Influencer Marketing Augmented Personalized Assortment Planning: A Two-stage Optimization Problem

Jing Yuan[1], Twumasi Mensah-Boateng[1], Shaojie Tang [2]

[1] Department of Computer Science and Engineering, University of North Texas
[2]Naveen Jindal School of Management, University of Texas at Dallas
jing.yuan@unt.edu, twumasimensah-boateng@my.unt.edu, shaojie.tang@utdallas.edu

Abstract

Assortment optimization presents a significant challenge for online retail platforms. Its primary objective is to create an optimal selection of products from a vast array of substitutes, which will be displayed to customers with the aim of maximizing expected revenue. The purchase behavior of customers is typically influenced by a choice model that determines the probability of purchasing each product from a given assortment. This paper extends traditional assortment optimization by introducing the integration of influencer marketing, a practice that involves enlisting influencers to promote products and enhance their appeal to customers. While conventional assortment optimization assumes fixed product attractiveness, our model enables platforms to strategically enhance the attractiveness of selected products through influencer marketing, thereby increasing revenue potential. Consequently, we present a novel problem formulation encompassing assortment and influencer marketing planning. Leveraging recent advancements in submodular optimization, we develop effective and efficient solutions for this joint optimization problem.

Introduction

Assortment optimization, which aims to identify a subset of products to maximize expected revenues, has been extensively studied in academic literature. Typically, customer purchase behavior within a given assortment is captured by a choice model, with the Multinomial Logit (MNL) model being widely used in both academia and industry. In the classic MNL model, each product, including the 'no-purchase' option, is assigned an attractiveness or preference weight. When presented with an assortment of products, customers choose their most-preferred item, and the probability of purchasing any product from the assortment increases with its attractiveness. Additionally, the platform earns revenue upon selling a product. It is important to note that customer purchase behavior and choice models can vary among individuals. Traditional assortment optimization involves determining the optimal assortment to offer, maximizing expected revenues.

Most existing studies on traditional assortment optimization assume that the attractiveness of each product is exogenous and fixed. However, in practice, platforms may employ sophisticated marketing strategies such as advertising and lotteries to enhance the attractiveness of certain products and, consequently, increase revenues (Wang, Wang, and Tang 2023). In this paper, we propose augmenting traditional assortment optimization with influencer marketing, a form of social media marketing that involves engaging influencers - individuals with large and engaged audiences or social followings - to promote products (Yuan and Tang 2017; Tang and Yuan 2016; Tang 2018). Influencer marketing effectively enhances brand awareness and product attractiveness by leveraging the reputation influencers have built with their followers. It is worth noting that influencers come in various types based on their follower count or categories, making it crucial for the platform to select the appropriate influencers to promote specific products. We define influencer marketing planning as the assignment of influencers to products, specifying which influencers are hired to promote particular products. Our problem formulation involves joint influencer marketing and assortment planning, with the objective of determining (1) an influencer marketing plan subject to various constraints and (2) a personalized assortment for each customer to maximize expected revenues. While our paper focuses on influencer marketing, our findings can be applied to a wide range of scenarios where product attractiveness can be strategically modified.

We next provide a summary of our key contributions. We extend traditional assortment optimization by introducing a joint influencer marketing and assortment planning problem. Leveraging recent advancements in submodular order optimization, we develop a solution that demonstrates if there exists an α-approximation algorithm for maximizing a constrained submodular order function, then we can find a $\frac{\alpha}{1+\alpha}$-approximation algorithm for our problem, subject to the same constraint for influencer marketing planning. Specifically for two important constraints - the knapsack constraint and the matroid constraint - we devise algorithms with improved approximation ratios.

Related Works

Our work belongs to the category of assortment optimization (Li, Rusmevichientong, and Topaloglu 2015; Davis, Gallego, and Topaloglu 2014; Blanchet, Gallego, and Goyal 2016; Farias, Jagabathula, and Shah 2013; Aouad, Farias, and Levi 2015). Majority of existing studies assume that the attractiveness of each product is fixed, hence, their focus is on selecting a group of products to carry. However, in practise, the platform can use additional levers to change the attractiveness of some products to increase revenues (Wang, Wang, and Tang 2023). For example, (Berbeglia, Flores, and Gallego 2021) study the refined assortment optimization problem for several regular choice models. Their main finding is that under many models, making certain products unattractive to consumers can increase revenues. Our work differs from theirs in that we focus on how to *increase* the attractiveness of some products to increase revenues. Moreover, while they assume that modifying a product's attractiveness is costless, our setting incorporates various types of constraints for IM planning. For example, there is often a budget constraint for hiring influencers. In addition, our model allows for personalization (El Housni and Topaloglu 2021), i.e., we can offer a personalized group of products to each customer.

Our work is built on the recent progress in submodular function maximization and its variants such as streaming submodular maximization (Badanidiyuru et al. 2014). Very recently, (Udwani 2021) introduced a class of stochastic functions which admit a very limited form of submodularity called *submodular order*. They show that this family of functions encompasses monotone submodular functions. They developed a series of approximation algorithms for maximizing a submodular order function subject to various types of constraints such as a matroid constraint and a knapsack constraint. Thanks to these results, we can use their algorithms as subroutines to design our algorithms. Specifically, we can transform our problem to a new optimization problem whose utility function satisfies the property of submodular order. As discussed earlier, our main contribution is to show that if there exists an α-approximation algorithm for maximizing a constrained submodular order function, then we can develop a $\frac{\alpha}{1+\alpha}$-algorithm for our problem subject to the same constraint for IM planning.

Multinomial Logit Model and Problem Formulation

In the rest of this paper, we use $[i]$ to denote the set $\{1, \cdots, i\}$ for any positive integer i. Given a set A and an element a, we use $A + a$ to denote the union $A \cup \{a\}$. Given a function f, we define the marginal value function, $f(X \mid S) = f(X + S) - f(S)$.

IM-Augmented Multinomial Logit Model

We first introduce our choice model, called IM-augmented Multinomial Logit Model. Our model can be viewed as a generalization of the classical multinomial logit model (MNL) (Anderson, De Palma, and Thisse 1992; McFadden et al. 1973) by incorporating the influencer marketing effect. The input is composed of a group of l customer types $\mathcal{L}$, a set of n products $\mathcal{N}$ and a group of m influencers $\mathcal{M}$. To represent an IM planning, we introduce a set of elements $\mathcal{X} \subseteq \mathcal{N} \times \mathcal{M}$. Each element $x = (i, t) \in \mathcal{X}$ defines an IM planning of hiring an influencer $t \in \mathcal{M}$ to promote a product $i \in \mathcal{N}$. For each element $x = (i, t) \in \mathcal{X}$, let $\mathcal{N}(x) = i$ and $\mathcal{M}(x) = t$. Selecting a group of elements $X \subseteq \mathcal{X}$ translates to assigning product $\mathcal{N}(x)$ to influencer $\mathcal{M}(x)$, or hiring influencer $\mathcal{M}(x)$ to promote product $\mathcal{N}(x)$, for all $x \in X$. For example, if $X = \{(1, 1), (2, 1), (2, 2)\}$, then we hire influencer 1 to promote products 1 and 2, and hire influencer 2 to promote product 2.

In the remainder of this paper, for the sake of simplicity, we will use "customer" to refer to the term "customer type." Given any customer $u \in \mathcal{L}$, any IM planning $X \subseteq \mathcal{X}$ and any product $i \in \mathcal{N}$, let $\Delta_u(i, X)$ denote the marginal utility of product i brought by an IM planning X with respect to a customer u. Intuitively, $\Delta_u(i, X)$ quantifies the improvement of the attractiveness of a product i with respect to a customer u due to the presence of an IM planning X. In cases where a product is ill-suited for influencer marketing, we can assign a lower value (or even zero) to $\Delta_u(i, X)$, discouraging the allocation of influencer marketing resources to that particular product type. Without loss of generality, we make the following assumption:

Assumption 0.1 $\Delta_u(i, \emptyset) = 0$ *for all* $i \in \mathcal{N}$ *and* $u \in \mathcal{L}$.

Following the classical MNL model, we assume that the probability $\phi_u(i, S, X)$ that a customer u chooses product i from assortment S given an IM planning X is proportional to $v_{u,i} + \Delta_u(i, X)$ where $v_{u,i}$ is the base attractiveness of a product i to a customer u in the absence of any IM planning. Formally,

$$\phi_u(i, S, X) := \frac{v_{u,i} + \Delta_u(i, X)}{v_0 + \sum_{j \in S}(v_{u,j} + \Delta_u(j, X))} \tag{1}$$

where v_0 represents the weight of the outside option. It is easy to verify that if we set $X = \emptyset$, then our model reduces to the classical MNL model.

We assume that if a customer chooses a product $i \in \mathcal{N}$, then the seller earns r_i from this transaction, where r_i is called the revenue of product i. Hence, the expected revenue $R_u(S, X)$ obtained from a customer u from an assortment S and an IM planning X is

$$R_u(S, X) := \sum_{i \in S} r_i \phi_u(i, S, X). \tag{2}$$

It follows that for a given IM planning X, the largest possible revenue gained from a customer u is given by

$$f_u(X) := \max_{S \subseteq \mathcal{N}} R_u(S, X). \tag{3}$$

Note that solving $\max_{S \subseteq \mathcal{N}} R_u(S, X)$ for a fixed customer u and a fixed IM planning X is a standard assortment optimization problem under MNL. (Talluri and Van Ryzin 2004) showed that the above problem can be solved optimally by simply including all products above a revenue threshold. In other words, once an IM planning X is given, we can easily find out the best assortment for each customer u and compute the value of $f_u(X)$.

Now we are in position to formally introduce our problem. Consider a constraint family over $\mathcal{X}$: $\mathcal{I} \subseteq 2^{\mathcal{X}}$. A set $X \in \mathcal{I}$ is called feasible, while a set $X \notin \mathcal{I}$ is called infeasible. The formal definition of the joint assortment and IM planning problem subject to a constraint $\mathcal{I}$ is listed in **P.0**. Our objective is to find the best feasible IM planning and its associated assortment for each customer to maximize the expected revenue.

P.0 *Maximize*$_{X \subseteq \mathcal{X}}$ $\sum_{u \in \mathcal{L}} f_u(X)$ *subject to* $X \in \mathcal{I}$

Remark: In **P.0**, we make the assumption that all customer types are equally important. However, in reality, it is possible that different customer types may have varying populations. It is worth noting that our model can readily accommodate this heterogeneity by assigning a weight to each customer type. Importantly, incorporating these weights will not impact the results presented in this study.

Assumptions

To facilitate our study, we make two additional mild assumptions in this paper. The first assumption states that promoting one product does not affect the attractiveness of other products.

Assumption 0.2 *For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any group of elements $X \subseteq \mathcal{X}$, and any element x such that $x \notin X$ and $i \neq \mathcal{N}(x)$, we have $\Delta_u(i, X + x) = \Delta_u(i, X)$.*

The second assumption states that the marginal utility brought by IM satisfies the property of diminishing returns, moreover, promoting one product does not decrease its attractiveness.

Assumption 0.3 *For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any two IM plannings $A \subseteq \mathcal{X}, B \subseteq \mathcal{X}$ such that $A \subseteq B$, and any element x such that $x \notin B$ and $i = \mathcal{N}(x)$, we have $\Delta_u(i, A+x) - \Delta_u(i, A) \geq \Delta_u(i, B+x) - \Delta_u(i, B)$. Moreover, $\Delta_u(i, B + x) \geq \Delta_u(i, B)$.*

All missing proofs are moved to our technical report.

Approximate Algorithm for General Constraint

In this section, we present an approximate solution to **P.0**. We first introduce some additional notations which are defined over $\mathcal{X}$. For any IM planning $X \subseteq \mathcal{X}$, any product $i \in \mathcal{N}(X)$ and any customer $u \in \mathcal{L}$, define

$$\phi'_u(i, X) := \frac{\Delta_u(i, X)}{v_0 + \sum_{i \in \mathcal{N}(X)} \Delta_u(i, X)}. \tag{4}$$

Intuitively, $\phi'_u(i, X)$ represents the probability that a customer u chooses a product i from an assortment $\mathcal{N}(X)$, assuming that X is the IM planning and all products have zero base utilities, i.e., $v_{u,i} = 0$ for all $i \in \mathcal{N}$. We next introduce the expected revenue of X from customer u given that $v_{u,i} = 0$ for all $i \in \mathcal{N}$,

$$R'_u(X) := \sum_{i \in \mathcal{N}(X)} r_i \phi'_u(i, X). \tag{5}$$

Consider an IM planning $X \subseteq \mathcal{X}$ and a customer $u \in \mathcal{L}$, assuming $v_{u,i} = 0$ for all $i \in \mathcal{N}$, we can represent the largest expected revenue from u under X as

$$f'_u(X) := \max_{S \subseteq \mathcal{N}(X)} R'_u(X_{\cap S}) \tag{6}$$

where $X_{\cap S} = \{x \in X \mid \mathcal{N}(x) \in S\}$ denotes the set of all elements from X that involve products from S. Again, once X is given, it is easy to compute the value of $f'_u(X)$ since it reduces to the standard assortment optimization problem under MNL.

Before presenting our algorithm to **P.0**, we first introduce a surrogate problem **P.1** whose solution is a key ingredient of our final algorithm. The objective of **P.1** is to find a feasible IM planning Z and its associated assortments for each customer to maximize the expected revenue $\sum_{u \in \mathcal{L}} f'_u(Z)$.

P.1 *Maximize*$_{Z \in \mathcal{I}}$ $\sum_{u \in \mathcal{L}} f'_u(Z)$

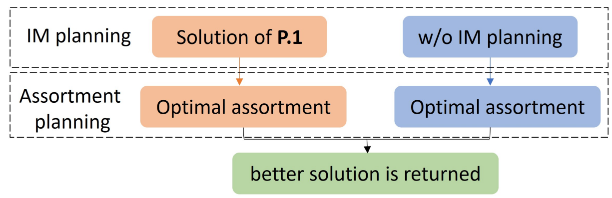

Figure 1: A framework of our final algorithm.

At a high level, our final algorithm to **P.0** consists of two candidate solutions:

1. Candidate Solution 1 (labeled by the orange path in Figure 1): In this approach, we compute an IM plan by solving problem **P.1**. Subsequently, we derive an optimal assortment based on this IM plan.
2. Candidate Solution 2: In contrast, the second candidate solution (represented by the blue path in Figure 1) abstains from incorporating any IM planning. Instead, it computes an optimal assortment under the assumption of zero influence from influencer marketing.

Subsequently, we select the superior solution from these two candidates as our final output. In the following section, we focus on solving **P.1**.

Approximate Solution to P.1

Before explaining our solution to **P.1**, we introduce the concept of *submodular order* and give a few relevant properties from (Udwani 2021).

Definition 0.4 *A permutation σ of elements in $\mathcal{N}$ is a submodular order with respect to a function $h : 2^{\mathcal{N}} \to \mathbb{R}$ if for all sets $B \subseteq A \subseteq \mathcal{N}$ and $C \subseteq \mathcal{N}$ to the right of A, we have $h(C \mid A) \leq h(C \mid B)$.*

Lemma 0.5 *A monotone function h is σ submodular ordered if for any two sets $B \subseteq A$ and an element i to the right of A, we have $h(i \mid A) \leq h(i \mid B)$.*

Lemma 0.6 *Given p monotone functions $h_1, h_2, h_3, \cdots, h_p$ with the same submodular order σ, then function $\sum_{i \in [p]} h_i$ is also monotone with submodular order σ.*

The main result (Corollary 0.10) proved in this section is to show that the order given by sorting elements in $\mathcal{X}$ in non-increasing order of revenues $r_{\mathcal{N}(x)}$ (breaking ties arbitrarily) is a submodular order with respect to function $\sum_{u\in\mathcal{L}} f'_u : 2^{\mathcal{X}} \to \mathbb{R}$. This, together with existing results on maximizing a constrained submodular order function, gives us an approximation algorithm to **P.1**. We next summarize some important results on constrained submodular order function maximization. (Udwani 2021) gives a $0.5-\epsilon$ approximation algorithm for cardinality constrained maximization of a submodular order function. For the case of budget constrains, they propose a local search algorithm that achieves a $0.5-\epsilon$ approximation ratio. They also develop a 0.25 approximation algorithm for maximizing a matroid constrained submodular order function.

Before proving Corollary 0.10, we first present three technical lemmas.

Lemma 0.7 *Given any IM planning $X \subseteq \mathcal{X}$ and any element $x \notin X$, for each customer $u \in \mathcal{L}$, we have*

$$R'_u(x \mid X) \geq 0 \Leftrightarrow r_{\mathcal{N}(x)} \geq R'_u(X)$$

where $R'_u(x \mid X) == R'_u(x+X) - R'_u(X)$.

Proof: For each $X \subseteq \mathcal{X}$, $x \notin X$ and $u \in \mathcal{L}$, define $\Phi_u(x \mid X) = \Delta_u(\mathcal{N}(x), X+x) - \Delta_u(\mathcal{N}(x), X)$. In the rest of the proof, we drop the subscript u if it is clear from the context.

$$\begin{aligned}
&R'(x \mid X) == R'(x+X) - R'(X) \\
&= \sum_{i\in\mathcal{N}(X+x)} \frac{r_i\Delta(i, X+x)}{v_0 + \sum_{j\in\mathcal{N}(X+x)} \Delta(j, X+x)} - R'(X) \\
&= \frac{\sum_{i\in\mathcal{N}(X)} r_i\Delta(i,X) + r_{\mathcal{N}(x)}\Phi(x \mid X)}{v_0 + \sum_{j\in\mathcal{N}(X)} \Delta(j,X) + \Phi(x \mid X)} - R'(X) \\
&= \frac{R'(X)(\sum_{j\in\mathcal{N}(X)} \Delta(j,X) + v_0) + r_{\mathcal{N}(x)}\Phi(x \mid X)}{v_0 + \Phi(x \mid X) + \sum_{j\in\mathcal{N}(X)} \Delta(j,X)} \\
&\qquad -R'(X) \\
&= \frac{-R'(X)\Phi(x \mid X) + r_{\mathcal{N}(x)}\Phi(x \mid X)}{v_0 + \Phi(x \mid X) + \sum_{j\in\mathcal{N}(X)} \Delta(j,X)} \\
&= \Phi(x \mid X)\frac{(-R'(X) + r_{\mathcal{N}(x)})}{v_0 + \Phi(x \mid X) + \sum_{j\in\mathcal{N}(X)} \Delta(j,X)} \qquad (7)
\end{aligned}$$

where the third equality is due to $R'(X) = \sum_{i\in\mathcal{N}(X)} \frac{r_i\Delta(i,X)}{v_0+\sum_{j\in\mathcal{N}(X)}\Delta(j,X)}$. Because $\Phi(x \mid X) \geq 0$ (due to Assumption 0.3) and $v_0 + \Phi(x \mid X) + \sum_{j\in\mathcal{N}(X)} \Delta(j,X) \geq 0$, this implies that $R'(x \mid X) \geq 0 \Leftrightarrow r_{\mathcal{N}(x)} \geq R'(X)$. □

Lemma 0.8 *For any customer $u \in \mathcal{L}$ and any two IM planning $A \subseteq \mathcal{X}, B \subseteq \mathcal{X}$ with $B \subseteq A$, $f'_u(A) \geq f'_u(B)$.*

Proof: We first introduce a new notation $g_u(X)$ for any customer $u \in \mathcal{L}$ and any $X \subseteq \mathcal{X}$,

$$g_u(X) := \max_{Y\subseteq X} R'_u(Y). \qquad (8)$$

We drop the subscript u if it is clear from the context. To prove this lemma, it suffices to show that $f'(X) = g(X)$ for any $X \subseteq \mathcal{X}$. This is because $g(A) \geq g(B)$ for any A, B with $B \subseteq A$. The rest of the proof is devoted to proving $f'(X) = g(X)$ for any $X \subseteq \mathcal{X}$.

Repeatedly applying Lemma 0.7 we have that there exists a $Y^* \in \arg\max_{Y\subseteq X} R'(Y)$ such that $Y^* = \{x \in X \mid r_{\mathcal{N}(x)} \geq \gamma\}$ for some constant γ, i.e., Y^* contains all and only elements from X whose revenue is no less than some constant γ. It follows that

$$\mathcal{N}(Y^*) \cap \mathcal{N}(X \setminus Y^*) = \emptyset. \qquad (9)$$

Recall that $X_{\cap\mathcal{N}(Y^*)} = \{x \in X \mid \mathcal{N}(x) \in \mathcal{N}(Y^*)\}$. (9) implies that $X_{\cap\mathcal{N}(Y^*)} = Y^*$. Hence,

$$R'\Big(X_{\cap\mathcal{N}(Y^*)}\Big) = R'(Y^*). \qquad (10)$$

Now we are ready to show that $f'(X) = g(X)$. First,

$$\begin{aligned}
f'(X) \quad &:= \max_{V\subseteq\mathcal{N}(X)} R'(X_{\cap V}) \geq R'\Big(X_{\cap\mathcal{N}(Y^*)}\Big) \\
&= \quad R'(Y^*) = g(X). \qquad (11)
\end{aligned}$$

The first equality is due to (10) and the second equality is due to the assumption that $Y^* \in \arg\max_{Y\subseteq X} R'(Y)$. Moreover, we have $f'(X) \leq g(X)$ for any $X \subseteq \mathcal{X}$ according to the definition of $f'(\cdot)$. This, together with (11), implies that $f'(X) = g(X)$. □

Lemma 0.9 *The order σ given by sorting elements in $\mathcal{X}$ in non-increasing order of revenues $r_{\mathcal{N}(x)}$ (breaking ties arbitrarily) is a submodular order with respect to function $f'_u : 2^{\mathcal{X}} \to \mathbb{R}$ for any customer $u \in \mathcal{L}$.*

Lemma 0.9, together with Lemma 0.6, implies the following corollary.

Corollary 0.10 *The order σ given by sorting elements in $\mathcal{X}$ in non-increasing order of revenues $r_{\mathcal{N}(x)}$ (breaking ties arbitrarily) is a submodular order with respect to function $\sum_{u\in\mathcal{L}} f'_u : 2^{\mathcal{X}} \to \mathbb{R}$.*

Algorithm Design and Analysis

Now we are ready to present an approximation algorithm ALG to the original problem **P.0**.

Algorithm Design. Suppose there exists an α approximate solution X^{P1} to **P.1**, we build ALG as follows: It randomly picks an IM planning from $\emptyset$ and X^{P1} as the final solution such that $\emptyset$ is picked with probability $\frac{\alpha}{1+\alpha}$ and X^{P1} is picked with probability $\frac{1}{1+\alpha}$. Recall that once the IM planning is picked, we can call the algorithm in (Talluri and Van Ryzin 2004) to find the optimal assortment for each customer.

Performance Analysis. To analyze the performance bound of ALG, we first introduce a lemma that builds a quantitative relation between the optimal solution to **P.1** and the optimal solution to our original problem **P.0**.

Lemma 0.11 *Let $X^* = \max_{X\in\mathcal{I}} \sum_{u\in\mathcal{L}} f_u(X)$ denote the optimal IM planning to **P.0**, and for each customer $u \in \mathcal{L}$, let $S^*_u = \max_{S\subseteq\mathcal{N}} R_u(S, X^*)$ denote the optimal assortment for u under the optimal IM planning X^*, i.e., $\sum_{u\in\mathcal{L}} f_u(X^*) = \sum_{u\in\mathcal{L}} R_u(S^*_u, X^*)$. Then we have*

$$\sum_{u\in\mathcal{L}} f_u(X^*) \leq \sum_{u\in\mathcal{L}} \sum_{i\in S^*_u} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S^*_u} v_{u,j}} + \max_{Z\in\mathcal{I}} \sum_{u\in\mathcal{L}} f'_u(Z)$$

*where $\max_{Z\in\mathcal{I}} \sum_{u\in\mathcal{L}} f'_u(Z)$ is the optimal solution to **P.1**.*

Proof: The following chain proves this lemma,

$$\sum_{u\in\mathcal{L}} f_u(X^*) = \sum_{u\in\mathcal{L}} R_u(S_u^*, X^*)$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \phi_u(i, S_u^*, X^*)$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i} + \Delta_u(i, X^*)}{v_0 + \sum_{j\in S_u^*}(v_{u,j} + \Delta_u(j, X^*))}$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \Big(\frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*}(v_{u,j} + \Delta_u(j, X^*))} + \frac{\Delta_u(i, X^*)}{v_0 + \sum_{j\in S_u^*}(v_{u,j} + \Delta_u(j, X^*))}\Big)$$
$$\leq \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \Big(\frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \frac{\Delta_u(i, X^*)}{v_0 + \sum_{j\in S_u^*} \Delta_u(j, X^*)}\Big)$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{\Delta_u(i, X^*)}{v_0 + \sum_{j\in S_u^*} \Delta_u(j, X^*)}$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*\cap\mathcal{N}(X^*)} r_i \frac{\Delta_u(i, X^*)}{v_0 + \sum_{j\in S_u^*\cap\mathcal{N}(X^*)} \Delta_u(j, X^*)} \quad (12)$$
$$\leq \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \sum_{u\in\mathcal{L}} \max_{S\subseteq\mathcal{N}(X^*)} \sum_{i\in S} r_i \frac{\Delta_u(i, X^*)}{v_0 + \sum_{j\in S} \Delta_u(j, X^*)}$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \sum_{u\in\mathcal{L}} \max_{S\subseteq\mathcal{N}(X^*)} \sum_{i\in S} r_i \frac{\Delta_u(i, X^*_{\cap S})}{v_0 + \sum_{j\in S} \Delta_u(j, X^*_{\cap S})} \quad (13)$$
$$= \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \sum_{u\in\mathcal{L}} f'_u(X^*)$$
$$\leq \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} + \max_{Z\in\mathcal{I}} \sum_{u\in\mathcal{L}} f'_u(Z). \quad (14)$$

Equality (12) is due to the observation that $\Delta_u(j, X^*) = \Delta_u(j, \emptyset) = 0$ for all $u \in \mathcal{L}$ and $j \in S^* \setminus \mathcal{N}(X^*)$ where the first equality is due to Assumption 0.2 and the second equality is due to Assumption 0.1, equality (13) is due to Assumption 0.2, and equality (14) is due to $X^* \in \mathcal{I}$. □

Now we are in position to present the main theorem of this paper.

Theorem 0.12 *Let $X^* = \max_{X\in\mathcal{I}} \sum_{u\in\mathcal{L}} f_u(X)$ denote the optimal IM planning to* ***P.0****, the expected revenue of* ALG *is at least $\frac{\alpha}{1+\alpha}\sum_{u\in\mathcal{L}} f_u(X^*)$.*

Proof: Recall that ALG picks $\emptyset$ with probability $\frac{\alpha}{1+\alpha}$ and picks X^{P1} with probability $\frac{1}{1+\alpha}$. Hence, the expected revenue of ALG is $\frac{\alpha}{1+\alpha}\sum_{u\in\mathcal{L}} f_u(\emptyset) + \frac{1}{1+\alpha}\sum_{u\in\mathcal{L}} f_u(X^{P1})$. Moreover, because $\sum_{u\in\mathcal{L}} f_u(X^*) \leq \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0+\sum_{j\in S_u^*} v_{u,j}} + \max_{Z\in\mathcal{I}}\sum_{u\in\mathcal{L}} f'_u(Z)$ (Lemma 0.11), to prove this theorem, it suffices to show that

$$\sum_{u\in\mathcal{L}} f_u(\emptyset) \geq \sum_{u\in\mathcal{L}}\sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}} \quad (15)$$

and

$$\sum_{u\in\mathcal{L}} f_u(X^{P1}) \geq \alpha \max_{Z\in\mathcal{I}} \sum_{u\in\mathcal{L}} f'_u(Z). \quad (16)$$

The proof of (15) is trivial,

$$f_u(\emptyset) := \max_{S\subseteq\mathcal{N}} \sum_{i\in S} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S} v_{u,j}} \geq \sum_{i\in S_u^*} r_i \frac{v_{u,i}}{v_0 + \sum_{j\in S_u^*} v_{u,j}}.$$

We next prove (16). Let $S' \in \arg\max_{S\subseteq\mathcal{N}(X^{P1})} \sum_{i\in S} r_i \frac{\Delta_u(i, X^{P1})}{v_0+\sum_{j\in S}\Delta_u(j, X^{P1})}$ denote the optimal assortment for customer u, assuming the IM planning is X^{P1} and the base utilities of all products are zero. Assume $|S'| = m$, index products in S' in on-increasing order of prices (breaking ties arbitrarily): $r_1 \geq r_2 \cdots \geq r_m$. By the definition of S' and Lemma 0.7, we have $r_m \geq f'_u(X^{P1})$. Hence,

$$f'_u(X^{P1}) := \sum_{i\in[m]} r_i \frac{\Delta_u(i, X^{P1})}{v_0 + \sum_{j\in[m]} \Delta_u(j, X^{P1})} \leq \frac{\sum_{i\in[m]} r_i \Delta_u(i, X^{P1}) + r_m v_{u,m}}{v_0 + \sum_{j\in[m]} \Delta_u(j, X^{P1}) + v_{u,m}} \quad (17)$$

where the inequality is due to $r_m \geq f'_u(X^{P1})$. Because the right hand side (RHS) of (17) is upper bounded by r_m and $r_m \leq r_{m-1}$, we have RHS of (17) $\leq r_{m-1}$. Hence,

$$f'_u(X^{P1}) \leq \text{ RHS of (17)} \leq \frac{\sum_{i\in[m]} r_i \Delta_u(i, X^{P1}) + r_m v_{u,m} + r_{m-1} v_{u,m-1}}{v_0 + \sum_{j\in[m]} \Delta_u(j, X^{P1}) + v_{u,m} + v_{u,m-1}}$$

where the second inequality is due to RHS of (17) $\leq r_{m-1}$. Repeatedly applying the same argument to the remaining products $m-2, m-3, \cdots, 1$, we have

$$f'_u(X^{P1}) \leq \sum_{i\in[m]} r_i \frac{v_{u,i} + \Delta_u(i, X^{P1})}{v_0 + \sum_{j\in[m]}(v_{u,j} + \Delta_u(j, X^{P1}))} == f_u(X^{P1}). \quad (18)$$

This combined with the assumption that $\sum_{u\in\mathcal{L}} f'_u(X^{P1}) \geq \alpha\max_{Z\in\mathcal{I}}\sum_{u\in\mathcal{L}} f'_u(Z)$ implies (16). □

Note that the approximation ratio derived in Theorem 0.12 is dependent on α whose value is constraint-specific. As discussed earlier, there exists an $\alpha = 0.5 - \epsilon$ (resp. an $\alpha = 0.25$) approximation algorithm for knapsack constrained (resp. matroid constrained) submodular order maximization. This, together with Theorem 0.12, implies that ALG achieves a $\frac{0.5-\epsilon}{1.5-\epsilon} \approx 0.33 - \epsilon$ (resp. $\frac{0.25}{1.25} = 0.2$) approximation for our original problem **P.0** subject to a knapsack constraint (resp. a matroid constraint).

Enhanced Algorithms for Knapsack and Matroid Constraints

Apart from the framework developed in the previous section for general constraints, in this section, we show that it is possible to design better algorithms for two special but important constraints: knapsack and matroid constraints. Specifically, we improve the approximation ratio from $0.33 - \epsilon$ (resp. 0.2) to $0.5 - \epsilon$ (resp. 0.25) for the knapsack constraint (resp. the matroid constraint).

Problem Formulation and Notations

We first study the knapsack constrained **P.0**, then apply a similar approach to solving the case of matroid constraints. Assume assigning an influencer $t \in \mathcal{M}$ to a product $i \in \mathcal{N}$ has a fixed cost $c((t, i))$, the total cost of an IM planning $X \subseteq \mathcal{X}$ is $c(X) = \sum_{x \in X} c(x)$. The objective of the knapsack constrained **P.0** is to find the best of group influencers subject to a knapsack constraint B.

(knapsack constrained) P.0
Maximize$_{X \subseteq \mathcal{X}}$ $\sum_{u \in \mathcal{L}} f_u(X)$ *subject to* $c(X) \leq B$

To facilitate our algorithm design, we introduce a *super* influencer s such that the cost of hiring s is zero, i.e., $c((s, i)) = 0$ for all $i \in \mathcal{N}$. Let $\mathcal{X}^{\mathsf{E}} = \mathcal{X} \cup \{(s, i) \mid i \in \mathcal{N}\}$ denote the expanded set of elements by including possible assignments of the super influencer s. For each product $i \in \mathcal{N}$ and each expanded IM planning $Y \subseteq \mathcal{X}^{\mathsf{E}}$, let $\mathbf{1}(i, Y)$ be an indicator that $(s, i) \in Y$. Let $Y_{-s} = Y \setminus \cup_{i \in \mathcal{N}} \{(s, i)\}$ denote a subset of Y excluding all assignments of the super influencer s. It will become clear later that we use s to capture the base attractiveness of each product without influencer marketing. For each customer $u \in \mathcal{L}$, we can represent the attractiveness of a product $i \in \mathcal{N}$ under an expanded IM planning $Y \subseteq \mathcal{X}^{\mathsf{E}}$ as:

$$\Delta_u^{\mathsf{E}}(i, Y) := \mathbf{1}(i, Y) \times v_{u,i} + \Delta_u(i, Y_{-s}).$$

By abuse of notation, for each customer $u \in \mathcal{L}$, each product $i \in \mathcal{N}$ and each expanded IM planning $Y \subseteq \mathcal{X}^{\mathsf{E}}$, define

$$\phi'_u(i, Y) := \frac{\Delta_u^{\mathsf{E}}(i, Y)}{v_0 + \sum_{i \in \mathcal{N}(Y)} \Delta_u^{\mathsf{E}}(i, Y)},$$

$$R'_u(Y) := \sum_{i \in \mathcal{N}(Y)} r_i \phi'_u(i, Y), f'_u(Y) := \max_{S \subseteq \mathcal{N}(Y)} R'_u(Y_{\cap S}).$$

Intuitively, $\phi'_u(i, Y)$ is the probability that a product i is selected by a customer u under an expanded IM planning Y and assortment $\mathcal{N}(Y)$; $R'_u(Y)$ is the expected revenue from a customer u under an expanded IM planning Y and an assortment $\mathcal{N}(Y)$; $f'_u(Y)$ is the revenue of the best assortment from customer u under an expanded IM planning Y. We next introduce a new optimization problem **P.2** subject to a knapsack constraint B. The objective of **P.2** is to find the best IM planning from $\mathcal{X}^{\mathsf{E}}$ to maximize $\sum_{u \in \mathcal{L}} f'_u(\cdot)$.

P.2 *Maximize*$_{Z \subseteq \mathcal{X}^{\mathsf{E}}}$ $\sum_{u \in \mathcal{L}} f'_u(Z)$ subject to $c(Z) \leq B$

Approximate Solution to P.2

We next show that **P.2** is a knapsack constrained submodular order maximization problem. Hence, there exists an $0.5 - \epsilon$ approximation algorithm to this problem.

Lemma 0.13 *The order σ given by sorting elements in $\mathcal{X}^{\mathsf{E}}$ in non-increasing order of price $r_{\mathcal{N}(x)}$ (breaking ties arbitrarily) is a submodular order with respect to function $\sum_{u \in \mathcal{L}} f'_u : 2^{\mathcal{X}^{\mathsf{E}}} \to \mathbb{R}$.*

To prove this lemma, it suffices to show that for each $u \in \mathcal{L}$, the function $\Delta_u^{\mathsf{E}}(\cdot, \cdot)$ satisfies all properties specified in Assumptions 0.1, 0.2 and 0.3 in the "expanded" world. This is because if all assumptions apply to $\Delta_u^{\mathsf{E}}(\cdot, \cdot)$, one can follow the proof of Lemma 0.9 to prove Lemma 0.13. The proof of Assumption 0.1 is trivial: for all $i \in \mathcal{N}$ and $u \in \mathcal{L}$, we have $\Delta_u^{\mathsf{E}}(i, \emptyset) := \mathbf{1}(i, \emptyset) \times v_{u,i} + \Delta_u(i, \emptyset_{-s}) = 0$. In the following proposition, we prove that $\Delta_u^{\mathsf{E}}(\cdot, \cdot)$ satisfies all properties specified in Assumptions 0.2 and 0.3.

Proposition 0.14 *For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any expanded IM planning $Y \subseteq \mathcal{X}^{\mathsf{E}}$, and any element $a \in \mathcal{X}^{\mathsf{E}}$ such that $a \notin Y$ and $i \neq \mathcal{N}(a)$,*

$$\Delta_u^{\mathsf{E}}(i, Y + a) = \Delta_u^{\mathsf{E}}(i, Y). \tag{19}$$

For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any two groups of elements $A \subseteq \mathcal{X}^{\mathsf{E}}, B \subseteq \mathcal{X}^{\mathsf{E}}$ such that $A \subseteq B$, and any element a such that $a \notin B$ and $i = \mathcal{N}(a)$,

$$\Delta_u^{\mathsf{E}}(i, A + a) - \Delta_u^{\mathsf{E}}(i, A) \geq \Delta_u^{\mathsf{E}}(i, B + a) - \Delta_u^{\mathsf{E}}(i, B). \tag{20}$$

Proof: We first prove (19). For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any group of elements $Y \subseteq \mathcal{X}^{\mathsf{E}}$, and any element $a \in \mathcal{X}^{\mathsf{E}}$ such that $y \notin Y$ and $i \neq \mathcal{N}(a)$, we have

$$\Delta_u^{\mathsf{E}}(i, Y + a) = \mathbf{1}(i, Y + a) \times v_{u,i} + \Delta_u(i, (Y + a)_{-s}). \tag{21}$$

We prove (19) in two cases.

- The case when $(s, i) \in Y$: Because $i \neq \mathcal{N}(a)$, we have $i \neq \mathcal{N}(\{a\}_{-s})$. This, together with Assumption 0.2, implies that $\Delta_u(i, (Y_{-s} + \{a\}_{-s})) = \Delta_u(i, Y_{-s})$. Hence,

$$\Delta_u(i, (Y + a)_{-s}) = \Delta_u(i, (Y_{-s} + \{a\}_{-s})) = \Delta_u(i, Y_{-s}). \tag{22}$$

  Moreover, if $(s, i) \in Y$, then $\mathbf{1}(i, Y) = \mathbf{1}(i, Y + a) = 1$. Hence, $\Delta_u^{\mathsf{E}}(i, Y + a) == \mathbf{1}(i, Y + a) \times v_{u,i} + \Delta_u(i, (Y + a)_{-s}) = v_{u,i} + \Delta_u(i, (Y + a)_{-s}) = v_{u,i} + \Delta_u(i, (Y)_{-s}) = \mathbf{1}(i, Y) \times v_{u,i} + \Delta_u(i, (Y)_{-s}) == \Delta_u^{\mathsf{E}}(i, Y)$.
- The case when $(s, i) \notin Y$: If $(s, i) \notin Y$, then because $i \neq \mathcal{N}(a)$, we have $\mathbf{1}(i, Y) = \mathbf{1}(i, Y + a) = 0$. Moreover, because $i \neq \mathcal{N}(a)$, (22) still holds. Hence, $\Delta_u^{\mathsf{E}}(i, Y + a) == \mathbf{1}(i, Y + a) \times v_{u,i} + \Delta_u(i, (Y + a)_{-s}) = \Delta_u(i, (Y + a)_{-s}) = \Delta_u(i, (Y)_{-s}) = \mathbf{1}(i, Y) \times v_{u,i} + \Delta_u(i, (Y)_{-s}) == \Delta_u^{\mathsf{E}}(i, Y)$.

This finishes the proof of (19). We next prove (20). For any customer $u \in \mathcal{L}$, any product $i \in \mathcal{N}$, any two groups of elements $A \subseteq \mathcal{X}^{\mathsf{E}}, B \subseteq \mathcal{X}^{\mathsf{E}}$ such that $A \subseteq B$, and any element a such that $a \notin B$ and $i = \mathcal{N}(a)$,

$$\begin{aligned}
&\Delta_u^{\mathsf{E}}(i, A+a) - \Delta_u^{\mathsf{E}}(i, A) \\
&= \big(\mathbf{1}(i, A+a) \times v_{u,i} + \Delta_u(i, (A+a)_{-s})\big) \\
&\qquad - \big(\mathbf{1}(i, A) \times v_{u,i} + \Delta_u(i, A_{-s})\big) \\
&= \big(\mathbf{1}(i, A+a) - \mathbf{1}(i, A)\big) \times v_{u,i} \\
&\qquad + \big(\Delta_u(i, (A+a)_{-s}) - \Delta_u(i, A_{-s})\big). \quad (23)
\end{aligned}$$

$$\begin{aligned}
&\Delta_u^{\mathsf{E}}(i, B+a) - \Delta_u^{\mathsf{E}}(i, B) \\
&= \big(\mathbf{1}(i, B+a) \times v_{u,i} + \Delta_u(i, (B+a)_{-s})\big) \\
&\qquad - \big(\mathbf{1}(i, B) \times v_{u,i} + \Delta_u(i, B_{-s})\big) \\
&= \big(\mathbf{1}(i, B+a) - \mathbf{1}(i, B)\big) \times v_{u,i} \\
&\qquad + \big(\Delta_u(i, (B+a)_{-s}) - \Delta_u(i, B_{-s})\big). \quad (24)
\end{aligned}$$

We prove (20) in three cases.

- The case when $(s, i) \in A$: If $(s, i) \in A$, then $(s, i) \in A + a$, and because $A \subseteq B$, we have $(s, i) \in B$, $(s, i) \in B + a$. Hence,

$$\mathbf{1}(i, A) = \mathbf{1}(i, A+a) = \mathbf{1}(i, B) = \mathbf{1}(i, B+a) = 1. \quad (25)$$

  Moreover,

$$\begin{aligned}
&\Delta_u(i, (A+a)_{-s}) - \Delta_u(i, A_{-s}) \\
&= \Delta_u(i, A_{-s} + a_{-s}) - \Delta_u(i, A_{-s}) \\
&\geq \Delta_u(i, B_{-s} + a_{-s}) - \Delta_u(i, B_{-s}) \\
&= \Delta_u(i, (B+a)_{-s}) - \Delta_u(i, B_{-s}) \quad (26)
\end{aligned}$$

  where the inequality is due to $A_{-s} \subseteq B_{-s}$ and Assumption 0.3. (23),(24),(25) and (26) together imply that $\Delta_u^{\mathsf{E}}(i, A+a) - \Delta_u^{\mathsf{E}}(i, A) \geq \Delta_u^{\mathsf{E}}(i, B+a) - \Delta_u^{\mathsf{E}}(i, B)$.
- The case when $(s, i) \notin A$ and $(s, i) \in B$: If $(s, i) \in B$, then $\mathbf{1}(i, B+a) - \mathbf{1}(i, B) = 0$. Hence, $\mathbf{1}(i, A+a) - \mathbf{1}(i, A) \geq \mathbf{1}(i, B+a) - \mathbf{1}(i, B)$. Moreover, because (26) still holds, (23) and (24) imply that $\Delta_u^{\mathsf{E}}(i, A+a) - \Delta_u^{\mathsf{E}}(i, A) \geq \Delta_u^{\mathsf{E}}(i, B+a) - \Delta_u^{\mathsf{E}}(i, B)$.
- The case when $(s, i) \notin A$ and $(s, i) \notin B$: We consider two subcases. If $(s, i) \notin A$, $(s, i) \notin B$ and $a = (s, i)$, then $\mathbf{1}(i, A) = \mathbf{1}(i, A+a) = \mathbf{1}(i, B) = \mathbf{1}(i, B+a) = 1$. Otherwise, if $(s, i) \notin A$, $(s, i) \notin B$ and $a \neq (s, i)$, then $\mathbf{1}(i, A) = \mathbf{1}(i, A+a) = \mathbf{1}(i, B) = \mathbf{1}(i, B+a) = 0$. Hence, $\mathbf{1}(i, A+a) - \mathbf{1}(i, A) = \mathbf{1}(i, B+a) - \mathbf{1}(i, B)$. Moreover, because (26) still holds, (23) and (24) imply that $\Delta_u^{\mathsf{E}}(i, A+a) - \Delta_u^{\mathsf{E}}(i, A) \geq \Delta_u^{\mathsf{E}}(i, B+a) - \Delta_u^{\mathsf{E}}(i, B)$.

This finishes the proof of (20). □

The next lemma shows that the optimal solution of **P.2** is no less than the optimal solution of the knapsack constrained **P.0**.

Lemma 0.15 *Let $X^* = \max_{X: c(X) \leq B} \sum_{u \in \mathcal{L}} f_u(X)$ denote the optimal IM planning to the knapsack constrained **P.0**, and for each customer $u \in \mathcal{L}$, let $S_u^* = \max_{S \subseteq \mathcal{N}} R_u(S, X^*)$ denote the optimal assortment for u under the optimal IM planning X^*, i.e., $\sum_{u \in \mathcal{L}} f_u(X^*) = \sum_{u \in \mathcal{L}} R_u(S_u^*, X^*)$. Then we have*

$$\sum_{u \in \mathcal{L}} f_u(X^*) \leq \max_{Z \subseteq \mathcal{X}^{\mathsf{E}}; c(Z) \leq B} \sum_{u \in \mathcal{L}} f'_u(Z) \quad (27)$$

*where $\max_{Z \subseteq \mathcal{X}^{\mathsf{E}}; c(Z) \leq B} \sum_{u \in \mathcal{L}} f'_u(Z)$ represents the optimal solution to **P.2**.*

Algorithm Design and Analysis

Algorithm Design. Now we are ready to present an approximation algorithm to the original knapsack constrained **P.0**. We first apply a $0.5 - \epsilon$ approximation (Algorithm 5 in (Udwani 2021)) to solve **P.2** and obtain a solution $Y^{\text{P2-approx}}$, then $Y_{-s}^{\text{P2-approx}}$ is returned as the final solution.

Performance Analysis. In the following theorem, we show that $Y_{-s}^{\text{P2-approx}}$ achieves an approximation ratio of $0.5 - \epsilon$ for the knapsack constrained **P.0**.

Theorem 0.16 *Assume $Y^{\text{P2-approx}}$ is a $0.5-\epsilon$ approximation solution for **P.2** and X^* is the optimal IM planning to the knapsack constrained **P.0**,*

$$\sum_{u \in \mathcal{L}} f_u(Y_{-s}^{\text{P2-approx}}) \geq (0.5 - \epsilon) \sum_{u \in \mathcal{L}} f_u(X^*). \quad (28)$$

Matroid Constraints

In this section, we develop enhanced algorithms for matroid constraints. Note that the cardinality constraint is an important special case of matroid constraint. Formally, a matroid is a pair $(\mathcal{N}, \mathcal{I})$, where $\mathcal{N}$ is a ground set, and $\mathcal{I}$ is a collection of subsets of $\mathcal{N}$. The collection $\mathcal{I}$ must satisfy the following three properties: $\mathcal{I} \neq \emptyset$; If $A \subseteq B \in \mathcal{I}$, then $A \in \mathcal{I}$; If $A, B \in \mathcal{I}$ and $|A| < |B|$, then there is an element $u \in B \setminus A$ for which $A + u \in \mathcal{I}$. Given a matroid $(\mathcal{N}, \mathcal{I})$, we first formalize the **(matroid constrained) P.0**, where one wants to find an IM planning X to maximize the revenue given that $X \in \mathcal{I}$, as follows:

(matroid constrained) P.0
Maximize$_{X \subseteq \mathcal{X}}$ $\sum_{u \in \mathcal{L}} f_u(X)$ *subject to* $X \in \mathcal{I}$

Similar to the approach used to solve the **(knapsack constrained) P.0**, we first introduce a *super* influencer s. In this section, we utilize all the notations introduced in the previous section. In addition, we introduce an expanded matroid $(\mathcal{X}^{\mathsf{E}}, \mathcal{I}^{\mathsf{E}})$ which is defined over the expanded set $\mathcal{X}^{\mathsf{E}} = \mathcal{X} \cup \{(s, i) \mid i \in \mathcal{N}\}$ such that $Y \in \mathcal{I}^{\mathsf{E}}$ if and only if $Y_{-s} \in \mathcal{I}$.

We next introduce a new optimization problem **P.3** subject to an expanded matroid $(\mathcal{N}^{\mathsf{E}}, \mathcal{I}^{\mathsf{E}})$. The objective of **P.3** is to find the best expanded IM planning to maximize $\sum_{u \in \mathcal{L}} f'_u(\cdot)$.

P.3 *Maximize*$_{Z \subseteq \mathcal{X}^{\mathsf{E}}}$ $\sum_{u \in \mathcal{L}} f'_u(Z)$ subject to $Z \in \mathcal{I}^{\mathsf{E}}$

It is easy to verify that Lemma 0.13 still holds, hence, **P.3** is a matroid constrained submodular order maximization problem which admits a 0.25 approximation. We next describe our solution to the **(matroid constrained) P.0**. We first apply a 0.25 approximation (Algorithm 7 in (Udwani 2021)) to solve **P.3** and obtain a solution $Y^{\text{P3-approx}}$, then $Y^{\text{P3-approx}}_{-s}$ is returned as the final solution. The proof of the following theorem is same as that of Theorem 0.16 except that $0.5-\epsilon$ is replaced by 0.25.

Theorem 0.17 *Assume $Y^{\text{P3-approx}}$ is a 0.25 approximation solution for **P.3** and X^* is the optimal IM planning to the **(matroid constrained) P.0**,*

$$\sum_{u\in\mathcal{L}} f_u(Y^{\text{P3-approx}}_{-s}) \geq 0.25 \sum_{u\in\mathcal{L}} f_u(X^*). \tag{29}$$

Performance Evaluation

We conduct experiments to evaluate the performance of our proposed algorithms for augmented personalized assortment planning on both synthetic data and real-world datasets. The performance of considered algorithms is evaluated in terms of their expected revenue with respect to changes in customer purchase behavior, influence probability and various constraints. All algorithms are implemented using Java and all experiments are run on a Linux server with Intel Xeon 2.40GHz CPU and 128GB memory.

Experimental Setup. We follow the widely used methods in (Gao et al. 2020) to obtain the experimental data. The revenue of each product, r_i, is uniformly sampled from $[1, 10]$. For $v_{u,i}$, the preference weight of customer (type) u for product i, we first sample $\eta_{u,i}$ from the uniform distribution $[0, 1]$, then set $v_{u,i} = \eta_{u,i}/\delta_u$, where $\delta_u = P_0 \sum_i v_{u,i}/(1-P_0)$. Here P_0 represents the weight of the outside (no-purchase) option. It is easy to verify that if we display all products to a customer, the no-purchase probability for the customer is $1/(1+\sum_{i\in[n]} v_{u,i}) = 1/(1 + (1-P_0)/P_0) = P_0$. We vary the value of P_0 in our experiments to evaluate the impact of the outside option associated with customer purchase behavior on the performance of the algorithms. The cost of hiring an influencer t to promote a product i is uniformly sampled from $[1, 2]$.

Hotel Recommendations. To evaluate the performance of our proposed algorithms on real world applications, we run experiments on the benchmark hotel transaction dataset from Expedia.com [1]. The travel booking platform recommends a list of hotels for each hotel search request from the customers. This dataset is widely used for predicting which hotel a customer is going to book, given a list of recommendations. We preprocess the dataset so that the extracted dataset contains the recommendation lists of 4922 search requests, where 3050 requests finally led to a choice and the customers of the remaining 1872 requests left without a purchase. For each search request, 2 to 30 hotels were recommended. There are $N = 30$ unique hotels in the extracted dataset and we use a 6-dimension feature vector, x_i, to represent each hotel. We partition the customers into different types based on their purchasing power, measured by the mean price of the recommended hotels of their search requests. In our experiments, we consider four customer types that correspond to mean price range (in US dollars) of $[50, 151)$, $[151, 155)$, $[155, 157)$, and $[157, 300)$, respectively. The distribution over these customer types is $(0.278, 0.338, 0.134, 0.250)$. We learn an MNL choice model on the extracted dataset using the standard Maximum Likelihood Estimation (MLE) method to capture the choice behavior of each customer type.

We adopt a large-scale benchmark social network, Slashdot, to obtain our influence probability data. Slashdot is a technology-related news website, containing $81,867$ nodes and $545,671$ edges. Each node represents a user and each edge represents a friend relationship. We adopt a commonly used approach (Jung, Heo, and Chen 2012) to set the influence probability of each neighbor on a user v as $1/d_v$, where d_v denotes the in-degree of v. We derive the distribution of influence probability over all users in the social network. The influence probability of influencer t on customer u is then sampled from this distribution. The cost of hiring an influencer t is set as $\sqrt[3]{d'_t}$ where d'_t is the out-degree of t in the social network.

Algorithms. We implement and evaluate our proposed algorithms against several baselines. For cardinality constrained augmented assortment planning, we implement our algorithm, and call it ALGC. It includes a subroutine of finding the IM planning and finding the optimal assortment for each customer given the planning. We also implement two algorithms TPC and RDMC as our benchmarks. TPC stands for Top Pairs with Cardinality constraint. It first sorts the elements in $\mathcal{X}$ in non-increasing order of revenues (breaking ties arbitrarily). Then it selects the top k elements as the IM planning. Given this planning, it calls the standard assortment optimization algorithm to finalize the assortment for each customer. RDMC stands for RanDoM pairs with Cardinality constraint. It randomly selects k elements from $\mathcal{X}$ as the IM planning, and then find the assortment for each customer using the standard assortment optimization algorithm. We also show in Figure 2-5 the expected revenue achieved by the standard assortment optimization without IM planning, labeled as MNL, standing for MultiNomial Logit.

For budget constrained augmented assortment planning, we implement our algorithm, and call it ALGK. We also implement two algorithms TPK and RDMK as our benchmarks. TPK stands for Top Pairs with Knapsack constraint. It first sorts the elements in $\mathcal{X}$ in non-increasing order of revenues (breaking ties arbitrarily). Then it scans the sorted list and for each element, it selects the element if selecting this element does not violate the budget constraint. Given this planning, it calls the standard assortment optimization algorithm to finalize the assortment for each customer. RDMK stands for RanDoM pairs with Knapsack constraint. It randomly selects elements from $\mathcal{X}$ until the budget is exhausted, and then based on this IM planning, it finds the optimal assortment for each customer.

Parameter Settings. For our algorithms ALGC and ALGK, we set the parameter $\epsilon = 0.01$ for obtaining the IM planning. In our experiments on the synthetic data, we first consider 10 customer types and 5 influencers, and set the default size limit of the assortment to be 5. We vary the value

[1] https://www.kaggle.com/competitions/expedia-hotel-recommendations/data

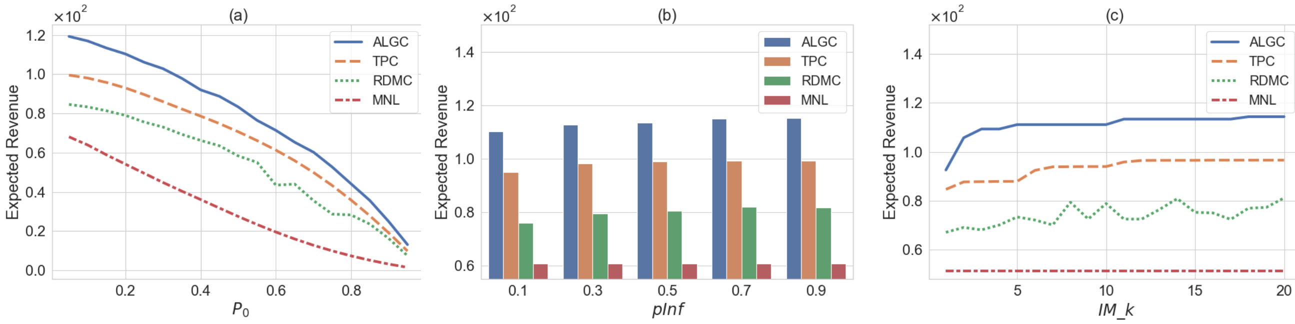

Figure 2: ALGC achieves superior expected revenue for cardinality constrained augmented assortment planning (20 products).

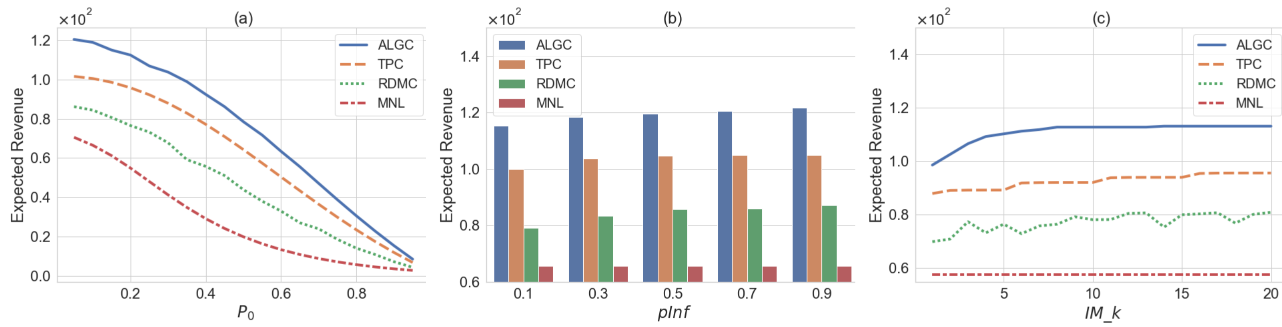

Figure 3: ALGC achieves superior expected revenue for cardinality constrained augmented assortment planning (30 products).

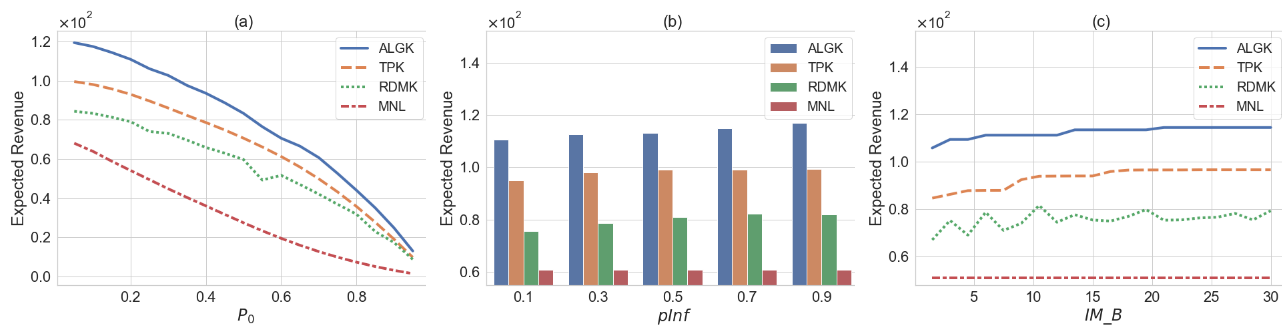

Figure 4: ALGK achieves superior expected revenue for knapsack constrained augmented assortment planning (20 products).

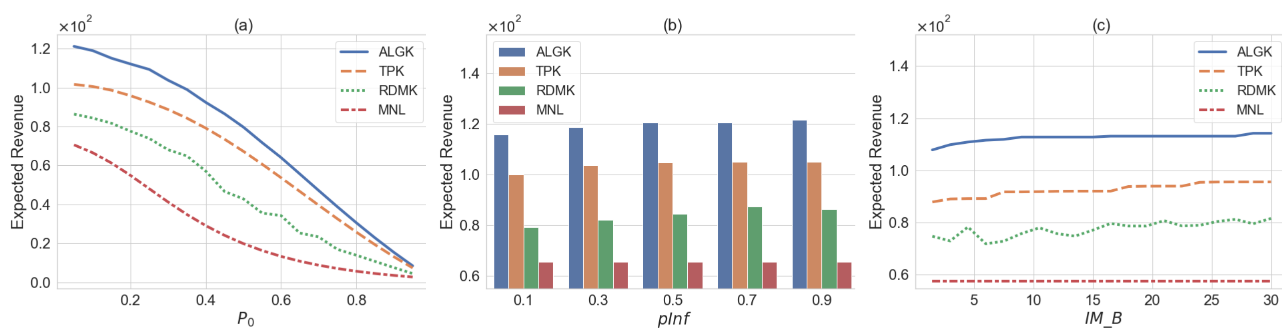

Figure 5: ALGK achieves superior expected revenue for knapsack constrained augmented assortment planning (30 products).

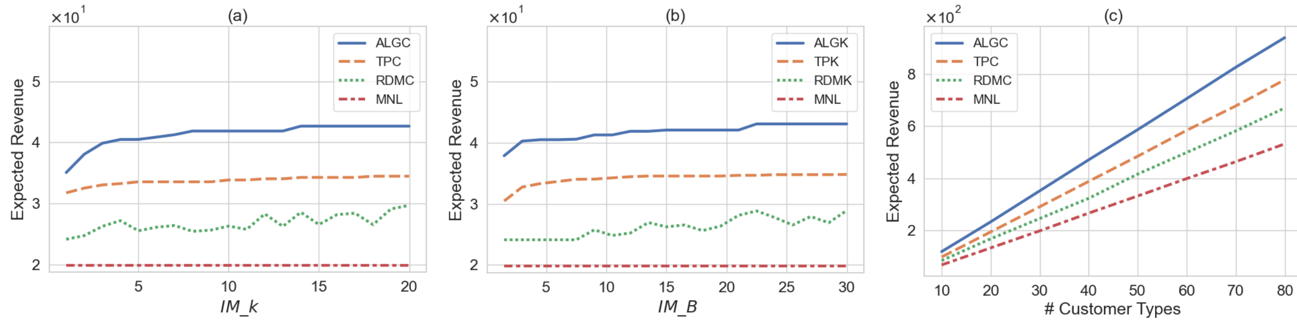

Figure 6: (a)-(b) ALGC and ALGK achieve superior expected revenue for augmented assortment planning on real dataset. (c) ALGC achieves superior expected revenue for augmented assortment planning under varying number of customer types.

of no-purchase probability P_0, the influence probability of influencers *pInf*, number of products, cardinality constraint *IM_k*, and budget *IM_B* to evaluate the performance of the algorithms under different parameter settings. For each setting, we run the experiments for 1000 rounds and the average results are presented as below. We also perform a scalability test on synthetic data by increasing the number of influencers and customer types, and illustrate their impact on the expected revenue as well as the running time.

Experimental Results. We measure the performance of the algorithms in terms of the expected revenue yielded. The results on synthetic data are shown in Figure 2-5. Figure 2 and 3 show the results for cardinality constrained augmented assortment planning under various parameter settings that consider 20 and 30 products, respectively. Figure 4 and 5 present the results for budget constrained augmented assortment planning under various parameter settings that consider 20 and 30 products, respectively. The results on real datasets are illustrated in Figure 6(a)-(b). Figure 6(c)-7 and Table 1-2 show the results of the scalability test.

As shown in Figure 2(a), the expected revenue generated by the algorithms decreases as P_0, the weight of the no-purchase option, increases. The reason is that a larger P_0 indicates a lower probability of the customers buying products from the assortment. We observe that ALGC outperforms the benchmarks by at least 20% in terms of the expected revenue. This result confirms the effectiveness of our proposed algorithm. We also observe that all algorithms with IM planning outperform MNL, the standard assortment optimization without IM planning. This validates the benefit of augmenting the traditional assortment optimization with influencer marketing in terms of increasing the expected revenue.

Figure 2(b) shows the performance of the algorithms with respect to changes in the influence probability of the influencers. We use $p_{t,u,i}$ to denote the influence probability with respect to customer u if influencer t is hired to promote product i. This probability has a direct impact on the amount of improvement of the attractiveness of a product i with respect to a customer u due to the presence of influencer t's promotion. For each tuple (t, u, i), the value of $p_{t,u,i}$ is randomly sampled from $\{0.1, pInf\}$. For example, for *pInf* = 0.5, the value of $p_{t,u,i}$ is randomly sampled from $\{0.1, 0.5\}$. As

# Influencer	# Customer Types				
	10	20	30	50	100
10	0.272	0.494	0.680	1.012	1.842
20	0.422	0.896	1.132	1.942	4.085
30	0.704	1.282	1.994	2.786	5.958
50	1.155	1.855	3.320	4.556	10.416
100	2.174	3.797	6.689	11.543	19.980

Table 1: Running Time for ALGC Algorithm (in seconds).

shown in Figure 2(b), we observe that the expected revenue yielded by the algorithms (except MNL) increases as *pInf* increases. The underlying reason is that with higher expected influence probability, our algorithm and benchmarks with IM planning will benefit from enhanced product attractiveness. Again we observe that ALGC outperforms benchmarks by at least 20% in terms of the expected revenue, which shows the superiority of our proposed algorithm.

Figure 2(c) presents the performance of the algorithms with respect to changes in the value of *IM_k*, the cardinality limit of IM planning. We observe that the expected revenue achieved by the algorithms (except MNL) increases as *IM_k* increases. This demonstrates that by incorporating more well planned influencer promotions into assortment planning, we are able to generate a higher expected revenue thanks to enhanced attractiveness of the products. We observe that among all algorithms, ALGC yields the highest expected revenue. This again verifies the effectiveness of our proposed algorithm compared with the benchmarks.

Figure 3 shows the results of the performance of the algorithms with respect to changes in various parameters when 30 products are considered, where we observe a similar pattern on the structure of the algorithms. Figure 4 and 5 show the results for the budget constrained augmented assortment planning with 20 products and 30 products, respectively. We observe that ALGK outperforms the benchmarks by at least 20%, and that the algorithms with IM planning outperforms MNL. This again demonstrates the superiority of our proposed algorithm in terms of improving the expected revenue.

Figure 6(a) and (b) illustrate the performance of the algo-

# Influencer	# Customer Types				
	10	20	30	50	100
10	0.275	0.569	0.804	1.286	2.585
20	0.481	1.149	1.496	2.670	6.040
30	0.886	1.867	2.741	4.229	9.191
50	1.544	2.812	5.012	7.107	16.499
100	3.257	6.224	10.885	18.166	34.272

Table 2: Running Time for ALGK Algorithm (in seconds).

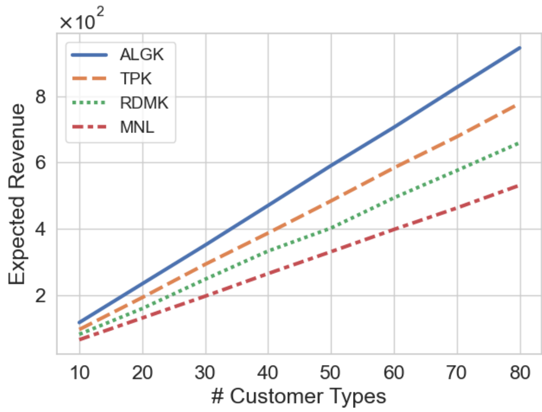

Figure 7: ALGK achieves superior expected revenue for augmented assortment planning under varying number of customer types.

rithms on real datasets for cardinality constrained and knapsack constrained augmented assortment planning, respectively. As shown in Figure 6(a), the expected revenue yielded by the algorithms (except MNL) increases as *IM_k* increases. We observe that among all algorithms, ALGC yields the highest expected revenue, gaining the benchmarks by more than 10% across all test cases. As shown in Figure 6(b), the expected revenue yielded by the algorithms (except MNL) increases as the budget increases. We observe that among all algorithms, ALGK yields the highest expected revenue, gaining the benchmarks by more than 20% across all test cases. This again demonstrates that incorporating well planned influencer promotions into assortment planning helps generate a higher expected revenue thanks to enhanced attractiveness of the products.

Finally, we present the results of the scalability test. We set the number of influencers to 50 and vary the number of customer types from 10 to 80. The size constraint is set as 50 and the budget is set as 100. As shown in Figure 6(c), the algorithms yield higher expected revenue as more customer types are considered, since the total revenue is calculated as the sum of revenue over all customer types. We observe that ALGC outperforms the benchmarks by more than 20% in expected revenue, and the performance gap increases as more customer types are considered. Figure 7 shows the superiority of ALGK in terms of yielded expected revenue, gaining at least 20% over the benchmarks. We report the running time of ALGC and ALGK in Table 1 and 2, respectively. In our experiments, the benchmarks return their results in 0.003 seconds. We observe that as the number of influencers and customer types increase, it takes longer for the algorithms to return the results, but still in a reasonable range, i.e., less than a minute across the test cases. This demonstrates that ALGC and ALGK achieve superior performance with outstanding efficiency.

Conclusion and Future Work

In this paper, we study a novel joint assortment and influencer marketing planning problem. While our current research framework relies on Assumption 0.3 , which assumes that promoting a product through IM does not decrease its attractiveness, empirical studies (Evans et al. 2017) have shown that the impact of IM on brand attitudes can be complex and context-dependent. We leave the task of addressing this challenge for future research.

References

Anderson, S. P.; De Palma, A.; and Thisse, J.-F. 1992. *Discrete choice theory of product differentiation*. MIT press.

Aouad, A.; Farias, V. F.; and Levi, R. 2015. Assortment optimization under consider-then-choose choice models. *Available at SSRN 2618823*.

Badanidiyuru, A.; Mirzasoleiman, B.; Karbasi, A.; and Krause, A. 2014. Streaming submodular maximization: Massive data summarization on the fly. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 671–680.

Berbeglia, G.; Flores, A.; and Gallego, G. 2021. The Refined Assortment Optimization Problem. *arXiv preprint arXiv:2102.03043*.

Blanchet, J.; Gallego, G.; and Goyal, V. 2016. A markov chain approximation to choice modeling. *Operations Research*, 64(4): 886–905.

Davis, J. M.; Gallego, G.; and Topaloglu, H. 2014. Assortment optimization under variants of the nested logit model. *Operations Research*, 62(2): 250–273.

El Housni, O.; and Topaloglu, H. 2021. Joint assortment optimization and customization under a mixture of multinomial logit models: On the value of personalized assortments. *Available at SSRN 3830082*.

Evans, N. J.; Phua, J.; Lim, J.; and Jun, H. 2017. Disclosing Instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of interactive advertising*, 17(2): 138–149.

Farias, V. F.; Jagabathula, S.; and Shah, D. 2013. A nonparametric approach to modeling choice with limited data. *Management science*, 59(2): 305–322.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Gao, P.; Ma, Y.; Chen, N.; Gallego, G.; Li, A.; Rusmevichientong, P.; and Topaloglu, H. 2020. Assortment optimization and pricing under the multinomial logit model with impatient customers: Sequential recommendation and selection. *Operations Research*.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Jung, K.; Heo, W.; and Chen, W. 2012. Irie: Scalable and robust influence maximization in social networks. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, 918–923. IEEE.

Li, G.; Rusmevichientong, P.; and Topaloglu, H. 2015. The d-level nested logit model: Assortment and price optimization problems. *Operations Research*, 63(2): 325–342.

McFadden, D.; et al. 1973. Conditional logit analysis of qualitative choice behavior.

Talluri, K.; and Van Ryzin, G. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1): 15–33.

Tang, S. 2018. When social advertising meets viral marketing: Sequencing social advertisements for influence maximization. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Tang, S.; and Yuan, J. 2016. Optimizing Ad Allocation in Social Advertising. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 1383–1392. ACM.

Udwani, R. 2021. Submodular Order Functions and Assortment Optimization. *arXiv preprint arXiv:2107.02743*.

Wang, C.; Wang, Y.; and Tang, S. 2023. When advertising meets assortment planning: Joint advertising and assortment optimization under multinomial logit model. In *International Conference on Combinatorial Optimization and Applications*, 16–42. Springer.

Yuan, J.; and Tang, S. 2017. No time to observe: adaptive influence maximization with partial feedback. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 3908–3914. AAAI Press.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? N/A
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? N/A
   (g) Did you discuss any potential misuse of your work? N/A
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? N/A
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? N/A
   (b) Have you provided justifications for all theoretical results? N/A
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? N/A
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? N/A
   (e) Did you address potential biases or limitations in your theoretical framework? N/A
   (f) Have you related your theoretical results to the existing literature in social science? N/A
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? N/A
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? Yes
   (b) Did you include complete proofs of all theoretical results? Yes
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? N/A
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? N/A
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? N/A
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? N/A
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? N/A
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? N/A
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? N/A
   (b) Did you mention the license of the assets? N/A
   (c) Did you include any new assets in the supplemental material or as a URL? N/A
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? N/A
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? N/A
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? N/A
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? N/A
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? N/A
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? N/A
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? N/A
   (d) Did you discuss how data is stored, shared, and deidentified? N/A