Online Influence Maximization with Local Observations

Julia Olkhovskaya
Universitat Pompeu Fabra
Barcelona, Spain
julia.olkhovskaya@gmail.com

Gergely Neu
Universitat Pompeu Fabra
Barcelona, Spain
gergely.neu@gmail.com

Gábor Lugosi
ICREA & Universitat Pompeu Fabra
Barcelona, Spain
gabor.lugosi@gmail.com

Abstract

We consider an online influence maximization problem in which a decision maker selects a node among a large number of possibilities and places a piece of information at the node. The node transmits the information to some others that are in the same connected component in a random graph. The goal of the decision maker is to reach as many nodes as possible, with the added complication that feedback is only available about the degree of the selected node. Our main result shows that such local observations can be sufficient for maximizing global influence in two broadly studied families of random graph models: stochastic block models and Chung–Lu models. With this insight, we propose a bandit algorithm that aims at maximizing local (and thus global) influence, and provide its theoretical analysis in both the subcritical and supercritical regimes of both considered models. Notably, our performance guarantees show no explicit dependence on the total number of nodes in the network, making our approach well-suited for large-scale applications.

1 Introduction

Finding most influential nodes in social networks has received increasing attention in the last few years. The problem has been cast in a variety of different ways according to the notion of influence and the information available to a decision maker. We refer the reader to [18, 9, 10, 27, 7, 29, 28] and the references therein for recent progress in various directions. In the present paper, we consider the problem of maximizing influence in a sequential setup where the learner has only partial information about the set of influenced nodes.

Specifically, we define and explore a sequential decision-making model in which the goal of a decision maker is to find a node among a set V of n possible nodes with maximal (expected) influence. In our model, at every time instance $t = 1, \dots, T$, the n nodes form the vertex set of a random graph G_t such that node i and node j are connected in G_t by an (undirected) edge with probability $p_{i,j}$. All edges are present independently of each other and the graphs $G_1, \dots, G_T$ are independent and identically distributed. If the decision maker selects a node $v_t \in V$ at time t, then the information placed at the node spreads to the entire connected component of v_t in the graph G_t. The goal of the decision maker is to spread the information as much as possible, which can be formulated as maximizing a notion of *rewards* corresponding to the number of vertices in the connected component containing the selected node.

Preprint. Work in progress.

In this paper, we study a setting where the decision maker has no prior knowledge of the distribution of G_t, so it has to learn about this distribution on the fly, while simultaneously attempting to maximize the total rewards. This gives rise to a dilemma of *exploration versus exploitation*, which is commonly studied within the framework of *multi-armed bandits* [5]. Indeed, if the decision maker could observe the set of all influenced nodes in every round, the sequential influence maximization problem outlined above could be naturally formulated as a *stochastic multi-armed bandit* problem [20, 2]. However, this direct approach has multiple setbacks. First off, in most practical applications, the number n of nodes is so large that one cannot even hope to maintain individual statistics about each of them, let alone expect any algorithm to identify the most influential node in any reasonable time. More importantly, in most cases of practical interest, tracking down the set of *all* influenced users may be difficult or downright impossible due to privacy and computational considerations. This motivates the study of a more restrictive setting where the decision maker has to manage with only partial observations of the set of influenced nodes.

Formally, we address this latter challenge by considering a more realistic observation model, where after selecting a node v_t to be influenced, the learner only observes the number of *immediate neighbors* of v_t in the realized random graph G_t (i.e., the degree of v_t in G_t). This model brings up the following important question: is it possible to maximize *global* influence while only having access to such local measurements? Our key technical result is answering this question in the positive for some broadly studied random graph models. Specifically, we show that, assuming that the graphs G_t are generated from certain stochastic block models [1] or a Chung–Lu model [14], maximizing local influence is equivalent to maximizing global influence.

This observation motivates our algorithmic approach that applies ideas from the multi-armed bandit literature to try and maximize the local influence of each selected node. In order to analyze the performance of our algorithms, we adapt the standard notion of *regret* from said literature to fit our needs. The traditional notion of regret measures the difference between the cumulative reward of choosing a maximally influential node in each round and the cumulative reward the decision maker achieves during the T rounds of the game. This definition, however, is rather problematic in our problem setup: as mentioned above, the number n of all nodes is typically so large that finding a maximally influential node is computationally infeasible, thus making the task of competing with this benchmark unreasonably complicated. To resolve this issue, we consider the notion of *quantile regret* that compares the learner's performance to the top α-fraction of all nodes ([8, 12, 22, 19]). Our main result is showing both instance-dependent bounds of order $\log T$ and worst-case bounds of order $\sqrt{T}$ on the quantile regret of our algorithm. Notably, our bounds hold for both the subcritical and supercritical regimes of the random-graph models considered, and show no explicit dependence on the number of nodes n.

Related online influence maximization algorithms consider more general classes of networks, but make more restrictive assumptions about the interplay between rewards and feedback. One line of work explored by Wen et al. [29], Wang and Chen [28] assumes that the algorithm receives *full feedback* on where the information reached in the previous trials (i.e., not only the number of influenced nodes, but their exact identities and influence paths, too). Clearly, such detailed measurements are nearly impossible to obtain in practice, as opposed to the local observations that our algorithm requires. Another related setup was considered by Carpentier and Valko [7], whose algorithm only receives feedback about the nodes that were directly influenced by the chosen node, but the model does not assume that neighbors in the graph share the information to further neighbors and counts the reward only by the nodes directly connected to the selected one. That is, in contrast to our work, this work does not attempt to show any relation between local and global influence maximization. One downside to all the above works is that they all provide rather conservative performance guarantees: On one hand, Wen et al. [29] and Carpentier and Valko [7] are concerned with worst-case regret bounds that uniformly hold for all problem instances for a fixed time horizon T. On the other hand, the bounds of Wang and Chen [28] depend on topological (rather than probabilistic) characteristics of the underlying graph structure, which inevitably leads to conservative results. For example, their bounds instantiated in our graph model lead to a regret bound of order $n^3 \log T$, which is virtually void of meaning in our regime of interest where n is very large (e.g, in the order of billions). In contrast, our bounds do not show explicit dependence on n. In this light, our work can be seen as the first one that takes advantage of a specific graph structure to obtain strong instance-dependent global performance guarantees, all while having access to only local observations.

The rest of the paper is organized as follows. In Section 2 we formally introduce the regret minimization problem and the notation. In Section 3, we introduce our key technical results that show the connection between local and global influence maximization. We describe our algorithm and state its performance guarantees in Section 4. We describe the main structure of the analysis in Section 5 and discuss our results in Section 6.

2 Preliminaries

We now describe our problem and model assumptions more formally. We consider the problem of sequential influence maximization on a fixed finite graph (V, E), formalized as a repeated interaction scheme between a learner and its environment. In this setup, the following steps are repeated for each round $t = 1, 2, \ldots$:

1. the learner picks a vertex $A_t \in V$,
2. the environment generates a subgraph G_t of (V, E),
3. the learner observes the degree of the node A_t, denoted as $Y_{A_t,t}$,
4. the learner earns the reward $r_{A_t,t} = |C_{A_t,t}|$, where the set $C_{a,t}$ is the connected component associated with vertex a:

$$C_{a,t} = \{v \in V : v \text{ is connected to } a \text{ by a path in } G_t\} \ .$$

We stress that the learner does *not* observe the reward, only the number of its immediate neighbors. Define c_a as the expected size of the connected component associated with the node a: $c_a = \mathbb{E}|C_{a,t}|$. Ideally, we would be interested in designing algorithms that minimize the *expected regret* defined as

$$R_T = \max_{a \in V} \sum_{t=1}^{T} (c_a - c_{A_t}) \,. \tag{1}$$

We would ideally aspire to design algorithms guaranteeing that the regret grows sublinearly in T. However, as we are interested in settings where the total number of nodes n is very large, this goal can be seen as far too ambitious: even with a fully known random graph model, finding the optimal node maximizing c_a is computationally infeasible. Such computational issues have lead to alternative definitions of the regret such as the *approximation regret* [17, 11, 25] or the *quantile regret* [8, 12, 22, 19].

In the present paper, we consider the α-quantile regret as our performance measure, which, instead of measuring the learner's performance against the single best decision, uses a near-optimal action as a baseline. For a more technical definition, let $a_1, a_2, \ldots, a_n$ be an ordering of the nodes satisfying $c_{a_1} \le c_{a_2} \le \cdots \le c_{a_n}$, and denote the α-quantile over the mean rewards as $c^*_\alpha = c_{a_{\lceil (1-\alpha)n \rceil}}$. Then, also defining the set $V^*_\alpha = \{a_{\lceil (1-\alpha)n \rceil}, \ldots, a_n\}$ as the set of α-near-optimal nodes, we define the α-quantile regret as

$$R^\alpha_T = \min_{a \in V^*_\alpha} \sum_{t=1}^{T} (c_a - c_{A_t}) = \sum_{t=1}^{T} (c^*_\alpha - c_{A_t}) \,. \tag{2}$$

We will make the assumption that each G_t is drawn from a fixed (and unknown) distribution of *inhomogeneous random graphs* (IRG, see, e.g.,[4]). In this model, we assume that (V, E) is the complete graph over n vertices and each edge (i, j) is present with probability $p_{ij} (= p_{ji})$, independently of all other edges. The inhomogeneous random graph can be parametrized by the symmetric positive matrix $\overline{A}$, such that the probability of i and j being connected is given by $p_{ij} = \overline{A}_{ij}/n$. We will assume that each element $\overline{A}_{ij}$ of the matrix is $O(1)$ as n grows large. This assumption corresponds to assuming that the graphs G_t are *sparse*, meaning that the expected degree of each vertex remains bounded as n grows. This assumption makes the problem both more realistic and challenging: denser graphs are connected with high probability, making the problem essentially vacuous. We will also use the notation $A = \overline{A}/n$. The random graph from the above distribution is denoted as $G(n, A)$.

We consider two fundamentally different regimes of the parameters $G(n, A)$: the *subcritical* case in which the size of the largest connected component is sublinear in n (with high probability), and the *supercritical* case where the largest connected component is at least of size cn for some $c > 0$ with high probability. (We say that an event holds *with high probability* if its probability converges to one as $n \to \infty$.) Such a connected component of linear size is called a *giant component*. These regimes can be formally characterized with the help of the matrix A. Letting λ be the the largest eigenvalue of A, we call $G(n, A)$ subcritical if $\lambda < 1$ and supercritical if $\lambda > 1$. It follows from [4, Theorem 3.1] that, with high probability, $G(n, A)$ has a giant component if it is supercritical, while the number of vertices in the largest component is $o(n)$ with high probability if it is subcritical.

Within the class of inhomogeneous random graphs, we will focus on two families of random graphs: *stochastic block models* and *Chung–Lu models*, as discussed below.

2.1 Stochastic block models

First, we make the assumption that each G_t is drawn from a *stochastic block model* (SBM). In this random-graph model, the probabilities p_{ij} are defined through the notion of *communities*, defined as elements of a partition $H_1, \dots, H_S$ of the set of vertices V. We will refer to the index m of community H_m as the *type* of vertices belonging to H_m. Each community H_m contains $\alpha_m n$ nodes (assuming without loss of generality that $\alpha_m n$ is an integer). With the help of the community structure, the probabilities p_{ij} are constructed as follows: if $i \in H_\ell$ and $j \in H_m$, the probability of i and j being connected is given by $p_{ij} = \frac{K_{\ell,m}}{n}$, where K is a symmetric matrix of size $S \times S$, with positive elements. The random graph from the above distribution is denoted as $G(n, \alpha, K)$.

In an SBM, identifying a node with maximal reward amounts to finding a node from the most influential community. Consequently, it is easy to see that choosing α such that $\alpha > \min_m \alpha_m$, the near-optimal set V_α^* will exactly correspond to the set of optimal nodes, and thus the quantile regret (2) will coincide with the regret (1).

Throughout the paper, we will consider SBM's satisfying the following assumption:

Assumption 1. $K_{i,j} = k > 0$ *for all* $i \neq j$.

Intuitively, this assumption requires that nodes i, j belonging to different communities are connected with the same probability, regardless of the exact identity of $m(i)$ or $m(j)$. Additionally, our analysis in the supercritical case will make the following natural assumption:

Assumption 2. *For all* i, $K_{i,i} \geq k$.

In plain words, this assumption requires that the density of edges within communities is larger than the density of edges between communities.

2.2 Chung–Lu models

We will also consider another natural IRG model that is closely related to many random graph models. This is the so-called *Chung–Lu model* (sometimes referred to as *rank-1 model*) as first defined by Chung and Lu [14] (see also [13, 4]), where the edge probabilities are defined through the positive vector $w \in \mathbb{R}^n$, with elements of the matrix given by $\overline{A}_{ij} = w_i w_j$. In other words, the Chung–Lu model considers rank-1 matrices of the form $\overline{A} = ww^\top$.

The random graph from the induced IRG distribution is denoted as $G(n, w)$. Such models can be shown to exhibit several interesting properties. For instance, if w is a sequence satisfying a power law, then $G(n, w)$ is a power law model, which allows one to model various real world networks including social networks [13].

3 From local to global influence maximization

Having described the setting, we can finally ask the question: is it possible to maximize global influence using only local observations? Our main technical results show that, perhaps surprisingly, the most influential nodes are actually identifiable from such feedback in the models discussed in Sections 2.1 and 2.2.

To be specific, we recall that $Y_{a,t}$ stands for the degree of node a in the realized graph G_t, and define $\mu_a = \mathbb{E}Y_{a,1}$ as the expected degree of node a. We also define $c^* = \max_a c_a$ and $\mu^* = \max_a \mu_a$. Our

```
**Algorithm 1** $d$-UCB$(V_0)$
**Parameters:** A set of nodes $V_0 \subseteq V$.
**Initialization:** Select each node in $V_0$ once. Observe the degree $X_{a,a}$ of vertex $a$ in the graph $G_a$ for $a = 1, \dots, |V_0|$. For each $a \in V_0$, set $N_a(|V_0|) = 1$ and $\widehat{\mu}_a(|V_0|) = X_{a,a}$.
**For** $t = |V_0|, \dots T$, **repeat**
    1. For each node, compute
       $U_a(t) = \sup \left\{ \mu : \mu - \widehat{\mu}_a(t) + \widehat{\mu}_a(t) \log\left(\frac{\widehat{\mu}_a(t)}{\mu}\right) \le \frac{3\log(t)}{N_a(t)} \right\}.$
    2. Select any node $A_{t+1} \in \arg\max_a U_a(t)$.
    3. Observe degree $Y_{A_{t+1},t+1}$ of node $A_{t+1}$ in $G_{t+1}$ and update
       $\widehat{\mu}_{A_{t+1}}(t+1) = \frac{N_{A_{t+1}}(t)\widehat{\mu}_{A_{t+1}}(t+1) + Y_{A_{t+1},t+1}}{N_{A_{t+1}}(t)+1}.$
       Update $N_{A_{t+1}}(t+1) = N_{A_{t+1}}(t) + 1$.
```

main technical result is proving that nodes with the largest expected degrees μ^* are exactly the ones with the largest influence c^*, in both the SBM and the Chung–Lu models, across both the subcritical and supercritical regimes. We formally state these results below.

Proposition 1. *Suppose that*

1. *G is generated from a subcritical $G(n, \alpha, K)$ satisfying Assumption 1, or*
2. *G is generated from a subcritical $G(n, w)$.*

Then, for any a satisfying $\mu_a < \mu^*$, we have $c^* - c_a \le 2c^* (\mu^* - \mu_a) + O(1/n)$. In particular, for n sufficiently large, we have $\arg\max_a c_a = \arg\max_a \mu_a$.

Proposition 2. *Suppose that*

1. *G is generated from a supercritical $G(n, \alpha, K)$ satisfying Assumptions 1 and 2, or*
2. *G is generated from a supercritical $G(n, w)$.*

Then, for any a satisfying $\mu_a < \mu^*$, we have $c^* - c_a \le c^* (\mu^* - \mu_a) + o(n)$. In particular, for n sufficiently large, we have $\arg\max_a c_a = \arg\max_a \mu_a$.

These propositions are proved in Appendix B and C, respectively. To the best of our knowledge, these results are novel and can be of independent interest. The proofs rely on the concept of multi-type Galton–Watson branching processes, which are briefly introduced alongside some of their main properties in Appendix A.

4 Algorithm and main results

We now present our learning algorithm, and provide its performance guarantees for the two regimes. Inspired by the observation that in the models that we consider, it is sufficient to identify nodes with maximal degree in order to maximize influence, we design a bandit algorithm that attempts to maximize the degrees of the influenced nodes. We propose to achieve this goal by employing a variant of the kl-UCB algorithm proposed and analyzed by [15, 23, 6, 21]. More precisely, we propose to use the observed degrees as rewards, and feed them to an instance of kl-UCB originally designed for Poisson-distributed rewards. A key technical challenge arising in the analysis is that the degree distributions do not actually belong to the Poisson family for finite n. We overcome this difficulty by showing that the degree distributions have a moment generating function bounded by those of Poisson distributions, and that this fact is sufficient for most of the kl-UCB analysis to carry through without changes.

A minor challenge is that, since we are interested in very large values of n, it is computationally infeasible to use *all* nodes as separate actions in our bandit algorithm. To address this challenge, we

```
Algorithm 2 $d$-UCB-DOUBLE($\beta$)
Parameters: $\beta \geq 2$.
Initialization: $V_0 = \emptyset$.
For $k = 1, 2 \ldots,$ repeat
    1. Sample subset of nodes $U_k$ uniformly such that $|U_k| = \left\lceil \frac{\log \beta}{\log(1/(1-\alpha))} \right\rceil$.
    2. Update action set $V_k = V_{k-1} \cup U_k$.
    3. For rounds $t = \beta^{k-1}, \beta^{k-1} + 1, \ldots, \beta^k - 1$, run a new instance of $d$-UCB $(V_k)$.
```

propose to *subsample* a set of representative nodes for kl-UCB to play on. The size of the subsampled nodes depends on the quantile α targeted in the regret definition (2) and the time horizon T. For clarity of presentation, we first propose a simple algorithm that assumes prior knowledge of T, and then move on to construct a more involved variant that adds new actions on the fly.

We first present our kl-UCB variant for a fixed set of nodes V_0 as Algorithm 1. We refer to this algorithm as d-UCB(V_0) (short for "degree-UCB on V_0"). Our two algorithms mentioned above use d-UCB (V_0) as a subroutine: they are both based on uniformly sampling a large enough set V_0 of nodes so that the subsample includes at least one node from the top α-quantile.

To simplify the presentation of our main results, let us introduce some more notation. Analogously to the α-optimal reward c^*_α, we define the α-optimal degree $\mu^*_\alpha = \min_{a \in V^*_\alpha} \mu_a$, and the corresponding gap parameters $\Delta_{\alpha,i} = (c_i - c^*_\alpha)_+$ and $\delta_{\alpha,i} = (\mu_i - \mu^*_\alpha)_+$. Finally, define $\Delta_{\alpha,\max} = \max_i \Delta_{\alpha,i}$. We first present a performance guarantee of our simpler algorithm that assumes knowledge of T. This method uniformly samples a subset of size

$$|V_0| = \left\lceil \frac{\log T}{\log(1/(1-\alpha))} \right\rceil \tag{3}$$

and plays d-UCB(V_0) on the resulting set. This algorithm satisfies the following performance guarantee:

Theorem 1. *Let V_0 be a uniform subsample of V with size given in Equation* (3) *and define the event $\mathcal{E} = \{V_0 \cap V^*_\alpha \neq \emptyset\}$. If the number of vertices n is sufficiently large, then the expected α-quantile regret of d-UCB(V_0) simultaneously satisfies*

$$R^\alpha_T \leq \mathbb{E}\left[\sum_{i \in V_0} \Delta_{\alpha,i} \left(\frac{\mu^*_\alpha \left(18 + 27 \log T\right)}{\delta^2_{\alpha,i}} + 3 \right) \middle| \mathcal{E} \right] + \Delta_{\alpha,\max},$$

where the expectation is taken over the random choice of V_0, and

$$R^\alpha_T \leq 18 c^* \sqrt{\frac{T \mu^* \left(2 + 3 \log T\right)^2}{\log(1/(1-\alpha))}} + \left(\frac{3 \log T}{\log(1/(1-\alpha))} + 4 \right) \Delta_{\alpha,\max}.$$

For unknown values of T, we propose the d-UCB-DOUBLE(β) algorithm (presented as Algorithm 2) that uses a doubling trick to estimate T. The following theorem gives a performance guarantee for this algorithm:

Theorem 2. *Fix T, let $k_{\max}$ be the value of k on which d-UCB-DOUBLE(β) terminates, and define the event $\mathcal{E} = \{V_{k_{\max}} \cap V^*_\alpha = \emptyset\}$. If the number of vertices n is sufficiently large, then the α-quantile regret of d-UCB-DOUBLE(β) simultaneously satisfies*

$$R^\alpha_T \leq \mathbb{E}\left[\sum_{i \in V_{k_{\max}}} \Delta_i \left(\left(\frac{18 \mu^*}{\delta^2_{\alpha,i}} + 3 \right) (\log_\beta T + 1) + \frac{27 \log \beta (\log_\beta T + 1)^2}{2 \delta^2_{\alpha,i}} \right) \middle| \mathcal{E} \right] + \Delta_{\alpha,\max} \log_\beta T,$$

where the expectation is taken over the random choice of the sets $V_1, V_2, \ldots,$ and

$$R^\alpha_T \leq 36 c^* \sqrt{\frac{T \left(\mu^* + \log\left(\beta T\right)\right) \log^2 T}{\log(1/(1-\alpha))}} + \left(\frac{3 \log^2 T}{\log(1/(1-\alpha))} + 4 \right) \Delta_{\alpha,\max}.$$

We discuss the key features of the above regret bounds in Section 6.

5 Analysis

This section outlines the main ideas of the proofs of our main results, Theorems 1 and 2. Having established that, in order to minimize regret in our setting, it is sufficient to design an algorithm that quickly identifies the nodes with the highest degree, it remains to show that our algorithms indeed achieve this goal. We do this below by providing a bound on the expected number of times $\mathbb{E}N_a(T) = \mathbb{E}[\sum_{t=1}^{T} \mathbb{I}_{\{A_t=a\}}]$ that our algorithm picks suboptimal node a such that $c_a \leq c^*$, and then using this guarantee to bound the regret.

Without loss of generality, we assume that $V_0 = \{1, 2, \ldots, |V_0|\}$. The key to our regret bounds is the following guarantee on the number of suboptimal actions taken by d-UCB(V_0).

Theorem 3 (Number of suboptimal node plays in d-UCB)**.** *Define $\eta_i = \left(\max_{j \in V_0} \mu_j - \mu_i\right)/3$. The number of times that any node $i \in \{a : \mu_a < \max_{j\in V_0} \mu_j\}$ is chosen by d-UCB(V_0) satisfies*

$$\mathbb{E}N_i(T) \leq \frac{\mu^* \left(2 + 6 \log T\right)}{\eta_i^2} + 3\,. \tag{4}$$

The proof is largely based on the analysis of the kl-UCB algorithm due to Cappé et al. [6], with some additional tools borrowed from Ménard and Garivier [24], crucially using that the degree distribution of each node is stochastically dominated by an appropriately chosen Poisson distribution. Specifically, letting Z_i be a Poisson random variable with mean $\mathbb{E}Y_{i,t}$, we have $\mathbb{E}e^{sY_{i,t}} \leq \mathbb{E}e^{sZ_i}$ for all s. Turns out that this property is sufficient for the kl-UCB analysis to go through in our case, which is an observation that may be of independent interest.

Due to space constraints, the proof of Theorem 3 is deferred to Appendix D. The remainder of the section uses Theorem 3 to prove our first main result, Theorem 1. The proof of Theorem 2 follows from similar ideas and some additional technical arguments, and is presented in Appendix E.

Proof of Theorem 1. We first note that, with high probability, the size of V_0 guarantees that the subset holds at least one node from the set V_α^*: $\mathbb{P}\left[\mathcal{E}\right] \geq 1 - 1/T$. Then, the regret can be bounded as

$$\begin{aligned}\mathbb{E}\left[R_T^\alpha\right] &\leq \mathbb{P}\left[\mathcal{E}^c\right] T\Delta_{\alpha,\max} + \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in V_0} \mathbb{I}[A_t = i]\Delta_{\alpha,i} \middle| \mathcal{E}\right] \mathbb{P}\left[\mathcal{E}\right] \\ &\leq \Delta_{\alpha,\max} + \mathbb{E}\left[\sum_{i\in V_0} \Delta_{\alpha,i}\mathbb{E}\left[N_i(T)\right] \middle| \mathcal{E}\right] .\end{aligned}$$

Now, observing that $\delta_{\alpha,i} \leq 3\eta_i$ holds under event $\mathcal{E}$, we appeal to Theorem 3 to obtain

$$R_T^\alpha \leq \Delta_{\alpha,\max} + \mathbb{E}\left[\sum_{i\in V_0} \Delta_{\alpha,i}\left(\frac{\mu^* \left(18 + 27\log T\right)}{\delta_{i,\alpha}^2} + 3\right)\middle| \mathcal{E}\right], \tag{5}$$

thus proving the first statement.

Next, we turn to proving the second statement regarding worst-case guarantees. To do this, we appeal to Propositions 1 and 2 that respectively show $\Delta_i \leq 2c^*\delta_i + O(1/n)$ and $\Delta_i \leq c^*\delta_i + o(n)$ for the sub- and supercritical settings, and we use our assumption that n is large enough so that we have $\Delta_i \leq 3c^*\delta_i$ in both settings. Specifically, we observe that $\delta_i = \Theta_n(1)$ by our sparsity assumption and c^* is $\Theta_n(1)$ in the subcritical and $\Theta_n(n)$ supercritical settings, so, for large enough n, the superfluous $O(1/n)$ and $o(n)$ terms can be respectively bounded by $c^*\delta_i$. To proceed, let us fix an arbitrary $\varepsilon > 0$ and split the set V_0 into two subsets: $U(\varepsilon) = \{a \in V_0 : \delta_{\alpha,i} \leq \varepsilon\}$ and $W(\varepsilon) = V_0 \setminus U(\varepsilon)$. Then, under event $\mathcal{E}$, we have

$$\begin{aligned}\sum_{i\in V_0}\Delta_{\alpha,i}\mathbb{E}\left[N_i(T)\right] &= \sum_{i\in U(\varepsilon)}\Delta_{\alpha,i}\mathbb{E}\left[N_i(T)\right]+\sum_{i\in W(\varepsilon)}\Delta_{\alpha,i}\mathbb{E}\left[N_i(T)\right]\\ &\le 3c^*\varepsilon\sum_{i\in U(\varepsilon)}\mathbb{E}\left[N_i(T)\right]+3c^*\sum_{i\in W(\varepsilon)}\delta_{\alpha,i}\left(\frac{\mu^*\left(18+27\log T\right)}{\delta_{\alpha,i}^2}\right)+3|W(\varepsilon)|\Delta_{\alpha,\max}\\ &\le 3c^*\varepsilon T+3c^*\sum_{i\in W(\varepsilon)}\frac{\mu^*\left(18+27\log T\right)}{\delta_{\alpha,i}}+3|V_0|\Delta_{\alpha,\max}\\ &\le 3c^*\left(\varepsilon T+|V_0|\frac{\mu^*\left(18+27\log T\right)}{\varepsilon}\right)+3|V_0|\Delta_{\alpha,\max}\\ &\le 6c^*\sqrt{T|V_0|\mu^*\left(18+27\log T\right)}+3|V_0|\Delta_{\alpha,\max},\end{aligned}$$

where the last step uses the choice $\varepsilon=\sqrt{|V_0|\mu^*\left(18+27\log T\right)/T}$. Plugging in the choice of $|V_0|$ concludes the proof. □

6 Discussion

Here we highlight some features of our results and discuss directions for future work.

Instance-dependent and worst-case regret bounds. Both of our main theorems establish two types of regret bounds. The first set of these bounds are polylogarithmic[1] in the time horizon T, but show strong dependence on the parameters of the distribution of the graphs G_t. Such bounds are usually called *instance-dependent*, and they are typically interesting in the regime where T grows large. However, these bounds become vacuous for finite T as the gap parameters $\delta_{\alpha,i}$ approach zero. This issue is addressed by our second set of guarantees, which offers a bound of $\widetilde{O}\big(c^*\sqrt{|U|\mu^*T}\big)$ for some set $U\subseteq V$ that holds simultaneously for all problem instances without becoming vacuous in any regime. Such bounds are commonly called *worst-case*, and they are typically more valuable when optimizing performance over a fixed horizon T.

Dependence on graph parameters. A notable feature of all our bounds is that they show no explicit dependence on the number of nodes n. This is enabled by our notion of α-quantile regret, which allows us to work with a small subset of the total nodes as our action set. Instead of n, our bounds depend on the size of some suitably chosen set of nodes U, which is of the order $\operatorname{polylog} T/\log(1/(1-\alpha))$. Notice that this gives rise to an interesting tradeoff: choosing smaller values of α inflates the regret bounds, but, in exchange, makes the baseline of the regret definition stronger (thus strengthening the regret notion itself). While the exact tradeoff seems very complicated to quantify in general, it is clear that setting α as the proportion of the smallest community in SBMs strengthens the regret baseline as much as possible.

Of course, having no *explicit* dependence on n does not mean that our bounds are completely independent of the size of the graph. In fact, it is natural to expect that the regret scales with the general magnitude of the rewards. Our bounds precisely achieve such a natural dependence: all our bounds scale linearly with the maximal expected reward c^*, which is of $\Theta_n(1)$ in the subcritical case, but is $\Theta_n(n)$ in the supercritical case.

Tightness of our bounds. In terms of dependence on T, both our instance-dependent and worst-case bounds are near-optimal in their respective settings: even in the simpler stochastic multi-armed bandit problem, the best possible regret bounds are $\Omega_T(\log T)$ and $\Omega_T(\sqrt{T})$ in the respective settings [2, 3, 5]. The optimality of our bounds with respect to other parameters such as c^*, μ^* and n is less clear, but we believe that these factors cannot be improved substantially for the models that we studied in this paper. As for the subproblem of identifying nodes with the highest degrees, we believe that our bounds on the number of suboptimal draws is essentially tight, closely matching the classic lower bounds by Lai and Robbins [20].

[1] Upon first glance, the bound of Theorem 1 may appear to be logarithmic, however, notice that the sum involved in the bound has $\log T$ elements, thus technically resulting in a bound of order $\log^2 T$.

Our assumptions. One may wonder how far our argument connecting local and global influence maximization can be stretched. Clearly, not every random graph model enables establishing such a strong connection. In fact, even within the class of stochastic block models, one can construct an instance (not satisfying Assumption 1) that does not have the property we desire. It is a challenging problem to characterize the class of inhomogeneous random graphs in which maximizing local and global influences are equivalent. Nevertheless, we believe that our techniques can be generalized to maximize global influence with more informative local feedback structures (e.g., working with observations from a slightly broader neighborhood of the chosen nodes).

Finally, let us comment on our condition that the number of vertices n needs to be "sufficiently large". We regard this condition as a technical artifact due to our proofs relying on asymptotic analysis. We expect that the required monotonicity property holds for small values of n under mild conditions. Whenever this is the case, the regret bounds of Theorems 1 and 2 remain valid.

References

[1] E. Abbe. Community detection and stochastic block models: recent developments. *arXiv preprint arXiv:1703.10146*, 2017.

[2] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. ISSN 0097-5397.

[4] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms*, 31(1):3–122, August 2007. ISSN 1042-9832.

[5] S. Bubeck and N. Cesa-Bianchi. *Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems*. Now Publishers Inc, 2012.

[6] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, and G. Stoltz. Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation. *Annals of Statistics*, 41(3):1516–1541, 2013.

[7] A. Carpentier and M. Valko. Revealing graph bandits for maximizing local influence. In *Artificial Intelligence and Statistics*, pages 10–18, 2016.

[8] K. Chaudhuri, Y. Freund, and D. J. Hsu. A parameter-free hedging algorithm. In *Advances in neural information processing systems*, pages 297–305, 2009.

[9] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1029–1038, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1.

[10] W. Chen, L. V. Lakshmanan, and C. Castillo. Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 5(4):1–177, 2013.

[11] W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, pages 151–159, 2013.

[12] A. Chernov and V. Vovk. Prediction with advice of unknown number of experts. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 117–125. AUAI Press, 2010.

[13] F. Chung, F. Chung, F. Graham, and L. Lu. *Complex Graphs and Networks*. Number 107 in Cbms Regional Conference Serie. American Mathematical Society. ISBN 9780821836576.

[14] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[15] A. Garivier and O. Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 359–376, 2011.

[16] W. Huaming. On total progeny of multitype Galton-Watson process and the first passage time of random walk with bounded jumps. *ArXiv e-prints*, September 2012.

[17] S. M. Kakade, A. T. Kalai, and K. Ligett. Playing games with approximation algorithms. *SIAM Journal on Computing*, 39(3):1088–1106, 2009.

[18] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[19] W. M. Koolen and T. Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175, 2015.

[20] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[21] T. L. Lai. Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics*, pages 1091–1114, 1987.

[22] H. Luo and R. E. Schapire. A drifting-games analysis for online learning and applications to boosting. In *Advances in Neural Information Processing Systems*, pages 1368–1376, 2014.

[23] O.-A. Maillard, R. Munos, and G. Stoltz. A finite-time analysis of multi-armed bandits problems with Kullback–Leibler divergences. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 497–514, 2011.

[24] P. Ménard and A. Garivier. A minimax and asymptotically optimal algorithm for stochastic bandits. In *Proceedings of the 28th International Conference on Algorithmic Learning Theory*, pages 223–237, 2017.

[25] M. Streeter and D. Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems*, pages 1577–1584, 2009.

[26] R. van der Hofstad. *Random Graphs and Complex Networks*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2016. doi: 10.1017/9781316779422.

[27] S. Vaswani, L. Lakshmanan, and M. Schmidt. Influence maximization with bandits. *arXiv preprint arXiv:1503.00024*, 2015.

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

[29] Z. Wen, B. Kveton, M. Valko, and S. Vaswani. Online influence maximization under independent cascade model with semi-bandit feedback. In *Advances in Neural Information Processing Systems*, pages 3026–3036, 2017.

A Multi-type branching processes

One of the most important technical tools for analyzing the component structure of random graphs is the theory of *branching processes*, see [26]. Indeed, while the connected components C_a of an inhomogenous random graph $G(n, A)$ have a complicated structure, many of their key properties may be analyzed through the concept of multi-type Galton–Watson processes. Specifically, we use Poisson multi-type Galton–Watson branching processes with n types, parametrized by an $n \times n$ matrix A with strictly posive elements. The branching process tracks the evolution of a set of *individuals* of various types. Starting in round $n = 0$ from a single individual of type i, each further generation in the Galton–Watson process $W_A(i)$ is generated by each individual of each type i producing $X_{k,i} \sim \mathrm{Poi}(A_{i,j})$ new individuals of each type j. Therefore, the size of the offspring of the individual of type i is $\sum_{j=1}^{n} X_{i,j} \sim \mathrm{Poi}(\sum_{j=1}^{n} A_{i,j})$. We also define the vector $b \in \mathbb{R}^n$ with coordinates $b_i = \mathbb{E}\left[\sum_{j=1}^{n} X_{i,j}\right] = \sum_{j=1}^{n} A_{i,j}$, $i = 1, \dots, n$.

Our analysis below makes use of the following quantities associated with the multi-type branching process:

1. $Z_n(i)$ is the number of individuals in generation n of $W_A(i)$ (where $Z_0(i) = 1$);
2. $X(i)$ is the *total progeny*, that is, the total number of individuals generated by $W_A(i)$ and its expectation is denoted by $x_i = \mathbb{E}[X(i)]$;
3. ρ_i is the *probability of survival*, that is, the probability that $X(i)$ is infinite.

We finally define a non-linear operator $\Phi_A : \mathbb{R}^n \to \mathbb{R}^n$ that plays a central role in our analysis: for a vector $f \in \mathbb{R}^n$, define each coordinate of $\Phi_A(f)$ as

$$\left(\Phi_A(f)\right)_j = 1 - e^{-(Af)_j}, \quad j = 1, \dots, n\,, \tag{6}$$

where $(Af)_j$ denotes the j-th coordinate of Af. Abusing notation, we use the shorthand form $\Phi_A(f) = 1 - e^{-Af}$. Clearly, if f has nonnegative components, then $(\Phi_A(f))_j \in [0, 1]$ for all j.

Bollobás, Janson, and Riordan [4] establish a connection between the sizes of connected components of IRG, the survival probability of a branching process $W_A(i)$, and the norm of the matrix A.

As shown in [4], the operator Φ_A can be directly used for characterizing the probability ρ_i of survival of the process $W_A(i)$ for all i. By their Theorem 6.2, the vector $\rho = (\rho_1, \dots, \rho_n)$ is one of the solutions of the non-linear fixed-point equation $\Phi_A(f) = f$. Furthermore, if the largest eigenvalue of the matrix A satisfies $\lambda_{\max}(A) < 1$, then $\rho_i = 0$ for all $i = 1, \dots, n$. On the other hand, $\lambda_{max} > 1$ implies that the vector ρ is the *maximal* fixed point of the operator Φ_A [4, Lemma 5.8.] also implies that when $\lambda_{max} > 1$, all components of ρ are positive.

B The proof of Proposition 1

The proof consists of the following steps:

- proving that $c_i - c_j = x_i - x_j + O(1/n)$ (Lemmas 3, 4),
- proving that $x_i > x_j$ implies $b_i > b_j$ (Lemma 1, 2),
- observing that $b_i = \mu_i + O(1/n)$.

These facts together lead to Proposition 1, given that n is large enough to suppress the effects of the residual terms. Before stating and proving the lemmas, we state some useful technical tools. Since we suppose that $G(n, A)$ is subcritical, we have $\mathbb{P}[X(i) = \infty] = 0$ and $x_i = \mathbb{E}X(i)$ is finite. First observe that the vector x of expected total progenies satisfies the system of linear equations

$$x = e + Ax\,,$$

where e is the vector with $e_i = 1$ for all i. Notice that, by its definition, the vector b can be succinctly written as $b = Ae$.

Armed with this notation, we can analyze the relation between b_i and x_i in a straightforward way:

Lemma 1 (Coordinate order for mean of the total progeny in the SBM)**.** *Assume that $G(n, \alpha, K)$ is subcritical and that $K_{m\ell} = k > 0$ holds for all $m \neq \ell$. If two coordinates of b are such that $b_i > b_j$, then we have $x_i > x_j$, and $x_i - x_j \leq 2x^* (b_i - b_j)$.*

Proof. For and SBM with S blocks, the system of equations $x = e + Ax$ can be equivalently written as $x' = e + Mx'$, for $M = K\text{diag}(\alpha) \in \mathbb{R}^{S \times S}$, and $x' \in \mathbb{R}^S$, with x'_m now standing for the expected total progeny associated with any node of type m. Similarly, we define b'_m as the expected degree of any node of type m. Notice that the system of equations $x' = e + Mx'$ satisfied by x' can be rewritten as $(I - M)x' = e$, where I is the $S \times S$ identity matrix. By exploiting our assumption on the matrix K and defining $\gamma_m = K_{m,m} - k$, this can be further rewritten as

$$\left(\begin{pmatrix} 1 - \alpha_1\gamma_1 & & \\ & \ddots & \\ & & 1 - \alpha_S\gamma_S \end{pmatrix} - k \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_S \\ \alpha_1 & \alpha_2 & \cdots & \alpha_S \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_S \end{pmatrix}\right) x' = e,$$

which means that for any m, x'_m satisfies

$$x'_m = \frac{1 + k(\alpha^\top x')}{1 - \alpha_m\gamma_m}.$$

Also observe that

$$b'_m = k(\alpha^T \bar{1}) + \alpha_m\gamma_m,$$

so, for any pair of types m and ℓ, we have

$$x'_m - x'_\ell = \frac{(1 + k(\alpha^\top x'))(\alpha_m\gamma_m - \alpha_\ell\gamma_\ell)}{(1 - \alpha_m\gamma_m)(1 - \alpha_\ell\gamma_\ell)},$$

which proves the first statement.

To prove the second statement, observe that for any pair ℓ and m of communities, we have either $\alpha_m \leq \frac{1}{2}$ or $\alpha_\ell \leq \frac{1}{2}$ (otherwise we would have $\alpha_m + \alpha_\ell > 1$). To proceed, let ℓ and m be such that $x'_m \geq x'_\ell$, and let us study the case $\alpha_\ell \leq \frac{1}{2}$ first. Here, we get

$$\begin{aligned} x'_m - x'_\ell &\leq \frac{(1 + k(\alpha^\top x'))(\alpha_m\gamma_m - \alpha_\ell\gamma_\ell)}{(1 - \alpha_m\gamma_m)(1 - \alpha_\ell\gamma_\ell)} = \frac{(\alpha_m\gamma_m - \alpha_\ell\gamma_\ell)}{(1 - \alpha_\ell\gamma_\ell)} x'_m \\ &\leq \frac{(\alpha_m\gamma_m - \alpha_\ell\gamma_\ell)}{(1 - \gamma_\ell/2)} x'_m \leq 2x'_m(b'_m - b'_\ell). \end{aligned}$$

In the other case where $\alpha_m \leq \frac{1}{2}$, we can similarly obtain

$$x'_m - x'_\ell \leq 2x'_\ell(b'_m - b'_\ell) \leq 2x'_m(b'_m - b'_\ell).$$

This concludes the proof. □

Lemma 2 (Coordinate order for mean of the total progeny in the Chung–Lu model)**.** *Assume that $G(n, w)$ is subcritical. If two coordinates of b are such that $b_i > b_j$, then we have $x_i > x_j$ and $x_i - x_j \leq x^*(b_i - b_j)$.*

Proof. For the system of equations $x = e + Ax$ the coordinates x_i have the form

$$x_i = 1 + \frac{1}{n} \cdot w_i \left(\sum_{j=1}^{n} w_j x_j \right),$$

which implies that $w_i \geq w_j$ holds if and only if $x_i \geq x_j$. This observation implies for $x^* = \max_i x_i$

$$x_i - x_j \leq \frac{1}{n} \cdot (w_i - w_j) \left(\sum_{j=1}^{n} w_j \right) x^* = (b_i - b_j)\, x^*,$$

thus concluding the proof. □

The next two lemmas establish the relationship between the expected component size c_i of vertex i and the expected total progeny x_i of the multi-type branching process seeded at vertex i.

Lemma 3. *For any i, the mean of the connected component associated with type i is bounded by the mean of the total progeny: $c_i \leq x_i$.*

Proof. The proof of the lemma uses the concept of *stochastic dominance* between random variables. We say that the random variable X is *stochastically dominated* by the random variable Y when, for every $x \in \mathbb{R}$, the inequality $\mathbb{P}[X \leq x] \geq \mathbb{P}[Y \leq x]$ holds. We denote this by $X \preceq Y$.

Now fix an arbitrary $i \in [n]$ and let $Y_{i,1}, Y_{i,2}, \dots, Y_{i,n}$ be independent Bernoulli random variables with respective parameters $(\overline{A}_{i,1}/n, \overline{A}_{i,2}/n, \dots, \overline{A}_{i,i}/n, \dots, \overline{A}_{i,n}/n)$. Consider a multitype binomial branching process where the individual of type i produce $Y_{i,j}$ individuals of type j, and let $X_{\text{Ber}}(i)$ denote its total progeny when started from an individual of type i. Recalling the Poisson branching process defined in Section A with offspring-distributions $X_{i,j}$, we can show $X_{\text{Ber}}(i) \preceq X(i)$ using the relation $Y_{i,j} \preceq X_{i,j}$.

Considering a node a of type i, we can use Theorem 4.2 of [26] to bound the size of the the connected component C_a as $|C_a| \preceq X_{\text{Ber}}(i)$, which implies by transitivity of $\preceq$ that $|C_{a_i}| \preceq X(i)$. The proof is concluded by appealing to Theorem 2.15 of [26] that shows that stochastic domination implies an ordering of the means. □

Next we upper bound the surplus that appears in the domination by the branching process:

Lemma 4. $x_i - c_i = O(\frac{1}{n})$.

Proof. Consider an exploration process in the realization G_t of a random graph $G(n, A)$ starting from a node a of type i. The process explores the nodes in a sequential way, by first visiting the neighboring nodes of the initial node, then moving on to the neighbors of the neighbors, and so on. The process stops after having explored the whole connected component C_a.

Also the Bernoulli multitype branching process $W_{\text{Ber}}(i)$ with the set of parameters $\mathbb{B}_j$, for $j \in [n]$, where parameters $\mathbb{B}$ correspond to the $G(n, A)$. Denote the tree naturally defined by the exploration process of the connected component by $\mathcal{T}$, and also the tree defined by the analogously defined Poisson exploration process of the branching process tree by $\mathcal{T}_{\text{Poi}}$. The proof relies on the fact that the total number of nodes visited by the exploration process can be upper bounded by the total progeny of the corresponding branching process [26, Section 4.1].

In order to estimate the difference $|\mathcal{T}| - |C_a|$, note that for each step of the exploration process, the number of nodes that have been already explored can be upper bounded by $|C_a|$ so we have $|C_a| \leq |\mathcal{T}|$. Let $\mathcal{S}$ be a set of nodes counted more than once. We call $|\mathcal{S}|$ the *surplus* whose expectation may be written as follows:

$$\mathbb{E}[|\mathcal{S}|] = \mathbb{E}\left[\sum_{v \in V} \mathbb{I}\{v \in \mathcal{S}\}\right] = \sum_{k=1}^{\infty} \mathbb{P}[|\mathcal{T}| = k] \sum_{v \in C_a} \mathbb{E}[\mathbb{I}\{v \in \mathcal{S}\} | \, |\mathcal{T}| = k] \, .$$

Define $\overline{A}_{\max} = \max_{i,j} \overline{A}_{i,j}$ be the maximal element of the matrix $\overline{A}$.

Then, by the union bound, the probability of an arbitrary node a' is counted more than once can be upper bounded as

$$\mathbb{P}[a' \in \mathcal{S}] \leq \frac{\overline{A}_{\max}|C_a|}{n} \, ,$$

and we also have

$$\mathbb{E}[\mathbb{I}\{a' \in \mathcal{S}\} | \, |\mathcal{T}| = k] \leq \frac{\overline{A}_{\max} k}{n} \, .$$

Since $|C_a| \leq |\mathcal{T}|$, we may upper bound the sum as

$$\sum_{v \in C_a} \mathbb{E}[\mathbb{I}\{v \in \mathcal{S}\} | \, |\mathcal{T}| = k] \leq \frac{\overline{A}_{\max} k^2}{n} \, .$$

Using our expression for $\mathbb{E}|\mathcal{S}|$, we get

$$\mathbb{E}|\mathcal{S}| \leq \sum_{k=1}^{\infty} \mathbb{P}\left[|\mathcal{T}| = k\right] \frac{\overline{A}_{\max} k^2}{n} = \frac{\overline{A}_{\max} \mathbb{E}|\mathcal{T}|^2}{n} \ .$$

Now we notice that, by the Le Cam's theorem, the total variation distance between the sum of Bernoulli distributed random variables with parameters $(\overline{A}_{i,1}/n, \dots, \overline{A}_{i,n/n})$ and the Poisson distribution Poi$(\sum_{j=1}^{n} \overline{A}_{i,j}/n)$ is at most $2(\sum_{j=1}^{n} \overline{A}_{i,j}^2)/n$. Using this fact and that the moments of the total progeny do not scale with n (cf. Theorem 1 of 16), we obtain the result as

$$\mathbb{E}|\mathcal{S}| \leq \frac{\overline{A}_{\max} \mathbb{E}|\mathcal{T}|^2}{n} \leq \frac{\overline{A}_{\max} \mathbb{E}|\mathcal{T}_{\text{Poi}}|^2}{n} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right) \ .$$

□

C The proof of Proposition 2

The proof relies on some known properties of the largest connected component in $G(n, A)$ in the supercritical regime. We denote the largest and second-largest connected components of G_t by $C_1(G_t)$ and $C_2(G_t)$, respectively. Recall that the survival probability of the branching process $W_A(i)$ is denoted as ρ_i. The following properties are proved by [4]:

- If $G(n, A)$ is supercritical, then, with high probability, $C_1 = \Theta(n)$;
- $C_1(G_n)/n \to \sum_{i \in S} \alpha_i \rho_i$ in probability;
- $C_2(G_n) = o(n)$ with high probability.

The expected size of the connected component of a vertex i is

$$c_i = \rho_i \mathbb{E}\left[C_1(G)\right] + o(n) \ . \tag{7}$$

Proposition 2 follows from the following lemmas for the SBM and the Chung–Lu models.

Lemma 5 (Coordinate order preserving in the SBM)**.** *Assume the conditions of Proposition 2 and let $i_* =$ arg max$_i$ b_i. Let $a = (a_1, \dots, a_S)$ be such that $a_j \in [0, a_{i_*}]$ for all j. Then $(\Phi_A(a))_{i_*} \geq (\Phi_A(a))_j$.*

Proof. Let us fix two arbitrary indices i and i'. By the definition of Φ_M, we have

$$(\Phi_A(a))_i = 1 - e^{-((\sum_{j \neq i} \alpha_j a_j) k + \alpha_i k_{i,i} a_i)} \ ,$$
$$(\Phi_A(a))_{i'} = 1 - e^{-((\sum_{j \neq i'} \alpha_j a_j) k + \alpha_{i'} k_{i',i'} a_{i'})} \ .$$

Notice that if i and i' satisfy

$$\left(\sum_{j \neq i} \alpha_j a_j\right) k + \alpha_i k_{i,i} a_i \geq \left(\sum_{j \neq i'} \alpha_j a_j\right) k + \alpha_{i'} k_{i',i'} a_{i'},$$

we have $(\Phi_A(a))_i \geq (\Phi_A(a))_{i'}$. Now, using the facts that

- $\sum_{j \neq i} \alpha_j a_j - \sum_{j \neq i'} \alpha_j a_j = \alpha_{i'} a_{i'} - \alpha_i a_i$,
- $\alpha_i k_{i,i} \geq \alpha_i k$,
- $\alpha_i k_{i,i} + \alpha_{i'} k \geq \alpha_{i'} k_{i',i'} + \alpha_i k$ and
- $a_i - a_{i'} \geq 0$,

we can verify that

$$\begin{aligned} &\alpha_i k_{i,i} a_i + \alpha_{i'} k a_{i'} - \alpha_i k a_i - \alpha_{i'} k_{i',i'} a_{i'} \\ &\quad = \ (\alpha_i k_{i,i} + \alpha_{i'} k) a_{i'} + (a_i - a_{i'}) \alpha_i k_{i,i} - (\alpha_{i'} k_{i',i'} + \alpha_i k) a_{i'} - (a_i - a_{i'}) \alpha_i k \geq 0, \end{aligned}$$

thus proving the lemma. □

Lemma 6 (Order of coordinates of eigenvector in the SBM)**.** *Let a be the eigenvector corresponding to the largest eigenvalue λ of the matrix $M = K diag(\alpha)$. Then if $i_* = \arg\max_m b_m$, we have $a_{i_*} \geq a_j$ for $j \neq i_*$.*

Proof. If a is an eigenvector of M, then for coordinates i, i':

$$\begin{cases} \left(\sum_{j\neq i} \alpha_j a_j\right) k + \alpha_i k_{i,i} a_i = \lambda a_i, \\ \left(\sum_{j\neq i'} \alpha_j a_j\right) k + \alpha_i k_{i',i'} a_{i'} = \lambda a_{i'} \end{cases}$$

By the Perron-Frobenius theorem and our conditions on matrix M, λ is a real number larger than one. Denote $C = k \sum_{j\neq i, j\neq i'} \alpha_j a_j$, $x = a_i$, $y = a_{i'}$, $a = \alpha_i k_{i,i}$, $b = \alpha_{i'} k$, $c = \alpha_i k$, $d = \alpha_{i'} k_{i',i'}$. Then,

$$\begin{cases} C + ax + by = \lambda x, \\ C + cx + dy = \lambda y \end{cases} \tag{8}$$

Let $r = 1 + \epsilon$ be such that $y = rx = (1+\epsilon)x$. Then

$$\begin{cases} \frac{C}{x} + a + b + b\epsilon = \lambda, \\ \frac{C}{x} + c + d + d\epsilon = \lambda + \lambda\epsilon \end{cases}$$

and therefore

$$\frac{C}{x} + c + d + d\epsilon = \frac{C}{x} + a + b + b\epsilon + \lambda\epsilon \ .$$

Rearranging the terms and using the fact that $a + b \geq c + d$, we have

$$0 \leq (a+b) - (c+d) = (d - b - \lambda)\epsilon \ .$$

Since $k_{i,i} \geq k$, we have $\alpha_i k_{i,i} \geq \alpha_i k$ and $a \geq c$.

We consider two cases separately: First, if $b \geq d$, we have $d - b - \lambda < 0$, which implies $\epsilon < 0$ and $y < x$, therefore proving $a_i > a_{i'}$ for this case. In the case when $b < d$, we have $a + b \geq c + d$ and $\frac{d-b}{a-c} \leq 1$. Subtracting the two equalities of the linear system 8, we get

$$\lambda(1 - r) = (a - c)\left(1 - \frac{d-b}{a-c} r\right) \ .$$

Now, since $\frac{d-b}{a-c} \leq 1$, we have $\lambda \geq a - c$, which implies $\lambda \geq d - b$ and $d - b - \lambda \leq 0$, thus leading to $\epsilon \leq 0$ and $y \leq x$, therefore proving $a_i \geq a_{i'}$ for this case. □

Lemma 7 (Order of coordinates of eigenvector in the Chung–Lu model)**.** *Let a be the eigenvector corresponding to the largest eigenvalue λ of the matrix A. Then if $i_* = \arg\max_m b_m$, we have $a_{i_*} \geq a_j$ for $j \neq i_*$.*

Proof. It is easy to see that the only eigenvector of A corresponding to a non-zero eigenvalue is $a = w$ with $\lambda_{max} = w^\top w / n$:

$$Aw = \frac{1}{n} \cdot (ww^\top)w = \frac{w^\top w}{n} \cdot w.$$

The proof is concluded by observing that the maximum coordinate of the vector b corresponds to the maximum coordinate of w, due to the equality

$$b_i = \frac{1}{n} \cdot w_i \sum_{j=1}^{n} w_j.$$

□

Lemma 8 (Coordinate order preserving in the Chung–Lu model)**.** *Assume the conditions of Proposition 2 and let $i_* = \arg\max_i b_i$. Let $a = (a_1, \ldots, a_n)$ be such that $a_j \in [0, a_{i_*}]$ for all j. Then $(\Phi_A(a))_{i_*} \geq (\Phi_A(a))_j$.*

Proof. Let us fix two arbitrary indices i and i'. By the definition of Φ_A, we have

$$(\Phi_A(a))_i = 1 - e^{-w_i(\sum_{j=1}^n w_j a_j)} ,$$
$$(\Phi_A(a))_{i'} = 1 - e^{-w_{i'}(\sum_{j=1}^n w_j a_j)} .$$

Then we have $(\Phi_A(a))_i \geq (\Phi_A(a))_{i'}$ thus proving the lemma. □

We finally study the maximal fixed point of the operator Φ_A, keeping in mind this fixed point is exactly the survival-probability vector ρ of the multi-type Galton–Watson branching process [4]. By Lemma 5.9 of [4], this is the unique fixed point satisfying $\rho_i > 0$ for all i. The following lemma shows that ρ_i takes its maximum at $i_* = \arg\max_i b_i$, concluding the proof of Proposition 2.

Lemma 9 (Fixed point coordinate domination)**.** *Let ρ be the unique non-zero fixed point of Φ_A, and let $i_* = \arg\max_i b_i$. Then, $\rho_{i_*} \geq \rho_j$ and $\rho_{i_*} - \rho_j \leq \rho^* (b_{i_*} - b_j)$ holds for all $j \neq i_*$.*

Proof. Letting a be the eigenvector of A that corresponds to the largest eigenvalue λ, Lemma 7, 6 guarantees $a_{i_*} \geq a_j$ for $j \neq i^*$. Let $\epsilon > 0$ be such that $\epsilon \leq \frac{1-1/\lambda}{a^*}$, where $a^* = \max_{i=1,\ldots,S} a_i$. Then by Lemma 5.13 of [4], $\Phi_M(\epsilon a) \geq \epsilon a$ holds elementwise for the two vectors.

Since the coordinates of the vector ϵa are positive, we can appeal to Lemma 5.12 of [4] to show that iterative application of Φ_A converges to the fixed point ρ: letting Φ_A^m be the operator obtained by iterative application of Φ_A for m times, we have $\lim_{m\to\infty} \Phi_A^m(\epsilon a) = \rho$, where ρ satisfies $\rho \geq \epsilon a \geq 0$ and $\Phi_A(\rho) = \rho > 0$. By the respective Lemmas 7, 6 we have $\rho_{i_*} \geq \rho_j$, for $i_* \neq j$ for both the SBM and the Chung–Lu models, proving the first statement.

The second statement can now be proven directly as

$$\begin{aligned} \rho_{i_*} - \rho_i = e^{-(A\rho)_j} - e^{-(A\rho)_{i*}} &= e^{-\sum_j^n A_{i*j}\rho_j} - e^{-\sum_j^n A_{ij}\rho_j} \\ &= \quad e^{-\sum_j^n A_{i*j}\rho_j}(1 - e^{-\sum_j^n A_{ij}\rho_j - A_{i*j}\rho_j}) \leq e^{-\sum_j^n A_{i*j}\rho_j} \left(\sum_j^n (A_{i_*j} - A_{ij})\rho_{i_*} \right) \\ &\leq \quad \rho^*(b_{i_*} - b_i), \end{aligned}$$

where the first inequality uses the relation $1 - e^{-z} \leq z$ that holds for all $z \in \mathbb{R}$, and the last step uses the fact that $A\rho$ has positive elements. □

D The proof of Theorem 3

Before delving into the proof, we introduce some useful notation. We start by defining $Y_{a,1}, Y_{a,2}, \ldots, Y_{a,n}$ as independent Bernoulli random variables with respective parameters $\mathbb{B} = (A_{a,1}, A_{a,2}, \ldots, A_{a,n})$, and noticing that the degree $Y_{t,a}$ can be written as a sum $Y_a = \sum_{i\neq a} Y_{a,i}$. The following lemma, used several times in our proofs, relates this quantity to a Poisson distribution with the same mean.

Lemma 10. *Let $i \in [S]$ and let $Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n}$ be independent Bernoulli random variables with respective parameters $k_{i,1}/n, k_{i,2}/n, \ldots, k_{i,n}/n$, and let X_i be a Poisson random variable with parameter $\mu_i = \sum_{j\neq i} k_{i,j}/n$. Defining $Y_i = \sum_{j\neq i} Y_{i,j}$, we have $\mathbb{E}\left[e^{sY_i}\right] \leq \mathbb{E}\left[e^{sX_i}\right]$ for all $s \in \mathbb{R}$.*

Proof. Fix an arbitrary $s \in \mathbb{R}$ and $i \in [n]$. By direct calculations, we obtain

$$\mathbb{E}e^{sY_i} = \prod_{j=1}^n \left(\mathbb{E}e^{sY_{i,j}}\right) \leq \prod_{j=1}^n \left(1 + \frac{k_{i,j}}{n}(e^s - 1)\right) \leq \prod_{j=1}^n \exp\left((k_{i,j}/n)\cdot(e^s - 1)\right),$$

where the last step follows from the elementary inequality $1 + x \leq e^x$ that holds for all $x \in \mathbb{R}$. The proof is concluded by observing that $\mathbb{E}e^{sX_i} = \exp\left(\mu\left(e^s - 1\right)\right)$ and using the definition of μ. □

For simplicity, we also introduce the notation $\psi_{\mathbb{B}}(s) = \log \mathbb{E}e^{sY_i}$ and $\phi_\lambda(s) = \log \mathbb{E}e^{sX_i} = \lambda(e^s - 1)$. The proof below repeatedly refers to the Fenchel conjugate of ϕ_λ defined as

$$\phi_\lambda^*(z) = \sup_{s\in\mathbb{R}}\{sz - \phi(s)\} = z\log\left(\frac{z}{\lambda}\right) + \lambda - z$$

for all $z \in \mathbb{R}$. Finally, we define $d(\mu, \mu') = \mu' - \mu + \mu \log\left(\frac{\mu}{\mu'}\right)$ for all $\mu, \mu' > 0$, noting that $\phi^*_\lambda(z) = d(z, \lambda)$.

As for the actual proof of the theorem, the statement is proven in four steps. Within this proof, we refer to nodes as *arms* and use K to denote the size of V_0. We use the notation $f(t) = 3 \log t$.

Step 1. We begin by rewriting the expected number of draws $\mathbb{E}N_a$ for any suboptimal arm a as

$$\mathbb{E}N_a = \mathbb{E}\left[\sum_{t=K}^{T-1} \mathbb{I}\{A_{t+1} = a\}\right] = \sum_{t=K}^{T-1} \mathbb{P}\{A_{t+1} = a\}.$$

By definition of our algorithm, at rounds $t > K$, we have $A_{t+1} = a$ only if $U_{a(t)} > U_{a^*(t)}$. This leads to the decomposition:

$$\begin{aligned} \{A_{t+1} = a\} &\subseteq \{\mu^* \geq U_{a^*}(t)\} \cup \{\mu^* < U_{a^*}(t) \text{ and } A_{t+1} = a\} \\ &\subseteq \{\mu^* \geq U_{a^*}(t)\} \cup \{\mu^* < U_a(t) \text{ and } A_{t+1} = a\} \end{aligned}$$

Steps 2 and 3 are devoted to bounding the probability of the two events above.

Step 2. Here we aim to upper bound

$$\sum_{t=K}^{T-1} \mathbb{P}\left[\mu^* \geq U_{a^*}(t)\right]. \tag{9}$$

Note, that $\{U_{a^*}(t) \leq \mu^*\} = \{\hat{\mu}_{a^*}(t) \leq U_{a^*}(t) \leq \mu^*\}$. Since $d(\mu, \mu') = \mu' - \mu + \mu \log(\frac{\mu}{\mu'})$ is non-decreasing in its second argument on $[\mu, +\infty)$, and by definition of $U_{a^*} = \sup\{\mu : d(\hat{\mu}_{a^*}(t), \mu) \leq \frac{f(t)}{N_{a^*(t)}}\}$ we have

$$\{\mu^* \geq U_{a^*}(t)\} \subseteq \left\{\hat{\mu}_{a^*}(t) \leq U_{a^*}(t) \leq \mu^* \text{ and } d(\hat{\mu}_{a^*}(t), \mu^*) \geq \frac{f(t)}{N_{a^*}(t)}\right\},$$

Taking a union bound over the possible values of $N_{a^*}(t)$ yields

$$\{\mu^* \geq U_{a^*}(t)\} \subseteq \bigcup_{n=1}^{t-K+1} \left\{\mu^* \geq \hat{\mu}_{a^*,n} \text{ and } d(\hat{\mu}_{a^*,n}, \mu^*) \geq \frac{f(t)}{n}\right\} = \bigcup_{n=1}^{t-K+1} D_n(t),$$

where the event $D_n(t)$ is defined through the last step. Since $d(\mu, \mu^*)$ is decreasing and continuous in its first argument on $[0, \mu^*)$, either $d(\hat{\mu}_{a^*,n}, \mu^*) < \frac{f(t)}{n}$ on this interval and $D_n(t)$ is the empty set, or there exists a unique $z_n \in [0, \mu^*)$ such that $d(z_n, \mu^*) = \frac{f(t)}{n}$. Thus, we have

$$\bigcup_{n=1}^{t-K+1} D_n(t) \subseteq \bigcup_{n=1}^{t-K+1} \{\hat{\mu}_{a^*,n} \leq z_n\}.$$

For $\lambda < 0$, let us define $\psi(\lambda)$ as the cumulant-generating function of the sum of binomials with parameters $\mathbb{B}$, and let $\phi(\lambda)$ be the cumulant-generating function of a Poisson random variable with parameter μ^*. With this notation, we have for *any* $\lambda < 0$ that

$$\begin{aligned} \mathbb{P}\left[\hat{\mu}_{a^*,n} \leq z_n\right] &= \mathbb{P}\left[\exp(\lambda \hat{\mu}_{a^*,n}) \geq \exp(\lambda z_n)\right] \\ &= \mathbb{P}\left[\exp\left(\lambda \sum_{i=1}^{n} Y_{a^*,i} - n\psi(\lambda)\right) \geq \exp(n\lambda z_n - n\psi(\lambda))\right] \\ &\leq \left(\frac{\mathbb{E}e^{\lambda Y_{a^*,1}}}{e^{\psi(\lambda)}}\right)^n e^{-n(\lambda z_n - \psi(\lambda))} \leq e^{-n(\lambda z_n - \psi(\lambda))}, \end{aligned}$$

where the last step uses the definition of $\psi(\lambda)$. Now fixing $\lambda^* = \arg\max_\lambda\{\lambda z_n - \phi(\lambda)\} = \log(z_n/\mu^*) < 0$, we get by Lemma 10 that

$$e^{-n(\lambda^* z_n - \psi(\lambda^*))} \leq e^{-n(\lambda^* z_n - \phi(\lambda^*))} = e^{-n\phi^*_{\mu^*}(z_n)} = e^{-nd(z_n,\mu^*)}.$$

In view of the definition of z_n and $f(t)$, this gives the bound

$$e^{-nd(z_n,\mu^*)} = e^{-f(t)} = \frac{1}{t^3},$$

which leads to

$$\sum_{t=K}^{T-1} \mathbb{P}\left[\mu^* \geq U_{a^*}(t)\right] \leq \sum_{t=K}^{T-1} \sum_{n=1}^{t-K+1} \frac{1}{t^3} < 2,$$

thus concluding this step.

Step 3. In this step, we borrow some ideas by [24, Proof Theorem 2, step 2] to upper-bound the sum

$$B = \sum_{t=K}^{T-1} \mathbb{P}\left[\mu^* < U_a(t) \text{ and } A_{t+1} = a\right]. \tag{10}$$

Writing $\eta = \eta_a = \{\mu^* - \mu_a\}/3$ for ease of notation, we have

$$\begin{aligned}\{\mu^* < U_a(t) \text{ and } A_{t+1} = a\} &\subseteq \{\mu^* - \eta < U_a(t) \text{ and } A_{t+1} = a\} \\ &\subseteq \{d(\hat{\mu}_a(t), \mu^* - \eta) \leq f(t)/N_a(t) \text{ and } A_{t+1} = a\}.\end{aligned}$$

Thus, we have

$$\begin{aligned} B &\leq \sum_{t=K}^{T-1} \mathbb{P}\left[d(\hat{\mu}_a(t), \mu^* - \eta) \leq f(t)/N_a(t) \text{ and } A_{t+1} = a\right] \\ &\leq \sum_{n=1}^{T} \mathbb{P}\left[d(\hat{\mu}_{a,n}, \mu^* - \eta) \leq f(T)/n\right] \end{aligned}$$

Defining the integer $n(\eta)$ as

$$n(\eta) = \left\lceil \frac{f(T)}{d(\mu_a + \eta, \mu^* - \eta)} \right\rceil,$$

we have $f(T)/n \leq d(\mu_a + \eta, \mu^* - \eta)$ for all $n \geq n(\eta)$. Thus, we may further upper-bound B as

$$\begin{aligned} B &\leq n(\eta) - 1 + \sum_{n=n(\eta)}^{T} \mathbb{P}\left[d(\hat{\mu}_{a,n}, \mu^* - \eta) \leq f(T)/n\right] \\ &\leq \frac{f(T)}{d(\mu_a + \eta, \mu^* - \eta)} + \sum_{n=n(\eta)}^{T} \mathbb{P}\left[d(\hat{\mu}_{a,n}, \mu^* - \eta) \leq d(\mu_a + \eta, \mu^* - \eta)\right]. \end{aligned}$$

By definition of η, we have

$$\{\hat{\mu}_{a,n}, \mu^* - \eta) \leq d(\mu_a + \eta, \mu^* - \eta)\} \subseteq \{\hat{\mu}_{a,n} \geq \mu_a + \eta\},$$

which implies

$$\sum_{n=n(\eta)}^{T} \mathbb{P}\left[d(\hat{\mu}_{a,n}, \mu^* - \eta) \leq d(\mu_a + \eta, \mu^* - \eta)\right] \leq \sum_{n=n(\eta)}^{T} \mathbb{P}\left[\hat{\mu}_{a,n} \geq \mu_a + \eta\right].$$

By an argument analogous to the one used in the previous step, we get for a well-chosen λ that

$$\begin{aligned}\sum_{n=n(\eta)}^{T} \mathbb{P}\left[\hat{\mu}_{a,n} \geq \mu_a + \eta\right] &\leq \mathbb{P}\left[\exp(\lambda\hat{\mu}_{a,n}) \geq \exp(\lambda(\mu_a+\eta))\right] \\ &= \sum_{n=n(\eta)}^{T} \mathbb{P}\left[\exp(\lambda\sum_{i=1}^{n} Y_{a,i} - n\psi(\lambda)) \geq \exp(n\lambda(\mu_a+\eta) - n\psi(\lambda))\right] \\ &\leq \sum_{n=n(\eta)}^{T} \left(\frac{\mathbb{E}\left[e^{\lambda Y_{a,i}}\right]}{e^{\psi(\lambda)}}\right)^{n} e^{-n(\lambda(\mu_a+\eta)-\psi(\lambda))} \\ &\leq \sum_{n=n(\eta)}^{T} e^{-n(\lambda(\mu_a+\eta)-\phi(\lambda))} = \sum_{n=n(\eta)}^{T} e^{-nd(\mu_a+\eta,\mu_a)} \\ &\leq \sum_{n=n(\eta)}^{\infty} e^{-nd(\mu_a+\eta,\mu_a)} \leq \frac{1}{e^{d(\mu_a+\eta,\mu_a)}-1} \leq \frac{1}{d(\mu_a+\eta,\mu_a)},\end{aligned}$$

where the last step uses the elementary inequality $1 + x \leq e^x$ that holds for all $x \in \mathbb{R}$.

Step 4. Putting together the results from the first three steps, we get

$$\mathbb{E}N_a \leq 3 + \frac{1}{d(\mu_a+\eta,\mu_a)} + \frac{3\log T}{d(\mu_a+\eta,\mu^*-\eta)}.$$

We conclude by taking a second-order Taylor-expansion of $d(\mu_a+\eta,\mu_a)$ in η to obtain for some $\eta' \in [0,\eta]$ that

$$d(\mu_a+\eta,\mu_a) = \frac{\eta^2}{2(\mu_a+\eta')} \geq \frac{\eta^2}{2(\mu_a+\eta)}.$$

Taking into account the definition of η, we get

$$\frac{1}{d(\mu_a+\eta,\mu_a)} \leq \frac{2\mu^*}{\eta^2}.$$

An identical argument can be used to bound $\left(d(\mu_a+\eta,\mu^*-\eta)\right)^{-1} \leq 2\mu^*/\eta^2$. □

E The proof of Theorem 2

We start by assuming that $\alpha < 1/2$. Also notice that for a uniformly sampled set of nodes U, the probability of U not containing a vertex from V^*_α is bounded as

$$\mathbb{P}\left[U \cap V^*_\alpha = \emptyset\right] \leq (1-\alpha)^{|U|}.$$

By the definition of V_k, this gives that the probability of not having sampled a node from V^*_α in period k of the algorithm is bounded as

$$\mathbb{P}\left[V_k \cap V^*_\alpha = \emptyset\right] \leq (1-\alpha)^{|V_k|} \leq \beta^{-k}.$$

For each period k, the expected regret can bounded as the weighted sum of two terms: the expected regret of d-UCB (V_k) in period k whenever $V_k \cap V^*_\alpha$ is not empty, and the trivial bound $\Delta_{\alpha,\max}\beta^k$ in the complementary case. Using the above bound on the probability of this event and appealing to Theorem 3 to bound the regret of d-UCB (V_k), we can bound the expected regret as

$$\begin{aligned}\mathbb{E}\left[R^\alpha_T\right] &\leq \sum_{k=1}^{k_{\max}} \left(\beta^k \frac{1}{\beta^k}\Delta_{\alpha,\max} + \sum_{i\in V_k} \Delta_{\alpha,i}\left(\frac{\mu^*\left(2+3\log\beta^k\right)}{\delta^2_{\alpha,i}} + 3\right)\right) \\ &\leq k_{\max}\Delta_{\alpha,\max} + \sum_{k=1}^{k_{\max}}\left(\sum_{i\in V_k}\Delta_{\alpha,i}\left(\frac{\mu^*\left(2+3k\log\beta\right)}{\delta^2_{\alpha,i}} + 3\right)\right) \\ &\leq k_{\max}\Delta_{\alpha,\max} + \sum_{i\in\overline{V}}\Delta_{\alpha,i}\left(\left(3+\frac{2\mu^*}{\delta^2_{\alpha,i}}\right)(k_{\max}+1) + \frac{3\log\beta(k_{\max}+1)^2}{2\delta^2_{\alpha,i}}\right).\end{aligned}$$

The proof of the first statement is concluded by upper-bounding the number of restarts up to time T as $k_{\max} \leq \frac{\log T}{\log \beta}$.

The second statement is proven by an argument analogous to the one used in the proof of Theorem 1, and straightforward calculations. □