Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Content Promotion for Online Content Platforms with the Diffusion Effect

Yunduan Lin, Mengxin Wang, Heng Zhang, Renyu Zhang, Zuo-Jun Max Shen

To cite this article:
Yunduan Lin, Mengxin Wang, Heng Zhang, Renyu Zhang, Zuo-Jun Max Shen (2024) Content Promotion for Online Content Platforms with the Diffusion Effect. Manufacturing & Service Operations Management

Published online in Articles in Advance 13 Mar 2024

. https://doi.org/10.1287/msom.2022.0172

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Articles in Advance, pp. 1–20
ISSN 1523-4614 (print), ISSN 1526-5498 (online)
https://pubsonline.informs.org/journal/msom

Content Promotion for Online Content Platforms with the Diffusion Effect

Yunduan Lin,[a] Mengxin Wang,[b] Heng Zhang,[c] Renyu Zhang,[d,*] Zuo-Jun Max Shen[e]

[a] Civil and Environmental Engineering Department, University of California, Berkeley, Berkeley, California 94720; [b] Naveen Jindal School of Management, The University of Texas at Dallas, Richardson, Texas 75080; [c] W. P. Carey School of Business, Arizona State University, Tempe, Arizona 85287; [d] Chinese University of Hong Kong Business School, The Chinese University of Hong Kong, Hong Kong, China; [e] Faculty of Engineering and Faculty of Business and Economics, University of Hong Kong, Hong Kong, China

*Corresponding author

Contact: yunduan_lin@berkeley.edu, https://orcid.org/0000-0001-9155-3652 (YL); mengxin.wang@utdallas.edu, https://orcid.org/0000-0002-3378-9402 (MW); hengzhang24@asu.edu, https://orcid.org/0000-0002-6105-6994 (HZ); philipzhang@cuhk.edu.hk, https://orcid.org/0000-0003-0284-164X (RZ); maxshen@hku.hk, https://orcid.org/0000-0003-4538-8312 (Z-JMS)

Received: April 17, 2022
Revised: June 29, 2023; December 26, 2023
Accepted: February 2, 2024
Published Online in Articles in Advance: March 13, 2024

https://doi.org/10.1287/msom.2022.0172

Abstract. ***Problem definition*:** Content promotion policies are crucial for online content platforms to improve content consumption and user engagement. However, traditional promotion policies generally neglect the diffusion effect within a crowd of users. In this paper, we study the candidate generation and promotion optimization (CGPO) problem for an online content platform, emphasizing the incorporation of the diffusion effect. ***Methodology/results*:** We propose a diffusion model that incorporates platform promotion decisions to characterize the adoption process of online content. Based on this diffusion model, we formulate the CGPO problem as a mixed-integer program with nonconvex and nonlinear constraints, which is proved to be NP-hard. 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 prove the submodularity of the objective function for the CGPO problem, which enables us to find an efficient $(1-1/e)$-approximation greedy solution. Furthermore, we demonstrate that the D-OLS estimators are consistent and have smaller asymptotic variances than traditional ordinary least squares 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 with the policy implemented on the platform, our proposed promotion policy increases total adoptions by 49.90%. ***Managerial implications*:** Our research 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.

Funding: R. Zhang is grateful for the financial support from the Hong Kong Research Grants Council General Research Fund [Grants 14502722 and 14504123] and the National Natural Science Foundation of China [Grant 72293560/72293565].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2022.0172.

Keywords: online content • diffusion modeling • promotion optimization • approximation algorithms

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: Whereas 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), whereas 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 that millions of new videos are uploaded every day.

iii. User consumption behavior: Unlike in an e-commerce setting in which 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 in which 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.

Example 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 repost 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 (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, 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 in which 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 (CG) and promotion optimization (PO) (Davidson et al. 2010, Covington et al. 2016). The candidate generation stage selects a small subset of content from a large corpus, whereas 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 on which we focus 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. First, 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. Second, leveraging the high-granularity data commonly available on online platforms, we design a novel estimation method for the parameters in the P-BDM. Finally, our modeling framework, optimization algorithm, and estimation method are demonstrated to be effective through counterfactual analyses

based on real online content data. In the next section, we detail our contribution.

1.1. Contribution and Organization

- 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 data set.
- 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 that they yield smaller asymptotic variances compared with the ordinary least squares (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%.

The remainder of the paper is structured as follows: In Section 2, we review the related literature. In Section 3, we discuss the formulation of the P-BDM. In Section 4, we formulate the CGPO problem and propose the AGA for solving it. In Section 5, we discuss the estimation issues for the P-BDM and propose the D-OLS method. Section 6 presents our numerical studies based on real-world data, followed by concluding statements in Section 7.

2. Literature Review

As we discuss 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), and it 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 that 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). Whereas 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 also explores diffusion in the context of social networks, in which each user is influenced only by the user's 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. Whereas users exhibit different behaviors, the process of candidate selection and promotion optimization shares some similarities with assortment and pricing problems. Golrezaei et al. (2020) 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 in which purchase decisions can be influenced by earlier purchases. Du et al. (2016) and Wang and Wang (2017) propose a variant of the multinomial logit model incorporating the network effect in an assortment optimization problem. Follow-up works (Chen and Chen 2021, Nosrat et al. 2021) also involve different choice models. This line of research inspires us to consider similar problems on online content platforms.

3. Promotion Bass Diffusion Model

In this section, we introduce the 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.

3.1. Background and Motivation from a Large-Scale Video-Sharing Platform

We begin with a brief overview of the 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, whereas 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.

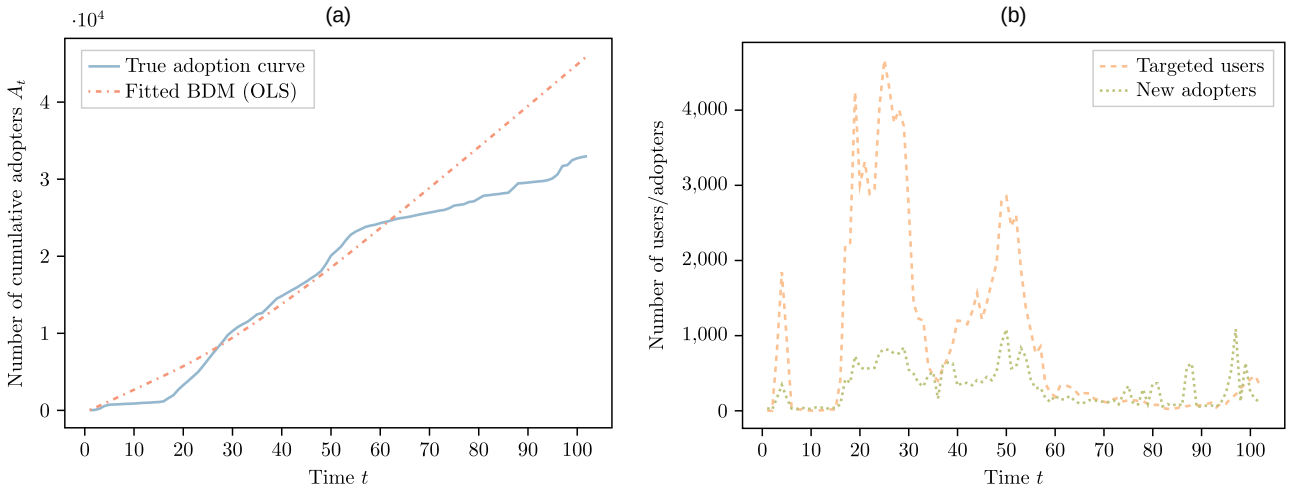
The discrete-time BDM models the adoption process of a product over a discrete finite time horizon $t = 1, 2, \ldots, T$ in a market of population size m. The initial number of adopters is denoted by A_0, and the number of new adopters at each time period t is given by

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

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 nonadopters 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 6.1. In Figure 1, we use a single video to showcase the typical pattern of the content diffusion process. In Figure 1(a), we present

Figure 1. (Color online) 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.

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 Online Section B.1.1.

A detailed explanation of our fitting method can be found in Section 5.1. Specifically, we first estimate the parameters of the BDM using data from the initial 60 periods and 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 periods 0 to 60 and 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 nonadopters are impacted by the innovation effect, as outlined in (1). However, this is unlikely to hold for online content because of limited user time and targeted promotion strategies. Figure 1(b) 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 are 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.

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 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) the promotion effect, which reflects the intrinsic preference of users toward 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) the diffusion effect, which represents the influence of the adopted population on others, that is, 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 nonadopters 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 nonadopters in the market at time $t-1$ (i.e., $x_t \le 1 - A_{t-1}/m$). Denoting A_0 as the initial number of adopters, the P-BDM assumes that the number of new adopters at time t is

$$a_t = \underbrace{\left(p + q\frac{A_{t-1}}{m}\right)mx_t}_{\text{Direct adopters}} + \underbrace{q\frac{A_{t-1}}{m}(m - A_{t-1} - mx_t)}_{\text{Indirect adopters}}$$
$$= \underbrace{pmx_t}_{\text{Promotion effect}} + \underbrace{q\frac{A_{t-1}}{m}(m - A_{t-1})}_{\text{Diffusion effect}}. \quad (2)$$

In the P-BDM dynamics (2), we categorize adopters based on whether they receive the promotion or not. For those not exposed to promotion, categorized as indirect adopters, 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 in which all nonadopters 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 life span 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 Online Section B.1.2. It is important to note that we consider the time-decay factor as an external influence, which means that 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 6), in which we explore the effect of time-decay factor γ further.

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 because of 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, whereas 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 CGPO problem as an optimization task to incorporate the diffusion effect into content promotion, using the P-BDM as a basis.

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

$$\max_{U \subseteq \mathcal{V}: |U| \le K} \quad R(U; C) + R(\mathcal{V} \setminus U; 0), \quad (3)$$

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 zero. 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 PO subproblem:

$$\max_{\boldsymbol{x} \ge \boldsymbol{0}, \boldsymbol{A}_{U,1:L}} \quad \sum_{v \in U} A_{v,L} \quad (4a)$$

$$\text{s.t.} \quad A_{v,t} = A_{v,t-1} + p_v m x_{v,t} + \frac{q_v}{m} A_{v,t-1}(m - A_{v,t-1}), \quad \forall v \in U \quad \forall t = 1, \ldots, L, \quad (4b)$$

$$x_{v,t} \le 1 - \frac{A_{v,t-1}}{m}, \quad \forall v \in U \quad \forall t = 1, \ldots, L, \quad (4c)$$

$$m\sum_{t=1}^{L}\sum_{v\in U} x_{v,t} \le C. \tag{4d}$$

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. Objective (4a) represents the total adoptions of all candidates in set U at the end of the L-period planning interval; (4b) mandates the cumulative adopters follow the P-BDM diffusion dynamics as defined in (2); (4c) ensures that the number of users receiving the promotion does not exceed the remaining nonadopters; and (4d) ensures, 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 $\overline{C} := C/mL$ 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) when $x_{v,t}$ is set to zero for all $v \in \mathcal{V} \setminus U$ and $t = 1, 2, \ldots L$.

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, $\boldsymbol{x} = (x_{v,t} : v \in \mathcal{V}, t = 1, \ldots, L)$, $\boldsymbol{x}_{v,:} = (x_{v,t} : t = 1, \ldots, L)$, 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, $\boldsymbol{x}_{v,2:L} = (x_{v,t} : t = 2, \ldots, L)$. 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 because of 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 an L-period window. We can reoptimize it periodically to account for the environmental changes over the extended time horizon.

4.2. Promotion Optimization Given the Content Set

The CGPO problem inherently comprises two stages, namely, CG and PO. The CG stage as represented in Problem (3) is a combinatorial optimization problem that embeds the PO stage as shown in Problem (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 (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 1 (NP-Hardness). *The CGPO problem (3) is NP-hard.*

For a detailed proof of Theorem 1, please refer to Online Section A.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 (4) remains difficult to solve because of its nonconvex 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.

4.2.1. Convex Relaxation.

The nonconvexity of Problem (4) originates from the set of equality constraints (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 (4b) as inequalities as follows:

$$A_{v,t} \le A_{v,t-1} + p_v m x_{v,t} + \frac{q_v}{m} A_{v,t-1}(m - A_{v,t-1}), \quad \forall v \in U \quad \forall t = 1, \ldots, L. \tag{5}$$

We denote the relaxed problem as PO-CR, which uses (4a) as the objective and includes (5), (4c), and (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 (4). Moreover, as we illustrate in Theorem 2, the PO-CR problem is, in fact, equivalent to the original PO problem (4).

Theorem 2 (Relaxation). *The PO problem (4) and relaxed problem PO-CR are equivalent.*

We remark that the equivalence is nontrivial because the decision variables $\boldsymbol{A}$ and $\boldsymbol{x}$ have opposing relationships in Constraints (4b) and (4c). Specifically, increasing $A_{v,t}$ seems to increase the objective value because of (4b), but it lowers $x_{v,s}$ for $s \ge t + 1$ because of (4c). To establish the equivalence, we show that the optimal solutions of the PO-CR problem are feasible solutions to the PO problem (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 (4b) is more beneficial than maintaining the constraint as a strict inequality for a larger upper bound of $\boldsymbol{x}$ in Constraint (4c). Complete proofs are in the online 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 (3) as a whole, we need to utilize the optimality condition of the PO-CR problem as detailed in Section 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.

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 $\boldsymbol{x}$:

$$\max_{\boldsymbol{x}\geq \boldsymbol{0}} \quad \sum_{v\in U} f_v(\boldsymbol{x}_{v,:}) \tag{6a}$$

$$\text{s.t.} \quad m\sum_{t=1}^{L}\sum_{v\in U} x_{v,t} \leq C, \tag{6b}$$

$$x_{v,t} \leq 1-\frac{A_{v,0}}{m}, \quad \forall\, v\in U \quad \forall\, t=1,\ldots,L, \tag{6c}$$

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

$$f_v(\boldsymbol{x}_{v,:}) := \max_{\boldsymbol{A}_{v,1:L}} \; A_{v,L} \quad \text{s.t.} \quad (5),(4c). \tag{7}$$

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 (6c), which ensures that Problem (7) always has a feasible solution. We elaborate on the rationale behind these constraints and the process of constructing a feasible solution to Problem (7) for any given policy in Online Section A.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 is 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 1.

Lemma 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 1 derives from the fact that Problem (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 (6) as the dual problem (8) with θ being the Lagrangian multiplier for Constraint (6b):

$$\min_{\theta\geq 0} \; \sum_{v\in U} h_v(\theta) + \theta C. \tag{8}$$

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

$$h_v(\theta) := \max_{\boldsymbol{x}_{v,:}\in[0,1-A_{v,0}/m]^L} f_v(\boldsymbol{x}_{v,:}) - \theta m\sum_{t=1}^{L} x_{v,t}. \tag{9}$$

For any candidate set $U \subseteq \mathcal{V}$, let $\theta^*(U)$ denote the optimal dual variables of dual problem (8). In Lemma 2, we provide a comparison of optimal dual variables for nested candidate sets $U_1 \subseteq U_2 \subseteq \mathcal{V}$.

Lemma 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 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 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.

4.3. Candidate Generation

In this section, we address the 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 (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.

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 $w\in\mathcal{V}\setminus U_2$,

$$\begin{aligned}&R(U_1\cup\{w\};C)-R(U_1;C)-R(\{w\};0)\\&\quad\geq R(U_2\cup\{w\};C)-R(U_2;C)-R(\{w\};0).\end{aligned}\tag{10}$$

The left and right sides of (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 the 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 the 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:

$$R(U\cup\{w\};C):=\max_{0\leq c\leq C}[R(U;c)+R(\{w\};C-c)],\tag{11}$$

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 (11), the marginal gain from including content piece w given content set U can be expressed as

$$\begin{aligned}&R(U\cup\{w\};C)-R(U;C)-R(\{w\};0)\\&=[R(U;c^*(U))-R(U;C)]+[R(\{w\};C-c^*(U))-R(\{w\};0)].\end{aligned}\tag{12}$$

The marginal gain is decomposed into two parts: the adoption loss of set U because of the cannibalization of the 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 (8), we formulate the PO dual problem for set U and promotion budget c as

$$R(U;c)=\min_{\theta\geq0}\sum_{v\in U}h_v(\theta)+\theta c,\tag{13}$$

where $h_v(\theta)$ adheres to the same definition in (9). By the envelope theorem, we can express the difference between two optimal values as an integral of the optimal dual variable, such as

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

where $\Theta^*(U;z)$ is the set of optimal dual variables to Problem (13) when the budget is z.

Consequently, the first term in (12) can be represented as

$$\begin{aligned}R(U;c^*(U))-R(U;C)&=\int_{z=0}^{c^*(U)}\theta^*(U;z)dz\\&\quad-\int_{z=0}^{C}\theta^*(U;z)dz\\&=-\int_{z=c^*(U)}^{C}\theta^*(U;z)dz.\end{aligned}$$

In a similar manner, we can express the second term in (12). Therefore, the marginal gain of adding piece w to the candidate set U can be represented by the optimal dual variables as

$$(12)=-\int_{z=c^*(U)}^{C}\theta^*(U;z)dz+\int_{z=0}^{C-c^*(U)}\theta^*(\{w\};z)dz.$$

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 in a comparison between the optimal dual variables of PO problems with two nested sets. This leads us directly to Theorem 3.

Theorem 3 (Submodularity). *The CGPO objective, $R(U;C)+R(\mathcal{V}\setminus U;0)$, is a monotone submodular set function with respect to content set $U\subseteq\mathcal{V}$.*

The proof of Theorem 3 relies on transforming marginal gain and utilizing the monotonic property with nested sets. The complete proof is included in Online Section A.1.4. As a result, the CGPO problem (3) can be viewed as a monotone submodular maximization problem with a cardinality constraint.

4.3.2. Accelerated Greedy Algorithm. The well-known greedy algorithm (Nemhauser et al. 1978) provides a $(1-1/e)$-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 Problem)

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 AGA, which is detailed as follows.

4.3.3. 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 the Lagrangian dual variable as $\max\{\theta^*(U_k), \theta^*(U_{k-1} \cup \{v\})\}$.

As indicated by Lemma 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.

5. Parameter Estimation

In this section, we discuss how to estimate the parameters of the P-BDM by adapting the classic 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.

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 OLS method. Schmittlein and Mahajan (1982) and Srinivasan and Mason (1986) apply maximum likelihood estimation (MLE) and nonlinear least square to obtain better estimates. However, these methods and their analyses are complicated by the diffusion nature, such as autocorrelation, which exists among observations, because diffusion happens as a dynamic process. 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). Estimating the parameters of the P-BDM introduces additional challenges because of 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, whereas there are inherent deficiencies in estimating diffusion models as mentioned, we can largely alleviate these issues and improve the estimation results on online platforms. In fact, compared with 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- and MLE-based estimators. In the following discussions, we focus on a fixed $v \in \mathcal{V}$, omit the subscript v, and treat the market size m as fixed.

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 Online Section A.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 zero 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, because 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 $\{(a_t^{\mathrm{d}}, a_t^{\mathrm{i}}, A_t, x_t)\}_{t=1}^T$. We propose a straightforward D-OLS method based on the following relationships:

$$a_t^{\mathrm{d}} = p \cdot mx_t + q \cdot A_{t-1}x_t + \epsilon_t^{\mathrm{d}} \quad \text{and}$$
$$a_t^{\mathrm{i}} = q \cdot \frac{A_{t-1}}{m}(m - A_{t-1} - mx_t) + \epsilon_t^{\mathrm{i}}, \tag{14}$$

where the first equation in (14) focuses on the direct adopters targeted by promotion, whereas 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 (14), resulting in

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

ii. We use the OLS method again, but this time, we substitute q in the first equation in (14) with the D-OLS estimator $\hat{q}^{\text{D-OLS}}$ to compute $\hat{p}^{\text{D-OLS}}$, which is given by

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

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.

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 because of 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 that the observations come from a fixed-design triangular array wherein the nth 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 nth diffusion process, let $\{A_{i,(n)}\}_{i=1}^n$ denote the adopters at n different time steps and $\{x_{i,(n)}\}_{i=1}^n$ 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:

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

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

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

where $\overline{A}_{i,(n)} = A_{i,(n)}/m_{(n)}$ 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 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 that the scaled random noise for the nth diffusion process $\overline{\epsilon} := \epsilon/m_{(n)}$ has variance σ^2. The following theorems, Theorems 4 and 5, show the asymptotic properties of D-OLS estimators. The detailed proof is given in Online Section A.2.2.

Theorem 4 (Consistency). *Suppose that the scaled random noise $\overline{\epsilon}_i^{\mathrm{i}} := \epsilon_i^{\mathrm{i}}/m_{(n)}$ and $\overline{\epsilon}_i^{\mathrm{d}} := \epsilon_i^{\mathrm{d}}/m_{(n)}$ are independently and identically distributed with mean zero and finite variance for all $i = 1, \ldots, n$; then, D-OLS estimators $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$ 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 4 implies that, with sufficient observations, the true values of p and q can be uncovered.

Theorem 5 (Asymptotic Normality). *Suppose that the scaled random noise* $\overline{\epsilon}_i^{\mathrm{i}}$ *and* $\overline{\epsilon}_i^{\mathrm{d}}$ *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* $\hat{p}_{(n)}^{\text{D-OLS}}$ *and* $\hat{q}_{(n)}^{\text{D-OLS}}$ *are asymptotically normal. Specifically,*

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

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

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

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

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

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

We draw two insights based on Theorem 5. 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 $\overline{A}(1-\overline{A})$, κ approaches infinity, whereas ξ_1 and ξ_2 remain bounded. Therefore, D-OLS estimators are more robust against correlations than OLS estimators. Second, when $\eta \le \tilde{Q}_{22}/Q_{22}$, D-OLS estimators have smaller asymptotic variances than OLS estimators (see Proposition EC.1 in Online Section A.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 $\eta \le \tilde{Q}_{22}/Q_{22}$ to be readily fulfilled in our setting, suggesting that D-OLS estimators present smaller asymptotic variances than OLS estimators.

5.2. MLE Estimators

Whereas 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 $(m - A_{t-1})$ nonadopters, 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

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

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

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

and the derived estimators are named D-MLE estimators. In Online Section A.2.3, we show that both log-likelihood functions are concave, allowing us to use the gradient method for estimation.

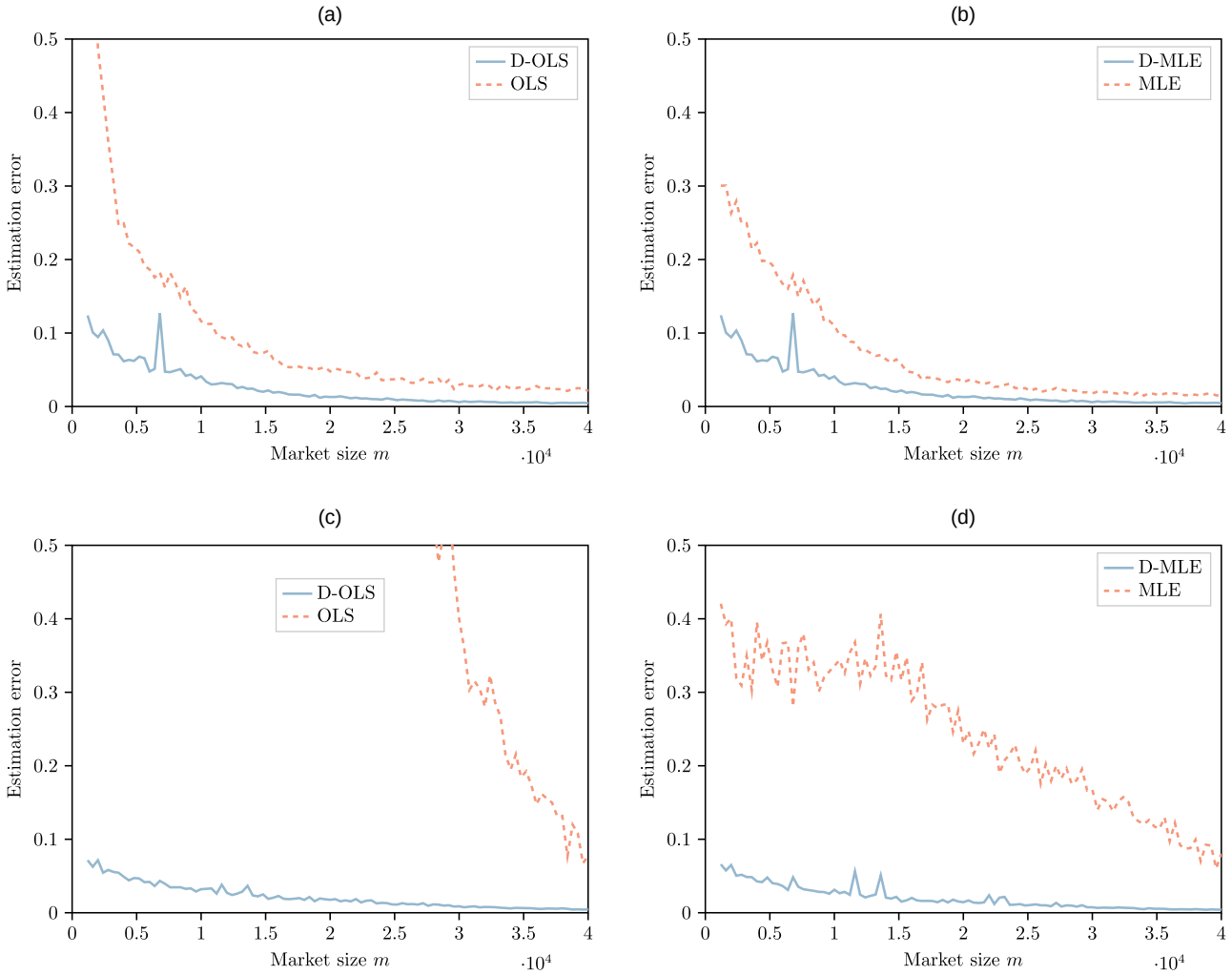
5.3. Comparing OLS- and MLE-Based Estimators with Simulation

We create a synthetic data set 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 5.2 when adopter types can be distinguished. We assess the estimators under two promotion schemes: (i) Const: promotion fraction x_t remains constant, and (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 5.1.1. Performance is measured using the 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 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. We offer two more observations. First, comparing Figure 2(a) with (c) and

Figure 2. (Color online) Estimation Errors for Different Methods Against Market Size (Scale with the Number of Observations)



Notes. (a) Const: OLS-based methods. (b) Const: MLE-based methods. (c) Linear: OLS-based methods. (d) Linear: MLE-based methods.

Figure 2(b) with (d), we see larger relative improvements under the Linear scheme compared with 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 5 and highlights the effectiveness of our proposed estimators. Second, comparing Figure 2, (a) and (b), or Figure 2, (c) and (d), we notice that D-OLS and D-MLE perform similarly when adopter types can be distinguished, whereas 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. Whereas 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.

6. Numerical Results

In this section, we conduct a comprehensive analysis using data from a large-scale video-sharing platform. To help readers better understand our numerical results, we provide the code for our analyses in a GitHub repository (https://github.com/YunduanLin/Content_Promotion).

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 more than 360 million daily active users and more than 20 billion videos. Effective content promotion plays an important role in platform operations. Whereas machine learning–based algorithms offer personalized recommendations curated based on user interests, promoting content that has the potential to go viral is challenging because of 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 (July 1–20, 2020). The logs contain time-stamped 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. Because of 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 time-stamped 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 according to its topic. The data set includes videos from 61 category labels provided by the platform, ranging from 155 to 2,759 videos per category.

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 Online Section B.1.2 for more details.

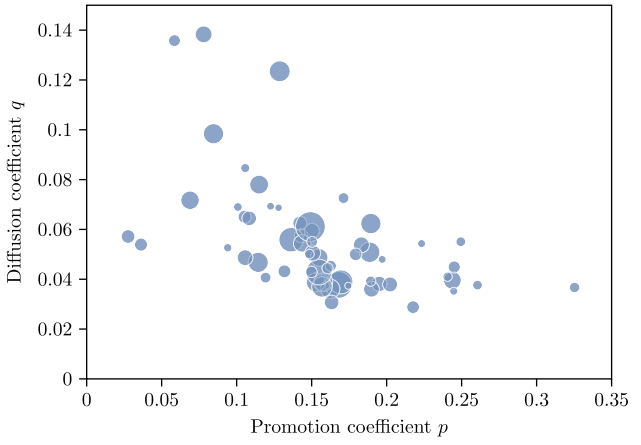
ii. Group estimation: We estimate the same p and q values for each video category. We highlight that promotion decisions are often made at the early stages of a video's life cycle when limited data are available for estimation. Consequently, group-wise estimation is typically utilized to guarantee generalizability. In principle, we can adopt a contextual approach given the availability of the featured information of each content piece. The group-based estimation can be seen as a special case of this approach, in which 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 Online Section B.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 Online Section B.1.4. Next, we present the calibration results under the best time-decay factor $\gamma = 0.983$.

6.2.1. Distribution of p and q. Figure 3 depicts the distribution of estimated coefficients across 61 different categories. Notably, a negative correlation between p and q is observed 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.

6.2.2. 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. Second, 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

Figure 3. (Color online) Distribution of Estimated Promotion Coefficient p and Diffusion Coefficient q



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

moderate yet noticeable enhancement compared with the OLS method. In contrast to the simulation in Section 5.3, the colinearity issue brought up in Section 5.1.1 is not severe in this data set. 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, whereas the remaining six categories do not show significant changes.

To further illustrate the effectiveness of the P-BDM, we present two examples in Figure 4. Figure 4(a) uses the same video as the example in Section 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% of data samples to estimate coefficients and generate the fitted curves for the entire time horizon using the estimated coefficients. Compared with the BDM, the P-BDM fits not only the overall adoption trend, but also the curve shape. Whereas 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 4(b), the estimated coefficient can even be negative, which lacks a valid real-world interpretation. These observations underscore the effectiveness of the P-BDM.

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.

6.3.1. Long-Term Performance with Different Planning Intervals.

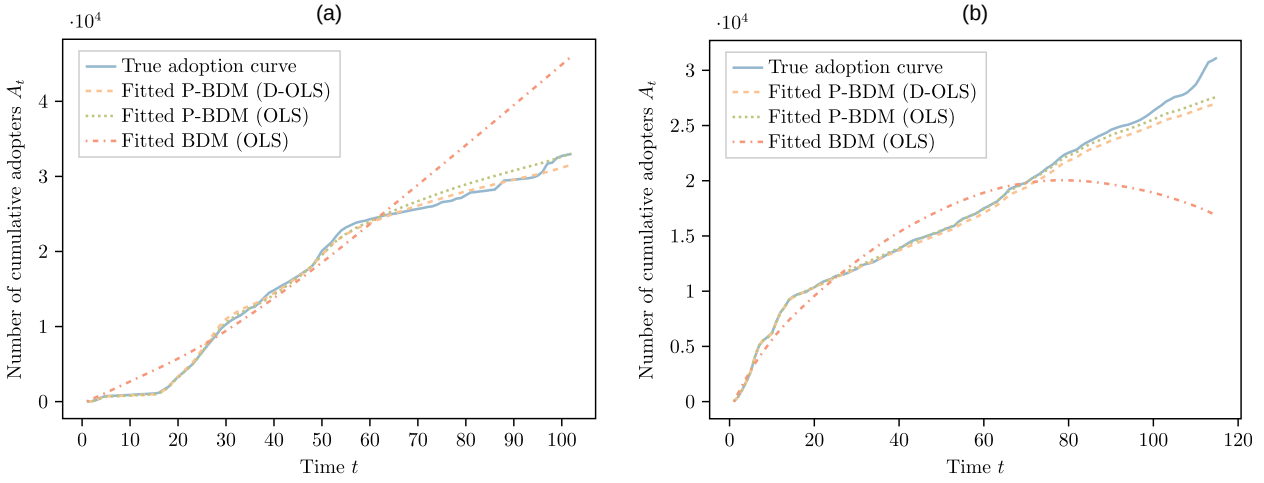
In practice, platforms are concerned with the long-term efficacy of promotion policies. Accordingly, we solve the CGPO problem every L periods using the AGA policy in the most recent platform environment.

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 6.2.
- User behavior: At each time, users act according to the D-MLE stochastic counterpart of the P-BDM as described in Section 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

Figure 4. (Color online) 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 3.1. (b) Example of negative fitted BDM coefficient.

effects but may delay the promotion of new videos because of less frequent policy updates. 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 5, we fix the average promotion budget per user per period $\overline{C}$ at six, and vary the size of the candidate set size to be $K \in \{30, 50, 70\}$. As shown in Figure 5(a), we notice that this trade-off 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 because of 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 5(b) 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.

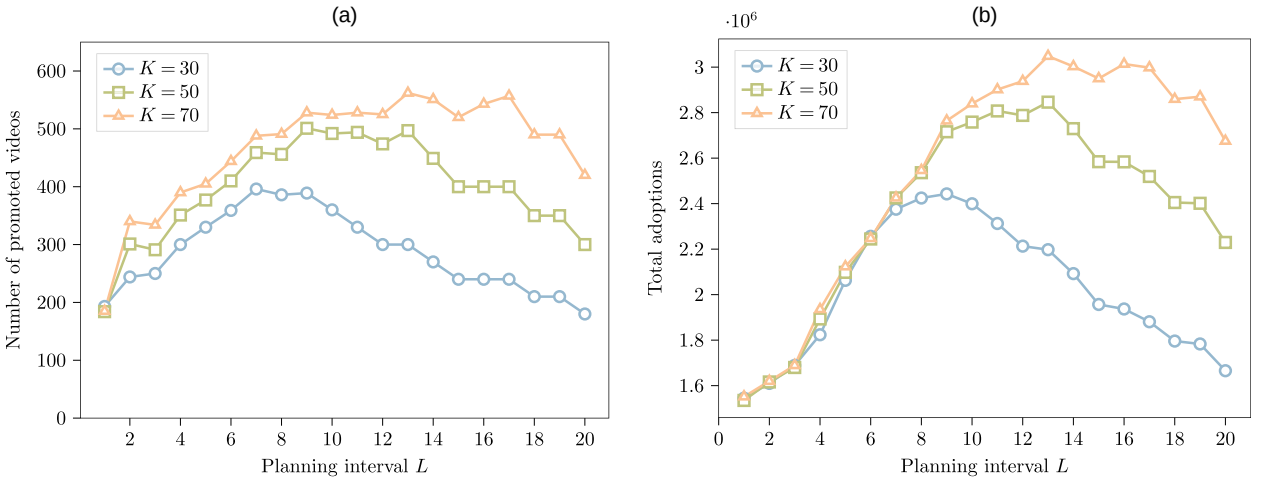
In Figure 6, we fix the candidate set size to be 50 and vary the average promotion budget to be $\overline{C} \in \{2, 4, 6, 8, 10\}$. From Figure 6(a), we observe a similar trade-off akin to our previous findings except that an increase in C leads to a smaller optimal planning horizon. As shown in Figure 6(b), 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 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 because of 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 7 shows the average promotion times received by videos in different categories. Although videos with a large p value tend to receive more 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.

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.

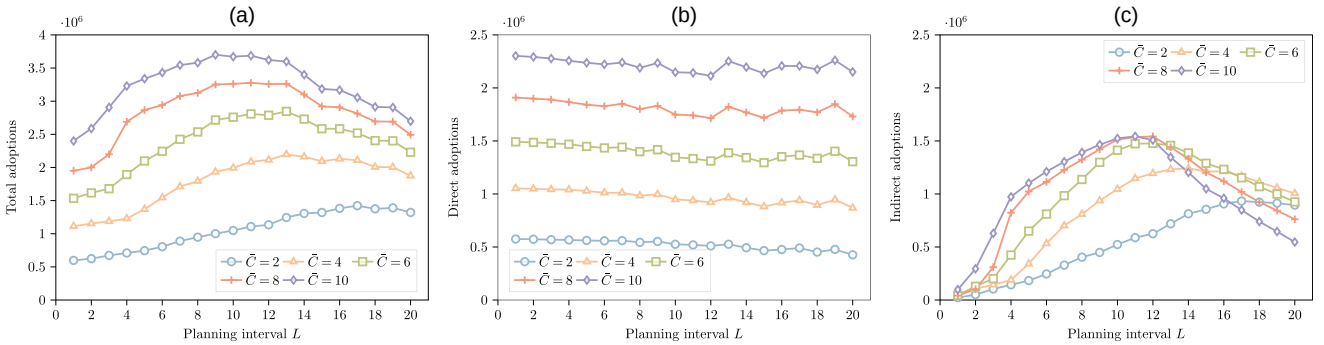
We aim to understand how a video's configuration (p_v and q_v) and lifetime ($A_{v,t-1}$) affect the promotion fraction $x_{v,t}$ in the AGA policy. We use our experimental results as observations, in which each observation represents a promotion fraction $x_{v,t}$ allocated to a video v at the beginning 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 $i = \lceil 5A_{v,t-1}/m \rceil$.

Figure 5. (Color online) Illustration of the AGA Policy for Different Selections of Candidate Set Size K



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

Figure 6. (Color online) Illustration of the AGA Policy for Different Selections of Promotion Budget

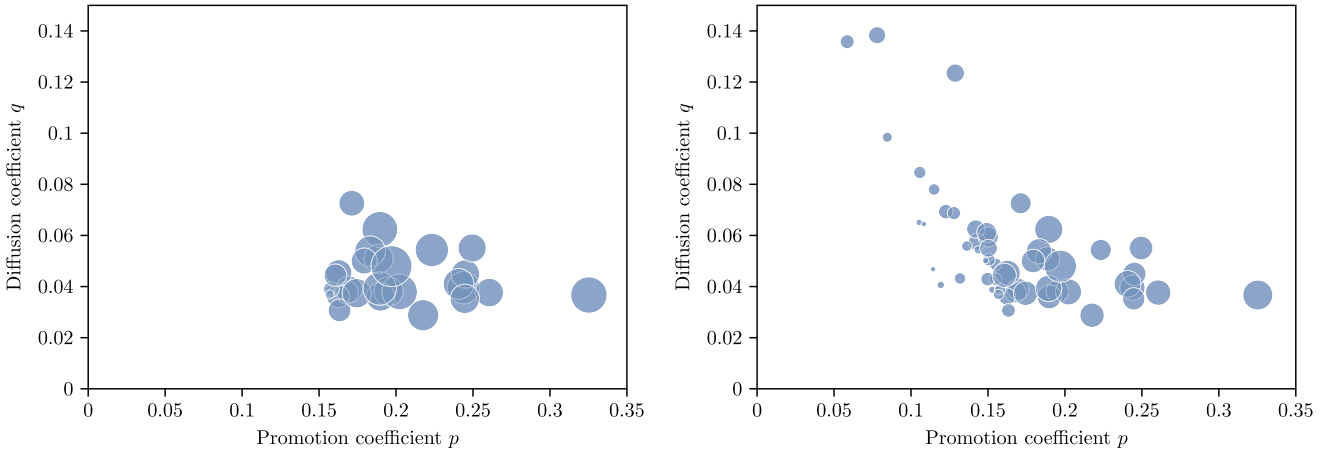


Notes. (a) Total adoptions. (b) Direct adoptions. (c) Indirect adoptions.

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, controlling other relevant covariates. For a detailed explanation of this analysis, please refer to Online Section B.2.2. The regression coefficients of p and q, which we interpret as their impacts, are presented in Figure 8. Figure 8(a) 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 8(b) shows the impact ratio between p and q, indicating that p carries more weight than q, particularly during the intermediate stages.

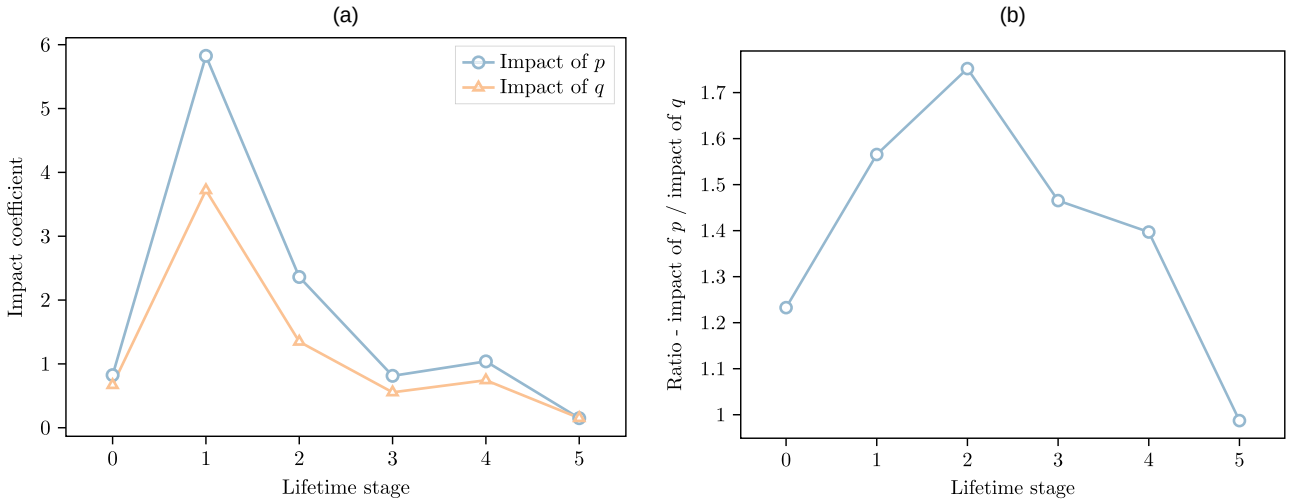
Then, we use K-means clustering to group video configurations into four clusters based on their lifetime promotion policies. The clustering procedure is described in Online Section B.2.3. Figure 9(a) displays the clustering results according to video configuration, whereas Figure 9(b) showcases the average 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 9(a), is distinct because of its notably lower promotion at stage 0. These are the videos discarded by the AGA policy. Roughly speaking, moving toward the right of Figure 9(a), videos receive more promotions. Echoing the insights from our

Figure 7. (Color online) Illustration of the AGA Policy Across Different Video Categories



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

Figure 8. (Color online) 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.

sensitivity analysis, we notice a trend in which the points of peak promotion shift toward later stages as p increases. In contrast, videos in clusters 1 and 2 that have larger q values and smaller p values need to be promoted early to take advantage of their diffusion potential.

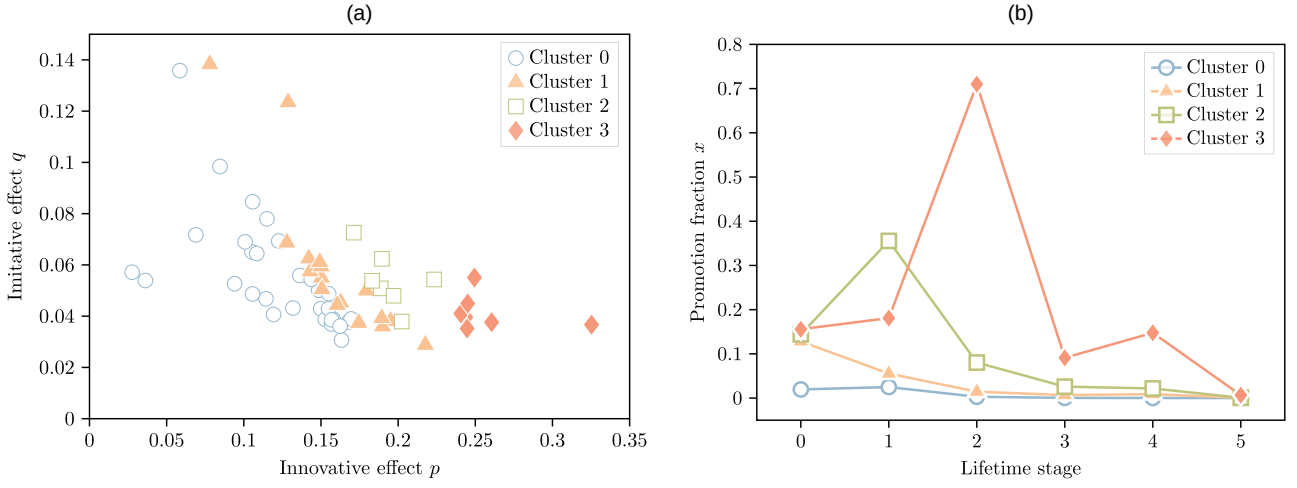
In summary, the AGA policy operates based on two main principles: (i) Promotion and diffusion effects, p and q, positively influence the promotion intensity with the most profound impact during the intermediate lifetime stages. Among these, the promotion effect has a more notable impact. (ii) Videos with a small p but large q mainly receive their promotion in the early stages, in which 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.

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 6.3.1.

6.3.3.1. 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 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 to ignore the diffusion effects when promoting content.

Figure 9. (Color online) Illustration of the AGA Policy Clusters Corresponding to Different Video Configurations



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

- 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.

6.3.3.2. Experiment Result. Table 1 compares the performance of all benchmarks. We draw three key observations from the table. First, AGA consistently performs well, ranking second only when $\overline{C} = 2$. The margin of AGA over others is remarkable. Second, 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. Third, POT performs even worse than NoD in most cases, indicating the drawbacks of the rich-get-richer principle. These observations highlight the importance of both candidate generation and diffusion effects for content promotion and support the effectiveness of our proposed promotion policy.

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 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 data set 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 off-line setting in this paper, in which parameters are estimated beforehand. It can be quite interesting to consider the online version in which 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.

Table 1. Total Adoptions of Different Promotion Policies

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

Note. 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.

Acknowledgments

The authors gratefully acknowledge the constructive feedback from the department editor, Melvyn Sim, the anonymous associate editor and three anonymous reviewers.

References

Agrawal S, Yin S, Zeevi A (2021) Dynamic pricing and learning under the Bassmodel. Bir P, Chawla S, Echenique F, eds. *Proc. 22nd ACM Conf. Econom. Comput.* (ACM, New York), 2–3.

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

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

Bass FM (1969) A new product growth for model consumer durables. *Management Sci.* 15(5):215–227.

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

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

Chen Y, Shi C (2019) Joint pricing and inventory management with strategic customers. *Oper. Res.* 67(6):1610–1627.

Covington P, Adams J, Sargin E (2016) Deep neural networks for YouTube recommendations. Sen S, Geyer W, Freyne J, Castells P, eds. *Proc. 10th ACM Conf. Recommender Systems* (ACM, New York), 191–198.

Davidson J, Liebald B, Liu J, Nandy P, Van Vleet T, Gargi U, Gupta S, et al. (2010) The YouTube video recommendation system. Amatriain X, Torrens M, Resnick P, Zanker M, eds. *Proc. Fourth ACM Conf. Recommender Systems* (ACM, New York), 293–296.

Du C, Cooper WL, Wang Z (2016) Optimal pricing for a multinomial logit choice model with network effects. *Oper. Res.* 64(2):441–455.

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

Feng J, Bhargava HK, Pennock DM (2007) Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS J. Comput.* 19(1):137–148.

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

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

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

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

Granovetter M (1978) Threshold models of collective behavior. *Amer. J. Sociol.* 83(6):1420–1443.

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

Jiang Z, Jain DC (2012) A generalized Norton–Bass model for multigeneration diffusion. *Management Sci.* 58(10):1887–1897.

Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. Getoor L, Senator TE, Domingos PM, Faloutsos C, eds. *Proc. Ninth ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 137–146.

Kiesling E, Günther M, Stummer C, Wakolbinger LM (2012) Agent-based simulation of innovation diffusion: A review. *Central Eur. J. Oper. Res.* 20(2):183–230.

Kitts B, Freed D, Vrieze M (2000) Cross-sell: A fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities. Ramakrishnan R, Stolfo SJ, Bayardo RJ, Parsa I, eds. *Proc. Sixth ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM, New York), 437–446.

Liu J, Dolan P, Pedersen ER (2010) Personalized news recommendation based on click behavior. Rich C, Yang Q, Cavazza M, Zhou MX, eds. *Proc. 15th Internat. Conf. Intelligent User Interfaces* (ACM, New York), 31–40.

Lu W, Chen S, Li K, Lakshmanan LV (2014) Show me the money: Dynamic recommendations for revenue maximization. *Proc. VLDB Endowment* 7(14):1785–1796.

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions I. *Math. Programming* 14(1):265–294.

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

Nosrat F, Cooper WL, Wang Z (2021) Pricing for a product with network effects and mixed logit demand. *Naval Res. Logist.* 68(2):159–182.

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

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

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

Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. *Adv. Artificial Intelligence* 2009:421425:1–421425:19.

Sultan F, Farley JU, Lehmann DR (1990) A meta-analysis of applications of diffusion models. *J. Marketing Res.* 27(1):70–77.

Teng JT, Grover V, Guttler W (2002) Information technology innovations: General diffusion patterns and its relationships to innovation characteristics. *IEEE Trans. Engrg. Management* 49(1):13–27.

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

Vahabi H, Koutsopoulos I, Gullo F, Halkidi M (2015) DifRec: A social-diffusion-aware recommender system. Bailey J, Moffat A, Aggarwal CC, de Rijke M, Kumar R, Murdock V, Sellis TK, Yu JX, eds. *Proc. 24th ACM Internat. Conf. Inform. Knowledge Management* (ACM, New York), 1481–1490.

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

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

Zhang H, Vorobeychik Y (2019) Empirically grounded agent-based models of innovation diffusion: A critical review. *Artificial Intelligence Rev.* 52:707–741.