Revealing graph bandits for maximizing local influence

Alexandra Carpentier, Michal Valko

Revealing graph bandits for maximizing local influence

Alexandra Carpentier
Universität Potsdam

Michal Valko
SequeL team, INRIA Lille - Nord Europe

Abstract

We study a graph bandit setting where the objective of the learner is to detect the *most influential node* of a graph by requesting as little information from the graph as possible. One of the relevant applications for this setting is marketing in social networks, where the marketer aims at finding and taking advantage of the most influential customers. The existing approaches for bandit problems on graphs require either partial or complete knowledge of the graph. In this paper, we do not assume any knowledge of the graph, but we consider a setting where it can be gradually discovered in a *sequential and active way*. At each round, the learner chooses a node of the graph and the only information it receives is a *stochastic* set of the nodes that the chosen node is currently influencing. To address this setting, we propose BARE, a bandit strategy for which we prove a regret guarantee that scales with the *detectable dimension*, a problem dependent quantity that is often much smaller than the number of nodes.

1 Introduction

Bandit problems on graphs (Mannor & Shamir, 2011; Caron et al., 2012) are sequential decision problems with *limited feedback*, where the learner can take advantage of a *graph structure* of the actions. This allows the learner to attain faster learning rates when compared to treating all the nodes independently. The recent popularity of this setting is due to its applications in marketing and advertising. In a typical case, the graph represents a social network, where the nodes are users and the edges encode the intensity of the social links between them. One marketing application that we target is *product placement*. An advertiser can offer a product to some users in a hope that they will recommend the product to their contacts, i.e., to the neighboring nodes in the social network. The advertiser then observes the set of contacts that these users have influenced and that have bought the product. The objective of the advertiser is to target *influential users*, the nodes of the graph whose influence is the most important. Ideally, the advertiser would only offer products to the users with maximal influence.

What plays the key role, when it comes to effective detection of the most influential users? It is the graph structure, as it gives *side information*, in particular, on the proximity between the nodes and their *influence* on others. The learner can leverage this side information and *learn faster*. The magnitude of the gains in the learning *rates* naturally depends on the graph. Consequently, the performance guarantees can be expressed with graph quantities such as the *clique partition number* (Caron et al., 2012), the *independence number* (Alon et al., 2013), or the *minimum dominating set* (Buccapatnam et al., 2014). Furthermore, there are many models of influence and some of the known ones were introduced in the seminal work on spreading the influence through a social network (Kempe et al., 2003, 2015). In the present paper, we consider *local influence*, where a node on the graph influences only its *immediate neighborhood*.

Most of the existing approaches for active learning on graphs assume that either the *entire graph* is known in advance, or at least that a substantial *part of the graph* is revealed to the learner after it selected the node. Typically, the algorithms require at least the knowledge of the set of neighbors of the neighbors of the nodes (*second neighborhood*). This knowledge of the graph is crucial for existing learning algorithms (Mannor & Shamir, 2011; Yu & Mannor, 2011; Caron et al., 2012; Cesa-Bianchi et al., 2013; Alon et al., 2013; Gentile et al., 2014; Kocák et al., 2014a; Gu & Han, 2014; Valko et al., 2014; Alon et al., 2015; Wu et al., 2015; Kocák et al., 2016) to help them learn faster than in the case if no structure existed. However, in some realistic scenarios, the graph information is *not*

Appearing in Proceedings of the 19^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2016, Cadiz, Spain. JMLR: W&CP volume 51.

available to the learner beforehand. Typically, the operator of the social network would not freely reveal the social links and therefore the graph is not known to the advertiser. On the other hand, for instance in the advertising example presented above, the advertiser has some local access to the social network in the sense that it can get information of the set of users that were influenced to purchase products through the other targeted customers. This information can be gathered through *promotional codes* when the goal is the *product purchase* or through "likes" in the case of an *information campaign* (Caron et al., 2012).

However, the existing graph bandit approaches do not allow to treat this scarce side information setting. Therefore, with the known tools, one can either (i) first thoroughly explore the graph and then apply existing graph bandit strategies, or (ii) forget about the underlying graph structure and apply existing multi-arm bandit algorithms to the nodes of the graph. In both cases, it is necessary that the learner substantially explores the graph and therefore *samples many nodes*, if not all of them. This is not very reasonable, for instance, in our marketing example, since graphs corresponding to social networks are usually large. Moreover, the advertiser is unlikely to have a large enough budget to target all the nodes of the graph in order to learn which ones are the most influential.

More formally, in this paper we consider a sequential graph learning problem where *at the beginning, there is no information* about the underlying graph. The edges are only *revealed progressively* as a result of the choices of the learner. Specifically, the learner observes the set of nodes that have been influenced by the chosen node. The objective of the learner is to find and target the *most influential node* of the graph. We consider a *local influence structure*. In this simple model, each node can only influence its neighbors. In this paper, we aim at finding a strategy for this problem that would be practical on very large graphs and in particular, that does not scale with the number of nodes if the graph has some structure. We propose a learning strategy that we call `BARE`, a bandit revelation algorithm. The performance guarantees for this algorithm do not scale with the number of nodes, but with the *detectable dimension*, a quantity that is often much smaller than the number of nodes. Specifically, `BARE` does not require to sample the entire graph when the *detectable dimension* is small.

Stochastic bandits on graphs were inspired by Mannor & Shamir (2011) and were first studied by Caron et al. (2012) that proposed `UCB-N` and `UCB-MaxN` that closely follow `UCB`, but in addition, they use side observations for better reward estimates (`UCB-N`) or choose one of the neighboring nodes with a better empirical estimate (`UCB-MaxN`). These improvements enable to improve guarantees, i.e., the regret does not scale with the number of nodes but with the *clique partition number*. Later, Buccapatnam et al. (2014) improved the results of Caron et al. (2012) with LP-based solutions and guarantees scaling only with the *minimum dominating set*. Spectral bandits (Valko et al., 2014; Kocák et al., 2014b; Gu & Han, 2014) assume that each node of the known graph has a mean reward that is *smooth* on the graph, which means that the connected nodes give similar rewards. Moreover, the gang of bandits (Cesa-Bianchi et al., 2013) addresses the problem of multiple users, where each of the nodes possesses a contextual linear bandit itself and the linear weight vectors of neighboring nodes are assumed to be similar. Furthermore, online clustering of bandits (Gentile et al., 2014) assumes that the nodes of the graph can be clustered with respect to some unknown underlying clustering and the nodes within a cluster exhibit similar behavior. Yet another assumption of a special graph reward structure is exploited by unimodal bandits (Yu & Mannor, 2011; Combes & Proutière, 2014) and networked bandits (Fang & Tao, 2014).

Bandits with side observations were first studied in the more difficult *non-stochastic setting* (Mannor & Shamir, 2011), where the rewards do not follow a fixed distribution. Their first algorithm, `ELP`, comes with the guarantee expressed as a function of the *clique number*. `ELP` was later followed by `Exp3-SET` (Alon et al., 2013), `Exp3-DOM` (Alon et al., 2013), and `Exp3-IX` (Kocák et al., 2014a), whose guarantees were proved to be functions of the *independence number*, which gives either equal or a better guarantee. This line of work was recently extended to the setting beyond bandit feedback (Alon et al., 2015), where the learner may not observe the reward of the chosen node and to *noisy* side observations (Wu et al., 2015; Kocák et al., 2016).

The common feature of all prior approaches is the need of having access to knowledge of the portions of the graph, in order to get faster learning rates. These portions are larger than what our setting permits. In particular, in our setting, the information revealed is just the *first neighborhood* of the chosen node.

Recently, Lei et al. (2015); Chen et al. (2015); Vaswani & Lakshmanan (2015) investigated the combinations of offline influence maximization approaches with multi-arm bandit strategies for the online influence maximization in the independent cascade model of Kempe et al. (2003) with *semi-bandit* feedback. Finally, another set of approaches considered restricted exploration constraints on a graph, modeling the crawling of the network (Bnaya et al., 2013a,b; Singla et al., 2015).

2 Local influence bandit settings

2.1 Description of the problem

Let $\mathcal{G}$ be a graph with d nodes. When a node i is selected, it can influence the nodes of $\mathcal{G}$, including itself. Node i influences each node j with *fixed* but *unknown* probability $p_{i,j}$. Let $\mathbf{M} = (p_{i,j})_{i,j}$ be the $d \times d$ matrix that represents $\mathcal{G}$.

We consider the following online, active setting. At each round (time) t, the learner chooses a node k_t and observes which nodes are influenced by k_t, i.e., the set $S_{k_t,t}$ of influenced nodes is *revealed*. Let us also write $S_{k_t,t}(r)$ for the rth coordinate of $S_{k_t,t}$, i.e., it is 1 if k_t influences r at time t and 0 otherwise. Given a budget of n rounds, the objective is to maximize the number of *influences* that the selected node exerts. Formally, our goal is to find the strategy maximizing the performance

$$L_n = \sum_{t=1}^{n} |S_{k_t,t}| \,.$$

The *influence* of node k, i.e., the expected number of nodes that node k exerts influence on, is by definition

$$r_k = \mathbb{E}\left[|S_{k,t}|\right] = \sum_{j \le d} p_{k,j}.$$

We also define the *dual influence* of node k as

$$r_k^\circ = \sum_{j \le d} p_{j,k}.$$

This quantity is the expected number of nodes that exert influence on node k. For an undirected graph $\mathcal{G}$, $\mathbf{M}$ is symmetric and $r_k^\circ = r_k$. However, in general, this is not the case, but we assume that the influence is up to a certain degree mutual. In other words, we assume that if a node is very influential, it also is subject to the influence of many other nodes. We make this precise in Section 3.

As the performance measure, we compare any *adaptive strategy* for this setting with the optimal oracle that knows $\mathbf{M}$. The oracle strategy always chooses one of the most influential nodes, which are the nodes whose expected number of influences r_k is maximal. We call one of these node $k^\star$, such that

$$k^\star = \arg\max_k \mathbb{E}\left[\sum_{t=1}^{n} |S_{k,t}|\right] = \arg\max_k n r_k.$$

Let the reward of this node be

$$r_\star = r_{k^\star}.$$

Then, its expected performance, if it consistently sampled $k^\star$ over n rounds, is equal to

$$\mathbb{E}\left[L_n^\star\right] = n r_\star.$$

The expected *regret* of any adaptive strategy that is unaware of $\mathbf{M}$, with respect to the oracle strategy, is defined as the expected difference of the two,

$$\mathbb{E}\left[R_n\right] = \mathbb{E}\left[L_n^\star\right] - \mathbb{E}\left[L_n\right].$$

Dually, we define $r_\star^\circ$ as the average number of influences received by the most influenced node,

$$r_\star^\circ = \max_k r_k^\circ.$$

2.2 Baseline comparison: Observing only $|S_{k_t}|$, the number of influenced nodes

For a meaningful baseline comparison that shows the benefit of the graph structure, we first consider a *restricted version* of the setting from Section 2.1. The restriction is that the learner, at round t, does not observe the set of influenced nodes $S_{k_t,t}$, but only the number number of elements in $S_{k_t,t}$, denoted by $|S_{k_t,t}|$. In other words, once we select a node, we receive as a feedback only the number of influenced nodes, *but not their identity*. In this setting, we do not observe enough information about the graph structure to exploit it, since we do not observe the *links* between the nodes. As a result, this setting can be mapped to a classic *multi-arm bandit* setting without underlying graph structure, where the reward that the learner observes for node k_t is equal to $|S_{k_t,t}|$.

If $n \ge d$, it is possible to directly apply classic multi-arm bandit reasoning. Since we never receive any information about the graph structure, we cannot exploit it and we can only consider the quantity $|S_{k_t,t}|$ as the standard bandit reward, which is a noisy version of r_{k_t}. Such problem is a standard bandit problem with rewards $|S_{k_t,t}|$, that are integers between 0 and d and have a variance bounded by r_{k_t}.

Directly building on upper and lower bounds arguments for the classic bandit strategies (Lai & Robbins, 1985; Audibert & Bubeck, 2009), we give the following result. This result's upper bound holds for a specific bandit algorithm that we call `GraphMOSS`, a slight adaptation of the `MOSS` algorithm by Audibert & Bubeck (2009) to our specific setting.

Theorem 1 (proof in Appendix A)**.** *In the graph bandit problem from Section 2.2, with the reward equal to the number of influenced nodes $|S_{k_t,t}|$ instead of $S_{k_t,t}$, the regret is bounded as follows.*

- Lower bound. *If for some fixed $\varepsilon > 0$, we have $\varepsilon d < r_\star < (1-\varepsilon)d$, then there exists a constant*

$\upsilon > 0$ such that for n large enough, depending on ε, we have that

$$\inf \sup \mathbb{E}\left[R_n\right] \geq \upsilon \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right),$$

where inf sup *means the best possible algorithm on the worst possible graph bandit problem.*

- Upper bound. *There exists a constant $U > 0$ such that the regret of Algorithm 1 is bounded as*

$$\mathbb{E}\left[R_n\right] \leq U \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right).$$

Algorithm 1 `GraphMOSS`

Input
d: the number of nodes
n: time horizon
Initialization
Sample each arm twice
Update $\widehat{r}_{k,2d}$, $\widehat{\sigma}_{k,2d}$, and $T_{k,2d} \leftarrow 2$, for $\forall k \leq d$
for $t = 2d+1, \ldots, n$ **do**
$C_{k,t} \leftarrow 2\widehat{\sigma}_{k,t}\sqrt{\frac{\max(\log(n/(dT_{k,t})),0)}{T_{k,t}}} + \frac{2\max(\log(n/(dT_{k,t})),0)}{T_{k,t}}$, for $\forall k \leq d$
$k_t \leftarrow \arg\max_k \widehat{r}_{k,t} + C_{k,t}$
Sample node k_t and receive $|S_{k_t,t}|$
Update $\widehat{r}_{k,t+1}$, $\widehat{\sigma}_{k,t+1}$, and $T_{k,t+1}$, for $\forall k \leq d$
end for

The lower bound holds also in the specific case where the graph $\mathcal{G}$ is undirected (i.e., symmetric $\mathbf{M}$), as is explained in the proof. This is an important remark as the undirected graphs are a canonical and "perfect" example of graphs where influencing and being influenced is correlated and where the dual influence is equal to the influence for each node.

3 The BARE algorithm and results

In this section we treat the *unrestricted* setting described in Section 2.1 where we *get revealed the identity of the influenced nodes*, while the reward stays the same as in Section 2.2. First, note that the minimax-optimal rate in this setting is the same as in the restricted information case above. To see that, one can, for instance, consider a network composed of isolated nodes with only a very small clique of most influential nodes, connected only to each other. Another example is a graph where the fact of being influential is uncorrelated with the fact of being influenced and where, for instance, the most influential node is not influenced by any node. For the same reasons as the ones described in Theorem 1, when $n \leq d$, there is no adaptive strategy in a minimax sense, also in this unrestricted setting.

However, the cases where the identity of the influenced nodes does not help, are somewhat pathological. Intuitively, they correspond to cases where the graph structure is not very informative for finding the most influential node. This is the case when there are many isolated nodes, and also in the case where observing nodes that are very influenced does not provide information on these nodes' influence. In many typical and more interesting situations, this is not the case. First, in these problems, the nodes that have high influence are also very likely to be subject being influenced, for instance, many interesting networks are symmetric and then it is immediately the case. Second, in the realistic graphs, there is typically a small portion of the nodes that are noticeably more connected than the others (Barabási & Albert, 1999).

In order to rigorously define these non-degenerate cases, let us first define function D that controls the number of nodes with a given *dual gap*, i.e., a given suboptimality with respect to the most influenced node

$$D(\Delta) \stackrel{\text{def}}{=} \left|\{i \leq d : r_\star^\circ - r_i^\circ \leq \Delta\}\right|.$$

The function $D(\Delta)$ is a non-decreasing quantity dual to the arm gaps. Note that $D(r) = d$ for any $r \geq r_\star^\circ$ and that $D(0)$ is the number of most influenced nodes. We now define the *problem dependent* quantities that express the difficulty of the problem and allow us to state our results.

Definition 1. *We define the **detectable horizon** as the smallest integer $T_\star > 0$ such that*

$$T_\star r_\star^\circ \geq \sqrt{D_\star n r_\star^\circ},$$

*when such $T_\star$ exists and $T_\star = n$ otherwise. Here, $D_\star$ is the **detectable dimension** defined as*

$$D_\star \stackrel{\text{def}}{=} D(\Delta_\star),$$

*where the **detectable gap** $\Delta_\star$ is defined as*

$$\Delta_\star \stackrel{\text{def}}{=} 16\sqrt{\frac{r_\star^\circ d \log(nd)}{T_\star}} + \frac{144 d \log(nd)}{T_\star}.$$

Remark 1. *From the definitions above, the detectable dimension is the $D_\star$ that corresponds to the smallest integer $T_\star > 1$ such that*

$$T_\star r_\star^\circ \geq \sqrt{D\left(16\sqrt{\frac{r_\star^\circ d \log(nd)}{T_\star}} + \frac{144 d \log(nd)}{T_\star}\right) n r_\star^\circ},$$

or $D_\star = d$ if such $T_\star$ does not exist. It is therefore a well defined quantity. Moreover, since D is nondecreasing and $D(0)$ is the number of most influenced

nodes, then $D_\star$ converges to the number of most influenced nodes as n tends to infinity.

Finally let us write the influential-influenced gap as

$$\varepsilon_\star \stackrel{\text{def}}{=} r_\star - \max_{k \in \mathcal{D}^\circ} r_k,$$

where $\mathcal{D}^\circ \stackrel{\text{def}}{=} \{i : r_i^\circ = \max_k r_k^\circ\}$. The quantity $\varepsilon_\star$ quantifies the gap between the most influential node overall vs. the most influential node in the set of most influenced nodes.

Remark 2. *The quantity $\varepsilon_\star$ is small when one of the most influenced nodes is also very influential. It is exactly zero when one of the most influential nodes happens to also be one of the most influenced nodes. For instance, the case $\varepsilon_\star = 0$ appears in undirected social network models with mutual influence.*

The graph structure is helpful when the D function decreases quickly with n. To give an intuition about how is D linked to the graph topology, consider a star-shaped graph which is the most helpful and can have $D_\star = 1$ even for a small n. On the other hand, a bad case is a graph with many small cliques. The worst case is where all nodes are disconnected except two, where $D_\star$ will be of order d even for a large n.

The detectable dimension $D_\star$ is a problem dependent quantity that represents the *complexity of the problem* instead of d. In real networks, $D_\star$ is typically smaller than the number of nodes d and we give several examples of the empirical value of $D_\star$ in Section 5 and Appendix C. As our analysis will show, $D_\star$ represents the number of nodes that we can *efficiently extract* from d nodes in less than n rounds of the time budget. Our *bandit revelator* algorithm, BARE (Algorithm 2), starts by the *global-exploration* phase and extracts a subset of cardinality less than or equal to a constant multiple of $D_\star$, that contains a very influential node, that is at most $\varepsilon_\star$ away from the most influential node. BARE does this extraction *without scanning all the d nodes*, which could be impossible anyway, since we do not restrict to $d \leq n$. In the subsequent *bandit* phase, BARE proceeds with scanning this smaller set of selected nodes to find the most influential one.

We now state our main theoretical result that proves a bound on the regret of BARE.

Theorem 2 (proof in Section 4)**.** *In the unrestricted local influence setting with information about the neighbors, BARE satisfies, for a constant $C > 0$,*

$$\mathbb{E}\left[R_n\right] \leq C \min\left(r_\star n, D_\star r_\star + \sqrt{r_\star n D_\star} + n\varepsilon_\star\right).$$

Remark 3. *Note that BARE does not need preliminary information about $\mathcal{G}$, as a classic multi-arm bandit strategy described in Section 2.2 would require in order to attain this rate.*

Algorithm 2 BARE: Bandit revelator

Input
d: the number of nodes
n: time horizon
Initialization
$T_{k,t} \leftarrow 0$, for $\forall k \leq d$
$\widehat{r_{k,t}^\circ} \leftarrow 0$, for $\forall k \leq d$
$t \leftarrow 1$, $\widehat{T}_\star \leftarrow 0$, $\widehat{D}_{\star,t} \leftarrow d$, $\widehat{\sigma}_{\star,1} \leftarrow d$
Global exploration phase
while $t\left(\widehat{\sigma}_{\star,t} - 4\sqrt{d\log(dn)/t}\right) \leq \sqrt{\widehat{D}_{\star,t} n}$ **do**
Influence a node at random (choose k_t uniformly at random) and get $S_{k_t,t}$ from this node
$\widehat{r_{k,t+1}^\circ} \leftarrow \frac{t}{t+1}\widehat{r_{k,t}^\circ} + \frac{d}{t+1} S_{k_t,t}(k)$
$\widehat{\sigma}_{\star,t+1} \leftarrow \max_{k'} \sqrt{\widehat{r_{k',t+1}^\circ} + 8d\log(nd)/(t+1)}$
$w_{\star,t+1} \leftarrow 8\widehat{\sigma}_{\star,t+1}\sqrt{\frac{d\log(nd)}{t+1}} + \frac{40d\log(nd)}{t+1}$
$\widehat{D}_{\star,t+1} \leftarrow \left|\left\{k : \max_{k'} \widehat{r_{k',t+1}^\circ} - \widehat{r_{k,t+1}^\circ} \leq w_{\star,t+1}\right\}\right|$
$t \leftarrow t+1$
end while
$\widehat{T}_\star \leftarrow t$.
Bandit phase
Run minimax-optimal bandit algorithm on the $\widehat{D}_{\star,\widehat{T}_\star}$ chosen nodes (e.g., Algorithm 1)

Corollary 1. *For an undirected social network model the expected regret of BARE is*

$$\mathbb{E}\left[R_n\right] \leq C \min\left(r_\star n, D_\star r_\star + \sqrt{r_\star n D_\star}\right),$$

which is the minimax-optimal regret in the case where there are $D_\star$ instead of d nodes. This highlights the dimensionality reduction potential of our method.

Finally, we state a lower bound for our setting. Notice that the influential-influence gap also appears here.

Theorem 3 (proof in Appendix B)**.** *Let $d \geq Cn > 0$, where $C > 0$ is a universal constant. Consider the set of unrestricted settings and the set of all problems that have maximal influence bounded by r (where for some fixed $u > 0$, we have $u\overline{D} < r < (1-u)\overline{D}$), $D_\star \leq \overline{D}$ and influential-influence gap smaller than ε (with $\varepsilon \leq u\sqrt{dr/n}$ for some small $u > 0$ if $d \leq n$). Then the expected regret of the best possible algorithm in the worst case of these problems is lower bounded as*

$$C'' \min\left(rn, \overline{D}r + \sqrt{rn\overline{D}} + n\varepsilon\right),$$

where C'' is a universal constant.

4 Proof of Theorem 2

For any node $k \leq d$ and any round t that is during the *global exploration phase*, let us define the following

estimator of reward $r^\circ_{k,t}$,

$$\widehat{r^\circ_{k,t}} = \frac{1}{t}\sum_{t'=1}^{t} dS_{k_t,t'}(k).$$

Notice that during the global exploration phase, the nodes are chosen uniformly at random among all the nodes. This means that for any k, the $(S_{k_t,t'}(k))_{t'}$ are i.i.d. Bernoulli random variables with parameter r°_k/d. By Bernstein inequality, this implies that with probability larger than $1-1/n^2$, for any node $k \leq d$ and for any round t within the global exploration phase,

$$\left|\widehat{r^\circ_{k,t}} - r^\circ_k\right| \leq 4\sqrt{\frac{dr^\circ_k \log(nd)}{t}} + \frac{4d\log(nd)}{t}. \quad (1)$$

Let ξ be the event such that Equation 1 holds. Note that on ξ, we have that for any t of the global exploration phase and for any $k \leq d$,

$$r^\circ_k - 4\sqrt{\frac{dr^\circ_k \log(nd)}{t}} + \frac{4d\log(nd)}{t} \leq \widehat{r^\circ_{k,t}} + \frac{8d\log(nd)}{t} \leq r^\circ_k + 4\sqrt{\frac{dr^\circ_k \log(nd)}{t}} + \frac{12d\log(nd)}{t},$$

which implies that on ξ, by factorizing the left hand side and right hand side, for any $k \leq d$,

$$\left(\sqrt{r^\circ_k} - 2\sqrt{\frac{d\log(nd)}{t}}\right)^2 \leq \widehat{r^\circ_{k,t}} + \frac{8d\log(nd)}{t} \leq \left(\sqrt{r^\circ_k} + 4\sqrt{\frac{d\log(nd)}{t}}\right)^2,$$

which implies

$$\sqrt{r^\circ_k} - 2\sqrt{\frac{d\log(nd)}{t}} \leq \sqrt{\widehat{r^\circ_{k,t}} + \frac{8d\log(nd)}{t}} \leq \sqrt{r^\circ_k} + 4\sqrt{\frac{d\log(nd)}{t}}.$$

We deduce from this that

$$\left|\sqrt{r^\circ_k} - \sqrt{\widehat{r^\circ_{k,t}} + \frac{8d\log(nd)}{t}}\right| \leq 4\sqrt{\frac{d\log(nd)}{t}}.$$

In particular, this implies that on ξ,

$$\left|\widehat{\sigma}_{\star,t} - \sqrt{r^\circ_\star}\right| \leq 4\sqrt{\frac{d\log(nd)}{t}}. \quad (2)$$

On ξ, we also have by Equation 1,

$$\left|\left(\max_{k'} \widehat{r^\circ_{k',t}} - \widehat{r^\circ_{k,t}}\right) - (r^\circ_\star - r^\circ_k)\right| \leq 8\sqrt{\frac{dr^\circ_\star \log(nd)}{t}} + \frac{8d\log(nd)}{t},$$

which implies that on ξ, by Equation 2,

$$\left|\left(\max_{k'} \widehat{r^\circ_{k',t}} - \widehat{r^\circ_{k,t}}\right) - (r^\circ_\star - r^\circ_k)\right| \leq 8\widehat{\sigma}_{\star,t}\sqrt{\frac{d\log(nd)}{t}} + \frac{40d\log(nd)}{t}.$$

Note that by the definition of the global exploration phase, we know that for any round $t \leq \widehat{T}_\star$, the set of most influenced nodes $\mathcal{D}^\circ$ will be on ξ in the set of the $\widehat{D}_{\star,t}$ kept nodes. Note that by Equation 2, this also implies that on ξ,

$$\widehat{D}_{\star,t} \leq D\left(16\widehat{\sigma}_{\star,t}\sqrt{\frac{d\log(nd)}{t}} + \frac{80d\log(nd)}{t}\right) \leq D\left(16\sqrt{\frac{dr^\circ_\star \log(nd)}{t}} + \frac{144d\log(nd)}{t}\right). \quad (3)$$

First case: the global exploration phase finishes before $3T_\star$ We consider the case $\widehat{T}_\star \leq 3T_\star$. If the exploration finishes at $\widehat{T}_\star$, then on ξ, by Equation 2, and by the definition of BARE,

$$3T_\star\sqrt{r^\circ_\star} \geq \widehat{T}_\star\sqrt{r^\circ_\star} \geq \sqrt{\widehat{D}_{\star,\widehat{T}_\star} n}.$$

By the definition of $D_\star$ we also have that

$$\sqrt{D_\star n r^\circ_\star} \geq (T_\star - 1)r^\circ_\star \geq T_\star r^\circ_\star/2,$$

which together implies

$$\widehat{D}_{\star,\widehat{T}_\star} \leq 36D_\star.$$

Also, on ξ, the optimal arm is among the $\widehat{D}_{\star,\widehat{T}_\star}$ arms.

Second case: the global exploration phase finishes after $3T_\star$ The detectable gap $\Delta_\star$ is equal to

$$\Delta_\star = 16\sqrt{\frac{r^\circ_\star d\log(nd)}{T_\star}} + \frac{144d\log(nd)}{T_\star}.$$

Since the detectable dimension $D_\star$ is smaller or equal to d, then $\Delta_\star \leq r^\circ_\star$. This implies that $T_\star$ must satisfy

$$r^\circ_\star \geq 16\sqrt{\frac{r^\circ_\star d\log(nd)}{T_\star}} + \frac{144d\log(nd)}{T_\star},$$

which implies that

$$T_\star \geq \frac{144d\log(nd)}{r^\circ_\star}. \quad (4)$$

In the case we consider $\left(\widehat{T}_\star \geq 3T_\star\right)$, the exploration phase does not stop at $3T^\star$ and we have that on ξ,

$$3T_\star\left(\widehat{\sigma}_{\star,T^\star} - 4\sqrt{\frac{d\log(nd)}{T_\star}}\right) \leq \sqrt{\widehat{D}_{\star,T_\star} n},$$

and thus by Equation 2, we have that

$$3T_\star\left(\sqrt{r_\star^\circ}-8\sqrt{\frac{d\log(nd)}{T_\star}}\right)\leq\sqrt{\widehat{D}_{\star,T_\star}n},$$

which implies in turn by Equation 4 that

$$\frac{3T_\star\sqrt{r_\star^\circ}}{3}\leq\sqrt{\widehat{D}_{\star,T_\star}n}.$$

Combining Equation 3, with the fact that D is a non-decreasing function, we get that on ξ,

$$\begin{aligned}T_\star\sqrt{r_\star^\circ}&\leq\sqrt{D\left(16\sqrt{\frac{dr_\star^\circ\log(nd)}{T_\star}}+\frac{144d\log(nd)}{T_\star}\right)n}\\&\leq\sqrt{D_\star n},\end{aligned}$$

which is false by definition of $D_\star$ and $T_\star$. Therefore, we know that on ξ, $3T_\star\geq\widehat{T}_\star$.

Conclusion To sum up, we know that on ξ,

$$\widehat{T}_\star\leq 3T_\star\quad\text{and}\quad\widehat{D}_{\star,\widehat{T}_\star}\leq 36D_\star,$$

and that the set of most influenced nodes $\mathcal{D}^\circ$ is among the nodes that are kept at the end of the global exploration phase. In particular, this implies that the gap with respect to the most influential node on this set is at most $\varepsilon_\star$.

Taking the $\widehat{D}_{\star,\widehat{T}_\star}\leq 36D_\star$ kept arms and running a minimax bandit algorithm, such as `GraphMOSS`, we can upper bound the regret incurred in the remaining rounds using Theorem 1. Since there are $n-T_\star$ remaining rounds, this implies that the expected regret on these last rounds, on ξ, for a given constant $C'>0$, bounded by

$$C'D_\star r_\star+C'\sqrt{r_\star\widehat{D}_{\star,\widehat{T}_\star}\left(n-\widehat{T}_\star\right)}\leq C'D_\star r_\star+C'\sqrt{r_\star D_\star n},$$

with respect to the optimal nodes in the set of kept nodes. Now, since $\mathcal{D}^\circ$ is in the set of kept nodes, and since the maximal gap of most influential nodes with respect to this set is at most $\varepsilon_\star$, the regret with respect to the most influential node $r_\star$ is

$$\begin{aligned}&C'D_\star r_\star+C'\sqrt{r_\star\widehat{D}_{\star,\widehat{T}_\star}\left(n-\widehat{T}_\star\right)}+n\varepsilon_\star\\&\qquad\leq C'D_\star r_\star+C'\sqrt{r_\star D_\star n}+n\varepsilon_\star.\end{aligned}$$

We can now conclude the proof by bounding the expected regret as

$$\begin{aligned}\mathbb{E}\left[R_n\right]&\leq T_\star r_\star+C'D_\star r_\star+C'\sqrt{r_\star D_\star n}+n\varepsilon_\star+\frac{r_\star}{n}\\&\leq(C'+2)\sqrt{r_\star D_\star n}+2C'r_\star D_\star+n\varepsilon_\star\\&\leq C\left(\sqrt{r_\star D_\star n}+r_\star D_\star\right)+n\varepsilon_\star.\end{aligned}$$

4.1 Discussion

Lower bound Theorem 3 holds in the case $Cn\leq d$ and makes the quantity $\varepsilon_\star$ appear. But we emphasize that in the case $n\geq d$, even if the oracle provides the learner with a set of nodes such that the optimal node belongs to this set, the minimax-optimal rate for the bandit problem becomes

$$C\min\left(r_\star n, D_\star r_\star+\sqrt{r_\star n D_\star}\right),$$

for a constant C. This can be seen from an argument similar to Theorem 1, together with the example with isolated nodes, given above. This argument holds even for undirected graphs with $\varepsilon_\star=0$. In this sense, `BARE` is minimax-optimal over the set of problems with detectable dimension $D_\star$.

Large scale setting The quantity $D_\star$ and `BARE` become particularly appealing when we consider an interesting practical situation with a large number of graph nodes. For instance, even in a medium-sized social network, the advertiser would not have enough budget to target all the users and discover the most influential one, i.e., $n\leq d$. Notice again, that in the restricted setting of Section 2.2, the regret of bandit strategies in this problem for $n\ll d$ is of order $nr_\star$, which is larger than the regret of `BARE`.

However, in the unrestricted setting, the situation is different when $D_\star\leq n$. This is the case where a small number of nodes is noticeably more influential than the others and the regret of `BARE` is of order

$$D_\star r_\star+\sqrt{r_\star n D_\star}+n\varepsilon_\star,$$

which is smaller than $nr_\star$, and the problem becomes *learnable*.

5 Experiments

The purpose of our experiments is to show that `BARE` can do better in the regime $n\leq d$, compared to the algorithms ignoring the graph structure. For the minimax optimal algorithm during the bandit phase of `BARE`, we used `GraphMOSS`, defined in Section 2.2 and analyzed in Appendix A, which is a close variation of the `MOSS` algorithm (Audibert & Bubeck, 2009). We also used `GraphMOSS` as the baseline algorithm that does not use the graph structure.

The confidence parameter δ was set to 0.01 and $p_{i,j}$ to 0.8 for all i and j. This means that whenever a node is chosen, each of its neighbors is influenced and revealed with probability 0.8. Since the confidence terms of `BARE` are conservative, in the experiments we multiplied them by 0.01. All figures show the results averaged over 100 trials.

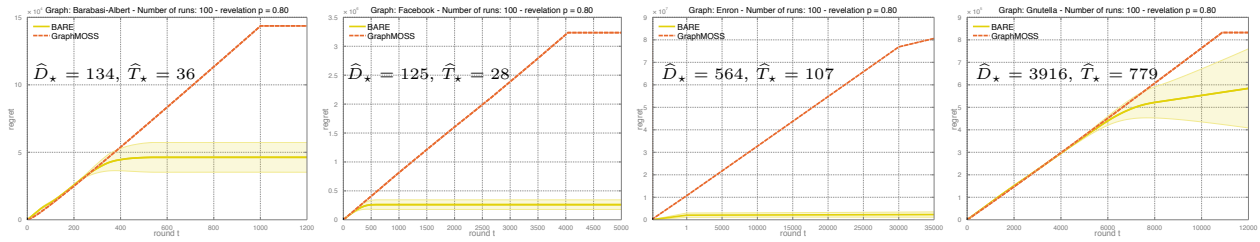

Figure 1: *Left*: Barabási-Albert. *Middle left*: Facebook. *Middle right*: Enron. *Right*: Gnutella.

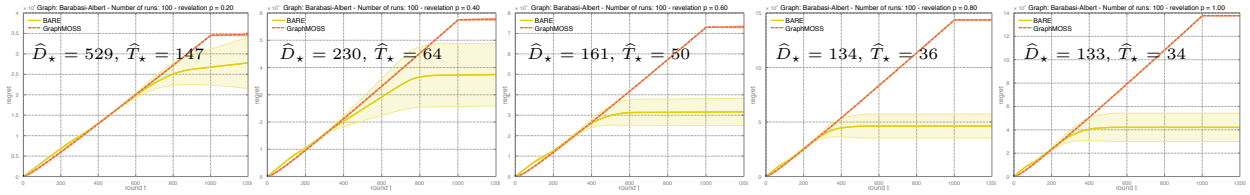

Figure 2: Barabási-Albert model with varying p between 0.2 and 1

We first performed an experiment on a graph generated by 10-out-degree Barabási-Albert model with $d = 1000$ nodes. Figure 1 (left) compares BARE with GraphMOSS. As expected, GraphMOSS suffers linear regret up to time $t = d$, since there is no sharing of information and for $t \leq d$, GraphMOSS pulls each arm once. While the regret of GraphMOSS is no longer linear for $t > d$ and eventually detects the best node, BARE is able to detect promising nodes much sooner during its global exploration phase and we can see the benefit of revealed information already around $t = 300$.

In Figure 2, we varied the probability of revelation p for a Barabási-Albert graph. When p close is to one, the more of the graph structure is revealed and the problem becomes easier. On the other hand, with p close to zero we do not get as much information about the structure and the performance of BARE and GraphMOSS are similar.

We also performed the experiments on Enron mail graph (Klimt & Yang, 2004) with $d = 36692$ and the snapshot of symmetrized version of Gnutella network from August 4th, 2002 (Ripeanu et al., 2002) with $d = 10879$, obtained from Stanford Large Network Dataset Collection (Leskovec & Krevl, 2014). Furthermore, we evaluated BARE on a subset of Facebook network with $d = 4039$ (Viswanath et al., 2009). We used the same parameters as for the Barabási-Albert case.

As expected, Figure 1 (middle left, middle right, right) shows that the performance gains of BARE over GraphMOSS depend heavily on the structure. In Enron and Facebook, the gain of BARE is significant which suggests that the graphs from these networks feature a relatively small number of influential nodes. On the other hand, the gain of BARE on Gnutella was much smaller which again suggests that this network is more decentralized.

In all the plots we include also the empirical estimate of the detectable dimension $\widehat{D}_\star$ and the detectable horizon $\widehat{T}_\star$. Notice that the smaller $\widehat{D}_\star$, as compared to d, and the smaller $\widehat{T}_\star$ is as compared to n, the sooner is BARE able to learn the most influential node as compared to GraphMOSS.

6 Conclusion

We hope that out work on local revelation incites the extensions on more elaborate propagation models on graphs (Kempe et al., 2015). One way to directly extend to more general propagation models is to consider that a more distant neighbor is a direct neighbor with contamination probability being the sum of the path products. Moreover, if we allow for more feedback, e.g., the identity of the influencing paths, our results could extend more efficiently. Note that in our setting, we were completely agnostic to the graph structure. Realistic networks often exhibit some additional structural properties that are captured by graph generator models, such as various *stochastic block models* (Girvan & Newman, 2002). In future, we would like to extend our approach to cases where we can take advantage of the assumptions stemming from these models and consider the subclasses of graph structures where we can further improve the learning rates.

Acknowledgements We thank Alan Mislove for the Facebook dataset. The research presented in this paper was supported by French Ministry of Higher Education and Research, Nord-Pas-de-Calais Regional Council, French National Research Agency project ExTra-Learn (n.ANR-14-CE24-0010-01), and by German Research Foundation's Emmy Noether grant MuSyAD (CA 1488/1-1).

References

Alon, Noga, Cesa-Bianchi, Nicolò, Gentile, Claudio, and Mansour, Yishay. From bandits to experts: A tale of domination and independence. In *Neural Information Processing Systems*, 2013.

Alon, Noga, Cesa-Bianchi, Nicolò, Dekel, Ofer, and Koren, Tomer. Online learning with feedback graphs: Beyond bandits. In *Conference on Learning Theory*, 2015.

Audibert, Jean-Yves and Bubeck, Sébastien. Minimax policies for adversarial and stochastic bandits. In *Conference on Learning Theory*, 2009.

Barabási, Albert-László and Albert, Réka. Emergence of scaling in random networks. *Science*, 286:11, 1999.

Bnaya, Zahy, Puzis, Rami, Stern, Roni, and Felner, Ariel. Bandit algorithms for social network queries. In *International Conference on Social Computing*, 2013a.

Bnaya, Zahy, Puzis, Rami, Stern, Roni, and Felner, Ariel. Social network search as a volatile multi-armed bandit problem. *Human Journal*, 2(2):84–98, 2013b.

Buccapatnam, Swapna, Eryilmaz, Atilla, and Shroff, Ness B. Stochastic bandits with side observations on networks. In *International Conference on Measurement and Modeling of Computer Systems*, 2014.

Caron, Stéphane, Kveton, Branislav, Lelarge, Marc, and Bhagat, Smriti. Leveraging side observations in stochastic bandits. In *Uncertainty in Artificial Intelligence*, 2012.

Cesa-Bianchi, Nicolò, Gentile, Claudio, and Zappella, Giovanni. A gang of bandits. In *Neural Information Processing Systems*, 2013.

Chen, Wei, Wang, Yajun, and Yuan, Yang. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. Technical report, 2015.

Combes, Richard and Proutière, Alexandre. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, 2014.

Fang, Meng and Tao, Dacheng. Networked bandits with disjoint linear payoffs. In *Internattional Conference on Knowledge Discovery and Data Mining*, 2014.

Gentile, Claudio, Li, Shuai, and Zappella, Giovanni. Online clustering of bandits. In *International Conference on Machine Learning*, 2014.

Girvan, Michelle and Newman, Mark E J. Community structure in social and biological networks. *National Academy of Sciences of the United States of America*, 99(12):7821–6, 2002.

Gu, Quanquan and Han, Jiawei. Online spectral learning on a graph with bandit feedback. In *International Conference on Data Mining*, 2014.

Kempe, David, Kleinberg, Jon, and Tardos, Éva. Maximizing the spread of influence through a social network. *Knowledge Discovery and Data mining*, pp. 137, 2003.

Kempe, David, Kleinberg, Jon, and Tardos, Éva. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2015.

Klimt, Bryan and Yang, Yiming. Introducing the Enron corpus. In *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, 2004.

Kocák, Tomáš, Neu, Gergely, Valko, Michal, and Munos, Rémi. Efficient learning by implicit exploration in bandit problems with side observations. In *Neural Information Processing Systems*, 2014a.

Kocák, Tomáš, Valko, Michal, Munos, Rémi, and Agrawal, Shipra. Spectral Thompson sampling. In *AAAI Conference on Artificial Intelligence*, 2014b.

Kocák, Tomáš, Neu, Gergely, and Valko, Michal. Online learning with noisy side observations. In *International Conference on Artificial Intelligence and Statistics*, 2016.

Lai, Tze L and Robbins, Herbert. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Lei, Siyu, Maniu, Silviu, Mo, Luyi, Cheng, Reynold, and Senellart, Pierre. Online influence maximization. In *Knowledge Discovery and Data mining*, 2015.

Leskovec, Jure and Krevl, Andrej. SNAP datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, 2014.

Mannor, Shie and Shamir, Ohad. From bandits to experts: On the value of side-observations. In *Neural Information Processing Systems*, 2011.

Ripeanu, Matei, Iamnitchi, Adriana, and Foster, Ian. Mapping the Gnutella network. *IEEE Internet Computing*, 6(1):50–57, 2002.

Singla, Adish, Horvitz, Eric, Kohli, Pushmeet, White, Ryen, and Krause, Andreas. Information gathering in networks via active exploration. In *International Joint Conferences on Artificial Intelligence*, 2015.

Valko, Michal, Munos, Rémi, Kveton, Branislav, and Kocák, Tomáš. Spectral bandits for smooth graph functions. In *International Conference on Machine Learning*, 2014.

Vaswani, Sharan and Lakshmanan, Laks. V. S. Influence maximization with bandits. Technical report, http://arxiv.org/abs/1503.00024, 2015.

Viswanath, Bimal, Mislove, Alan, Cha, Meeyoung, and Gummadi, Krishna P. On the evolution of user interaction in facebook. In *ACM Workshop on Online Social Networks*, 2009.

Wu, Yifan, György, András, and Szepesvári, Csaba. Online learning with Gaussian payoffs and side observations. In *Neural Information Processing Systems*, 2015.

Yu, Jia Yuan and Mannor, Shie. Unimodal bandits. In *International Conference on Machine Learning*, 2011.

A Proof of Theorem 1

Theorem 1. *In the graph bandit problem from Section 2.2, with the reward equal to the number of influenced nodes $|S_{k_t,t}|$ instead of $S_{k_t,t}$, the regret is bounded as follows.*

- Lower bound. *If for some fixed $\varepsilon > 0$, we have $\varepsilon d < r_\star < (1-\varepsilon)d$, then there exists a constant $\upsilon > 0$ such that for n large enough, depending on ε, we have that*

$$\inf \sup \mathbb{E}\left[R_n\right] \geq \upsilon \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right),$$

  where inf sup *means the best possible algorithm on the worst possible graph bandit problem.*

- Upper bound. *There exists a constant $U > 0$ such that the regret of Algorithm 1 is bounded as*

$$\mathbb{E}\left[R_n\right] \leq U \min\left(r_\star n, r_\star d + \sqrt{r_\star n d}\right).$$

The upper bound follows immediately from the results of Audibert & Bubeck (2009) using the Bernstein bound in the confidence term in the `MOSS` algorithm (Audibert & Bubeck, 2009). The confidence term in the resulting Algorithm 1 becomes

$$C_{k,t} = 2\widehat{\sigma}_{k,t}\sqrt{\frac{\max\left(\log(n/(dT_{k,t})),0\right)}{T_{k,t}}} + \frac{2\max(\log(n/(dT_{k,t})),0)}{T_{k,t}},$$

where $T_{k,t}$ is the number of pulls of node k at round t, and $\widehat{\sigma}^2_{k,t}$ is the empirical variance of node k at round t. This resulting scaling is then bounded by $r_\star$, since $|S_{k_t,t}|$ is a sum of Bernoulli random variables such that the sum of their variances is $\sigma^2_{k_t} \leq r_{k_t} \leq r_\star$, which implies that the variance of $|S_{k_t,t}|$ is bounded by $r_\star$. The result is obtained following the scheme of Audibert & Bubeck (2009).

The lower bound can be deduced from the one of Lai & Robbins (1985), replicating their d-armed problems with gap $\sqrt{d/n}$ by an equivalent problem with the graph structure where we define the following setup:

- One arm influences any arm with probability $r_\star/d$ with $1-\varepsilon > r_\star/d > \varepsilon$.
- The other arms influence each other with probability $r_\star/d - \sqrt{r_\star/(dn)}$ that is between $1-\varepsilon$ and $\varepsilon/2$ for n large enough, i.e., larger than d times a universal constant.

The gap with respect to any suboptimal arm is $\sqrt{dr_\star/n}$ and the variance for each arm is of order $r_\star$ for n large enough (larger than d times a universal constant). The statement of the lower bound by the adaptation of the proof of Lai & Robbins (1985) to this specific sub-Gaussian reward, i.e., sum of Bernoulli random variables with a parameter between $1-\varepsilon$ and $\varepsilon/2$.

For both the upper and the lower bound, the quantity $r_\star d$ corresponds to the fact that each arm must be sampled at least once in the case of $n \geq d$ and the quantity $r_\star n$ corresponds to the *unlearnable* case where $n \leq d$. Otherwise, one can always consider the worst case permutation of the arms.

B Proof of Theorem 3

Theorem 3. *Let $d \geq Cn > 0$, where $C > 0$ is a universal constant. Consider the set of unrestricted settings and the set of all problems that have maximal influence bounded by r (where for some fixed $u > 0$, we have $u\overline{D} < r < (1-u)\overline{D}$), $D_\star \leq \overline{D}$ and influential-influence gap smaller than ε (with $\varepsilon \leq u\sqrt{dr/n}$ for some small $u > 0$ if $d \leq n$). Then the expected regret of the best possible algorithm in the worst case of these problems is lower bounded as*

$$C'' \min\left(rn, \overline{D}r + \sqrt{rn\overline{D}} + n\varepsilon\right),$$

where C'' is a universal constant.

If $\sqrt{nr\overline{D}} \geq an\varepsilon$ for a small $a > 0$, we consider the following construction. We consider the construction of the lower bound in Theorem 1, where the player also receives the position of the $\overline{D}$ best nodes as an additional information. This situation is therefore easier than the full problem. Let $\mathcal{S}$ be the set of $\overline{D}$ best nodes. For any $k, l \in \mathcal{S}^2$, if neither k or l is the optimal node then we set $p_{k,l} = r/\overline{D} - \sqrt{r/(\overline{D}n)}$, if either k or l is the optimal node then $p_{k,l} = r/\overline{D}$. For the remaining nodes k and l, we define $p_{k,l} = 0$. In this situation, $D_\star = \overline{D}$, $\varepsilon_\star = 0$, and $r_\star = r$ can be chosen arbitrarily. Since in this case the graph structure does not significantly help,[1] we can use the bound of Theorem 1 with the knowledge of which arms are in $\mathcal{S}$. The lower bound in this case implies the result when $\sqrt{nr\overline{D}} \geq an\varepsilon$.

Alternatively, if $\sqrt{nr\overline{D}} \leq an\varepsilon$, we consider two cases: $n \leq d$ and $n \geq d$. If $n \leq d$, the result clearly holds because we can, similarly to the proof of Theorem 1, consider an asymmetric graph such that the graph structure does not help and where the detectable dimension is 1 for n large enough.[2] If $n \geq d$, we consider a related construction as in the first part of Theorem 1, setting $p_{l,k_0} = 1$ for some suboptimal node k_0 and any l so that the detectable dimension is 1 for n large enough, and setting $p_{l,k} = r_l/d$ for $k \neq k_0$ and any l so that the graph is asymmetric, therefore making the graph structure completely useless. Then, by Theorem 1, the regret is at least of order $\sqrt{drn}$ and since $\sqrt{drn}$ is higher than εn, we get the result.

C Detectable dimension

In this appendix, we provide additional plots that help understand the behavior of detectable dimension $D_\star$ as a function of number of rounds n.

As a sanity check, in Figure 3, we show the behaviour on an easy *star* graph where $D_\star = 1$ even for a small n and a difficult *complete* graph, where $D_\star = d$, even for a large n. In the case of the empty graph — classic bandit setting — $D_\star = d$ even for a large n as well.

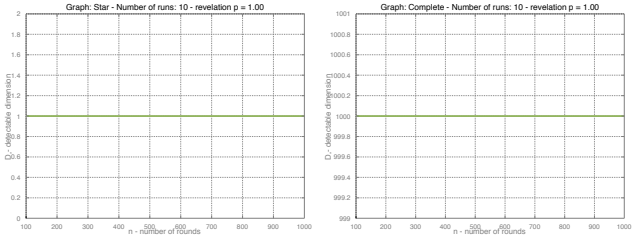

Figure 3: *Left*: Star graph. *Right*: Complete graph.

In Figure 4, we plot the empirical value of $D_\star$ as a function of n, for the graphs used in Section 5 to give intuition on how $D_\star$ decreases for different graphs.

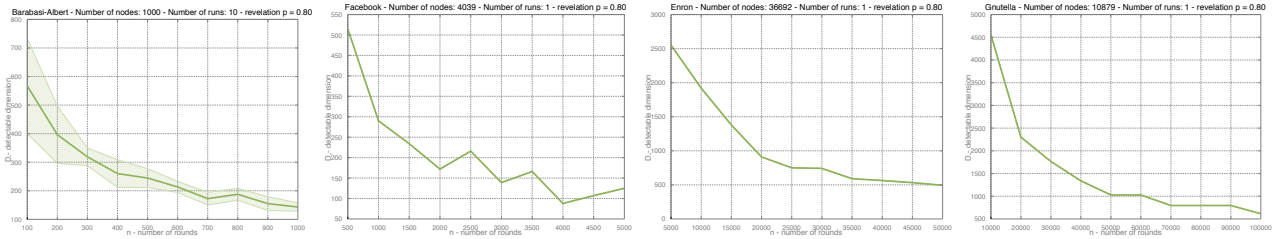

Figure 4: *Left*: Barabási-Albert. *Middle left*: Facebook. *Middle right*: Enron. *Right*: Gnutella.

[1] If we consider only the information we obtain for a node from the graph, and not the information we obtain from *pulling* the node, the error is at least $d\sqrt{\varepsilon/n}$ even at the end of the budget, which is higher than the gap $\sqrt{r\overline{D}/n}$, and therefore the information coming from the graph structure does not significantly help.

[2] We can set $p_{l,k_0} = 1$ for some suboptimal node k_0 and for any l; and set $p_{l,k} = r_l/d$ for any k, l and r_l as in the second part of Theorem 1.