in \{30, 50, 70\}$.

Figure 2.5: Illustration of the AGA Policy for Different Selections of Candidate Set Size K.



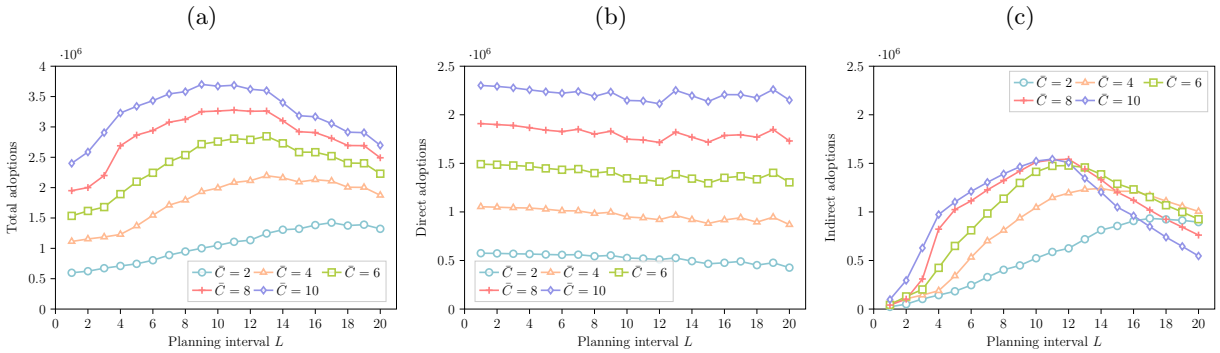
Notes. (a) Promoted videos. (b) Total adoptions.

As shown in Figure 2.5a, we notice that this tradeoff is evident by an initial increase followed by a decrease in the number of unique promoted videos as L increases. Notably, we observe that an increase in K leads to a rise in the number of promoted videos. This trend suggests that the capacity constraint becomes more restricted in scenarios with a larger planning horizon L, primarily due to the increased complexity of the diffusion trajectory in such cases. On average, 61.37% of instances face a binding capacity constraint, underscoring its significant impact on the outcomes of the promotion strategy. Figure 2.5b further sheds light on the total adoptions during the process, which mirrors the pattern observed in the number of promoted videos. Especially, as K increases, the optimal planning horizon L also tends to be larger.

In Figure 2.6, we fix the candidate set size to be 50, and vary the average promotion budget to be $\bar{C} \in \{2, 4, 6, 8, 10\}$.

From Figure 2.6a, we observe a similar tradeoff akin to our previous findings, except that an increase in C leads to a smaller optimal planning horizon. As shown in Figure 2.6b, direct adoptions exhibit a consistent decrease with an increase in L. This is expected because a longer planning interval reduces direct adoptions from promotion to potentially

Figure 2.6: Illustration of the AGA Policy for Different Selections of Promotion Budget.

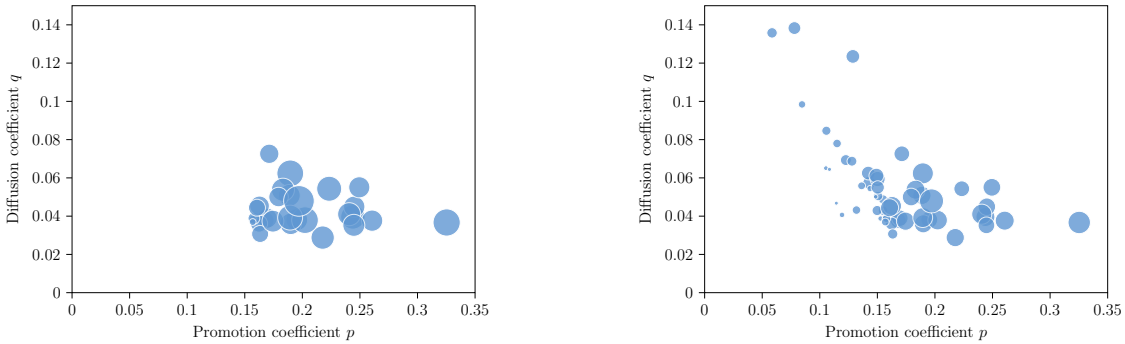


Notes. (a) Total adoptions. (b) Direct adoptions. (c) Indirect Adoptions.

increase indirect adoptions driven by diffusion. Similar to the total adoptions, the indirect adoption curve assumes an inverted U-shape, a phenomenon driven by the fact that when L is too large, the algorithm suffers from infrequent updates, losing the diffusion power for new videos due to timeliness.

Hereafter, we fix the cardinality constraint $K = 50$ to conduct the simulation. We then investigate how the AGA policy with different L values distributes the promotion budget among videos based on p and q. Figure 2.7 shows the average promotion times received by videos in different categories. Although videos with a large p value tend to receive more

Figure 2.7: Illustration of the AGA Policy across Different Video Categories



Notes. Each point in the scatter plot represents a video category. The size of points represents the average promotion times. Left: $L = 1$. Right: $L = 20$.

promotion, clearly as L increases, the AGA policy increases the budget allocated to the

videos with a large q value to trigger more long-term diffusion. The judicious allocation of limited resources is governed by our algorithm.

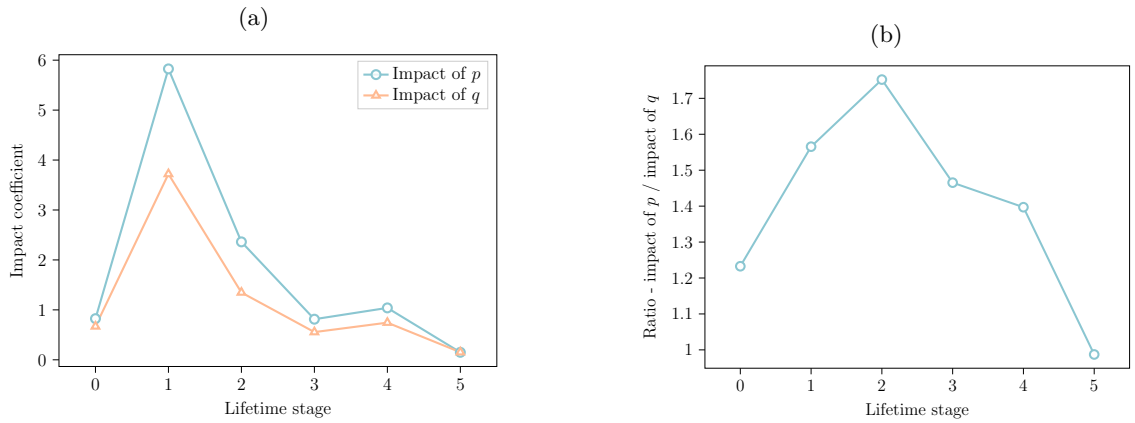
2.6.3.2 The Underlying Mechanism of the AGA Policy

To gain deeper insights into the mechanism underlying the AGA policy and promote a qualitative understanding of how to manage the interactions of promotion and diffusion effects, we conduct additional analysis of the promotion fraction for videos, focusing on different model primitives. For illustration, we select $L = 13$, which consistently performs well across different promotion budgets in our experiments.

We aim to understand how a video's configuration (p_v and q_v) and lifetime ($A_{v,t-1}$) affect the promotion fraction $x_{v,t}$ in the AGA policy. We use our experimental results as observations, where each observation represents a promotion fraction $x_{v,t}$ allocated to a video v at the beginning of time t. We divide the observations into six stages based on video lifetime. Stage 0 includes observations with $A_{v,t-1} = 0$, and stages $i \in \{1, 2, 3, 4, 5\}$ include observations when the video v has an adopter number $A_{v,t-1}$ at the start of time t such that $i = \lceil 5A_{v,t-1}/m \rceil$.

We first conduct a sensitivity analysis to examine how video configurations impact the promotion fraction x. This analysis uses linear regression to study the effects of p and q while controlling other relevant covariates. For a detailed explanation of this analysis, please refer to Appendix A.2.2.2. The regression coefficients of p and q, which we interpret as their impacts, are presented in Figure 2.8.

Figure 2.8: The Impacts of Video Configurations on the AGA Policy Across Different Lifetime Stages

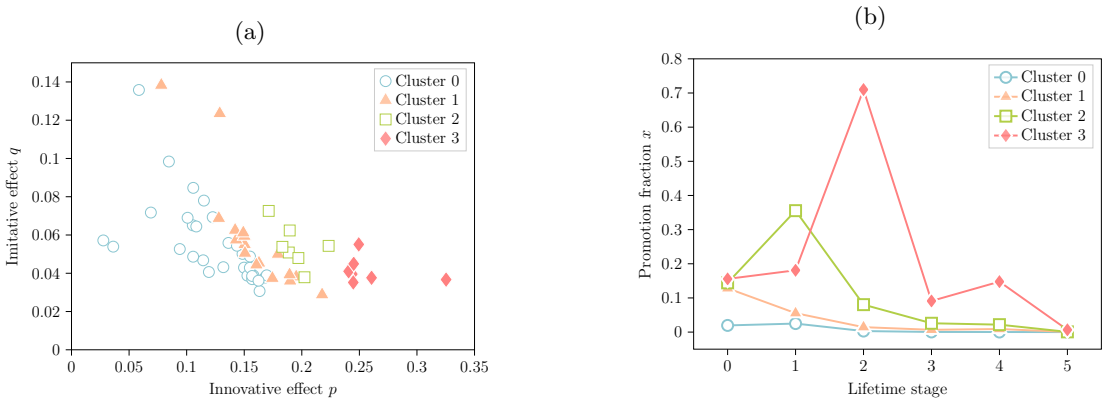


Notes. (a) Impacts of p and q. (b) Impact ratio between p and q.

Figure 2.8a demonstrates that both p and q positively influence x. Contrary to the intuitive expectation that the impact should decrease gradually after the initial stage, our findings suggest otherwise. The impact of p and q is most pronounced during the intermediate stages (i.e., stages 1 and 2). This is because, in the initial stage, the policy aims to kickstart the diffusion processes for a large pool of videos; so that it does not heavily differentiate between video configurations. In other words, by accounting for the diffusion effect, the AGA policy promotes a diverse range of videos in their initial stages, thereby making efficient use of the promotion budget. In the intermediate stages, however, the policy becomes more selective, filtering out noncompetitive videos and favoring videos with greater potential. Furthermore, Figure 2.8b shows the impact ratio between p and q, indicating that p carries more weight than q, particularly during the intermediate stages.

Then, we use K-Means clustering to group video configurations into four clusters based on their lifetime promotion policies. The clustering procedure is described in Appendix A.2.2.3.

Figure 2.9: Illustration of the AGA Policy Clusters Corresponding to Different Video Configurations



Notes. (a) Clusters of different video configurations. (b) Cluster centers.

Figure 2.9a displays the clustering results according to video configuration, while Figure 2.9b showcases the "average" promotion policy for each cluster. Several observations can be made. Despite the clusters being based on promotion policy, they strongly correlate with video configurations. Cluster 0, located in the bottom left of Figure 2.9a, is distinct due to its notably lower promotion at stage 0. These are the videos "discarded" by the AGA policy. Roughly speaking, moving towards the right of Figure 2.9a, videos receive more promotions. Echoing the insights from our sensitivity analysis, we notice a trend where the points of peak promotion shift towards later stages as p increases. In contrast, videos in Cluster 1 and 2

that have larger q values and smaller p values need to be promoted early to take advantage of their diffusion potential.

In summary, the AGA policy operates based on two main principles: (i) Promotion and diffusion effects, p and q, positively influence the promotion intensity, with the most profound impact during the intermediate lifetime stages. Among these, the promotion effect has a more notable impact. (ii) Videos with a small p but large q mainly receive their promotion in the early stages, where the promotion acts as a trigger for diffusion. In contrast, videos with a large p but small q continue to be promoted, serving as both a trigger for diffusion and an attraction for direct adoptions.

2.6.3.3 Comparison with Benchmarks

Finally, we compare the total adoptions of the AGA policy with benchmark policies, using the same experimental setting as in Section 2.6.3.1.

Benchmark policies: To ensure a fair comparison, we simulate the benchmark policies using the same diffusion process as our algorithm and compare the generated virtual rewards.

- *CGPO with accelerated greedy algorithm* (AGA). Our proposed algorithm as discussed in Section 2.4.3.2, with a planning horizon of $L = 13$.
- *CGPO without the diffusion effect* (NoD). This benchmark ignores the network effect. This is equivalent to the CGPO formulation when $L = 1$. It is a common practice in the industry that ignore the diffusion effects when promoting content.
- *Candidate Generation by Attractiveness* (ATT). This benchmark considers a heuristic CG strategy by selecting content that has the largest promotion potential $p_v(m - A_v)$. This benchmark speeds up the CG procedure but overemphasizes the promotion effect.
- *Candidate Generation by Timeliness* (TIM). This benchmark considers a heuristic CG strategy by selecting content that is most recently added to the platform. This benchmark takes the timeliness of online content into account but overlooks the promotion effect.
- *Candidate Generation by Potential* (POT). This benchmark considers a heuristic CG strategy by selecting content that has the most number of new adopters at the previous time step. This benchmark illustrates the "rich-get-richer" principle.

Experiment result: Table 2.1 compares the performance of all benchmarks. We draw three key observations from the table. Firstly, AGA consistently performs well, ranking second only when $\bar{C} = 2$. The margin of AGA over others is remarkable. Secondly, ATT and TIM also show notable improvement over NoD, suggesting the benefits of considering diffusion effects in promotion decisions. These benchmarks can be practical alternatives to AGA in real-world scenarios. Thirdly, POT performs even worse than NoD in most cases, indicating the drawbacks of the "rich-get-richer" principle. These observations highlight the

Table 2.1: Total adoptions of different promotion policies

	NoD	AGA	ATT	TIM	POT
$\bar{C} = 2$	597,774	1,244,845	1,086,217	1,276,566	603,698
	-	108.25%	81.71%	113.44%	0.99%
$\bar{C} = 4$	1,113,254	2,195,978	1,984,325	2,127,953	916,924
	-	97.26%	78.25%	91.15%	-17.64%
$\bar{C} = 6$	1,534,981	2,846,019	2,691,889	2,664,396	1,048,988
	-	85.41%	75.37%	73.58%	-31.66%
$\bar{C} = 8$	1,950,074	3,262,392	3,172,585	3,081,788	1,202,599
	-	67.30%	62.69%	58.03%	-38.33%
$\bar{C} = 10$	2,400,101	3,597,649	3,534,973	3,401,598	1,265,783
	-	49.90%	47.28%	41.74%	-47.26%

Notes. In each table cell, the first number represents the total adoptions under the benchmark, and the second number represents its relative improvement from NoD policy.

importance of both candidate generation and diffusion effects for content promotion and support the effectiveness of our proposed promotion policy.

2.7 Conclusion

In this study, we address the content promotion problem in online content platforms with the diffusion effect. We introduce a novel diffusion model to capture the platform's policy and the timeliness factor in online content diffusion. Based on this model, we formulate the candidate generation and promotion optimization (CGPO) problem. The problem is proved to be NP-hard, and we offer an efficient approximation algorithm that exploits the problem structure. We also propose a double OLS method to estimate model parameters, leveraging the online platform data. Finally, we use a real-world dataset to validate the model, evaluate the performance, and provide managerial insights. Our empirical evidence underscores the importance of considering the diffusion effect in promotion optimization and supports the effectiveness of our proposed promotion policy.

There are several future directions for this study. First, we could investigate the impact of externalities between content pieces. For instance, similar content could potentially substitute for each other, complicating the CGPO problem as the current submodularity results no longer apply. Second, we focus on an offline setting in this paper, where parameters are estimated beforehand. It can be quite interesting to consider the online version where parameters for new videos are estimated simultaneously with promotion optimization. The wealth of user and content information available on online platforms offers opportunities to explore this setting.

Chapter 3

Nonprogressive Diffusion on Social Networks: Approximation and Applications

3.1 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.

The remainder of this chapter is structured as follows: In the following of this section, we review the related literature. Section 3.2 introduces the nonprogressive diffusion model and characterizes the limiting adoption rate. In Section 3.3, we describe the FPA scheme and demonstrate our main theoretical results. Then in Section 3.4, we establish the order-optimal error bound followed by extensive numerical experiments in Section 3.5. We study the IM and pricing problem using our FPA scheme in Section 3.6. Section 3.7 concludes this

paper. Throughout this paper, we use increasing and decreasing in the non-strict sense.

3.1.1 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. 2024). 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 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-

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

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.

3.2 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.

3.2.1 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

$$u_i(t) := v_i + \beta \cdot \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t). \tag{3.1}$$

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 (3.1), $u_i(t)$ 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 3.2.1 (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 3.2.1 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 3.2.2 (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 3.2.2. It not only excludes divergent or periodic behavior in the long run in our diffusion model (Proposition 3.2.1) but also guarantees a valid fixed-point characterization of the limiting adoption probabilities (Proposition 3.3.1). In Section 3.5.1, we also extend our discussion

through extensive numerical experiments to investigate the implications of Assumption 3.2.2 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.

3.2.2 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

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

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

$$q_i(t) \;:=\; \mathbb{P}(Y_i(t) = 1) \;\equiv \sum_{\mathbf{y} \in \{0,1\}^n} \mathbb{1}\{y_i = 1\} \cdot \mathbb{P}(\mathbf{Y}(t) = \mathbf{y}). \tag{3.2}$$

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

Proposition 3.2.1 (Limiting Adoption Probability) *Under Assumptions 3.2.1 and 3.2.2, 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 3.2.1, 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

$$\lim_{t\to\infty} \frac{1}{t} \cdot \sum_{s=1}^{t} Y_i(s) \,=\, q_i^* \quad \text{a.s.} \qquad \text{and} \qquad \lim_{t\to\infty} \frac{1}{t} \cdot \sum_{s=1}^{t} q_i(s) \,=\, q_i^*, \quad \forall i \in V, \tag{3.3}$$

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 3.6.1) and the pricing problem on a social network (Section 3.6.2), can be generally framed as:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \qquad g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta), \mathbf{x}\Big). \tag{3.4}$$

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 (3.4) 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 3.6.

Solving the optimization problem (3.4) 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 (3.4) is generally intractable either analytically or computationally, which motivates us to develop our approximation scheme for $\mathbf{q}^*$ presented in Section 3.3.

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.

3.3 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 (3.6). In this section, we assume that Assumptions 3.2.1 and 3.2.2 hold.

3.3.1 Overview and Motivating Example

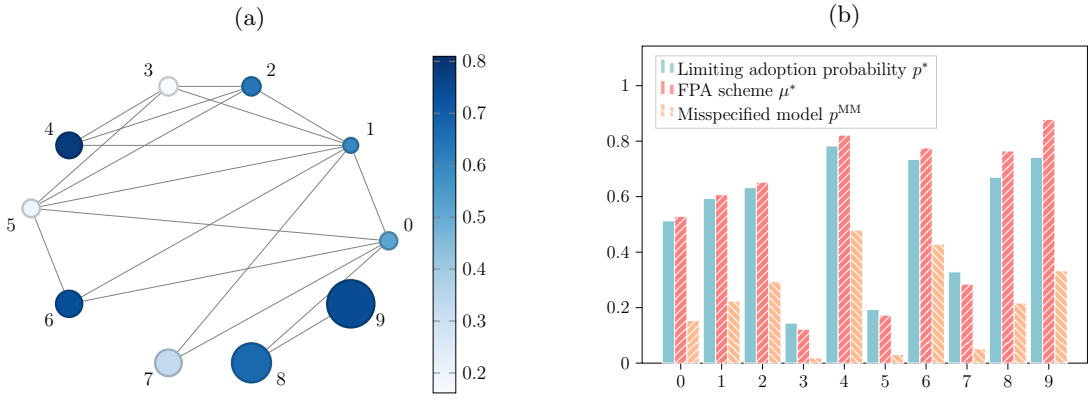
For a given diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, we will show that the limiting adoption probability $\mathbf{q}^*$ can be reasonably approximated by the solution $\boldsymbol{\mu}^*$ of the following simple system of equations:

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i}\right), \text{ for all } i \in V. \tag{3.5}$$

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 $\mathbb{E}\left[\mathbb{1}\{v_i + \epsilon_i \geq 0\}\right]$ so the network effects are ignored. For detailed information about this example instance, including numerical results, please refer to Appendix B.2.1.

Figure 3.1a 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 3.1b, 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

Figure 3.1: The 10-node Example to Illustrate the FPA Scheme



Notes. In the left subfigure, the color denotes the true value of $\mathbf{q}^*$ and the size denotes the absolute error between $\mathbf{q}^*$ and $\boldsymbol{\mu}^*$. (a) Approximation results on the network structure. (b) Adoption probability values.

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.

3.3.2 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 3.2.2). 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:

$$\mathcal{C}(G;\rho) := (1-\rho)\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^{\ell} \tilde{\mathbf{A}}^{\ell}\right)\mathbf{b} = (1-\rho)\left(\mathbf{I} - \rho\tilde{\mathbf{A}}\right)^{-1}\mathbf{b} \quad \text{and} \quad \mathcal{D}(G) := \frac{\mathbf{e}^{\top}\mathbf{b}}{n}. \tag{3.6}$$

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 3.2.2. 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 3.3.1, our primary focus is to characterize the entry-wise error between the limiting probability $\mathbf{q}^*$ and our FPA solution $\boldsymbol{\mu}^*$.

Theorem 3.3.1 (Entry-wise Error Bound of the FPA Scheme) *Under Assumptions 3.2.1 and 3.2.2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$*, the absolute difference between the limiting adoption probability* $\mathbf{q}^*$ *and the fixed-point solution* $\boldsymbol{\mu}^*$ *can be upper bounded by*

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

Theorem 3.3.1 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 3.1.

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 3.3.1, it then holds that

$$\mathcal{C}_i(G;\rho) \leq (1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)} \;\leq\; \frac{1}{d_{\min}}, \text{ for all } i \in V, \tag{3.7}$$

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

$$|q_i^* - \mu_i^*| \;\leq\; C_\rho \cdot \sqrt{(1-\rho)\sum_{\ell=0}^{\infty}\frac{\rho^\ell}{d_{\min,i}(\ell)}}, \tag{3.8}$$

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 (3.7).

Corollary 3.3.1 (ℓ_∞-Norm Bound) *Under Assumptions 3.2.1 and 3.2.2, 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*

$$\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\leq\; C_\rho \cdot \sqrt{\frac{1}{d_{\min}}}. \tag{3.9}$$

Corollary 3.3.1 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 (3.8) and (3.9) are both intuitively appealing, these bounds are looser than the one based on the inverse in-degree centrality presented in Theorem 3.3.1.

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 3.3.2 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 3.3.2 (Scaled ℓ_1-Norm Bound) *Under Assumptions 3.2.1 and 3.2.2, 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*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\leq\; C_\rho \cdot \sqrt{\frac{\|\mathcal{C}(G,\rho)\|_1}{n}}. \tag{3.10}$$

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

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\leq\; C_\rho \cdot \sqrt{\frac{1-\rho}{1-\rho r(G)} \cdot \mathcal{D}(G)}. \tag{3.11}$$

In light of (3.7), 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 3.3.2 provides more meaningful bounds. Particularly, (3.10) 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 (3.10) 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 (3.11) 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 3.3.2 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 (3.1) 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 3.5. 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 (3.4). Instead of solving (3.4) directly, we can replace $\mathbf{q}^*$ with $\boldsymbol{\mu}^*$ and approximate (3.4) as follows:

$$\underset{\mathbf{x}\in\mathcal{X},\boldsymbol{\mu}}{\text{maximize}} \qquad g(\boldsymbol{\mu},\mathbf{x}) \quad \text{s.t. } \boldsymbol{\mu} = \mathbf{h}\Big(\boldsymbol{\mu}\,;\, G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta\Big), \tag{3.12}$$

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 (3.13) in Section 3.3.3. The approximate problem (3.12) offers an explicit formulation by incorporating the FPA scheme as a constraint. This stands in contrast to the implicit variable $\mathbf{q}^*$ in problem (3.4), which emerges from a complex stochastic process. The formulation (3.12) thus simplifies the optimization problem and facilitates its solution.

We advocate for the use of this approximate optimization problem (3.12). 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 3.4.2 and 3.5.2). From a practical standpoint, many real-world networks are large and continuously expanding, making the FPA scheme a promising tool (see Section 3.5.3). (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 (3.12) is considerably more tractable than (3.4), which paves the way for developing more efficient algorithms tailored for specific problems (see Section 3.6).

3.3.3 Proof Sketch of Theorem 3.3.1

In this section, we sketch the proof of Theorem 3.3.1, 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:

$$\mu_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} \mu_j(t-1)}{d_i}\right) & t > 0 \end{cases}, \quad \text{for all } i \in V. \tag{3.13}$$

By Proposition 3.2.1, $\mathbf{q}^*$ 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 3.2.1, 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 3.3.1 (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 3.3.1(i) follows from the definition of $\mathbf{h}(\cdot)$. Proposition 3.3.1(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 3.2.1 and 3.2.2, 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 (3.1). 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 3.3.1 provides an upper bound of this variance for each period over the entire time horizon.

Lemma 3.3.1 (In-Neighbor Variance Bound) *Under Assumptions 3.2.1 and 3.2.2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *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 3.3.1 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:

$$\boldsymbol{\kappa}(t) \;\leq\; \lim_{t\to\infty}\frac{1}{4}\left[\mathbf{I} + \sum_{\tau=1}^{t-1}\left(\frac{\rho^2}{2}\right)^{\tau}\tilde{\mathbf{A}}^{\tau}\right]\mathbf{b} \;=\; \frac{1}{4}\left(\mathbf{I} - \frac{\rho^2}{2}\tilde{\mathbf{A}}\right)^{-1}\mathbf{b}, \text{ for all } t \geq 0. \tag{3.14}$$

With the variance bound (3.14), 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 3.3.2 (Deviation of Adoption Probability) *Under Assumptions 3.2.1 and 3.2.2, for any diffusion instance* $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ *and* $t \geq 1$*, we have*

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

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 3.3.1, Lemma 3.3.2 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 3.3.1. 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 3.3.1 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 $\mathbf{q}^*$.

3.4 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 3.4.1.

Assumption 3.4.1 (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.

3.4.1 Improved Upper Bounds

Recall that Theorem 3.3.1 establishes an upper bound for the error of FPA at the order of $\left[\mathcal{C}(G;\rho)\right]^{\frac{1}{2}}$. Under Assumption 3.4.1, Theorem 3.4.1 below refines this upper bound to a lower

order of $\mathcal{C}(G;\rho) \leq \left[\mathcal{C}(G;\rho)\right]^{\frac{1}{2}}$, where the inequality follows immediately from $\mathcal{C}(G;\rho) \leq \mathbf{e}$. 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 3.4.1 (Improved Entry-Wise Error Bound of the FPA Scheme) *Under Assumptions 3.2.1, 3.2.2 and 3.4.1, for any diffusion instance $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$, we have*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\leq\; \tilde{C} \cdot \mathcal{C}(G;\rho) \quad \text{and} \quad \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\leq\; \tilde{C} \cdot \frac{1}{d_{\min}}. \tag{3.15}$$

Our inverse indegree centrality $\mathcal{C}(G;\rho)$ remains essential in the improved bounds. In light of the refined bound (3.15), we also sharpen the scaled ℓ_1-norm of the approximation error.

Corollary 3.4.1 (Improved Bound for Scaled ℓ_1-Norm) *Under Assumptions 3.2.1, 3.2.2 and 3.4.1, for any diffusion instance $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ with $r(G) < 1/\rho$, we have*

$$\frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\leq\; \frac{(1-\rho)\tilde{C}}{1-\rho r(G)} \cdot \mathcal{D}(G). \tag{3.16}$$

Finally, we remark that the proof of Theorem 3.4.1 largely parallels that of Theorem 3.3.1, with the primary difference in that we employ a second-order Taylor expansion of $F_\epsilon(\cdot)$, enabled by Assumption 3.4.1 to bound the deviation $|\mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t)|$. Lemma 3.4.1 refines Lemma 3.3.2 by removing the square root in the upper bound.

Lemma 3.4.1 (Improved Fixed-Point Deviation of Adoption Probability) *Under Assumptions 3.2.1, 3.2.2 and 3.4.1, for any diffusion instance $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ and $t \geq 1$, we have*

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

3.4.2 A Matching Lower Bound

Our results in Section 3.4.1 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 3.4.2 (A Lower Bound with Regular Graphs) *For any $(n,d) \in \mathbb{Z}_+^2$ with $d \leq n$, there exists a diffusion instance $\big(G, \mathbf{v}, F_\epsilon(\cdot), \beta\big)$ satisfying Assumptions 3.2.1, 3.2.2 and 3.4.1 such that $|V| = n$, G is d-regular and*

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\geq\; c \cdot \frac{1}{d}\mathbf{e}, \tag{3.17}$$

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

$$|\mathbf{q}^* - \boldsymbol{\mu}^*| \;\geq\; c \cdot \mathcal{C}(G;\rho), \quad \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_\infty \;\geq\; c \cdot \frac{1}{d_{\min}}, \quad \text{and} \quad \frac{1}{n}\|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \;\geq\; c \cdot \mathcal{D}(G). \tag{3.18}$$

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 (3.18) follows from (3.17). 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 (3.15) and (3.16) 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 (3.17), 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 $\epsilon_i(t) \sim \text{Logistic}(0,1)$[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 (3.17).

3.5 Numerical Experiments

In this section, we conduct a series of numerical studies, spanning across the 10-node example (i.e., Figure 3.1a), 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 3.2.2, 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 3.2, 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 B.2.3 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

[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 B.2.2 for details.

a convergence criterion of 10^{-5}. The noise distribution follows $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ 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 $\text{MAPE} = \frac{1}{n}\sum_{i\in V}\frac{|\mu_i^* - q_i^*|}{q_i^*} \times 100\%$, 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.

3.5.1 Revisiting the Motivating Example

In this subsection, we revisit the 10-node motivating example introduced in Section 3.3.1 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 3.2.2, 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 $\boldsymbol{\mu}(0) = \mathbf{0}$.

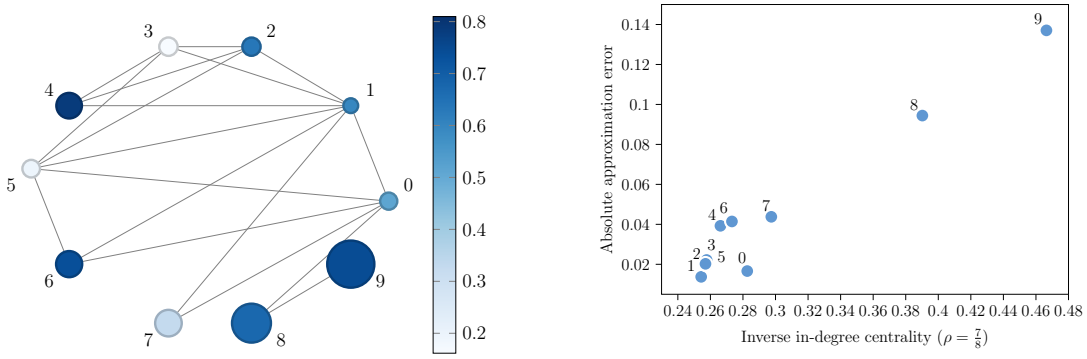
3.5.1.1 Discussions on the Inverse In-degree Centrality

As highlighted in Theorem 3.3.1, 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 3.2, 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.

3.5.1.2 Discussions on Assumption 3.2.2

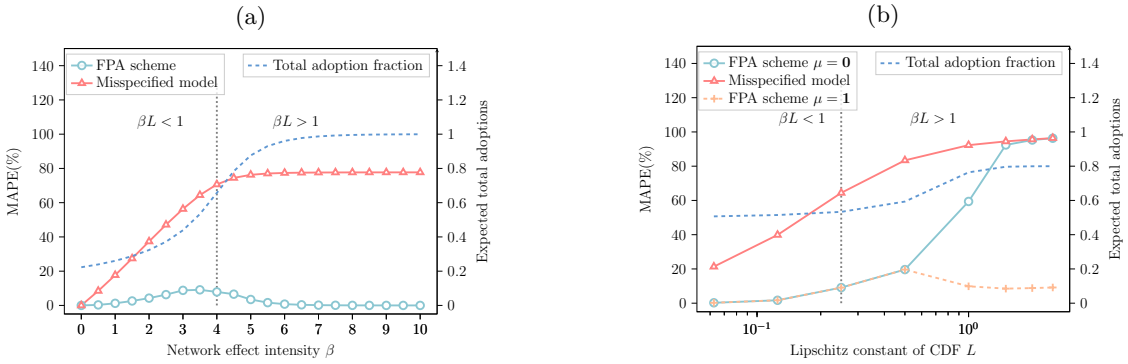
Although similar assumptions as $\rho = \beta L < 1$ are commonly made in the network literature (e.g., see Huang et al. 2022), its implications on the FPA scheme warrant further exploration—especially when this assumption is not satisfied. The parameter ρ has two key elements, namely the network effect 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 3.3.1). Figure 3.3 plots the MAPE of the FPA scheme for instances with different values of β and L.

Figure 3.2: Analysis of the 10-Node Example Instance



Notes. Left: Reproduction of the network diagram from Figure 3.1a. Right: Illustration of the relationship between the approximation error and inverse in-degree centrality.

Figure 3.3: Sensitivity of the FPA Error Against the Discount Parameter ρ



Notes. In right subfigure, FPA scheme 0 (1, resp.) represents the FPA solution initialized with $\boldsymbol{\mu}(0) = \mathbf{0}$ ($\boldsymbol{\mu}(0) = \mathbf{1}$, resp.). (a) Sensitivity regarding network effect intensity. (b) Sensitivity regarding noise distribution.

(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 $L = f_\epsilon(0) = 0.25$, our experiments encompass both scenarios when Assumption 3.2.2 is satisfied or violated. As illustrated in Figure 3.3a, 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 $\rho = \beta L > 1$.

Our analysis can then be divided into two segments based on the value of ρ in relation

to 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 3.2.2, 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 3.2.2 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 3.3.1.

In Figure 3.3b, 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 3.3.1, 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 3.3b. 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 3.2.2 is breached in this context, the FPA scheme has multiple solutions, complicating its application. In short, the effects of violating Assumption 3.2.2

on the FPA's quality are multifaceted, contingent on the specific circumstances. We leave further explorations of these phenomena to future studies.

3.5.2 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 B.2.4. 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.

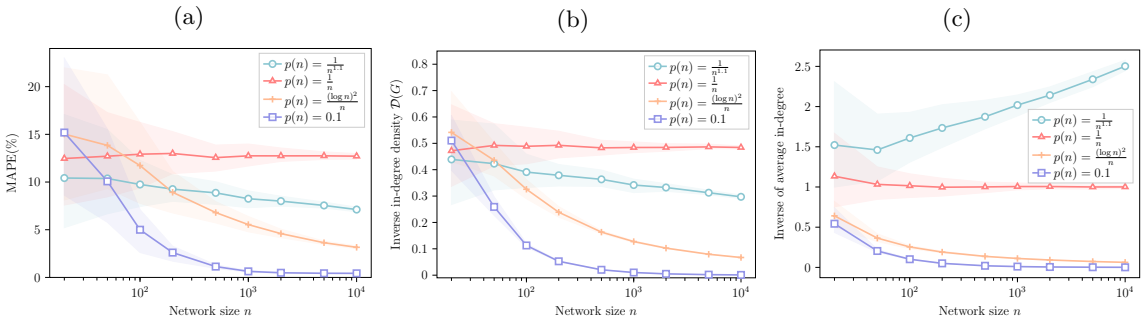
3.5.2.1 The FPA Scheme's Accuracy with Regard to Network Structure

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 B.2.5 (see, also, Huang et al. 2022).

In Figure 3.4a, 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 3.3 and 3.4, 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

Figure 3.4: Performance of the FPA Scheme on Erdős-Rényi Networks of Different Network Sizes



Notes. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval. (a) MAPE against n. (b) $\mathcal{D}(G)$ against n. (c) Inverse of average in-degree against n.

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.

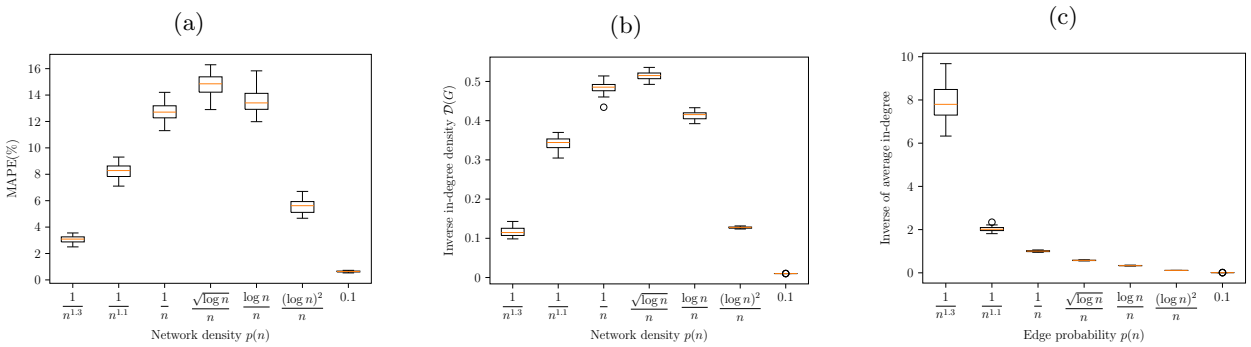
The observation in Figure 3.4a reveals that some traditional measures of network density, such as $p(n)$ and average degree, may not effectively evaluate the FPA scheme's performance, because critical information regarding network configuration is absent in these measures. To confirm this intuition, Figures 3.4b and 3.4c 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 3.5a, the MAPE first increases and then decreases as the network gets denser, reflecting a trend that aligns with Figure 3.4a.

The peak MAPE 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 3.5b and 3.5c 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 3.4b and 3.4c, again confirming that $\mathcal{D}(G)$ is a robust indicator of FPA'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 3.3.2, we show

Figure 3.5: Performance of the FPA Scheme on Erdős-Rényi Networks of Different Network Densities



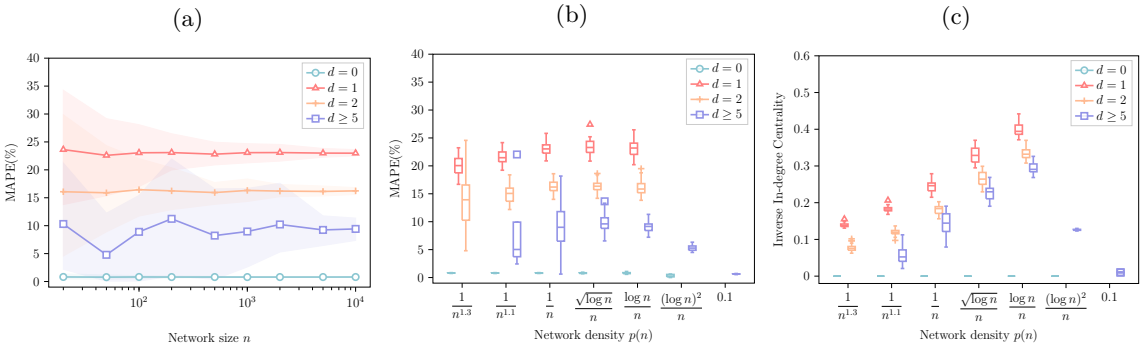
Notes. (a) MAPE against $p(n)$. (b) $\mathcal{D}(G)$ against $p(n)$. (c) Inverse of average in-degree against $p(n)$.

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 B.2.4).

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 3.6, 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 3.4a and 3.5a. These findings reaffirm our node-level theoretical predictions (Theorem 3.3.1 and 3.4.1), 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 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 3.6c. 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.

Figure 3.6: Degree-level MAPE of the FPA Scheme



Notes. (a) MAPE against 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. (b) MAPE against $p(n)$ and n is fixed to be $1,000$. (c) $\mathcal{C}(G;\rho)$ against $p(n)$.

3.5.2.2 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 100 to 10,000. As shown in Table 3.1, 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 3.1: The CPU Time Required for the MCMC Simulation and the FPA Scheme

Network size n	100	200	500	1,000	2,000	5,000	10,000
MCMC time (s)	0.5296	2.4152	18.9790	97.4521	286.7179	1315.3010	2366.8788
FPA time (s)	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.

3.5.3 Real-World Networks

Our numerical experiments in Section 3.5.2 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 3.2. For the raw data and additional summary statistics, readers can refer to the Network Repository website.[4]

Table 3.2: Experiment Results for Real-World Networks

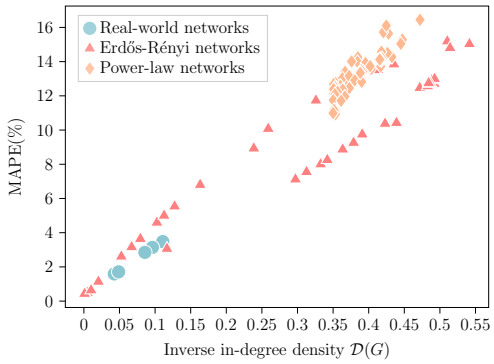
Instance	n	Avg. in-degree	$\mathcal{D}(G)$	MAPE(%)	MCMC time (s)	FPA time (s)
Caltech36	770	43.2623	0.1108	3.48	4.5335	0.0636
Reed98	963	39.0696	0.0962	3.14	5.6228	0.0623
Haverford76	1,447	82.3621	0.0427	1.59	23.2347	0.1009
Simmons81	1,519	43.4338	0.0857	2.85	11.9185	0.1426
Amherst41	2,236	81.3542	0.0488	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 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 3.7 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 3.7: Performance of the FPA Scheme Against the Inverse In-degree Density for Different Networks



Notes. Each node of the random networks represents the average values of the same pair of parameters across all repetitions.

3.6 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 (3.4) and its approximate formulation (3.12), 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:

$$\text{Regret}(\mathbf{x}) = g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}^*), F_\epsilon, \beta), \mathbf{x}^*\Big) - g\Big(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}), F_\epsilon, \beta), \mathbf{x}\Big), \tag{3.19}$$

where $\mathbf{x}^*$ is the optimal solution derived from original problem (3.4).

3.6.1 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 3.2.1, 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

$$\underset{S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} q_i^* = \lim_{t\to\infty} \sum_{i\in V} \mathbb{E}[Y_i(t)] \tag{3.20a}$$

$$\text{subject to} \quad Y_i(t) = \begin{cases} 1 & \forall i \in S, t \geq 1, \\ \mathbb{1}\left\{ v_i + \beta \frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0 \right\} & \forall i \in V \setminus S, t \geq 1. \end{cases} \tag{3.20b}$$

The objective (3.20a) is the limiting total expected adoptions. Constraint (3.20b) 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:

$$\underset{\boldsymbol{\mu}, S\subseteq V:|S|=K}{\text{maximize}} \quad \sum_{i\in V} \mu_i \tag{3.21a}$$

$$\text{subject to} \quad \mu_i = 1, \qquad \forall i \in S, \tag{3.21b}$$

$$\mu_i = 1 - F_\epsilon\left(-v_i - \beta \frac{\sum_{j\in\mathcal{N}_i} \mu_j}{d_i} \right), \ \forall i \in V \setminus S. \tag{3.21c}$$

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 (3.21) compared with the optimal solution to (3.20), S^*.

Proposition 3.6.1 (IM Regret Bound) *Under Assumptions 3.2.1 and 3.2.2, 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 3.4.1, can be extended to the approximate IM problem as well. For example, following (3.7), the worst-case regret bound in Proposition 3.6.1 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 (3.21) 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 3.6.1 (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 3.6.1 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 B.3.1.1. Under Assumption 3.6.1, the following theorem shows the submodularity of $g(\boldsymbol{\mu}^*(S), S)$.

Theorem 3.6.1 (Submodularity of Approximate IM Problem) *Under Assumptions 3.2.1, 3.2.2 and 3.6.1, $g(\boldsymbol{\mu}^*(S), S)$ is a submodular set function of seed set S.*

As an immediate corollary of Theorem 3.6.1, 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 3.6.1, the greedy algorithm provides a $(1 - 1/e)$-approximation solution to the approximate IM problem (3.21). Together with Proposition 3.6.1, the simple greedy approach also provides a high-quality solution to the original IM problem (3.20).

In summary, our approximate IM formulation presents several notable advantages. It facilitates establishing clear-cut conditions, such as Assumption 3.6.1, 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 B.3.1.2. Importantly, our results indicate that the greedy algorithm for the approximate IM problem achieves near-optimal solutions irrespective of whether Assumption 3.6.1 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.

3.6.2 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

$$\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{i\in V}\left(\sum_{k=1}^{m} W_{ik}p_k\right)\cdot q_i^* = \sum_{i\in V}\left(\sum_{k=1}^{m} W_{ik}p_k\right)\cdot \lim_{t\to\infty}\mathbb{E}[Y_i(t)] \tag{3.22a}$$

$$\text{subject to} \quad Y_i(t) = \mathbb{1}\left\{v_i - \gamma\sum_{k=1}^{m} W_{ik}p_k + \beta\frac{\sum_{j\in\mathcal{N}_i} Y_i(t-1)}{d_i} + \epsilon_i(t) \geq 0\right\} \quad \forall i \in V, t \geq 1. \tag{3.22b}$$

The objective (3.22a) represents the total profit and constraint (3.22b) 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:

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top\mathbf{W}\mathbf{p} \tag{3.23a}$$

$$\text{subject to} \quad \mu_i = 1 - F_\epsilon\left(-v_i + \gamma\sum_{k=1}^{m} W_{ik}p_k - \beta\frac{\sum_{j\in\mathcal{N}_i}\mu_j}{d_i}\right),\ \forall i \in V. \tag{3.23b}$$

We use $\boldsymbol{\mu}$ as an explicit decision variable and use constraint (3.23b) to link $\mathbf{p}$ and $\boldsymbol{\mu}$. The approximate problem (3.23) is generally nonconvex due to the constraint (3.23b) and challenging to solve. Similar to the IM problem, we first establish the regret bound for the optimal solution to (3.23), denoted by $\mathbf{p}^{\text{FPA}}$, before discussing how to solve (3.23).

Proposition 3.6.2 (Regret Bound for Approximate Pricing Problem) *Under Assumptions 3.2.1 and 3.2.2, 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 3.6.2 establishes a similar regret bound to Proposition 3.6.1, 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 3.6.1, our bound in Proposition 3.6.1 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. 2020a, 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.

3.6.2.1 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]

$$\underset{\boldsymbol{\mu},\mathbf{p}}{\text{maximize}} \quad \boldsymbol{\mu}^\top \mathbf{p} \tag{3.24a}$$

$$\text{subject to} \quad \mu_i = 1 - \frac{1}{1+\exp\{v_i + \beta \sum_{j\in\mathcal{N}_i} \mu_j/d_i - \gamma p_i\}},\ \forall i \in V. \tag{3.24b}$$

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

$$\underset{\boldsymbol{\mu}}{\text{maximize}} \quad \sum_{i\in V} \frac{1}{\gamma}\left(v_i + \beta \sum_{j\in\mathcal{N}_i} \frac{\mu_j}{d_i} + \ln\frac{1-\mu_i}{\mu_i}\right)\mu_i \tag{3.25a}$$

$$\text{subject to} \quad 0 \le \mu_i \le 1,\ \forall i \in V. \tag{3.25b}$$

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 3.6.2 (Concavity of Price Optimization) *The objective of the pricing problem* (3.25) *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 3.6.2 states that when the network diffusion intensity satisfies $0 < \beta \leq 3.375$, problem (3.25) 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 3.6.2 and Assumption 3.2.2 require the network effect parameter to be relatively small, a condition consistently made in the related literature (e.g., Candogan et al. 2012).

3.6.2.2 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:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}), \tag{3.26}$$

where the gradient of $\boldsymbol{\mu}(\mathbf{p})$ is not explicitly given. To obtain this gradient, we apply the implicit function theorem to (3.24b) (i.e., $\boldsymbol{\mu}(\mathbf{p}) = \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))$; see Appendix B.3.2.1 for details) and rewrite (3.26) as

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{-1} \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}). \tag{3.27}$$

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 B.3.2.1 and Appendix B.3.2.2 for more detailed discussions of the implementation and results of the experiments, respectively.

3.7 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.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this dissertation, I have explored the interface between social network analytics and operations management, addressing a range of challenges and opportunities presented by this interdisciplinary field. By employing various analytical tools and methods, I have managed to uncover significant insights that facilitate practical implementations in operations management.

The integration of the diffusion effect into platform policies, as demonstrated in Chapter 2, underscores the transformative potential of leveraging social network effects in operational strategies. The observed significant performance improvements validate the power of social dynamics to enhance operational efficiencies and outcomes. Moreover, the development of a general framework in Chapter 3 provides a means to streamline network operations, enabling more personalized and effective management practices that account for the complex web of social interactions.

In conclusion, this dissertation not only illuminates the complexities at the intersection of social network analytics and operations management but also sets the stage for future research and development in this promising field. As digital platforms continue to evolve and the volume of social network data expands, the methodologies and insights presented here will remain relevant and increasingly crucial.

4.2 Future Work

Inspired by my current research—which underscores the far-reaching influence of network interactions—I am committed to contributing to a more interconnected and enlightened future. This unified vision will guide my future research along two main avenues:

Optimizing network operations in practical contexts. The inherent complexities of networks demand innovative approaches for their practical operation. Broadly, I am driven to explore fundamental questions such as: *Can we quantify the relationship between network*

complexity and operational efficiency? and *How can insights from one type of network be efficiently transferred to another?* One compelling application that brings this focus to life is the realm of experimental design within social networks. Traditional A/B testing approaches often fall short when network effects are at play, introducing biases into strategy impact estimations. A case in point is the enduring challenge of attributing similar behaviors within a community to shared demographics (homophily) or social influence. Fueled by this challenge, I am eager to integrate social network analytics, machine learning, and statistical methods to design and execute robust large-scale randomized network experiments. I aim to strike a delicate balance between robust data collection for precise social influence inferences and practical, cost-effective real-world implementations.

Extending the reach of social goods through social network. The transformative power of social networks extends beyond the confines of private platforms—it holds the potential to reshape society itself. This potential drives my future research endeavors, as I aim to harness it for broad social benefits. For instance, I plan to apply a game-theoretical approach to explore how social networks can promote organic farming, comparing these influences with traditional farming subsidies to offer actionable insights that can guide more effective policymaking and foster sustainable agriculture.

In Section 4.2.1, I will present the ongoing work that serves as a bridge connecting the available social network data with complex network diffusion models.

4.2.1 Learning User Behavior in a Social Network with Limited Data

The role of network diffusion in analyzing user behavior data is profound, leading to a paradigm shift in its interpretation and application. This shift moves away from considering user behaviors as isolated individual actions, instead framing them within a context of intricate interdependence. Such a transition necessitates a focused approach toward calibrating network diffusion models. Traditional analytical approaches, which ignore the network effects, can lead to misinterpretations, incomplete insights, and erroneous decision-making.

In network studies, addressing the calibration challenge is also crucial, as it forms a vital link between theoretical results and real-world applications. The effectiveness of network-related models and algorithms is contingent upon accurately calibrating their parameters with available data. Such calibration not only facilitates the practical deployment of these models but also supports their use in downstream applications. These applications include, but are not limited to, predicting user behavior in a new social network structure and optimizing network operations. Take the network diffusion model in Chapter 3 as an example, in order to apply all of the results, we first need to learn the true value of v and β in reality. Therefore, calibration is an essential, practical step that transcends theoretical analysis, rendering network studies pertinent and actionable in diverse real-world settings.

Effective calibration involves the understanding of both individual characteristics and the network effects arising from interactions among individuals. There are primarily two chal-

lenges. Firstly, the network diffusion process is dynamic and governed by intricate network structures. This complexity means that the outcomes of diffusion often lack straightforward, closed-form expressions that would facilitate easy analysis. Additionally, theoretical analysis becomes even more complicated due to correlations resulting from network interactions. Secondly, calibrating network diffusion models is complicated by limitations in available data, both temporally and spatially. From the temporal perspective, although network diffusion is a dynamic process, capturing its entire trajectory may require extensive efforts and thus is often not feasible. This limitation often necessitates calibrating the network diffusion model with long-term behavior. From the spatial perspective, network structures are typically specific and limited. While individual-level data might be abundant, sampling network-level structures as desired is usually not possible. Consequently, the data collected may be overly specific to a certain network structure, thereby limiting their generalizability.

In the literature, several methods have been suggested to calibrate the network diffusion model. Some scholars propose the use of surrogate models to simplify complex dynamics and employ conventional estimation methods for calibration. Others propose the collection of panel data to compensate for data insufficiency, with this data offering richer and more diverse information for calibration. Nevertheless, these methods either compromise the precision of the diffusion dynamics or demand significant data collection efforts. Compromising on the diffusion dynamics can lead to loss of critical insights, and excessive data collection can impose logistical constraints and significant resource costs which may not always be feasible in reality.

In this work, we aim to present a comprehensive framework for learning user behavior in social networks, which involves calibrating diffusion models and generalizing to new diffusion instances. We aim at calibrating parameterized diffusion model based on one single observed network instances. Our approach assumes prior knowledge of the model but with unknown parameters and the parameters indicate (i) the dependence of users' intrinsic utility and features and (ii) the intensity of network effects.

Due to the aforementioned challenges of complex dynamics, directly obtaining the explicit dependence of the output data on the network behavior parameters is a formidable challenge. This necessitates the employment of simulation as a vital tool to evaluate the dependence. Specifically, simulation enables the modeling of complex systems and the observation of their behavior under various conditions. By constructing a simulation model of the network, we can generate output data corresponding to different network parameter values, circumventing the need for an explicit relationship between the parameters and the output data.

In addition, we also leverage simulation to generate output data for diverse combinations of the network parameters. Subsequently, we compare this simulated data to the observed data in an attempt to identify the parameter values that yield the closest match. This comparison is facilitated by the definition of an objective function, which quantifies the distance between the given data and the simulated data. Through the execution of simulation experiments, we can explore the parameter space and pinpoint the values that most accurately explain the diffusion dynamics in our network. This approach enables us to systematically address the problem, despite the intricate nature of the system and the scarcity of data.

Bibliography

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 Biró, P., Chawla, S., and Echenique, F., editors, *EC '21: The 22nd ACM Conference on Economics and Computation, Budapest, Hungary, July 18-23, 2021*, pages 2–3. ACM.

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.

AMZScout (2021). Amazon statistics for 2021 and the latest facts. `https://amzscout.net/blog/amazon-statistics/`. Accessed: 2022-03-11.

Anandhan, A., Shuib, L., Ismail, M. A., and Mujtaba, G. (2018). Social media recommender systems: review and open research issues. *IEEE Access*, 6:15608–15628.

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.

Besbes, O., Gur, Y., and Zeevi, A. (2016). Optimization in online content recommendation services: Beyond click-through rates. *Manufacturing & Service Operations Management*, 18(1):15–33.

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. (2019). Joint pricing and inventory management with strategic customers. *Operations Research*, 67(6):1610–1627.

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

Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In Sen, S., Geyer, W., Freyne, J., and Castells, P., editors, *Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016*, pages 191–198. ACM.

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

Davidson, J., Liebald, B., Liu, J., Nandy, P., Vleet, T. V., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., and Sampath, D. (2010). The youtube video recommendation system. In Amatriain, X., Torrens, M., Resnick, P., and Zanker, M., editors, *Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010, Barcelona, Spain, September 26-30, 2010*, pages 293–296. ACM.

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.

Easingwood, C. J., Mahajan, V., and Muller, E. (1983). A nonuniform influence innovation diffusion model of new product acceptance. *Marketing Science*, 2(3):273–295.

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

Feng, J., Bhargava, H. K., and Pennock, D. M. (2007). Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing*, 19(1):137–148.

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.

FOX32 (2020). Missing child alert canceled after 2 chipley girls found safe; mom in custody. `https://www.fox32chicago.com/news/missing-child-alert-canceled-after-2-chipley-girls-found-safe-mom-in-custody`. Accessed: 2022-03-16.

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. (2020a). No-regret learning in price competitions under consumer reference effects. *Advances in Neural Information Processing Systems*, 33:21416–21427.

Golrezaei, N., Nazerzadeh, H., and Randhawa, R. (2020b). Dynamic pricing for heterogeneous time-sensitive customers. *Manufacturing & Service Operations Management*, 22(3):562–581.

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

Graffius, S. M. (2022). Lifespan (life span) of social media posts. `https://www.scottgraffius.com/blog/files/tag-lifespan-0028half-life0029-of-social-media-posts003a-update-for-2022.html`. Accessed: 2023-05-11.

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., Milner, J., and Wu, J. (2016). Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science*, 62(3):867–879.

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.

Jiang, Z. and Jain, D. C. (2012). A generalized norton–bass model for multigeneration diffusion. *Management Science*, 58(10):1887–1897.

Kempe, D., Kleinberg, J. M., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In Getoor, L., Senator, T. E., Domingos, P. M., and Faloutsos, C., editors, *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003*, pages 137–146. ACM.

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.

Kiesling, E., Günther, M., Stummer, C., and Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, 20(2):183–230.

Kitts, B., Freed, D., and Vrieze, M. (2000). Cross-sell: a fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities. In Ramakrishnan, R., Stolfo, S. J., Bayardo, R. J., and Parsa, I., editors, *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, MA, USA, August 20-23, 2000*, pages 437–446. ACM.

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., Zhang, H., Zhang, R., and Shen, Z.-J. M. (2024). Content promotion for online content platforms with the diffusion effect. *Manufacturing & Service Operations Management*.

Lin, Y., Zhang, H., Zhang, R. P., and Shen, Z.-J. M. (2022). Nonprogressive diffusion on social networks: Approximation and applications. *Available at SSRN 4232670*.

Liu, J., Dolan, P., and Pedersen, E. R. (2010). Personalized news recommendation based on click behavior. In Rich, C., Yang, Q., Cavazza, M., and Zhou, M. X., editors, *Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI 2010, Hong Kong, China, February 7-10, 2010*, pages 31–40. ACM.

Lu, W., Chen, S., Li, K., and Lakshmanan, L. V. (2014). Show me the money: dynamic recommendations for revenue maximization. *Proceedings of the VLDB Endowment*, 7(14):1785–1796.

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.

Norton, J. A. and Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management science*, 33(9):1069–1086.

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.

Schmittlein, D. C. and Mahajan, V. (1982). Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing science*, 1(1):57–78.

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.

Srinivasan, V. and Mason, C. H. (1986). Nonlinear least squares estimation of new product diffusion models. *Marketing science*, 5(2):169–178.

Statista (2021). Youtube average video length by category 2018. `https://www.statista.com/statistics/1026923/youtube-video-category-average-length/`. Accessed: 2022-03-11.

Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Adv. Artif. Intell.*, 2009:421425:1–421425:19.

Sultan, F., Farley, J. U., and Lehmann, D. R. (1990). A meta-analysis of applications of diffusion models. *Journal of marketing research*, 27(1):70–77.

Teng, J. T., Grover, V., and Guttler, W. (2002). Information technology innovations: general diffusion patterns and its relationships to innovation characteristics. *IEEE transactions on engineering management*, 49(1):13–27.

TikTok (2021). #squidgame hashtag videos on tiktok. `https://www.tiktok.com/tag/squidgame`.

Vahabi, H., Koutsopoulos, I., Gullo, F., and Halkidi, M. (2015). Difrec: A social-diffusion-aware recommender system. In Bailey, J., Moffat, A., Aggarwal, C. C., de Rijke, M., Kumar, R.,

Murdock, V., Sellis, T. K., and Yu, J. X., editors, *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, pages 1481–1490. ACM.

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.

YouTube (2021). Youtube for press. `https://blog.youtube/press/`.

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.

Zhang, H. and Vorobeychik, Y. (2019). Empirically grounded agent-based models of innovation diffusion: a critical review. *Artificial Intelligence Review*, pages 1–35.

Appendix A

Supplemental Materials for Chapter 2

A.1 Proofs and Supplements

A.1.1 Proofs and Supplements for Section 2.4

A.1.1.1 NP-Hardness of the CGPO Problem

Proof of Theorem 2.4.1: We prove the NP-hardness of the CGPO problem by reduction from the SUBSET-SUM problem. Let the positive integers $b_1, \dots b_n$, and B form an instance of SUBSET-SUM. Without loss of generality, we assume that $b_i \leq B$ for all $i \in \{1, 2, \dots, n\}$. Let $s_i = b_i/B$. We have $s_i \in [0, 1]$ for all i.

We construct a special instance of the CGPO problem where $\mathcal{V} = \{1, 2, \dots, n\}$, $L = 2$, $C = m$, $K \in \mathbb{Z}_+$, $A_{v,0} = 0$ and $p_v \leq 1/2$ for all $v \in \mathcal{V}$. By this construction, the optimal solution satisfies $x^*_{v,2} = 0$ for all $v \in \mathcal{V}$. We show this claim by contradiction. Assume the optimal solution $x^*_{v,2} > 0$. We construct a feasible solution: $x'_{v,1} = x^*_{v,1} + x^*_{v,2}$ and $x'_{v,2} = 0$. Let $x^*_{v,1} + x^*_{v,2} = c_v \leq 1$. The optimal adoption is given by

$$A^*_{v,2} = -p_v^2 q_v m {x^*_{v,1}}^2 + p_v m q_v x^*_{v,1} + p_v m (x^*_{v,1} + x^*_{v,2}) = -p_v^2 q_v m {x^*_{v,1}}^2 + p_v m q_v x^*_{v,1} + p_v m c_v \,.$$

Similarly, the adoptions with $x'_{v,1}$ and $x'_{v,2}$ is $A'_{v,2} = -p_v^2 q_v m {x'_{v,1}}^2 + p_v m q_v x'_{v,1} + p_v m c_v$. Given $p_v \leq 1/2$, we have $A'_{v,2} - A^*_{v,2} = p_v q_v m x^*_{v,2}(1 - p_v(x'_{v,1} + x^*_{v,1})) \geq p_v q_v m x^*_{v,2}(1 - 2p_v) \geq 0$, contradicting with the assumption.

Therefore, we can omit the variables $\boldsymbol{x}_{:,2}$ and write $\boldsymbol{x} = \boldsymbol{x}_{:,1}$ for simplicity. Let $\alpha_v = p_v^2 q_v m$ and $\beta_v = p_v m(q_v + 1)$. Under this construction, the CGPO problem can be expressed as

$$\begin{aligned} \max_{\boldsymbol{x}, U \subseteq \mathcal{V}: |U| \leq K} \quad & \sum_{v \in \mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v \\ \text{s.t.} \quad & \sum_{v \in \mathcal{V}} x_v \leq 1, \\ & 0 \leq x_v \leq \mathbb{1}\{v \in U\}, \forall v \in \mathcal{V}. \end{aligned}$$

We assume that there exists values of $\boldsymbol{p}$, $\boldsymbol{q}$ and m such that for all $v \in \mathcal{V}$, $-2\alpha_v s_v + \beta_v = d_1$ and $\alpha_v s_v^2 = d_2$ for some $d_1, d_2 \in \mathbb{R}_+$. We justify the existence of $\boldsymbol{p}$, $\boldsymbol{q}$, and m at the end of this proof. Here, we let $\ell(\boldsymbol{x}) = \sum_{v\in\mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v$. We claim that

$$\ell(\boldsymbol{x}^*) \geq d_1 + d_2 K \,, \iff \text{ there exists } U \subseteq \mathcal{V} \text{ with } |U| = K \text{ and } \sum_{v\in U} s_v = 1 \,.$$

To prove this claim, we first express the objective value as

$$\begin{aligned}\ell(\boldsymbol{x}) = \sum_{v\in\mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v &\overset{(a)}{=} \sum_{v\in U} -\alpha_v x_v^2 + \beta_v x_v \\ &\overset{(b)}{=} \sum_{v\in U} -\alpha_v (x_v - s_v)^2 + d_1 x_v + d_2 \overset{(c)}{\leq} d_1 \sum_{v\in U} x_v + d_2 |U| \overset{(d)}{\leq} d_1 + d_2 K \,.\end{aligned}$$

where (a) follows from constraint $x_v \leq \mathbb{1}\{v \in U\}$; (b) follows from the definition of d_1 and d_2; (c) follows since $\alpha_v \geq 0$; (d) follows from the constraint $\sum_{v\in\mathcal{V}} x_v \leq 1$. If there exists $U \subseteq \mathcal{V}$ such that $|U| = K$ and $\sum_{v\in U} s_v = 1$, then we can let $U^* = U$ and $x_v^* = s_v \mathbb{1}\{v \in U\}$ for all $v \in \mathcal{V}$. Then, it is easy to verify that $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$. On the other hand, if $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$, then the (c) implies that we must have $x_v^* = s_v \mathbb{1}\{v \in U^*\}$; (d) implies that $\sum_{v\in U^*} x_v^* = \sum_{v\in U^*} s_v = 1$ and $|U^*| = K$.

This claim allows us to reduce SUBSET-SUM to our problem as follows. If there exists a subset $I \subseteq \{1, \ldots, n\}$ such that $\sum_{i\in I} s_i = 1$, then the objective value $\ell(\boldsymbol{x}^*)$ is equal to $d_1 + d_2 K$ for $K = |I|$. If there exists $K \in \{1, 2, \ldots, n\}$ such that $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$, one can find $I = U^*$ such that $\sum_{i\in I} s_i = 1$.

We then proceed by providing an example of the values of m, d_1, and d_2, which ensures that our construction serves as a valid CGPO instance. Let $\underline{s} = \min_{v\in\mathcal{V}} s_v$, $m = 64/\underline{s}^2$, $d_1 = m/2$ and $d_2 = 1$. Our construction assumes

$$\begin{cases} -2p_v^2 q_v m s_v + p_v m (q_v + 1) & = d_1 \\ p_v^2 q_v m s_v^2 & = d_2 \end{cases} , \Longrightarrow \begin{cases} p_v q_v + p_v & = \frac{1}{2} + \frac{2}{m s_v} \\ p_v q_v \cdot p_v & = \frac{1}{m s_v^2} \end{cases} .$$

Therefore, for any given s_v, we can solve p_v and q_v as demonstrated in (A.1):

$$p_v = \frac{\frac{1}{2} + \frac{2}{m s_v} + \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}}{2} \text{ and } q_v = \frac{\frac{1}{2} + \frac{2}{m s_v} - \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}}{\frac{1}{2} + \frac{2}{m s_v} + \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}} . \tag{A.1}$$

To ensure that p_v and q_v provided in (A.1) are valid within the context of P-BDM, we need further justifications. First, we need to ensure (A.1) has real value solutions, that is,

$$\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2} = \frac{1}{4} + \frac{2}{m s_v} + \left(\frac{4}{m^2} - \frac{4}{m}\right)\frac{1}{s_v^2} \geq \frac{1}{4} + \frac{2}{m} + \left(\frac{4}{m^2} - \frac{4}{m}\right)\frac{1}{\underline{s}^2} = \frac{3}{16} + \frac{33}{16m} \geq 0 \,.$$

Next, we need to validate that $p_v, q_v \geq 0$, $p_v \leq 1/2$, and $p_v + q_v \leq 1$. It is obvious that $p_v, q_v \geq 0$ by (A.1). In order to show $p_v + q_v \leq 1$, it suffices to show that $p_v, q_v \leq 1/2$. To show $p_v \leq 1/2$, we have

$$\sqrt{\left(\frac{1}{2}+\frac{2}{ms_v}\right)^2-\frac{4}{ms_v^2}} \leq \frac{1}{2}-\frac{2}{ms_v} \Longleftrightarrow \frac{1}{2}-\frac{2}{ms_v} \geq 0,$$
$$\text{and } \sqrt{\left(\frac{1}{2}-\frac{2}{ms_v}\right)^2+\frac{4}{ms_v}-\frac{4}{ms_v^2}} \leq \frac{1}{2}-\frac{2}{ms_v},$$

where the inequalities follow since $0 \leq s_v \leq 1$. To show $q_v \leq 1/2$, we have

$$\frac{1}{2}+\frac{2}{ms_v} \leq 3\sqrt{\left(\frac{1}{2}+\frac{2}{ms_v}\right)^2-\frac{4}{ms_v^2}} \Longleftrightarrow \left(\frac{1}{2}+\frac{2}{ms_v}\right)^2-\frac{9}{2ms_v^2} \geq \frac{23}{128}+\frac{33}{16m} \geq 0\,.$$

Therefore, the construction of this specific CGPO instance is valid. Such construction exists for every instance of SUBSET-SUM. In conclusion, the CGPO problem is NP-hard. □

A.1.1.2 Proofs for the PO-CR Problem

Proof of Theorem 2.4.2: To establish the equivalence of two problems, we need to show that for all optimal solutions of the PO-CR problem, equalities hold for all constraints (2.5). We will show this by contradiction.

Let $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ and R^* be the optimal solution and optimal value of the PO-CR problem. Assume there exists $u \in U$ and $\tau \in \{1, \ldots, L\}$ such that the following inequality holds: $A^*_{u,\tau} < A^*_{u,\tau-1} + p_v m x^*_{u,\tau} + \frac{q_u}{m} A^*_{u,\tau-1}(m - A^*_{u,\tau-1})$.

By the optimality of $(\boldsymbol{x}^*, \boldsymbol{A}^*)$, it is straightforward to confirm that $(\boldsymbol{x}^*_{U,\tau+1:L}, \boldsymbol{A}^*_{U,\tau+1:L})$ is also the optimal solution of the following problem (A.2). Consequently, the optimal value of problem (A.2) is equal to R^*.

$$\max_{\boldsymbol{x}\geq \boldsymbol{0}, \boldsymbol{A}_{U,\tau+1:L}} \quad \sum_{v\in U} A_{v,L} \tag{A.2a}$$
$$\text{s.t.} \quad A_{v,\tau+1} \leq A^*_{v,\tau} + p_v m x_{v,\tau+1} + q_v \frac{A^*_{v,\tau}}{m}(m - A^*_{v,\tau}), \quad \forall v \in U, \tag{A.2b}$$
$$x_{v,\tau+1} \leq 1 - \frac{A^*_{v,\tau}}{m}, \quad \forall v \in U, \tag{A.2c}$$
$$A_{v,t} \leq A_{v,t-1} + p_v m x_{v,t} + q_v \frac{A_{v,t-1}}{m}(m - A_{v,t-1}), \forall v \in U \ \ \forall t = \tau+2, \ldots, L, \tag{A.2d}$$
$$x_{v,t} \leq 1 - \frac{A_{v,t-1}}{m}, \quad \forall v \in U \ \ \forall t = \tau+2, \ldots, L, \tag{A.2e}$$
$$m \sum_{t=\tau+1}^{L} \sum_{v\in U} x_{v,t} \leq C - m \sum_{t=1}^{\tau} \sum_{v\in U} x^*_{v,t}. \tag{A.2f}$$

We then construct a feasible solution $(\boldsymbol{x}', \boldsymbol{A}')$ for the PO-CR problem and let R' be the objective value:

(i) When $t = 1, 2, \ldots, \tau - 1$, $\boldsymbol{x}'_{U,t} := \boldsymbol{x}^*_{U,t}$; When $t = 0, 1, \ldots, \tau - 1$, $\boldsymbol{A}'_{U,t} := \boldsymbol{A}^*_{U,t}$.

(ii) When $t = \tau$, $\boldsymbol{x}'_{U,t} := \boldsymbol{x}^*_{U,t}$ and

$$A'_{v,t} := \begin{cases} A^*_{v,t-1} + p_v m x^*_{u,t-1} + q_v \frac{A^*_{v,t-1}}{m}(m - A^*_{v,t-1}), & \text{when } v \in \{u\}, \\ A^*_{v,t}, & \text{when } v \in U \setminus \{u\}. \end{cases}$$

(iii) When $t = \tau + 1, \ldots, L$, $(\boldsymbol{x}'_{U,t}, \boldsymbol{A}'_{U,t})$ is defined as the optimal solution of the following problem (A.3):

$$\max_{\boldsymbol{x} \geq \boldsymbol{0}, \boldsymbol{A}_{U,\tau+1:L}} \quad \text{(A.2a)}$$

$$\text{s.t.} \quad A_{v,\tau+1} \leq A'_{v,\tau} + p_v m x_{v,\tau+1} + \frac{q_v}{m} A'_{v,\tau}(m - A'_{v,\tau}), \forall v \in U, \tag{A.3a}$$

$$x_{v,\tau+1} \leq 1 - \frac{A'_{v,\tau}}{m}, \qquad \forall v \in U, \tag{A.3b}$$

$$\text{(A.2d)} - \text{(A.2f)}.$$

As a consequence, the optimal value of problem (A.3) is also equal to R'.

In the following, we focus on problems (A.2) and (A.3). We aim to show that $R' > R^*$, which contradicts with the optimality of $(\boldsymbol{x}^*, \boldsymbol{A}^*)$. Firstly, to characterize the optimal solution for problem (A.2), we use Lagrangian multipliers to introduce the constraints (A.2b), (A.2c) and (A.2f) into the objective function. Let Ω denote the feasible region constructed by constraints (A.2d), (A.2e). The dual problem can thus be expressed as

$$\min_{\boldsymbol{\lambda} \geq \boldsymbol{0}, \boldsymbol{\mu} \geq \boldsymbol{0}, \theta \geq 0} r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta) + \sum_{v \in U} \lambda_v \left(A^*_{v,\tau} + \frac{q_v}{m} A^*_{v,\tau}(m - A^*_{v,\tau}) \right) + \sum_{v \in U} \mu_v \left(1 - \frac{A^*_{v,\tau}}{m} \right) + \theta C \,, \tag{A.4}$$

where $r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta)$ is the optimal value function of a parameterized maximization problem as shown in (A.5):

$$r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta) := \max_{\substack{\boldsymbol{x}_{U,\tau+1:L} \geq 0, \\ (\boldsymbol{x}, \boldsymbol{A})_{U,\tau+1:L} \in \Omega}} \sum_{v \in U} \Bigg[A_{v,L} - \lambda_v A_{v,\tau+1} + (\lambda_v m p_v - \mu_v - \theta m) x_{v,\tau+1} - \theta m \left(\sum_{t=1}^{\tau} x^*_{v,t} + \sum_{t=\tau+2}^{L} x_{v,t} \right) \Bigg] . \tag{A.5}$$

Notice that, in the maximization problem (A.5), variable $\boldsymbol{x}_{U,\tau+1}$ is not constrained by any other variables, but only by a constant 0. Thus, we can split problem (A.5) into two subproblems:

$$\max_{\substack{\boldsymbol{x}_{U,\tau+2:L} \geq 0, \\ (\boldsymbol{x}, \boldsymbol{A})_{U,\tau+1:L} \in \Omega}} \sum_{v \in U} \left[A_{v,L} - \lambda_v A_{v,\tau+1} - \theta m \left(\sum_{t=1}^{\tau} x^*_{v,t} + \sum_{t=\tau+2}^{L} x_{v,t} \right) \right]$$

$$\text{and} \max_{\boldsymbol{x}_{U,\tau+1}\geq 0} \sum_{v\in U} \left(\lambda_v m p_v - \mu_v - \theta m\right) x_{v,\tau+1}\,.$$

We then optimize two subproblems separately. For the former subproblem, we denote its optimal value as $s(\boldsymbol{\lambda}, \theta)$. For the latter subproblem, the maximization problem can be further decomposed for each content piece $v \in U$ and we can easily derive the optimal solution and the value of function $r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta)$ as follows:

$$x^*_{v,\tau+1} = \begin{cases} 0 & \text{when } \mu_v \geq m(\lambda_v p_v - \theta), \\ +\infty & \text{when } \mu_v < m(\lambda_v p_v - \theta), \end{cases}$$

$$\text{and } r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta) = \begin{cases} s(\boldsymbol{\lambda}, \theta) & \text{when } \theta > \min_{v\in U}(\lambda_v p_v - \frac{\mu_v}{m}), \\ +\infty & \text{o.w.} \end{cases}$$

We then substitute the value function of $r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta)$ into the dual problem (A.4). To minimize the dual value, it is obvious that $r(\boldsymbol{\lambda}, \boldsymbol{\mu}, \theta)$ should not be infinity. Hence, the dual problem (A.4) can be reformulated as (A.6),

$$\min_{\boldsymbol{\lambda}\geq 0, \theta\geq 0} \Bigg[s(\boldsymbol{\lambda}, \theta) + \sum_{v\in U} \lambda_v \left(A^*_{v,\tau} + \frac{q_v}{m} A^*_{v,\tau}(m - A^*_{v,\tau}) \right) + \theta C + \sum_{v\in U} \min_{\substack{\mu_v \geq 0, \\ \mu_v \geq m(\lambda_v p_v - \theta)}} \mu_v \left(1 - \frac{A^*_{v,\tau}}{m}\right) \Bigg]. \tag{A.6}$$

The inner minimization of problem (A.6) can be represented as (A.7). For any $v \in U$,

$$\min_{\substack{\mu_v \geq 0, \\ \mu_v \geq m(\lambda_v p_v - \theta)}} \mu_v \left(1 - \frac{A^*_{v,\tau}}{m}\right). \tag{A.7}$$

Since $A^*_{v,\tau} \leq m$ always holds, when given the value of $(\boldsymbol{\lambda}, \theta)$, the optimal solution of μ_v should be $\mu^*_v(\boldsymbol{\lambda}, \theta) = \max\{0, m(\lambda_v p_v - \theta)\}$. Finally, we define the dual value function as

$$u(\boldsymbol{\lambda}, \theta; \boldsymbol{A}^*_{U,\tau}) := s(\boldsymbol{\lambda}, \theta) + \sum_{v\in U} \lambda_v \left(A^*_{v,\tau} + \frac{q_v}{m} A^*_{v,\tau}(m - A^*_{v,\tau}) \right) + \theta C + \sum_{v\in U} \left(1 - \frac{A^*_{v,\tau}}{m}\right) \max\{0, m(\lambda_v p_v - \theta)\}.$$

Similarly, for problem (A.3), we can also have the dual value function as

$$u(\boldsymbol{\lambda}, \theta; \boldsymbol{A}'_{U,\tau}) := s(\boldsymbol{\lambda}, \theta) + \sum_{v\in U} \lambda_v \left(A'_{v,\tau} + \frac{q_v}{m} A'_{v,\tau}(m - A'_{v,\tau}) \right) + \theta C + \sum_{v\in U} \left(1 - \frac{A'_{v,\tau}}{m}\right) \max\{0, m(\lambda_v p_v - \theta)\}.$$

Let $(\boldsymbol{\lambda}^*, \theta^*)$ and $(\boldsymbol{\lambda}', \theta')$ be the optimal dual variables for the dual problems of (A.2) and (A.3). Given that problems (A.2) and (A.3) are convex optimization problems, by invoking Slater's condition, strong duality holds for both problems. Thus, the optimal values for problems (A.2) and (A.3) can be represented as $R^* = u(\boldsymbol{\lambda}^*, \theta^*; \boldsymbol{A}^*_{U,\tau})$ and $R' = u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_{U,\tau})$. Finally, we compare between R^* and R' and show that $R^* \leq R'$.

$$
\begin{aligned}
R^* - R' &= u(\boldsymbol{\lambda}^*, \theta^*; \boldsymbol{A}^*_\tau) - u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_\tau) \leq u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}^*_\tau) - u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_\tau) && \text{(A.8a)}\\
&= \sum_{v\in U} \lambda'_v \left(A^*_{v,\tau} + \frac{q_v}{m} A^*_{v,0}(m - A^*_{v,\tau}) \right) + \sum_{v\in U} \left(1 - \frac{A^*_{v,\tau}}{m} \right) \max\{0, m(\lambda'_v p_v - \theta')\}\\
&\quad - \left[\sum_{v\in U} \lambda'_v \left(A'_{v,\tau} + \frac{q_v}{m} A'_{v,\tau}(m - A'_{v,\tau}) \right) + \sum_{v\in U} \left(1 - \frac{A'_{v,\tau}}{m} \right) \max\{0, m(\lambda'_v p_v - \theta')\} \right]\\
&= \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \left[\lambda'_v \left(1 + q_v - \frac{q_v}{m}(A^*_{v,\tau} + A'_{v,\tau}) \right) - \frac{1}{m} \max\{0, m(\lambda'_v p_v - \theta')\} \right]\\
&\leq \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \left[\lambda'_v \left(1 + q_v - \frac{q_v}{m}(A^*_{v,\tau} + A'_{v,\tau}) \right) - \lambda'_v p_v \right] && \text{(A.8b)}\\
&< \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \lambda'_v (1 - q_v - p_v) && \text{(A.8c)}\\
&\leq 0, && \text{(A.8d)}
\end{aligned}
$$

where (A.8a) follows since $(\boldsymbol{\lambda}^*, \theta^*)$ is the optimal solution to dual problem (A.4); (A.8b) follows due to the construction of $\boldsymbol{A}^*_\tau \leq \boldsymbol{A}'_\tau$ and the inherent non-negativity of dual variables (i.e., $\boldsymbol{\lambda}' \geq \mathbf{0}$ and $\theta' \geq 0$); (A.8c) follows since by definition $q_v \geq 0$, $A^*_{u,\tau} < A'_{u,\tau} \leq m$ and $A^*_{v,\tau} = A'_{u,\tau} \leq m$ for all $v \in U \setminus \{u\}$; (A.8d) follows from the definition $p_v + q_v \leq 1$.

As a result, $(\boldsymbol{x}', \boldsymbol{A}')$ is a feasible solution to the PO-CR problem while the resulting objective value R' is greater than R^*. This indicates that $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ cannot be an optimal solution to the PO-CR problem.

In conclusion, the original PO problem (2.4) and the relaxed problem are equivalent. □

A.1.1.3 Lemmas and Proofs for the Single-Variable Reformulation

In the following, we include lemmas and proofs to validate the single-variable reformulation and to show that it remains a convex program.

Lemma A.1.1 (Redundant Constraint) *Constraint* (2.6c) *is redundant to the reformulation* (2.6).

Proof of Lemma A.1.1: Let $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ be the optimal solution of the PO problem (2.4). By P-BDM dynamics (2.4b), we can deduce that the optimal adoption number $\boldsymbol{A}^*$ is non-decreasingly evolves over time (i.e., $A_{v,0} \leq A^*_{v,1} \leq \cdots \leq A^*_{v,L}$ for all $v \in U$). By Theorem

2.4.2, $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ is also the optimal solution to the PO-CR problem. Therefore, for all $v \in U, t = 1, 2, \ldots, L$, constraint $x_{v,t} \leq 1 - A_{v,0}/m$ is redundant compared with $x_{v,t} \leq 1 - A_{v,t-1}/m$ in the PO-CR problem. □

Construction of Feasible Solution to Problem (2.7)**.**

For any $v \in U$, given $\boldsymbol{x}_{v,:} \in [0, 1 - A_{v,0}/m]^L$, a feasible solution can be constructed by the following three steps:

Step 1: Set $t := 1$. Let $A_{v,1} := A_{v,0} + p_v m x_{v,1} + q_v A_{v,0}(m - A_{v,0})/m$. Then, increment t by 1, i.e., $t := t + 1$.

Step 2: Let $A_{v,t} = \min\{m(1 - x_{v,t+1}), A_{v,t-1} + p_v m x_{v,t} + q_v A_{v,t-1}(m - A_{v,t-1})/m\}$.

Step 3: Increment t by 1, i.e., $t := t + 1$. Repeat step 2, until $t = L + 1$. □

To demonstrate that single-variable reformulation (2.6) remains a convex program, it is sufficient to show that the objective is concave, given that all constraints are linear.

Proof of Lemma 2.4.1: For simplicity, we omit the subscript v here. For any $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, let $\boldsymbol{x}^{(\lambda)} = \lambda \boldsymbol{x}^{(1)} + (1 - \lambda)\boldsymbol{x}^{(2)}$, and we want to show that $\lambda f(\boldsymbol{x}^{(1)}) + (1 - \lambda) f(\boldsymbol{x}^{(2)}) \leq f(\boldsymbol{x}^{(\lambda)})$ holds for any $0 \leq \lambda \leq 1$.

Suppose $\boldsymbol{A}^{(1*)}$, $\boldsymbol{A}^{(2*)}$ and $\boldsymbol{A}^{(\lambda*)}$ are the optimal solutions to problem (2.7) with regard to $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$, and $\boldsymbol{x}^{(\lambda)}$, respectively. Let $\boldsymbol{A}^{(\lambda)} = \lambda \boldsymbol{A}^{(1*)} + (1 - \lambda)\boldsymbol{A}^{(2*)}$. We first show that $\boldsymbol{A}^{(\lambda)}$ is a feasible solution to problem (2.7) with regard to $\boldsymbol{x}^{(\lambda)}$ by verifying that it satisfies constraints (2.5) and (2.4c).

For constraint (2.5), we have

$$
\begin{aligned}
& A_{t-1}^{(\lambda)} + pmx_t^{(\lambda)} + \frac{q}{m} A_{t-1}^{(\lambda)}(m - A_{t-1}^{(\lambda)}) - A_t^{(\lambda)} \\
=& \lambda A_{t-1}^{(1*)} + (1 - \lambda) A_{t-1}^{(2*)} + pm \left[\lambda x^{(1)} + (1 - \lambda) x^{(2)}\right] \\
& \quad + \frac{q}{m} \left[\lambda A_{t-1}^{(1*)} + (1 - \lambda) A_{t-1}^{(2*)}\right] \left[m - \lambda A_{t-1}^{(1*)} - (1 - \lambda) A_{t-1}^{(2*)}\right] - \left[\lambda A_t^{(1*)} + (1 - \lambda) A_t^{(2*)}\right] \\
=& \lambda \left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + qA_{t-1}^{(1*)} - A_t^{(1*)}\right] + (1 - \lambda) \left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + qA_{t-1}^{(2*)} - A_t^{(2*)}\right] \\
& \qquad - \frac{q}{m} \left[\lambda A_{t-1}^{(1*)} + (1 - \lambda) A_{t-1}^{(2*)}\right]^2 \\
\geq& \lambda \left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + qA_{t-1}^{(1*)} - A_t^{(1*)}\right] + (1 - \lambda) \left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + qA_{t-1}^{(2*)} - A_t^{(2*)}\right] \\
& \qquad - \frac{q}{m} \left[\lambda {A_{t-1}^{(1*)}}^2 + (1 - \lambda) {A_{t-1}^{(2*)}}^2\right] (\lambda + 1 - \lambda) \\
=& \lambda \left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + \frac{q}{m} A_{t-1}^{(1*)} \left(m - A_{t-1}^{(1*)}\right) - A_t^{(1*)}\right] \\
& \qquad + (1 - \lambda) \left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + \frac{q}{m} A_{t-1}^{(2*)} \left(m - A_{t-1}^{(2*)}\right) - A_t^{(2*)}\right] \\
\geq& 0\,,
\end{aligned}
$$

where the first inequality follows from Cauchy–Schwarz inequality, the second inequality follows since $\boldsymbol{A}^{(1*)}$ and $\boldsymbol{A}^{(2*)}$ satisfy the constraint (2.5) with regard to $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$.

For constraint (2.4c), we have

$$1 - \frac{A_{t-1}^{(\lambda)}}{m} - x_t^{(\lambda)} = \lambda \left[1 - \frac{A_{t-1}^{(1*)}}{m} - x_t^{(1)}\right] + (1-\lambda)\left[1 - \frac{A_{t-1}^{(2*)}}{m} - x_t^{(2)}\right] \geq 0,$$

where the inequality follows since $\boldsymbol{A}^{(1*)}$ and $\boldsymbol{A}^{(2*)}$ satisfy the constraint (2.4c) with regard to $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$.

Next, by the optimality of $\boldsymbol{A}^{(\lambda*)}$, we have $f(\boldsymbol{x}^{(\lambda)}) = A_L^{(\lambda*)} \geq A_L^{(\lambda)} = \lambda A_L^{(1*)} + (1-\lambda) A_L^{(2*)} = \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda) f(\boldsymbol{x}^{(2)})$.

In conclusion, $f_v(\boldsymbol{x}_{v,:})$ is a concave function on the range $[0, 1 - A_{v,0}/m]^L$. $\square$

We then outline the optimality condition of the single-variable reformulation in Lemma A.1.2. To facilitate the characterization of subgradient, we introduce the convex function $\tilde{f}_v := -f_v$. Let $\partial \tilde{f}_v(\boldsymbol{x}_v)$ be the subgradient set at $\boldsymbol{x}_v$:

Lemma A.1.2 (Optimality Condition) *Given $\theta \geq 0$, the optimal solution $\boldsymbol{x}^*(\theta)$ to the inner maximization problem* (2.9) *satisfies the following condition.*

$$\forall v \in U, t = 1, 2, \ldots, L, \boldsymbol{g}_{v,:}(\theta) \in \partial \tilde{f}_v(\boldsymbol{x}_v^*(\theta)), \begin{cases} g_{v,t}(\theta) \geq -\theta & \text{when } x_{v,t}^*(\theta) = 0, \\ g_{v,t}(\theta) = -\theta & \text{when } 0 < x_{v,t}^*(\theta) < A_{v,0}, \\ g_{v,t}(\theta) \leq -\theta & \text{when } x_{v,t}^*(\theta) = A_{v,0}. \end{cases} \tag{A.9}$$

Proof of Lemma A.1.2: Define $r_v(\boldsymbol{x}_{v,:}; \theta) := -\tilde{f}_v(\boldsymbol{x}_{v,:}) - \theta m \sum_{t=1}^L x_{v,t}$ for all $v \in U$. The problem is separable by content piece v, so we focus on a specific $v \in U$ in the following and omit the subscription for clarity.

When $x_t^*(\theta) = 0$, we construct a feasible solution $\boldsymbol{x}'(\theta) := \boldsymbol{x}^*(\theta) + \epsilon \boldsymbol{e}_t$ where ϵ is a sufficiently small positive constant and $\boldsymbol{e}_t$ is a vector with 1 in the t-th entry and 0 in all other entries. By the concavity of r, we have

$$r(\boldsymbol{x}'(\theta); \theta) \geq r(\boldsymbol{x}^*(\theta); \theta) - (\boldsymbol{g}(\theta) + \theta \mathbf{1})^\top (\boldsymbol{x}'(\theta) - \boldsymbol{x}^*(\theta)) = r(\boldsymbol{x}^*(\theta); \theta) - (g_t(\theta) + \theta)\epsilon\,,$$

where $\mathbf{1}$ is the all one vector. Given the optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) \geq -\theta$ should hold.

When $x_t^*(\theta) = A_{v,0}$, we construct a feasible solution $\boldsymbol{x}''(\theta) := \boldsymbol{x}^*(\theta) - \epsilon \boldsymbol{e}_t$. By concavity of r, we have

$$r(\boldsymbol{x}''(\theta); \theta) \geq r(\boldsymbol{x}^*(\theta); \theta) - (\boldsymbol{g}(\theta) + \theta \mathbf{1})^\top (\boldsymbol{x}''(\theta) - \boldsymbol{x}^*(\theta)) = r(\boldsymbol{x}^*(\theta); \theta) + (g_t(\theta) + \theta)\epsilon\,.$$

Given the optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) \leq -\theta$ should hold.

When $0 < x_t^*(\theta) < A_{v,0}$, we simultaneously construct two feasible solutions $\boldsymbol{x}'(\theta)$ and $\boldsymbol{x}''(\theta)$ as previous. Similarly, by optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) = -\theta$ should hold.

In conclusion, we can characterize this optimality condition based on the subgradient $\boldsymbol{g}(\theta)$ and θ. $\square$

For any $\theta \geq 0$, we denote $\left(\boldsymbol{x}^*_{v,:}(\theta) : v \in \mathcal{V}\right)$ as the optimal solution to (2.9), and let function $s(\theta; U) := m \sum_{v\in U} \sum_{t=1}^{L} x^*_{v,t}(\theta)$ describe the total optimal promotion times with dual variable θ. In the following, we establish the property of $s(\theta; U)$ for any given candidate set $U \subseteq \mathcal{V}$ in Lemma A.1.3.

Lemma A.1.3 (Monotonicity) *For any $U \subseteq \mathcal{V}$, $s(\theta; U)$ is a nonincreasing function in θ.*

Proof of Lemma A.1.3: For all $\theta_1, \theta_2 \geq 0$, let

$$\boldsymbol{g}_v(\theta_1) \in \partial \tilde{f}_v(\boldsymbol{x}^*_{v,:}(\theta_1)) \text{ and } \boldsymbol{g}_v(\theta_2) \in \partial \tilde{f}_v(\boldsymbol{x}^*_{v,:}(\theta_2)),$$

for all $v \in U$. By Lemma A.1.2, for all $v \in U$, we have

$$\begin{aligned}
&\left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \left(-\boldsymbol{g}_v(\theta_1) + \boldsymbol{g}_v(\theta_2)\right) \\
&= \left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v}(\theta_2)\right)^\top (\theta_1 \cdot \mathbf{1} - \theta_2 \cdot \mathbf{1}) + \sum_{t=1}^{L} \left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(-\theta_1 - g_{v,t}(\theta_1)\right) \\
&\qquad\qquad + \sum_{t=1}^{L} \left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(\theta_2 + g_{v,t}(\theta_2)\right) \\
&= (\theta_1 - \theta_2) \cdot \left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \mathbf{1} \\
&\quad + \sum_{t=1}^{L} \left(\mathbb{1}\{x^*_{v,t}(\theta_1) = 0\} + \mathbb{1}\{x^*_{v,t}(\theta_1) = A_{v,0}\}\right)\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(-\theta_1 - g_{v,t}(\theta_1)\right) \\
&\quad + \sum_{t=1}^{L} \left(\mathbb{1}\{x^*_{v,t}(\theta_2) = 0\} + \mathbb{1}\{x^*_{v,t}(\theta_2) = A_{v,0}\}\right)\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(\theta_2 + g_{v,t}(\theta_2)\right).
\end{aligned}$$

We further discuss the latter two terms. We have

$$\begin{cases}
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = -x^*_{v,t}(\theta_2) \leq 0, \text{ and } -\theta_1 - g_{v,t}(\theta_1) \leq 0, & \text{when } x^*_{v,t}(\theta_1) = 0, \\
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = A_{v,0} - x^*_{v,t}(\theta_2) \geq 0, \text{ and } -\theta_1 - g_{v,t}(\theta_1) \geq 0, & \text{when } x^*_{v,t}(\theta_1) = A_{v,0};
\end{cases}$$

and

$$\begin{cases}
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = x^*_{v,t}(\theta_1) \geq 0, \text{ and } \theta_2 + g_{v,t}(\theta_2) \geq 0, & \text{when } x^*_{v,t}(\theta_2) = 0, \\
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = x^*_{v,t}(\theta_2) - A_{v,0} \leq 0, \text{ and } \theta_2 + g_{v,t}(\theta_2) \leq 0, & \text{when } x^*_{v,t}(\theta_2) = A_{v,0}.
\end{cases}$$

Given concavity of f_v, we have $0 \geq \left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \left(-\boldsymbol{g}_v(\theta_1) + \boldsymbol{g}_v(\theta_2)\right) \geq (\theta_1 - \theta_2) \cdot \left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \mathbf{1}$.

By summing up over $v \in U$, we have $(\theta_1 - \theta_2) \cdot m \sum_{v\in U} \sum_{t=1}^{L} \left[x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right] = (\theta_1 - \theta_2) \cdot (h(\theta_1) - h(\theta_2)) \leq 0$.

In conclusion, $s(\theta; U)$ is nonincreasing. $\square$

Proof of Lemma 2.4.2: We begin by demonstrating that $s(\theta^*(U_2); U_1) \leq s(\theta^*(U_1); U_1)$. We can decompose the function $s(\theta; U_2)$ as $s(\theta; U_1) + s(\theta; U_2 \setminus U_1)$. We consider two cases based on the value of $s(\theta^*(U_1); U_1)$:

(i) $s(\theta^*(U_1); U_1) = C$. We directly have

$$s^*(\theta^*(U_2); U_1) \leq s^*(\theta^*(U_2); U_2) \leq C = s(\theta^*(U_1); U_1).$$

(ii) $s(\theta^*(U_1); U_1) < C$. We show this by contradiction. Assume that $s(\theta^*(U_2); U_1) > s(\theta^*(U_1); U_1)$. We can construct a feasible solution $\boldsymbol{x}'_{U_2,:}$ for the PO problem given candidate set U_2 as

$$\boldsymbol{x}'_{v,:} = \begin{cases} \boldsymbol{x}^*_{v,:}(\theta^*(U_1)) & \text{when } v \in U_1, \\ \boldsymbol{x}^*_{v,:}(\theta^*(U_2)) & \text{when } v \in U_2 \setminus U_1. \end{cases}$$

The objective value generated by $\boldsymbol{x}'_{U_2,:}$ is larger than $\boldsymbol{x}^*_{U_2,:}(\theta^*(U_2))$, given the optimality of $\boldsymbol{x}^*_{U_1,:}(\theta^*(U_1))$ for the PO problem with candidate set U_1. This contradicts with the optimality of $\boldsymbol{x}^*_{U_2,:}(\theta^*(U_2))$ for the PO problem with candidate set U_2.

Consequently, $s(\theta^*(U_2); U_1) \leq s(\theta^*(U_1); U_1)$. By Lemma A.1.3, we conclude that $\theta^*(U_1) \leq \theta^*(U_2)$. □

A.1.1.4 Submodularity of the CGPO Objective

Proof of Theorem 2.4.3: It is easy to verify that $R(U; C) + R(\mathcal{V} \setminus U; 0)$ is a monotone function. By (2.11), we have

$$\begin{aligned} &R(U \cup \{w\}; C) + R(\mathcal{V} \setminus (U + \{w\}); 0) \\ =&\max_{0 \leq c \leq C} R(U; c) + R(\{w\}; C - c) + R(\mathcal{V} \setminus (U + \{w\}); 0) \\ \geq&R(U; C) + R(\{w\}; 0) + R(\mathcal{V} \setminus (U + \{w\}); 0) = R(U; C) + R(\mathcal{V} \setminus U; 0). \end{aligned}$$

Next, we focus on the proof of submodularity. To prove that $R(U; C) + R(\mathcal{V} \setminus U; 0)$ is a submodular function, we need to demonstrate that for any given $U_1 \subseteq U_2 \subseteq \mathcal{V}$ and $w \in \mathcal{V} \setminus U_2$, (2.10) holds. Therefore, we compare the marginal gain of content piece w when given nested content sets U_1 and U_2 as follows:

$$\begin{aligned} &R(U_1 \cup \{w\}; C) - R(U_1; C) - R(\{w\}; 0) - [R(U_2 + \{w\}; C) - R(U_2; C) - R(\{w\}; 0)] \\ =&R(U_1; c^*(U_1)) + R(\{w\}; C - c^*(U_1)) - R(U_1; C) - R(\{w\}; 0) \\ &\qquad - [R(U_2; c^*(U_2)) + R(\{w\}; C - c^*(U_2)) - R(U_2; C) - R(\{w\}; 0)] \\ \geq&R(U_1; c^*(U_2)) + R(\{w\}; C - c^*(U_2)) - R(U_1; C) - R(\{w\}; 0) \\ &\qquad - [R(U_2; c^*(U_2)) + R(\{w\}; C - c^*(U_2)) - R(U_2; C) - R(\{w\}; 0)] \\ =&[R(U_1; c^*(U_2)) - R(U_1; C)] - [R(U_2; c^*(U_2)) - R(U_2; C)] \end{aligned}$$

$$= -\int_{z=c^*(U_2)}^{C} \theta^*(U_1; z)dz + \int_{z=c^*(U_2)}^{C} \theta^*(U_2; z)dz$$
$$= \int_{z=c^*(U_2)}^{C} [\theta^*(U_2; z) - \theta^*(U_1; z)]\, dz \geq 0.$$

where the first inequality follows since $c^*(U_1)$ is the optimal solution that maximizes problem (2.11) and the last inequality follows by Lemma 2.4.2.

In conclusion, $R(U; C) + R(\mathcal{V} \setminus U; 0)$ is a monotone submodular set function. □

A.1.2 Proofs and Supplements for Section 2.5

A.1.2.1 The OLS Estimation Method for the BDM

According to Bass (1969), the OLS method of the BDM works as follows. Given a sequence of adoption data $\{(a_t, A_t)\}_{t=0}^{T}$, it assumes the following relationship: $a_t = \beta_1 + \beta_2 \cdot A_{t-1} + \beta_3 A_{t-1}^2 + \epsilon_t$, where $\beta_1 = pm$, $\beta_2 = q - p$ and $\beta_3 = -q/m$ are three different parameters to estimate; 1, A_{t-1}, and A_{t-1}^2 are considered as three observed covariates; ϵ_t is the independent white noise. For notation simplicity, we denote the covariate matrix and dependent variable as

$$Z = \begin{pmatrix} 1 & A_0 & A_0^2 \\ 1 & A_1 & A_1^2 \\ \vdots & \vdots & \vdots \\ 1 & A_{T-1} & A_{T-1}^2 \end{pmatrix} \text{ and } \boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix}.$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ can then be derived as $\hat{\boldsymbol{\beta}} = (Z^\top Z)^{-1} Z^\top \boldsymbol{a}$. Consequently, the estimators can be obtained as

$$\hat{m} = \frac{\hat{\beta}_2 \pm \sqrt{\hat{\beta}_2^2 - 4\hat{\beta}_1\hat{\beta}_2}}{2\hat{\beta}_3},\ \hat{p} = \frac{\hat{\beta}_1}{\hat{m}},\ \text{and } \hat{q} = -\hat{\beta}_3\hat{m}.$$

However, these estimators suffer from large variances. In extreme cases (e.g., $\hat{m} = 0$), they even become invalid.

A.1.2.2 Proofs for the Asymptotic Analysis of the OLS-Based Estimators

To streamline notation, we define the fixed-design covariate matrix for the n-th diffusion process as

$$Z_{(n)} = \begin{pmatrix} x_{1,(n)} & \bar{A}_{1,(n)}(1 - \bar{A}_{1,(n)}) \\ x_{2,(n)} & \bar{A}_{2,(n)}(1 - \bar{A}_{2,(n)}) \\ \vdots & \vdots \\ x_{n,(n)} & \bar{A}_{n,(n)}(1 - \bar{A}_{n,(n)}) \end{pmatrix}.$$

Proof of Theorem 2.5.1: We first consider the n-th D-OLS estimator for q, which is represented as

$$\hat{q}_{(n)}^{\text{D-OLS}} = q + \frac{\sum_{i=1}^{n}\left[A_{i,(n)}(1 - x_{i,(n)} - \frac{A_{i,(n)}}{m_{(n)}})\epsilon_{i,(n)}^{\text{i}}\right]}{\sum_{i=1}^{n}\left[A_{i,(n)}(1 - x_{i,(n)} - \frac{A_{i,(n)}}{m_{(n)}})\right]^2} = q + \frac{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\text{i}}}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2}\,.$$

Let $\bar{\boldsymbol{\epsilon}}^{\text{i}}$ denote the noise distribution with finite variance $\text{var}\left(\bar{\epsilon}^{\text{i}}\right) < \infty$, we can characterize the mean and variance of D-OLS estimator as follows:

$$\mathbb{E}\left[\hat{q}_{(n)}^{\text{D-OLS}}\right] = q + \frac{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)}) \cdot \mathbb{E}\left[\bar{\epsilon}_{i,(n)}^{\text{i}}\right]}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2} = q, \text{ and}$$

$$\text{var}\left(\hat{q}_{(n)}^{\text{D-OLS}}\right) = \frac{\frac{1}{n^2}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2\,\text{var}\left(\bar{\epsilon}_{i,(n)}^{\text{i}}\right)}{\left[\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2\right]^2} = \frac{\tilde{Q}_{22,(n)}}{nQ_{22,(n)}^2}\,\text{var}\left(\bar{\epsilon}^{\text{i}}\right)\,.$$

By Chebyshev's inequality, we have $\Pr\left(|\hat{q}_{(n)}^{\text{D-OLS}} - q| \geq k\right) \leq \frac{\tilde{Q}_{22,(n)}}{k^2 nQ_{22,(n)}^2}\,\text{var}\left(\bar{\epsilon}^{\text{i}}\right)$. Taking limits on both sides, we get

$$\begin{aligned}\lim_{n\to\infty}\Pr\left(|\hat{q}_{(n)}^{\text{D-OLS}} - q| \geq k\right) &\leq \lim_{N\to\infty}\frac{1}{k^2}\frac{\tilde{Q}_{22,(n)}\,\text{var}\left(\bar{\epsilon}^{\text{i}}\right)}{nQ_{22,(n)}^2} \\ &= \frac{\text{var}\left(\bar{\epsilon}^{\text{i}}\right)}{k^2}\cdot\frac{\lim_{n\to\infty}\tilde{Q}_{22,(n)}}{\lim_{n\to\infty}Q_{22,(n)}^2}\cdot\lim_{n\to\infty}\frac{1}{n} = \frac{\tilde{Q}_{22}\,\text{var}\left(\bar{\epsilon}^{\text{i}}\right)}{k^2Q_{22}^2}\cdot\lim_{n\to\infty}\frac{1}{n} = 0\,,\end{aligned}$$

which implies that $\lim_{n\to\infty}\hat{q}_{(n)}^{\text{D-OLS}} = q$. Similarly, we consider the n-th D-OLS estimator for p, which is

$$\begin{aligned}\hat{p}_{(n)}^{\text{D-OLS}} &= p + \frac{\sum_{i=1}^{n}\left[m_{(n)}x_{i,(n)}\left((q - \hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}A_{i,(n)} + \bar{\epsilon}_{i,(n)}^{\text{d}}\right)\right]}{\sum_{t=1}^{n}(m_{(n)}x_{i,(n)})^2} \\ &= p + \frac{\frac{1}{n}\sum_{i=1}^{n}\left[x_{i,(n)}\left((q - \hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}\bar{A}_{i,(n)} + \bar{\epsilon}_{i,(n)}^{\text{d}}\right)\right]}{\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}^2}.\end{aligned}$$

We consider the following two terms separately:

$$\lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}^2\bar{A}_{i,(n)}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right) \quad\text{and}\quad \lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right).$$

The first term converges to 0 since

$$\lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(q - \hat{q}^{\text{D-OLS}})x_{i,(n)}^2\bar{A}_{i,(n)}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right) = \lim_{n\to\infty}\left(q - \hat{q}_{(n)}^{\text{D-OLS}}\right)\cdot\lim_{n\to\infty}\left(\frac{\tilde{Q}_{11,(n)}}{Q_{11,(n)}}\right)$$

$$= \frac{\tilde{Q}_{11}}{Q_{11}} \lim_{n\to\infty} \left(q - \hat{q}_{(n)}^{\text{D-OLS}}\right) = 0\,.$$

The second term also converges to 0, similarly as in the proof of $\hat{q}_{(n)}^{\text{D-OLS}}$. As a result, we have $\lim_{n\to\infty} \hat{p}_{(n)}^{\text{D-OLS}} = p$.

In conclusion, D-OLS estimators $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$ are consistent estimators. □

Proof of Theorem 2.5.2: We consider two different estimation methods, respectively.

(i) The D-OLS method.

First, we characterize the limiting distribution of $\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)$. We can represent it as

$$\begin{aligned}\sqrt{n}\left(\hat{q}_{(n)}^{\text{D-OLS}} - q\right) &= \frac{\sqrt{n}\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\text{i}}\right]}{\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)})\right]^2} \\ &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\text{i}}\right]}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2}\,.\end{aligned}$$

We consider it as the sum of n independent random variables:

$$\sqrt{n}\left(\hat{q}_{(n)}^{\text{D-OLS}} - q\right) = \sum_{i=1}^{n} w_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{i}}, \text{ where } w_{i,(n)} = \frac{\frac{1}{\sqrt{n}}\left[\bar{A}_{i,(n)}(1 - x_{i,(n)} - \bar{A}_{i,(n)})\right]}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2}\,.$$

We then show that this sequence satisfies

$$\begin{aligned}\lim_{n\to\infty}\max_{i=1,2,\ldots,n}|w_{i,(n)}| &\le \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{1}{\tilde{Q}_{22,(n)}} \\ &= \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\lim_{n\to\infty}\frac{1}{\tilde{Q}_{22,(n)}} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{1}{\tilde{Q}_{22}} = 0\,.\end{aligned}$$

where the inequality follows since $0 \le x_{i,(n)}, \bar{A}_{i,(n)} \le 1$ for all $i = 1, 2, \ldots, n$.

This implies that Lindeberg's condition is satisfied. At last, the variance of $\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)$ is as follows:

$$\text{var}\left(\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2}{\left[\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1 - x_{i,(n)} - \bar{A}_{i,(n)})^2\right]^2}\,\text{var}\left(\bar{\epsilon}_{i,(n)}^{\text{i}}\right) = \frac{1}{\tilde{Q}_{22,(n)}}\eta\sigma^2.$$

Let $\xi_2 = \eta Q_{22}/\tilde{Q}_{22} - 1$. By Lindeberg's central limit theorem, we have

$$\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{22,(n)}}(1+\xi_2)\sigma^2\right).$$

Next, we characterize the limiting distribution of $\sqrt{n}(\hat{p}_{(n)}^{\text{D-OLS}} - p)$. We can represent it as

$$\begin{aligned}\sqrt{n}(\hat{p}_{(n)}^{\text{D-OLS}} - p) &= \frac{\sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)} + \sqrt{n}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\sum_{i=1}^n x_{i,(n)}^2} \\ &= \sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\frac{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} + \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2}.\end{aligned}$$

We consider the two terms separately. For the former term, we can easily derive that

$$\sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\frac{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{\eta\tilde{Q}_{11}^2}{\tilde{Q}_{22}Q_{11}^2}\sigma^2\right).$$

For the latter term, we perform a similar analysis as the previous one and get

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{1-\eta}{Q_{11}}\sigma^2\right).$$

Let $\xi_1 = \eta(\tilde{Q}_{11}^2/\tilde{Q}_{22}Q_{11} - 1)$. As these two terms are independent, therefore, we can conclude that

$$\sqrt{n}(\hat{p}^{\text{D-OLS}} - p) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{11}}(1+\xi_1)\sigma^2\right).$$

(ii) The OLS method. For notation simplicity, we write the OLS formulation in matrix form. Let $\boldsymbol{\beta} = (p, q)^\top$ and $\bar{\boldsymbol{\epsilon}}_{(n)} = (\bar{\epsilon}_{1,(n)}, \bar{\epsilon}_{2,(n)}, \ldots, \bar{\epsilon}_{n,(n)})^\top$. Consider the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \boldsymbol{\beta})$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \beta) = \sqrt{n}\left({Z_{(n)}}^\top Z_{(n)}\right)^{-1}{Z_{(n)}}^\top \bar{\boldsymbol{\epsilon}}_{(n)}.$$

Let $W_{(n)} = \sqrt{n}({Z_{(n)}}^\top Z_{(n)})^{-1}{Z_{(n)}}^\top$ and $\boldsymbol{w}_{i,(n)}$ be the i-th column of $W_{(n)}$ for $i = 1, 2, \ldots, n$. As a consequence, we can write $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \boldsymbol{\beta})$ as a sum of n independent random variables: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \beta) = \sum_{i=1}^n \boldsymbol{w}_{i,(n)}\bar{\epsilon}_{i,(n)}$. We then show that this sequence satisfies

$$\lim_{n\to\infty}\max_{i=1,2,\ldots,n}\|\boldsymbol{w}_{i,(n)}\|_2 = \lim_{n\to\infty}\|W_{(n)}\|_{\infty,2} = \lim_{n\to\infty}\left\|\left(\frac{1}{n}{Z_{(n)}}^\top Z_{(n)}\right)^{-1}\frac{1}{\sqrt{n}}{Z_{(n)}}^\top\right\|_{\infty,2}$$

$$= \lim_{n\to\infty} \left\| {Q_{(n)}}^{-1} \frac{1}{\sqrt{n}} {Z_{(n)}}^{\top} \right\|_{\infty,2} \leq \lim_{n\to\infty} \frac{1}{\sqrt{n}} \left\| {Q_{(n)}}^{-1} \right\|_{2,\infty} \left\| {Z_{(n)}}^{\top} \right\|_{2,2}$$
$$\leq \lim_{n\to\infty} \frac{1}{\sqrt{n}} \left\| {Q_{(n)}}^{-1} \right\|_{2,\infty} = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \cdot \lim_{n\to\infty} \left\| {Q_{(n)}}^{-1} \right\|_{2,\infty} = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \cdot \left\| Q^{-1} \right\|_{2,\infty} = 0,$$

where the first inequality follows since the definition of matrix operator norm and the second inequality follows since $0 \leq x_{i,(n)}, \bar{A}_{i,(n)} \leq 1$ for all $i = 1, 2, \ldots, n$.

This implies that Lindeberg's condition is satisfied. Then, we calculate the variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}^{\mathrm{OLS}} - \beta)$ as follows:

$$\begin{aligned} \operatorname{var}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\mathrm{OLS}} - \beta)\right) &= \mathbb{E}\left[W_{(n)} \bar{\boldsymbol{\epsilon}}_{(n)} {\bar{\boldsymbol{\epsilon}}_{(n)}}^{\top} {W_{(n)}}^{\top}\right] = \sigma^2 W_{(n)} {W_{(n)}}^{\top} \\ &= n \left({Z_{(n)}}^{\top} Z_{(n)}\right)^{-1} = {Q_{(n)}}^{-1}. \end{aligned}$$

By Lindeberg's central limit theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\mathrm{OLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(0, Q^{-1}\right),$$

where $Q^{-1} = \begin{bmatrix} \frac{1}{Q_{11}}\left(1 + \frac{{Q_{12}}^2}{Q_{11}Q_{22} - Q_{12}^2}\right) & -\frac{Q_{12}}{Q_{11}Q_{22} - Q_{12}^2} \\ -\frac{Q_{12}}{Q_{11}Q_{22} - Q_{12}^2} & \frac{1}{Q_{22}}\left(1 + \frac{Q_{12}^2}{Q_{11}Q_{22} - Q_{12}^2}\right) \end{bmatrix}$.

At last, we conclude that (i) The asymptotic variances of $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$ are $(1 + \xi_1)\sigma^2/Q_{11}$ and $(1 + \xi_2)\sigma^2/Q_{22}$, where $\xi_1 = \eta(\tilde{Q}_{11}^2/\tilde{Q}_{22}Q_{11} - 1)$ and $\xi_2 = \eta Q_{22}/\tilde{Q}_{22} - 1$. (ii) The asymptotic variances of $\hat{p}^{\mathrm{OLS}}$ and $\hat{q}^{\mathrm{OLS}}$ are $(1 + \kappa)\sigma^2/Q_{11}$ and $(1 + \kappa)\sigma^2/Q_{22}$, where $\kappa = Q_{12}^2/|Q|$. □

Proposition A.1.1 *When $\eta \leq \tilde{Q}_{22}/Q_{22}$, we can show that the asymptotic variances of D-OLS estimators are smaller than those of OLS estimators, that is, $\xi_1 \leq \kappa$ and $\xi_2 \leq \kappa$.*

Proof for Proposition A.1.1: Particularly, we have

$$\begin{aligned} \kappa - \xi_1 &= \frac{Q_{12}^2}{Q_{11}Q_{22} - Q_{12}^2} - \eta\left(\frac{\tilde{Q}_{11}^2}{Q_{11}\tilde{Q}_{22}} - 1\right) \geq \frac{Q_{12}^2}{Q_{11}Q_{22} - Q_{12}^2} - \frac{\tilde{Q}_{11}^2}{Q_{11}Q_{22}} \\ &\geq \frac{Q_{12}^2 - \tilde{Q}_{11}^2}{Q_{11}Q_{22}} = \frac{Q_{12} + \tilde{Q}_{11}}{Q_{11}Q_{22}}\left(Q_{12} - \tilde{Q}_{11}\right) \\ &= \frac{Q_{12} + \tilde{Q}_{11}}{Q_{11}Q_{22}} \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i,(n)} A_{i,(n)}(1 - A_{i,(n)}) - \frac{1}{n}\sum_{i=1}^{n} x_{i,(n)}^2 A_{i,(n)}\right) \\ &\geq \frac{Q_{12} + \tilde{Q}_{11}}{Q_{11}Q_{22}} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left(x_{i,(n)} A_{i,(n)} \cdot x_{i,(n)} - x_{i,(n)}^2 A_{i,(n)}\right) = 0. \end{aligned}$$

where the first inequality follows since $0 \leq \eta \leq \tilde{Q}_{22}/Q_{22}$, the second inequality follows since $Q_{12}^2 \geq 0$ and the third inequality follows since $x_{i,(n)} \leq 1 - \bar{A}_{i,(n)}$. Furthermore, we have

$$\kappa - \xi_2 = \frac{Q_{12}^2}{|Q|} - \left(\eta \frac{Q_{22}}{\tilde{Q}_{22}} - 1\right) \;\geq\; \kappa \;\geq\; 0.$$

where the first inequality follows since $\eta \leq \tilde{Q}_{22}/Q_{22}$ and the second inequality follows by definition. □

A.1.2.3 Proofs for the MLE Estimators

Proposition A.1.2 *When platforms cannot observe adopter types, the log-likelihood function $\mathcal{LL}^{MLE}(p,q)$ is concave.*

Proof of Proposition A.1.2: To show that $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave, it is sufficient to show that the corresponding Hessian matrix is negative semi-definite.

Let $g_t = px_t/(1-A_{t-1}/m)+qA_{t-1}/m$. The partial derivatives of $\mathcal{LL}^{\text{MLE}}(p,q)$ with regard to p and q are

$$\frac{\partial \mathcal{LL}^{\text{MLE}}(p,q)}{\partial p} = \sum_{t=1}^{T} \frac{\partial g_t}{\partial p}\left(\frac{a_t}{g_t} - \frac{m - A_{t-1} - a_t}{1-g_t}\right),$$

$$\text{and} \quad \frac{\partial \mathcal{LL}^{\text{MLE}}(p,q)}{\partial q} = \sum_{t=1}^{T} \frac{\partial g_t}{\partial q}\left(\frac{a_t}{g_t} - \frac{m - A_{t-1} - a_t}{1-g_t}\right).$$

Let $\boldsymbol{H}_{\mathcal{LL}}$ and $\boldsymbol{H}_{g_t}$ be the Hessian matrices of $\mathcal{LL}^{\text{MLE}}$ and g_t, respectively. We have

$$\begin{aligned}&\boldsymbol{H}_{\mathcal{LL}}\\ &= \sum_{t=1}^{T}\left[-\left(\frac{a_t}{g_t^2} + \frac{m - A_{t-1} - a_t}{(1-g_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial g_t}{\partial p}\\ \frac{\partial g_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial g_t}{\partial p} & \frac{\partial g_t}{\partial q}\end{pmatrix} + \left(\frac{a_t}{g_t} - \frac{m - A_{t-1} - a_t}{1-g_t}\right)\cdot \boldsymbol{H}_{g_t}\right].\end{aligned}$$

Since $\boldsymbol{H}_{g_t}$ is a zero matrix and $a_t/g_t^2 + (m - A_{t-1} - a_t)/(1-g_t)^2 \geq 0$ always holds, we can conclude that $\boldsymbol{H}_{\mathcal{LL}}$ is negative semi-definite which implies $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave.

In conclusion, the log-likelihood function $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave. □

Proposition A.1.3 *When platforms can observe adopter types, the log-likelihood function $\mathcal{LL}^{D\text{-}MLE}(p,q)$ is concave.*

Proof of Proposition A.1.3: To show that $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave, it is sufficient to show that the corresponding Hessian matrix is negative semi-definite.

Let $g_t = qA_{t-1}/m$ and $h_t = p + qA_{t-1}/m$. The partial derivatives of $\mathcal{LL}^{\text{D-MLE}}(p,q)$ with regard to p and q are

$$\frac{\partial \mathcal{LL}^{\text{D-MLE}}(p,q)}{\partial p} = \sum_{t=1}^{T}\left[\frac{\partial g_t}{\partial p}\left(\frac{a_t^{\text{i}}}{g_t} - \frac{m - A_{t-1} - mx_t - a_t^{\text{i}}}{1-g_t}\right) + \frac{\partial h_t}{\partial p}\left(\frac{a_t^{\text{d}}}{h_t} - \frac{mx_t - a_t^{\text{d}}}{1-h_t}\right)\right],$$

$$\text{and}\quad \frac{\partial \mathcal{LL}^{\text{D-MLE}}(p,q)}{\partial q} = \sum_{t=1}^{T}\left[\frac{\partial g_t}{\partial q}\left(\frac{a_t^{\text{i}}}{g_t} - \frac{m - A_{t-1} - mx_t - a_t^{\text{i}}}{1-g_t}\right) + \frac{\partial h_t}{\partial q}\left(\frac{a_t^{\text{d}}}{h_t} - \frac{mx_t - a_t^{\text{d}}}{1-h_t}\right)\right].$$

Let $\boldsymbol{H}_{\mathcal{LL}}$, $\boldsymbol{H}_{g_t}$ and $\boldsymbol{H}_{h_t}$ be the Hessian matrices of $\mathcal{LL}^{\text{D-MLE}}$, g_t and h_t, respectively. We have

$$\begin{aligned}\boldsymbol{H}_{\mathcal{LL}} = \sum_{t=1}^{T}\Bigg[&-\left(\frac{a_t^{\text{i}}}{g_t^2} + \frac{m - A_{t-1} - mx_t - a_t^{\text{i}}}{(1-g_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial g_t}{\partial p}\\ \frac{\partial g_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial g_t}{\partial p} & \frac{\partial g_t}{\partial q}\end{pmatrix}\\ &+\left(\frac{a_t^{\text{i}}}{g_t} - \frac{m - A_{t-1} - mx_t - a_t^{\text{i}}}{1-g_t}\right)\cdot\boldsymbol{H}_{g_t}\Bigg]\\ +\sum_{t=1}^{T}\Bigg[&-\left(\frac{a_t^{\text{d}}}{h_t^2} + \frac{mx_t - a_t^{\text{d}}}{(1-h_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial h_t}{\partial p}\\ \frac{\partial h_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial h_t}{\partial p} & \frac{\partial h_t}{\partial q}\end{pmatrix} + \left(\frac{a_t^{\text{d}}}{g_t} - \frac{mx_t - a_t^{\text{d}}}{1-g_t}\right)\cdot\boldsymbol{H}_{h_t}\Bigg].\end{aligned}$$

Since $\boldsymbol{H}_{g_t}$ and $\boldsymbol{H}_{h_t}$ are zero matrices and $a_t^{\text{i}}/g_t^2 + (m - A_{t-1} - mx_t - a_t^{\text{i}})/(1-g_t)^2 \geq 0$ and $a_t^{\text{d}}/h_t^2 + (mx_t - a_t^{\text{d}})/(1-h_t)^2 \geq 0$ always hold, we can conclude that $\boldsymbol{H}_{\mathcal{LL}}$ is negative semi-definite which implies $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave.

In conclusion, the log-likelihood function $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave. □

A.2 Supplements for Numerical Experiments

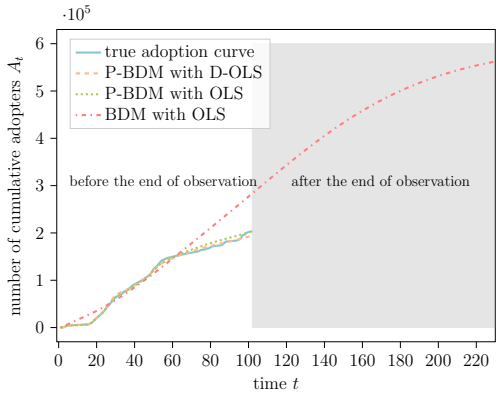
A.2.1 Model Calibration

A.2.1.1 The BDM for Online Content Adoption

This section complements our discussions on the discrepancy on the BDM and actual adoption data for online content. A common issue to fit the BDM is the notable underestimation of the diffusion coefficient q, exemplified in our case study shown in Figure 2.4a. In some cases, this coefficient is even negative, as seen in Figure 2.4b, leading to a deviation of the fitted BDM curve from its typical S-shaped configuration.

It is important to emphasize that these variations, while significant, do not contradict the theoretical foundation of the BDM. To clarify this point, we illustrate the complete trajectory of the fitted BDM curve, extended beyond the time horizon of our observation, for our motivating example in Figure A.1. In Figures 2.1a and 2.4a, we have only included a segment of the BDM curve that fits within the observed time frame. In Figure A.1, the complete fitted BDM curve gradually presents an S-shape. Despite this delayed emergence of the S-shape,

Figure A.1: Illustration of the Complete Fitted BDM Curve to the Actual Adoption for the Motivating Example



the BDM curve significantly diverges from the actual pattern of online content adoption. This discrepancy highlights the limitations of the BDM model in accurately capturing the pattern of online content adoption and underscores the necessity of adopting a modified model, such as the P-BDM, for a more precise representation of these dynamics.

A.2.1.2 Timeliness of Online Content Diffusion

In this section, we will explore the concept of timeliness in online content and how it affects the diffusion process, resulting in a time-decay factor. We will begin by presenting our findings from data analysis and then modify the P-BDM to incorporate the time-decay factor for a better fit.

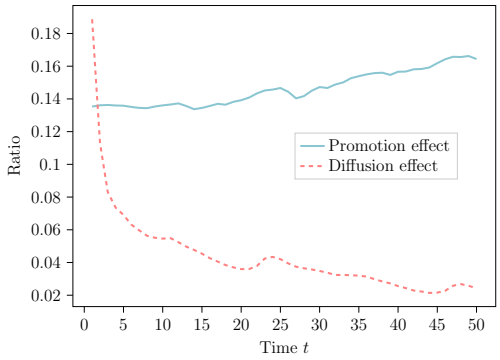
Online platforms operate in a highly dynamic and fast-paced environment, with new content being created and shared at a rapid rate. Compared to traditional markets, online platforms have a faster speed of information dissemination. As a result, the timeliness of online content plays a critical role. For example, a review video of a new movie will lose its relevance and generate fewer adoptions as the movie becomes older and less popular. Our analysis of the dataset from the video-sharing platform confirms this phenomenon. To demonstrate this, we calculate two ratios to characterize the promotion and diffusion effects:

$$\frac{a^{\mathrm{d}}_{v,t}}{m x_{v,t}} - \frac{a^{\mathrm{i}}_{v,t}}{m - A_{v,t-1}} \text{ and } \frac{a^{\mathrm{i}}_{v,t}}{\frac{A_{v,t-1}}{m}(m - A_{v,t-1})}. \tag{A.10}$$

We would like to remark that our goal with this analysis is not to calculate precise values of p and q, but rather to provide insight into the trends of both effects in the real world. As shown in Figure A.2, the average values of these two ratios among all content pieces are presented against the time from $t = 1$ to $t = 50$.

It is apparent from Figure A.2 that the diffusion effect exhibits a time-decay trend, while the promotion effect remains nearly constant throughout the entire horizon. The sensitivity

Figure A.2: Illustration of the Trends of Promotion and Diffusion Effects



Notes. The x-axis represents the time elapsed since the video was uploaded to the platform and the y-axis represents the average values of the ratios as shown in (A.10) among all videos at the same time step t.

to timeliness is primarily observed in the diffusion effect. As such, incorporating the time-decay factor in diffusion modeling is critical to accurately capture the content adoptions for online platforms.

Recall that in the P-BDM dynamic (2.2), the diffusion effect is proportional to the cumulative adopter number, given by qA_{t-1}/m. To incorporate the timeliness of online content diffusion, we introduce a time-decay multiplicative factor γ where $0 < \gamma \leq 1$. Specifically, we consider the diffusion effect to be $q\gamma^{t-1}A_{t-1}/m$ instead. Therefore, the P-BDM with a time-decay factor can be shown as follows:

$$\begin{aligned} a_t &= \underbrace{\left(p + q\frac{\gamma^{t-1}A_{t-1}}{m}\right)mx_t}_{\text{Direct adopters}} + \underbrace{q\frac{\gamma^{t-1}A_{t-1}}{m}(m - A_{t-1} - mx_t)}_{\text{Indirect adopters}} \\ &= \underbrace{pmx_t}_{\text{Promotion effect}} + \underbrace{q\frac{\gamma^{t-1}A_{t-1}}{m}(m - A_{t-1})}_{\text{Diffusion effect}} . \end{aligned} \tag{A.11}$$

This model uses the time-decay factor to characterize the decreasing incentive to diffuse the content as time elapses since its upload. When $\gamma = 1$, this model is equivalent to the original P-BDM.

We make two remarks here. First, the exponent of γ is related to the time elapsed since the content is uploaded. It should be distinguished from the subscript t in the CGPO problem, where the latter is used to denote the time since the beginning of the L planning period. Second, when γ is given, all the results in Sections 2.4 and 2.5 still hold. From the optimization perspective, it serves as a known parameter in the CGPO problem and does not change the underlying optimal structure. From the estimation perspective, it requires preprocessing of the observations, but the same estimation methods and analyses can be

applied. Therefore, the P-BDM with a time-decay factor does not add to the difficulty of the entire problem but provides flexibility in characterizing the true adoption processes.

A.2.1.3 Group Estimation

In the context of online platforms, estimating parameters for each content piece individually is usually impractical because of the mountainous amount of videos and the scarcity of data pertaining to a video. What makes things worse is that we often have to make promotion decisions at the early stage of a video's life cycle with minimal data available for estimation. Consequently, it is reasonable to group or cluster the videos using features, and then estimate the parameters to make sure that past estimates can be generalized to future videos and the results are precise. Due to the lack of contextual information, we focus on using category labels for estimation.

The group estimation procedure is similar to that of a single piece, except that the observations are expanded to include all content pieces in the group. Let $\mathcal{V}^c \subseteq \mathcal{V}$ be the set of content pieces in the group c. Therefore, the observations for a group can be represented as $\cup_{v\in\mathcal{V}^c}\{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$. The OLS-based and the MLE-based methods can be readily applied.

A.2.1.4 Calibration Process

For each video category $c \in \mathcal{C}$ within the dataset, we split the observations into training, validation and test sets, separately. To avoid data corruption, we split the data based on video granularity, using a 60-20-20 split. Specifically, for the video set $\mathcal{V}^c \subseteq \mathcal{V}$ corresponding to category c, we randomly select 60% of the videos $v \in \mathcal{V}^c$ and assign the associated observations $\{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$ to the training set. The remaining videos are also randomly split into 20% and 20% for validation and test sets, respectively.

To summarize the calibration process, we present Algorithm 2. We make a remark here, for each video $v \in \mathcal{V}$, we only include observations when the promotion fraction $x_{v,t}$ is positive in our training, validation, and test sets.

We evaluate the calibration performance using the weighted mean absolute percentage error (WMAPE). In Figure A.3, we show the WMAPE we obtain by calibrating the P-BDM with different time-decay factor γ using the D-OLS method. The minimum WMAPE is achieved when $\gamma = 0.983$.

For the sake of completeness, we also include the calibration results in Figure A.4 when the timeliness is ignored (i.e., $\gamma = 1$). We observe that the estimated diffusion coefficient q is smaller in this case to account for the time decay in diffusion. However, the average out-of-sample WMAPE is 42.92%, which is 10% larger than when $\gamma = 0.983$. The average out-of-sample WMAPEs of the P-BDM with OLS and the BDM are 43.53% and 81.25%, respectively.

Algorithm 2: Calibration process with time-decay factor and group estimation.

for $c \in \mathcal{C}$ **do**
  Randomly split the video set $\mathcal{V}^c$ into $\mathcal{V}^c_{\text{train}}$, $\mathcal{V}^c_{\text{valid}}$ and $\mathcal{V}^c_{\text{test}}$, using a 60-20-20 split.
  Training set $\mathcal{D}_{\text{train}} := \cup_{v \in \mathcal{V}^c_{\text{train}}} \{(a^{\text{d}}_{v,t}, a^{\text{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
  Valid set $\mathcal{D}_{\text{valid}} := \cup_{v \in \mathcal{V}^c_{\text{valid}}} \{(a^{\text{d}}_{v,t}, a^{\text{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
  Test set $\mathcal{D}_{\text{test}} := \cup_{v \in \mathcal{V}^c_{\text{test}}} \{(a^{\text{d}}_{v,t}, a^{\text{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
end
for $\gamma \in \{\gamma_1, \gamma_2, \ldots\}$ **do**
  for $c \in \mathcal{C}$ **do**
    Obtain $\hat{p}^c(\gamma)$ and $\hat{q}^c(\gamma)$ based on the training set $\mathcal{D}^{\text{train}}$ and time-decay factor γ.
    for $v \in \mathcal{V}^c_{valid}$ **do**
      Use $\hat{p}^c(\gamma)$ and $\hat{q}^c(\gamma)$ to predict adoptions as $\{\hat{a}_{v,t}\}_{t=1,2,\ldots T_v}$.
      $\text{WMAPE}_v(\gamma) := \sum_{t=1}^{T_v} |a_{v,t} - \hat{a}_{v,t}| / \sum_{t=1}^{T_v} a_{v,t} \times 100\%$.
    end
  end
  $\text{WMAPE}(\gamma) := \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \text{WMAPE}_v(\gamma)$.
end
$\gamma^* := \arg\max_\gamma \text{WMAPE}(\gamma)$.
for $c \in \mathcal{C}$ **do**
  for $v \in \mathcal{V}^c_{test}$ **do**
    Use $\hat{p}^c(\gamma^*)$ and $\hat{q}^c(\gamma^*)$ to predict adoptions as $\{\hat{a}_{v,t}\}_{t=1,2,\ldots T_v}$.
  end
end

A.2.2 Supplementary Analysis of the AGA Policy

In this section, we provide a supplementary analysis of the AGA policy with $L = 13$. We begin by examining the AGA policy across different lifetimes, followed by the detailed procedures of the sensitivity analysis and K-Means clustering analysis.

A.2.2.1 The AGA Policy Across Different Lifetime Stages

We begin by presenting the AGA policy across different lifetime stages in Figure A.5.

As shown in Figure A.5a, the policy primarily promotes videos in their initial stages, dedicating about 53% of the overall budget to videos that have no adopters (stage 0). This heavy initial promotion indicates the policy aims at sparking interest in new content. Figure A.5b further emphasizes this point by showing that, on average, the promotion fraction allocated to videos tends to decline as the lifetime progresses. However, an exception can be observed at stage 1. This anomaly occurs because only a subset of videos is advanced to later

Figure A.3: WMAPE of the Validation Set Against Time-decay Factor γ.

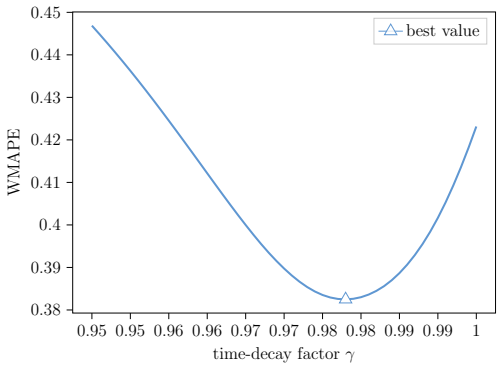
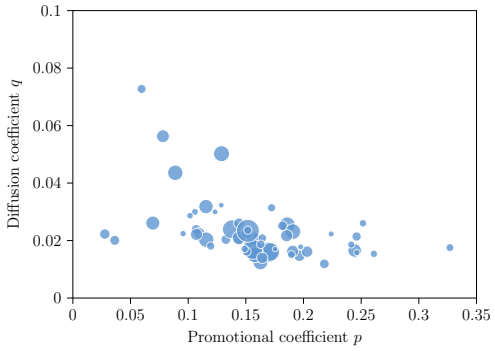


Figure A.4: Distribution of Estimated Coefficients When the Timeliness of Content Diffusion is Ignored ($\gamma = 1$).



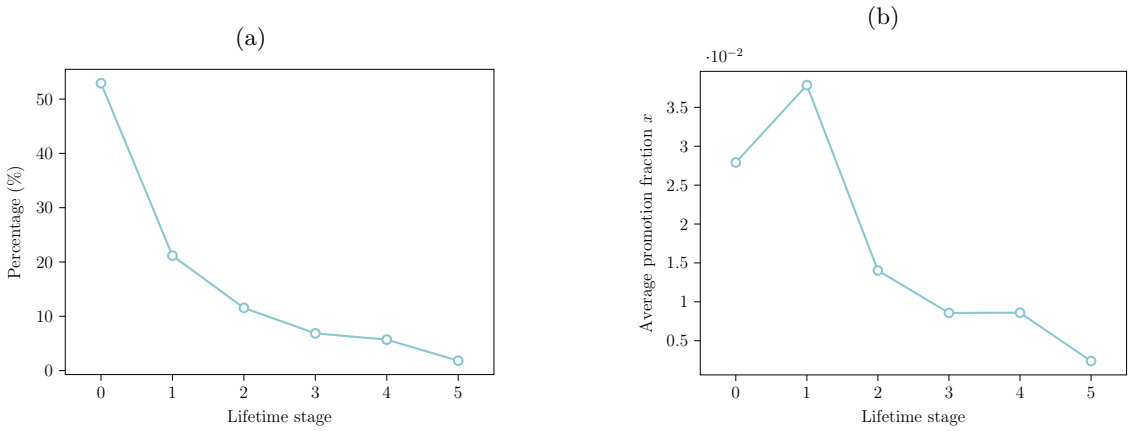
stages after the initial promotion at stage 0. It implies that the policy also acts as a filter or selection mechanism, deciding if a video shows enough promise for further promotion. As a result, the policy generally allocates the promotion budget to videos that show considerable potential in their early stages.

A.2.2.2 Sensitivity Analysis of Promotion and Diffusion Coefficients

In order to understand the relationship between the promotion policy and the characteristics of videos, we perform sensitivity analysis for some important characteristics. For this purpose, we specify the following regression model for the allocated promotion fraction $x_{c,v,t}$, where each observation is a video v at the beginning of time t in the experiment with promotion budget $\bar{C} = c$:

$$x_{c,v,t} = \beta_0 + \beta_1 p_v + \beta_2 q_v + \beta_3 \bar{A}_{v,t-1} + \beta_4 c + \epsilon_{v,t}. \tag{A.12}$$

Figure A.5: Promotion Budget Allocation of the AGA Policy for Videos at Different Lifetime Stages



Notes. (a) Percentage of allocated budget. (b) Average promotion fraction.

where the adoption number is normalized by the market size m before entering the regression. The coefficient $\boldsymbol{\beta}$ in (A.12) can therefore be used to represent the impact of each characteristic on the allocated promotion fraction.

We conduct the regression on the observations from the previous AGA experiments with $L = 13$. Furthermore, in order to illustrate the difference of policy for videos at different lifetime, we perform regression within each lifetime stage separately. Table A.1 reports the regression results. Although the R^2 is small for all the regressions, indicating that the linear regression model (A.12) is not a good representation of the complicated AGA policy, all coefficients are significant at the significance level of 0.0001. Therefore, we consider the values of regression coefficients can represent the impact of video characteristics on the promotion policy.

A.2.2.3 K-Means Clustering for the Promotion Policy

In order to further understand the AGA policy over the entire lifetime as a whole, we perform K-Means clustering on the average policy for different video configurations. Let $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$ be the set of lifetime stages. To summarize the clustering process, we present Algorithm 3.

Table A.1: Regression Results of the Promotion Fraction of the AGA Policy with Regard to Video Characteristics

	# of Obs.	R^2	β_0 (const)	β_1 (p)	β_2 (q)	β_3 ($\bar{A}$)	β_4 ($\bar{C}$)
stage 0	72,867	0.044	-0.1412****	0.8243****	0.6687****	-	0.0055****
			(0.0043)	(0.0216)	(0.0377)	-	(0.0002)
stage 1	21,490	0.418	-1.1621****	5.8253****	3.7211****	-0.7627****	0.0576****
			(0.0245)	(0.1223)	(0.1035)	(0.0195)	(0.0245)
stage 2	31,610	0.209	-0.4399****	2.3614****	1.3477****	-0.1820****	0.0121****
			(0.0145)	(0.0794)	(0.0527)	(0.0101)	(0.0004)
stage 3	30,733	0.125	-0.2060****	0.8130****	0.5548****	0.0563****	0.0039****
			(0.0079)	(0.0304)	(0.0248)	(0.0046)	(0.0002)
stage 4	25,460	0.213	-0.1825****	1.0395****	0.7441****	-0.0714****	0.0048****
			(0.0055)	(0.0298)	(0.0247)	(0.0055)	(0.0002)
stage 5	29,340	0.085	0.0395****	0.1499****	0.1519****	-0.0843****	0.0003****
			(0.0021)	(0.0068)	(0.0080)	(0.0039)	(0.0001)

Notes. Robust standard errors are reported in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. For stage 0, $\bar{A}$ is not included in the regression since it takes 0 value for all observations.

Algorithm 3: K-Means clustering for the promotion policy of different video categories.

for $c \in \mathcal{C}$ **do**
  Classify observations $\cup_{v \in \mathcal{V}^c} \{A_{v,t-1}, x_{v,t}\}_{t=1,2,\ldots,T}$ into different lifetime stages by $\bar{A}_{v,t-1}$.
  for $s \in \mathcal{S}$ **do**
    Let $\mathcal{X}_s$ be the set of promotion fractions for observations at stage s.
    $\tilde{x}_s^c := \sum_{x \in \mathcal{X}_s} x / |\mathcal{X}_s|$.
  end
  $\tilde{\boldsymbol{x}}^c := (\tilde{x}_s^c)_{s \in \mathcal{S}}^\top$. `// average promotion policy for category` c.
end
$\tilde{X} := (\tilde{\boldsymbol{x}}^c)_{c \in \mathcal{C}}$. `// feature matrix for all categories.`
Impute the missing values matrix $\tilde{X}$ using k-Nearest Neighbors, with $k = 2$.
Perform K-Means clustering based on $\tilde{X}$.

Appendix B

Supplemental Materials for Chapter 3

B.1 Proofs and Supplements

B.1.1 Proofs for Section 3.2

The following lemma is useful in the proof of Proposition 3.2.1.

Lemma B.1.1 *For each agent $i \in V$, at least one of following value*

$$(a) \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\boldsymbol{Y}(t-1)=\boldsymbol{y}) \text{ or } (b) \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=0|\boldsymbol{Y}(t-1)=\boldsymbol{y})$$

are positive.

Proof of Lemma B.1.1: When random noise $\epsilon_i(t)$ is not bounded on either side, it is obvious that the statement holds. In the following, we only consider the situation when $\epsilon_i(t))$ is with support on some bounded interval $[\underline{\epsilon}, \overline{\epsilon}]$.

If (a) is not positive, there exists $\boldsymbol{y} \in \{0,1\}^{|V|}$ such that $F_\epsilon(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}) = 1$. Hence, we have

$$-v_i - \inf_{\boldsymbol{y}\in\{0,1\}^{|V|}} \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i} = -v_i \geq \overline{\epsilon}.$$

Consequently, we can derive the following inequality

$$-v_i - \beta > -v_i - \frac{1}{L} \geq -v_i - (\overline{\epsilon} - \underline{\epsilon}) \geq \underline{\epsilon},$$

where the first inequality follows from Assumption 3.2.2, the second inequality follows from property (ii) in Assumption 3.2.1. As a direct result, for any $\boldsymbol{y} \in \{0,1\}^{|V|}$, $F_\epsilon(-v_i - \beta\frac{\sum_{j\in\mathcal{N}_i} y_j}{n_i}) > 0$ and the value of (b) is positive.

In conclusion, for each agent $i \in V$, at least one of (a) and (b) have positive value. □

Proof of Proposition 3.2.1: Our goal is to show that although the MC may not be irreducible, there is only one recurrent communication class and it is aperiodic.

We first show that there is only one recurrent communication class. Follow after Lemma B.1.1, we construct $\boldsymbol{y}'$ as follows

$$y_i' = \begin{cases} 0 & \text{when } \prod_{\boldsymbol{y}\in\{0,1\}^{|V|}} \mathbb{P}(Y_i(t)=1|\boldsymbol{Y}(t-1)=\boldsymbol{y})=0, \\ 1 & \text{o.w.} \end{cases}$$

Hence, we have $\mathbb{P}(Y_i(t) = y_i'|\boldsymbol{Y}(t-1)=\boldsymbol{y}) > 0$ for all $i \in V$ and $\boldsymbol{y} \in \{0,1\}^{|V|}$. When given the previous adoption state, each agent $i \in V$ makes their decision independently. Consequently, $p(\boldsymbol{y}, \boldsymbol{y}') = \prod_{i\in V} \mathbb{P}(Y_i(t) = y_i'|\boldsymbol{Y}(t-1)=\boldsymbol{y}) > 0$ holds for all $\boldsymbol{y} \in \{0,1\}^{|V|}$. It further implies that all states of this MC communicates with state $\boldsymbol{y}'$. As a result, the states that $\boldsymbol{y}'$ communicates with forms a recurrent communication class while other states are in the transient classes.

Further, we can notice that since $p(\boldsymbol{y}', \boldsymbol{y}') > 0$ also holds, state $\boldsymbol{y}'$ has period 1 which implies that the recurrent communication class is aperiodic.

In conclusion, this MC has a limiting distribution $\boldsymbol{\pi}$ that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}P$ and the limiting adoption probability of each agent is a linear transformation of $\boldsymbol{\pi}$ that follows (3.2). □

B.1.2 Proofs and Supplements for Section 3.3

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

Proof of Proposition 3.3.1: 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 $\boldsymbol{a} \le \mathbf{b}$, we have $\sum_{j\in\mathcal{N}_i} a_j \le \sum_{j\in\mathcal{N}_i} b_j$ for all $i \in V$. Since CDF $F_\epsilon(\cdot)$ is monotonically increasing, if $\boldsymbol{a} \le \mathbf{b}$,

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

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

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 3.2.1, 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 3.2.2.

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 B.1.2.

Lemma B.1.2 (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 B.1.2:

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

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 3.3.1: We can decompose $\mathrm{Cov}(Y_i(t), Y_{i'}(t))$ into two parts by using the law of total covariance:

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

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

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

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

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

where (B.1a) follows from Cauchy–Schwarz inequality, (B.1b) follows from Lemma B.1.2 and (B.1c) 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

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

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

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

where the first inequality comes from (B.1) 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

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

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

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

By induction, we can show that

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

The conclusion follows. □

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

Proof of Lemma 3.3.2: 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 $t \geq 0$, 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

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

$$\leq \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\Big[L^2\big|\Delta_i(t-1)\big|^2\Big]} = \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)}\left[(L\beta)^2\left(\frac{\sum_{j\in\mathcal{N}_i} q_j(t-1)}{d_i} - \frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right)^2\right]}$$
$$= \sqrt{\rho^2 \operatorname{Var}\left[\frac{\sum_{j\in\mathcal{N}_i} Y_j(t-1)}{d_i}\right]},$$

where the first inequality follows by Jensen's inequality and the second inequality follows by Assumption 3.2.1.

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 (3.14), we can obtain

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

which further leads to

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

In summary, we have

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

and this concludes the proof. □

Finally, we prove our main result, i.e., Theorem 3.3.1.

Proof of Theorem 3.3.1: We first show by induction that, $\underline{\boldsymbol{\mu}}(t) \leq \mathbf{q}(t) \leq \overline{\boldsymbol{\mu}}(t)$ for each $t \geq 0$.

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 $\underline{\boldsymbol{\mu}}(t) \leq \mathbf{q}(t) \leq \overline{\boldsymbol{\mu}}(t)$. Then we have

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

where the first and last inequalities follow from Proposition 3.3.1(i) while the other two follow from Lemma 3.3.2.

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,

$$\underline{\boldsymbol{\mu}}^* \leq \mathbf{q}^* \leq \overline{\boldsymbol{\mu}}^* \text{ and } \underline{\boldsymbol{\mu}}^* \leq \boldsymbol{\mu}^* \leq \overline{\boldsymbol{\mu}}^*. \tag{B.3}$$

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 $i \in V$,

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

where the inequality comes from Assumption 3.2.1.

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

$$(\mathbf{I} - \rho\tilde{\mathbf{A}})|\Delta\boldsymbol{\mu}| \leq 2\boldsymbol{\delta}. \tag{B.4}$$

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 (B.4) by $(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}$, we obtain the inequality

$$|\Delta\boldsymbol{\mu}| \;\leq\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\delta}. \tag{B.5}$$

Combining (B.3) and (B.5), we finally have the following chain of inequalities:

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

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. $\square$

In the following, we prove the corollary for Theorem 3.3.1.

Proof of Corollary 3.3.2: From Theorem 3.3.1, it holds that

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

where (a) follows due to Cauchy-Schwarz inequality. Further, (a) proves (3.10) 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

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

Then, it holds that

$$\begin{aligned}&\left\|\tilde{\mathbf{A}}^s\mathbf{b}\right\|_1 \;=\; \mathbf{e}^\top\mathbf{D}\mathbf{Q}(s)\mathbf{D}\mathbf{e} \\ &=\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{d_i d_j}Q_{ij}(s) \;\leq\; \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{2}\left(\frac{1}{d_i^2}+\frac{1}{d_j^2}\right)Q_{ij}(s) \;=\; \frac{1}{2}\left\|\mathbf{D}^2\mathbf{Q}(s)\mathbf{e}\right\|_1+\frac{1}{2}\left\|\mathbf{Q}(s)\mathbf{D}^2\mathbf{e}\right\|_1,\end{aligned} \tag{B.6}$$

where the inequality follows from the AM-GM inequality.

We then bound the two terms in (B.6) 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

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

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 (B.6), we get

$$\left\|\tilde{\mathbf{A}}^s\mathbf{b}\right\|_1 \leq \frac{1}{2}\left(1+r^s(G)\right)\left\|\mathbf{b}\right\|_1 \leq r^s(G)\left\|\mathbf{b}\right\|_1, \tag{B.7}$$

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

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

where the first inequality follows from the subadditivity of norms and the last inequality follows from (B.7). This concludes the proof. □

B.1.3 Proofs and Supplements for Section 3.4

For a refined upper bound on the approximation error, we first show the proof for Lemma 3.4.1, which is a refined version of Lemma 3.3.2.

Proof of Lemma 3.4.1: 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

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

With Assumption 3.4.1, we can apply Taylor expansion to $F_\epsilon(X_i)$ and get

$$\begin{aligned} \Big| \mathbb{E}_{X_i} \big[F_\epsilon (X_i) - F_\epsilon (\nu_i) \big] \Big| \ = \ & \left| \mathbb{E}_{X_i} \left[F_\epsilon (\nu_i) + f_\epsilon(\nu_i)(X_i - \nu_i) + \frac{1}{2} f'_\epsilon(C_i)(X_i - \nu_i)^2 - F_\epsilon (\nu_i) \right] \right| \\ = \ & \frac{1}{2} \left| \mathbb{E}_{X_i} \left[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2 \right] \right|, \end{aligned} \tag{B.8}$$

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 (B.8) by

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

where the first inequality comes from Jensen's inequality and the second inequality is from Assumption 3.4.1. 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 3.3.1, we can finally get

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

which further leads to

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

In conclusion, we have

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

□

Based on Lemma 3.4.1, we then show the proof for the refined Theorem 3.4.1 and Corollary 3.4.1.

Proof of Theorem 3.4.1: Following the same steps leading to (B.5), with Lemma 3.4.1, we obtain

$$|\Delta\boldsymbol{\mu}| \;\le\; 2(\mathbf{I} - \rho\tilde{\mathbf{A}})^{-1}\boldsymbol{\eta}. \tag{B.9}$$

Therefore, following the same line of analysis in the proof of Theorem 3.3.1, it holds that

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

We conclude the proof. □

Proof of Corollary 3.4.1: By Theorem 3.4.1, 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 3.3.2, the last term can be bounded by

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

and we conclude the proof. □

Proof of Theorem 3.4.2: 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

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

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 3.3.1 to be valid, it suffices to use the Lipschitz constant $L = u$, see (B.1b).

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 $\text{var}(Y_i(t)) = q_i(t)(1-q_i(t))$ 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

$$\text{var}\big(Y_i(t)\big) \ \ge\ \big(1-F_\epsilon(-v)\big)\cdot F_\epsilon(-v-\beta) \ =\ \big(1-F_\epsilon(1.5)\big)F_\epsilon(0.5) \ \approx\ 0.114. \tag{B.10}$$

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

By Lemma 3.3.1, we have

$$\boldsymbol{\kappa}(t) \ \le\ \frac{1}{4d}\left[\mathbf{I} + \sum_{\tau=1}^{t}\left(\frac{\rho^2}{2}\right)^\tau \tilde{\mathbf{A}}^\tau\right]\mathbf{e} \ =\ \frac{1}{4d}\left[1 + \sum_{\tau=1}^{t}\left(\frac{\rho^2}{2}\right)^\tau\right]\mathbf{e} \ \le\ \frac{1}{4\left(1-\frac{\rho^2}{2}\right)d}\boldsymbol{e}, \tag{B.11}$$

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 (B.2) as

$$\text{Cov}(Y_i(t), Y_{i'}(t)) \ \ge\ -\frac{\rho^2}{2}\left[\text{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t-1)\right) + \text{Var}\left(\frac{1}{d}\sum_{j'\in\mathcal{N}_{i'}} Y_{j'}(t-1)\right)\right]. \tag{B.12}$$

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

$$\begin{aligned}\text{Var}\left(\frac{1}{d}\sum_{j\in\mathcal{N}_i} Y_j(t)\right) \ &=\ \frac{1}{d^2}\sum_{j\in\mathcal{N}_i}\text{Var}\left(Y_j(t)\right) + \frac{1}{d^2}\sum_{j\in\mathcal{N}_i}\sum_{j'\in\mathcal{N}_i, j'\ne j}\text{Cov}(Y_j(t), Y_{j'}(t)) \\ &\ge\ c_1\frac{1}{d} - \frac{d-1}{d^2}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\text{Var}\left(\frac{1}{d}\sum_{k\in\mathcal{N}_j} Y_k(t-1)\right) \end{aligned} \tag{B.13a}$$

$$\geq\ c_1\frac{1}{d} - \frac{1}{d}\cdot\frac{\rho^2}{2}\sum_{j\in\mathcal{N}_i}\mathrm{Var}\left(\frac{1}{d}\sum_{k\in\mathcal{N}_j}Y_k(t-1)\right) \tag{B.13b}$$

$$\geq\ c_1\frac{1}{d} - \frac{\frac{\rho^2}{2}}{4\left(1-\frac{\rho^2}{2}\right)d}, \tag{B.13c}$$

where (B.13a) follows from (B.12) and (B.10), (B.13b) follows from the trivial fact that the variance is nonnegative, and (B.13c) follows from (B.11).

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 3.4.1, we obtain

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

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,

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

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

$$\mathbf{q}(t)\ \geq\ \mathbf{h}(\mathbf{q}(t-1)) + c_3\frac{1}{d}\boldsymbol{e}. \tag{B.14}$$

Finally, we lower bound the approximation error in a way analogous to Theorem 3.3.1. Let $\zeta = c_3/n$. We show $\mathbf{q}(t) \geq \overline{\overline{\boldsymbol{\mu}}}(t) \geq \boldsymbol{\mu}(t)$ by induction, where

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

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 (B.14), the second and third inequalities follow Proposition 3.3.1 and the induction hypothesis, and the last inequality is trivial because $\zeta > 0$.

Thus, the following result holds,

$$\mathbf{q}^* \;\geq\; \overline{\overline{\boldsymbol{\mu}}}^* \;\geq\; \boldsymbol{\mu}^*, \tag{B.15}$$

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 (B.15). We conclude the proof. □

B.2 Supplements for Numerical Experiments on the FPA Scheme

B.2.1 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. 3.1a, while the intrinsic values assigned to each agent are detailed in Table B.1. 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 B.1.

B.2.2 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 B.1a shows an example of network size $n = 6$.

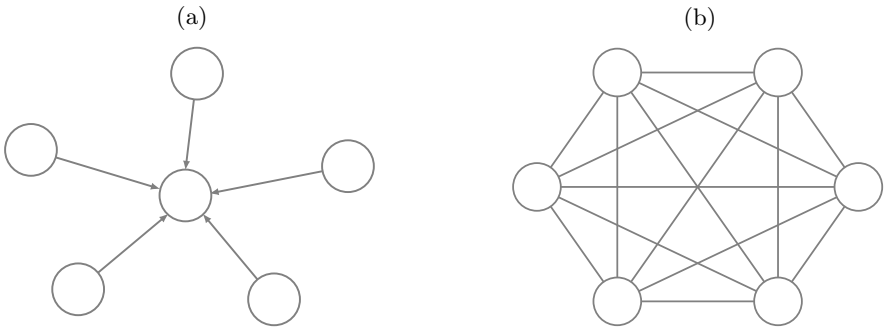
- *Complete network.* A complete network is the network where each node is directly connected to every other node. Figure B.1b shows an example of network size $n = 6$.

Table B.1: 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

Notes. p^* is calculated by first constructing a 1,024-state MC according to Section 3.2.2 and calculating the stationary distribution. μ^* is calculated by conducting fixed-point iteration according to (3.5), 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 $(\mu_i^{\mathrm{MM}} - q_i^*)$.

Figure B.1: Illustration of Highly Structured Symmetric Network Structure



Notes. (a) Directed star network. (b) Complete network.

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:

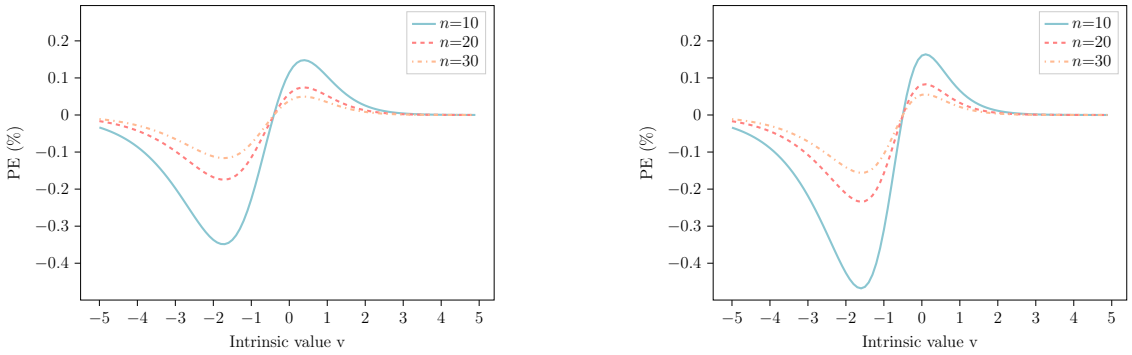
$$\text{PE} = \frac{\mu_i^* - q_i^*}{q_i^*} \cdot 100\%.$$

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 B.2 shows the PE of both network structures at different intrinsic values. Overall,

Figure B.2: PE Versus Intrinsic Value

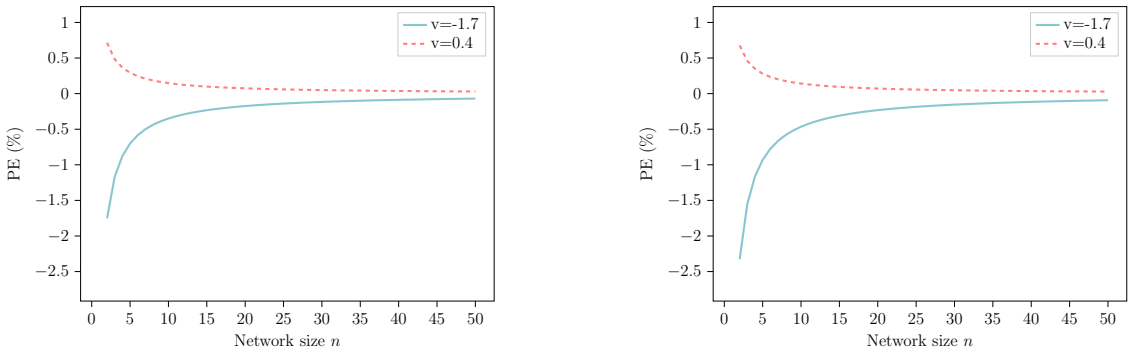


Notes. Left: directed star network. Right: complete network.

all instances have a small absolute percentage error (less than 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 B.3 shows the PE across these different network sizes. Regardless of the network structure and the intrinsic values, PE converges to 0 rapidly

Figure B.3: PE Versus Network Size



Notes. Left: directed star network. Right: complete network.

when the network size increases. This can be theoretically confirmed, and we have explicitly demonstrated it in Corollary 3.3.1 or a refined version in Theorem 3.4.1. Our findings show that, for highly-structured networks, the FPA scheme offers excellent approximation quality and exhibits asymptotic convergence as the network size grows.

B.2.3 MCMC Simulation Settings

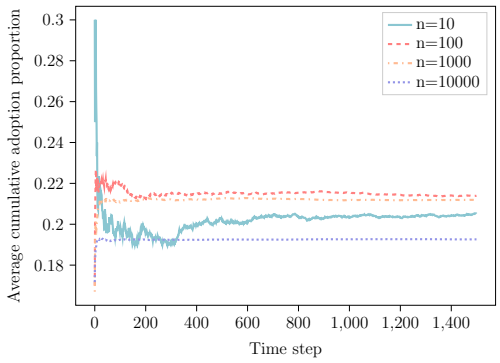
To estimate the limiting adoption probability for general instances where direct computation is impractical, we employ the MCMC simulation technique. The simulation initiates with all agents in a non-adopted state. We designate the first 1,000 time steps as the warm-up period to allow the system to reach a steady state, which we will elaborate on shortly. These initial steps are discarded from our analysis to avoid transient bias. In line with (3.3), 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 B.4, 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 $G(n, p(n))$. 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

Figure B.4: Average Cumulative Adoption Proportion Versus MCMC Time Steps



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.

B.2.4 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:

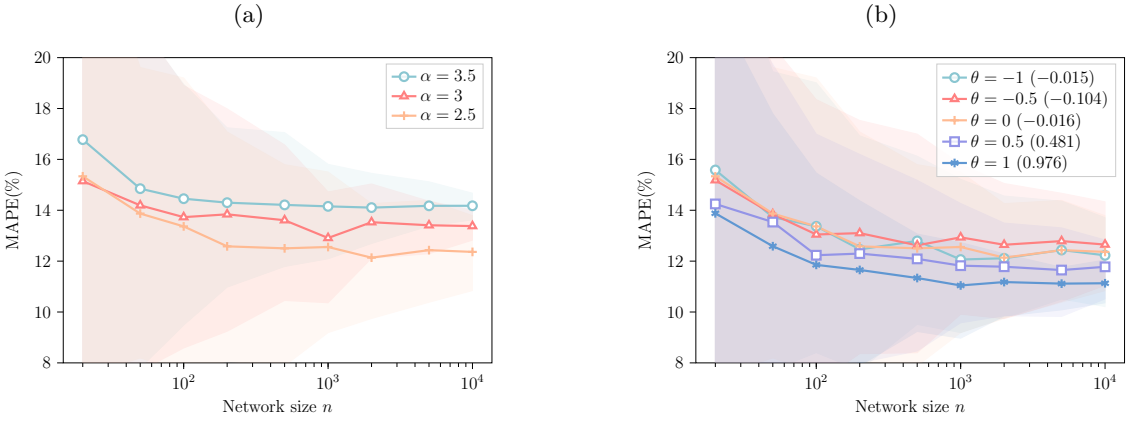
$$F_d(x;n) \;=\; \mathbb{P}(d \le x) \;=\; \frac{1-\left(\frac{x}{d_{\min}}\right)^{1-\alpha}}{1-\left(\frac{d_{\max}}{d_{\min}}\right)^{1-\alpha}} \quad \text{for} \quad d_{\min} \le x \le d_{\max} \;=\; n,$$

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 3.5) 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 B.2.5 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 B.5. In general, The FPA scheme still performs reasonably well. Also, we observe a consistent

Figure B.5: Performance of the FPA Scheme on Power-Law Networks of Different α and θ Values



Notes. 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. (a) MAPE against α. (b) MAPE against θ.

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-

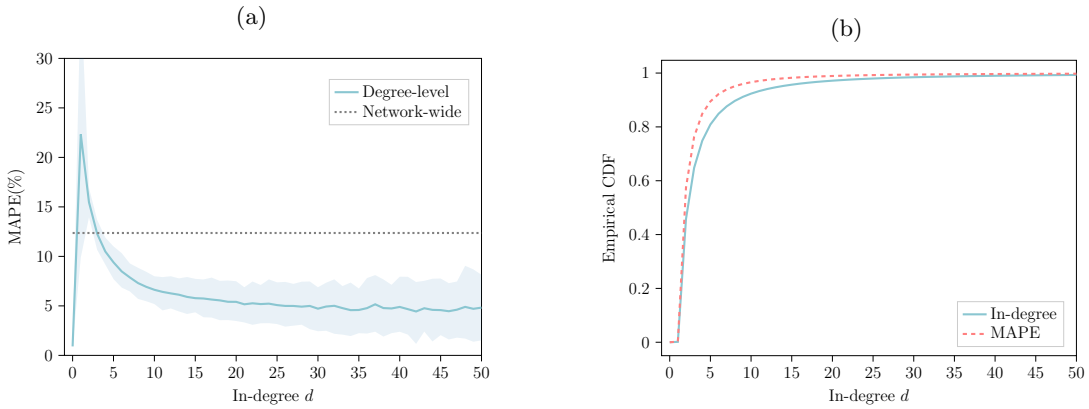
degrees. The increased variance, on the other hand, is primarily due to the more intricate structural variations inherent to power-law networks when specified parameters are used.

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 B.5a. 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 3.3.2 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 B.5b, 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 B.6a, we illustrate how the MAPE varies with respect to the in-degree d.

Figure B.6: Performance of the FPA Scheme on Power-Law Networks with Regard to the In-Degree Values



Notes. The dotted line in the left subfigure is the average network-wide MAPE. Shaded area represents 95% confidence interval. (a) MAPE against d. (b) Empirical CDF for in-degree and MAPE.

Aside from standalone nodes—which display zero error—the MAPE consistently de-

creases 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 B.6b, we extend our analysis by 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.

B.2.5 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 of networks.

(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 B.7.

B.3 Supplements for Applications

B.3.1 Proofs and Supplements for IM Problem

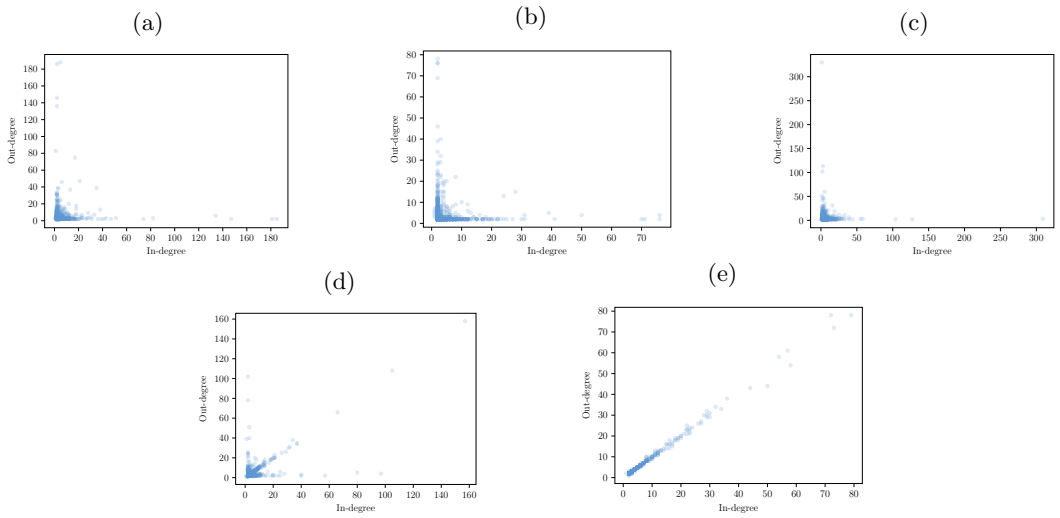
B.3.1.1 Proofs for Section 3.6.1

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

Proof of Proposition 3.6.1: The proof largely follows Corollary 3.3.2, where we bound the scaled ℓ_1-norm of the FPA error. Therefore, the regret can be bounded by

$$\begin{aligned}
&\text{Regret}(S^{\text{FPA}}) \\
=&\sum_{i\in V} q_i^*\left(S^*\right) - \sum_{i\in V} q_i^*\left(S^{\text{FPA}}\right) \\
=&\sum_{i\in V} q_i^*\left(S^*\right) - \sum_{i\in V} \mu_i^*\left(S^*\right) + \sum_{i\in V} \mu_i^*\left(S^*\right) - \sum_{i\in V} \mu_i^*\left(S^{\text{FPA}}\right) + \sum_{i\in V} \mu_i^*\left(S^{\text{FPA}}\right) - \sum_{i\in V} q_i^*\left(S^{\text{FPA}}\right)
\end{aligned}$$

Figure B.7: In-Degree and Out-Degree Distributions of Power-Law Networks with Different Pairwise Correlations θ



Notes. (a) $\theta = -1$. (b) $\theta = -0.5$. (c) $\theta = 0$. (d) $\theta = 0.5$. (e) $\theta = 1$.

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

where the first inequality holds trivially, the second inequality follows from the optimality of S^{FPA} for approximate IM problem (3.21), and the third inequality follows from Corollary 3.3.2. □

In the next, we discuss the applicability of Assumption 3.6.1.

Instances that satisfy Assumption 3.6.1. 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 $t \geq 1$. Therefore, for any $\beta > 0$, CDF F_ϵ can be expressed as $F_\epsilon(x) = \mathbb{1}\{x \geq -0.5\} \cdot (x + 0.5)$ on range $[0.5 - \beta, 0.5]$, which is convex. Additionally, some other diffusion instances related to common utility models can also meet Assumption 3.6.1. Some examples are: (i) Linear probability model: $v_i \geq -c$ and $\epsilon_i(t) \sim \mathcal{U}(-c, c)$ for all $i \in V$, $t \geq 0$. (ii) Logit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \text{Logistic}(0, s)$ and for all $i \in V$, $t \geq 0$. (iii) Probit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \mathcal{N}(0, s)$ and for all $i \in V$, $t \geq 0$. 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 3.6.1: 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 (3.21b) and (3.21c) 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 3.3.1(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 3.3.1(i) and the second inequality comes from Assumption 3.6.1.

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. □

B.3.1.2 Experiments for IM Problem

In the experiments, we consider two scenarios, one satisfies Assumption 3.6.1 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 $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim}$ 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 B.2, we show the numerical results of the greedy-FP algorithm.

Table B.2: 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 3.5.3 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.

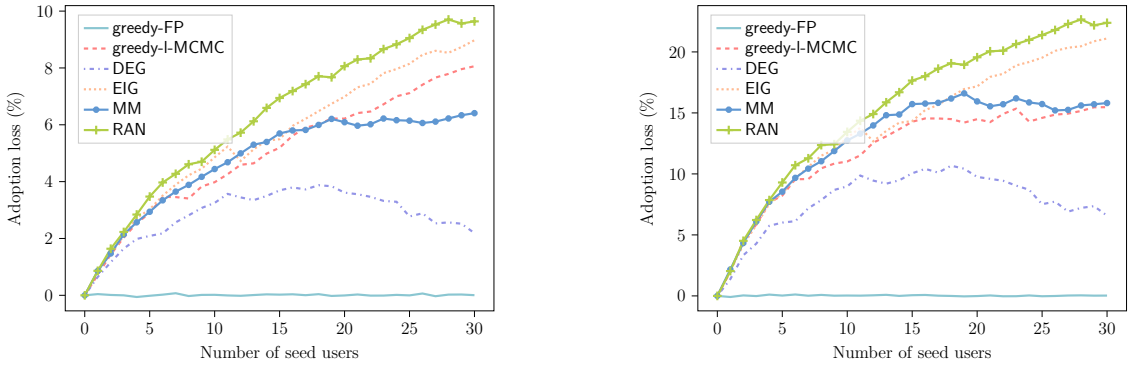
Benchmarks:

- *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 B.8 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.

Figure B.8: Performance of Different IM Algorithms



Notes. Left: submodular case. Right: non-submodular case.

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 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.

B.3.2 Proofs and Supplements for Pricing Problem

B.3.2.1 Proofs for Section 3.6.2

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

Proof of Proposition 3.6.2: The proof aligns with Corollary 3.6.1. Therefore, the regret can be bounded by

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

where the first inequality follows trivially, the second inequality follows since the optimality of $\mathbf{p}^{\text{FPA}}$ for approximate pricing problem (3.23), the third inequality follows from Corollary 3.3.2, 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 3.6.2: 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

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

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

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

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 $\beta \leq 3.375$. $\square$

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 3.3.1, 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 (3.27) 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:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} \approx \widetilde{\mathbf{G}}_k(\mathbf{p}) = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(I + \sum_{\ell=1}^{k} \left(\frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})}\right)^{\ell}\right) \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}), \quad \text{(B.16)}$$

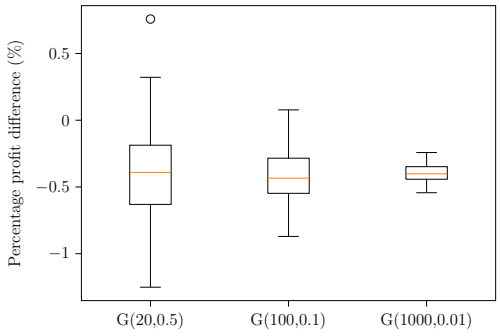
for the pricing problem. We expect such an easy-to-compute approximate gradient to lead to a significant efficiency gain, as is usually the case in the literature regarding approximate gradient descent (Ruder 2016). Previous works have applied similar low-order approximations for network effects for 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.

B.3.2.2 Experiments for Pricing Problem

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 B.9, we show the distribution of profit difference among all diffusion instances. We notice that the absolute profit difference is small. Furthermore, as the network

Figure B.9: Profit Difference between MCMC and the FPA Solution



becomes larger, 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.

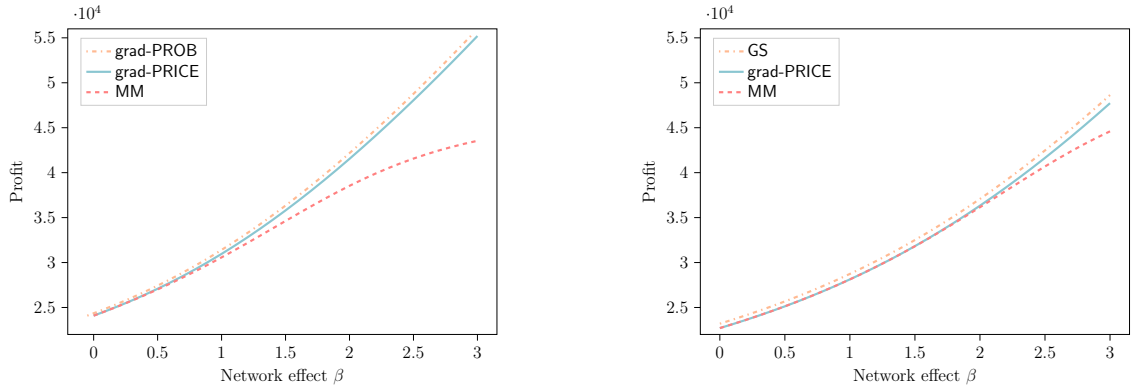
Hereafter, we compare the pricing scheme under the FPA scheme as default. We assume $\epsilon_i(t) \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$, which follows the theoretical analysis in Section 3.6.2. We study two extreme scenarios, the *perfect price discrimination* case, where each consumer 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 3.6.2, 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 3.5.3. 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 B.10 and B.11, 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 (3.27), which requires calculating the inverse of a n-by-n matrix. We resort to the second-order approximate gradient, $\widetilde{G}_2(\mathbf{p})$, as given in (B.16).

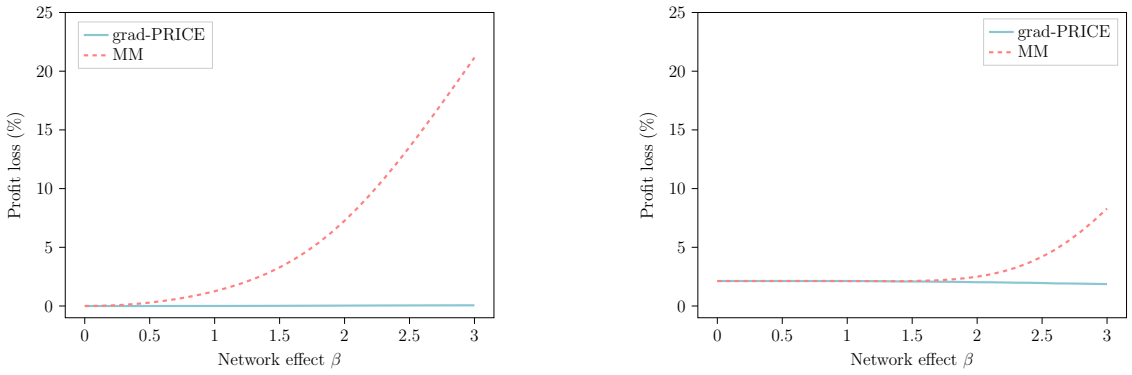
Figure B.10: Realized Profit Versus Network Effect.



Notes. 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.)

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.

Figure B.11: Profit Loss Compared with grad-PROB/GS Versus Network Effect.



Notes. Left: with price discrimination. Right: without price discrimination.

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 B.3 and B.4. 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 (B.16), grad-PRICE becomes a competitive price optimization technique. It can be efficiently implemented in various practical scenarios to find high-quality price solutions.

Table B.3: 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	19.034	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 B.4: 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