Multi-layered Network Exploration via Random Walks: From Offline Optimization to Online Learning

Xutong Liu [1] **Jinhang Zuo** [2] **Xiaowei Chen** [3] **Wei Chen** [4] **John C.S. Lui** [1]

Abstract

Multi-layered network exploration (MuLaNE) problem is an important problem abstracted from many applications. In MuLaNE, there are multiple network layers where each node has an importance weight and each layer is explored by a random walk. The MuLaNE task is to allocate total random walk budget B into each network layer so that the total weights of the unique nodes visited by random walks are maximized. We systematically study this problem from offline optimization to online learning. For the offline optimization setting where the network structure and node weights are known, we provide greedy based constant-ratio approximation algorithms for overlapping networks, and greedy or dynamic-programming based optimal solutions for non-overlapping networks. For the online learning setting, neither the network structure nor the node weights are known initially. We adapt the combinatorial multi-armed bandit framework and design algorithms to learn random walk related parameters and node weights while optimizing the budget allocation in multiple rounds, and prove that they achieve logarithmic regret bounds. Finally, we conduct experiments on a real-world social network dataset to validate our theoretical results.

1. Introduction

Network exploration is a fundamental paradigm of searching/exploring in order to discover information and resources available at nodes in a network, and random walk is often used as an effective tool for network exploration (Lv et al., 2002; Gleich, 2015; Wilder et al., 2018). In this paper, we study the multi-layered network exploration via random walks problem, which can model many real-world applications, including resource searching in peer-to-peer (P2P) networks and web surfing in online social networks (OSNs).

In P2P networks, a user wants to find resources that are scattered on nodes (i.e. peers) in different P2P networks via some platform-specified strategies. A commonly used search strategy is based on multiple random walks (Lv et al., 2002). Since different resources have different importance, the search quality of different random walkers varies. Moreover, the resource-search process typically has a life span, i.e., a total time-limit or hop-limit for random walks. So the user's goal is to decide how to allocate limited budgets to different random walkers to find as many important resources as possible. Another application is the web surfing, where users want to find information in different OSNs, by looking at posts via others' home-pages (Lerman & Jones, 2006). Once the user arrives at one of his friends' home-pages in a particular OSN, he could find some information and continues to browse the home-pages of his friends' friends, which can be regarded as a random walk process. Since the user only has a finite duration in web surfing, the similar question is how to allocate his time in exploring different OSNs so to get the maximum amount of useful information.

We abstract the above application scenarios as the **Mu**lti-**La**yered **N**etwork **E**xploration (MuLaNE) problem. In MuLaNE, we model the overall network to be explored (e.g., combining different OSNs) as a multi-layered network $\mathcal{G}$ that consists of m layers $L_1, ..., L_m$. Each layer L_i (e.g., a single OSN) is represented by a weighted directed graph $\mathcal{G}_i(\mathcal{V}_i, \mathcal{E}_i, w_i)$, where $\mathcal{V}_i$ is the set of nodes to be explored (e.g., users' home-pages with importance weight σ_u for $u \in \mathcal{V}_i$), $\mathcal{E}_i \subseteq \mathcal{V}_i \times \mathcal{V}_i$ is the set of directed edges (e.g., social links), and w_i is the edge weight function on edges $\mathcal{E}_i$. Each layer L_i is associated with an explorer (or random walker) W_i and a fixed starting distribution $\boldsymbol{\alpha}_i$ on nodes $\mathcal{V}_i$. Explorer W_i starts on a node in $\mathcal{G}_i$ following the distribution $\boldsymbol{\alpha}_i$, and then walks on network $\mathcal{G}_i$ following outgoing edges with probability proportional to edge weights. We assume that each random walk step will cost one unit of the bud-

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China [2]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA [3]Bytedance, Mountain View, CA, USA [4]Microsoft Research, Beijing, China. Correspondence to: Xutong Liu <liuxt@cse.cuhk.edu.hk>, Wei Chen <weic@microsoft.com>, John C.S. Lui <cslui@cse.cuhk.edu.hk>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

get[1]. Given a total budget constraint B and individual layer budget constraint c_i, the MuLaNE task is to find the optimal budget allocation $\boldsymbol{k}=(k_1,...,k_m)$ with $\sum_{i=1}^{m} k_i \leq B$, $0\leq k_i \leq c_i$ such that the expected total weights of *unique* nodes visited are maximized.

In real applications, the network structure $\mathcal{G}$, the node weights $\boldsymbol{\sigma}$ and the starting distributions $\boldsymbol{\alpha}_i$'s may not be known in advance, so we consider both the *offline optimization* cases when $\mathcal{G}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}_i$'s are known and the *online learning* case when $\mathcal{G}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}_i$'s are unknown. Moreover, different layers may have overlapping vertices (e.g., homepage of the same user may appear in different OSNs), and the starting distributions $\boldsymbol{\alpha}_i$'s may or may not be stationary distributions. In this paper, we provide a systematic study of all these case combinations.

For the offline optimization, we first consider the general overlapping setting and provide constant approximation algorithms based on the lattice submodularity property for non-stationary $\boldsymbol{\alpha}_i$'s and the diminishing-return (DR) submodularity property for stationary $\boldsymbol{\alpha}_i$'s. For the special non-overlapping setting, we design a dynamic programming algorithm that finds the exact optimal solution for MuLaNE.

In the online learning, we conduct multiple rounds of exploration, each of which has the same budget constraints B and c_i's. After each round of exploration, the total weights of unique nodes visited in this round are the reward for this round, and the trajectory of every explorer in every layer and the importance weights of visited nodes are observed as the feedback, which could be used to learn information about the network for the benefit of future explorations.

We adapt the combinatorial multi-armed bandit (CMAB) framework and the CUCB algorithm (Chen et al., 2016) to our setting, and design online learning algorithms to minimize the regret, which is the difference between the cumulative reward achieved by the exact or approximate offline algorithm and that achieved by our learning algorithm, over T rounds. We show that directly learning the graph structure $\mathcal{G}$ is inefficient. Instead, we define intermediate random variables in MuLaNE as the base arms in CMAB and learn these intermediate parameters, which can sufficiently determine the rewards to guide our budget allocation. Moreover, the node weight is not revealed until this node is first visited, which further complicates the design of online exploration algorithms. For the overlapping case, we adapt the CUCB algorithm to address the unrevealed node weights and the extra constraint of monotonicity for the intermediate parameters. We further improve the analysis by leveraging on special properties in our setting and show logarithmic regret bounds in this case. For the non-overlapping case, we define more efficient intermediate parameters and use the exact offline algorithm, and thus achieve a better regret bound. Finally, we conduct experiments on a real-world multi-layered social network dataset to validate the effectiveness of both our offline and online algorithms.

[1]Our model can be extended to move multiple steps with one unit of budget.

Our contributions can be summarized as follows: (1) We are the first to model the multi-layered network exploration via random walks problem (MuLaNE) as an abstraction for many real-world applications (2) We provide a systematic study of the MuLaNE problem via theoretical analysis and empirical validation by considering offline and online settings for both overlapping and non-overlapping multi-layered networks. Due to the space limit, proofs are included in the supplementary material.

1.1. Related Work

Network exploration via random walks has been studied in various application contexts such as community detection (Pons & Latapy, 2005), centrality measuring (Gleich, 2015), large-scale network sampling (Li et al., 2019), and influence maximization (Wilder et al., 2018). Multiple random walks has also been used, e.g. in (Lv et al., 2002) for query resolution in peer-to-peer networks. However, none of these studies address the budgeted MuLaNE problem. MuLaNE is also related to the influence maximization (IM) problem (Kempe et al., 2003; Chen et al., 2009) and can be viewed as a special variant of IM. Different from the standard Independent Cascade (IC) model (Wang et al., 2012) in IM, which randomly broadcasts to all its neighbors, the propagation process of MuLaNE is a random walk that selects one neighbor. Moreover, each unit of budget in MuLaNE is used to propagate one random-walk step, not to select one seed node as in IC.

The offline budget allocation problem has been studied by Alon et al. (2012); Soma et al. (2014), the latter of which propose the lattice submodularity and a constant approximation algorithm based on this property. Our offline overlapping MuLaNE setting is based on a similar approach, but we provide a better approximation ratio compared to the original analysis in (Alon et al., 2012). Moreover, we further show that the stationary setting enjoys DR-submodularity, leading to an efficient algorithm with a better approximation ratio.

The multi-armed bandit (MAB) problem is first studied by Robbins (1952) and then extended by many studies (cf. (Bubeck & Cesa-Bianchi, 2012)). Our online MuLaNE setting fits into the general Combinatorial MAB (CMAB) framework of (Gai et al., 2012; Chen et al., 2016). CUCB is proposed as a general algorithm for CMAB (Chen et al., 2016). Our study includes several adaptations, such as handling unknown node weights, defining intermediate random variables as base arms and so on.

Chen et al. (2018) study the community exploration problem,

which is essentially a special case of MuLaNE with non-overlapping complete graphs. As a result, their technique is different and much simpler than ours.

The rest of the paper is organized as follows. Section 2 states the settings of MuLaNE; Section 3 states the equivalent bipartite coverage model to derive the explicit form for our reward function; Section 4 states offline algorithms for four offline settings with provable approximation guarantee and running time analysis; Section 5 states two online learning algorithms with regret analysis; Empirical results are shown in Section 6; and Section 7 concludes the paper.

2. Problem Settings

Basic Notations. In this paper, we use $\mathbb{R}$ and $\mathbb{Z}$ to denote the sets of real numbers and integers, respectively, and the associated subscript ≥ 0 and > 0 denote their non-negative and positive subsets, respectively. Suppose we want to explore a multi-layered network $\mathcal{G}(\mathcal{V}, \mathcal{E}, \boldsymbol{\sigma})$, consisting of m layers $L_1, ..., L_m$ and N unique nodes $\mathcal{V}$ with fixed node weights $\boldsymbol{\sigma} \in [0,1]^{\mathcal{V}}$. Let $[m]$ denote the set $\{1, 2, ..., m\}$. Each layer $L_i, i \in [m]$, represents a subgraph of $\mathcal{G}$ we could explore and is modeled as a weighted digraph $\mathcal{G}_i(\mathcal{V}_i, \mathcal{E}_i; w_i)$, where $\mathcal{V}_i \subseteq \mathcal{V}$ is the set of vertices in layer L_i, $\mathcal{E}_i \subseteq \mathcal{V}_i \times \mathcal{V}_i$ is the set of edges in $\mathcal{G}_i$, and $w_i : \mathcal{E}_i \to \mathbb{R}_{>0}$ is the edge weight function associating each edge of $\mathcal{G}_i$. For any two layers L_i and L_j, we say they are *non-overlapping* if $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$. $\mathcal{G}$ is called a *non-overlapping* multi-layered network if any two layers are non-overlapping; otherwise $\mathcal{G}$ is an *overlapping* multi-layered network. Let $n_i = |\mathcal{V}_i|$ denote the size of layer L_i, and the adjacency matrix of $\mathcal{G}_i$ is defined as $\boldsymbol{A}_i \in \mathbb{R}_{\geq 0}^{n_i \times n_i}$, where $\boldsymbol{A}_i[u, v] = w_i((u, v))$ if $(u, v) \in \mathcal{E}_i$ and 0 otherwise.

Exploration Rule. Each layer L_i is associated with an explorer W_i, a budget $k_i \in \mathbb{Z}_{\geq 0}$ and an initial starting distribution $\boldsymbol{\alpha}_i = (\alpha_{i,u})_{u \in \mathcal{V}_i} \in \mathbb{R}_{\geq 0}^{n_i}$ with $\sum_{u \in \mathcal{V}_i} \alpha_{i,u} = 1$. The exploration process of W_i is identical to applying *weighted random walks* within layer L_i. Specifically, the explorer W_i starts the exploration from a random node $v_1 \in \mathcal{V}_i$ with probability α_{i,v_1}; it consumes one unit of budget and continues the exploration by walking from v_1 to one of its out-neighbors, say v_2, with probability $\boldsymbol{P}_i[v_1, v_2]$, where $\boldsymbol{P}_i \in \mathbb{R}_{\geq 0}^{n_i \times n_i}$ is the transition probability matrix, and $\boldsymbol{P}_i[u, v] := \boldsymbol{A}_i[u, v] / (\sum_{w:(u,w) \in \mathcal{E}_i} \boldsymbol{A}_i[u, w])$. Consider visiting the initial node v_1 as the first step, the process is repeated until the explorer W_i walks k_i steps and visits k_i nodes (with possibly duplicated nodes). Note that one can easily generalize our results by moving $\lambda_i \in \mathbb{Z}_{>0}$ steps with one unit of budget for W_i, and for simplicity, we set $\lambda_i = 1$.

Reward Function. We define the exploration trajectory for explorer W_i after exploring k_i nodes as $\Phi(i, k_i) := (X_{i,1}, \ldots, X_{i,k_i})$, where $X_{i,j} \in \mathcal{V}_i$ denotes the node visited at j-th step, and $\Pr(X_{i,1} = u) = \alpha_{i,u}, u \in \mathcal{V}_i$. The reward for $\Phi(i, k_i)$ is defined as the total weights of *unique* nodes visited by W_i, i.e., $\sum_{v \in \{X_{i,j}\}_{j=1}^{k_i}} \sigma_v$. Considering trajectories of all W_i's, the total reward is the total weights of *unique* nodes visited by all random walkers, i.e., $\sum_{v \in \bigcup_{i=1}^{m} \{X_{i,j}\}_{j=1}^{k_i}} \sigma_v$. For notational simplicity, let $\mathcal{G} := (\mathcal{G}_1, \ldots, \mathcal{G}_m)$, $\boldsymbol{\alpha} := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m)$ and $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_{|\mathcal{V}|})$ be the parameters of a problem instance and $\boldsymbol{k} := (k_1, \ldots, k_m)$ be the allocation vector. For overlapping multi-layered network $\mathcal{G}$, the total expected reward is

$$r_{\mathcal{G}, \boldsymbol{\alpha}, \boldsymbol{\sigma}}(\boldsymbol{k}) := \mathbb{E}_{\Phi(1,k_1), \ldots, \Phi(m,k_m)} \left[\sum_{v \in \bigcup_{i=1}^{m} \{X_{i,j}\}_{j=1}^{k_i}} \sigma_v \right], \quad (1)$$

where its explicit formula will be discussed later in Sec. 3. For non-overlapping $\mathcal{G}$, the reward function can be simplified and written as the summation over separated layers,

$$r_{\mathcal{G}, \boldsymbol{\alpha}, \boldsymbol{\sigma}}(\boldsymbol{k}) := \sum_{i=1}^{m} \mathbb{E}_{\Phi(i,k_i)} \left[\sum_{v \in \{X_{i,j}\}_{j=1}^{k_i}} \sigma_v \right]. \quad (2)$$

Problem Formulation. We are interested in the *budget allocation problem*: how to allocate the total budget $B \in \mathbb{Z}_{\geq 0}$ to the m explorers so as to maximize the total weights of unique nodes visited. Also, we assume there is a budget constraint $\boldsymbol{c} := (c_1, ..., c_m)$ such that the allocated budget k_i should not exceed c_i, i.e., $k_i \leq c_i$, for $i \in [m]$.

Definition 1. *Given graph structures $\mathcal{G} := (\mathcal{G}_1, \ldots, \mathcal{G}_m)$, starting distributions $\boldsymbol{\alpha} := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m)$, total budget B and budget constraints $\boldsymbol{c} := (c_1, ..., c_m)$, the **Mu**lti-**La**yered **N**etwork **E**xploration problem, denoted as **MuLaNE**, is formulated as the following optimization problem,*

$$\max_{\boldsymbol{k}} \; r_{\mathcal{G}, \boldsymbol{\alpha}, \boldsymbol{\sigma}}(\boldsymbol{k}) \; s.t. \; \boldsymbol{k} \in \mathbb{Z}_{\geq 0}^{m} \leq \boldsymbol{c}, \; \sum_{i=1}^{m} k_i \leq B, \quad (3)$$

where $r_{\mathcal{G}, \boldsymbol{\alpha}, \boldsymbol{\sigma}}(\boldsymbol{k})$ is given by Eq. (1) or Eq. (2). We also need to consider the following settings:

Offline Setting. For the offline setting, all problem instance parameters $(\mathcal{G}, \boldsymbol{\alpha}, \boldsymbol{\sigma}, \boldsymbol{c}, B)$ are given, and we aim to find the optimal budget allocation $\boldsymbol{k}$ determined by Eq. (3).

Online Setting. For the online setting, we consider T-round explorations. Before the exploration, we only know an upper bound of the total number of nodes in $\mathcal{G}$ and the number of layers m,[2] but we do not know about the network structure $\mathcal{G}$, the starting distributions $\boldsymbol{\alpha}$ or node weights $\boldsymbol{\sigma}$. In round $t \in [T]$, we choose the budget allocation $\boldsymbol{k}_t := (k_{t,1}, \ldots, k_{t,m})$ only based on observations from previous rounds, where $k_{t,i} \in \mathbb{Z}_{\geq 0}$ is the budget allocated to the i-th layer and $\sum_{i=1}^{m} k_{t,i} \leq B, \boldsymbol{k}_t \leq \boldsymbol{c}$. Define $\boldsymbol{k}_t$ as the *action* taken in round t. By taking the action $\boldsymbol{k}_t$ we mean that we interact with the environment and the random explorer W_i, which is part of the environment, would

[2] More precisely we only need to know the number of independent explorers. If two explorers explore on the same layer, it is equivalent as two layers with identical graph structures.

explore $k_{t,i}$ steps and generate the exploration trajectory $(i, k_{t,i}) = (X_{i,1}, \ldots, X_{i,k_{t,i}})$. After we take the action k_t, the exploration trajectory $(i, k_{t,i})$ for each layer V_i as well as the fixed importance weight w_u of $u \in (i, k_{t,i})$ is revealed as the feedback[2], which we leverage on to learn parameters related to the graph structure G, the starting distribution and the node weights, so that we can select better actions in future rounds. The reward we gain in round t is the total weights of unique nodes visited by all random explorers in round t. Our goal is to design an efficient online learning algorithm A to give us guidance on taking actions and gain as much cumulative reward as possible in T rounds.

In general, the online algorithm has to deal with the exploration-exploitation tradeoff. The cumulative regret is a commonly used metric to evaluate the performance of an online learning algorithm A. Formally, the T-round (α, β)-approximation) regret of A is:

$$\mathrm{Reg}_{G,\alpha,\beta}(T) = T \cdot r_{G,\alpha,\beta}(k^*) - \mathbb{E}\left[\sum_{t=1}^{T} r_{G,\alpha,\beta}(k_t^A)\right], \tag{4}$$

where $r_{G,\alpha,\beta}(k^*)$ is the reward value for the optimal budget allocation k^*, k_t^A is budget allocation selected by the learning algorithm A in round t, the expectation is taken over the randomness of the algorithm and the exploration trajectories in all T rounds, and (α, β) is the approximation guarantee of the offline oracle as explained below. Similar to other online learning frameworks (Chen et al., 2016; Wang & Chen, 2017), we assume that the online learning algorithm has access to an offline (α, β)-approximation oracle, which for problem instance (G, α, β, c, B) outputs an action k such that $\Pr(r_{G,\alpha,\beta}(k) \geq \alpha \cdot r_{G,\alpha,\beta}(k^*)) \geq \beta$. We also remark that the actual oracle we use takes certain intermediate parameters as inputs instead of G and α, β.

Submodularity and DR-Submodularity Over Integer lattices. To solve the MuLaNE problem, we leverage on the submodular and DR-submodular properties of the reward function. For any $x, y \in \mathbb{Z}_0^m$, we denote $x \vee y, x \wedge y \in \mathbb{Z}_0^m$ as the coordinate-wise maximum and minimum of these two vectors, i.e., $(x \vee y)_i = \max\{x_i, y_i\}$, $(x \wedge y)_i = \min\{x_i, y_i\}$. We define a function $f : \mathbb{Z}_0^m \to \mathbb{R}$ over the integer lattice $\mathbb{Z}_0^m$ as a submodular function if the following inequality holds for any $x, y \in \mathbb{Z}_0^m$:

$$f(x \vee y) + f(x \wedge y) \leq f(x) + f(y). \tag{5}$$

Let $\mathrm{supp}^+(x - y)$ denote the set $\{i \in [m] : x_i > y_i\}$, $\mathbf{1}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the one-hot vector whose i-th element is 1 and 0 otherwise, and $x \leq y$ means $x_i \leq y_i$ for all $i \in [m]$. We define a function $f : \mathbb{Z}_0^m \to \mathbb{R}$ as a DR-submodular (diminishing return submodular) function if the following inequality holds for any $x \leq y$ and $i \in [m]$,

$$f(y + \mathbf{1}_i) - f(y) \leq f(x + \mathbf{1}_i) - f(x). \tag{6}$$

We say a function f is monotone if for any $x \leq y$, $f(x) \leq f(y)$. Note that for a function $f : \mathbb{Z}_0^m \to \mathbb{R}$, submodularity does not imply DR-submodularity over integer lattices. In fact, the former is weaker than the latter, that is, a DR-submodular function is always a submodular function, but not vice versa. However, for a typical submodular function $f : \{0,1\}^m \to \mathbb{R}$ defined on a set, they are equivalent.

3. Equivalent Bipartite Coverage Model

In order to derive the explicit formulation of the reward function $r_{G,\alpha,\beta}(k)$ in Eq. (1), we construct an undirected bipartite coverage graph $BG(W, V, E')$, where $W = \{W_1, \ldots, W_m\}$ denotes m random explorers, $V = \bigcup_{i \in [m]} V_i$ denotes all possible distinct nodes to be explored in G, and the edge set $E' = \{(W_i, u) \mid u \in V_i, i \in [m]\}$ indicating whether node u could be visited by W_i. For each edge $(W_i, u) \in E'$, we associate it with $c_i + 1$ visiting probabilities denoted as $P_{i,u}(k_i)$ for $k_i \in \{0\} \cup [c_i]$. $P_{i,u}(k_i)$ represents the probability that the node u is visited by the random walker W_i given the budget k_i, i.e., $P_{i,u}(k_i) = \Pr(u \in \mathcal{V}(i, k_i))$. Since W_i can never visit the node outside the i-th layer (i.e., $u \notin G_i$), we set $P_{i,u}(k_i) = 0$ if $(W_i, u) \notin E'$. Then, given budget allocation k, the probability of a node u visited by at least one random explorer is $\Pr(u; k) = 1 - \prod_{i \in [m]} (1 - P_{i,u}(k_i))$ because each W_i has the independent probability $P_{i,u}(k_i)$ to visit u.

By summing over all possible nodes, the reward function is

$$r_{G,\alpha,\beta}(k) = \sum_{u \in V} w_u \left(1 - \prod_{i \in [m]} (1 - P_{i,u}(k_i))\right) \tag{7}$$

According to Eq. (2), for non-overlapping multi-layered network, we can rewrite the reward function as:

$$r_{G,\alpha,\beta}(k) = \sum_{i \in [m]} \sum_{u \in V_i} w_u P_{i,u}(k_i). \tag{8}$$

Remark. The bipartite coverage model is needed to integrate the graph structure and the random walk exploration mechanisms into $P_{i,u}(k_i)$ to determine the reward function. It also handles the scenario where each layer is explored by multiple random walkers (see Appendix A).

3.1. Properties of the visiting probability $P_{i,u}(k_i)$

The quantities $P_{i,u}(k_i)$'s in Eq. (7) and (8) are crucial for both our offline and online algorithms, and thus we provide their analytical formulas and properties here. In order to analyze the property of the reward function, we apply the absorbing Markov Chain technique to calculate $P_{i,u}(k_i)$. For $u \notin G_i$, $P_{i,u}(k_i) = 0$; For $u \in G_i$, we create an absorbing Markov Chain $P_i(u) \in \mathbb{R}^{n_i \times n_i}$ by setting the target

[2] We further consider random node weights with unknown mean vector, see the Appendix F.3 for details.

node u as the absorbing node. We derive the corresponding transition matrix $P_i(u)$ as $P_i(u)[v;\ ] = \mathbf{0}^\top_u$ if $v = u$ and $P_i(u)[v;\ ] = P_i[v;\ ]$ otherwise, where $P_i[v;\ ]$ denotes the row vector corresponding to the node v of P_i, and $\mathbf{0}_u = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ denotes the one-hot vector with 1 at the u-th entry and 0 elsewhere.[4] Intuitively, $P_i(u)$ corresponds to the transition probability matrix of G_i after removing all out-edges of u and adding a self loop to itself in the original graph G_i. The random walker W_i will be trapped in u if it ever visits u. We observe that $P_{i;u}(k_i)$ equals to the probability the random walker stays in the absorbing node u at step $k_i \in \mathbb{Z}_{>0}$ (trivially, $P_{i;u}(0) = 0$),

$$P_{i;u}(k_i) = \mathbf{0}^\top_u P_i(u)^{k_i - 1} \mathbf{1}_u. \tag{9}$$

Then define the marginal gain of $P_{i;u}(\cdot)$ at step $k_i - 1$ as,

$$g_{i;u}(k_i) = P_{i;u}(k_i) - P_{i;u}(k_i - 1). \tag{10}$$

The physical meaning of $g_{i;u}(k_i)$ is the probability that node u is visited exactly at the k_i-th step and not visited before k_i. Now, we can show $g_{i;u}(k_i)$ is non-negative,

Lemma 1. *$g_{i;u}(k_i) \geq 0$ for any $i \in [m]$, $u \in V_i$, $k_i \in \mathbb{Z}_{>0}$.*

This means $P_{i;u}(k_i)$ is non-decreasing with respect to step k_i. In other words, the more budgets we allocate to W_i, the higher probability W_i can visit u in k_i steps.

Starting from arbitrary distributions: Although $P_{i;u}(k_i)$ is monotone (non-decreasing) w.r.t. k_i, $g_{i;u}(k_i)$ may not be monotonic non-increasing under arbitrary staring distributions. Intuitively, there may exist one critical step k_i such that $P_{i;u}(k_i)$ suddenly increases by a large value (see Appendix B.1). This means that $P_{i;u}(k_i)$ lacks the diminishing return (or "discretely concave") property, which many problems rely on to provide good optimization (Kapralov et al., 2013; Soma & Yoshida, 2015). In other words, we are dealing with a more challenging non-concave optimization problem for the discrete budget allocation.

Starting from the stationary distribution: If our walkers start with the stationary distributions, the following property holds: $g_{i;u}(k_i)$ is non-increasing w.r.t. k_i.

Lemma 2. *$g_{i;u}(k_i + 1) \leq g_{i;u}(k_i) \leq 0$ for any $i \in [m]$, $u \in V_i$, $k_i \in \mathbb{Z}_{>0}$, if $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i$, where $\boldsymbol{\pi}_i^\top P_i = \boldsymbol{\pi}_i^\top$.*

4. Offline Optimization for MuLaNE

In this section, we first consider the general case, where layers are overlapping and starting distributions are arbitrary. Next, we consider the starting distribution is the stationary distribution, and give solutions with better solution quality and time complexity. Then we analyze special cases where layers are non-overlapping and give optimal solutions for arbitrary distributions. The summary for offline models and algorithmic results are presented in Table 1.

4.1. Overlapping MuLaNE

Starting from arbitrary distributions. Based on the equivalent bipartite coverage model, one can observe that our problem formulation is a generalization of the Probabilistic Maximum Coverage (PMC) (Chen et al., 2016) problem, which is NP-hard and has $(1 - 1/e)$ approximation based on submodular set function maximization. However, our problem is more general in that we want to select multi-sets from W with budget constraints, and the reward function in general does not have the DR-submodular property for an arbitrary starting distribution. Nevertheless, we have the following lemma to solve our problem.

Lemma 3. *For any network G, distribution $\boldsymbol{\pi}$ and weights $\mathbf{r}$, $r_{G;\boldsymbol{\pi};\mathbf{r}}(\cdot) : \mathbb{Z}^m_{\geq 0} \to \mathbb{R}$ is monotone and submodular.*

Leveraging on the monotone submodular property, we design a Budget Effective Greedy algorithm (Alg. 1). The core of Alg. 1 is the BEG procedure. Let $\Delta(i; b; \mathbf{k})$ be the per-unit marginal gain $(r_{G;\boldsymbol{\pi};\mathbf{r}}(\mathbf{k} + b\,\mathbf{1}_i) - r_{G;\boldsymbol{\pi};\mathbf{r}}(\mathbf{k}))/b$ for allocating b more budgets to layer i, which equals to

$$\sum_{u \in V_u} \prod_{j \neq i} (1 - P_{j;u}(k_j)) \left(P_{i;u}(k_i + b) - P_{i;u}(k_i)\right)/b. \tag{11}$$

BEG procedure consists of two parts and maintains a queue Q, where any pair $(i; b) \in Q$ represents a tentative plan of allocating b more budgets to layer i. The first part is built around the while loop (line 6-10), where each iteration greedily selects the $(i; b)$ pair in Q such that the per-unit marginal gain $\Delta(i; b; \mathbf{k})$ is maximized. The second part is a for loop (line 12-13), where in the i-th round we attempt to allocate all c_i budgets to layer i and replace the current best budget allocation if we have a larger reward.

Theorem 1. *Algorithm 1 obtains a $(1 - e^{-\beta}) \geq 0.357$-approximate solution, where β is the solution of equation $e^{\beta} = 2 - \beta$, to the overlapping MuLaNE problem.*

Line 1 uses $O(m\,k c_1\, n^3_{\max})$ time to pre-calculate visiting probabilities based on Eq. (9), where $n_{\max} = \max_i |V_i|$. In the BEG procedure, the while loop contains B iterations, the size of queue Q is $O(m\,k c_1)$, and line 7 uses $O(n_{\max})$ to calculate $\Delta(i; b; \mathbf{k})$ by proper pre-computation and update (see Appendix C.3), thus the time complexity of Algorithm 1 is $O(B\,k c_1\, mn_{\max} + m\,k c_1\, n^3_{\max})$.

Remark 1. The idea of combining the greedy algorithm with enumerating solutions on one layer is also adopted in previous works (Khuller et al., 1999; Alon et al., 2012), but they only give a $\frac{1}{2}(1 - e^{-1}) \geq 0.316$-approximation analysis. In this paper, we provide a novel analysis with a better $(1 - e^{-\beta}) \geq 0.357$-approximation, see Appendix C.1 for details.

[4] When related to matrix operations, we treat all vectors as column vectors by default.

Table 1. Summary of the of ine models and algorithms.

Overlapping?	Starting distribution	Algorithm	Apprx ratio	Time complexity
X	Arbitrary	Budget Effective Greedy	$(1 \quad e \quad)^5$	$O(B\ kck_1\ mn_{max} + kck_1\ mn^3_{max})$
X	Stationary	Myopic Greedy	$(1 \quad 1=e)$	$O(Bmn_{max} + kck_1\ mn^3_{max})$
	Arbitrary	Dynamic Programming	1	$O(B\ kck_1\ m + kck_1\ mn^3_{max})$
	Stationary	Myopic Greedy	1	$O(B \log m + kck_1\ mn^3_{max})$

Algorithm 1 Budget Effective Greedy (BEG) Algorithm for the Overlapping MuLaNE

Input: Network G, starting distributions , node weights , budget B, constraints c.
Output: Budget allocation k.
1: Compute visiting probabilities $(P_{i;u}(b))_{i\,2\,[m];u\,2V;b2\,[c_i]}$ according to Eq. (9).
2: k BEG$((P_{i;u}(b))_{i\,2\,[m];u\,2V;b2\,[c_i]}$, , B, c).
3: Procedure BEG$((P_{i;u}(b))_{i\,2\,[m];u\,2V;b2\,[c_i]}$, , B, c)
4: Let $k := (k_1; :::; k_m)$ $0, K$ B.
5: Let Q $f(i; b_i) j i\ 2\ [m]; 1$ b_i $c_i g$.
6: while $K > 0$ and Q 6= ; do
7: $(i ; b)$ $\arg\max_{(i;b)2Q}$ $(i; b; k)=b$ .Eq. (11)
8: k_i $k_i + b, K$ K b.
9: Modify all pairs $(i; b)\ 2\ Q$ to $(i; b$ $b)$.
10: Remove all pairs $(i; b)\ 2\ Q$ such that b 0.
11: end while
12: for $i\ 2\ [m]$ do
13: if $r_{G;\ ;}(c_i \ _i) > r_{G;\ ;}(k)$, then k $c_i \ _i$:
14: end for
15: return $k := (k_1; :::; k_m)$.
16: end Procedure

Remark 2. Another algorithm (Alon et al., 2012) with a better approximation ratio is to use partial enumeration techniques (i.e., BEGE), which can achieve $(1 \quad 1=e)$ approximation ratio. This is the best possible solution in polynomial time unless P=NP. But the time complexity is prohibitively high in $O(B^4m^4\ kck_1\ n_{max} + m\ kck_1\ n^3_{max})$. The algorithm and the analysis are provided in the Appendix D.2.

Starting from the stationary distribution. We also consider the special case where each random explorer starts from the stationary distribution $_i$ with $_i > P = _i > .$ In this case, we have the following stronger DR-submodularity

Lemma 4. For any network G, stationary distributions and node weights , function $r_{G;\ ;}(\ ) : Z^m_0 \ ! \ R$ is monotone and DR-submodular.

Since the reward function is DR-submodular, the BEG procedure can be replaced by the simple MG procedure in Alg. 2 with a better approximation ratio. The time complexity is also improved to $O(Bmn_{max} + m\ kck_1\ n^3_{max})$.

Theorem 2. Algorithm 2 obtains a $(1 \quad 1=e)$-approximate solution to the overlapping MulaNE with the stationary starting distributions.

Algorithm 2 Myopic Greedy (MG) Algorithm for MuLaNE

1: Same input, output and line 1-2 as in Alg. 1, except replacing BEG with MG procedure below.
2: Procedure MG$((P_{i;u}(b))_{i\,2\,[m];u\,2V;b2\,[c_i]}$, , B, c)
3: Let $k := (k_1; :::; k_m)$ $0, K$ B.
4: while $K > 0$ do
5: i $\arg\max_{i\,2\,[m];k_i+1\ c_i}$ $(i; 1; k)$: . Eq. (11)
6: k_i $k_i + 1, K$ K 1.
7: end while
8: return $k = (k_1; :::; k_m)$.
9: end Procedure

4.2. Non-overlapping MuLaNE

For non-overlapping MuLaNE, we are able to achieve the exact optimal solution: for the stationary starting distribution, a slight modi cation of the greedy algorithm Alg. 2 gives the optimal solution, while for an arbitrary starting distribution, we design a dynamic programming algorithm to compute the optimal solution . Due to the space constraint, the details are included in Appendix E.

5. Online Learning for MuLaNE

In the online setting, we continue to study both overlapping and non-overlapping MuLaNE problem. However, the network structure G, the distribution and node weights are not known a priori. Instead, the only information revealed to the decision maker includes total budget B, the number of layers m, the number of the target nodes jVj (or an upper bound of it) and the budget constraint c.

5.1. Online Algorithm for Overlapping MuLaNE

For the unknown network structure and starting distributions, we bypass the transition matrices P_is and directly estimate the visiting probabilities $P_{i;u}(b)\ 2\ [0; 1]$. This avoids the analysis for how the estimated P_i affects the performance of online algorithms, which could be unbounded since we consider general graph structures and the reward function given by Eq. (7) is highly non-linear in P_i. Moreover, we can save the matrix calculation by directly using $P_{i;u}(b)$,

which is efficient even for large networks.

Specifically, we maintain a set of base arms $A = \{(i,u,b) | i \in [m], u \in V, b \in [c_i]\}$, where the total number $|A| = \sum_{i \in [m]} c_i |V|$. For each base arm $(i,u,b) \in A$, we denote $\mu_{i,u,b}$ as the true value of each base arm, i.e., $\mu_{i,u,b} = P_{i,u}(b)$. For the unknown node weights, we maintain the optimistic weight $\bar{w}_u = 1$ if $u \in V$ has not been visited. After u is first visited and its true value w_u is revealed, we replace $\bar{w}_u$ with w_u. With a little abuse of the notation, we use $\boldsymbol{X}$ to denote the unknown intermediate parameters and $r_{S,\boldsymbol{w},\boldsymbol{X}}(k)$ to denote the reward $r_{G,\boldsymbol{w},\boldsymbol{X}}(k)$.

We present our algorithm in Alg. 3, which is an adaptation of the CUCB algorithm for the general Combinatorial Multi-arm Bandit (CMAB) framework (Chen et al., 2016) to our setting. Notice that we use A as defined above in the algorithm, and A is defined using V, which is the set of node ids and should not be known before the learning process starts. This is not an issue, because at the beginning we can create $|V|$ (which is known) placeholders for the node ids, and once a new node is visited, we immediately replace one of the placeholders with the new node id. Thus V appearing in the above definition of A is just for notational convenience.

Algorithm 3 CUCB-MAX Algorithm for the MuLaNE

Input: Budget B, number of layers m, number of nodes $|V|$, constraints $\boldsymbol{c}$, of line oracle BEG.

1: For each arm $(i,u,b) \in A$, $T_{i,u,b} \leftarrow 0$, $\hat{\mu}_{i,u,b} \leftarrow 0$.
2: For each node $v \in V$, $\bar{w}_v \leftarrow 1$.
3: **for** $t = 1, 2, 3, \ldots, T$ **do**
4: **for** $(i,u,b) \in A$ **do**
5: $\rho_{i,u,b} \leftarrow \sqrt{3 \ln t / (2T_{i,u,b})}$.
6: $\tilde{\mu}_{i,u,b} \leftarrow \min\{\hat{\mu}_{i,u,b} + \rho_{i,u,b}, 1\}$.
7: **end for**
8: For $(i,u,b) \in A$, $\tilde{\mu}_{i,u,b} \leftarrow \max_{j \in [b]} \tilde{\mu}_{i,u,j}$.
9: $\boldsymbol{k} \leftarrow \mathrm{BEG}((\tilde{\mu}_{i,u,b})_{(i,u,b) \in A}, (\bar{w}_v)_{v \in V}, B, \boldsymbol{c})$.
10: Apply budget allocation $\boldsymbol{k}$, which gives trajectories $\boldsymbol{X} := (X_{i,1}, \ldots, X_{i,k_i})_{i \in [m]}$ as feedbacks.
11: For any visited node $v \in \bigcup_{i \in [m]} \{X_{i,1}, \ldots, X_{i,k_i}\}$, receive its node weight w_v and set $\bar{w}_v \leftarrow w_v$.
12: For any $(i,u,b) \in S := \{(i,u,b) \in A \mid b \le k_i\}$, $Y_{i,u,b} \leftarrow 1$ if $u \in \{X_{i,1}, \ldots, X_{i,b}\}$ and 0 otherwise.
13: For $(i,u,b) \in S$, update $T_{i,u,b}$ and $\hat{\mu}_{i,u,b}$: $T_{i,u,b} \leftarrow T_{i,u,b} + 1$, $\hat{\mu}_{i,u,b} \leftarrow \hat{\mu}_{i,u,b} + (Y_{i,u,b} - \hat{\mu}_{i,u,b})/T_{i,u,b}$.
14: **end for**

In Alg. 3, we maintain an unbiased estimation of the visiting probability $P_{i,u}(b)$, denoted as $\hat{\mu}_{i,u,b}$. Let $T_{i,u,b}$ record the number of times arm (i,u,b) is played so far and $\bar{w}_v$ denote the optimistic importance weight. In each round t, we compute the confidence radius $\rho_{i,u,b}$ in line 5, which controls the level of exploration. The confidence radius is larger when the arm (i,u,b) is not explored often (i.e. $T_{i,u,b}$ is small), and thus motivates more exploration. Due to the randomness of the exploration process, the upper confidence bound (UCB) value $\tilde{\mu}_{i,u,b}$ could be decreasing with b, but our offline oracle BEG can only accept non-decreasing UCB values (otherwise the $(1-1/e)$ approximation is not guaranteed). Therefore, in line 8, we increase the UCB value $\tilde{\mu}_{i,u,b}$ and set it to be $\max_{j \in [b]} \tilde{\mu}_{i,u,j}$. This is the adaption of the CUCB algorithm to fix our case, and thus we name our algorithm as CUCB-MAX. After we apply the $(1-1/e)$ approximate solution $\boldsymbol{k}$ given by the BEG oracle (i.e., Alg. 1), we get m trajectories as feedbacks. In line 11, we can update the unknown weights of visited nodes. In line 13, for base arms (i,u,b) with $b \le k_i$, we update corresponding statistics by the Bernoulli random variable $Y_{i,u,b} \in \{0,1\}$ indicating whether node u is visited by W_i in first b steps.

Regret Analysis. We define the reward gap $_{\boldsymbol{k}} = \max(0, r_{S,\boldsymbol{w},\boldsymbol{X}}(\boldsymbol{k}) - r_{S,\boldsymbol{w},\boldsymbol{X}}(\boldsymbol{k}))$ for all feasible action $\boldsymbol{k}$ satisfying $\sum_{i=1}^{m} k_i = B$, $0 \le k_i \le c_i$, where $\boldsymbol{k}$ is the optimal solution for parameters, and $= 1 - e^{-}$ is the approximation ratio of the $(,)$-approximate offline oracle. For each base arm (i,u,b), we define $\mu^{\min}_{i,u,b} = \min_{\boldsymbol{k} > 0; k_i = b} {}_{\boldsymbol{k}}$ and $\mu^{\max}_{i,u,b} = \max_{\boldsymbol{k} > 0; k_i = b} {}_{\boldsymbol{k}}$. As a convention, if there is no action $\boldsymbol{k}$ with $k_i = b$ such that ${}_{\boldsymbol{k}} > 0$, we define $\mu^{\min}_{i,u,b} = 1$ and $\mu^{\max}_{i,u,b} = 0$. Let $_{\min} = \min_{(i,u,b) \in A} \mu^{\min}_{i,u,b}$ and $_{\max} = \max_{(i,u,b) \in A} \mu^{\max}_{i,u,b}$. The following theorem summarizes the regret bound for Alg. 3.

Theorem 3. Algorithm 3 has the following distribution-dependent $(1-e^{-})$ approximation regret,

$$\mathrm{Reg}_{S,\boldsymbol{w},\boldsymbol{X}}(T) \le \sum_{(i,u,b) \in A} \frac{108 m |V| \ln T}{\mu^{\min}_{i,u,b}} + 2|A| + \frac{2}{3}|A| \quad _{\max}.$$

Remark 1. Looking at the above distribution dependent bound, we have the $O(\log T)$ approximation regret, which is asymptotically tight. Coefficient $m|V|$ in the leading term corresponds to the number of edges in the complete bipartite coverage graph. Notice that we cannot use the true edge number $\sum_{i \in [m]} |V_i|$, because the learning algorithm does not know which nodes are contained in each layer, and has to explore all visiting possibilities given by the default complete bipartite graph. The set of base arms A has some redundancy due to the correlation between these base arms, and thus it is unclear if the summation over all base arms in the regret bound is tight. For the non-overlapping case, we further reduce the number of base arms to achieve better regret bounds, but for the overlapping case, how to further reduce base arms to achieve a tighter regret bound is a challenging open question left for the future work.

Remark 2. The $(1-e^{-})$ approximate regret is determined by the offline oracle BEG that we plug in Line 9 and can be replaced by $(1-1/e;1)$ regret using BEGE or even by the exact regret if the oracle can obtain the optimal budget allocation. The usage of BEG is a trade-off we

make between computational ef ciency and learning ef - ciency and empirically, it performs well as we shall see in Section 6.

Remark 3. The full proof of the above theorem is included in the Appendix F, where we rely on the following properties of $r_{;}(k)$ to bound the regret.

Property 1. (Monotonicity). The reward $_{;}(k)$ is monotonically increasing, i.e., for any budget allocation k, any two vectors $=(_{i;u;b})_{(i;u;b)2A}$, $^0=(^0_{i;u;b})_{(i;u;b)2A}$ and any node weights , ^{0}, we have $r_{;}(k)$ $r_{^0;^0}(k)$, if $_{i;u;b}$ $^0_{i;u;b}$ and $_v$ ^0_v, $8(i;u;b)$ 2 A ; v 2 V.

Property 2. (1-Norm Bounded Smoothness). The reward function $r_{;}(k)$ satis es the 1-norm bounded smoothness condition, i.e., for any budget allocation k, any two vectors $=(_{i;u;b})_{(i;u;b)2A}$, $^0=(^0_{i;u;b})_{(i;u;b)2A}$ and any node weights , ^{0}, we have $jr_{;}(k)$ $r_{^0;^0}(k)j$ $P_{i2[m];u2V;b=k_i}(_u j_{i;u;b}$ $^0_{i;u;b}j + j_u$ $^0_u j^0_{i;u;b})$.

We emphasize that our algorithm and analysis differ from the original CUCB algorithm (Chen et al., 2016) as follows. First, we have the additional regret caused by the overestimated weights for unvisited nodes, i je$_u$, $^0_u j^0_{i;u;b}$ term in property 2. We carefully bound this term based on the observation that $^0_{i;u;b}$ is small and decreasing quickly before u is rst visited. Next, we have to take the max (line 8) to guarantee the UCB value is monotone w.r.t. since our BEG oracle can only output $(1$ e ; 1)-approximation with monotone inputs. Due to the above operation, $(i;u;b)$'s UCB value depends on the feedback from all arms $(i;u;j)$ for j b (set in line 12). So we should update all these arms (line 13) to guarantee that the estimates to all these arms are accurate enough. Finally, directly following the standard CMAB result would have a larger regret, because arms in are de ned as triggered arms, but only arms in ^{0} $=$ $f(i;u;b)$ 2 $Aj k_i$ $=$ bg affect the rewards. So we conceptually view arms in ^{0} as triggered arms and use a tighter 1-Norm Bounded Smoothness condition as given above to derive a tighter regret bound. This improves the coef cient of the leading n T term in the distribution dependent regret by a factor of $jfj=j$ $^0j = O(B=m)$, and the $1=$ $_{i;u;b}$ $_{min}$ term is smaller since the original de nition would have $_{i;u;b}$ $_{min}$ $=$ $\min_{k>0;k_i}$ b k.

5.2. Online Algorithm for Non-overlapping Case

For the non-overlapping case, we say layer-wise marginal gains as our base arms. Concretely, we maintain a set of base arms A $=$ $f(i;b)ji$ 2 $[m];b$ 2 $[c_i]g$. For each base arm $(i;b)$ 2 A., let $_{i;b}$ $=$ $_{u2V}$ $_u(P_{i;u}(b)$ $P_{i;u}(b$ $1))$ be the true marginal gain of assigning budget b to layer i. We apply the standard CUCB algorithm to this setting and call the resulting algorithm CUCB-MG (Algorithm 9 in Appendix G). Note that in the non-overlapping setting we can solve the of ine problem exactly, so we can use the exact of- ine oracle to solve the online problem and achieve an exact regret bound. This is the major advantage over the overlapping setting where we can only achieve an approximate bound. De ne $_k$ $=$ $r_{;}(k$ $)$ $r_{;}(k)$ for all feasible action k, and $_{i;b}$ $_{min}$ $=$ $\min_{k>0;k_i}$ b k, CUCB-MG has a $O(P_{(i;b)2A}$ $48B^2 \ln T=$ $_{i;b}$ $_{min})$ regret bound.

(a) Of ine, overlapping. (b) Of ine, non-overlapping.

(c) Online, overlapping. (d) Online, non-overlapping.

Figure 1. Above: total weights of unique nodes visited for of ine algorithms. Below: regret for online algorithms when B $=$ 3000.

Table 2. Statistics for FF-TW-YT network

Layer	FriendFeed	Twitter	YouTube
# of vertices	5,540	5,702	663
# of edges	31,921	42,327	614

6. Experiments

Dataset and settings. We conduct experiments on a real-world multi-layered network FF-TW-YT, which contains m $=$ 3 layers representing users' social connections in FriendFeed (FF), Twitter (TW) and YouTube (YT) (Dickison et al., 2016). In total, FF-TW-YT has 6,407 distinct vertices representing users and 74,836 directed edges representing connections ("who follows whom") among users. The statistics for each layer is summarized in Table 2. We transform the FF-TW-YT network $G(V;E)$ into a symmetric directed network (by adding a new edge $(v;u)$ if $(u;v)$ 2 E but $(v;u)$ 2 E) because a user can be visited via her followers or followees. Each edge weight is set to 1 and the node weights are set to be $_v$ 2 $f0;0:5;1g$ uniformly at random. Each random walker always starts from the smallest node-id in each layer and we set constraints $_i$ equal to the total budget B. Note that in order to test the non-overlapping case, we use the same network but relabel node-ids so that

Table 3. Running time (seconds) for of ine and online algorithms.

	B=2.6k	B=2.8k	B=3.0k
BEG	0.274	0.316	0.363
BEGE	34.37	45.51	59.20
OPT	91.16	98.42	105.63

(a) Running time of of ine algorithms for the overlapping case with different budgets B.

	B=2.6k	B=2.8k	B=3.0k
DP	0.038	0.044	0.050
MG	0.008	0.009	0.010
OPT	70.10	75.78	81.11

(b) Running time of of ine algorithms for the non-overlapping case with different budgets B.

Overlapping?	X	
BEG	1.22	NA
BEGE	60.03	NA
DP	NA	0.86
OPT	107.33	82.01

(c) Per-round running time for CUCB-MAX (or CUCB-MG) with different oracles when B=3.0k.

they do not overlap between different layers. To handle the randomness, we repeat 10,000 times and present the averaged total weights of unique nodes visited for of ine optimization. We calculate the regret by comparing with the optimal solution, which is stronger than comparing with $(1 - e^{-1})$-approximate solution as de ned in Eq. (4). We average over 200 independent experiments to provide the mean regret with 95% con dence interval. To evaluate the computational ef ciency, we also present the running time for both of ine and online algorithms in Table 3.

Algorithms in comparison. For the of ine setting, we present the results for Alg. 1 (denoted as BEG), Alg. 2 (denoted as MG), BEG with partial enumeration (denoted as BEGE) and Alg. 7 (denoted as DP). We provide two baselines PROP-S and PROP-W, which allocates the budget proportional to the layer size and proportional to the total weights if we allocate $B=3$ budgets to that layer, respectively. The optimal solution (denoted as OPT) is also provided by enumerating all possible budget allocations. For online settings, we consider CUCB-MAX (Alg. 3) and CUCB-MG (Alg. 9) algorithms. We shrink the con dence interval by , i.e., $_{i;u;b}$ $_{i;u;b}$, to speed up the learning, though our theoretical regret bound requires = 1. For baselines, we consider the EMP algorithm which always allocates according to the empirical mean, and the -Greedy algorithm which allocates budgets according to empirical mean with probability 1 and allocates all B budgets to the i-th layer with probability $=m$. We also compare with the Thompson sampling (TS) method (Wang & Chen, 2018), which uses Beta distribution $Beta(\ ;\ )$ (where $=\ = 1$ initially) as prior distribution for each base arm.

Experimental results. We show the results for MuLaNE problems in Figure 1. For the of ine overlapping case, both BEG and MG outperform two baselines PROP-W and PROP-S in receiving total weights. Although not guaranteed by the theory, BEG are empirically close to BEGE and the optimal solution (OPT). As for the computational ef ciency, in Table 3a, the BEG is at least two orders of magnitude (e.g., 163 times when B=3.0k) faster than BEGE and OPT. Combining that the reward of BEG is empirically close to BEGE and the optimal solution, this shows that BEG is empirically better than BEGE and OPT. For the of ine non-overlapping case, the results are similar, but the difference is that we have the theoretical guarantee for the optimality of DP. For the online setting, all CUCB-MAX/CUCB-MG curves outperform the baselines. This demonstrates empirically that CUCB-MAX algorithm can effectively learn the unknown parameters while optimizing the objective. For the computational ef ciency of online learning algorithms, since the running time for algorithms with the same oracle is similar, we present the running time for CUCB-MAX with different oracles in Table 3c. CUCB-MAX with BEG is 50 times faster than BEGE, which is consistent with our theoretical analysis. The results for different budgets B are consistent with $B = 3000$, which are included in the Appendix H.2. Results and analysis for stationary starting distributions are also in the Appendix H.1.

7. Conclusions and Future Work

This paper formulates the multi-layered network exploration via random walks (MuLaNE) as a budget allocation problem, requiring that the total weights of distinct nodes visited on the multi-layered network is maximized. For the of ine setting, we propose four algorithms for MuLaNE according to the speci cation of multi-layered network (overlapping or non-overlapping) and starting distributions (arbitrary or stationary), each of which has a provable guarantee on approximation factors and running time. We further study the online setting where network structure and the node weights are not known a priori. We propose the CUCB-MAX algorithm for overlapping MuLaNE and the CUCB-MG algorithm for the non-overlapping case, both of which are bounded by a $O(\log T)$ (approximate) regret. Finally, we conduct experiments on a social network dataset to show the empirical performance of our algorithms.

There are many compelling directions for the future study. For example, it would be interesting to extend our problem where the decision maker can jointly optimize the starting distribution and the budget allocation. One could also study the adaptive MuLaNE by using the feedback from the exploration results of the previous steps to determine the exploration strategy for future steps.

8. Acknowledgement

The work of John C.S. Lui was supported in part by the GRF 14200420.

References

Alon, N., Gamzu, I., and Tennenholtz, M. Optimizing budget allocation among channels and in uencers. In Proceedings of the 21st international conference on World Wide Web, pp. 381–388. ACM, 2012.

Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends® in Machine Learning, 5(1):1–122, 2012.

Chen, W., Wang, Y., and Yang, S. Ef cient in uence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 199–208, 2009.

Chen, W., Wang, Y., Yuan, Y., and Wang, Q. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. The Journal of Machine Learning Research, 17(1):1746–1778, 2016.

Chen, X., Huang, W., Chen, W., and Lui, J. C. Community exploration: From of ine optimization to online learning. In Advances in Neural Information Processing Systems, pp. 5474–5483, 2018.

Dickison, M. E., Magnani, M., and Rossi, L. Multilayer social networks. Cambridge University Press, 2016.

Dubhashi, D. P. and Panconesi, A. Concentration of measure for the analysis of randomized algorithms. Cambridge University Press, 2009.

Gai, Y., Krishnamachari, B., and Jain, R. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. IEEE/ACM Transactions on Networking (TON), 20(5): 1466–1478, 2012.

Gleich, D. F. Pagerank beyond the web. SIAM Review, 57 (3):321–363, 2015.

Kapralov, M., Post, I., and Vondrák, J. Online submodular welfare maximization: Greedy is optimal. In Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms, pp. 1216–1225. SIAM, 2013.

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

Khuller, S., Moss, A., and Naor, J. S. The budgeted maximum coverage problem. Information processing letters, 70(1):39–45, 1999.

Kijima, M. Markov processes for stochastic modeling, volume 6. CRC Press, 1997.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. Ef cient sensor placement optimization for securing large water distribution networks. Journal of Water Resources Planning and Management, 134(6): 516–526, 2008.

Lerman, K. and Jones, L. Social browsing on icker. arXiv preprint cs/0612046, 2006.

Li, Y., Wu, Z., Lin, S., Xie, H., Lv, M., Xu, Y., and Lui, J. C. Walking with perception: Ef cient random walk sampling via common neighbor awareness. 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 962–973. IEEE, 2019.

Lv, Q., Cao, P., Cohen, E., Li, K., and Shenker, S. Search and replication in unstructured peer-to-peer networks. In Proceedings of the 16th international conference on Supercomputing, pp. 84–95, 2002.

Pons, P. and Latapy, M. Computing communities in large networks using random walks. In International symposium on computer and information sciences, pp. 284–293. Springer, 2005.

Raginsky, M., Sason, I., et al. Concentration of measure inequalities in information theory, communications, and coding. Foundations and Trends® in Communications and Information Theory, 10(1-2):1–246, 2013.

Robbins, H. Some aspects of the sequential design of experiments. Bulletin of the American Mathematical Society, 58(5):527–535, 1952.

Serfozo, R. Basics of applied stochastic processes. Springer Science & Business Media, 2009.

Soma, T. and Yoshida, Y. A generalization of submodular cover via the diminishing return property on the integer lattice. In Advances in Neural Information Processing Systems, pp. 847–855, 2015.

Soma, T., Kakimura, N., Inaba, K., and Kawarabayashi, K.-i. Optimal budget allocation: Theoretical guarantee and ef cient algorithm. In International Conference on Machine Learning, pp. 351–359, 2014.

Wang, C., Chen, W., and Wang, Y. Scalable in uence maximization for independent cascade model in large-scale social networks. Data Mining and Knowledge Discovery, 25(3):545–576, 2012.

Wang, Q. and Chen, W. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*, pp. 1161–1171, 2017.

Wang, S. and Chen, W. Thompson sampling for combinatorial semi-bandits. In *International Conference on Machine Learning*, pp. 5114–5122, 2018.

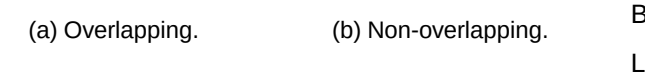
Wilder, B., Immorlica, N., Rice, E., and Tambe, M. Maximizing influence in an unknown social network. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Supplementary Material

The supplementary material is organized as follows.

We first discuss how we handle the multiple random walks in Section A. We provide proofs and examples for properties of the visiting probability in Section B. Proofs of of ine optimization for overlapping MuLaNE are provided in Section C. We provide the detailed budget effective greedy algorithm with partial enumeration (BEGE) and its analysis in Section D. Proofs of of ine optimization for non-overlapping MuLaNE are provided in Section E. We state the detailed analysis of online learning for overlapping MuLaNE in Section F. We state the detailed analysis of online learning for non-overlapping MuLaNE in Section G. Supplemental experiments are provided in Section H.

A. Handling the Multiple Random Walkers



(a) Overlapping. (b) Non-overlapping.

Figure 2. Two types of multi-layered networks.

We can handle the scenario where each layer is explored by multiple random walkers using the bipartite coverage model. Without loss of generality, suppose we want to add a new random walker W_1^0 to the layer L_1. A new node W_1^0 can be easily added to $\mathcal{V}$ representing the new random explorer, and new edges $\{(W_1^0, u) | u \in V_1\}$ are added to $\mathcal{E}^0$. Thus, we can use the same algorithms and analysis to solve the optimization problem on the newly constructed bipartite coverage graph.

B. Proofs and Examples for Properties of the Visiting Probability $P_{i,u}(k_i)$

B.1. Starting From Arbitrary Distributions

Lemma 1. $g_{i,u}(k_i) \geq 0$ for any $i \in [m], u \in V_i, k_i \in \mathbb{Z}_{>0}$.

Proof. For analysis, we use the following equivalent formulation for $k_i \geq 2$, (trivially $P_{i,u}(0) = 0$ and $P_{i,u}(1) = \pi_{i,u}$),

$$P_{i,u}(k_i) = ((\boldsymbol{\pi}_i)^\top_u (\mathbf{I}_i)_{u,k_i} + \pi_{i,u}); \quad (12)$$

where $(\mathbf{I}_i)_{u,k_i} = (\mathbf{I} + \mathbf{T}_i(u) + \cdots + \mathbf{T}_i(u)^{k_i-2})(\mathbf{p}_i)_u$, $(\mathbf{p}_i)_u = (\mathbf{P}_i[\cdot, u])^\top$, $\mathbf{T}_i(u) = (\mathbf{P}_i)_{-u,-u}$, (we use Eq. (2.67) in (Kijima, 1997) to derive $(\mathbf{I}_i)_{u,k_i}$). Note that $\mathbf{p}_u \in \mathbb{R}^{n-1}$ is the vector obtained by deleting the u-th element from $\mathbf{p} \in \mathbb{R}^n$, and $\mathbf{P}_{-u,-v} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the matrix obtained by deleting the u-th row and the v-th column from $\mathbf{P} \in \mathbb{R}^{n\times n}$.

We further derive the marginal gain of $P_{i,u}(\cdot)$ at step $k_i \geq 2$ (trivially $g_{i,u}(1) = \pi_{i,u}$) as,

$$g_{i,u}(k_i) = (\boldsymbol{\pi}_i)^\top_u \mathbf{T}_i(u)^{k_i-2}(\mathbf{p}_i)_u. \quad (13)$$

Now, we can show $g_{i,u}(k_i)$ is non-negative because any element of $(\boldsymbol{\pi}_i)^\top_u$, $\mathbf{T}_i(u)$ and $(\mathbf{p}_i)_u$ are non-negative, which means $P_{i,u}(k_i)$ is non-decreasing with respect to step k_i.

An example showing $g_{i,u}(k_i)$ is not monotone: Consider a path P_3 with three nodes as the i-th layer $G_i(V_i, E_i)$, where $V_i = \{u, v, w\}$ and $E_i = \{(u, v), (v, u), (v, w), (w, v)\}$. If the W_i always starts from the left-most node and chooses the right-most node w as our target node. Then $g_{i,w}(1) = g_{i,w}(2) = 0$ but $g_{i,w}(3) > 0$ since at least three steps are needed to visit the node w, which shows that $g_{i,w}(k_i)$ is not always non-increasing.

B.2. Starting From the Stationary Distribution

Lemma 2. $g_{i,u}(k_i + 1) - g_{i,u}(k_i) \geq 0$ for any $i \in [m]$, $u \in V_i, k_i \in \mathbb{Z}_{>0}$, if $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i$, where $\boldsymbol{\pi}_i^\top \mathbf{P}_i = \boldsymbol{\pi}_i^\top$.

Proof. Consider any layer L_i with transition probability matrix $\mathbf{P}_i$, if we start from the stationary distribution $\boldsymbol{\pi}_i$, the probability that node $u \in V_i$ is ever visited in the first k_i steps is $P_{i,u}(k_i) = (\boldsymbol{\pi}_i)^\top_u (\mathbf{I}_i)_{u,k_i} + \pi_{i,u}$, where $(\mathbf{I}_i)_{u,k_i} = (\mathbf{I} + \mathbf{T}_i(u) + \cdots + \mathbf{T}_i(u)^{k_i-2})(\mathbf{p}_i)_u$, $\mathbf{T}_i(u) = (\mathbf{P}_i)_{-u,-u}$, $(\mathbf{p}_i)_u = (\mathbf{P}_i[\cdot, u])^\top_{-u}$. Then the marginal gain for node u is $g_{i,u}(k_i) = P_{i,u}(k_i) - P_{i,u}(k_i - 1) = (\boldsymbol{\pi}_i)^\top_u \mathbf{T}_i(u)^{k_i-2}(\mathbf{p}_i)_u$ for $k_i \geq 2$ and $g_{i,u}(k_i) = \pi_{i,u}$ when $k_i = 1$.

Define the margin of the marginal gain as $\Delta_{i,u}(k_i) = g_{i,u}(k_i + 1) - g_{i,u}(k_i)$. When $k = 1$, $\Delta(u, 1) = g_{i,u}(2) - g_{i,u}(1) = (\boldsymbol{\pi}_i)^\top_u (\mathbf{p}_i)_u - \pi_{i,u} = \pi_{i,u} - \pi_{i,u} = \pi_{i,u}\mathbf{P}[u, u] - \pi_{i,u} = \pi_{i,u}\mathbf{P}[u, u] \geq 0$. When $k_i \geq 2$, $\Delta_{i,u}(k_i) = g_{i,u}(k_i + 1) - g_{i,u}(k_i) = (\boldsymbol{\pi}_i)_{-u}(\mathbf{T}_i(u)^{k_i-1} - \mathbf{T}_i(u)^{k_i-2})(\mathbf{p}_i)_u$. Because $\boldsymbol{\pi}_i^\top \mathbf{P}_i = \boldsymbol{\pi}_i^\top$, we have $(\boldsymbol{\pi}_i)^\top_u - (\boldsymbol{\pi}_i)^\top_u \mathbf{P}_i(u) = \pi_{i,u}(q_i)^\top_u$, where $(q_i)^\top_u = (\mathbf{P}_i[u, \cdot])_{-u}$. Thus, $\Delta_{i,u}(k_i) = (\boldsymbol{\pi}_{i,u})(q_i)^\top_u \mathbf{T}_i(u)^{k_i-2}(\mathbf{p}_i)_u \geq 0$ because any element in $\boldsymbol{\pi}_i$, $(q_i)^\top_u$, $(\mathbf{p}_i)_u$ and $\mathbf{P}_i(u)$ is non-negative.

More interestingly, the stationary distribution $\boldsymbol{\pi}_i$ is the only starting distribution for $\mathbf{P}_i$ such that any $u \in V_i; k_i \in \mathbb{Z}_{>0}$, $g_{i,u}(k_i + 1) - g_{i,u}(k_i) \geq 0$ when $\mathbf{P}_i$ is ergodic and there are no self loops in G_i.

Proof. With a little abuse of the notation, we use P to denote the transition probability matrix $P_i \in R^{n \times n}$ and let $P_{i;j}$ be the element in the i-th row and the j-th column. Since P is ergodic, according to Theorem 54 in (Serfozo, 2009), there exists a unique and positive stationary distribution $= (_1; :::; _n)$, i.e., $P = $ and > 0. Any starting distribution can be represented by $+$, where $= (_1; :::; _n)$ is a perturbation vector, and $_j \quad _j \quad 1 \quad _j; j \in [n]$. We have the following equation for the margin of marginal gains for node u in the first two steps,

$$_u = g_{i;u}(1) \quad g_{i;u}(2) = (_u + _u) \quad \sum_{j \in u} (_j + _j) P_{j;u}$$
$$= _u P_{uu} + _u \quad \sum_{j \in u} _j P_{j;u} :$$

Since G has no self loops, i.e., $P_{u;u} = 0$, we have $_u = \sum_{j \in u} P_{j;u}$, and we can verify that $\sum_{u \in [n]} _u = \sum_{u \in [n]} (1 \quad \sum_{j \in u} P_{u;j}) _u = 0$. Therefore, we have to guarantee $_u = 0$ for all u, otherwise there will exist a node u such that $_u < 0$. To ensure $_u = 0$, we need to ensure $P = $ by rephrasing the equations $_u = 0$ for all u. Again, since there exists a unique and positive stationary distribution for P and $\sum_{u \in [n]} _u = 0$, we can derive $= $, where has to be 0. Therefore, combined with Lemma 2, the stationary distribution is the only starting distribution such that for any P_i, any $u \in V_i; k_i \in Z_{>0}$, $g_{i;u}(k_i + 1) \quad g_{i;u}(k_i) \quad 0$.

C. Proofs of Of ine Optimization for Overlapping MuLaNE

C.1. Starting from the arbitrary distribution

Lemma 3. For any network G, distribution and weights , $r_{G; ;}( ) : Z^m_0 \to R$ is monotone and submodular.

Proof. By de nition, we need to show $r_{G; ;}(x \wedge y) + r_{G; ;}(x \vee y) \quad r_{G; ;}(x) + r_{G; ;}(y)$ for any $x; y \in Z^m_0$, and $r_{G; ;}(x) \quad r_{G; ;}(y)$ if $x \quad y$.

(Monotonicity.) By Eq. (7), $r_{G; ;}(k) = \sum_{v \in V} _v (1 \quad \prod_{i \in [m]} (1 \quad P_{i;v}(k_i)))$. Since $P_{j;v}(x_j) \quad P_{j;v}(x_j + 1)$, for any $j \in [m]; v \in V, x_j \in Z_{\quad 0}$, we have

$$r_{G; ;}(x + _j) \quad r_{G; ;}(x)$$
$$= \sum_{v \in V} _v \Big(\prod_{i \in j} (1 \quad P_{i;v}(x_i)) \Big) \big(P_{j;v}(x_j + 1) \quad P_{j;v}(k_j) \big)$$
$$0;$$

for any $x \in Z^m_0; j \in [m]$. Then we can use the above inequality repeatedly to show that $r_{G; ;}(x) \quad r_{G; ;}(y)$ when $x \quad y$.

(Submodularity.) For submodular property, it is sufficient to prove $(1 \quad \prod_{i \in [m]} (1 \quad P_{i;v}(k_i)))$ is submodular for any $v \in V$, because a positive weighted sum of submodular function is still submodular. We will rely on the following lemma to prove the submodularity of $r_{G; ;}(k)$.

Lemma 5. Function $f : Z^m_0 \to R$ is submodular if and only if

$$f(x + _i) \quad f(x) \quad f(x + _j + _i) \quad f(x + _j);$$
$$\text{for any } x \in Z^m_0 \text{ and } i \quad j. \tag{14}$$

Proof of Lemma 5. (If part.) We first prove, if inequality (14) holds, the following inequality holds,

$$f(x + _i) \quad f(x) \quad f(y + _i) \quad f(y) \tag{15}$$

for any $x \quad y$ and $i \in [m]$ such that $x_i = y_i$.

Let $I_0 = \{i \in [m] : x_i = y_i\}; I_1 = \{i \in [m] : x_i < y_i\}$. For any $x \quad y$, we denote the elements in I_1 by $i_1; :::; i_s$ and write $y = x + \sum_{j=1}^{s} _j _{i_j}$, where $_j = y_j \quad x_j$. For any $i \in I_0$, we have

$$\begin{aligned}
&f(x + _i) \quad f(x) \\
&\quad f(x + _{i_1} + _i) \quad f(x + _{i_1}) \\
&\quad f(x + 2 _{i_1} + _i) \quad f(x + 2 _{i_1}) \\
&\quad ::: \\
&\quad f(x + _{i_1} _{i_1} + _i) \quad f(x + _{i_1} _{i_1}) \\
&\quad f(x + _{i_1} _{i_1} + _{i_2} + _i) \quad f(x + _{i_1} _{i_1} + _{i_2}) \\
&\quad f(x + _{i_1} _{i_1} + _{i_2} _{i_2} + _i) \quad f(x + _{i_1} _{i_1} + _{i_2} _{i_2}) \\
&\quad ::: \\
&\quad f\Big(x + \sum_{j=1}^{s} _j _{i_j} + _i\Big) \quad f\Big(x + \sum_{j=1}^{s} _j _{i_j}\Big) \\
&= f(y + _i) \quad f(y):
\end{aligned} \tag{16}$$

Then for any $i \in I_0$ and $a \in Z_{\quad 0}$, we have

$$\begin{aligned}
&f(x + a _i) \quad f(x) \\
&= \sum_{j=1}^{a} (f(x + j _i) \quad f(x + (j \quad 1) _i)) \\
&\quad \sum_{j=1}^{a} (f(y + j _i) \quad f(y + (j \quad 1) _i)) \\
&\quad f(y + a _i) \quad f(y):
\end{aligned} \tag{17}$$

The first inequality holds because of Inequality (16), the fact $x + (j \quad 1) _i \quad y + (j \quad 1) _i$ and $x_i + j \quad 1 = y_i + j \quad 1$.

Then for any $x; y \in Z^m_0$, let $I_2 = \{i \in [m] : x_i > y_i\} =$

$\{i_1, \ldots, i_s\}$. We have

$$
\begin{aligned}
& f(x) - f(x \wedge y) \\
&= f\Big(x \wedge y + \sum_{l=1}^{s} \sum_{j=1}^{l} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j}\Big) \\
&\quad - f\Big(x \wedge y + \sum_{j=1}^{s-1} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j}\Big) \\
&\quad - f\Big(y + \sum_{l=1}^{s} \sum_{j=1}^{l} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j}\Big) \\
&\quad + f\Big(y + \sum_{j=1}^{s-1} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j}\Big) \\
&= f(x \vee y) - f(y).
\end{aligned}
$$

The inequality is derived from Inequality (17) because $y + \sum_{j=1}^{l-1} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j} \le x \wedge y + \sum_{j=1}^{l-1} (x_{i_j} - y_{i_j})\, \mathbf{1}_{i_j}$ for any $0 \le l \le s$ and $(x \wedge y)_{i_l} = y_{i_l}$, which concludes the if part.

(Only if part.) Assume f is submodular, let $a = x + \mathbf{1}_i$; $b = x + \mathbf{1}_j$; $i \ne j$, we have $a \vee b = x + \mathbf{1}_i + \mathbf{1}_j$, $a \wedge b = x$. $f(x + \mathbf{1}_i) - f(x) = f(a) - f(a \wedge b) \ge f(a \vee b) - f(b) = f(x + \mathbf{1}_j + \mathbf{1}_i) - f(x + \mathbf{1}_j)$.

Algorithm 4 Equivalent Budget Effective Greedy Algorithm (BEG) for the Overlapping MuLaNE.

Input: Network G, starting distributions $\cdot$, node weights $\cdot$, budget B, constraints c.
Output: Budget allocation k.

1: Compute visiting probabilities $(P_{i;u}(b))_{i \in [m];u \in V;b \in [c_i]}$ according to Eq. (9).
2: $k \leftarrow \mathrm{BEG}((P_{i;u}(b))_{i \in [m];u \in V;b \in [c_i]}, \cdot, B, c)$.
3: Procedure $\mathrm{BEG}((P_{i;u}(b))_{i \in [m];u \in V;b \in [c_i]}, \cdot, B, c)$
4: Let $k := (k_1, \ldots, k_m) \leftarrow 0$, $K \leftarrow B$.
5: Let $Q \leftarrow \{(i, b_i) \mid i \in [m], 1 \le b_i \le c_i\}$.
6: while $K > 0$ and $Q \ne \emptyset$ do
7: $(i^*, b^*) \leftarrow \arg\max_{(i,b) \in Q} \Delta(i, b; k)/b$. Eq. (11)
8: if $b^* \le K$ then
9: $k_{i^*} \leftarrow k_{i^*} + b^*$, $K \leftarrow K - b^*$.
10: Modify pairs $(i^*, b) \in Q$ to $(i^*, b - b^*)$.
11: Remove pairs $(i^*, b) \in Q$ such that $b \le 0$.
12: else
13: Remove (i^*, b^*) from Q.
14: end if
15: end while
16: for $i \in [m]$ do
17: if $r_{G;\cdot;\cdot}(c_i \mathbf{1}_i) > r_{G;\cdot;\cdot}(k)$, then $k \leftarrow c_i \mathbf{1}_i$.
18: end for return $k := (k_1, \ldots, k_m)$.
19: end Procedure

Then, by Lemma 5 and the explicit formula of the reward function given by Eq. (7), we can prove $g(x; v) - g(x + \mathbf{1}_l; v) \ge g(x + \mathbf{1}_j; v) - g(x + \mathbf{1}_l + \mathbf{1}_j; v)$ for any $x \in \mathbb{Z}^m_0$, $v \in V$ and $l \ne j \in [m]$, where $g(x; v) = \prod_{i \in [m]} (1 - P_{i;v}(x_i))$. This holds due to the fact that the left hand side equals to $\prod_{i \in [m] \setminus \{j,l\}} (1 - P_{i;v}(x_i))\,(1 - P_{j;v}(k_j))(P_{l;v}(k_l + 1) - P_{l;v}(k_l))$ and the right hand side equals to $\prod_{i \in [m] \setminus \{j,l\}} (1 - P_{i;v}(x_i))\,(1 - P_{j;v}(k_j + 1))(P_{l;v}(k_l + 1) - P_{l;v}(k_l))$, and the left hand side is larger or equal to the right hand side because $(1 - P_{j;v}(k_j)) \ge (1 - P_{j;v}(k_j + 1))$. By summation over all nodes $v \in V$ with node weights $\cdot_v \in [0, 1]$, we can prove the reward function is submodular.

Theorem 1. Algorithm 1 obtains a $(1 - e^{-\gamma}) \approx 0.357$-approximate solution, where γ is the solution of equation $e^{-\gamma} = 2 - 2\gamma$, to the overlapping MuLaNE problem.

Proof. For theoretical analysis, we first give a modified version of Alg. 1 in Alg. 4. Both algorithms provide the same solution k given the same problem instance $(G, \cdot, \cdot, B, c)$. To see this fact, Alg. 4 considers invalid tentative allocations (i^*, b^*) (adding it will exceed the total budget constraint B) and remove them in line 13 of Alg. 4, while Alg. 1 only considers valid allocations by directly removing invalid allocations in advance in line 10 of Alg. 1.

With a little abuse of the notation, we use $r(k)$ to represent the reward $r_{G;\cdot;\cdot}(k)$ of a given problem instance. Let $k^* \in \mathbb{Z}^m_0$ denote the optimal budget allocation and $k^j \in \mathbb{Z}^m_0$ denote the budget allocation before entering the j-th iteration of the while loop (line 6-13) in Alg. 4. After entering the j-th iteration, the algorithm tries to extend the current budget allocation k^j by choosing the pair (i^j, b^j) in line 7, which we denote as (i^j, b^j). Let s be the first iteration we can not extend the current solution, i.e., $k^s = k^{s+1}$ and $k^j < k^{j+1}$ for $j = 1, \ldots, s - 1$. If we can always extend the current solution, we set s to be $B + 1$. For analysis, we temporarily add (i^s, b^s) (in the algorithm, this pair is removed by line 13 in the s-th iteration) to form a "virtual" budget allocation $k^{s+1} = k^s + b^s \mathbf{1}_{i^s}$. Let $k_g \in \mathbb{Z}^m_0$ denote the solution returned by Alg. 4.

We first introduce lemmas describing two important properties given by the submodularity over the integer lattice.

Lemma 6. (Soma et al., 2014). Let $t : \mathbb{Z}^m_0 \to \mathbb{R}$ be a

submodular function. For any $x, y \in Z^m_0$, we have,

$$f(x \vee y) \leq f(x) + \sum_{i \in \mathrm{supp}^+(y-x)} (f(x + (y_i - x_i)e_i) - f(x)). \tag{18}$$

Lemma 7. (Soma et al., 2014). Let $f : Z^m_0 \to R$ be a monotone submodular function. For any $x, y \in Z^m_0$ with $x \leq y$ and $i \in [m]$ we have,

$$f(x \vee k e_i) - f(x) \geq f(y \vee k e_i) - f(y). \tag{19}$$

Then, we have the following lemma.

Lemma 8. For $j = 1, \ldots, s$,

$$r(k^{j+1}) \geq (1 - \frac{b^j}{B}) r(k^j) + \frac{b^j}{B} r(k^*). \tag{20}$$

Proof of Lemma 8. This is because

$$\begin{aligned} r(k^*) &\leq r(k^* \vee k^j) \\ &\leq r(k^j) + \sum_{i \in \mathrm{supp}^+(k^* - k^j)} r(k^j + (k^*_i - k^j_i)e_i) - r(k^j) \\ &= r(k^j) + \sum_{i \in \mathrm{supp}^+(k^* - k^j)} r(k^j + \alpha_i e_i) - r(k^j) \\ &\qquad (\text{Let } \alpha_i = k^*_i - k^j_i) \\ &\leq r(k^j) + \sum_{i \in \mathrm{supp}^+(k^* - k^j)} \frac{r(k^{j+1}) - r(k^j)}{b^j} \alpha_i \\ &\leq r(k^j) + B \frac{r(k^{j+1}) - r(k^j)}{b^j}. \end{aligned} \tag{21}$$

The second inequality comes from Lemma 6, the third inequality holds because of the greedy procedure in line 7 and the last inequality holds because $\sum_{i \in \mathrm{supp}^+(k^* - k^j)} \alpha_i \leq B$. By rearranging terms, Inequality (20) holds.

Next, We can prove the following lemma.

Lemma 9. For $l = 1, \ldots, s$,

$$r(k^{l+1}) \geq r(k^1) \prod_{j=1}^{l} \left(1 - \frac{b^j}{B}\right) + r(k^*) \left(1 - \prod_{j=1}^{l} \left(1 - \frac{b^j}{B}\right)\right). \tag{22}$$

Proof of Lemma 9. We can prove this lemma by induction on l. When $l = 1$, the lemma holds due to Lemma 8. Assume that the lemma holds for l, we have the following inequality holds,

$$\begin{aligned} r(k^{l+1}) &\geq (1 - \frac{b^l}{B}) r(k^l) + \frac{b^l}{B} r(k^*) \\ &\geq (1 - \frac{b^l}{B}) r(k^1) \prod_{j=1}^{l-1} \left(1 - \frac{b^j}{B}\right) \\ &\quad + (1 - \frac{b^l}{B}) r(k^*) \left(1 - \prod_{j=1}^{l-1} \left(1 - \frac{b^j}{B}\right)\right) + \frac{b^l}{B} r(k^*) \\ &= r(k^1) \prod_{j=1}^{l} \left(1 - \frac{b^j}{B}\right) + r(k^*) \left(1 - \prod_{j=1}^{l} \left(1 - \frac{b^j}{B}\right)\right); \end{aligned} \tag{23}$$

where the first inequality is due to Lemma 8 by setting $j = l$, and the second inequality is by the assumption for l. By induction, Lemma 9 holds.

We then consider the following cases.

Case 1. Suppose the total budget used for k^* is larger or equal to B, i.e., $\sum_{j=1}^{s-1} b^j \geq \beta B$, where $\beta \in [0, 1]$.

We have the following inequality.

$$r(k^s) \geq (1 - e^{-\beta}) r(k^*). \tag{24}$$

This is due to Lem. 9 by setting $l = s - 1$, combined with the fact that $k^1 = 0$ and $\prod_{j=1}^{s-1} \left(1 - \frac{b^j}{B}\right) \leq e^{-\beta}$. The later fact holds because,

$$\begin{aligned} \log\left(\prod_{j=1}^{s-1} \left(1 - \frac{b^j}{B}\right)\right) &= (s-1) \sum_{j=1}^{s-1} \frac{1}{s-1} \log\left(1 - \frac{b^j}{B}\right) \\ &\leq (s-1) \log\left(1 - \frac{1}{s-1} \sum_{j=1}^{s-1} \frac{b^j}{B}\right) \\ &\leq (s-1) \log(1 - \frac{\beta}{s-1}); \end{aligned}$$

where the first inequality holds because of the Jensen's Inequality (Raginsky et al., 2013) and the second inequality holds because $\sum_{j=1}^{s-1} \frac{b^j}{B} \leq \beta$. Then we can easily check $\prod_{j=1}^{s-1} \left(1 - \frac{b^j}{B}\right) \leq (1 - \frac{\beta}{s-1})^{s-1} \leq e^{-\beta}$.

Case 2. Suppose the total budget $\sum_{j=1}^{s-1} b^j \leq \beta B$. Then, we have $b^s > (1 - \beta)B$. We can prove the following inequality holds,

$$r(k_g) \geq (1 - \frac{1}{2}) r(k^*). \tag{25}$$

This is due to Inequality (21), we have

$$r(k) \quad r(k^s) + B \frac{r(k^{s+1}) \quad r(k^s)}{b^s}$$
$$r(k^s) + \frac{r(k^{s+1}) \quad r(k^s)}{1}$$
$$(1 + \frac{1}{1})r(k_g);$$

where the second inequality is due to $b^s > (1 \quad)B$ and the last equality is due to the fact $r(k^{s+1}) \quad r(k^s) \quad r(k_g)$, $r(k^s) \quad r(k_g)$. To see the above fact, we assume without loss of generality the pair $(i^s; b^s)$ improves k^s towards the optimal budget allocation, i.e. $k^{s+1}_{i^s} = b^s + k^s_{i^s} \quad k_{i^s}$. Otherwise, if $b^s + k^s_{i^s} > k_{i^s}$, we can safely delete $(i^s; b^s)$ in the queue Q and does not affect the greedy solution, the optimal solution and the analysis. Let $_{max} = \max_{i \in [m]} r(c_i \quad _i)$, we have $r(k^{s+1}) \quad r(k^s) \quad r(k^s \_ k_{i^s \quad i^s}) \quad r(k^s) \quad r(k_{i^s \quad i^s}) \quad r(0) \quad _{max} \quad r(k_g)$. Also, we can obtain that $r(k^s) \quad r(k_g)$ since $k^s \quad k_g$. By rearranging the terms, Inequality (25) holds.

Combining Inequality (24) and (25), we have

$$r(k_g) \quad \min_{\in [0;1]} \max(1 \quad e \quad ; 1 \quad \frac{1}{2})r(k)$$
$$(1 \quad e \quad)r(k);$$

where $$ is the solution for equation $e \quad = 2 \quad$.

C.2. Starting From the Stationary Distribution

Lemma 4. For any network G, stationary distributions $$ and node weights $$, function $r_{G; \; ;}() : Z^m_0 \; ! \; R$ is monotone and DR-submodular.

Proof. By definition, we need to show $r_{G; \; ;}(y + \_j) \quad r_{G; \; ;}(y) \quad r_{G; \; ;}(x + \_j) \quad r_{G; \; ;}(x)$, and $r_{G; \; ;}(x) \quad r_{G; \; ;}(y)$ for any $x \quad y, j \in [m]$.

We first introduce the following lemma to help us to prove the DR-submodularity.

Lemma 10. Function f is DR-submodular if and only if

$$f(x + \_i) \quad f(x) \quad f(x + \_j + \_i) \quad f(x + \_j);$$
$$\text{for any } x \in Z^m_0. \tag{26}$$

Proof of Lemma 10. (If part.) We can easily check this direction holds by using the similar argument for Inequality (16), where the only difference is we consider $i \in [m]$ instead of $i \in I_0$.

(Only if part.) We can set $y := x^0 + \_j$, $x := x^0 + \_i$, for any $i; j \in [m]; x^0 \in Z^m_0$, and use Inequality (15) to show the only if part holds.

Since $r_{G; \; ;}(k)$ is submodular for arbitrary starting distributions $$, Inequality (14) holds. Then consider any layer $v \in [m]$ and budget allocation $x \in Z^m_0$, it is sufficient to show $r_{G; \; ;}(x + 2\_j) \quad r_{G; \; ;}(x + \_j) \quad r_{G; \; ;}(x + \_j) \quad r_{G; \; ;}(x)$. Since the left hand side minus the right hand side equals to $\sum_{v \in V} \_v [\prod_{i \in [m]; i \neq j} (1 \quad P_{i;v}(x_i)) (P_{j;v}(x_j) \quad 2P_{j;u}(x_j + 1) + P_{j;v}(x_j + 2))]$, we only need to show $\sum_{v \in V} \_v [\prod_{i \in [m]; i \neq j} (1 \quad P_{i;v}(x_i)) (g_{j;v}(x_j + 2) \quad g_{j;v}(x_j + 1))] \quad 0$. The above inequality holds because of Lemma 2.

Theorem 2. Algorithm 2 obtains a $(1 \quad 1=e)$-approximate solution to the overlapping MulaNE with the stationary starting distributions.

Proof. We can observe that line 7 always select the pair $(i; b)$ with $b = 1$ because $r_{G; \; ;}(k + b\_i) \quad r_{G; \; ;}(k + (b \quad 1)\_i) \quad r_{G; \; ;}(k + \_i) \quad r_{G; \; ;}(k)$ for arbitrary k, i and b. Thus, we have $s = B + 1$ and by the similar argument for Inequality (24), we have $r(k_g) = r(k^{B+1}) \quad (1 \quad 1=e)r(k \quad)$, which completes the proof.

C.3. Efficiently Evaluating the Reward Function

One key issue to derive the budget allocation is to efficiently evaluate the reward function $r_{G; \; ;}(k)$ and its marginal gains $(i; b; k)$. Since we need to repetitively use the visiting probabilities $P_{i;u}(k_i)$, we pre-calculate $P_{i;u}(k_i)$ in $O(m \; kck_1 \; n^3_{max})$ time in our algorithms based on Eq. (9), where $n_{max} = \max_i |V_i|$. For the overlapping case, given the current budget allocation k, we maintain a value $p_u = \prod_{i \in [m]} (1 \quad P_{i;u}(k_i))$ for each node $u \in V$. The marginal gain $(i; b; k) = \sum_{u \in V} \_u p_u [1 \quad \frac{1 \quad P_{i;u}(k_i + b)}{1 \quad P_{i;u}(k_i)}]=b$ can be evaluated in $O(n_{max})$ time. Then, we update all $p_u = p_u \frac{1 \quad P_{i;u}(k_i + b)}{1 \quad P_{i;u}(k_i)}$ in $O(n_{max})$ after we allocated b more budgets to layer i. Therefore, we can use $O(n_{max})$ in total to evaluate $(i; b; k)$. In practice, lazy evaluation (Krause et al., 2008) and parallel computing can be used to further accelerate our algorithm.

D. Budget Effective Greedy Algorithm With Partial Enumeration and Its Analysis

D.1. Algorithm

The algorithm is shown in Alg. 5.

D.2. Analysis

Theorem 4. The Algorithm 5 obtains a $(1 \quad 1=e)$-approximate solution to the overlapping MuLaNE with ar-

Algorithm 5 Budget Effective Greedy Algorithm with Partial Enumeration (BEGE).

Input: Graph G, starting distributions α, node weights β, budget B, constraints c.
Output: Budget allocation k.

1: Compute visiting probabilities $(P_{i,u}(b))_{i\in[m],u\in V,b\in[c_i]}$ according to Eq. (7).
2: $k \leftarrow$ BEGE$((P_{i,u}(b))_{i\in[m],u\in V,b\in[c_i]}, \alpha, B, c)$.

3: Procedure BEGE$((P_{i,u}(b))_{i\in[m],u\in V,b\in[c_i]}, \alpha, B, c)$
4: Let $k_{max} \leftarrow 0$.
5: $S \leftarrow \{k = (k_1,\dots,k_m) \mid 0 \le k_i \le c_i; \sum_{i\in[m]} k_i \le B; \text{If } k_i > 0 \text{ then } k_j = c_j \text{ for } j < i\}$. // S contains all partial solutions which allocate partial budgets to at most three layers
6: for $k \in S$ do
7: $K \leftarrow B - \sum_{i\in[m]} k_i$.
8: Let $Q \leftarrow \{(i,b_i) \mid i \in [m], 1 \le b_i \le c_i - k_i\}$.
9: while $K > 0$ and $Q \neq \emptyset$ do
10: $(i^*, b^*) \leftarrow \arg\max_{(i,b)\in Q} \Delta(i,b;k)/b$. // Eq. (11)
11: if $b^* \le K$ then
12: $k_{i^*} \leftarrow k_{i^*} + b^*$, $K \leftarrow K - b^*$.
13: Modify all pairs $(i^*, b) \in Q$ to $(i^*, b - b^*)$.
14: Remove all pairs $(i^*, b) \in Q$ such that $b \le 0$.
15: else
16: Remove (i^*, b^*) from Q.
17: end if
18: end while
19: if $r_{G,\alpha,\beta}(k) > r_{G,\alpha,\beta}(k_{max})$, then $k_{max} \leftarrow k$.
20: end for return $k_{max} := (k_1,\dots,k_m)$.
21: end Procedure

bitrary starting distributions.

Proof. Suppose that we start from a partial solution $k^1 \in \mathbb{Z}^m_{\ge 0} \setminus 0$. Let us first reorder the optimal solution k according to a non-increasing marginal gain ordering. Namely, the marginal gain of the pair (s_1, b_1) with respect to the empty pair is the highest among all other pairs, the marginal gain of the pair (s_2, b_2) with respect to the solution consisting of pair (s_1, b_1) is the highest among all remaining pairs, and so on. To be concrete, $k = \{(s_1, b_1), \dots, (s_m, b_m)\}$, where $s_i \in [m] \setminus \{s_1, \dots, s_{i-1}\}$ is selected to maximize the following equation:

$$\left(r\left(\sum_{j=1}^{i-1} b_j \mathbf{1}_{s_j} + b \mathbf{1}_{s_i}\right) - r\left(\sum_{j=1}^{i-1} b_j \mathbf{1}_{s_j}\right)\right). \quad (27)$$

Then, we try to bound the "virtual" marginal gain $\Delta^s = r(k^{s+1}) - r(k^s)$, and recall s is the first iteration we can not extend the current solution. Without loss of generality, we assume the virtual pair (i^s, b^s) improves k^s towards the optimal budget allocation, i.e., $b^s + k^s_{i^s} \le k_{i^s}$. Otherwise, if $b^s + k^s_{i^s} > k_{i^s}$, we can safely delete (i^s, b^s) in the queue Q and does not affect the greedy solution, the optimal solution and the analysis.

If we start from the initial solution k^1 such that k^1 matches $(s_1, b_1), (s_2, b_2), (s_3, b_3)$, i.e., $k^1_{s_i} = b_i$ for $i = 1, 2, 3$, we have $\Delta^s = r(k^{s+1}) - r(k^s) \le r(k^s \vee k_{i^s} \mathbf{1}_{i^s}) - r(k^s) \le r(k_{i^s} \mathbf{1}_{i^s}) - r(0) \le r(b_1 \mathbf{1}_{s_1})$. Moreover, $\Delta^s = r(k^{s+1}) - r(k^s) \le r(k^s \vee k_{i^s} \mathbf{1}_{i^s}) - r(k^s) = r(k^s \vee b_1 \mathbf{1}_{s_1} \vee k_{i^s} \mathbf{1}_{i^s}) - r(k^s \vee b_1 \mathbf{1}_{s_1}) \le r(b_1 \mathbf{1}_{s_1} \vee k_{i^s} \mathbf{1}_{i^s}) - r(b_1 \mathbf{1}_{s_1}) \le r(b_1 \mathbf{1}_{s_1} + b_2 \mathbf{1}_{s_2}) - r(b_1 \mathbf{1}_{s_1})$. Similarly, we have $\Delta^s \le r(b_1 \mathbf{1}_{s_1} + b_2 \mathbf{1}_{s_2} + b_3 \mathbf{1}_{s_3}) - r(b_1 \mathbf{1}_{s_1} + b_2 \mathbf{1}_{s_2})$. By adding above inequalities, we have $3\Delta^s \ge r(k^1)$.

Now we can use Lemma 9 by setting $\ell = s$ and the fact $\sum_{j=1}^{s} b^j \le B$ to show $r(k^{s+1}) \ge r(k^1) + (1 - 1/e)(r(k^*) - r(k^1))$. Combining $3\Delta^s \ge r(k^1)$, we have $r(k^s) \ge (1 - 1/3) r(k^1) + (1 - 1/e)(r(k^*) - r(k^1)) \ge (1 - 1/e) r(k^*)$, which completes the proof.

The time complexity of Alg. 5 is $O(B^4 m^4 k c k_1 n_{max} + k c k_1 m n^3_{max})$. This is because the number of all partial solutions is $|S| = O(B^3 m^3)$, and for any partial enumeration $k \in S$, the time complexity is the same order as the Alg. 1, i.e., $O(B k c k_1 m n_{max})$, where $O(n_{max})$ is the time to evaluate $\Delta(i, b; k)$ as discussed before.

E. Algorithms and Analysis of Offline Optimization for Non-overlapping MuLaNE

E.1. Starting From the Stationary Distribution

According to Eq. (8) and (10), the reward function can be rewritten as

$$r_{G,\alpha,\beta}(k) = \sum_{i\in[m]} \sum_{b\in[k_i]} g_i(b), \quad (28)$$

where $g_i(b)$ represents the layer-level marginal gain when we allocate one more budget (from $b-1$ to b) to layer i, which is given by

$$g_i(b) = \sum_{v\in V_i} \beta_v (P_{i,v}(b) - P_{i,v}(b-1)). \quad (29)$$

From the definition of the reward function, we have two observations. First, budgets allocated to a layer will not affect the reward of other layers because layers are non-overlapping. Second, the layer-level marginal gain $g_i(b)$ for the i-th layer is non-increasing with with respect to budget b. Based on these two observation, we propose the myopic greedy algorithm (Alg. 6), which is a slight modification of Alg. 2. Alg. 6 takes the multi-layered network $G = (G_1, \dots, G_m)$, starting distributions $(\alpha_1, \dots, \alpha_m)$, node weights $\beta = (\beta_1, \dots, \beta_{|V|})$ and total budget B as inputs. It

consists of B rounds to compute the optimal budget allocation $\boldsymbol{k}$. In each round, line 7 in Alg. 6 selects the layer with the largest layer-level marginal gain and allocate one budget to layer i until total B budgets are used up.

Algorithm 6 Myopic Greedy Algorithm for the Non-overlapping MuLaNE

Input: Network G, starting distributions $\boldsymbol{\alpha}$, budget B, constraints $\boldsymbol{c}$.
Output: Budget allocation $\boldsymbol{k}$.

1: Compute visiting probabilities $(P_{i;u}(b))_{i\in[m];u\in V;b\in[c_i]}$ according to Eq. (8).
2: Compute layer-level marginal gain $(g_i(b))_{i\in[m];b\in[c_i]}$ according to Eq. (29).
3: $\boldsymbol{k} \leftarrow MG((g_i(b))_{i\in[m];b\in[c_i]}, B, \boldsymbol{c})$
4: **Procedure** $MG((g_i(b))_{i\in[m];b\in[c_i]}, B, \boldsymbol{c})$
5: Let $\boldsymbol{k} := \{k_1, \dots, k_m\} \leftarrow \boldsymbol{0}, K \leftarrow B$.
6: **while** $K > 0$ **do**
7: $i \leftarrow \arg\max_{i\in[m];k_i+1\le c_i} g_i(k_i+1)$: $O(\log m)$ using the priority queue
8: $k_i \leftarrow k_i + 1, K \leftarrow K - 1$.
9: **end while return** $\boldsymbol{k} = (k_1, \dots, k_m)$.
10: **end Procedure**

Algorithm 7 Dynamic Programming (DP) Algorithm for the Non-overlapping MuLaNE

Input: Network G, starting distributions $\boldsymbol{\alpha}$, node weights $\boldsymbol{\beta}$, budget B, constraints $\boldsymbol{c}$.
Output: Budget allocation $\boldsymbol{k}$.

1: Compute visiting probabilities $(P_{i;u}(b))_{i\in[m];u\in V;b\in[c_i]}$ according to Eq. (8).
2: Compute layer-level marginal gain $(g_i(b))_{i\in[m];b\in[c_i]}$ according to Eq. (29).
3: $\boldsymbol{k} \leftarrow DP((g_i(b))_{i\in[m];b\in[c_i]}, B, \boldsymbol{c})$
4: **Procedure** $DP((g_i(b))_{i\in[m];b\in c_i}, B, \boldsymbol{c})$
5: **for** $i \in 0[m]; b \in 0[B]$. **do**
6: Set $A[i;b] \leftarrow 0, V[i;b] \leftarrow 0$.
7: **end for**
8: **for** $i \leftarrow 0$ to $m-1$ **do**
9: **for** $b \in [B]$ **do**
10: $j \leftarrow \arg\max_{j\in 0[c_{i+1}]}(V[i;b-j] + r_{G;\alpha;\beta}(j_{i+1}))$:
11: $A[i+1;b] \leftarrow A[i;b-j] + j_{i+1}$.
12: $V[i+1;b] \leftarrow V[i;b-j] + r_{G;\alpha;\beta}(j_{i+1})$:
13: **end for**
14: **end for return** $A[m;B]$.
15: **end Procedure**

Theorem 5. The Algorithm 6 obtains the optimal budget allocation to the non-overlapping MuLaNE with the stationary starting distributions.

Proof. Define a two dimensional array $M \in R^{c_1+c_2+\dots+c_m}$, where $(i;j)$-th entry is the j-th step layer-level marginal gain for layer i, i.e., $M[i;j] = \sum_{v\in V_i} g_{i;v}(j)$, for $i \in [m]; j \in [c_i]$. Given the budget allocation $\boldsymbol{k} = \{k_1, \dots, k_m\}$, the expected reward $r_{G;\alpha;\beta}(\boldsymbol{k})$ can be written as the sum of elements in M, i.e., $r_{G;\alpha;\beta}(\boldsymbol{k}) = \sum_{i=1}^{m}\sum_{j=1}^{k_i} M[i;j]$. Because $g_{i;v}(k_i)$ is non-increasing with respect to k_i, the element in each row or the layer-level marginal gain is non-increasing. Hence, the greedy method at step s choose the s-th largest element in M. At step $s = B$, the greedy policy selects all B largest elements and the corresponding budget allocation maximizes the expected reward $r_{G;\alpha;\beta}(\boldsymbol{k})$.

Alg. 6 uses $O(kck_1 mn^3_{max})$ to compute visiting probabilities $P_{i;u}(k_i)$ and $O(kck_1 mn_{max})$ to compute layer-wise marginal gains $g_i(b)$. Then $(i; 1; k)$ can be evaluated in $O(\log m)$ using the priority queue, which is repeated for B iterations. Therefore, the time complexity of Alg. 6 is $O(B \log m + kck_1 mn^3_{max})$.

E.2. Starting From Arbitrary Distributions

When random explorers start from any arbitrary distribution, Alg. 6 can not obtain the optimal solution. This is because the layer-level marginal gain $g_i(b)$ is not non-increasing with respect to b. So we have to adopt a more general technique, dynamic programming (DP), to solve the budget allocation problem. The key idea is to keep a DP budget allocation table $A \in R^{(m+1)\times(B+1)}$, where the $(i;b)$-th entry $A[i;b]$ saves the optimal budget allocation by allocating b budgets to the first i layers. Another DP table $V \in R^{(m+1)\times(B+1)}$ saves the value of the optimal reward, where the $(i;b)$-th entry $V[i;b]$ corresponds to the optimal reward when setting $A[i;b]$ as the budget allocation. In the i'-th outer loop and b'-th inner loop in Alg 7, we have already obtained the optimal budget allocation $A[i;b]$ with $i = 0, \dots, i'$, $b \in [B]$, which help us to find the optimal amount of budget j to $(i'+1)$-th layer (in line 10), and thus we can obtain the $A[i'+1;b']$ and $V[i'+1;b']$ accordingly.

Theorem 6. Algorithm 7 obtains the optimal budget allocation to the non-overlapping MuLaNE with an arbitrary starting distribution.

Alg. 7 uses $O(kck_1 mn^3_{max})$ to compute visiting probabilities $P_{i;u}(k_i)$ and $O(kck_1 mn_{max})$ to compute layer-wise marginal gains $g_i(b)$. In the DP procedure, we can also precalculate all rewards $r_{G;\alpha;\beta}(j_{i+1})$ in $O(kck_1 m)$. Then DP procedure takes $O(kck_1)$ to find j, which is repeated for Bm iterations. Therefore, the Alg. 7 has the time complexity of $O(B\, kck_1 m + kck_1 mn^3_{max})$.

F. Analysis of Online Learning for Overlapping MuLaNE

F.1. Proof of 1-Norm Bounded Smoothness.

Property 1. (Monotonicity). The reward $r_{\cdot;\cdot}(k)$ is monotonically increasing, i.e., for any budget allocation k, any two vectors $=(_{i;u;b})_{(i;u;b)2A}$, $^0=(^0_{i;u;b})_{(i;u;b)2A}$ and any node weights $$, ^{0}, we have $r_{;}(k)$ $r_{^0;^0}(k)$, if $_{i;u;b}$ $^0_{i;u;b}$ and $_v$ ^0_v, $8(i;u;b)$ 2 A ; v 2 V.

Proof. According to Eq. (7) and since $_{i;u;b}$ $^0_{i;u;b}$, for $(i;u;b)$ 2 A, we have $(1$ $\prod_{i2[m]}(1$ $_{i;u;k_i}))$ $(1$ $\prod_{i2[m]}(1$ $^0_{i;u;k_i}))$. Therefore for any $_u$ ^0_u, we have $_u(1$ $\prod_{i2[m]}(1$ $_{i;u;k_i}))$ $^0_u(1$ $\prod_{i2[m]}(1$ $^0_{i;u;k_i}))$

By summing up both sides over $u2$ V, we have $r_k($ $;$ $)$ $r_k(^0;$ $^0)$.

Property 2. (1-Norm Bounded Smoothness). The reward function $r_{\cdot;\cdot}(k)$ satisfies the 1-norm bounded smoothness condition, i.e., for any budget allocation k, any two vectors $=(_{i;u;b})_{(i;u;b)2A}$, $^0=(^0_{i;u;b})_{(i;u;b)2A}$ and any node weights $$, ^{0}, we have $|r_{;}(k)$ $r_{^0;^0}(k)| \le \sum_{i2[m];u2V;b=k_i}(|_u|\,|_{i;u;b}$ $^0_{i;u;b}|+|_u|\,|_u$ $^0_u|\,^0_{i;u;b})$.

Proof. The left-hand side:

$$
\begin{aligned}
&|r_{;}(k) - r_{^0;^0}(k)| \\
&= \Big|\sum_{u\in V}\Big(1-\prod_{i\in[m]}(1-_{i;u;k_i})\Big)\Big(_u\prod_{i\in[m]}(1-_{i;u;k_i})-{}^0_u\prod_{i\in[m]}(1-{}^0_{i;u;k_i})\Big)\Big| \\
&\le \sum_{u\in V}\Big|_u\prod_{i\in[m]}(1-q_{i,u})-{}^0_u\prod_{i\in[m]}(1-q^0_{i,u})\Big| \\
&= \sum_{u\in V}\Big|_u\prod_{i\in[m]}(1-q_{i,u})-{}_u\prod_{i\in[m]}(1-q^0_{i,u})+{}_u\prod_{i\in[m]}(1-q^0_{i,u})-{}^0_u\prod_{i\in[m]}(1-q^0_{i,u})\Big| \\
&\le \sum_{u\in V}|_u|\Big|\prod_{i\in[m]}(1-q_{i,u})-\prod_{i\in[m]}(1-q^0_{i,u})\Big| \\
&\quad+\sum_{u\in V}|_u-{}^0_u|\prod_{i\in[m]}(1-q^0_{i,u}),
\end{aligned}
$$

where $q_{i;u}=(1$ $_{i;u;k_i})$; $q^0_{i;u}=(1$ $^0_{i;u;k_i})$.

For the first term, it can be derived that,

$$
\begin{aligned}
&\Big|\prod_{i\in[m]}(1-q_{i,u})-\prod_{i\in[m]}(1-q^0_{i,u})\Big| \\
&= \Big|\sum_{i=1}^{m}\Big(\prod_{j=1}^{i-1}(1-q^0_{j,u})\,(q_{i,u}-q^0_{i,u})\prod_{j=i+1}^{m}(1-q_{j,u})\Big)\Big| \\
&\le \sum_{i=1}^{m}|q_{i,u}-q^0_{i,u}| \\
&= \sum_{i=1}^{m}|_{i;u;k_i}-{}^0_{i;u;k_i}|,
\end{aligned}
$$

where last inequality holds because $q_{i;u}$; $q^0_{i;u}$ 1 and $_u$ 2 $[0;1]$, for any j 2 $[m]$; u 2 V.

For the second term, we have

$$
\begin{aligned}
&|_u-{}^0_u|\prod_{i\in[m]}(1-q^0_{i,u}) \\
&= |_u-{}^0_u|\prod_{i\in[m]}(1-(1-{}^0_{i;u;k_i})) \\
&\le |_u-{}^0_u|\sum_{i\in[m]}{}^0_{i;u;k_i},
\end{aligned}
$$

where the inequality holds because of the Weierstrass product inequality. Plugging the above two terms into the previous inequality, 1-Norm Bounded Smoothness is satisfied.

F.2. Regret Analysis for Overlapping MuLaNE

In this section, we give the regret analysis for overlapping MuLaNE.

F.2.1. Facts

We utilize the following tail bound in our analysis.

Fact 1 (Hoeffding's Inequality (Dubhashi & Panconesi, 2009).) Let $Y_1;:::;Y_n$ be independent and identically distributed (i.e., i.i.d) random variables with common suppport $[0;1]$ and mean $$. Let $Z = Y_1 + ::: + Y_n$: Then for all $$,

$$Prfj Z \quad n \; j \quad g \quad 2e^{\;2\;^2=n}:$$

F.2.2. Proof Details

Let A denote the set containing all base arms, i.e., $A=$ $f(i;u;b)ji$ 2 $[m];u$ 2 $V;b$ 2 $[c_i]g$. We add subscript t to denote the value of a variable at the end of round $t2[T]$, e.g. $_t;\hat{}_{i;u;b;t}$, where T is the total number of rounds. For example, $T_{i;u;b;t}$ denotes the total number of times that arm

(i,u,b) is played at the end of round t. Let us first introduce a definition of an unlikely event, where $\hat{\mu}_{i,u,b,t-1}$ is not accurate as expected.

Definition 2. We say that the sampling is nice at the beginning of round t, if for every arm $(i,u,b) \in A$, $|\hat{\mu}_{i,u,b,t-1} - \mu_{i,u,b}| < c_{i,u,b,t}$, where $c_{i,u,b,t} = \sqrt{\frac{3\ln t}{2T_{i,u,b,t-1}}}$. Let N_t be such event.

Lemma 11. For each round $t \geq 1$, $\Pr\{\neg N_t\} \leq 2|A|t^{-2}$.

Proof. For each round $t \geq 1$, we have

$$\begin{aligned}
&\Pr\{\neg N_t\} \\
&= \Pr\{\exists (i,u,b) \in A, |\hat{\mu}_{i,u,b,t-1} - \mu_{i,u,b}| \geq \sqrt{\frac{3\ln t}{2T_{i,u,b,t-1}}}\} \\
&\leq \sum_{(i,u,b)\in A} \Pr\{|\hat{\mu}_{i,u,b,t-1} - \mu_{i,u,b}| \geq \sqrt{\frac{3\ln t}{2T_{i,u,b,t-1}}}\} \\
&= \sum_{(i,u,b)\in A} \sum_{k=1}^{t} \Pr\{T_{i,u,b,t-1} = k \\
&\quad \wedge |\hat{\mu}_{i,u,b,t-1} - \mu_{i,u,b}| \geq \sqrt{\frac{3\ln t}{2T_{i,u,b,t-1}}}\} \\
&\leq \sum_{(i,u,b)\in A} \sum_{k=1}^{t} \frac{2}{t^3} \leq 2|A|t^{-2}.
\end{aligned}$$

Given $T_{i,u,b,t-1} = k$, $\hat{\mu}_{i,u,b,t-1}$ is the average of k i.i.d. random variables $Y_{i,u,b}^{[1]}, \ldots, Y_{i,u,b}^{[k]}$, where $Y_{i,u,b}^{[j]}$ is the Bernoulli random variable (computed in Alg. 3 line 12) when the arm (i,u,b) is played for the j-th time during the execution. That is, $\hat{\mu}_{i,u,b,t-1} = \sum_{j=1}^{k} Y_{i,u,b}^{[j]}/k$. Note that the independence of $Y_{i,u,b}^{[1]}, \ldots, Y_{i,u,b}^{[k]}$ comes from the fact we select a new starting position following the starting distribution $\mathbf{q}$ at the beginning of each round t. And the last inequality uses the Hoeffding Inequaility (Fact 1).

Then we can use monotonicity and 1-norm bounded smoothness properties (Property 1 and Property 2) to bound the reward gap $\Delta_{k_t} = r_{\mathbf{p},\mathbf{v}}(k^*) - r_{\mathbf{p},\mathbf{v}}(k_t)$ between the optimal action k^* and the action $k_t = (k_{t,1}, \ldots, k_{t,m})$ selected by our algorithm A. To achieve this, we introduce a positive real number $M_{i,u,b}$ for each arm (i,u,b) and define $M_{k_t} = \max_{(i,u,b)\in S_t} M_{i,u,b}$, where $S_t = \{(i,u,b) \in A \mid b = k_{t,i}\}$. Define $|E^0| = |S_t| = \sum_{i\in[m]} |V_i| = m|V|$, and let

$$\ell_T(M,s) = \begin{cases} 2, & \text{if } s = 0; \\ 3\sqrt{\frac{6\ln T}{s}}, & \text{if } 1 \leq s \leq \ell_T(M); \\ 0, & \text{if } s \geq \ell_T(M)+1; \end{cases}$$

where

$$\ell_T(M) = \left\lceil \frac{54|E^0|^2 \ln T}{M^2} \right\rceil.$$

Lemma 12. If $\{\Delta_{k_t} \leq M_{k_t}\} \wedge N_t$ holds, we have

$$\Delta_{k_t} \leq \sum_{(i,u,b)\in S_t} \ell_T(M_{i,u,b}, T_{i,u,b,t-1}). \quad (30)$$

Proof. First, we can observe, for $(i,u,b) \in A$, $c_{i,u,b}$ is upper bounded by $c_{i,u,b,t}$, i.e., $c_{i,u,b,t} \leq c_{i,u,b}$, when N_t holds. This is due to

$$\begin{aligned}
c_{i,u,b,t} &= \max_{j\in[b]} \tilde{c}_{i,u,j,t} \\
\tilde{c}_{i,u,b,t} &= \min\{\sqrt{\frac{3\ln t}{2T_{i,u,b,t-1}}}, 1\} \\
&= \min\{\hat{c}_{i,u,b,t-1} + c_{i,u,b,t}, 1\} \\
&\leq c_{i,u,b};
\end{aligned}$$

where the last inequality comes from the fact that $|\hat{\mu}_{i,u,b,t-1} - \mu_{i,u,b}| < c_{i,u,b,t}$ when N_t holds.

Then, notice that the right hand of Inequality (30) is non-negative, it is trivially satisfied when $\Delta_{k_t} = 0$. We only need to consider $\Delta_{k_t} > 0$. By N_t and Property 1, We have the following inequalities,

$$r_{\mathbf{p},\mathbf{v}}(k_t) \geq r_{\mathbf{p},\mathbf{v}}(k^*) - r_{\mathbf{p},\mathbf{v}}(k^*) = \Delta_{k_t} + r_{\mathbf{p},\mathbf{v}}(k_t).$$

The first inequality is due to the oracle outputs the (α,β)-approximate solution given parameters $\mathbf{p}$, where $\alpha = 1 - e^{-1}$, $\beta = 1$ by definition. The second inequality is due to Condition 1 (monotonicity).

Then, by Condition 2, we have

$$\begin{aligned}
\Delta_{k_t} &\leq r_{\mathbf{p},\mathbf{v}}(k_t) - r_{\mathbf{p},\hat{\mathbf{v}}}(k_t) \\
&\leq \sum_{(i,u,b)\in S_t} (|\mu_{i,u,b,t} - \mu_{i,u,b}| + |\mu_u - \hat{\mu}_{i,u,b,t}|) \\
&\leq \sum_{(i,u,b)\in S_t} (|\mu_{i,u,b,t} - \mu_{i,u,b}| + c_{i,u,b,t});
\end{aligned}$$

where the last inequality is due to the following observation: $\mu_u \neq \hat{\mu}_u$ if and only if u has not been visited so far, so $\hat{\mu}_{i,u,b,t-1} = 0$ for any $i \in [m], b \in [B]$ and $c_{i,u,b,t} = \max_{j\in[b]}\{\hat{c}_{i,u,j,t-1} + c_{i,u,j,t}\} = c_{i,u,b,t}$.

Next, we can bound Δ_{k_t} by bounding $|\mu_{i,u,b,t} - \mu_{i,u,b}| + c_{i,u,b,t}$ by applying a transformation. As $\{\Delta_{k_t} \leq M_{k_t}\}$ holds by assumption, so $\sum_{(i,u,b)\in S_t} |\mu_{i,u,b,t} - \mu_{i,u,b}| +$

$_{i;u;b;t}$ $_{k_t}$ M_{k_t}. We have

$$\sum_{k_t} \sum_{(i;u;b)\in S_t} (j\ _{i;u;b;t} - {}_{i;u;b}j + {}_{i;u;b;t})$$
$$\le \sum_{(i;u;b)\in S_t} M_{k_t} + 2\,|\ {}_{i;u;b;t} - {}_{i;u;b}| + {}_{i;u;b;t}$$
$$= 2 \sum_{(i;u;b)\in S_t} (j\ _{i;u;b;t} - {}_{i;u;b}j + {}_{i;u;b;t} - \frac{M_{k_t}}{2jE^0j})$$
$$\le 2 \sum_{(i;u;b)\in S_t} (j\ _{i;u;b;t} - {}_{i;u;b}j + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j}). \quad (31)$$

By N_t, for all $(i;u;b)\in A$, we have:

$$j\ _{i;u;b;t} - {}_{i;u;b}j + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j}$$
$$= \tilde{}_{i;u;l;t} - {}_{i;u;b} + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j} \quad (l = \arg\max_{j\in[b]} \tilde{}_{i;u;j;t})$$
$$\le \tilde{}_{i;u;l;t} - {}_{i;u;l} + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j}$$
$$\le \min\{2\ _{i;u;l;t} + {}_{i;u;b;t}; 1\} - \frac{M_{i;u;b}}{2jE^0j}$$
$$\le \min\{3\ _{i;u;b;t}; 1\} - \frac{M_{i;u;b}}{2jE^0j}$$
$$\le \left(\sqrt{\frac{3\ln T}{2T_{i;u;b;t-1}}}\right) \wedge 1 - \frac{M_{i;u;b}}{2jE^0j}.$$

where the first inequality is due to $_{i;u;j} \le {}_{i;u;b}$ for $j \le b$, the second inequality is due to the event $_t$, and the third inequality holds because $T_{i;u;j;t-1} \le T_{i;u;b;t-1}$ for $j \le b$.

Now consider the following cases:

Case 1. If $T_{i;u;b;t-1} \ge \grave{}_T(M_{i;u;b})$, we have $_{i;u;b} + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j} \le \min\left\{3\sqrt{\frac{3\ln T}{2T_{i;u;b;t-1}}}; 1\right\} \le \frac{1}{2}\ _T(M_{i;u;b}; T_{i;u;b;t-1})$.

Case 2. If $T_{i;b;t-1} < \grave{}_T(M_{i;u;b;t}) + 1$, then $3\sqrt{\frac{3\ln T}{2T_{i;u;b;t-1}}} \le \frac{M_{i;u;b}}{2jE^0j}$, so $_{i;u;b;t} - {}_{i;u;b} + {}_{i;u;b;t} - \frac{M_{i;u;b}}{2jE^0j} \le 0 = \frac{1}{2}\ _T(M_{i;u;b}; T_{i;u;b;t-1})$.

Combine these two cases and inequality (31), we have the following inequality, which concludes the lemma.

$$\sum_{k_t} \le \sum_{(i;u;b)\in S_t} \ _T(M_{i;u;b}; T_{i;u;b;t-1}).$$

Theorem 3. Algorithm 3 has the following distribution-dependent $(1 - e\ ; 1)$ approximation regret,

$$\mathrm{Reg}_{\ ;\ }(T) \le \sum_{(i;u;b)\in A} \frac{108m|V|\ln T}{{}^{2}_{i;u;b\,\min}} + 2|A| + \frac{2}{3}|A|\ _{\max}.$$

Proof of Theorem 3. By above lemmas, we have

$$\sum_{t=1}^{T} I(f\ _{k_t} \le M_{k_t}g \wedge N_t)\ _{k_t}$$
$$\le \sum_{t=1}^{T} \sum_{(i;u;b)\in S_t} \ _T(M_{i;u;b}; T_{i;u;b;t-1})$$
$$= \sum_{(i;u;b)\in A} \sum_{t\in[T]:k_{t;l}=b} \ _T(M_{i;u;b}; T_{i;u;b;t-1})$$
$$\le \sum_{(i;u;b)\in A} \sum_{s=0}^{T_{i;u;b;T}} \ _T(M_{i;u;b}; s)$$
$$\le \sum_{(i;u;b)\in A} \grave{}_T(M_{i;u;b}) \cdot 3\sqrt{\frac{6\ln T}{s}}$$
$$\le 2|A| + \sum_{(i;u;b)\in A} \sum_{s=1}^{\grave{}_T(M_{i;u;b})} 3\sqrt{\frac{6\ln T}{s}}$$
$$\le 2|A| + \sum_{(i;u;b)\in A} \int_{s=0}^{\grave{}_T(M_{i;u;b})} 3\sqrt{\frac{6\ln T}{s}}\,ds$$
$$\le 2|A| + \sum_{(i;u;b)\in A} \sqrt{6\cdot 6\ln T \cdot \grave{}_T(M_{i;u;b})}$$
$$\le 2|A| + \sum_{(i;u;b)\in A} \frac{108jE^0j\ln T}{M_{i;u;b}}.$$

For distribution dependent bound, let us set $_{i;u;b} = {}^{i;u;b}_{\min}$ and thus the event $_{k_t} \le M_{k_t}g$ always holds, i.e., If $_{k_t} < M_{k_t}g = 0$. We have

$$\sum_{t=1}^{T} I(N_t \wedge f\ _{k_t} \le M_{k_t}g)\ _{k_t}$$
$$\le 2|A| + \sum_{(i;u;b)\in A} \frac{108jE^0j\ln T}{{}^{i;u;b}_{\min}}.$$

By Lem. 11, $\Pr\{\ N_t g \le 2|A|\ ^2\}$, then

$$E\left[\sum_{t=1}^{T} I(\cdot N_t)\ _{k_t}\right] \le \sum_{t=1}^{T} 2|A|\ ^2\ _{\max} \le \frac{2}{3}|A|\ _{\max}.$$

By our choice of $M_{i;u;b}$,

$$\sum_{t=1}^{T} I(\ _{k_t} < M_{k_t})\ _{k_t} = 0.$$

By the definition of the regret,

$$\mathrm{Reg}(T) = E\left[\sum_{t=1}^{T} r_{k_t}\right] - E\left[\sum_{t=1}^{T} \mathbb{I}\{N_t\} \cdot \mathbb{I}\{r_{k_t} < M_{k_t}\}\right] + \mathbb{I}\{N_t\} \cdot f(k_t, M_{k_t}, g) \le \sum_{(i,u,b)\in A} \left(\frac{108 m |V| \ln T}{\Delta_{i,u,b}^{\min}} + 2|A| + \frac{2}{3}|A|\right) r_{\max}.$$

For distribution independent bound, we take $M_{i,u,b} = M = \sqrt{108 m |V| |A| \ln T / T}$, then we have $\sum_{t=1}^{T} \mathbb{I}\{r_{k_t} < M_{k_t}\} \le T M$.

Then

$$\mathrm{Reg}(T) \le \sum_{(i,u,b)\in A} \frac{108 m |V| \ln T}{M_{i,u,b}} + 2|A| + \frac{2}{3}|A| r_{\max} + T M = \frac{108 m |V| |A| \ln T}{M} + 2|A| + \frac{2}{3}|A| r_{\max} + T M = 2\sqrt{108 m |V| |A| T \ln T} + \frac{2}{3}|A| r_{\max} + 2|A| \le 21\sqrt{m |V| |A| T \ln T} + \frac{2}{3}|A| r_{\max} + 2|A|.$$

F.3. Extension for Random Node Weights

In this section, we extend the deterministic node weight w_u for node $u \in V$ to a random weight. Specifically, we denote $\tilde{w}_u \in [0,1]$ as the random weight when we first visit u (after the first visit we will not get any reward so on so forth), and denote $\mu_u := E[\tilde{w}_u]$.

For the offline setting, the reward function in Eq.(7) remains the same since $\tilde{w}_u$ is independent of all other random variables. For the online setting, since the reward function remains unchanged, property 1 and property 2 still hold. However, we need to change the way we maintain the optimistic node weight $\bar{w}_u$.

We present our CUCB-MAX-R algorithm in Alg. 8. Compared with Alg. 3, the major difference is that $\bar{w}_u$ now represent the upper confidence bound (UCB) value of $\tilde{w}_u$. In line 2, we denote T_v^0 as the number of times node v is played so far and $\hat{w}_v$ as the empirical mean of $\tilde{w}_v$'s weight. In line 9, ρ_v^0 is the confidence radius and $\bar{w}_u$ is the UCB value of node u. In line 15, we update the empirical mean $\hat{w}_v$ for any visited v.

Algorithm 8 CUCB-MAX-R: Combinatorial Upper Confidence Bound algorithm for the overlapping MuLaNE, with Random node weights

Input: Budget B, number of layers m, number of nodes $|V|$, constraints c, offline oracle BEG.

1: For each arm $(i,u,b) \in A$, $T_{i,u,b} \leftarrow 0$, $\hat{\pi}_{i,u,b} \leftarrow 0$.
2: For each node $v \in V$, $T_v^0 \leftarrow 0$, $\hat{w}_v \leftarrow 0$.
3: **for** $t = 1, 2, 3, \ldots, T$ **do**
4: **for** $(i,u,b) \in A$ **do**
5: $\rho_{i,u,b} \leftarrow \sqrt{3 \ln t / (2 T_{i,u,b})}$.
6: $\bar{\pi}_{i,u,b} \leftarrow \min\{\hat{\pi}_{i,u,b} + \rho_{i,u,b}, 1\}$.
7: **end for**
8: **for** $v \in V$ **do**
9: $\rho_v^0 \leftarrow \sqrt{3 \ln t / (2 T_v^0)}$.
10: $\bar{w}_v \leftarrow \min\{\hat{w}_v + \rho_v^0, 1\}$.
11: **end for**
12: For $(i,u,b) \in A$, $\mu_{i,u,b} \leftarrow \max_{j \in [b]} \tilde{\mu}_{i,u,j}$.
13: $k \leftarrow \mathrm{BEG}((\mu_{i,u,b})_{(i,u,b)\in A}, (\bar{w}_v)_{v\in V}, B, c)$.
14: Apply budget allocation k, which gives trajectories $X := (X_{i,1}, \ldots, X_{i,k_i})_{i\in[m]}$ as feedbacks.
15: For any visited node $v \in \bigcup_{i\in[m]} \{X_{i,1}, \ldots, X_{i,k_i}\}$, receive its random node weight $\tilde{w}_v$, update $T_v^0 \leftarrow T_v^0 + 1$, $\hat{w}_v \leftarrow \hat{w}_v + (\tilde{w}_v - \hat{w}_v)/T_v^0$
16: For any $(i,u,b) \in O := \{(i,u,b) \in A \mid b \le k_i\}$, $Y_{i,u,b} \leftarrow 1$ if $u \in \{X_{i,1}, \ldots, X_{i,b}\}$ and 0 otherwise.
17: For $(i,u,b) \in O$, update $T_{i,u,b}$ and $\hat{\pi}_{i,u,b}$: $T_{i,u,b} \leftarrow T_{i,u,b} + 1$, $\hat{\pi}_{i,u,b} \leftarrow \hat{\pi}_{i,u,b} + (Y_{i,u,b} - \hat{\pi}_{i,u,b})/T_{i,u,b}$.
18: **end for**

Regret analysis. Let $\mathcal{A}$ denote the set containing all base arms for visiting probabilities, i.e. $\mathcal{A} = \{(i,u,b) \mid i \in [m], u \in V, b \in [c_i]\}$. With a bit of abuse of notation, we also use V to denote the set of base arms corresponding to the node weights. Therefore, $\mathcal{A} \cup V$ is the set of all base arms, which is different from the Section F.2 that only has $\mathcal{A}$ as base arms and does not have base arms for random weights. Following the notations of (Wang & Chen, 2017), we define $X_a \in [0,1]$ be the random outcome of base arm $a \in \mathcal{A} \cup V$ and define $\mathcal{D}$ as the unknown joint distribution of random outcomes $\mathbf{X} = (X_a)_{a\in\mathcal{A}\cup V}$ with unknown mean $\mu_a = E_{\mathbf{X}\sim\mathcal{D}}[X_a]$. Let k be any feasible budget allocation, we denote $p_a^{\mathcal{D},k}$ the probability that action k triggers arm a (i.e. outcome X_a is observed) when the unknown environment distribution is $\mathcal{D}$. Specifically, $p_a^{\mathcal{D},k} = 1$ for $a \in \{(i,u,b) \mid i \in [m], u \in V, b = k_i\}$, $p_a^{\mathcal{D},k} = 1 - \prod_{i\in[m]}(1 - \mathbb{I}\{a = (i,u,k_i)\})$ for $a \in V$ and $p_i^{\mathcal{D},k} = 0$ for the rest of base arms. Therefore, we can rewrite the

inequality in property 2 as

$$|r_{\theta,\sigma}(k) - r_{\theta',\sigma}(k)| \le \sum_{i\in[m],u\in V,b=k_i} (\alpha_{u}|\theta_{i,u,b} - \theta'_{i,u,b}|) + \sum_{u\in V}|\alpha_u - \alpha'_u|\Big(1 - \prod_{i\in[m]}(1-\theta_{i,u,b})\Big) \le \sum_{a\in A} p_a^{D,k}|\beta_a - \beta'_a|, \quad (32)$$

where we abuse the notation to denote β_a as β_a and β'_a as β'_a if $a \in A_2$ in the last inequality. We can observe that the above rewritten property 2 can be viewed as a special case of the 1-norm TPM bounded smoothness condition in (Wang & Chen, 2017).

Now we can apply the similar proof in Section B.3 in (Wang & Chen, 2017), together with how we deal with the $|\theta_{i,u,b} - \theta_{i,u,b,t}|$ in Section F.2.2 to get a distribution-dependent regret bound.

Let $K = m|V| + |V|$ be the maximum number of triggered arms, $|A[V]| = Bm|V| + |V|$ be the number of base arms and $\max\{x, 0\} = \max\{x, 0\}$. For any feasible budget allocation k and any base arm $a \in A[V]$, the gap $\Delta_k = \max\{0, r_{\theta,\sigma}(k) - r_{\theta,\sigma}(k)\}$ and we define $\Delta_{\min}^a = \inf_{k:p_a^{D,k}>0,\ \Delta_k>0} \Delta_k$, $\Delta_{\max}^a = \sup_{k:p_a^{D,k}>0,\ \Delta_k>0} \Delta_k$. As convention, if there is no action k such that $p_a^{D,k} > 0$ and $\Delta_k > 0$, we define $\Delta_{\min}^a = +\infty$; $\Delta_{\max}^a = 0$. We define $\Delta_{\min} = \min_{a\in A} \Delta_{\min}^a$ and $\Delta_{\max} = \max_{a\in A} \Delta_{\max}^a$.

Theorem 7. Algorithm 8 has the following distribution-dependent $(1-\varepsilon, 1)$ approximation regret,

$$\mathrm{Reg}_{\theta,\sigma}(T) \le \sum_{a\in A[V]} \frac{576K^2 \ln T}{\Delta_{\min}^a} + 4|A[V]| + \sum_{a\in A[V]} \max\Big\{\log_2 \frac{2K}{\Delta_{\min}^a}, 0\Big\} + 2\Big(\frac{\pi^2}{6} + 1\Big)\Delta_{\max}.$$

We can also have a distribution-independent bound,

$$\mathrm{Reg}_{\theta,\sigma}(T) \le 12\sqrt{|A[V]|KT\ln T} + 2|A[V]| + \max\Big\{\log_2 \frac{T}{18\ln T}, 0\Big\} + 2\Big(\frac{\pi^2}{6} + 1\Big)\Delta_{\max}.$$

Remark. Compared with the regret bound in Thm. 3 in the main text, we can see the base arms now became $|V|$ instead of $|A|$, which is worse than the regret bound of the fixed unknown weight setting. This is a trade-off since we need more estimation to handle the uncertain random node weight, which incurs additional regrets. It also shows our non-trivial analysis of the fixed unknown weights carefully merge the error term into base arms and incurs only constant-factor of extra regrets.

G. Online Learning for Non-overlapping MuLaNE

G.1. Online Learning Algorithm for Non-overlapping MuLaNE

For non-overlapping MuLaNE, we estimate layer-level marginal gains as our parameters, rather than probabilities $P_{i,u}(b)$ and node weights α_u. By doing so, we can largely simply our analysis since we no longer need to handle the unknown weights and the extra constraint of $P_{i,u}(b)$ as in the overlapping setting. Concretely, we maintain a set of base arms $A = \{(i,b) | i \in [m], b \in [c_i]\}$, and let $|A| = \sum_{i\in[m]} c_i$ be the total number of these arms. For each base arm $(i,b) \in A$, denote $\theta_{i,b}$ as the truth value of each base arm, i.e. $\theta_{i,b} = \sum_{u\in V} \alpha_u(P_{i,u}(b) - P_{i,u}(b-1))$. Based on the definition of base arms, we can see that we do not need to explicitly estimate the node weights or probabilities $P_{i,u}(b)$, so we will use $r(k)$ to denote the reward function $r_{G,\sigma}(k)$.

We present our algorithm in Alg. 9, which has three differences compared with the Alg. 3 for overlapping MuLaNE. First, the expected reward of action k is $r(k) = \sum_{i=1}^m \sum_{b=1}^{k_i} \theta_{i,b}$, which is a linear reward function and satisfies two new Monotonicity and 1-Norm Bounded Smoothness conditions (see Condition 1 and 2). Second, we use the Dynamic Programming method (Alg. 7) as our Oracle, which does not require the parameters $\theta_{i,b}$ to be increasing. Moreover, the offline oracle always outputs optimal solutions. Third, we use a new random variable $Y_{i,b}$, which indicates the gain of weights which depends on whether a new node is visited by random walker from the i-th layer exactly at the b-th step, i.e., $Y_{i,b} = \sum_{u\in S_{j=1}^b X_{i,j}} \alpha_u - \sum_{u\in S_{j=1}^{b-1} X_{i,j}} \alpha_u$ for base arms (i,b) such that $b \le k_i$.

Condition 1 (Monotonicity). The reward $r(k)$ is monotonically increasing, i.e., for any budget allocation k and any two vectors $\theta = (\theta_{i,b})_{(i,b)\in A}$, $\theta' = (\theta'_{i,b})_{(i,b)\in A}$, we have $r(k) \le r_{\theta'}(k)$, if $\theta_{i,b} \le \theta'_{i,b}$, for $(i,b) \in A$.

Condition 2 (1-Norm Bounded Smoothness). The reward function $r(k)$ satisfies the 1-norm bounded smoothness, i.e., for any budget allocation k and any two vectors $\theta = (\theta_{i,b})_{(i,b)\in A}$, $\theta' = (\theta'_{i,b})_{(i,b)\in A}$, we have $|r(k) - r_{\theta'}(k)| \le \sum_{b\le k_i} |\theta_{i,b} - \theta'_{i,b}|$.

We define the reward gap $\Delta_k = r(k^*) - r(k)$ for all feasible action k satisfying $\sum_{i=1}^m k_i = B, 0 \le k_i \le c_i$. For each base arm (i,b), we define $\Delta_{\min}^{i,b} = \min_{k>0; k_i \ge b} \Delta_k$ and $\Delta_{\max}^{i,b} = \max_{k>0; k_i \ge b} \Delta_k$. As a convention, if there is no action with $k_i \ge b$ such that $\Delta_k > 0$, we define $\Delta_{\min}^{i,b} = +\infty$ and $\Delta_{\max}^{i,b} = 0$. Also, we define $\Delta_{\min} = \min_{(i,b)\in A} \Delta_{\min}^{i,b}$ and $\Delta_{\max} = \max_{(i,b)\in A} \Delta_{\max}^{i,b}$.

Theorem 8. CUCB-MG has the following distribution-

Algorithm 9 CUCB-MG: Combinatorial Upper Con dence Bound (CUCB) algorithm for non-overlapping MuLaNE, using Marginal Gains as the base arms

Input: Budget B, number of layers m, of ine oracle DP.

For each $(i,b) \in A$, $T_{i,b} \leftarrow 0$, $\hat{\theta}_{i,b} \leftarrow 1$.
for $t = 1, 2, 3, \ldots, T$ do
  for $(i,b) \in A$ do
    $\rho_{i,b} \leftarrow \sqrt{3 \ln t / (2T_{i,b})}$.
    $\theta_{i,b} \leftarrow \min\{\hat{\theta}_{i,b} + \rho_{i,b}, 1\}$.
  end for
  $k \leftarrow \mathrm{DP}((\theta_{i,b})_{i,b}, B, c)$
  Apply budget allocation k, which gives trajectories $X := (X_{i,1}, \ldots, X_{i,k_i})_{i \in [m]}$ as feedbacks.
  For $(i,b) \in S := \{(i,b) \in A \mid b \le k_i\}$, $Y_{i,b} = \sum_{u \in S_{b,j=1} f X_{i,j} g} u \Big/ \sum_{u \in S_{b,j=1} f X_{i,j} g} u$.
  For $(i,b) \in S$, update $T_{i,b}$ and $\hat{\theta}_{i,b}$:
  $T_{i,b} \leftarrow T_{i,b} + 1$, $\hat{\theta}_{i,b} \leftarrow \hat{\theta}_{i,b} + (Y_{i,b} - \hat{\theta}_{i,b})/T_{i,b}$.
end for

dependent regret bound,

$$\mathrm{Reg}(T) \le \sum_{(i,b) \in A} \frac{48B \ln T}{\Delta_{i,b,\min}} + 2|A| + \frac{2}{3}|A| \Delta_{\max}.$$

G.2. Regret Analysis for Non-overlapping MuLaNE

In this section, we give the regret analysis for CUCB-MG, which applies the standard CUCB algorithm (Wang & Chen, 2017) to this setting.

G.2.1. Proof Details

Let A denote the set containing all base arms, i.e., $A = \{(i,b) \mid i \in [m], b \in [c_i]\}$. We add subscript t to denote the value of a variable at the end of t. For example, $T_{i,b,t}$ denotes the total times of arm $(i,b) \in A$ is played at the end of round t. Let us rst introduce a de nition of an unlikely event that $\hat{\theta}_{i,b,t-1}$ is not accurate as expected.

De nition 3. We say that the sampling is nice at the beginning of round t, if for every arm $(i,b) \in A$, $|\hat{\theta}_{i,b,t-1} - \theta_{i,b}| < \rho_{i,b,t}$, where $\rho_{i,b,t} = \sqrt{3 \ln t / (2T_{i,b,t-1})}$. Let N_t be such event.

Lemma 13. For each round $t \ge 1$, $\Pr[\neg N_t] \le 2|A| t^{-2}$.

Proof. For each round $t \ge 1$, we have

$$\begin{aligned}
&\Pr[\neg N_t] \\
&= \Pr\Big[\exists (i,b) \in A, |\hat{\theta}_{i,b,t-1} - \theta_{i,b}| \ge \sqrt{\frac{3 \ln t}{2T_{i,b,t-1}}}\Big] \\
&\le \sum_{(i,b) \in A} \Pr\Big[|\hat{\theta}_{i,b,t-1} - \theta_{i,b}| \ge \sqrt{\frac{3 \ln t}{2T_{i,b,t-1}}}\Big] \\
&= \sum_{(i,b) \in A} \sum_{k=1}^{t-1} \Pr[T_{i,b,t-1} = k] \cdot \Big|\hat{\theta}_{i,b,k} - \theta_{i,b}\Big| \ge \sqrt{\frac{3 \ln t}{2k}}\Big] \\
&\le \sum_{(i,b) \in A} \sum_{k=1}^{t-1} \frac{2}{t^3} \le 2|A| t^{-2}.
\end{aligned}$$

When $T_{i,b,t-1} = k$, $\hat{\theta}_{i,b,t-1}$ is the average of k i.i.d. random variables $Y_{i,b}^{[1]}, \ldots, Y_{i,b}^{[k]}$, where $Y_{i,b}^{[j]}$ is the Bernoulli random variable of arm (i,b) when it is played for the j-th time during the execution. That is, $\hat{\theta}_{i,b,t-1} = \sum_{j=1}^{k} Y_{i,b}^{[j]}/k$. Note that the independence of $Y_{i,b}^{[1]}, \ldots, Y_{i,b}^{[k]}$ comes from the fact we select a new starting position following the starting distribution $\mathcal{D}_i$ at the beginning of each round t. The last inequality uses the Hoeffding Inequaility (Fact 1).

Then we can use monotonicity and 1-norm bounded smoothness propterty (Condition 1 and Condition 2) to bound the reward gap $\Delta_{k_t} = r(k^*) - r(k_t)$ between the optimal action k^* and the action $k_t = (k_{t,1}, \ldots, k_{t,m})$ selected by our algorithm A. To achieve this, we introduce a positive real number $M_{i,b}$ for each arm (i,b) and de ne $M_{k_t} = \max_{(i,b) \in S_t} M_{i,b}$, where $S_t = \{(i,b) \in A \mid b \le k_{t,i}\}$ and $|S_t| = B$. De ne,

$$\tau_T(M, s) = \begin{cases} \infty, & \text{if } s = 0; \\ 2\sqrt{\frac{6 \ln T}{s}}, & \text{if } 1 \le s \le \ell_T(M); \\ 0, & \text{if } s \ge \ell_T(M) + 1; \end{cases}$$

where

$$\ell_T(M) = \Big\lfloor \frac{24B^2 \ln T}{M^2} \Big\rfloor.$$

Lemma 14. If $\neg\{\Delta_{k_t} \ge M_{k_t}\} \cap N_t$ holds, we have

$$\Delta_{k_t} \le \sum_{(i,b) \in S_t} \tau_T(M_{i,b}, T_{i,b,t-1}). \tag{33}$$

Proof. First, we can observe $\theta_{i,b,t} \le \theta_{i,b}$, for $(i,b) \in A$, when N_t holds. This comes from the fact $\theta_{i,b,t} = \min\{\hat{\theta}_{i,b,t-1} + \rho_{i,b,t}, 1\}$ and $|\hat{\theta}_{i,b,t-1} - \theta_{i,b}| < \rho_{i,b,t}$.

Notice that the right hand of Inequality (33) is non-negative, it is trivially satisfied when $\Delta_{k_t} = 0$. We only need to consider $\Delta_{k_t} > 0$. By $\mathcal{N}_t$ and Condition 1, We have the following inequalities,

$$r_{t-1}(\boldsymbol{k}_t) \geq r_{t-1}(\boldsymbol{k}^*) \geq r(\boldsymbol{k}^*) = \Delta_{k_t} + r(\boldsymbol{k}_t).$$

The first inequality is due to the *oracle* outputs the optimal solution given parameters $\bar{\boldsymbol{\mu}}_t$,the second inequality is due to Condition 1 (monotonicity).

Then, by Condition 2, we have

$$\Delta_{k_t} \leq r_{t-1}(\boldsymbol{k}_t) - r(\boldsymbol{k}_t) \leq \sum_{(i,b)\in S_t} |\bar{\mu}_{i,b,t} - \mu_{i,b}|.$$

Next, we can bound Δ_{k_t} by bounding $\bar{\mu}_{i,b,t} - \mu_{i,b}$ by applying a transformation. As $\{\Delta_{k_t} \geq M_{k_t}\}$ holds by assumption, so $\sum_{(i,b)\in S_t} |\bar{\mu}_{i,b,t} - \mu_{i,b}| \geq \Delta_{k_t} \geq M_{k_t}$. We have

$$\begin{aligned}
\Delta_{k_t} &\leq \sum_{(i,b)\in S_t} |\bar{\mu}_{i,b,t} - \mu_{i,b}| \\
&\leq -M_{k_t} + 2\sum_{(i,b)\in S_t} |\bar{\mu}_{i,b,t} - \mu_{i,b}| \\
&= 2\sum_{(i,b)\in S_t} (|\bar{\mu}_{i,b,t} - \mu_{i,b}| - \frac{M_{k_t}}{2B}) \\
&\leq 2\sum_{(i,b)\in S_t} (|\bar{\mu}_{i,b,t} - \mu_{i,b}| - \frac{M_{i,b}}{2B}). \qquad (34)
\end{aligned}$$

By $\mathcal{N}_t$, we have:

$$\begin{aligned}
|\bar{\mu}_{i,b,t} - \mu_{i,b}| - \frac{M_{i,b}}{2B} &\leq \min\{2\rho_{i,b,t}, 1\} - \frac{M_{i,b}}{2B} \\
&\leq \min\left\{2\sqrt{\frac{3\ln T}{2T_{i,b,t-1}}}, 1\right\} - \frac{M_{i,b}}{2B}.
\end{aligned}$$

Now consider the following cases:

Case 1. If $T_{i,b,t-1} \leq \ell_T(M_{i,b})$, we have $\bar{\mu}_{i,b,t} - \mu_{i,b} - \frac{M_{i,b}}{2B} \leq \min\left\{2\sqrt{\frac{3\ln T}{2T_{i,b,t-1}}}, 1\right\} \leq \frac{1}{2}\kappa_T(M_{i,b}, T_{i,b,t-1})$.

Case 2. If $T_{i,b,t-1} \geq \ell_T(M_{i,b,t}) + 1$, then $2\sqrt{\frac{3\ln T}{2T_{i,b;t-1}}} \leq \frac{M_{i,b}}{2B}$, so $\bar{\mu}_{i,b,t} - \mu_{i,b} - \frac{M_{i,b}}{2B} \leq 0 = \frac{1}{2}\kappa_T(M_{i,b}, T_{i,b,t-1})$.

Combine these two cases and inequality (34), we have the following inequality, which concludes the lemma.

$$\Delta_{k_t} \leq \sum_{(i,b)\in S_t} \kappa_T(M_{i,b}, T_{i,b,t-1}).$$

Theorem 8. *CUCB-MG has the following distribution-dependent regret bound,*

$$Reg_\mu(T) \leq \sum_{(i,b)\in A} \frac{48B\ln T}{\Delta^{i,b}_{\min}} + 2|\mathcal{A}| + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{\max}.$$

Proof of Theorem 8. By above lemmas, we have

$$\begin{aligned}
&\sum_{t=1}^{\infty} \mathbb{I}(\{\Delta_{k_t} \geq M_{k_t}\} \wedge \mathcal{N}_t)\Delta_{k_t} \\
&\leq \sum_{t=1}^{\infty}\sum_{(i,b):(i,b)\in S_t} \kappa_T(M_{i,b}, T_{i,b,t-1}) \\
&= \sum_{(i,b)\in A}\sum_{t'\in[T],k_{t';i}=b} \kappa_T(M_{i,b}, T_{i,b,t'-1}) \\
&\leq \sum_{(i,b)\in A}\sum_{s=0}^{T_{i,b;T}} \kappa_T(M_{i,b}, s) \\
&\leq 2|\mathcal{A}| + \sum_{(i,b)\in A}\sum_{s=1}^{\ell_T(M_{i,b})} 2\sqrt{6\frac{\ln T}{s}} \\
&\leq 2|\mathcal{A}| + \sum_{(i,b)\in A}\int_{s=0}^{\ell_T(M_{i,b})} 2\sqrt{6\frac{\ln T}{s}}ds \\
&\leq 2|\mathcal{A}| + \sum_{(i,b)\in A} 4\sqrt{6\ln T\ell_T(M_{i,b})} \\
&\leq 2|\mathcal{A}| + \sum_{(i,b)\in A} \frac{48B\ln T}{M_{i,b}}.
\end{aligned}$$

Let us set $M_{i,b} = \Delta^{i,b}_{\min}$ and thus the event $\{\Delta_{k_t} \geq M_{k_t}\}$ always holds. We have

$$\begin{aligned}
&\sum_{t=1}^{\infty} \mathbb{I}(\mathcal{N}_t \wedge \{\Delta_{k_t} \geq M_{k_t}\})\Delta_{k_t} \\
&\leq 2|\mathcal{A}| + \sum_{(i,b)\in A} \frac{48B\ln T}{\Delta^{i,b}_{\min}}.
\end{aligned}$$

By Lem. 13, $Pr\{\neg N_t\} \leq 2|\mathcal{A}|t^{-2}$, then we have

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \mathbb{I}(\neg\mathcal{N}_t)\Delta_{k_t}\right] \leq \sum_{t=1}^{\infty} 2|\mathcal{A}|t^{-2}\Delta_{\max} \leq \frac{\pi^2}{3}|\mathcal{A}|\Delta_{\max}.$$

By our choice of $M_{i,b}$,

$$\sum_{t=1}^{\infty} \mathbb{I}(\Delta_{k_t} < M_{k_t})\Delta_{k_t} = 0.$$

By the definition of the regret,

$$
\begin{aligned}
Reg\ (T) &= \mathsf{E}\left[\sum_{t=1}^{T} \Delta_{k_t}\right] \\
&\le \mathsf{E}\left[\sum_{t=1}^{T} \mathsf{I}\ \neg\mathcal{N}_t\ + \mathsf{I}\ \Delta_{k_t} < M_{k_t}\right] \\
&+ \mathsf{I}\ \mathcal{N}_t\ \wedge \{\Delta_{k_t} \ge M_{k_t}\}\quad \Delta_{k_t} \\
&\le \sum_{(i,b)\,2\,A} \frac{48B\ln T}{\Delta^{i,b}_{min}} + 2|\mathcal{A}| + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{max}.
\end{aligned}
$$

For distribution independent bound, we take $M_{i,b} = M = \sqrt{48B|\mathcal{A}|\ln T/T}$, then we have $\sum_{t=1}^{T} \mathsf{I}\{\Delta_{k_t} \le M_{k_t}\} \le TM$.

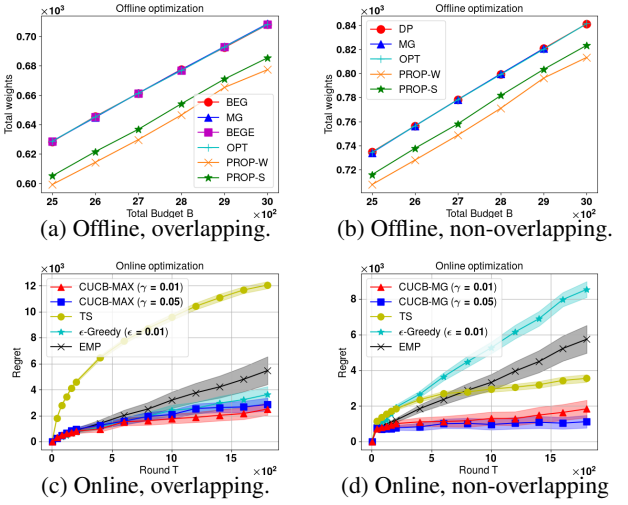
Then

$$
\begin{aligned}
Reg_\mu(T) &\le \sum_{(i,b)\,2\,A} \frac{48B\ln T}{M_{i,b}} + 2|\mathcal{A}| + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{max} + TM \\
&\le \frac{48B|\mathcal{A}|\ln T}{M} + 2|\mathcal{A}| + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{max} + TM \\
&= 2\sqrt{48B|\mathcal{A}|T\ln T} + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{max} + 2|\mathcal{A}| \\
&\le 14\sqrt{B|\mathcal{A}|T\ln T} + \frac{\pi^2}{3}|\mathcal{A}|\Delta_{max} + 2|\mathcal{A}|.
\end{aligned}
$$

H. Supplemental Experiments

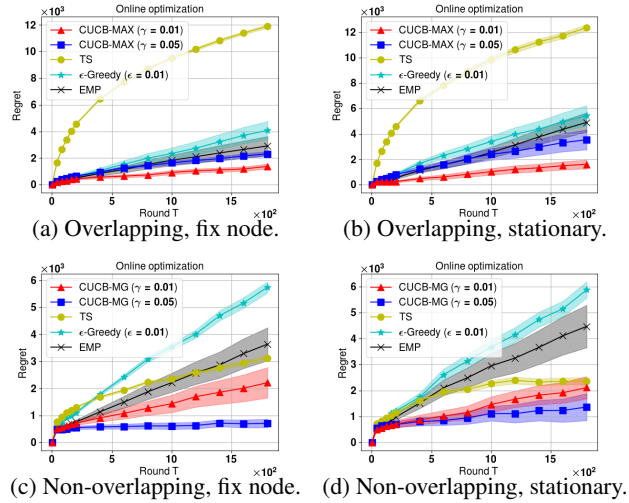
H.1. Experimental Results for Stationary Distributions

In this section, we show experimental results with explorers starting from stationary distributions in Fig. 3. Specifically, each explorer W_i starts from the node u in layer L_i with probability proportional to the out-degree of u. For offline overlapping case, the total weights of unique nodes visited for BEG and MG are the same, and outperforms two baselines, which accords with our theoretical analysis. Similar to the Sec. 6, BEG and MG are empirically close to the optimal solution. For offline non-overlapping case, DP and MG give us the same *optimal* solution. For the online learning algorithms, note that we still use BEG rather than MG as the offline oracle, this is due to the UCB values do not have the diminishing return properties. The same applies for the non-overlapping case, where DP (instead of MG) is used as the offline oracle. The trend of the curves and the analysis are similar to that in the Sec. 6.



(a) Offline, overlapping. (b) Offline, non-overlapping.
(c) Online, overlapping. (d) Online, non-overlapping

Figure 3. Explorers start with stationary distributions. Above: total weights of unique nodes visited given by different offline algorithms. Below: regret for different online algorithms when $B = 3000$.



(a) Overlapping, fix node. (b) Overlapping, stationary.
(c) Non-overlapping, fix node. (d) Non-overlapping, stationary.

Figure 4. Regret for different online algorithms, when total budget $B = 2000$.

H.2. Experimental Results for Online Learning Algorithms with Different Budgets B

As shown in Fig. 4 and Fig. 5, although there are some fluctuations, the trend of the curves and the analysis are almost the same as that when $B = 3000$, which shows our experimental results are consistent for different budgets.

(a) Overlapping, fix node. (b) Overlapping, stationary.

(c) Non-overlapping, fix node. (d) Non-overlapping, stationary.

Figure 5. Regret for different online algorithms, when total budget $B = 4000$.