Influence Maximization in Unknown Social Networks: A Contextual Bandit Approach

Vincenzo Auletta
University of Salerno
Fisciano, Italy
auletta@unisa.it

Diodato Ferraioli
University of Salerno
Fisciano, Italy
dferraioli@unisa.it

Grazia Ferrara
University of Salerno
Fisciano, Italy
grferrara@unisa.it

ABSTRACT

The Influence Maximization (IM) problem is a fundamental problem on social networks where you are required to choose a set of few seeds from which starting an information campaign aiming to reach as many nodes as possible in the network. In this work, we consider the IM problem in a setting where neither network nodes nor their relationships are known, except for very few samples. Thus, you have to orchestrate the campaign while learning information about the network. This problem has been recently showed to have applications in public health: e.g., to maximize the diffusion of HIV prevention information among marginalized people, such as homeless. In this work we propose a two-level bandit approach to address the IM problem with partially observed networks: the lower layer implements a contextual bandit that selects nodes to query based on the current observed subgraph, available nodes, and edge discovery rewards; the upper meta-layer dynamically chooses between two exploration strategies: a global approach maximizing immediate edge discovery, and a component-focused strategy targeting the least-explored connected component. This dual approach prevents local over-exploitation while maintaining efficient global exploration. The proposed method outperforms the state-of-the-art method across diverse network topologies, particularly under tight budget constraints.

KEYWORDS

Influence Maximization, Contextual Bandit, Health Prevention Campaigns

1 INTRODUCTION

Influence Maximization (IM) [13] is defined as the problem of finding within a social or contact network a set of *seeds* for initiating a diffusion campaign able to reach as many network members as possible. The problem has been highlighted to be relevant to many application contexts, such as marketing [9, 13, 14], opinion formation [3–5, 10], voting [6, 7, 26]. Recently, this problem has been observed to be relevant even to public health. Prevention is regarded as a foundational practice for ensuring individual health and, more broadly, for safeguarding public health through the mitigation of infectious diseases. While this imperative was significantly underscored by the recent COVID-19 pandemic, its relevance is consistently reflected across the breadth of medical literature concerning communicable pathologies. For instance, the 2024 report by the Joint United Nations Programme on HIV/AIDS (UNAIDS) (https://crossroads.unaids.org/) asserts that "much more effort and urgency is required to accelerate prevention", particularly regarding "marginalized people", in order to fulfill the objective of ending "AIDS as a public health threat by 2030".

This underscores the challenge of identifying key intermediaries capable of catalyzing broader engagement with prevention initiatives, particularly among marginalized populations that are frequently underserved by conventional communication channels (e.g., homeless people, sex workers, incarcerated individuals, ...). Given these constraints, this objective can be formally conceptualized as an influence maximization problem.

Leveraging this framework facilitates the integration of established influence maximization literature. While seminal IM research typically assumes a known topology to develop scalable heuristics and approximation algorithms with provable guarantees [18], recent efforts address uncertainty regarding edge weights. Specifically, when network structure is known but relationship strengths are latent, online learning frameworks have emerged as a primary solution, yielding several efficient algorithms [1, 8].

Wilder et al. [25] identified a key limitation of standard IM for public health: its reliance on a known network topology causes it to fail on real-world networks, missing marginalized nodes. Their CHANGE algorithm addresses this by providing a network exploration protocol via node queries instead of assuming a known structure. However, CHANGE still requires knowing all nodes *a priori*, which is unlikely for marginalized groups, and its performance often approximates a random querying baseline on some network topologies [16].

Algorithms for IM under partial observability have evolved from computationally heavy precursors [27, 28] to heuristic approaches like ARISEN [24], that guided discovery by leveraging community structure through random walks, and the deployable CHANGE [25] algorithm, which provided a robust heuristic based on the Friendship Paradox. Later, a shift to data-driven methods emerged: NeuGreedy and NeuMax [29] learned influence functions from diffusion cascades; reinforcement learning methods like DQN [12] and CLAIM [16] learned transferable exploration policies; and the recent IM-META [22] exploited node metadata and a Siamese network to infer missing links and improve exploration quality.

Existing approaches predominantly rely on various forms of side-information (e.g., network assumptions, metadata, training data). By contrast, scenarios operating with the same limited information as CHANGE, or even more restrictive conditions, remain largely unexplored (cf., [22]).

Our Contribution. We introduce CANCEL, an end-to-end framework for Influence Maximization (IM) under extreme network unobservability. Unlike existing methods, CANCEL requires no auxiliary metadata or global topological priors, operating only via a localized query protocol that iteratively expands a minimal initial subgraph.

Drawing on the established success of bandit-based strategies for structural discovery [19, 23], we model the network exploration process through the *Contextual Multi-Armed Bandit (ConMAB)* framework [20] where nodes are arms, context is derived from the revealed topology, and rewards are based on node discovery.

Crucially, we identify a specific *locality trap* in single-layer ConMABs: because the agent lacks a global map, it tends to over-exploit dense, early-discovered clusters where rewards are certain, failing to traverse sparse "bridges" to distant, influential components.

To overcome this, CANCEL employs a hierarchical, two-level bandit architecture. A meta-layer manages the exploration-exploitation trade-off by alternating between a global discovery strategy and a component-focused strategy targeting under-explored regions. A lower-level ConMAB then selects nodes within the chosen scope.

Empirical evaluations on real-world networks demonstrate that CANCEL consistently outperforms state-of-the-art baselines. Our findings highlight several key insights:

- CANCEL achieves higher expected influence spread, particularly under tight budget constraints, by effectively prioritizing high-centrality nodes even when total node discovery is lower than baselines.
- The hierarchical structure is essential for performance; on several topologies, flat ConMAB architectures fail to surpass simple baselines due to topological local optima.

To the best of our knowledge, this is the first work to address influence maximization under extreme network unobservability by employing a hierarchical bandit framework that operates independently of auxiliary metadata, pre-trained models, or global topological assumptions.

Note on earlier venue: A preliminary version of this work was published in PRIMA 2025 [2]. This extended version adds: (i) ablation study isolating the hierarchical structure; (ii) sensitivity analysis across four algorithmic instantiations; (iii) robustness evaluation with expanded initial information; (iv) analysis decoupling discovery volume from influence spread; and (v) complete results across all eight datasets. The original publication: https://link.springer.com/chapter/10.1007/978-3-032-13562-9_16.

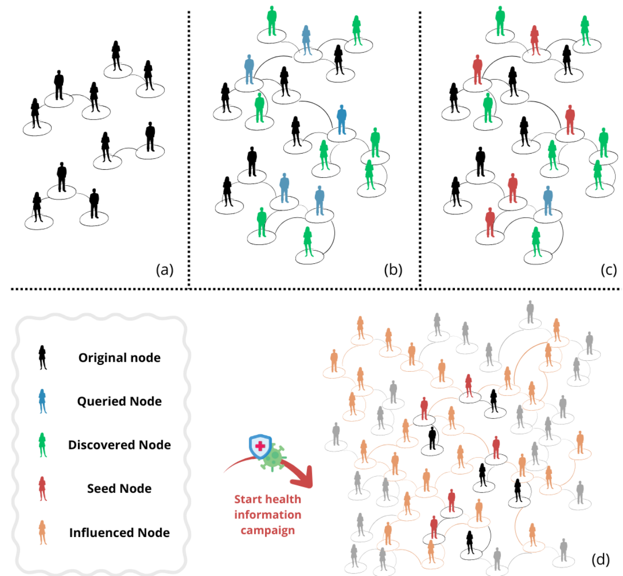

Figure 1: Visual representation of the framework. (a) The initial graph G_0 consisting of known seeds (black) and their immediate neighbors. (b) The *Network Exploration Phase*: the agent queries nodes (blue) to discover new parts of the network (green). (c) *Seed Selection*: after the budget is exhausted, a seed set S (red) is computed on the discovered proxy graph. (d) *Influence Diffusion*: the campaign propagates from the seeds, activating influenced nodes (orange) according to the IC model.

2 RELATED WORKS

Influence Maximization. Influence Maximization (IM) was formally introduced by Kempe et al. [13], who established the submodularity of the problem under different information diffusion models. This foundation paved the way for a vast literature focused on improving the scalability of the greedy algorithm and developing theoretical approximation guarantees [18]. However, the majority of these classical methods assume that the social network's global topology is fully visible to the optimizer.

IM under Partial Observability. Recent research has focused on the more realistic "unknown network" setting. Early approaches assumed a known structure with hidden edge weights, learning influence probabilities through repeated campaigns [8, 11]. More recently, the focus has shifted toward *structural* unobservability. CHANGE [25] pioneered this space by using a querying protocol based on the Friendship Paradox. Subsequent works, such as DQN [12] and CLAIM [16] and IM-META [22], have employed data-driven techniques, including reinforcement learning and methods based on node metadata exploitation to guide network exploration.

MABs and Contextual MABs. *Multi-Armed Bandits* (MABs) provide a simple method to face the problem of decision-making under uncertainty and they have been proved to be successful in many different contexts, including news recommendation, dynamic pricing and ad placement [15]. In MABs, a learner has to learn which arm is the most effective in an environment within a limited number of tries. Efficient and effective algorithms for multi-armed bandits in a variety of settings have been proposed (we refer to [15, 20]).

When the decision done by a bandit algorithm during the sequential decision process changes the context, i.e., changes how arms fit within the environment, then we need to adopt another MAB model which is known as *contextual bandit*. Effective algorithms have been provided in the case in which the change in the fit of an arm within the environment is bounded by the amount that the context has changed. Examples are kNN-UCB [19], assuming that the change in the arms' fitness is bounded by a Lipschitz constant, and LinUCB [17], that instead assumes that arms' fitness linearly depends on context.

3 DEFINITIONS

The problem. Let $G^* = (V^*, E^*)$ denote the ground-truth *oracle graph*, where V^* and E^* represent the set of nodes and edges, respectively. We operate under a partial observability regime where G^* is initially unknown, except for a small seed set $\Theta \subseteq V^*$. The exploration begins with an initial observable graph $G_0 = (V_0, E_0)$, where $V_0 = \Theta \cup \mathcal{N}_{G^*}(\Theta)$ comprises the initial nodes and their neighbors, and $E_0 = \{(u, v) \in E^* \mid u \in \Theta \vee v \in \Theta\}$ represents the set of edges incident to Θ (see Figure 1a).

To explore G^*, we are allocated a budget of B queries. The process proceeds in discrete steps $t = 0, \ldots, B-1$. Let Q_t denote the set of nodes fully queried prior to step t, initialized as $Q_0 = \Theta$. At each step t, we select a node v_t from the current *frontier* of observable but unqueried nodes, $V_t \setminus Q_t$. Querying v_t reveals its neighborhood $\mathcal{N}_{G^*}(v_t)$ and incident edges, updating the observed topology to $G_{t+1} = (V_{t+1}, E_{t+1})$, where $V_{t+1} = V_t \cup \mathcal{N}_{G^*}(v_t)$ and $E_{t+1} = E_t \cup \{(v_t, u) \mid u \in \mathcal{N}_{G^*}(v_t)\}$. This process yields a final proxy graph G_B, as illustrated in the progression to Figure 1b.

Upon expiring the budget, we employ G_B as a structural proxy to identify a seed set S_B of size k for the influence campaign (Figure 1c). We use the IMM algorithm [21] for this task, denoted as $S_B = \mathsf{IMM}(k, G_B)$. IMM is chosen for its ability to provide $(1 - 1/e - \epsilon)$-approximation guarantees while circumventing the prohibitive cost of Monte Carlo simulations required by standard greedy approaches [13] through the use of Reverse-Reachable (RR) set samples.

The information diffusion is modeled via the *Independent Cascade (IC)* model [13] (Figure 1d). The process initializes with an active set $A_0 = S_B$. At each discrete time step τ, every newly activated node $u \in A_{\tau-1}$ attempts to activate its currently inactive neighbor $v \in \mathcal{N}_{G^*}(u)$ with probability p_{uv}. The process terminates when no new activations occur. We denote the expected number of influenced nodes at the end of this process on the oracle graph G^* as $\hat{\sigma}_{G^*}(S_B)$.

Our objective is to identify an exploration policy that builds a proxy graph G_B such that the derived seed set S_B maximizes the expected spread on the unknown oracle graph. Formally, we seek to maximize $\hat{\sigma}_{G^*}(S_B)$ given the initial information G_0 and the query budget B.

The Bandit Framework. We adopt the Multi-Armed Bandit (MAB) framework to model the sequential decision-making process inherent in network exploration. While the fundamental *stochastic bandit* model, often solved via algorithms like ε-Greedy or UCB [15, 20], optimizes cumulative rewards under Independent and Identically Distributed (IID) assumptions, our problem necessitates the *Contextual Multi-Armed Bandit* (ConMAB) model, as querying nodes fundamentally alters the system state. In this setting, the learner observes a context $\mathcal{X}_t$ before selecting an arm a_t based on a policy π. In compliance with previous work [19], we formally cast the network exploration phase into this paradigm by defining the **Context** at step t as the current observable topology $G_t = (V_t, E_t) \in \mathcal{G}$; the set of available **Arms** A_t as the *frontier* of discovered but unqueried nodes (i.e., $A_t = V_t \setminus Q_t$); and the **Reward** r_t as the information gain, specifically the count of newly discovered edges $r_t = |E_{t+1}| - |E_t|$. The algorithmic interaction proceeds iteratively: at each step t, the learner observes G_t and A_t, then samples a node a_t from a policy-defined distribution $\mathcal{P}_t$. Upon querying, the environment reveals the neighborhood $\mathcal{N}_{G^*}(a_t)$, triggering the state update $G_{t+1} = (V_t \cup \mathcal{N}_{G^*}(a_t),\ E_t \cup \{(a_t, u) \mid u \in \mathcal{N}_{G^*}(a_t)\})$ and $Q_{t+1} = Q_t \cup \{a_t\}$. This cycle repeats until the budget B is exhausted, with the overarching objective of defining a distribution $\mathcal{P}_t$ that constructs a proxy graph G_B capable of maximizing the expected influence spread $\hat{\sigma}_{G^*}(S_B)$.

4 PROPOSED SOLUTION

The design of a network exploration algorithm in this setting faces severe challenges, primarily stemming from the scarcity of initial data. The prevention center initializes the campaign knowing only a small set of individuals Θ and their immediate social circles. Consequently, the initial observable graph G_0 is typically highly fragmented, consisting of multiple disconnected components, essentially, isolated "islands" of known nodes centered around the initial seeds. A secondary challenge arises from the disconnect between the exploration process and the ultimate objective. While our goal is to identify optimal seeds for influence maximization, we lack influence-based feedback during the exploration phase. We are forced to rely on structural expansion, discovering as many nodes and edges as possible, as a proxy for identifying influential candidates in the underlying oracle graph G^*.

However, using a standard contextual Multi-Armed Bandit (ConMAB) for this task carries a significant risk of topological entrapment, driven by the interaction between the initial fragmentation and the underlying community structure of G^*. Consider a ground-truth network G^* composed of dense communities connected by sparse bridges. If the initial seeds are scattered such that one belongs to a dense community and another to a sparse region (or a bridge), a standard bandit algorithm will quickly bias its policy. The seed within the dense community will yield high immediate rewards (numerous newly discovered neighbors), incentivizing the policy to exclusively exploit that component. Consequently, the algorithm risks becoming an "expert" on a single community while neglecting other observable components that initially appear less promising. These neglected components, however, might act as crucial bridges to vast, undiscovered parts of the graph. Given the limited query budget B, the agent may never traverse these structural bottlenecks, resulting in a biased proxy graph G_B that fails to represent the global network topology. To mitigate this, an effective policy must balance the exploitation of high-yield regions with the exploration of structurally distinct components, even when the immediate rewards for doing so are lower.

To address this, we introduce the CANCEL (**C**omponents-**A**ided **N**etwork **C**ontextual **E**xploration **L**earning) framework. CANCEL adopts a hierarchical approach by stacking two distinct bandit layers. The upper level, termed *StrategyBandit*, acts as a meta-learner that dynamically selects the exploration mode. At each step, it chooses between two strategies: a *Global* strategy, which behaves as a standard exploration agent considering all available nodes in the frontier ($V_t \setminus Q_t$), and a *Component-Focused* strategy. The latter restricts the search space specifically to the connected component within the current graph G_t that possesses the fewest queried nodes. By selecting the second option, the system forces the lower-level agent, termed *NodeBandit*, to select nodes from under-explored observable components, thereby preventing the policy from fixating on a single high-density area of the ground truth graph.

Algorithm 1 CANCEL

Input: $G_0 = (V_0, E_0)$, $Q_0 = \Theta$, budget B, oracle G^*, seed set size k
Initialize *NodeBandit*, *StrategyBandit*
for $t = 0, \ldots, B-1$ **do**
  $\zeta \leftarrow \textsc{SelectArm}(\textit{StrategyBandit})$
  if $\zeta \equiv \textsc{Global}$ **then**
    $A_t \leftarrow V_t \setminus Q_t$
  else
    $C_t \leftarrow \textsc{FindLeastQueriedComponent}(G_t, Q_t)$
    $A_t \leftarrow V(C_t) \setminus Q_t$
  $a_t \leftarrow \textsc{SelectArm}(\textit{NodeBandit}, A_t, G_t)$
  $N_{G^*}(a_t) \leftarrow \textsc{GetNeighbors}(G^*, a_t)$
  $V_{t+1} \leftarrow V_t \cup N_{G^*}(a_t)$
  $E_{t+1} \leftarrow E_t \cup \{(a_t, u) \mid u \in N_{G^*}(a_t)\}$
  $G_{t+1} \leftarrow (V_{t+1}, E_{t+1})$
  $Q_{t+1} \leftarrow Q_t \cup \{a_t\}$
  $r_t \leftarrow |E_{t+1}| - |E_t|$
  $\textsc{Update}(\textit{NodeBandit}, a_t, G_t, r_t)$
  $\textsc{Update}(\textit{StrategyBandit}, \zeta, r_t)$
return $\mathsf{IMM}(k, G_B)$

The complete procedure is detailed in Algorithm 1. The process initializes with the starting graph G_0, the set of initially known nodes $Q_0 = \Theta$, and the query budget B. Both bandit instances are initialized, with the *NodeBandit* managing node-level features and the *StrategyBandit* managing the two high-level strategies. In each round t, the *StrategyBandit* determines the current strategy ζ. If "Global" is selected, the set of available arms A_t includes the entire unqueried frontier; if "Component-Focused" is selected, A_t is restricted to the least-queried component C_t. The *NodeBandit* then selects a specific node a_t from A_t, the oracle reveals its neighborhood $N_{G^*}(a_t)$, and the reward r_t (the number of newly discovered edges) is computed. Crucially, this reward is used to update *both* the *NodeBandit* (to learn which nodes are valuable) and the *StrategyBandit* (to learn which strategy yields better discovery).

It is important to note that the CANCEL framework is model-agnostic regarding the specific implementations of the learners. We leave the initialization and update rules of the *NodeBandit* and *StrategyBandit* unspecified, allowing the use of whichever MAB (e.g., UCB1, ϵ-greedy [15, 20]) or Contextual MAB algorithms (e.g., LinUCB [17], kNN-UCB [19]) best fit the specific application context. We have tested the framework with various algorithmic choices; while the hierarchical structure generally outperforms previous non-hierarchical approaches, the optimal combination of internal algorithms may vary depending on the specific topology of the oracle network.

5 EVALUATION

To assess the efficacy of the CANCEL algorithm, we conducted a systematic empirical evaluation. This section first provides an overview of the selected baselines and datasets, with their salient characteristics. Subsequently, we provide a comprehensive description of the experimental configuration. Finally, we present and analyze the results derived from our comparative study. All experiments were executed on a workstation equipped with an Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz and 64 GB of RAM, running a Windows operating system. The proposed framework was implemented using Python 3.13.

Baselines. We compare CANCEL to the state-of-the-art method for the setting of interest (i.e., with no side information available) and to some other natural benchmarks (also used in previous works [22]): *(i)* Opt, the optimal baseline which select seeds by knowing the entire network, so providing an upper bound on performances; *(ii)* CHANGE [25], the SOTA method that queries a random node and then one of its random friends; *(iii)* RANDOM, that queries nodes uniformly at random; and *(iv)* DFS, a benchmark that explores the network in a depth-first fashion.

Datasets. Similarly to previous work [12, 16], we decided to select datasets from different domains:

- *Retweet Networks.* In these networks, the nodes represent Twitter users and edges represent retweets. These were collected from various social and political hashtags.
- *Animal Interaction Networks.* These networks are real-world animal interaction network datasets. In particular, in this case, node represent voles, while there's a edge whenever two voles are caught in at least one common trap.
- *Human Interaction Networks.* In these networks, nodes represent humans and edges between them represent proximity (i.e., contacts in the physical world).

Notice that the first two categories are the same considered in [16], while the other one is different because the private dataset they used, representing a physical network between female sex workers, was not publicly available. Still, with the aim of preserving the idea of a dataset representing contact networks between human beings, we decided to take as reference some human interaction networks. The details about the considered datasets are summarized in Table 1 and all the datasets are publicly available on Network Repository.

5.1 Experimental Configuration

Following established protocols in the literature [12, 16], the initial graph G_0 was generated by randomly sampling five nodes and their respective neighborhoods. To ensure statistical robustness, we evaluated the mean performance across 20 distinct initializations of G_0. Furthermore, to account for the stochastic nature of some exploration protocols, we performed 20 independent executions of each of them for every initialization. In accordance with prior studies [12, 16], information diffusion was simulated using the Independent Cascade (IC) model [13], with the activation probability p for each edge set to 0.1. The resulting influence spread was averaged over 200 independent Monte Carlo simulations.

As each arm a must be contextualized within the set $\mathcal{X}_t$ at each query step t, every node a is represented by a feature vector $x_{a,t} \in \mathbb{R}^d$. The dynamic nature of the environment requires these representations to be recomputed following each network update, as new neighborhoods are integrated into G_{t-1}. Adopting the feature engineering approach proposed for network exploration in [19], we utilize four handcrafted topological features:

Table 1: Structural properties of each selected reference dataset G^*. $|V^*|$ and $|E^*|$ respectively represent the number of nodes and the number of edges of the network. $\bar{d}$ represents the average degree, γ the clustering coefficient, $|C|$ the number of connected components, and $|C_{max}|$ the size of the largest connected component.

G^*	Retweet Networks				Animal Networks		Human Contact Networks	
	israel	damascus	obama	assad	bhp	rob	crime-moreno	infect-dublin
$\|V^*\|$	3698	3052	3212	2139	1686	1480	829	410
$\|E^*\|$	4165	3869	3423	2788	4623	3935	1475	2765
$\bar{d}$	2.25	2.53	2.13	2.61	5.48	5.32	3.56	13.49
γ	0.0037	0.012	0.0037	0.016	0.46	0.44	0.0059	0.46
$\|C\|$	1	1	1	1	31	21	1	1
$\|C_{max}\|$	-	-	-	-	1613	1430	-	-

- *Observed degree* $d(u)$: the degree of node u within the current graph G_t;
- *Average neighbor degree* $\bar{d}(N_{G_t}(u))$: the mean observed degree of the neighborhood of u;
- *Median neighbor degree* $\tilde{d}(N_{G_t}(u))$: the median observed degree of the neighborhood of u;
- *Probed neighbor fraction* $\rho_p(u)$: the ratio of explored neighbors to the total observed neighborhood of u.

The expected influence spread $\hat{\sigma}_{G^*}(S_B)$ for the identified seed sets was estimated using the Influence Maximization via Martingales (IMM) algorithm on the discovered subgraph G_B. The IMM parameters were configured with a budget $k = 10$, approximation parameter $\epsilon = 0.5$, and sampling parameter $l = 1$ under the IC model [13].

Algorithm Instantiation. To demonstrate the model-agnostic nature of the CANCEL framework, we evaluated it using different combinations of algorithms for the hierarchical layers.

For the *StrategyBandit* (the meta-learner selecting between Global and Component-focused strategies), we employed two standard stochastic bandit algorithms [15, 20]:

- UCB1, with an exploration parameter $\alpha = 1.0$;
- ε-Greedy, with exploration rate $\varepsilon = 0.1$.

For the *NodeBandit* (the contextual agent selecting specific nodes), we implemented two distinct approaches to handle the feature space defined above:

- LinUCB [17]: A parametric approach assuming a linear relationship between the feature vector $x_{a,t}$ and the reward. We configured the exploration parameter $\alpha = 1.0$.
- kNN-UCB [19]: A non-parametric approach that estimates rewards based on the history of similar nodes. We configured the method with $k = 25$ nearest neighbors and a sliding history window of $\tau = 500$ observations to adapt to the evolving graph structure. The exploration parameter was set to $\alpha = 1.0$.

In all contextual implementations, feature vectors were normalized using a standard scaler updated online as new nodes were discovered.

In the following section, unless otherwise specified, the notation CANCEL refers to the specific instantiation of the framework employing UCB1 for the *StrategyBandit* and LinUCB for the *NodeBandit*. All other configurations will be explicitly denoted by their respective component algorithms.

5.2 Experimental Results and Analysis

In this section, we validate the performance and robustness of the proposed framework. Our experimental evaluation is designed to answer the following four research questions:

- **RQ1.** Does CANCEL outperform state-of-the-art network exploration baselines in maximizing influence spread under strict budget constraints?
- **RQ2.** Is the maximization of the discovered graph volume a necessary condition for influence maximization, or can a smaller, representative subgraph yield superior results?
- **RQ3.** Does the hierarchical stacking of bandits provide a tangible benefit over single-level contextual strategies or pure component-based heuristics?
- **RQ4.** How sensitive is the framework to the choice of internal bandit algorithms and the size of the initial sample?

Influence Spread Comparative Analysis (RQ1). To address RQ1, we evaluate the effectiveness of the CANCEL framework in maximizing the expected influence spread $\hat{\sigma}_{G^*}(S_B)$ for seed sets S_B identified via IMM [21] on the sampled subgraph G_B. Figure 2 depicts the performance trajectories across a query budget B ranging from 1 to 30. Empirical evidence indicates that CANCEL yields a significant advantage over baseline methods across the majority of datasets, with the performance gap most pronounced under tight budget constraints ($B \leq 10$). Even at $B = 5$, CANCEL achieves substantial gains over state-of-the-art benchmarks. While the performance gap is narrower in the *Animal Networks* due to their highly fragmented topology with dense local clustering and isolated components, which hinder inter-component transitions and provide sparse feedback, CANCEL maintains a consistent lead. As budget increases to $B = 30$, the relative advantage diminishes on some networks as baselines accumulate sufficient structural information, though CANCEL maintains superior or competitive performance across all topologies.

Discovery Efficiency vs. Volume (RQ2). A critical insight emerging from this analysis is the decoupling of influence maximization from

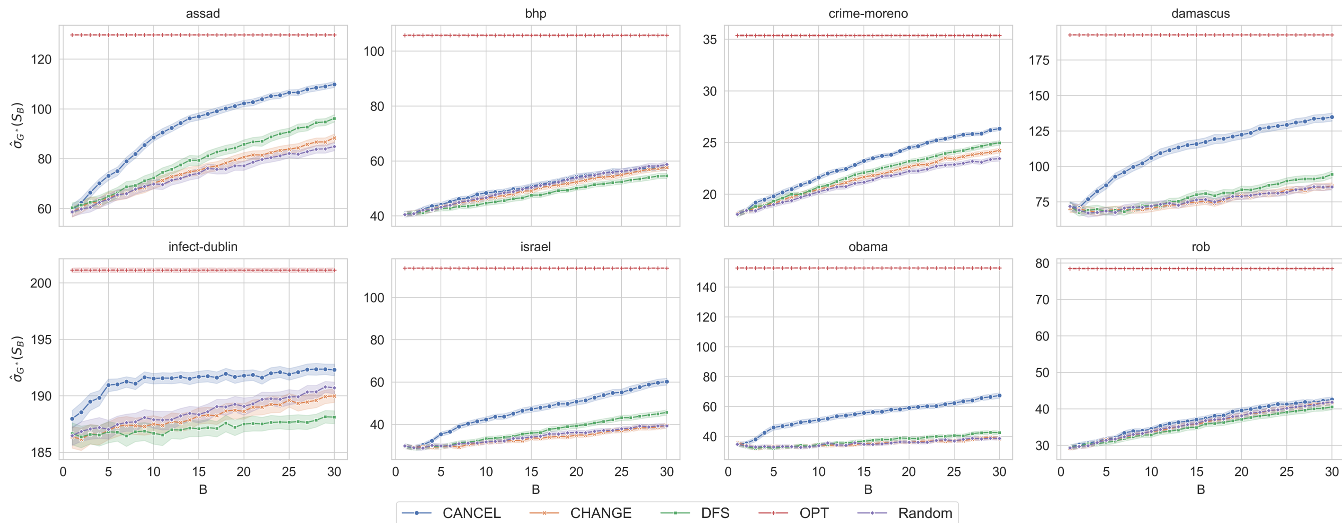

Figure 2: Expected influence spread $\hat{\sigma}_{G^*}(S_B)$ comparison between CANCEL and the baselines. Results demonstrate consistent outperformance across diverse network topologies.

exhaustive network discovery. While conventional approaches often equate exploration success with the volumetric expansion of the observed subgraph, CANCEL prioritizes the acquisition of a *topologically representative* proxy. Figure 3 highlights this distinction. In the *infect-dublin* network, for example, CANCEL discovers fewer total nodes than several baselines yet achieves a superior influence spread. This confirms our core hypothesis: the algorithm effectively targets high-centrality substructures containing influential seeds, rather than merely maximizing the cardinality of V_B. Conversely, on other topologies, CANCEL excels in both discovery rate and influence, suggesting that its hierarchical framework dynamically adapts its objective, shifting between representative sampling and exhaustive exploration, based on the underlying topological feedback. Although Figure 3 depicts results for a representative subset of the benchmark datasets, these trends remain consistent across all evaluated topologies, with CANCEL demonstrating sustained superiority in both network discovery and influence spread.

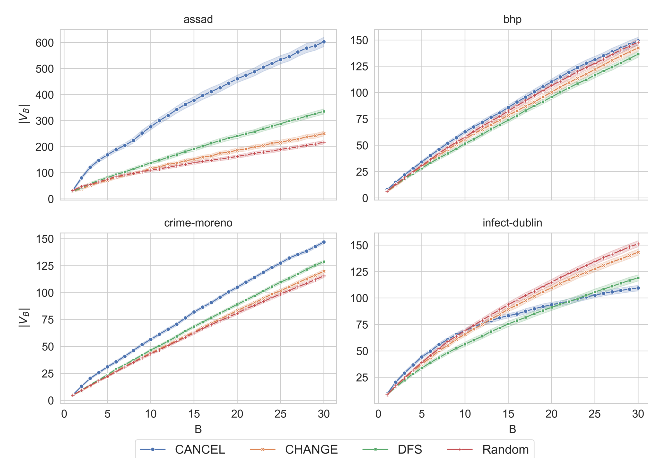

Figure 3: Growth of the discovered subgraph $|V_B|$ across some representative datasets. CANCEL demonstrates that maximizing the number of discovered nodes is not a prerequisite for optimal influence spread.

Ablation Study: Hierarchical Architecture (RQ3). To isolate the contribution of the hierarchical framework, we conducted an ablation study comparing CANCEL against two restricted variants: *(i)* a standard single-level contextual bandit (LinUCB) and *(ii)* a strategy exclusively focused on component exploration (CANCEL-ComponentsOnly). As illustrated in Figure 4, the full hierarchical model demonstrates superior robustness. Compared to the standard LinUCB, our approach achieves an average improvement in influence spread of 2.35% at $B = 5$, scaling to 6.57% by $B = 30$. This confirms that integrating component-level discovery rewards provides a more effective learning signal than traditional contextual bandits alone, particularly as the budget allows for deeper exploitation of the learned structure. Furthermore, while the CANCEL-ComponentsOnly variant remains competitive on specific datasets (e.g., *Infect-Dublin*, *Obama*), it exhibits significant volatility on others due to its rigid focus. For instance, at $B = 5$, CANCEL outperforms the component-only variant by 8.6% on the *Assad* network and 4.9% on *Israel*. These results demonstrate that while component-aware exploration is a powerful heuristic, it is insufficient as a standalone strategy. The hierarchical meta-bandit is therefore essential for generalizability, dynamically arbitrating between strategies to capture the benefits of component targeting while mitigating the performance risks inherent in purely heuristic approaches.

Sensitivity to Algorithmic Instantiations (RQ4). We further examined the impact of specific bandit configurations by evaluating

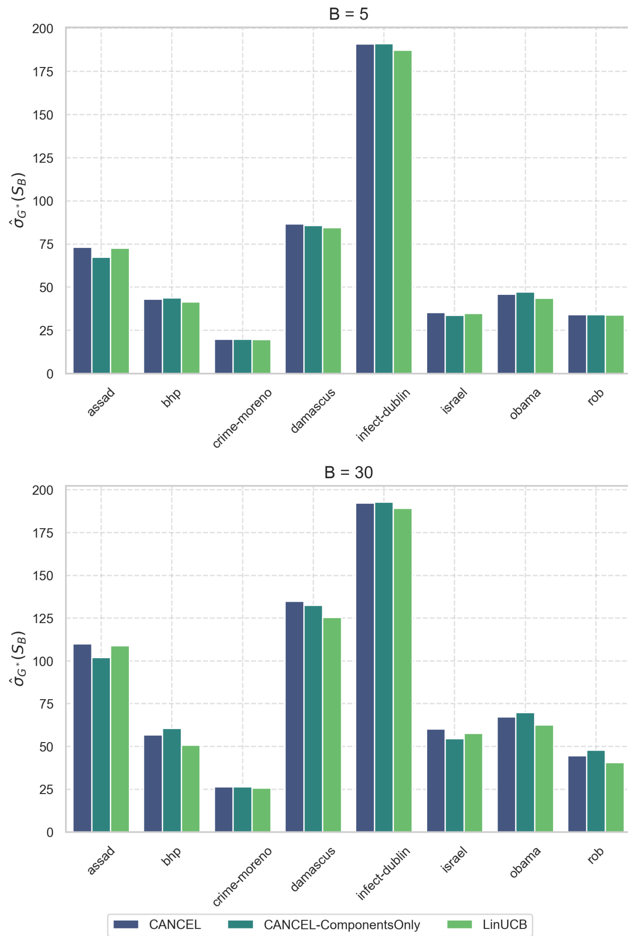

Figure 4: Ablation analysis comparing the hierarchical CANCEL framework against single-level and component-only baselines at $B = 5$ (top) and $B = 30$ (bottom). CANCEL's adaptive selection consistently yields superior or more robust performance.

four combinations of *StrategyBandit* (Meta-level) and *NodeBandit* (Contextual-level). To facilitate a direct comparison across heterogeneous network scales, the results in Figure 5 are presented as a relative performance heatmap. For each dataset, the expected influence spread of a specific variant is normalized by the maximum influence achieved by the best-performing configuration on that same topology.

The primary CANCEL configuration (UCB1 + LinUCB) proves to be the most resilient, attaining the highest influence spread in the majority of datasets. As evidenced by the consistently high color intensity across the corresponding column in Figure 5, it maintains near-optimal performance across diverse topologies and exhibits minimal regret even in its least favorable scenarios. In contrast, variants relying on kNN-UCB exhibit higher volatility, suffering noticeable performance degradation on complex networks such as *obama* and *assad*. This confirms that the linear approximation of LinUCB, paired with the rigorous uncertainty bounds of UCB1,

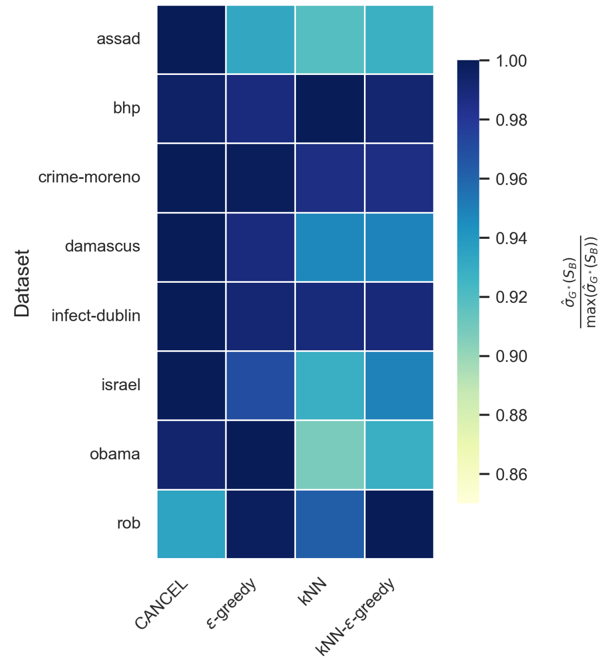

Figure 5: Robustness analysis of algorithmic variants at $B = 5$. The heatmap displays the performance ratio $\hat{\sigma}_{G^*}(S_B)/\max(\hat{\sigma}_{G^*}(S_B))$, where 1.0 indicates the best result achieved by any variant for a given dataset. The CANCEL (UCB1+LinUCB) configuration demonstrates superior stability across diverse topologies.

offers the most effective balance for high-dimensional network exploration, avoiding the instability often associated with nearest-neighbor approaches in sparse data regimes.

Robustness to Initial Sample Size (RQ4). Finally, to assess the algorithm's dependency on initial data scarcity, we evaluated performance when the initial random sample size $|S_0|$ is increased from 5 to 50 nodes. Figure 6 presents the comparative results across all datasets with this expanded initialization ($B = 5$). The results indicate that CANCEL maintains its superiority even when provided with significantly more prior information. Specifically, CANCEL outperforms the baselines on 7 out of 8 datasets. The gains are particularly pronounced on structurally complex networks such as *Obama* and *Israel*, where CANCEL improves over the best baseline by 64.1% and 39.5%, respectively. On *infect-dublin*, performance is comparable to the baselines, likely due to saturation effects where the larger observable graph sample already captures the primary influence paths. This robustness confirms that the algorithm effectively leverages the available budget to identify high-influence nodes regardless of the starting sample size.

6 CONCLUSION

This study introduces CANCEL, a hierarchical bandit framework for *Influence Maximization* (IM) under extreme network unobservability. While motivated by public health interventions among

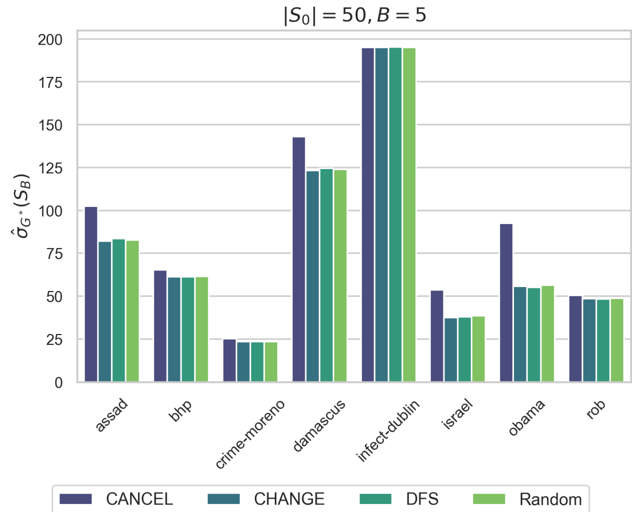

Figure 6: Expected influence spread $\hat{\sigma}_{G^*}(S_B)$ comparison between CANCEL and the baselines across different datasets, with an original sample size of 50 nodes and a budget $B = 5$.

marginalized populations, our approach is domain-agnostic for settings where global topological data is unavailable.

Our empirical evaluation demonstrates that CANCEL significantly outperforms state-of-the-art methods with minimal initial information. The algorithm prioritizes high-centrality nodes over volumetric expansion, a "strategic discovery" phenomenon validated by our ablation study showing the hierarchical architecture is essential for mitigating component-bias in single-level contextual models.

Limitations and Future Work. Several constraints merit acknowledgment and present avenues for future research. All evaluated networks contain fewer than 3,700 nodes; extending validation to massive-scale synthetic and real-world networks is necessary to demonstrate scalability. We substitute generic human-contact networks for unavailable marginalized-population datasets, preserving relevant topological characteristics but not directly validating the target domain. The exploration phase optimizes edge discovery as a proxy for influence spread; while empirically effective, a theoretical analysis quantifying this gap remains open. We provide no regret bounds and pursuing rigorous convergence and regret analysis is a priority. Finally, addressing truthful neighborhood disclosure through privacy-preserving mechanisms, and moving beyond reach maximization to incentivize long-term behavioral change, represents a vital step toward real-world public health deployment.

ACKNOWLEDGMENTS

This work was supported by the PNRR project FAIR – Future AI Research (PE00000013) under the NRRP MUR program funded by NextGenerationEU.

REFERENCES

[1] Vincenzo Auletta, Francesco Carbone, Diodato Ferraioli, and Cosimo Vinci. 2025. Adaptive Multi-Round Influence Maximization with Limited Information. In *AAMAS '25*. 2408–2410.

[2] Vincenzo Auletta, Diodato Ferraioli, and Grazia Ferrara. 2025. Influence Maximization in Unknown Social Networks: A Contextual Bandit Approach. In *International Conference on Principles and Practice of Multi-Agent Systems*. Springer, 224–231.

[3] Vittorio Bilò, Diodato Ferraioli, and Cosimo Vinci. 2022. General Opinion Formation Games with Social Group Membership. In *IJCAI*. 88–94.

[4] David Bindel, Jon M. Kleinberg, and Sigal Oren. 2015. How bad is forming your own opinion? *Games Econ. Behav.* 92 (2015), 248–265.

[5] Robert Bredereck, Lilian Jacobs, and Leon Kellerhals. 2020. Maximizing the Spread of an Opinion in Few Steps: Opinion Diffusion in Non-Binary Networks. In *IJCAI*. 1622–1628.

[6] Matteo Castiglioni, Diodato Ferraioli, Nicola Gatti, and Giulia Landriani. 2021. Election manipulation on social networks: Seeding, edge removal, edge addition. *Journal of Artificial Intelligence Research* 71 (2021), 1049–1090.

[7] Federico Corò, Emilio Cruciani, Gianlorenzo D'Angelo, and Stefano Ponziani. 2019. Exploiting Social Influence to Control Elections Based on Scoring Rules. In *IJCAI*. 201–207.

[8] Debojit Das, Shweta Jain, and Sujit Gujar. 2022. Budgeted Combinatorial Multi-Armed Bandits. In *AAMAS*. 345–353.

[9] Pedro M. Domingos and Matthew Richardson. 2001. Mining the network value of customers. In *SIGKDD*. ACM, 57–66.

[10] Diodato Ferraioli, Paul W. Goldberg, and Carmine Ventre. 2016. Decentralized dynamics for finite opinion games. *Theor. Comput. Sci.* 648 (2016), 96–115.

[11] Victor Gabillon, Branislav Kveton, Zheng Wen, Brian Eriksson, and Shanmugavelayutham Muthukrishnan. 2013. Adaptive submodular maximization in bandit setting. *NeurIPS* 26 (2013).

[12] Harshavardhan Kamarthi, Priyesh Vijayan, Bryan Wilder, Balaraman Ravindran, and Milind Tambe. 2020. Influence maximization in unknown social networks: Learning policies for effective graph sampling. In *AAMAS*. 575–583.

[13] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *KDD*. 137–146.

[14] David Kempe, Jon M. Kleinberg, and Éva Tardos. 2005. Influential Nodes in a Diffusion Model for Social Networks. In *ICALP*. 1127–1138.

[15] Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.

[16] Dexun Li, Meghna Lowalekar, and Pradeep Varakantham. 2021. CLAIM: Curriculum learning policy for influence maximization in unknown social networks. In *UAI*. PMLR, 1455–1465.

[17] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *WWW*. 661–670.

[18] Yandi Li, Haobo Gao, Yunxuan Gao, Jianxiong Guo, and Weili Wu. 2023. A survey on influence maximization: From an ml-based combinatorial optimization. *ACM Transactions on Knowledge Discovery from Data* 17, 9 (2023), 1–50.

[19] Kaushalya Madhawa and Tsuyoshi Murata. 2019. A multi-armed bandit approach for exploring partially observed networks. *Applied Network Science* 4 (2019), 1–18.

[20] Aleksandrs Slivkins et al. 2019. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning* 12, 1-2 (2019), 1–286.

[21] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence maximization in near-linear time: A martingale approach. In *SIGMOD*. 1539–1554.

[22] Cong Tran, Won-Yong Shin, and Andreas Spitz. 2024. IM-META: Influence maximization using node metadata in networks with unknown topology. *IEEE Transactions on Network Science and Engineering* 11, 3 (2024), 3148–3160.

[23] Yingfei Wang, Inbal Yahav, and Balaji Padmanabhan. 2024. Smart testing with vaccination: A bandit algorithm for active sampling for managing covid-19. *Information Systems Research* 35, 1 (2024), 120–144.

[24] Bryan Wilder, Nicole Immorlica, Eric Rice, and Milind Tambe. 2018. Maximizing influence in an unknown social network. In *AAAI*, Vol. 32.

[25] Bryan Wilder, Laura Onasch-Vera, Juliana Hudson, Jose Luna, Nicole Wilson, Robin Petering, Darlene Woo, Milind Tambe, and Eric Rice. 2018. End-to-End Influence Maximization in the Field.. In *AAMAS*, Vol. 18. 1414–1422.

[26] Bryan Wilder and Yevgeniy Vorobeychik. 2018. Controlling Elections through Social Influence. In *AAMAS*. 265–273.

[27] Bryan Wilder, Amulya Yadav, Nicole Immorlica, Eric Rice, and Milind Tambe. 2017. Uncharted but not Uninfluenced: Influence Maximization with an uncertain network. In *AAMAS*. 1305–1313.

[28] Amulya Yadav, Hau Chan, Albert Xin Jiang, Haifeng Xu, Eric Rice, and Milind Tambe. 2016. Using Social Networks to Aid Homeless Shelters: Dynamic Influence Maximization under Uncertainty.. In *AAMAS*, Vol. 16. 740–748.

[29] Bo Yan, Kexiu Song, Jiamou Liu, Fanku Meng, Yiping Liu, and Hongyi Su. 2019. On the Maximization of Influence Over an Unknown Social Network.. In *AAMAS*, Vol. 19. 13–17.