

# A Unified Online-Offline Framework for Co-Branding Campaign Recommendations

Xiangxiang Dai  
The Chinese University of Hong  
Kong, Hong Kong SAR, China  
xxdai23@cse.cuhk.edu.hk

Xiaowei Sun  
Fudan University,  
Shanghai, China  
xwsun24@m.fudan.edu.cn

Jinhang Zuo  
City University of Hong Kong,  
Hong Kong SAR, China  
jinhang.zuo@cityu.edu.hk

Xutong Liu\*  
Carnegie Mellon University,  
PA, USA  
xutongl@andrew.cmu.edu

John C.S. Lui  
The Chinese University of Hong  
Kong, Hong Kong SAR, China  
cslui@cse.cuhk.edu.hk

## Abstract

Co-branding has become a vital strategy for businesses aiming to expand market reach within recommendation systems. However, identifying effective cross-industry partnerships remains challenging due to resource imbalances, uncertain brand willingness, and ever-changing market conditions. In this paper, we provide the first systematic study of this problem and propose a unified online-offline framework to enable co-branding recommendations. Our approach begins by constructing a bipartite graph linking “initiating” and “target” brands to quantify co-branding probabilities and assess market benefits. During the online learning phase, we dynamically update the graph in response to market feedback, while striking a balance between exploring new collaborations for long-term gains and exploiting established partnerships for immediate benefits. To address the high initial co-branding costs, our framework mitigates redundant exploration, thereby enhancing short-term performance while ensuring sustainable strategic growth. In the offline optimization phase, our framework consolidates the interests of multiple sub-brands under the same parent brand to maximize overall returns, avoid excessive investment in single sub-brands, and reduce unnecessary costs associated with over-prioritizing a single sub-brand. We present a theoretical analysis of our approach, establishing a highly nontrivial sublinear regret bound for online learning in the complex co-branding problem, and enhancing the approximation guarantee for the NP-hard offline budget allocation optimization. Experiments on both synthetic and real-world co-branding datasets demonstrate the practical effectiveness of our framework, with at least 12% improvement.

## CCS Concepts

• **Information systems** → **Recommender systems**; • **Theory of computation** → **Online learning algorithms**; **Approximation algorithms analysis**.

\*Xutong Liu is the corresponding author.



This work is licensed under a Creative Commons Attribution 4.0 International License. *KDD '25, Toronto, ON, Canada*

© 2025 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1454-2/2025/08  
<https://doi.org/10.1145/3711896.3736824>

## Keywords

Co-branding recommendation, Online learning, Multi-armed bandit, Offline optimization, Approximation guarantee

## ACM Reference Format:

Xiangxiang Dai, Xiaowei Sun, Jinhang Zuo, Xutong Liu, and John C.S. Lui. 2025. A Unified Online-Offline Framework for Co-Branding Campaign Recommendations. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25)*, August 3–7, 2025, Toronto, ON, Canada. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3711896.3736824>

## KDD Availability Link:

The source code of this paper has been made publicly available at <https://doi.org/10.5281/zenodo.15532548>.

## 1 Introduction

Co-branding recommendation is a popular strategy that businesses and consumers alike embrace. By leveraging collaborations between two different brands, this recommendation strategy involves co-creating new products or services to maximize user satisfaction and achieve joint market benefits [13]. For example, the launch of CASETiFY’s collaborative line of electronic accessories in partnership with Netflix’s television series “Squid Game 2” strongly resonates with fans of the series, showcasing the success of such brand collaborations<sup>1</sup>. These collaborations not only provide brands with enhanced exposure and increased market gains, but also cater to consumer demand for novelty. According to SocialBeta’s data [23] covering the period from March 2022 to February 2023, approximately 15% of the 4,428 recorded brand marketing cases involved co-branding. Further analysis reveals that this proportion never dropped below 10% in any given month, with co-branding seeing notable spikes during key promotional periods like summer sales and the year-end holiday season (e.g., Christmas and New Year). In this context, co-branding has now become a trend. Despite its potential, a key challenge remains: *How can brands identify the most suitable partners from the set of target brands to ensure successful co-branding and maximize market returns?* This question remains systematically unresolved and cannot be overlooked. For instance, the collaboration between Heytea and the Jingdezhen China Ceramic Museum resulted in reduced profits, highlighting the importance of selecting the right target brands [14, 33].

<sup>1</sup><https://www.casetify.com/co-lab/squid-game-2024>

However, this is a challenging task. *The co-branding market is fraught with uncertainty.* Firstly, overemphasizing certain popular co-branding collaborations can lead to a resource imbalance, i.e., “Matthew Effect” [11], where smaller, long-tail brands are overlooked despite their potential to deliver greater market benefits. The willingness of target brands to participate in collaborations is also uncertain, influenced by multiple factors such as brand positioning and financial commitment from the initiating brand. Additionally, market changes make over-reliance on historical data risky. For instance, the continued use of certain types of co-branding may lead to consumer fatigue or shifting preferences, which may be overlooked by brands overly dependent on past performance. In such cases, collaborations with new brands can disrupt the status quo and introduce a sense of novelty [3, 33].

In such a dynamic online recommendation market, static strategies are insufficient for determining appropriate co-branding partnerships. Instead, *a more dynamic approach that continuously adapts to uncertainty is required.* This involves addressing the “*exploration-exploitation trade-off*”, where brands must balance the known willingness and market benefits of established partnerships while exploring potential collaborations with new, promising target brands that could deliver better market performance. Recent studies have shown that multi-armed bandit (MAB) methods can effectively address this trade-off challenge in recommendation systems in a lightweight, online manner [9, 15, 22, 37]. *From a decision-making perspective, co-branding can be framed as a recommendation problem, where the goal is to suggest partnerships that maximize long-term market impact.* However, unlike traditional recommendation tasks (e.g., recommending a movie to a user), the costs associated with co-branding decisions are significantly high [8, 24, 39]. Implementing a co-branding partnership requires substantial upfront efforts, including extensive market research and creating branded materials, all of which incur significant costs. As a result, while long-term brand objectives may benefit from the online adjustment of co-branding strategies to maximize collaboration gains, exploration must be conducted with caution to mitigate risks—especially in the early stages—to prevent brands from losing confidence in the algorithm before its benefits materialize.

In addition to the inherent uncertainty surrounding the willingness or feasibility of co-branding and market gains, brands must also manage budget constraints when planning co-branding initiatives. In practice, the parent brand that initiates co-branding campaigns often oversees “*multiple sub-brands*” (e.g., the Coca-Cola company owns sub-brands like *Sprite* and *Fanta*, or *Marvel* has various superhero intellectual properties (IPs), such as Spider-Man and Iron Man). An illustration of this structure is shown in (a) part of Fig. 1, with further details provided in Section 2. Due to financial limitations, it may not be feasible for every sub-brand to engage in co-branding with target brands. Consequently, when planning and allocating budgets for a co-branding initiative, the parent brand must adopt a holistic view of the entire brand ecosystem, rather than focusing on a single sub-brand. Focusing solely on co-branding efforts for individual sub-brands could reduce overall collaboration benefits, potentially affecting the profitability of the parent brand as a whole. In such cases, allocating excessive resources to a single sub-brand’s co-branding efforts may lead to diminishing returns, whereas a more cost-effective sub-brand might generate higher

overall value for the parent brand. Therefore, *it is crucial to assess and align the interests of all sub-brands within the parent brand’s portfolio.* Optimizing budget allocation for maximum collective benefit, rather than focusing on individual sub-brands, can help maximize overall market outcomes.

To the best of our knowledge, no existing work systematically addresses the complex problem of co-branding campaign recommendations. We present *the first* systematic approach to model and tackle this challenge. Motivated by the aforementioned practical needs, we propose a unified online-offline framework, beginning with the construction of a co-branding bipartite graph model. This model captures the structural relationships between the initiating parent brand—along with its multiple sub-brands—and target brands, while incorporating key factors such as the willingness or feasibility of co-branding and market gains to accurately represent brand interactions. Our framework enables two key processes: online graph learning and offline budget optimization, leveraging their synergy to continuously enhance co-branding strategies.

**Online Phase.** The model dynamically updates itself based on online market feedback rather than remaining static. Extending beyond deterministic arm selection in traditional MAB methods, our approach explicitly addresses uncertainties in co-branding, including budget-dependent co-branding willingness and market gains that are revealed only upon successful co-branding. Given the high upfront investment in co-branding, we also emphasize short-term performance alongside long-term performance by strategically minimizing redundant exploration. Theoretically, through a highly non-trivial analysis, we merge the error terms introduced by these new uncertainties while maintaining a near-optimal regret bound.

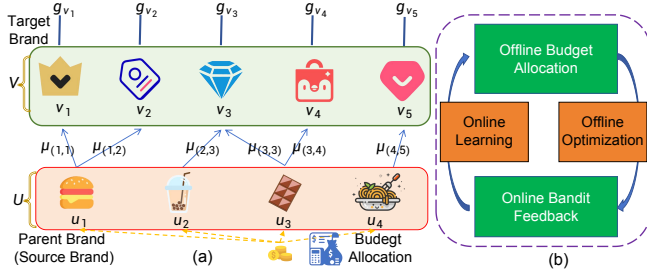
**Offline Phase.** We allocate budgets based on the insights learned from online feedback, optimizing future collaboration outcomes from the parent brand’s perspective. This approach avoids the pitfalls of overemphasizing individual sub-brands and mitigates unnecessary premium costs. Given the high expenses of co-branding campaigns—e.g., joint marketing, material production, and negotiation efforts—we enhance the approximation ratio by introducing a partial enumeration technique. This enables more flexible allocations for the NP-hard budget allocation problem while ensuring computational traceability compared to previous works.

Our contributions are listed as follows.

- **Model Formulation:** We propose a novel co-branding bipartite graph model that captures critical features based on the bidirectional relationships between “initiating” and “target” brands. This model not only incorporates the uncertainty of successful collaborations, but also incorporates market gains, enabling a comprehensive representation of the potential value created by different co-branding partnerships.

- **Algorithm Design:** To balance exploration and exploitation in an online market, we extend the MAB framework and adapt it to the co-branding scenario during the online learning phase, which emphasizes both long-term gains and short-term performance. Concurrently, in the offline optimization phase, we allocate budgets at the parent brand level, ensuring equitable investment among sub-brands and preventing over-allocation to a single sub-brand.

- **Theoretical Analysis:** We establish theoretical analysis and interoperability for our framework. For the online learning phase, we provide a near-optimal sub-linear regret analysis under complex



**Figure 1: Co-Branding Bipartite Graph with Online-Offline Iterative Framework (with  $U = 4, V = 5$  in this example).**

uncertainty with refined exploration. In the offline optimization phase, we guarantee a  $1 - 1/e$  approximation for the NP-hard optimal solutions under budget considerations.

• **Experimental Evaluation:** We perform extensive evaluations on both synthetic and real-world datasets, including curated datasets that we will publicly release. The results show that our approach achieves 39% improvements on average over traditional methods, delivering superior short-term and long-term returns.

## 2 Model and Problem Formulation

This section presents the co-branding problem formulation, comprising both the online graph learning scenario and the offline budget allocation problem<sup>2</sup>. For  $\forall z \in \mathbb{N}_0$ , let  $[z]$  denote the set  $\{1, 2, \dots, z\}$ , with  $|\cdot|$  denoting the size of a set.

### 2.1 Co-Branding Bipartite Graph Model

We adopt a bipartite graph  $\mathcal{G} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$  as the foundational framework to model co-branding opportunities between a parent brand and potential partner brands. This choice is motivated by the natural division in co-branding scenarios: sub-brands initiating co-branding and potential partner brands that could participate in these initiatives. As shown in the (a) part of Fig. 1, the set  $\mathcal{U}$  denotes co-branding initiators, representing a parent brand system with multiple sub-brands, where  $|\mathcal{U}| = U$  is the number of sub-brands involved in co-branding. These sub-brands maintain diverse market positions and brand images within the brand architecture, enabling various co-branding initiatives. For companies without sub-brands,  $U$  reduces to 1, representing a single-brand entity. The set  $\mathcal{V}$  represents target co-branding partner brands, where  $|\mathcal{V}| = V$  denotes the number of available potential partner brands. These target brands allow individual assessment of each co-branding opportunity's feasibility and success probability.

Each edge  $e := (u, v) \in \mathcal{E}$  signifies a co-branding pair between sub-brand  $u \in \mathcal{U}$  and target partner brand  $v \in \mathcal{V}$ . The weight vector  $\mu = \{\mu_e\}_{e \in \mathcal{E}}$  represents the probability of successful co-branding for each edge. These probabilities are influenced by multiple factors in the complex business environment, e.g., market positioning alignment, brand image compatibility, and profit-sharing arrangements. These probabilities are typically unknown a priori and require continuous online learning to update. Moreover, the probability positively correlates with the parent brand's financial budget  $\mathbf{b}$  (introduced in detail shortly), i.e., higher investment generally increases the likelihood of successful co-branding. Since any

$u$  belongs to the same parent company  $\mathcal{U}$ , there is no bidding agreement between sub-brand  $u$  and any target brand  $v \in \mathcal{V}$ , meaning  $v$  is not restricted to co-branding with only one specific  $u$  from  $\mathcal{U}$ . The graph  $\mathcal{G}$  also incorporates the brand market gain vector  $\mathbf{g} = \{g_v\}_{v \in \mathcal{V}}$ , which represents the revenues of selling products or services from a target brand  $v$  to the entire brand company  $\mathcal{U}$  upon the launch of a co-branding initiative. Note that the market gain is primarily driven by consumer preferences rather than budget allocation, as a complex market environment does not necessarily imply a direct correlation between higher investment and higher returns. Each  $g_v \in \mathbf{g}$  denotes the consumer market gain from target brand  $v$  to the whole parent brand  $\mathcal{U}$ , which is initially unknown.

### 2.2 Online Feedback Mechanism

While brands may leverage historical collaboration data to construct a static bipartite graph  $\mathcal{G}$ , this approach is ill-suited to the ever-changing nature of market conditions. To address the inherent dynamics and uncertainties in the probabilities of successful brand partnerships and market gains, we adopt an online learning framework. This framework dynamically updates the bipartite graph  $\mathcal{G} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$  based on sequentially observed outcomes.

The online learning process unfolds over  $T$  rounds, each corresponding to a co-branding season. Initially, the structure of  $\mathcal{G}$  is unclear and incomplete. During each co-branding season  $t$  (i.e., round), a budget allocation  $\mathbf{b}_t = (b_{t,1}, \dots, b_{t,U})$  is determined based on observations from previous rounds (the budget allocation model will be introduced later). Under this allocation, co-branding initiators  $\mathcal{U}$  engage in co-branding negotiations with target brands  $\mathcal{V}$ . The resulting co-branding pairings between sets  $\mathcal{U}$  and  $\mathcal{V}$  are denoted as action  $S_t^{b_t} \subseteq \mathcal{E}$  for the current co-branding season. The parent company can determine  $S_t$  under  $\mathbf{b}_t$  using various business strategies, such as prioritizing sub-brands with higher cooperation potential, distributing the budget progressively, or adopting a hybrid approach, though these heuristics are beyond our focus. For brevity, this is abbreviated as  $S_t$  when no ambiguity arises.

Upon executing actions on co-branding pairs, the decision-maker observes a feedback vector  $X_{t,S_t} = (X_{t,1}, \dots, X_{t,|S_t|}) \in [0, 1]^{|S_t|}$ , where each  $X_{t,e,b_u}$  reflects the propensity for co-branding success for edge  $e = (u, v)$  under budget allocation  $b_u$ . Specifically, if brand  $v$  accepts a co-branding invitation from brand  $u$ ,  $X_{t,e,b_u} = 1$ , signaling a successful outcome. Conversely,  $X_{t,e,b_u} = 0$  represents a failed invitation. The expectation of  $X_{t,e}$  is denoted as  $\mu_e$ , representing the probability of successful co-branding between  $u$  and  $v$ . Leveraging this feedback, the decision-maker can update its estimates of the cooperation probabilities  $\mu$ , enabling it to prioritize promising brand pairings while avoiding those with uncertain prospects. Furthermore, for each successful co-branding pair  $e = (u, v)$ , where  $X_{t,e} = 1$ , the market gain  $Y_{t,v}$  from the target brand  $v$  is observed under the budget allocation  $\mathbf{b}$  for the parent brand  $\mathcal{U}$  in season  $t$ . The expectation of  $Y_{t,v}$ , denoted as  $g_v$ , represents the brand market gain, quantifying the economic benefits of the collaboration. Notably, this feedback is available only for successful co-branding actions. We denote the vector of observed market gain feedback at season  $t$  as  $Y_{t,\mathcal{V}}$ . By analyzing these feedback data, the decision-maker acquires valuable insights into the tangible economic benefits of

<sup>2</sup>Further discussions and extensions about the model can be found in Appendix A.

partnerships, enabling more effective and informed decisions for future co-branding strategies.

As a result, when action  $S_t$  is executed under the given budget allocation, the decision-maker observes a reward  $R_{\mathcal{G}}(\mathbf{b}_t)$ , which is influenced by both the co-branding probabilities and the market gains under the graph  $\mathcal{G}$ :

$$R_{\mathcal{G}}(\mathbf{b}_t) = \sum_{v \in \mathcal{V}} \mathbb{I}\{\exists e = (u, v) \in S_t \text{ s.t. } X_{t,e,b_u} = 1\} Y_{t,v}, \quad (1)$$

where  $\mathbb{I}\{\cdot\}$  is an indicator function that evaluates to 1 if one successful co-branding action  $X_{t,e} = 1$  exists, and 0 otherwise. For the online phase, the decision-maker needs to accurately and efficiently estimate and learn the structure of the graph  $\mathcal{G}$  to maximize the cumulative reward over time.

### 2.3 Offline Strategic Budget Allocation

As described earlier, the financial investments of the parent brand play a pivotal role in co-branding negotiations and planning for the upcoming season. Specifically, based on the online-learned market graph  $\mathcal{G}$ , which includes the co-branding probabilities  $\mu$  of co-branding and the corresponding market gains  $\mathbf{g}$ , the parent brand must allocate a total budget  $B \in \mathbb{N}$  (typically a large integer, e.g., in millions) across its sub-brands. The allocation is represented as  $\mathbf{b} = (b_1, \dots, b_U) \in \mathbb{N}_0^U$ , where  $b_u$  denotes the budget assigned to sub-brand  $u$ , subject to the constraint  $\sum_{u=1}^U b_u \leq B$ . Subbrands  $u \in \mathcal{U}$  that do not participate in co-branding (e.g. due to poor past performance) are assigned a zero budget, i.e.,  $b_u = 0$ . Furthermore, each subbrand  $u$  is assigned a predetermined budget cap  $c_u$ , set by the parent brand, ensuring  $b_u \leq c_u$  for all  $u \in \mathcal{U}$ . This ensures a controlled and strategic allocation of financial resources while accounting for the individual potential of each sub-brand.

Based on Eq. (1) and the definitions of  $\mu$  and  $\mathbf{g}$ , the expected reward  $r_{\mathcal{G}}(\mathbf{b})$  for season  $t$ , derived from  $R_{\mathcal{G}}(\mathbf{b})$  with action  $S$  on  $X_{t,S}$  and  $Y_{t,\mathcal{V}}$  is expressed as:

$$r_{\mathcal{G}}(\mathbf{b}) = \sum_{v \in \mathcal{V}} g_v \left( 1 - \prod_{e=(u,v) \in S} (1 - \mu_{e,b_u}) \right). \quad (2)$$

The offline budget allocation problem involves distributing the total budget  $B$  among  $U$  sub-brands to maximize the reward  $r_{\mathcal{G}}(\mathbf{b})$ . To simplify marketing budget decisions, brand managers often prefer predefined strategies over continuous, fine-tuned adjustments [28, 30]. Accordingly, we define the set of tentative spending plans for each sub-brand  $u$  as  $\mathcal{N}_u \subseteq \mathbb{N}_0$ , where  $s_u \in \mathcal{N}_u$  represents a specific tentative spending plan, still constrained by  $s_u \leq c_u$ . For instance, a sub-brand may offer three tiers of budget plans, such as low, medium, and high, to provide flexibility while remaining within the predefined cap. The optimization problem is formalized as:

$$\begin{aligned} \max_{\mathbf{b}} \quad & r_{\mathcal{G}}(\mathbf{b}) = \mathbb{E}[R_{\mathcal{G}}(\mathbf{b}_t)] \\ \text{s.t.} \quad & \mathbf{b} \in \mathbb{N}_0^U, \sum_{u=1}^U b_u \leq B, \\ & b_u \leq c_u, b_u \in \mathcal{N}_u, \forall u \in \mathcal{U}. \end{aligned} \quad (3)$$

Eq. (3) aims to optimally allocate resources among sub-brands for maximizing the total revenue, prioritizing those with higher potential returns while adhering to budget constraints. However, the expected reward function  $r_{\mathcal{G}}(\mathbf{b})$  is highly non-linear, rendering the

problem NP-hard [7, 16, 34] and extending beyond the standard knapsack problem framework [1]. As solving it to optimality is computationally infeasible, we pursue an offline  $\alpha$ -approximation solution, where  $\alpha \in (0, 1]$  denotes the approximation ratio. Specifically, the solution  $\mathbf{b}$  satisfies:  $r_{\mathcal{G}}(\mathbf{b}) \geq \alpha \cdot r_{\mathcal{G}}(\mathbf{b}^*)$ , where  $\mathbf{b}^*$  denotes the computationally intractable optimal budget allocation. This approach provides a practical and efficient method to approximate the optimal solution while maintaining acceptable accuracy. Our goal is to maximize  $\alpha$ , achieving a solution as close to optimal as feasible within computational constraints.

### 2.4 Unified Problem Formulation

In practical applications, complete information about the underlying graph  $\mathcal{G}$  is rarely available. The online learning problem addresses this challenge by iteratively learning the underlying graph parameters through interactions with the market environment. This decision-making process involves a critical trade-off between “exploitation” (leveraging historical empirical probabilities of successful co-branding and market gains) and “exploration” (identifying potentially better but yet undiscovered co-branding opportunities and market gains). In the offline phase, at the beginning of each season  $t$ , the decision-maker must determine the budget allocation  $\mathbf{b}_t$  based on the online-learned graph  $\mathcal{G}$ . As shown in the (b) part of Fig. 1, these two components mutually reinforce each other, fostering continuous improvement in co-branding strategies.

The performance of an algorithm  $A$ , which operates over both offline and online phases, is evaluated using the concept of “ $\alpha$ -approximate regret” [5, 20, 25]. Over  $T$  rounds, the cumulative regret is defined as:

$$Reg(T) = \alpha T \cdot r_{\mathcal{G}}(\mathbf{b}^*) - \mathbb{E} \left[ \sum_{t=1}^T r_{\mathcal{G}}(\mathbf{b}_t^A) \right], \quad (4)$$

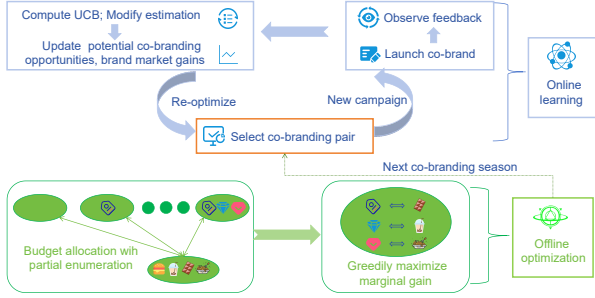
where  $\mathbf{b}_t^A$  represents the budget allocation chosen by algorithm  $A$  in season  $t$ . The goal is to design an algorithm  $A$  that minimizes  $Reg(T)$ , ensuring that the cumulative rewards closely approximate those of the optimal allocation over  $T$  rounds. Minimizing  $Reg(T)$  presents significant challenges. First, uncertainty in partnerships arises as leading brands often dominate resource allocation, overshadowing long-tail brands with untapped potential. Similarly, dynamic market conditions make over-reliance on historical data risky. These uncertainties jointly highlight the need for a careful balance between exploration and exploitation. Beyond long-term performance, early-stage failures pose a risk to stakeholder trust in online learning methods, requiring a more careful exploration strategy. Lastly, parent brands managing multiple sub-brands face intricate resource distribution challenges, where the objective shifts from optimizing individual sub-brand performance to maximizing collective benefits across the entire brand portfolio.

## 3 Algorithmic Design

As shown in Fig. 2, this section outlines our algorithm design, which seamlessly integrates both online and offline processes.

### 3.1 Graph Learning via Online Feedback

**Exploration-Exploitation Trade-off.** In practical markets, the graph  $\mathcal{G}$  information structure is often *partially or entirely unknown*,



**Figure 2: Workflow of Hybrid Online-Offline Algorithm:** It begins with the estimation of the co-branding bipartite graph, followed by offline budget allocation for selecting optimal co-branding pairs. After executing initial campaigns, market feedback is integrated to refine the estimations. These updated insights then guide the re-optimization of subsequent campaigns.

leading to uncertainties in identifying potential co-branding opportunities and estimating their potential gains. A naive approach that always selects the current best-known partner based on empirical feedback risks becoming trapped in a *local optimum*, as it emphasizes exploitation over exploration. To overcome this pitfall, we introduce a confidence-based multi-armed bandit (MAB) strategy that effectively balances exploration and exploitation [19, 21].

**Algorithm Workflow.** We introduce our online algorithm in Algorithm 1, an adaptation of the combinatorial upper confidence bound (CUCB) algorithm tailored specifically to the co-branding context within the general combinatorial multi-armed bandit (CMAB) framework [4–6]. Unlike the deterministic arm selection employed in the original CUCB algorithm [5], co-branding introduces unique challenges: the co-branding probability  $\mu$  is influenced by the specific budget allocation rather than being fixed, and the unknown market gain  $g$  depends on the co-branding probability, which is only observable upon success. To address these challenges, we extend the concept of the base arm in [4, 5]. Specifically, we define a set of partial base arms, denoted as  $\mathcal{A} = \{(e, s) \mid e \in \mathcal{E}, s \in \mathcal{N}_u\}$ , where  $\mu_{e,s}$  denotes the probability of successful co-branding for each base arm  $(e, s)$ . Here,  $(e, s)$  corresponds to a co-branding pair  $e$  combined with a tentative spending plan  $s$ . Furthermore, we introduce an additional set of base arms  $\mathcal{A}'$  to represent the remaining base arms corresponding to the unknown market gain  $g$ , which contribute to additional regret in the system. The complete set of base arms is therefore given by  $\mathcal{A} \cup \mathcal{A}'$ , allowing us to address the co-branding dynamics within our algorithmic framework.

We estimate the co-branding probability  $\mu_{e,s}$  using the empirical mean  $\hat{\mu}_{e,s}$ , where  $T_{t,e,s}$  record the number of times the base arm  $(e, s) \in \mathcal{A}$  has been selected up to season  $t$ ; Similarly,  $\hat{g}_v$  represents the empirical mean of  $v$ 's market gain, with  $T'_v$  recording the number of times the target brand  $v \in \mathcal{A}'$  has been selected (line 1). Algorithm 1 leverages confidence bounds to balance exploitation and exploration. Specifically, for each season  $t$ , the confidence radius  $\rho_{e,s}$ , computed in line 3, regulates the level of exploration for potentially better co-branding pairs. This radius is larger for under-explored arms  $(e, s)$  (i.e., when  $T_{t,e,s}$  is small), thereby incentivizing exploration. Similarly, in line 4,  $\hat{g}_v$  is the optimistic estimate of the market gain  $g_v$  with confidence radius  $\rho'_v$ . Co-branding feedback is collected after executing the action  $S_t$  under budget allocation  $\mathbf{b}$ , as

#### Algorithm 1 Confidence-Based Online Learning for Co-Branding (CBOL)

**Require:** Set of co-branding initiators  $\mathcal{U}$ , set of target brands  $\mathcal{V}$ .

- 1: Initialize  $T_{t,e,s}$ ,  $\hat{\mu}_{e,s}$  for each  $(e, s) \in \mathcal{A}$ , and  $T'_v$ ,  $\hat{g}_v$  for each  $v \in \mathcal{A}'$  using historical dataset  $\mathcal{D}$ .
- 2: **for** season  $t = 0, 1, 2, \dots, T$  **do**
- 3: **For**  $(e, s) \in \mathcal{A}$ ,  $\rho_{e,s} \leftarrow \text{Eq. (5)}$ ,  $\tilde{\mu}_{e,s} \leftarrow \hat{\mu}_{e,s} + \rho_{e,s}$ ,  $\tilde{\mu}_{e,s} \leftarrow \max_{j \in \mathcal{N}_u, j \leq s} \tilde{\mu}_{e,j}$ .
- 4: **For**  $v \in \mathcal{A}'$ ,  $\rho'_v \leftarrow \sqrt{\frac{6\hat{V}_{e,s} \log t}{T_{t,v}}} + \frac{9 \log t}{T_{t,v}}$ ,  $\hat{g}_v \leftarrow \hat{g}_v + \rho'_v$ .
- 5: Budget allocation  $\mathbf{b} \leftarrow \text{GPE (Algorithm 2)}$ .
- 6: Observe co-branding intention feedback  $X_{t,S_t}$  under budget allocation  $\mathbf{b}$ .
- 7: **For each**  $(e, s)$  that receives feedback  $X_{e,s}$ , update  $T_{t,e,s} \leftarrow T_{t,e,s} + 1$ ,  $\hat{\mu}_{e,s} \leftarrow \hat{\mu}_{e,s} + (X_{e,s} - \hat{\mu}_{e,s})/T_{t,e,s}$ ,  $\hat{V}_{e,s} \leftarrow \frac{T_{t,e,s}-1}{T_{t,e,s}} \left( \hat{V}_{e,s} + \frac{1}{T_{t,e,s}} (\hat{\mu}_{e,s} - X_{t,e,s})^2 \right)$ .
- 8: **For any successful co-branding pair**  $e \in S_t$  with  $X_{t,e,s} = 1$ , observe market gain  $Y_{t,v}$  and update  $T'_{t,v} \leftarrow T'_{t,v} + 1$ ,  $\hat{g}_v \leftarrow \hat{g}_v + (Y_{t,v} - \hat{g}_v)/T'_{t,v}$ ,  $\hat{V}'_v \leftarrow \frac{T'_{t,v}-1}{T'_{t,v}} \left( \hat{V}'_v + \frac{1}{T'_{t,v}} (\hat{g}_v - Y_{t,v})^2 \right)$ .
- 9: **end for**

determined by Algorithm 2 (introduced subsequently). In line 7, the corresponding estimations are updated based on the online bandit feedback regarding co-branding outcomes  $X_{t,S_t}$ . Finally, the unknown market gain is refined in line 8 following the implementation of the co-branding campaign.

**Reduce Redundant Exploration.** Note that the original confidence radius in [5, 35] relies solely on the empirical mean, neglecting variance and resulting in a loose confidence radius for each arm. This can lead to unnecessary exploration. To address this, inspired by [27], we adopt a stronger Bernstein-type concentration bound [19], which incorporates the empirical variance  $\hat{V}_{t-1,i}$  to construct an improved confidence radius (line 3):

$$\rho_{e,s} = \sqrt{\frac{6\hat{V}_{e,s} \log t}{T_{t,e,s}}} + \frac{9 \log t}{T_{t,e,s}}. \quad (5)$$

In the first term of Eq. (5),  $\hat{V}_{e,s}$  serves as an approximation of the true variance  $V_{e,s}$ . Since  $V_{e,s} \leq (1 - \mu_{e,s})\mu_{e,s}$ , the estimation of  $\mu_{e,s}$  becomes more precise when  $\mu_{e,s}$  approaches boundary values. The second term of Eq. (5) compensates for using the empirical variance  $\hat{V}_{e,s}$  instead of  $V_{e,s}$ . A similar tighter confidence radius strategy is applied to estimate the unknown market gains (line 4).

To further reduce unnecessary exploration, we leverage historical dataset  $\mathcal{D}$  consisting of past season data with a scale of  $D$  to initialize observed potential co-branding probabilities  $\mu$ , brand market gains  $g$ , and selecting counts  $T_{t,e,s}$  (and  $T'_{t,v}$ ) (line 1). Note that naively incorporating historical data can degrade short-term performance, such as over-exploring low-reward arms or under-exploring high-reward ones, thereby incurring high regret [38]. Thus, Algorithm 1 excludes historical arm selections when computing the confidence radius  $\rho_{e,s}$  (and  $\rho'_v$ ), relying only on online feedback. This balances long-term benefits of historical data with reduced short-term exploration.

**Non-Decreasing UCB Values.** In line 3, we compute the upper confidence bound (UCB) value  $\tilde{\mu}_{e,s}$ . Although higher spending



**Algorithm 2** Greedy Partial Enumeration for Budget Optimization (GPE)

---

**Require:** Co-Branding graph  $\mathcal{G}$ , total budget  $B$ , budget cap  $c_u$ , tentative spending plans  $\mathcal{N}_u, u \in \mathcal{U}$ , operational constraint  $K$ .

- 1: Initialize  $\mathbf{b}_{max} \leftarrow \mathbf{0}$ .
- 2:  $\mathcal{B} \leftarrow \{\mathbf{b} = (b_1, \dots, b_U) \mid 0 \leq b_u \leq c_u, b_u \in \mathcal{N}_u, \sum_{u \in \mathcal{U}} b_u \leq B, \sum_{u \in \mathcal{U}} \mathbb{I}\{b_u > 0\} \leq K\}$ .
- 3: **for**  $\mathbf{b} \in \mathcal{B}$  **do**
- 4:    $B' \leftarrow B - \sum_{u \in \mathcal{U}} b_u$ .
- 5:   Let  $Q \leftarrow \{(u, s_u) \mid u \in \mathcal{U}, s_u \in \mathcal{N}_u, 1 \leq s_u \leq c_u - b_u\}$ .
- 6:   **while**  $B' > 0$  and  $Q \neq \emptyset$  **do**
- 7:      $(u^*, s^*) \leftarrow \arg \max_{(u,s) \in Q} \delta(u, s, \mathbf{b})/s$ .
- 8:     **if**  $s^* \leq B'$  **then**
- 9:        $s_{u^*} \leftarrow s_{u^*} + s^*, B' \leftarrow B' - s^*$ .
- 10:       Adjust all pairs  $(u^*, s) \in Q$  to  $(u^*, s - s^*)$ .
- 11:       Remove all pairs  $(u^*, s) \in Q$  such that  $s \leq 0$ .
- 12:     **else**
- 13:       Remove  $(u^*, s^*)$  from  $Q$ .
- 14:     **end if**
- 15:   **end while**
- 16:   **if**  $r_{\mathcal{G}}(\mathbf{b}) > r_{\mathcal{G}}(\mathbf{b}_{max})$ , **then**  $\mathbf{b}_{max} \leftarrow \mathbf{b}$ .
- 17: **end for**

---

typically increases (or at least does not decrease) collaboration probability in real-world co-branding scenarios, random exploration can cause  $\tilde{\mu}_{e,s}$  to paradoxically drop with larger  $s$ . To reflect realistic spending–collaboration trends, we ensure non-decreasing UCBs by setting each  $\tilde{\mu}_{e,s}$  to the maximum of all  $\tilde{\mu}_{e,j}$  with  $j \leq s$ .

### 3.2 Budget Optimization via Offline Planning

**Submodular Property Basis.** To address the budget allocation problem defined in Eq. (3), we exploit the *submodular property of the reward function*, which intuitively reflects the diminishing marginal returns as the budget increases [2, 26] (Formal proof of this property is in Section 4.1). While increasing the budget for a single sub-brand  $u$  within a co-branding strategy may not directly exhibit a diminishing marginal effect, it is noted that the overall marginal gain for the entire parent brand  $\mathcal{U}$  decreases as the total budget is distributed. This occurs because the available budget for other sub-brands within the same parent brand becomes limited as more budget is allocated to  $u$ . We now formally introduce this definition. For any  $\mathbf{x}, \mathbf{y} \in \mathbb{N}_0^U$ , we denote  $\mathbf{x} \vee \mathbf{y}, \mathbf{x} \wedge \mathbf{y} \in \mathbb{N}_0^U$  as the coordinate-wise maximum and minimum of these two vectors:  $(\mathbf{x} \vee \mathbf{y})_i = \max\{x_i, y_i\}, (\mathbf{x} \wedge \mathbf{y})_i = \min\{x_i, y_i\}$ . A function  $f: \mathbb{N}_0^U \rightarrow \mathbb{R}$  is defined as *submodular* over the integer lattice  $\mathbb{N}_0^U$  if it satisfies the following inequality for any  $\mathbf{x}, \mathbf{y} \in \mathbb{N}_0^U$ :  $f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ . Let  $\mathbf{x}_i = (0, \dots, 1, \dots, 0)$  represent the one-hot vector with the  $i$ -th equal to 1 and all other elements equal to 0.

**Refining Approximation Ratio.** Building on strategies that combine greedy algorithms with submodular property [2, 18, 26], which mainly give  $\alpha = \frac{1}{2}(1 - e^{-1}) \approx 0.316$ -approximation and  $\alpha = (1 - e^{-\eta}) \approx 0.357$ -approximation (where  $\eta$  satisfies  $e^\eta = 2 - \eta$ ), we integrate partial enumeration methods to further refine the approximation ratio. In the context of co-branding, the high cost

of creating co-brands for parent brands requires significant preliminary efforts, such as extensive market research and the costs associated with creating co-branding materials. Unlike traditional problems focused on reducing time complexity, the need for such upfront investments in co-branding makes it essential to optimize the approximation ratio without sacrificing the quality of the outcomes. Leveraging these insights, we focus on improving the approximation ratio for better co-branding results, while ensuring that the approach remains computationally tractable within polynomial time. The detailed procedure is presented in Algorithm 2.

**Algorithm Workflow.** Algorithm 2 constructs an initial solution set  $\mathcal{B}$ , which stores all partial solutions that allocate budgets to at most  $K$  sub-brands (line 2). Each initial solution corresponds to a potential final solution, from which we must select the optimal one. The parameter  $K$  serves as an operational constraint, and as  $K$  increases, the number of possible initial solutions grows, potentially exponentially. While increasing  $K$  can improve the approximation ratio by permitting more flexible allocations (e.g., distributing budgets evenly or dedicating the entire budget  $c_u$  is allocated to a single sub-brand  $u$ , it also significantly increases the computational complexity. Based on the theoretical analysis in Appendix B.3,  $K = 3$  provides a relatively reasonable trade-off between runtime efficiency and approximation quality. For any partial solution, Algorithm 2 maintains a set  $Q$ , where any pair  $(u, s) \in Q$  represents a tentative plan to allocate an additional budget of  $s$  to sub-brand  $u$ . The algorithm follows a *greedy iteration strategy*: a while loop (line 6–11) that iteratively selects the pair  $(u, s)$  from  $Q$  such that that maximizes the *per-unit marginal gain*  $\delta(u, s, \mathbf{b})$ . Specifically, the per-unit marginal gain is defined as  $\delta(u, s, \mathbf{b}) = (r_{\mathcal{G}}(\mathbf{b} + s\mathbf{x}_u) - r_{\mathcal{G}}(\mathbf{b}))/s$ , where  $r_{\mathcal{G}}$  is the reward function,  $\mathbf{b}$  is the current budget allocation, and  $\mathbf{x}_u$  represents the one-hot allocation vector for sub-brand  $u$ . If  $s$  is within the remaining budget, it is allocated to sub-brand  $u$ ; otherwise, the algorithm proceeds to the next iteration. Note that Algorithm 2 ensures monotonicity of the input due to the maximum operator (line 3 in Algorithm 1)<sup>3</sup>.

**Integration with Online Learning.** After budget allocation, sub-brands use the learned co-branding graph  $\mathcal{G}$  to engage with target brands for executing co-branding campaigns. If a sub-brand  $u$  has multiple successful co-branding partners from the target brand set  $\mathcal{V}$ , it allocates its budget  $b_u$  in a balanced manner, thus ensuring the exploration of diverse co-branding opportunities without over-investing in a single target brand. Alternatively, the budget can be proportionally allocated based on market returns, target brand performance, and strategic priorities, which depend on the brand's economic context and are beyond the primary focus of our research. Executing a co-branding campaign involves forming partnerships, deploying marketing resources, and monitoring performance. The corresponding feedback is used to update the estimates in future co-branding seasons, as described in Algorithm 1.

## 4 Theoretical Analysis

In this section, we provide a detailed theoretical analysis of the performance of our proposed algorithms. The key results are summarized here, with formal proofs presented in Appendix B.

<sup>3</sup>A function  $f$  is monotone if  $f(\mathbf{x}) \leq f(\mathbf{y})$  for all vectors  $\mathbf{x} \leq \mathbf{y}$ , where  $\mathbf{x} \geq \mathbf{y}$  means  $x_i \geq y_i$  for all elements  $i$ .

#### 4.1 Monotone Submodular Property

**LEMMA 1** (GRAPH-BASED LATTICE SUBMODULARITY). *For any graph  $\mathcal{G}$ ,  $r_{\mathcal{G}}(\cdot) : \mathbb{N}_0^U \rightarrow \mathbb{R}$  is monotone and submodular, i.e.,  $r_{\mathcal{G}}(\mathbf{x} \wedge \mathbf{y}) + r_{\mathcal{G}}(\mathbf{x} \vee \mathbf{y}) \leq r_{\mathcal{G}}(\mathbf{x}) + r_{\mathcal{G}}(\mathbf{y})$  for any  $\mathbf{x}, \mathbf{y} \in \mathbb{N}_0^U$ , and  $r_{\mathcal{G}}(\mathbf{x}) \leq r_{\mathcal{G}}(\mathbf{y})$  if  $\mathbf{x} \leq \mathbf{y}$ .*

**PROOF.** (**Monotone**) Recall the definition of expected reward  $r_{\mathcal{G}}(\mathbf{b}) = \sum_{v \in \mathcal{V}} g_v \left(1 - \prod_{e=(u,v) \in S} (1 - \mu_{e,b_u})\right)$  and  $\chi_i$  represents the one-hot vector with the  $i$ -th equal to 1 and all other elements equal to 0. Since the fact that the co-branding willingness  $\mu$  does not decrease as budget increases, we have

$$r_{\mathcal{G}}(\mathbf{x} + \chi_j) - r_{\mathcal{G}}(\mathbf{x}) = \sum_{v \in \mathcal{V}} g_v \left( \prod_{e=(j,v), i \neq j} (1 - \mu_{(i,v), \chi_i}) \right) (\mu_{e, \chi_{j+1}} - \mu_{e, \chi_j}) \geq 0,$$

for any  $\mathbf{x} \in \mathbb{N}_0^U$ ,  $j \in \mathcal{U}$ . Then we can use the above inequality repeatedly to show that  $r_{\mathcal{G}}(\mathbf{x}) \leq r_{\mathcal{G}}(\mathbf{y})$  when  $\mathbf{x} \leq \mathbf{y}$ .

(**Submodular**) For submodular property of  $r_{\mathcal{G}}(\mathbf{b})$ , we will first prove that  $(1 - \prod_{e=(u,v) \in S} (1 - \mu_{e,b_u}))$  is submodular for any  $v \in \mathcal{V}$ .

Denote  $h(\mathbf{x}) = \prod_{e=(u,v) \in S} (1 - \mu_{e,x_u})$  for simplicity. We aim to show that  $h(\mathbf{x})$  satisfies the following inequality for any  $\mathbf{x} \in \mathbb{N}_0^U$ ,  $v \in \mathcal{V}$ , and distinct  $l, j \in \mathcal{U}$ :

$$h(\mathbf{x}) - h(\mathbf{x} + \chi_l) \geq h(\mathbf{x} + \chi_j) - h(\mathbf{x} + \chi_l + \chi_j). \quad (6)$$

The proof is as follows. Consider the left-hand side in Eq. (6):

$$h(\mathbf{x}) - h(\mathbf{x} + \chi_l) = \left( \prod_{e \in S \setminus \{(l,v), (j,v)\}} (1 - \mu_{e,x_u}) \right) (1 - \mu_{(j,v), \chi_j}) (\mu_{(l,v), \chi_{l+1}} - \mu_{(l,v), \chi_l}).$$

Similarly, the right-hand side in Eq. (6) is:

$$h(\mathbf{x} + \chi_j) - h(\mathbf{x} + \chi_l + \chi_j) = \left( \prod_{e \in S \setminus \{(l,v), (j,v)\}} (1 - \mu_{e,x_u}) \right) (1 - \mu_{(j,v), \chi_{j+1}}) (\mu_{(l,v), \chi_{l+1}} - \mu_{(l,v), \chi_l}).$$

The common factor in both expressions is:

$$\left( \prod_{e=(u,v) \in S, e \neq (l,v), (j,v)} (1 - \mu_{e,x_u}) \right) (\mu_{(l,v), \chi_{l+1}} - \mu_{(l,v), \chi_l}).$$

Thus, Eq. (6) reduces to comparing:  $(1 - \mu_{(j,v), \chi_j}) \geq (1 - \mu_{(j,v), \chi_{j+1}})$ . Since  $\mu_{e,b_u}$  is non-decreasing in  $b_u$  (i.e.,  $\mu_{(j,v), \chi_{j+1}} \geq \mu_{(j,v), \chi_j}$ ), it follows that:  $1 - \mu_{(l,v), \chi_l} \geq 1 - \mu_{(j,v), \chi_{j+1}}$ .

We invoke the following lemma to formalize the submodularity:

**LEMMA 2** (LATTICE SUBMODULARITY CONDITION). *A function  $f : \mathbb{N}_0^U \rightarrow \mathbb{R}$  is submodular if and only if  $f(\mathbf{x} + \chi_i) - f(\mathbf{x}) \geq f(\mathbf{x} + \chi_j + \chi_i) - f(\mathbf{x} + \chi_j)$ , for any  $\mathbf{x} \in \mathbb{N}_0^U$  and  $i \neq j$ .*

Then Lemma 2 (proved in Appendix B.4) and the holding of Eq. (6) confirm the submodularity of  $h(\mathbf{x}) = (1 - \prod_{e=(u,v) \in S} (1 - \mu_{e,b_u}))$  for any  $v \in \mathcal{V}$ . With the explicit formula of the reward function given by Eq. (2), it shows the reward function is a weighted sum over all  $v \in \mathcal{V}$  with  $g_v$ . Since a non-negative weighted sum of submodular functions is submodular,  $r_{\mathcal{G}}(\mathbf{b})$  is submodular.  $\square$

#### 4.2 Online Learning

**THEOREM 1** (REGRET BOUND). *Algorithm 1 has the following instance-independent sub-linear regret bound  $O(V\sqrt{(NU+1)T \log T} + \log(UVT + VT) \log T)$ , where  $N = \max_{u \in \mathcal{U}} |\mathcal{N}_u|$ .*

**Remark 1.** We also provide an instance-dependent regret analysis for Algorithm 1, as detailed in Appendix B.1. Compared to prior work [4, 26], our base arms expand from  $\mathcal{A}$  to  $\mathcal{A} \cup \mathcal{A}'$ , introducing higher uncertainty and additional regret relative to scenarios with known market gains. Through careful analysis, we incorporate the error term arising from the unknown and uncertain market gain into the base arms. In standard CMAB formulations [5, 7, 27], the set  $\tau$  is typically defined as the set of triggered arms. In our case, this would imply a size of  $NUV + V$ . However, since only the arms in  $\{(e, s) \in \mathcal{A} | b_u = s\}$ , representing actual spending decisions, and their corresponding target brands in  $\mathcal{A}'$  affect the rewards, we redefine  $\tau$  as the set of observed base arms within a determined budget. This modification reduces the size of  $\tau$  to  $UV + V$ , thereby improving the leading  $\ln T$  term in the regret by a factor of  $O(\frac{U+1}{NU+1})$ . Additionally, we treat all past selections of a base arm as a single pull, using the average reward from historical dataset  $\mathcal{D}$ , which bounds the regret from  $\mathcal{D}$  by a time-independent constant.

#### 4.3 Offline Optimization

**THEOREM 2** (APPROXIMATION). *Algorithm 2 achieves a  $(1 - 1/e)$ -approximate solution, i.e., the approximation ratio is  $\alpha = 1 - 1/e$ .*

**Remark 2.** By combining partial enumeration techniques with a greedy algorithm, we exploit the unique properties and requirements of the co-branding optimization problem to achieve a  $(1 - 1/e)$  approximation ratio. This is the best possible solution in polynomial time unless  $P = NP$  [2, 18]. Specifically, for any partial enumeration  $\mathbf{b} \in \mathcal{B}$ , the while loop contains  $B$  iterations in the worst case. The size of the queue  $Q$  is  $O(N^2UV)$ , and line 7 requires  $O(V)$  to calculate  $\delta(u, s, \mathbf{b})$  through pre-computation and update. As a result, the time complexity of any partial enumeration in Algorithm 2 is  $O(BN^2UV^2)$ , which scales linearly with  $U$  and quadratically with  $V$ , avoiding exponential growth. Note that in practice, budget allocations are typically limited to a finite set of candidate plans,  $\mathcal{N}_u$ , which significantly reduces the scale of the problem. For instance, when only three candidate allocation tiers (e.g., low, medium, high) are provided, a single sub-brand would have at most three possible budget allocations, further optimizing computational efficiency. Additionally, for large graphs, the parameter  $K$ , which balances time and performance, can be set to 1 to further accelerate processing.

### 5 Experiments

In this section, we conduct extensive experiments. Specifically, we seek to address the following research questions (RQs):

**RQ1:** In high-uncertainty co-branding recommendations, can our online learning algorithm achieve superior both long-term & short-term revenue performance compared to existing methods?

**RQ2:** Can our unified offline budget allocation strategy across multiple sub-brands significantly enhance overall revenue?

**RQ3:** Across diverse scenarios (e.g., different budget caps, seasons, and tentative plans) can the proposed framework maintain stable performance improvements?

## 5.1 Implementation Details

All experiments were conducted on a MacBook Pro with an M3 8-core CPU. To ensure statistical reliability, each experiment was repeated 10 times, and we reported the average results along with confidence intervals derived from these repetitions.

**Baseline Algorithms.** We select the following algorithms from existing studies as baselines for comparison with Algorithm 1: 1) EMP [19], which allocates budgets and co-brands according to the empirical mean without exploration. 2)  $\epsilon$ -Greedy [31], which distributes the budget uniformly, randomly selecting co-branding pairs with probability  $\epsilon$ , or otherwise allocates budgets according to empirical mean ( $\epsilon$  set to 0.1 in all tests). 3) Thompson Sampling (TS) [36], which uses Beta distribution  $\text{Beta}(\beta_1, \beta_2)$  ( $\beta_1 = \beta_2 = 1$  initially) as a prior for estimating the co-branding bipartite graph model. 4) CUCB [5, 35], a combinatorial bandit algorithm that does not mitigate redundant exploration. In addition, we also specifically compare Algorithm 2 with the following offline algorithms: GBO [2], a greedy algorithm for budget optimization without partial enumeration, and the proportional methods PROP-S and PROP-W [26], which allocate budgets for co-branding based on the number of sub-brands and the market gains, respectively.

**Data Generation and Pre-processing.** Our synthetic dataset follows prior studies [27, 35] and is structured as a  $10 \times 60$  matrix, where rows represent ten sub-brands under a parent brand initiating co-branding, and columns correspond to sixty potential target brands. To evaluate the co-branding strategy under varying budgets, we synthesize collaboration probabilities between sub-brands  $\mathcal{U}$  and target brands  $\mathcal{V}$ , considering both inherent brand compatibility and budget impact. Specifically, we generate a compatibility parameter  $v_{(u,v)} \sim U(-1, 1)$  for each pair  $(u, v)$ , representing baseline affinity, where negative values indicate minimal collaboration intent without budget support. Given a sub-brand  $u$  with budget  $s_u$  (budget allocation scheme detailed later), the probability of successful collaboration is defined as  $\mu_{(u,v),s_u} = \sigma(v_{(u,v)} + s_u)$ , where  $\sigma(x) = 1/(1 + e^{-x})$  [10, 24]. As  $s_u$  increases,  $v_{(u,v)} + s_u$  rises, leading to a higher  $\sigma(\cdot)$ , this ensures that larger budgets enhance partnership success [17]. To reflect market uncertainty where higher investment does not always yield higher returns, we introduce a stochastic factor  $g_v \sim U(0, 1)$  for each  $v \in \mathcal{V}$ .

We further conduct experiments on real-world datasets. As there are no publicly available datasets for co-branding evaluation, we developed a Python-based web crawler using keywords (e.g., “co-branding campaigns,” “brand collaboration cases,” “cross-category partnerships”) to collect 3,500 co-branding cases up to 2024 from SocialBeta [23], a platform dedicated to brand marketing insights. We then categorized these cases by the types of brands involved, constructing three real-world datasets corresponding to the initiators of the co-branding efforts, namely “*diet*,” “*apparel*,” and “*IP-themed*” brands. Each dataset comprises multiple categories of potential partner brands and their historical co-branding frequencies; due to commercial protection policies, these categories represent general groupings rather than specific brands. We then establish a baseline affinity  $v_{(u,v)}$  for each pair of sub-brand  $u$  and target brand  $v$  by calculating the proportion of co-branding occurrences relative to the total cases, and we incorporate standardized market values from

a cloud-native data catalog platform, dataworld<sup>4</sup>, to quantify the market gain  $g$ . In the Diet dataset, the set of initiating brands  $\mathcal{U}$  spans 10 categories totaling 269 entities (with category sizes of 32, 18, 22, 38, 25, 36, 21, 19, 28, and 30) and the set of target brands  $\mathcal{V}$  covers 12 categories totaling 608 entities (64, 56, 52, 44, 58, 49, 72, 40, 38, 42, 28, and 65). The Apparel dataset has  $\mathcal{U}$  split into 12 categories totaling 192 entities (22, 14, 16, 24, 18, 20, 15, 12, 9, 14, 11, 17) and  $\mathcal{V}$  split into 10 categories totaling 471 entities (64, 56, 52, 44, 58, 49, 40, 38, 42, 28). The IP-themed dataset has  $\mathcal{U}$  with 8 categories (17, 24, 26, 12, 18, 27, 23, 14) totaling 161 entities and  $\mathcal{V}$  with 8 categories (64, 56, 72, 40, 38, 42, 28, 65) totaling 405 entities. We then apply the same procedures as used for the synthetic dataset.

**Other Setups.** We set the historical dataset  $\mathcal{D}$  to cover the past  $D = 50$  seasons, similar to [38]. The initial values of  $\mu, g, T_{t,e,s}, T'_{t,v}$  are determined by averaging over this dataset. For the synthetic dataset and the three real-world datasets, we define the base budget  $B_0$  as one-hundredth of the market revenue observed in each dataset’s respective historical dataset  $\mathcal{D}$ . The total budget  $B$  is then set to be ten times  $B_0$  to ensure the profitability of the co-branding campaign. Similarly, the budget cap for any sub-brand  $u$ , denoted as  $c_u$ , is set proportional to its market share, which is defined as the ratio of brand  $u$ ’s revenue to the total market revenue in its historical dataset  $\mathcal{D}$ , scaled by twice the total budget  $B$ . Since the total budget caps exceed  $B$ , some brands may not receive their full allocation  $c_u$ . The budget allocation is chosen from a predefined set of tentative spending plans  $\mathcal{N}_u$ , consisting of three candidate levels: low, medium, and high. Specifically, each sub-brand  $u$  can be allocated one of three budget values:  $\lfloor \frac{1}{3}c_u \rfloor$ ,  $\lfloor \frac{2}{3}c_u \rfloor$ , and  $c_u$ . We determine  $S$  by sorting the predefined levels in  $\mathcal{N}_u$  in increasing order under a given  $b$ . Additionally, we conduct exploratory experiments on different  $B, c_u$ , and  $\mathcal{N}_u$ .

## 5.2 Evaluation Results

**Co-branding Online Performance (RQ1).** We test Algorithm 1 on the received market gain of co-branding pairs, as in Eq. (1). The evaluation spans  $T = 2,000$  rounds, with results shown in Fig. 3. Across all datasets, our algorithm consistently outperforms the baselines, achieving the fastest convergence and improving the online received revenue by 12% to 73% compared to the baselines.

**Offline Budget Allocation (RQ2).** We evaluate Algorithm 2 across different total budgets  $B$  to assess whether our proposed GPE algorithm enhances performance across multiple sub-brands. As shown in Fig. 4, GPE outperforms three baselines GPO, PROP-S, and PROP-W in revenue acquisition within the same  $\mathcal{G}$ , achieving improvements of 13%, 42%, and 29%, respectively, showing its effectiveness from a holistic parent brand perspective.

**Tradeoff Between Performance and Computational Cost (RQ2 & RQ3).** We test the impact of different  $K$  (see line 2 of GPE). As shown in Fig. 5, when  $K$  increases from 1 to 3, the number of possible initial solutions expands, leading to a significant increase in the co-branding revenue. However, when  $K$  reaches or exceeds 4, the revenue gain becomes relatively marginal. On the other hand, Tab. 1 shows that the average running time of different budget settings for  $K = 4$  and  $K = 5$  increases by more than a thousand times compared to  $K = 1$  which aligns with our theoretical analysis.

<sup>4</sup><https://data.world/>, 2025



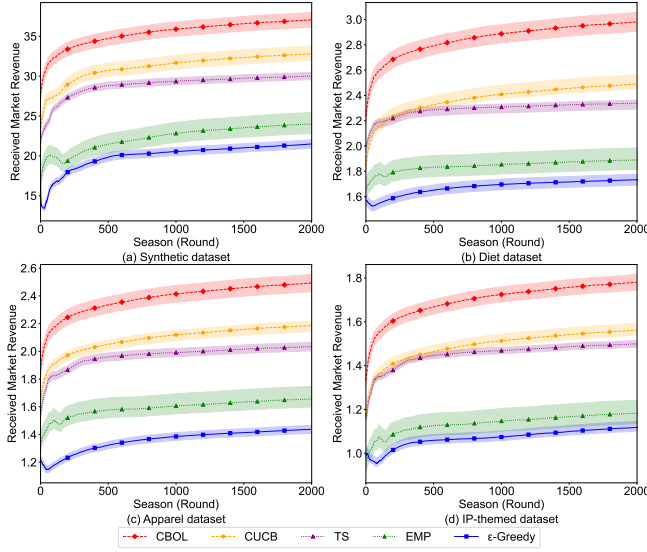


Figure 3: Comparison of average received market revenue.

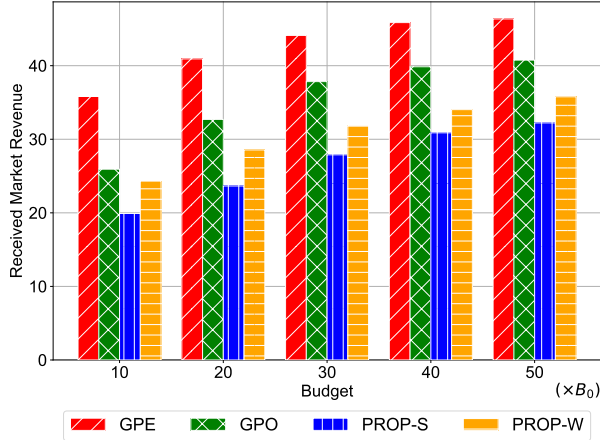
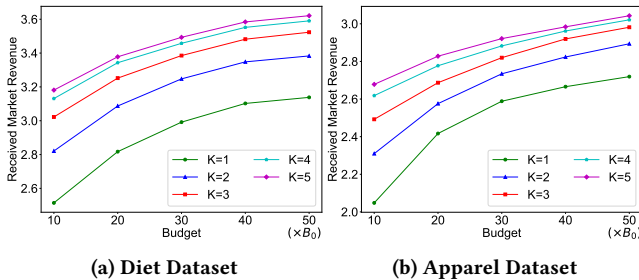


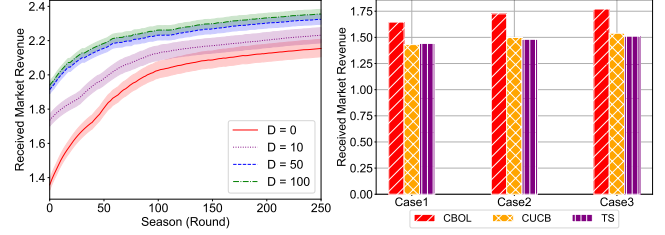
Figure 4: Comparison of offline optimization across different total budgets on Synthetic dataset.

Figure 5: Influence of operational constraint  $K$  under different  $B$  on Diet and Apparel datasets.

**Impact of Historical Dataset (RQ1).** We test the impact of historical dataset  $\mathcal{D}$  of various sizes  $D$  on Apparel dataset. As shown in Fig. 6, the influence of  $\mathcal{D}$  is particularly significant in the early stages of online co-branding seasons compared to  $D = 0$ . This

Table 1: Computational Influence of  $K$  on Diet dataset.

Value of $K$	1	2	3	4	5
Average running time (s)	0.1304	0.3214	0.8921	1.9052	3.5116
Average reward	2.9126	3.1772	3.3329	3.4148	3.4515
Increase of running time	0	146.1%	584.0%	1361.6%	2592.7%

Figure 6: Influence of varying size  $D$  on Apparel dataset. Figure 7: Sensitivity analysis on IP-themed dataset.

highlights its advantage in ensuring short-term algorithmic performance, thus avoiding the performance loss typically associated with the early exploration phase in traditional MAB methods.

**Ablation Study (RQ3).** We evaluate the performance on IP-themed dataset with different cases. Specifically, in Case 1, the budget cap  $c_u$  for each brand  $u$  is set to twice the total budget  $B$ , scaled by  $u$ 's market share, and the tentative spending plan  $\mathcal{N}'$  consists of two levels:  $\lfloor c_u \rfloor$  and  $\lfloor c_u/2 \rfloor$ , with  $T = 500$ . In Case 2,  $c_u$  is increased to three times  $B$ , and  $\mathcal{N}'$  comprises three levels:  $\lfloor c_u \rfloor$ ,  $\lfloor 2c_u/3 \rfloor$ , and  $\lfloor c_u/3 \rfloor$ , with  $T = 1000$ . In Case 3,  $c_u$  is increased to four times  $B$ , and  $\mathcal{N}'$  includes four levels:  $\lfloor c_u \rfloor$ ,  $\lfloor 3c_u/4 \rfloor$ ,  $\lfloor c_u/2 \rfloor$ , and  $\lfloor c_u/4 \rfloor$ , with  $T = 1500$ . We use the two best-performing algorithms from Fig. 3 as baselines, and Fig. 7 shows that Algorithm 1 consistently achieves the best performance, confirming its robustness.

## 6 Conclusion

We presented the first unified online-offline framework for the co-branding problem, systematically tackling challenges such as resource imbalance, uncertain brand willingness, and evolving market trends. Our approach online dynamically updates a co-branding bipartite graph model to balance the exploration of new collaborations with the exploitation of proven ones, while minimizing the high initial costs often associated with co-branding. An offline optimization phase further refines sub-brand budget allocations to maximize overall returns. Through rigorous and highly non-trivial theoretical analysis, we provided regret bounds for the online learning component and approximation guarantees for the offline allocation strategy, offering interpretability and clarity regarding the underlying mechanisms. Empirical results on synthetic and real-world datasets highlight our framework's effectiveness for guiding co-branding strategies. Future work will focus on refining the model to handle even more complex market environments.

## Acknowledgement

The work of Jinhang Zuo was supported by CityUHK 9610706. The work of Xutong Liu was partially supported by a fellowship award from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK PDFS2324-4S04). The work of John C.S. Lui was supported in part by the RGC GRF-14215722.

## References

- [1] Shipra Agrawal, Morteza Zadimoghaddam, and Vahab Mirrokni. 2018. Proportional allocation: Simple, distributed, and diverse matching with high entropy. In *International Conference on Machine Learning*. PMLR, 99–108.
- [2] Noga Alon, Iftah Gamzu, and Moshe Tennenholtz. 2012. Optimizing budget allocation among channels and influencers. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 381–388.
- [3] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-based systems* 46 (2013), 109–132.
- [4] Wei Chen, Yajun Wang, and Yang Yuan. 2013. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*. PMLR, 151–159.
- [5] Wei Chen, Yajun Wang, Yang Yuan, and Qinshi Wang. 2016. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *The Journal of Machine Learning Research* 17, 1 (2016), 1746–1778.
- [6] Xiangxiang Dai, Jin Li, Xutong Liu, Anqi Yu, and John Lui. 2024. Cost-Effective Online Multi-LLM Selection with Versatile Reward Models. *arXiv preprint arXiv:2405.16587* (2024).
- [7] Xiangxiang Dai, Xutong Liu, Jinhang Zuo, Hong Xie, Carlee Joe-Wong, and John C. S. Lui. 2025. Variance-Aware Bandit Framework for Dynamic Probabilistic Maximum Coverage Problem With Triggered or Self-Reliant Arms. *IEEE Transactions on Networking* (2025), 1–12.
- [8] Xiangxiang Dai, Zhiyong Wang, Jize Xie, Xutong Liu, and John CS Lui. 2024. Conversational Recommendation with Online Learning and Clustering on Misspecified Users. *IEEE Transactions on Knowledge and Data Engineering* 36, 12 (2024), 7825–7838.
- [9] Xiangxiang Dai, Zhiyong Wang, Jize Xie, Tong Yu, and John CS Lui. 2024. Online Learning and Detecting Corrupted Users for Conversational Recommendation Systems. *IEEE Transactions on Knowledge and Data Engineering* 36, 12 (2024), 8939–8953.
- [10] Louis Fauray, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. 2020. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*. PMLR, 3052–3060.
- [11] Chongming Gao, Kexin Huang, Jiawei Chen, Yuan Zhang, Biao Li, Peng Jiang, Shiqin Wang, Zhong Zhang, and Xiangnan He. 2023. Alleviating Matthew Effect of Offline Reinforcement Learning in Interactive Recommendation. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2023).
- [12] Dileep Reddy Goda, Vishal Reddy Vadiyala, Sridhar Reddy Yerram, and Suman Reddy Mallipeddi. 2023. Dynamic Programming Approaches for Resource Allocation in Project Scheduling: Maximizing Efficiency under Time and Budget Constraints. *ABC Journal of Advanced Research* 12, 1 (2023), 1–16.
- [13] Sonal Gogri. 2022. Co-branding: A strategic decision in a competitive world. *EPRA International Journal of Economics, Business and Management Studies* 9 (2022), 20–23.
- [14] Bernd Helmig, Jan-Alexander Huber, and Peter Leeflang. 2007. Explaining behavioural intentions toward co-branded products. *Journal of Marketing Management* 23, 3–4 (2007), 285–304.
- [15] Marco Heyden, Vadim Arzamasov, Edouard Fouché, and Klemens Böhm. 2024. Budgeted Multi-Armed Bandits with Asymmetric Confidence Intervals. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1073–1084.
- [16] Dorit S Hochba. 1997. Approximation algorithms for NP-hard problems. *ACM Sigact News* 28, 2 (1997), 40–52.
- [17] Christopher C Johnson et al. 2014. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems* 27, 78 (2014), 1–9.
- [18] Samir Khuller, Anna Moss, and Joseph Seffi Naor. 1999. The budgeted maximum coverage problem. *Information processing letters* 70, 1 (1999), 39–45.
- [19] Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.
- [20] Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. 2016. Contextual combinatorial cascading bandits. In *International conference on machine learning*. PMLR, 1245–1253.
- [21] Zhuohua Li, Maoli Liu, Xiangxiang Dai, and John C.S. Lui. 2025. Demystifying Online Clustering of Bandits: Enhanced Exploration Under Stochastic and Smoothed Adversarial Contexts. In *Proceedings of the International Conference on Learning Representations*.
- [22] Zhuohua Li, Maoli Liu, Xiangxiang Dai, and John CS Lui. 2025. Towards Efficient Conversational Recommendations: Expected Value of Information Meets Bandit Learning. In *Proceedings of the ACM on Web Conference 2025*. 4226–4238.
- [23] SocialBeta Lite. 2024. <https://socialbeta.com>.
- [24] Xutong Liu, Xiangxiang Dai, Xuchuang Wang, Mohammad Hajiesmaili, and John Lui. 2024. Combinatorial Logistic Bandits. *arXiv preprint arXiv:2410.17075* (2024).
- [25] Xutong Liu, Xiangxiang Dai, Jinhang Zuo, Siwei Wang, Carlee Joe-Wong, John Lui, and Wei Chen. 2025. Offline Learning for Combinatorial Multi-armed Bandits. *arXiv preprint arXiv:2501.19300* (2025).
- [26] Xutong Liu, Jinhang Zuo, Xiaowei Chen, Wei Chen, and John CS Lui. 2021. Multi-layered Network Exploration via Random Walks: From Offline Optimization to Online Learning. In *International Conference on Machine Learning*. PMLR, 7057–7066.
- [27] Xutong Liu, Jinhang Zuo, Siwei Wang, Carlee Joe-Wong, John Lui, and Wei Chen. 2022. Batch-size independent regret bounds for combinatorial semi-bandits with probabilistically triggered arms or independent arms. *Advances in Neural Information Processing Systems* 35 (2022), 14904–14916.
- [28] George S Low and Jakki J Mohr. 1998. *Brand managers' perceptions of the marketing communications budget allocation process*. Marketing Science Institute.
- [29] George L Nemhauser and Zev Ullmann. 1969. Discrete dynamic programming and capital allocation. *Management Science* 15, 9 (1969), 494–505.
- [30] Hubert Pun and H Sebastian Heese. 2015. A note on budget allocation for market research and advertising. *International Journal of Production Economics* 166 (2015), 85–89.
- [31] Anupam Singh. 2021. Reinforcement learning based empirical comparison of ucb, epsilon-greedy, and thompson sampling. *Int. J. of Aquatic Science* 12, 2 (2021), 2961–2969.
- [32] Tasuku Soma, Naonori Kakimura, Kazuhiro Inaba, and Ken-ichi Kawarabayashi. 2014. Optimal budget allocation: Theoretical guarantee and efficient algorithm. In *International Conference on Machine Learning*. 351–359.
- [33] Ceyda Paydas Turan. 2022. Deal or deny: The effectiveness of crisis response strategies on brand equity of the focal brand in co-branding. *Journal of Business Research* 149 (2022), 615–629.
- [34] Zongqi Wan, Jialin Zhang, Wei Chen, Xiaoming Sun, and Zhijie Zhang. 2023. Bandit multi-linear DR-submodular maximization and its applications on adversarial submodular bandits. In *International Conference on Machine Learning*. PMLR, 35491–35524.
- [35] Qinshi Wang and Wei Chen. 2017. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*. 1161–1171.
- [36] Siwei Wang and Wei Chen. 2018. Thompson Sampling for Combinatorial Semi-Bandits. In *International Conference on Machine Learning*. 5114–5122.
- [37] Yu Xia, Junda Wu, Tong Yu, Sungchul Kim, Ryan A. Rossi, and Shuai Li. 2023. User-Regulation Deconfounded Conversational Recommender System with Bandit Feedback. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2694–2704.
- [38] Lily Xu, Elizabeth Bondi, Fei Fang, Andrew Perrault, Kai Wang, and Milind Tambe. 2021. Dual-mandate patrols: Multi-armed bandits for green security. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 14974–14982.
- [39] Shuhua Yang, Hui Yuan, Xiaoying Zhang, Mengdi Wang, Hong Zhang, and Huazheng Wang. 2024. Conversational Dueling Bandits in Generalized Linear Models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3806–3817.

## A Model Extensions

### A.1 Incorporating Sub-Brand Market Gains

In the main text, our model considers brand market gain as the impact of the target brand on the overall profitability of the parent company’s brand portfolio, without distinguishing specific sub-brands. Here, we extend this concept to analyze the influence of the target brand on individual sub-brands. Specifically, the graph  $\mathcal{G}$  incorporates the brand market gain vector  $\mathbf{g} = \{g_e\}_{e \in \mathcal{G}}$ , representing the revenue earned by the parent brand from co-branding collaborations. Each element  $g_e \in \mathbf{g}$  denotes the consumer market gain for partner brand pair  $e$ . These gains are initially unknown and depend on the parent brand’s financial investments in marketing and promotion. Additionally, for each successful co-branding pair  $e := (u, v)$ , i.e.,  $X_{t,e} = 1$ , the market gain  $Y_{t,e}$  is observed based on the budget allocation  $\mathbf{b}$  for  $e$  during season  $t$ . The expected value of  $Y_{t,e}$  is the brand market gain  $g_e$ , quantifying the economic benefits of collaboration. This feedback component is observable only after successful co-branding actions. We denote the set of successful co-branding pair as  $S'_t \in S_t$ , with corresponding brand market gain feedback denoted as  $Y_{t,S'_t} = (Y_{t,1}, \dots, Y_{t,|S'_t|})$ .

Therefore, when action  $S_t$  is determined under a given budget allocation, the decision-maker receives reward  $R_{\mathcal{G}}(\mathbf{b}_t)$  determined by both co-branding success probability and market gain:  $R_{\mathcal{G}}(\mathbf{b}_t) = \sum_{e \in S_t} \mathbb{I}\{\exists e \in S_t \text{ s.t. } X_{t,e} = 1\} Y_{t,e}$ . Given the definitions of  $\mu$  and  $\mathbf{g}$ , the expected reward  $r_{\mathcal{G}}(\mathbf{b})$  is given by:  $r_{\mathcal{G}}(\mathbf{b}) = \sum_{e \in S} g_e \mu_e$ . This formulation reveals a linear summation. Consequently, our algorithm remains applicable to this simpler model and its formulation. Specifically, the elements of  $\mathcal{A}'$  are modified to include  $e \in \mathcal{A}'$ . Lines 4 and 8 in Algorithm 1 are updated as follows:  $\rho'_e \leftarrow \sqrt{3 \ln t / (2T'_{t,e})}$ ,  $\hat{g}_e \leftarrow \hat{g}_e + \rho'_e$  and  $T'_e \leftarrow T'_{t,e} + 1$ ,  $\hat{g}_e \leftarrow \hat{g}_e + (Y_e - \hat{g}_e) / T'_{t,e}$ ,  $\hat{V}'_e \leftarrow \frac{T'_{t,e}-1}{T'_{t,e}} \left( \hat{V}'_e + \frac{1}{T'_{t,e}} (\hat{g}_e - Y_{t,e})^2 \right)$ , respectively. These updates allow the algorithm to incorporate brand market gains effectively in this extended scenario.

## A.2 Target Brand Exclusivity

We consider scenarios where multiple sub-brands  $u$  under the same parent company  $\mathcal{U}$  can collaborate with any target brand  $v \in \mathcal{V}$  without exclusivity agreements in the main text. This implies that a target brand  $v$  is not restricted to partnering with a single specific sub-brand from  $\mathcal{U}$ . For example, target brand  $v_1$  can collaborate with both  $u_1$  and  $u_3$ , without violating exclusivity agreements. In contrast, we also consider a simplified yet commonly encountered case where each target brand is limited to partnering with a single sub-brand from a given parent company. In this context, the reward function can be decomposed into separate components on each source sub-brand, and the expected reward can be expressed as:  $r_{\mathcal{G}}(\mathbf{b}) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}_u} g_u \mu_{u,v,b_u}$ , where  $\mathcal{V}_u$  represents the set of target brands associated with the  $u$ -th sub-brand. In the absence of exclusivity agreements, multiple base arms (i.e., co-branding pairs) may share resources or be influenced by the same target brand's market performance. This overlap introduces complexities in coordination and estimation, as the interactions between different co-branding pairs become interdependent. Conversely, when a target brand can collaborate with only one sub-brand from a parent brand, the marginal gains of each base arm become independent. This independence eliminates the need for complex joint distribution modeling, thereby simplifying the budget allocation process. Under this exclusivity constraint, classical dynamic programming techniques [12, 29] can be employed to achieve the optimal offline budget allocation, ensuring an approximation ratio of  $\alpha = 1$ .

## B Proof Appendix

We begin by introducing some definitions. Recall that  $\mathcal{A} \cup \mathcal{A}'$  represents the set of all base arms. Without loss of generality, assume  $X_{t,e} \in [0, 1]$ ,  $Y_{t,v} \in [0, 1]$  for any  $t$ . Following the notation in [35] and with slight abuse of notation, we use  $X_a \in [0, 1]$  to consistently represent both  $X_e$  and  $Y_v$  as the random outcome of base arm  $a \in \mathcal{A} \cup \mathcal{A}'$ . Define  $D$  as the unknown joint distribution of random outcomes  $\mathbf{X} = (X_a)_{a \in \mathcal{A} \cup \mathcal{A}'}$  with unknown mean  $\mu_a = \mathbb{E}_{\mathbf{X} \sim D} [X_a]$ . The probability that a budget allocation  $\mathbf{b}$  triggers arm  $a$  (i.e.  $X_a$  is observed) under the unknown environment distribution  $D$  is denoted as  $p_a^b$  the probability that budget allocation. For simplicity, we omit the dependency on  $D$  when there is no ambiguity. Specifically,  $p_a^b = 1$  for  $a \in \mathcal{A}$ ,  $p_a^b = 1 - \prod_{e \in \mathcal{E}} (1 - \mu_{e,a,s_u})$  for  $a \in \mathcal{A}'$ , and  $p_a^b = 0$  for the rest of base arms. We define the reward gap

$\Delta_{\mathbf{b}}$  as  $\max(0, \text{arg}(\mathbf{b}^*) - r_{\mathcal{G}}(\mathbf{b}))$  for all feasible budget allocation  $\mathbf{b}$ . For each base arm  $a$ , we further define  $\Delta_{\min}^a = \min_{\Delta_{\mathbf{b}} > 0, b_u = s} \Delta_{\mathbf{b}}$  and  $\Delta_{\max}^a = \max_{\Delta_{\mathbf{b}} > 0, b_u = s} \Delta_{\mathbf{b}}$ . If there is no budget allocation  $\mathbf{b}$  with  $b_u = s$  such that  $\Delta_{\mathbf{b}} > 0$ , we adopt the convention:  $\Delta_{\min}^a = \infty$  and  $\Delta_{\max}^a = 0$ . Let  $\Delta_{\min} = \min_{a \in \mathcal{A} \cup \mathcal{A}'} \Delta_{\min}^a$  and  $\Delta_{\max} = \max_{a \in \mathcal{A} \cup \mathcal{A}'} \Delta_{\max}^a$ . Define the set of observed base arms at  $t$  as  $\tau_t$ , which includes both co-branding pairs with allocated budgets  $\{(e, s) \in \mathcal{A} | b_u = s\}$  and the corresponding arms in  $\mathcal{A}'$ .

### B.1 Instance-Dependent Regret Analysis

**LEMMA 3.** *Algorithm 1 has the following instance-dependent regret bound  $O\left((NUV + V) \left(\frac{\log(UV+V)}{\Delta_{\min}^a} + \log\left(\frac{UV+V}{\Delta_{\min}^a}\right)\right) \log T\right)$ .*

**PROOF.** Recall that  $S = \{(e, s) \in \mathcal{A} | s = b_u, s \in \mathcal{N}_u\}$  represents the set of co-branding pairings under a specific budget allocation. Let  $\tilde{S}$  denote the set of base arms that can be triggered by  $S$ , i.e.,  $\{a \in \mathcal{A} \cup \mathcal{A}' | p_a^b > 0\}$ . Using the Lemma 9 from [27], we decompose the regret  $\text{Reg}(T)$  into two event-filtered components,  $\text{Reg}(T, E_{t,1})$  and  $\text{Reg}(T, E_{t,2})$ :  $\text{Reg}(T) \leq \text{Reg}(T, E_{t,1}) + \text{Reg}(T, E_{t,2})$ , where event  $E_{t,1} = \{\Delta_{\mathbf{b}_t} \leq 2e_{t,1}(\mathbf{b}_t)\}$ , event  $E_{t,2} = \{\Delta_{\mathbf{b}_t} \leq 2e_{t,2}(\mathbf{b}_t)\}$ , with error terms defined as  $e_{t,1}(\mathbf{b}_t) = 4\sqrt{3.75V} p_a^b \sqrt{\sum_{a \in \tilde{S}_t} \left(\frac{\log t}{T_{t-1,a}} \wedge \frac{1}{28}\right)}$ ,  $e_{t,2}(\mathbf{b}_t) = 28 \sum_{a \in \tilde{S}_t} p_a^b \left(\frac{\log t}{T_{t-1,a}} \wedge \frac{1}{28}\right)$ . Since the confidence radius for each base arm does not incorporate feedback from the online market, we treat all past selections of a base arm as a single pull. The feedback for this single selection can be reviewed as the average reward over all historical pulls. Consequently, regardless of the distribution of the finite historical data  $\mathcal{D}$ , the regret from historical arm selections is bounded by a time-independent constant.

For  $\text{Reg}(T, E_{t,1})$ , denote  $\tilde{\Delta}_{t,b_t} = \frac{\Delta_{t,b_t}}{p_a^b}$  for budget allocation  $\mathbf{b}_t$  at  $t$ , and  $\tilde{\Delta}_{\min}^a = \min_{\mathbf{b}: p_a^b > 0, \Delta_{t,b} > 0} \Delta_{t,b} / p_a^b$ . Regarding  $\Delta_{\mathbf{b}_t}$  on the selected budget allocation  $\mathbf{b}_t$  at  $t$ , we have:  $\Delta_{\mathbf{b}_t} \leq \sum_{a \in \mathcal{A} \cup \mathcal{A}'} \frac{4(4\sqrt{3})^2 V (p_a^b)^2 \frac{\log t}{T_{t-1,a}}}{\Delta_{\mathbf{b}_t}} \leq \sum_{a \in \mathcal{A} \cup \mathcal{A}'} p_a^b \left( \frac{384V \frac{\log t}{T_{t-1,a}}}{\Delta_{\mathbf{b}_t}} - \frac{\tilde{\Delta}_{t,b_t}}{UV+V} \right)$ . Using a modified version of Lemma 10 in [27], we define a regret allocation function as follows:  $\kappa_{a,T}(\ell) = \frac{48V}{\Delta_{\min}^a}$  for  $\ell = 0$ ;  $2\sqrt{\frac{192V \log T}{\ell}}$  for  $1 \leq \ell \leq \frac{192V \log T}{(\Delta_{\min}^a)^2}$ ;  $\frac{384V \log T}{\ell \Delta_{\min}^a}$  for  $\frac{192V \log T}{(\Delta_{\min}^a)^2} < \ell \leq \frac{384V^2 (U+1) \log T}{(\Delta_{\min}^a)^2}$ ; 0 otherwise. Using  $\Delta_{\mathbf{b}_t} = \mathbb{E}_t[\Delta_{\mathbf{b}_t}]$ , we obtain:  $\Delta_{\mathbf{b}_t} = \mathbb{E}_t[\Delta_{\mathbf{b}_t}] \leq \mathbb{E}_t \left[ \sum_{a \in \mathcal{A} \cup \mathcal{A}'} \left( \frac{384V \frac{\log t}{T_{t-1,a}}}{\Delta_{\mathbf{b}_t}} - \frac{\tilde{\Delta}_{t,b_t}}{UV+V} \right) p_a^b \right] \leq \mathbb{E}_t \left[ \sum_{a \in \tau_t} \kappa_{a,T}(T_{t-1,a}) \right]$ . By the tower rule and the fact that  $T_{t-1,a}$  is increased by 1 if and only if  $a \in \tau_t$ , we can derive:  $\mathbb{E} \left[ \sum_{t \in [T]} \mathbb{E}_t \left[ \sum_{a \in \tau_t} \kappa_{a,T}(T_{t-1,a}) \right] \right] = \mathbb{E} \left[ \sum_{a \in \mathcal{A} \cup \mathcal{A}'} \sum_{j=0}^{T_{t-1,a}} \kappa_{a,T}(j) \right]$ . Applying the definition of  $\kappa_{a,T}(\ell)$ , the expected regret  $\text{Reg}(T, E_{t,1}) = \mathbb{E} \left[ \sum_{t=1}^T \Delta_{\mathbf{b}_t} \right]$  under event  $E_{t,1}$  is bounded as:  $\text{Reg}(T, E_{t,1}) \leq \sum_{a \in \mathcal{A} \cup \mathcal{A}'} \left( \frac{48V}{\Delta_{\min}^a} + \frac{384V \log T}{\Delta_{\min}^a} (1 + \log(UV + V)) + \frac{768V \log T}{\Delta_{\min}^a} \right)$ . Here,  $|\mathcal{A}| = NUV$ ,  $|\mathcal{A}'| = V$ .

In the analysis of  $\text{Reg}(T, E_{t,2})$  under event  $E_{t,2}$ , similar to that of  $\text{Reg}(T, E_{t,1})$ , we can derive:  $\Delta_{\mathbf{b}_t} \leq \sum_{a \in \tilde{S}_t} 56p_a^b \min \left\{ 1/28, \frac{\log T}{T_{t-1,a}} \right\} \leq -\Delta_{\mathbf{b}_t} + 2 \sum_{a \in \tilde{S}_t} 56p_a^b \min \left\{ 1/28, \frac{\log T}{T_{t-1,a}} \right\} \leq \sum_{a \in \mathcal{A} \cup \mathcal{A}'} p_a^b \left( - \right.$

$\frac{\Delta_{b_t}}{UV+V} + 112 \min \left\{ \frac{1}{28}, \frac{\log T}{T_{t-1,a}} \right\}$ . Different from [27] that using  $p_a^{b_t} = \mathbb{E}_t[\mathbb{I}\{a \in \tau_t\}]$ , we directly establish the probability equivalence:  $\mathbb{E}_t \left[ \sum_{a \in \mathcal{A} \cup \mathcal{A}'} p_a^{b_t} \left( -\frac{\Delta_{b_t}}{UV+V} + 112 \min \left\{ \frac{1}{28}, \frac{\log T}{T_{t-1,a}} \right\} \right) \right] = \mathbb{E}_t \left[ \sum_{a \in \tau_t} \left( -\frac{\Delta_{b_t}}{UV+V} + 112 \min \left\{ \frac{1}{28}, \frac{\log T}{T_{t-1,a}} \right\} \right) \right]$ , which minimizes the observation randomness. It then follows that:  $\Delta_{b_t} \leq \mathbb{E}_t \left[ \sum_{a \in \tau_t} \kappa_{a,T}(T_{t-1,a}) \right]$ , where the regret allocation function  $\kappa_{a,T}(\ell)$  is adjusted as:  $\kappa_{a,T}(\ell) = \Delta_a^{\max}$  if  $0 \leq \ell \leq \frac{112 \log T}{\Delta_a^{\max}}$ ;  $\frac{112 \log T}{\ell}$  if  $\frac{112 \log T}{\Delta_a^{\max}} < \ell \leq \frac{112(UV+V) \log T}{\Delta_a^{\min}}$ ; 0 if  $\ell > \frac{112(UV+V) \log T}{\Delta_a^{\min}} + 1$ . Consequently, the expected regret  $\text{Reg}(T, E_{t,2})$  under event  $E_{t,2}$  can be bounded as:  $\text{Reg}(T, E_{t,2}) \leq \mathbb{E} \left[ \sum_{t \in [T]} \mathbb{E}_t \left[ \sum_{a \in \tau_t} \kappa_{a,T}(T_{t-1,a}) \right] \right] = \mathbb{E} \left[ \sum_{a \in \mathcal{A} \cup \mathcal{A}'} \sum_{s=0}^{T_{t-1,a}} \kappa_{a,T}(s) \right] \leq \sum_{a \in \mathcal{A} \cup \mathcal{A}'} 112 \left( 1 + \log \left( \frac{(UV+V) \Delta_a^{\max}}{\Delta_a^{\min}} \right) \right) \log T + UV \Delta_{\max}$ .  $\square$

## B.2 Proof of Theorem 3

Building on the instance-dependent result in Lemma 1, we now present the proof of the instance-independent regret bound in Theorem 3 for Algorithm 1.

**PROOF.** Since the minimum gap  $\Delta_{\min}^a$  can be extremely small in some instances, we derive regret bounds for Algorithm 1 that are independent of specific co-branding problem instances and do not directly rely on  $\Delta_{t,b}$ . To achieve this, we introduce a fixed gap parameter  $\Delta$ , the value of which will be specified later. Based on Lemma 1, We analyze the regret under two scenarios:  $\{\Delta_{t,b_t} \leq \Delta\}$  and  $\{\Delta_{t,b_t} > \Delta\}$ . For the first scenario where  $\Delta_{t,b_t} \leq \Delta$ , the regret is trivially upper bounded by:  $\text{Reg}(T, \{\Delta_{t,b_t} \leq \Delta\}) \leq T\Delta$ . For the second scenario where  $\Delta_{t,b_t} > \Delta$ , we simplify the regret analysis by approximating all minimal gaps  $\Delta_{\min}^a$  with  $\Delta$ . This results in the following bound:  $\text{Reg}(T, \{\Delta_{t,b_t} > \Delta\}) \leq O((NUV+V) \log T (\frac{V}{\Delta} \log(UV+V) + \log(\frac{UV+V}{\Delta})))$ . To balance these two scenarios, we carefully choose  $\Delta$  as:  $\Delta = \Theta(\sqrt{\frac{(NU+1)V^2 \log T \log(UV+V)}{T}} + \frac{(NUV+V) \log(UV+V) \log T}{T})$ . By substituting this choice of  $\Delta$ , we derive a new regret bound for Algorithm 1 that is independent of the variations in gaps:  $\text{Reg}(T) \leq O(V\sqrt{(NU+1)(\log(U+1)V)T \log T} + \log((U+1)VT) \log T)$ . This formulation ensures that the regret bound does not depend on specific gap values, making it more robust to variations in different co-branding instances.  $\square$

## B.3 Proof of Theorem 2

**PROOF.** Consider a starting value given by a partial solution  $b^1 \in \mathbb{N}_0^U$ . We proceed by reordering the optimal solution  $b^*$  based on a non-increasing ordering of marginal gains. Specifically, the pair  $(i_1^*, s_1^*)$  is chosen such that its marginal gain concerning the empty set is the highest among all possible pairs. Similarly, the pair  $(i_2^*, s_2^*)$  is selected such that its marginal gain, evaluated concerning the solution containing only the pair  $(i_1^*, s_1^*)$ , is the highest among the remaining pairs, and so forth. In formal terms, the optimal solution can be expressed as  $b^* = \{(i_1^*, s_1^*), \dots, (i_U^*, s_U^*)\}$ , where each index  $i_u^* \in U \setminus \{i_1^*, \dots, i_{u-1}^*\}$  is selected to maximize the following expression:  $\left( r_{\mathcal{G}} \left( \sum_{j=1}^{u-1} s_j^* \chi_{i_j^*} + s^u \chi_{i_u^*} \right) - r_{\mathcal{G}} \left( \sum_{j=1}^{u-1} s_j^* \chi_{i_j^*} \right) \right)$ . Next, we aim to bound the “virtual” marginal gain  $\Delta^i = r_{\mathcal{G}}(b^{i+1}) - r_{\mathcal{G}}(b^i)$ .

Recall that  $i$  represents the first iteration where the current solution cannot be extended. Without loss of generality, we assume the virtual pair  $(u^i, s^i)$  contributes to improving  $b^i$  towards the optimal budget allocation, i.e.,  $s^i + b_{u^i}^i \leq b_{u^i}^*$ . Otherwise, if  $s^i + b_{u^i}^i > b_{u^i}^*$ , we can safely remove  $(u^i, s^i)$  from the queue  $Q$  without affecting the greedy solution, the optimal solution, or the analysis.

Starting from the initial solution  $b^1$  such that  $b^1$  matches  $(i_1^*, s_1^*)$ ,  $(i_2^*, s_2^*)$ , and  $(i_3^*, s_3^*)$ , i.e.,  $b_{i_u}^1 = s_i^*$  for  $i = 1, 2, 3$ , we aim to bound  $\Delta^i = r_{\mathcal{G}}(b^{i+1}) - r_{\mathcal{G}}(b^i) \leq r_{\mathcal{G}}(b^i \vee b_{u^i}^* \chi_{u^i}) - r_{\mathcal{G}}(b^i) \leq r_{\mathcal{G}}(b_{u^i}^* \chi_{u^i}) - r_{\mathcal{G}}(0) \leq r_{\mathcal{G}}(s_1^* \chi_{i_1^*})$ . Using marginal gain properties, we have:  $\Delta^i = r_{\mathcal{G}}(b^{i+1}) - r_{\mathcal{G}}(b^i) \leq r_{\mathcal{G}}(b^i \vee b_{u^i}^* \chi_{u^i}) - r_{\mathcal{G}}(b^i) = r_{\mathcal{G}}(b^i \vee s_1^* \chi_{i_1^*} \vee b_{u^i}^* \chi_{u^i}) - r_{\mathcal{G}}(b^i \vee s_1^* \chi_{i_1^*}) \leq r_{\mathcal{G}}(s_1^* \chi_{i_1^*} \vee b_{u^i}^* \chi_{u^i}) - r_{\mathcal{G}}(s_1^* \chi_{i_1^*}) \leq r_{\mathcal{G}}(s_1^* \chi_{i_1^*} + s_2^* \chi_{i_2^*}) - r_{\mathcal{G}}(s_1^* \chi_{i_1^*})$ . By repeating the above reasoning, we also have  $\Delta^i \leq r_{\mathcal{G}}(s_1^* \chi_{i_1^*} + s_2^* \chi_{i_2^*} + s_3^* \chi_{i_3^*}) - r_{\mathcal{G}}(s_1^* \chi_{i_1^*} + s_2^* \chi_{i_2^*})$ . Adding these inequalities together yields:  $\Delta^i \leq r_{\mathcal{G}}(b^1)/3$ . Now, applying Lemma 9 from [26] with  $l = i$  and noting that  $\sum_{j=1}^u s_j^i \geq B$ , we establish:  $r_{\mathcal{G}}(b^{i+1}) \geq r_{\mathcal{G}}(b^1) + (1 - 1/e)(r_{\mathcal{G}}(b^*) - r_{\mathcal{G}}(b^1))$ . Combining this with  $\Delta^i \leq r_{\mathcal{G}}(b^1)/3$ , we derive:  $r_{\mathcal{G}}(b^i) \geq (1 - 1/3)r_{\mathcal{G}}(b^1) + (1 - 1/e)(r_{\mathcal{G}}(b^*) - r_{\mathcal{G}}(b^1)) \geq (1 - 1/e)r_{\mathcal{G}}(b^*)$ . Note that if  $K < 3$  (i.e.,  $1/K > 1/3$ ), the inequality does not hold, and the  $1 - 1/e$  approximation ratio cannot be achieved. Therefore, we set  $K = 3$  to ensure the approximate ratio while minimizing the time complexity as much as possible.  $\square$

## B.4 Proof of Lemma 2

**PROOF. “If” Direction.** We assume that  $f(x + \chi_i) - f(x) \geq f(x + \chi_j + \chi_i) - f(x + \chi_j)$  holds for all  $x \in \mathbb{N}_0^U$  and distinct  $i, j \in \mathcal{U}$ . Our goal is to prove that  $f(x + \chi_i) - f(x) \geq f(y + \chi_i) - f(y)$  for any  $x \leq y$  (i.e.,  $x_i \leq y_i$  for all  $i \in \mathcal{U}$ ) and  $i \in \mathcal{U}$  such that  $x_i = y_i$ .

Define  $I_0 = \{i \in \mathcal{U} : x_i = y_i\}$  and  $I_1 = \{i \in \mathcal{U} : x_i < y_i\}$ . Since  $x \leq y$ , we can write  $y = x + \sum_{j=1}^s \alpha_j \chi_{i_j}$ , where  $I_1 = \{i_1, \dots, i_s\}$  and  $\alpha_j = y_{i_j} - x_{i_j} \geq 1$ . For any  $i \in I_0$ , we start with  $f(x + \chi_i) - f(x)$ . Using the assumption with  $j = i_1$ , we have  $f(x + \chi_i) - f(x) \geq f(x + \chi_{i_1} + \chi_i) - f(x + \chi_{i_1})$ . Applying the assumption again to  $x + \chi_{i_1}$ , we get  $f(x + \chi_{i_1} + \chi_i) - f(x + \chi_{i_1}) \geq f(x + 2\chi_{i_1} + \chi_i) - f(x + 2\chi_{i_1})$ . We continue this process incrementally, adding  $\alpha_1 \chi_{i_1}$ , then  $\alpha_2 \chi_{i_2}$ , and so on, up to  $\alpha_s \chi_{i_s}$ , arriving at  $f(x + \sum_{j=1}^s \alpha_j \chi_{i_j} + \chi_i) - f(x + \sum_{j=1}^s \alpha_j \chi_{i_j}) = f(y + \chi_i) - f(y)$ . Thus, the chain of inequalities gives  $f(x + \chi_i) - f(x) \geq f(y + \chi_i) - f(y)$ . Next, for  $i \in I_0$  and any  $a \in \mathbb{N}_0^U$ , we express  $f(x + a\chi_i) - f(x) = \sum_{k=1}^a [f(x + k\chi_i) - f(x + (k-1)\chi_i)]$  and similarly  $f(y + a\chi_i) - f(y) = \sum_{k=1}^a [f(y + k\chi_i) - f(y + (k-1)\chi_i)]$ . For each  $k$ , since  $x + (k-1)\chi_i \leq y + (k-1)\chi_i$  with equal  $i$ -th coordinates, the previous result implies  $f(x + k\chi_i) - f(x + (k-1)\chi_i) \geq f(y + k\chi_i) - f(y + (k-1)\chi_i)$ . Summing over  $k$  from 1 to  $a$ , we obtain  $f(x + a\chi_i) - f(x) \geq f(y + a\chi_i) - f(y)$ . By applying the submodularity equivalence from Lemma 2.2 in [32], submodularity follows from the diminishing returns property, thus establishing the “if” direction.

**“Only If” Direction.** Assume  $f$  is submodular, meaning  $f(a) + f(b) \geq f(a \vee b) + f(a \wedge b)$  for all  $a, b \in \mathbb{N}_0^U$ . Let  $a = x + \chi_i$  and  $b = x + \chi_j$  for distinct  $i, j \in \mathcal{U}$ . Then,  $a \vee b = x + \chi_i + \chi_j$  and  $a \wedge b = x$ . By submodularity,  $f(x + \chi_i) + f(x + \chi_j) \geq f(x + \chi_i + \chi_j) + f(x)$ , which rearranges to  $f(x + \chi_i) - f(x) \geq f(x + \chi_i + \chi_j) - f(x + \chi_j)$ , as required.  $\square$