Supplementary material

A An NP-Hard Problem

A.1 Proof of Theorem 4.1

Theorem A.1. *Consider a given matrix $\mathbf{M}_\star \in \mathbb{R}^{d\times d}$ and a finite arm set $\mathcal{X} \subset \mathbb{R}^d$. Unless P=NP, there is no polynomial time algorithm guaranteed to find the optimal solution of*

$$\max_{\left(x^{(1)},\dots,x^{(n)}\right)\in\mathcal{X}^n} \sum_{(i,j)\in E} x^{(i)\top}\mathbf{M}_\star\, x^{(j)} \ .$$

Proof. We prove the statement by reduction to the Max-Cut problem. Let $\mathcal{G} = (V, E)$ be a graph with $V = \{1, \dots, n\}$. Let $\mathcal{X} = \{e_0, e_1\}$, where $e_0 = (1, 0)^\top$ and $e_1 = (0, 1)^\top$. Let $\mathbf{M}_\star = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. For any joint arm assignment $\left(x^{(1)} \dots x^{(n)}\right) \in \mathcal{X}^n$, let $F \subseteq E$ be defined as $F = \left\{i : x^{(i)} = e_1\right\}$. Note that

$$\sum_{(i,j)\in E} x^{(i)\top}\mathbf{M}_\star x^{(j)} = \sum_{(i,j)\in E} \mathbf{1}\left[x^{(i)} \neq x^{(j)}\right] = 2 \times \sum_{(i,j)\in E} \mathbf{1}\left[i \in F, j \notin F\right],$$

where $\mathbf{1}[\cdot]$ is the indicator function. The assignement $\left(x^{(1)}, \dots, x^{(n)}\right)$ induces a cut $(F, V\backslash F)$, and the value of the assignment is *precisely* twice the value of the cut. Thus, if there was a polynomial time algorithm solving our problem, this algorithm would also solve the Max-Cut problem. □

A.2 Proof of Theorem 4.2

Theorem A.2. *Let us consider the graph $\mathcal{G} = (V, E)$, a finite arm set $\mathcal{X} \subset \mathbb{R}^d$ and the matrix $\mathbf{M}_\star$ given as input to Algorithm 1. Then, the expected global reward $r = \sum_{(i,j)\in E} x^{(i)\top}\mathbf{M}_\star x^{(j)}$ associated to the returned allocation $\mathbf{x} = \left(x^{(1)}, \dots, x^{(n)}\right) \in \mathcal{X}^n$ verifies:*

$$\frac{r - r_{\min}}{r_\star - r_{\min}} \geq \frac{1}{2} \ .$$

where $r_\star$ and $r_{\min}$ are respectively the highest and lowest global reward one can obtain with the appropriate joint arm. Finally, the complexity of the algorithm is in $\mathcal{O}(K^2 + n)$.

Proof. Given the matrix $\mathbf{M}_\star$, the algorithm obtains the two node-arms $(x_\star, x'_\star) \in \mathcal{X}$ solution of

$$\max_{(x,x')\in\mathcal{X}} x^\top \mathbf{M}_\star x' \ .$$

Note that it is equivalent to obtain $z_\star$ solution of

$$\max_{z\in\mathcal{Z}} z^\top \operatorname{vec}(\mathbf{M}_\star) \ .$$

Let us analyze a round of Algorithm 1 where we assign the arm of a node in V. For sake of simplicity, we assume that node i is assigned at round i. At round i, we count the number $n_1^{(i)}$ of neighbors of i that have

already been assigned the arm $x_\star$ and we count the number $n_2^{(i)}$ of neighbors of i that have already been assigned the arm $x'_\star$. Then, node i is assigned the arm least represented among its neighbors, that is arm $x_\star$ if $n_2^{(i)} \geq n_1^{(i)}$ and $x'_\star$ otherwise. Eventually, the optimal edge-arm $z_\star$ has been assigned $\max(n_1^{(i)}, n_2^{(i)})$ times among node i's neighborhood. Hence, for each node i, if we denote r_i the sum of all the rewards obtained with the edge-arms constructed only during the round i, we have

$$\begin{aligned} r_i &= \max\left(n_1^{(i)}, n_2^{(i)}\right) z_\star^\top \theta_\star + \min\left(n_1^{(i)}, n_2^{(i)}\right) z^\top \theta_\star \\ &\geq \frac{n_1^{(i)} + n_2^{(i)}}{2} (z_\star^\top \theta_\star + z^\top \theta_\star) \ . \end{aligned}$$

One can notice that the arm z can only be equal to $\text{vec}\left(x_\star x_\star^\top\right)$ or $\text{vec}\left(x_\star'^\top x_\star'^\top\right)$. Let assume that $\text{vec}\left(x_\star x_\star^\top\right)^\top \theta_\star \leq \text{vec}\left(x'_\star x'^\top_\star\right)^\top \theta_\star$ without loss of generality and let consider the worst case where z is always equal to $\text{vec}\left(x_\star x_\star^\top\right)$. Since z is constructed with the same node-arm $x_\star$, the allocation that constructs at each edge the edge-arm z exists (which is allocating $x_\star$ to all the nodes), thus $m \times z^\top \theta_\star \geq r_{\min}$.

Moreover one can also notice that $m \times z_\star^\top \theta_\star \geq r_\star$. We thus have,

$$r_i \geq \frac{n_1^{(i)} + n_2^{(i)}}{2m} (r_\star + r_{\min}) \ .$$

Now let us sum all the rewards obtained with the constructed edge-arms at each round of the algorithm, that is the global reward r of the graph allocation returned by the proposed algorithm:

$$\begin{aligned} r &= \sum_{i=1}^{n} r_i \\ &\geq \sum_{i=1}^{n} \frac{n_1^{(i)} + n_2^{(i)}}{2m} (r_\star + r_{\min}) \\ &= \frac{1}{2}(r_\star + r_{\min}) \\ &= \frac{1}{2}(r_\star - r_{\min}) + r_{\min} \ . \end{aligned}$$

Moreover, the algorithm does K^2 estimation to find the best couple $(x_\star, x'_\star) \in \mathcal{X}^2$, and each of the n rounds of the algorithm is in $O(1)$. Hence the complexity is equal to $O(K^2 + n)$. □

B Deriving the stopping condition

In this section, we remind key results to derive the stopping condition. We refer the reader to Soare et al. (2014) and references therein for additional details. Let $\mathcal{Z} \subset \mathbb{R}^{d^2}$ be the set of edge-arms and let $K^2 = |\mathcal{Z}|$. For $m, t > 0$, we consider a sequence of edge-arms $\mathbf{z}_t = (z_1, \ldots, z_{mt}) \in \mathcal{Z}^{mt}$ and the corresponding noisy rewards $(r_1, \ldots, r_{mt})$. We assume that the noise terms in the rewards are i.i.d., following a σ-sub-Gaussian distribution. Let $\hat{\theta}_t = \mathbf{A}_t^{-1} b_t \in \mathbb{R}^{d^2}$ be the solution of the ordinary least squares problem with $\mathbf{A}_t = \sum_{i=1}^{mt} z_i z_i^\top \in \mathbb{R}^{d^2 \times d^2}$ and $b_t = \sum_{i=1}^{k} z_i r_i \in \mathbb{R}^{d^2}$. We first recall the following property.

Proposition B.1 (Proposition 1 in Soare et al. (2014))**.** *Let $c = 2\sigma\sqrt{2}$. For every fixed sequence $\mathbf{z}_t$, with probability $1 - \delta$, for all $t > 0$ and for all $z \in \mathcal{Z}$, we have*

$$\left| z^\top \theta_\star - z^\top \hat{\theta}_t \right| \leq c \|z\|_{\mathbf{A}_t^{-1}} \sqrt{\log\left(\frac{6m^2 t^2 K^2}{\delta \pi}\right)} \ .$$

Our goal is to find the arm $z_\star$ that has the optimal expected reward $z_\star^\top \theta_\star$. In other words, we want to find an arm $z \in \mathcal{Z}$, such that for all $z' \in \mathcal{Z}$, $(z - z')^\top \theta_\star \geq 0$. However, one does not have access to $\theta_\star$, so we have to use its empirical estimate.

Let us consider a confidence set $\hat{S}(\mathbf{z}_t)$ centered at $\hat{\theta}_t \in \hat{S}(\mathbf{z}_t)$ and such that $\mathbb{P}\left(\theta_\star \notin \hat{S}(\mathbf{z}_t)\right) \leq \delta$, for some $\delta > 0$. Since $\theta_\star$ belongs to $\hat{S}(\mathbf{z}_t)$ with probability at least $1 - \delta$, one can stop pulling arms when an arm has been found, such that the above condition is verified for any $\theta \in \hat{S}(\mathbf{z}_t)$. More formally, the best arm identification task will be considered successful when an arm $z \in \mathcal{Z}$ will verify the following condition for any $z' \in \mathcal{Z}$ and any $\theta \in \hat{S}(\mathbf{z}_t)$:

$$(z - z')^\top (\hat{\theta}_t - \theta) \leq \hat{\Delta}_t(z, z') \ ,$$

where $\hat{\Delta}_t\left(z, z'\right) = \left(z - z'\right)^\top \hat{\theta}_t$ is the empirical gap between z and z'.

Using the upper bound in Proposition B.1, one way to ensure that $\mathbb{P}\left(\theta_\star \in \hat{S}(\mathbf{z}_t)\right) \geq 1 - \delta$ is to define the confidence set $\hat{S}(\mathbf{z}_t)$ as follows

$$\hat{S}\left(\mathbf{z}_t\right) = \left\{ \theta \in \mathbb{R}^d,\ \forall z \in \mathcal{Z},\ \forall z' \in \mathcal{Z}, \left(z - z'\right)^\top \left(\hat{\theta}_t - \theta\right) \leq c \|z - z'\|_{\left(\mathbf{A}_t\right)^{-1}} \sqrt{\log\left(\frac{6m^2t^2K^4}{\delta\pi}\right)} \right\} \ .$$

Then, the stopping condition can be reformulated as follows:

$$\exists z \in \mathcal{Z},\ \forall z' \in \mathcal{Z},\ c\|z - z'\|_{\mathbf{A}_t^{-1}} \sqrt{\log\left(\frac{6m^2t^2K^4}{\delta\pi}\right)} \leq \hat{\Delta}_t\left(z, z'\right) \ . \tag{S1}$$

C Estimation of the unknown parameter

C.1 Proof of Theorem 5.1

To prove Theorem 5.1, we first state some useful propositions and lemmas. For any finite set $X \subset \mathbb{R}^d$, we define the function $h_X : \mathcal{S}_X \to \mathbb{R} \cup \{+\infty\}$ as follows: for any $\lambda \in \mathcal{S}_X$,

$$h_X(\lambda) = \begin{cases} \max_{x' \in X} x'^\top \Sigma_X(\lambda)^{-1} x' & \text{if } \Sigma_X(\lambda) \text{ is invertible} \\ +\infty & \text{otherwise} \ . \end{cases}$$

Lemma C.1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite set spanning $\mathbb{R}^d$ and let $\mathcal{Z} = \{\mathrm{vec}\left(xx'^\top\right),\ (x, x') \in \mathcal{X}^2\}$. If $\mu^\star \in \mathcal{S}_\mathcal{X}$ is a minimizer of $h_\mathcal{X}$, then $\mu^\star$ is a solution of*

$$\min_{\mu \in \mathcal{S}_\mathcal{X}} \max_{z \in \mathcal{Z}} z^\top \left(\sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \mu_x \mu_{x'} \mathrm{vec}\left(xx'^\top\right) \mathrm{vec}\left(xx'^\top\right)^\top \right)^{-1} z \ .$$

Proof. First, let us notice that, for any $\mathcal{X} \subset \mathbb{R}^d$, one has $h_\mathcal{X} \geq 0$. Thus, $\mu^\star$ is also a minimizer of $h_\mathcal{X}^2$. In addition, $\mathcal{X}$ is spanning $\mathbb{R}^d$ so $h_\mathcal{X}(\mu^\star) < +\infty$. Developing $h_\mathcal{X}(\mu^\star)^2$ yields:

$$\begin{aligned}
h_\mathcal{X}(\mu^\star) \times h_\mathcal{X}(\mu^\star) &= \left(\max_{x \in \mathcal{X}} x^\top \Sigma_X(\mu^\star)^{-1} x \right) \times \left(\max_{x \in \mathcal{X}} x^\top \Sigma_X(\mu^\star)^{-1} x \right) \\
&= \max_{x \in \mathcal{X}} \max_{x' \in \mathcal{X}} x^\top \Sigma_X(\mu^\star)^{-1} x x'^\top \Sigma_X(\mu^\star)^{-1} x' \\
&= \max_{x \in \mathcal{X}} \max_{x' \in \mathcal{X}} \mathrm{vec}\left(xx'^\top\right)^\top \mathrm{vec}\left(\Sigma_X(\mu^\star)^{-1} x x'^\top \Sigma_X(\mu^\star)^{-1}\right) \\
&= \max_{x \in \mathcal{X}} \max_{x' \in \mathcal{X}} \mathrm{vec}\left(xx'^\top\right)^\top \left(\Sigma_X(\mu^\star)^{-1} \otimes \Sigma_X(\mu^\star)^{-1}\right) \mathrm{vec}\left(xx'^\top\right) \\
&= \max_{z \in \mathcal{Z}} z^\top \left(\Sigma_X(\mu^\star)^{-1} \otimes \Sigma_X(\mu^\star)^{-1}\right) z \ ,
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. We can now focus on the central term:

$$\begin{aligned}
\Sigma_X(\mu^\star)^{-1} \otimes \Sigma_X(\mu^\star)^{-1} &= \left(\sum_{x\in\mathcal{X}} \mu_x^\star x x^\top\right)^{-1} \otimes \left(\sum_{x\in\mathcal{X}} \mu_x^\star x x^\top\right)^{-1} \\
&= \left(\sum_{x\in\mathcal{X}} \mu_x^\star x x^\top \otimes \sum_{x\in\mathcal{X}} \mu_x^\star x x^\top\right)^{-1} \\
&= \left(\sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} \mu_x^\star \mu_{x'}^\star \left(x x^\top \otimes x' x'^\top\right)\right)^{-1} \\
&= \left(\sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} \mu_x^\star \mu_{x'}^\star \operatorname{vec}\left(x x'^\top\right) \operatorname{vec}\left(x x'^\top\right)^\top\right)^{-1} ,
\end{aligned}$$

and the result holds. □

Theorem C.2. *Let $\mu^\star \in \mathcal{S}_\mathcal{X}$ be a minimizer of $h_\mathcal{X}$. Let $\lambda^\star \in \mathcal{S}_\mathcal{Z}$ be the distribution defined from $\mu^\star$ such that, for all $z = \operatorname{vec}(x x'^\top)$, $\lambda_z^\star = \mu_x^\star \mu_{x'}^\star$. Then $\lambda^\star$ is a minimizer of $h_\mathcal{Z}$.*

Proof. From Kiefer and Wolfowitz (1960), we know that $\min_{\lambda\in\mathcal{S}_\mathcal{Z}} h_\mathcal{Z}(\lambda) = d^2$ and $\min_{\mu\in\mathcal{S}_\mathcal{X}} h_\mathcal{X}(\mu) = d$. Then, using Proposition C.1, one has

$$\begin{aligned}
d^2 &= h_\mathcal{X}(\mu^\star) \times h_\mathcal{X}(\mu^\star) \\
&= \max_{z\in\mathcal{Z}} z^\top \left(\sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} \mu_x^\star \mu_{x'}^\star \operatorname{vec}\left(x x'^\top\right) \operatorname{vec}\left(x x'^\top\right)^\top\right)^{-1} z \ .
\end{aligned}$$

This result implies that $h_\mathcal{Z}(\lambda^\star) = d^2$. Since $\min_{\lambda\in\mathcal{S}_\mathcal{Z}} h_Z(\lambda) = d^2$, $\lambda^\star$ is a minimizer of $h_\mathcal{Z}$.

□

C.2 Proof of Theorem 5.2

To prove our confidence bound, we need the two following proposition. The first one is from Tropp et al. (2015).

Proposition C.3 (Tropp et al. (2015), Chapter 5 and 6)**.** *Let $\mathbf{Z}_1, \dots, \mathbf{Z}_t$ be i.i.d. positive semi-definite random matrices in $\mathbb{R}^{d^2\times d^2}$, such that there exists $L > 0$ verifying $\mathbf{0} \preceq \mathbf{Z}_1 \preceq mL\mathbf{I}$. Let $\mathbf{A}_t$ be defined as $\mathbf{A}_t \triangleq \sum_{s=1}^t \mathbf{Z}_s$. Then, for any $0 < \varepsilon < 1$, one can lowerbound $\lambda_{\min}(\mathbf{A}_t)$ as follows:*

$$\mathbb{P}(\lambda_{\min}(\mathbf{A}_t) \le (1-\varepsilon)\lambda_{\min}(\mathbb{E}\mathbf{A}_t)) \le d^2 e^{-\frac{t\varepsilon^2\lambda_{\min}(\mathbb{E}\mathbf{Z}_1)}{2mL}} .$$

If in addition, there exists some $v > 0$, such that $\|\mathbb{E}\big[(\mathbf{Z}_1 - \mathbb{E}\mathbf{Z}_1)^2\big]\| \le v$, then for any $u > 0$, one has

$$\mathbb{P}\left(\|\mathbf{S}_t\| \ge u\right) \le 2d^2 e^{-\frac{u^2}{2mLu/3 + 2tv}},$$

From the second inequality, Rizk et al. (2019) derived a slightly different inequality that we use here :

Proposition C.4 (Rizk et al. (2019), Appendix A.3)**.** *Let $\mathbf{Z}_1, \dots, \mathbf{Z}_t$ be t i.i.d. random symmetric matrices in $\mathbb{R}^{d^2\times d^2}$ such that there exists $L > 0$ such that $\|\mathbf{Z}_1\| \le mL$, almost surely. Let $\mathbf{A}_t \triangleq \sum_{i=1}^t \mathbf{Z}_i$. Then, for any $u > 0$, one has:*

$$\mathbb{P}\left(\|\mathbf{A}_t - \mathbb{E}\mathbf{A}_t\| \ge \sqrt{2tvu} + \frac{mLu}{3}\right) \le d^2 e^{-u} \ .$$

where $v \triangleq \big\|\mathbb{E}\big[(\mathbf{Z}_1 - \mathbb{E}\mathbf{Z}_1)^2\big]\big\|$.

Finally, to prove our main theorem, we need the following lemma.

Lemma C.5. *One has* $\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\| \leq \frac{d^2}{\nu_{\min}}$, *where* $\nu_{\min}$ *is the smallest eigenvalue of the covariance matrix* $\frac{1}{K^2}\sum_{z\in\mathcal{Z}} z^\top z$.

Proof. Define $\mathcal{B} = \left\{z \in \mathbb{R}^{d^2} : \|z\| = 1\right\}$. First, for any semi-definite matrix $\mathbf{A} \in \mathbb{R}^{d^2\times d^2}$, we have $\|\mathbf{A}\| = \max_{z\in\mathcal{B}}\ z^\top \mathbf{A} z$. Because $\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}$ is positive definite and symmetric, and by Rayleigh-Ritz theorem,

$$\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\| = \max_{z\in\mathcal{B}}\ \frac{z^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z}{z^\top z} = \max_{z\in\mathcal{B}}\ z^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z\ .$$

Let $\mathbf{Z} \in \mathbb{R}^{K^2\times d^2}$ be the matrix whose rows are vectors of $\mathcal{Z}$ in an arbitrary order. Notice that $\mathcal{Z}$ spans $\mathbb{R}^{d^2}$, since $\mathcal{X}$ spans $\mathbb{R}^d$. Now for any $z \in \mathcal{B}$, define $\beta^{(z)} \in \mathbb{R}^{K^2}$ as a vector such that $z = \mathbf{Z}^\top\beta^{(z)}$. Then,

$$\begin{aligned}
\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\| &= \max_{z\in\mathcal{B}}\ {\beta^{(z)}}^\top\mathbf{Z}\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\mathbf{Z}^\top\beta^{(z)} \\
&= \max_{z\in\mathcal{B}}\ \sum_{i=1}^{d^2}\sum_{j=1}^{d^2}\beta_i^{(z)}\beta_j^{(z)}z_i^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z_j \\
&\leq \max_{z\in\mathcal{B}}\ \left\|\beta^{(z)}\right\|_1^2 \times \max_{i,j}\ z_i^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z_j\ .
\end{aligned}$$

Define $\tilde{z}_i = \Sigma_{\mathcal{Z}}(\lambda^\star)^{-\frac{1}{2}}z_i$. Clearly, $\max_{i,j}\ z_i^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z_j = \max_{i,j}\ \tilde{z}_i^\top\tilde{z}_j = \max_i\ \tilde{z}_i^2$. So we have

$$\begin{aligned}
\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\| &\leq \max_{z\in\mathcal{B}}\ \left\|\beta^{(z)}\right\|_1^2 \times \max_{z'\in\mathcal{Z}}\ z'^\top\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}z' \\
&\leq \max_{z\in\mathcal{B}}\ \left\|\beta^{(z)}\right\|_1^2 d^2\ .
\end{aligned}$$

The last inequality comes from Kiefer and Wolfowitz equivalence theorem (Kiefer and Wolfowitz, 1960). Now observe that $\beta^{(z)}$ can be obtained by least square regression : $\beta^{(z)} = \left(\mathbf{Z}\mathbf{Z}^\top\right)^{-1}\mathbf{Z}z = \left(\mathbf{Z}^\top\right)^\dagger z$ where $(\cdot)^\dagger$ is the Moore-Penrose pseudo-inverse. Note that $\mathbf{Z}\mathbf{Z}^\top$ is a Gram matrix. It is known that for a matrix having singular values $\{\sigma_i\}_i$, its pseudo-inverse has singular values $\begin{cases}\frac{1}{\sigma_i} & \text{if } \sigma_i \neq 0\\ 0 & \text{otherwise}\end{cases}$ for all i. So for $z \in \mathcal{B}$, we have:

$$\left\|\beta^{(z)}\right\|_1^2 \leq K^2\left\|\beta^{(z)}\right\|_2^2 \leq K^2\left\|\left(\mathbf{Z}^\top\right)^\dagger\right\|^2 \leq \frac{K^2}{\sigma_{\min}\left(\mathbf{Z}\right)^2}\ ,$$

where $\sigma_{\min}(\cdot)$ refers to the smallest singular value. Let $\nu_{\min}(\cdot)$ refer to the smallest eigenvalue. Noting that

$$\sigma_{\min}\left(\mathbf{Z}\right)^2 = \nu_{\min}\left(\mathbf{Z}^\top\mathbf{Z}\right) = K^2\nu_{\min}\left(\frac{1}{K^2}\sum_{z\in\mathcal{Z}}zz^\top\right)\ ,$$

yields the desired result.

□

We are now ready to state the bound on the random sampling error, relatively to the objective value $\Sigma_{\mathcal{Z}}(\lambda^\star)$ of the convex relaxation solution.

Theorem C.6. *Let* $\lambda^\star \in \mathcal{S}_{\mathcal{Z}}$ *be a minimizer of* $h_{\mathcal{Z}}$. *Let* $0 \leq \delta \leq 1$ *and let* $t_0 > 0$ *be such that*

$$t_0 = 2Ld^2\log(2d^2/\delta)/\nu_{\min}\ ,$$

where $L = \max_{z \in \mathcal{Z}} \|z\|^2$ *and* $\nu_{\min}$ *is the smallest eigenvalue of the covariance matrix* $\frac{1}{K^2} \sum_{z \in \mathcal{Z}} z^\top z$*. Then, at each round* $t \geq t_0$*, with probability at least* $1-\delta$*, the randomized G-allocation strategy for graphical bilinear bandit in Algorithm 2 produces a matrix* $\mathbf{A}_t$ *such that:*

$$h_{\mathcal{Z}}(\mathbf{A}_t) \leq (1+\alpha) h_{\mathcal{Z}}(mt \times \Sigma_{\mathcal{Z}}(\lambda^\star)) \ ,$$

where

$$\alpha = \frac{Ld^2}{m\nu_{\min}^2} \sqrt{\frac{2v}{t} \log\left(\frac{2d^2}{\delta}\right)} + o\left(\frac{1}{\sqrt{t}}\right),$$

and $v \triangleq \|\mathbb{E}\big[(\mathbf{A}_1 - \mathbb{E}\mathbf{A}_1)^2\big]\|$.

Proof. Let $(X_s^{(1)})_{s=1,\dots,t}, \dots, (X_s^{(n)})_{s=1,\dots,t}$ be nt i.i.d. random vectors in $\mathbb{R}^d$ such that for all $x \in \mathcal{X}$, $\mathbb{P}\left(X_1^{(1)} = x\right) = \mu_x^\star$. For $(i,j) \in E$ and $1 \leq s \leq t$, we define the random matrix $\mathbf{Z}_s^{(i,j)}$ by

$$\mathbf{Z}_s^{(i,j)} = \mathrm{vec}\left(X_s^i X_s^{j\top}\right) \mathrm{vec}\left(X_s^i X_s^{j\top}\right)^\top \ .$$

Finally, let us define for all $1 \leq s \leq t$, the edge-wise sum $\mathbf{Z}_s \in \mathbb{R}^{d^2 \times d^2}$, that is

$$\mathbf{Z}_s = \sum_{(i,j) \in E} \mathbf{Z}_s^{(i,j)} \ .$$

One can easily notice that $\mathbf{Z}_1, \dots, \mathbf{Z}_t$ are i.i.d. random matrices. We define the overall sum $\mathbf{A}_t = \sum_{s=1}^t \mathbf{Z}_s$ and our goal is to measure how close $f_{\mathcal{Z}}(\mathbf{A}_t)$ is to $f_{\mathcal{Z}}(mt \times \Sigma_{\mathcal{Z}}(\lambda^\star))$, where mt corresponds to the total number of sampled arms $z \in \mathcal{Z}$ during the t rounds of the learning procedure. By definition of $\mathbf{A}_t$, one has

$$\begin{aligned}
\max_{z \in \mathcal{Z}} z^\top \left(\mathbb{E}\mathbf{A}_t\right)^{-1} z &= \max_{z \in \mathcal{Z}} z^\top \left(\sum_{s=1}^t \sum_{(i,j) \in E} \mathbb{E}\left[\mathbf{Z}_s^{(i,j)}\right] \right)^{-1} z \\
&= \max_{z \in \mathcal{Z}} z^\top \left(\sum_{s=1}^t \sum_{(i,j) \in E} \sum_{x,x' \in \mathcal{X}} \mu_x^\star \mu_{x'}^\star \, \mathrm{vec}\left(xx'^\top\right) \mathrm{vec}\left(xx'^\top\right)^\top \right)^{-1} z \\
&= \max_{z \in \mathcal{Z}} z^\top \left(\sum_{s=1}^t \sum_{(i,j) \in E} \sum_{z' \in \mathcal{Z}} \lambda_{z'}^\star z' z'^\top \right)^{-1} z \\
&= f_{\mathcal{Z}}(mt \Sigma_{\mathcal{Z}}(\lambda^\star)) \ .
\end{aligned}$$

This allows us to bound the relative error as follows:

$$\begin{aligned}
\alpha &= \frac{f_{\mathcal{Z}}(\mathbf{A}_t)}{f_{\mathcal{Z}}(mt \times \Sigma_{\mathcal{Z}}(\lambda^\star))} - 1 \\
&= \frac{\max_{z \in \mathcal{Z}} \ z^\top \left(\mathbf{A}_t^{-1} - \left(\mathbb{E}\mathbf{A}_t\right)^{-1} + \left(\mathbb{E}\mathbf{A}_t\right)^{-1}\right) z}{f_{\mathcal{Z}}(mt \times \Sigma_{\mathcal{Z}}(\lambda^\star))} - 1 \\
&\leq \frac{\max_{z \in \mathcal{Z}} \ z^\top \left(\mathbf{A}_t^{-1} - \left(\mathbb{E}\mathbf{A}_t\right)^{-1}\right) z}{f_{\mathcal{Z}}(mt \times \Sigma_{\mathcal{Z}}(\lambda^\star))} \ .
\end{aligned}$$

Using the fact that $f_{\mathcal{Z}}(mt\Sigma_{\mathcal{Z}}(\lambda^\star)) = d^2/mt$ (Kiefer and Wolfowitz, 1960), we obtain

$$\begin{aligned}\alpha &\leq \frac{mt}{d^2} \times \max_{z\in\mathcal{Z}} z^\top \left(\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\right) z \\ &\leq \frac{mt}{d^2} \times \max_{z\in\mathcal{Z}} \|z\|^2 \|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\| \\ &\leq \frac{mtL}{d^2} \times \|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\| \ .\end{aligned}$$

Therefore, controlling the quantity $\|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\|$ will allow us to provide an upper bound on the relative error. Notice that

$$\begin{aligned}\|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\| &= \|\mathbf{A}_t^{-1}(\mathbb{E}\mathbf{A}_t - \mathbf{A}_t)(\mathbb{E}\mathbf{A}_t)^{-1}\| \\ &\leq \|\mathbf{A}_t^{-1}\| \ \|\mathbb{E}\mathbf{A}_t - \mathbf{A}_t\| \ \|(\mathbb{E}\mathbf{A}_t)^{-1}\| \ .\end{aligned}$$

Using Proposition C.3, we know that for any $d^2 e^{-\frac{t\lambda_{\min}(\mathbb{E}\mathbf{Z}_1)}{mL}} < \delta_h < 1$, the following holds:

$$\|\mathbf{A}_t^{-1}\| \leq \frac{\|(\mathbb{E}\mathbf{A}_t)^{-1}\|}{1 - \sqrt{\frac{2mL}{t}\|(\mathbb{E}\mathbf{Z}_1)^{-1}\|\log(d^2/\delta_h)}} \ ,$$

with probability at least $1-\delta_h$. Similarly, using Proposition C.4, for any $0 < \delta_b < 1$, we have

$$\|\mathbf{A}_t - \mathbb{E}\mathbf{A}_t\| \leq \frac{mL}{3}\log\frac{d^2}{\delta_b} + \sqrt{2tv^2\log\frac{d^2}{\delta_b}} \ ,$$

with probability at least $1-\delta_b$. Combining these two results with a union bound leads to the following bound, with probability $1-(\delta_b+\delta_h)$:

$$\left\|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\right\| \leq \left\|(\mathbb{E}\mathbf{A}_t)^{-1}\right\|^2 \frac{(mL/3)\log(d^2/\delta_b) + \sqrt{2tv\log(d^2/\delta_b)}}{1 - \sqrt{(2mL/t)\left\|(\mathbb{E}\mathbf{Z}_1)^{-1}\right\|\log(d^2/\delta_h)}} \ .$$

In order to obtain a unified bound depending on one confidence parameter $1-\delta$, one could optimize over δ_b and δ_h, subject to $\delta_b + \delta_h = \delta$. This leads to a messy result and a negligible improvement. One can use simple values $\delta_b = \delta_h = \delta/2$, so the overall bound becomes, with probability $1-\delta$:

$$\|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\| \leq \frac{1}{tm^2}\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\|^2 \sqrt{\frac{2v}{t}\log\left(\frac{2d^2}{\delta}\right)} \left(\frac{1 + \sqrt{\frac{m^2L^2\log(2d^2/\delta)}{18vt}}}{1 - \sqrt{\frac{2L\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\|\log(2d^2/\delta)}{t}}}\right) \ .$$

This can finally be formulated as follows:

$$\left\|\mathbf{A}_t^{-1} - (\mathbb{E}\mathbf{A}_t)^{-1}\right\| \leq \frac{1}{tm^2}\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\|^2 \sqrt{\frac{2v}{t}\log\left(\frac{2d^2}{\delta}\right)} + o\left(\frac{1}{t\sqrt{t}}\right) \ .$$

Using the obtained bound on $\|\mathbf{A}_t^{-1} - \mathbb{E}(\mathbf{A}_t)^{-1}\|$ yields

$$\begin{aligned}\frac{f_{\mathcal{Z}}(\mathbf{A}_t)}{f_{\mathcal{Z}}(mt\times\Sigma_{\mathcal{Z}}(\lambda^\star))} - 1 &\leq \frac{mtL}{d^2} \times \left(\frac{1}{tm^2}\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\|^2 \sqrt{\frac{2v}{t}\log\left(\frac{2d^2}{\delta}\right)} + o\left(\frac{1}{t\sqrt{t}}\right)\right) \\ &\leq \frac{L}{md^2}\left\|\Sigma_{\mathcal{Z}}(\lambda^\star)^{-1}\right\|^2 \sqrt{\frac{2v}{t}\log\left(\frac{2d^2}{\delta}\right)} + o\left(\frac{1}{\sqrt{t}}\right) \ ,\end{aligned}$$

By noticing that $f_{\mathcal{Z}}(mt\times\Sigma_{\mathcal{Z}}(\lambda^\star)) \leq f_{\mathcal{Z}}(\mathbf{A}_t^\star)$ and by using Lemma C.5, the result holds. □

D Variance analysis

Star graph. The covariance matrix of the star graph can be bounded as follows:

$$\mathrm{Var}(\mathbf{A}_1) \preceq m \times P \cdot \mathbf{I} + (n_{\mathrm{S}} - 1)(n_{\mathrm{S}} - 2)M \cdot \mathbf{I} + n_{\mathrm{S}}(n_{\mathrm{S}} - 1)N \cdot \mathbf{I} \ .$$

Since the star graph of m edges has a number of nodes $n_{\mathrm{S}} = m/2 + 1$, we have

$$\|\mathrm{Var}(\mathbf{A}_1)\| \leq m \times P + (M + N) \times O\left(m^2\right) \ .$$

Complete graph. As for the star graph,

$$\mathrm{Var}(\mathbf{A}_1) \preceq m \times P \cdot \mathbf{I} + n_{\mathrm{Co}}(n_{\mathrm{Co}} - 1)(n_{\mathrm{Co}} - 2)M \cdot \mathbf{I} + n_{\mathrm{Co}}(n_{\mathrm{Co}} - 1)(n_{\mathrm{Co}} - 1)N \cdot \mathbf{I} \ .$$

Since the complete graph of m edges has a number of nodes $n_{\mathrm{Co}} = \left(1 + \sqrt{4m + 1}\right)/2$, we have

$$\|\mathrm{Var}(\mathbf{A}_1)\| \leq m \times P + (M + N) \times O\left(m\sqrt{m}\right) \ .$$

Circle graph. Again,

$$\mathrm{Var}(\mathbf{A}_1) \preceq m \times P \cdot \mathbf{I} + 2n_{\mathrm{Ci}}M \cdot \mathbf{I} + 4n_{\mathrm{Ci}}N \cdot \mathbf{I} \ .$$

Since the circle graph of m edges has a number of nodes $n_{\mathrm{Ci}} = m/2$, we have

$$\|\mathrm{Var}(\mathbf{Z}_1)\| \leq m \times P + (M + N) \times O\left(m\right) \ .$$

Matching graph. Finally,

$$\mathrm{Var}(\mathbf{A}_1) \preceq m \times P \cdot \mathbf{I} + n_{\mathrm{M}}N \cdot \mathbf{I} \ .$$

Since the matching graph of m edges has a number of nodes $n_{\mathrm{M}} = m$, we have

$$\|\mathrm{Var}(\mathbf{A}_1)\| \leq m \times P + m \times N \ .$$

E Generalization

In this section, we provide some insights into the generalization to broader reward settings.

E.1 When $\mathbf{M}_\star$ is not symmetric

Consider the same graphical bilinear bandit setting as the one explained in the paper with the only difference that $\mathbf{M}_\star$ is not symmetric. We recall here that in the graph $\mathcal{G} = (V, E)$ associated to the graphical bilinear bandit setting, $(i, j) \in E$ if and only if $(j, i) \in E$. Hence, for a given allocation $(x^{(1)}, \ldots, x^{(n)}) \in \mathcal{X}^n$, one can write the associated expected global reward as follows :

$$
\begin{aligned}
\sum_{(i,j)\in E} x^{(i)\top}\mathbf{M}_\star x^{(j)} &= \sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}} x^{(i)\top}\mathbf{M}_\star x^{(j)} + x^{(j)\top}\mathbf{M}_\star x^{(i)} \\
&= \sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}} x^{(i)\top}\mathbf{M}_\star x^{(j)} + \left(x^{(j)\top}\mathbf{M}_\star x^{(i)}\right)^\top \\
&= \sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}} x^{(i)\top}\mathbf{M}_\star x^{(j)} + x^{(i)\top}\mathbf{M}_\star^\top x^{(j)} \\
&= \sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}} x^{(i)\top}\left(\mathbf{M}_\star x^{(j)} + \mathbf{M}_\star^\top x^{(j)}\right) \\
&= \sum_{i=1}^{n}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}} x^{(i)\top}\left(\mathbf{M}_\star + \mathbf{M}_\star^\top\right) x^{(j)} \ .
\end{aligned}
$$

Let us denote $\bar{\mathbf{M}}_\star = \mathbf{M}_\star + \mathbf{M}_\star^\top$. One can notice that $\bar{\mathbf{M}}_\star$ is symmetric. Solving the graphical bilinear bandit with the matrix $\bar{\mathbf{M}}_\star$ is exactly what we propose throughout the main paper.

E.2 When the reward captures more information than the interactions between agents

Consider the real world problems introduced in the paper, but with the difference that instead of a reward only related to the interaction between two neighboring agents/nodes, there is an additional term that informs about the absolute quality of the arm chosen by the agent itself. More formally we consider the following reward $r_t^{(i,j)}$ for the node i:

$$
r_t^{(i,j)} = x_t^{(i)\top}\mathbf{M}_\star x_t^{(j)} + x_t^{(i)\top}\beta_\star + \eta_t^{(i,j)} \ .
$$

where $\beta_\star \in \mathbb{R}^d$ is a second unknown parameter that allows to capture the quality of the arm chosen by the node i independently of its neighbors.

In order to add a constant term in the reward, let us construct the set $\tilde{\mathcal{X}} \subset \mathbb{R}^{d+1}$ such that each arm $x \in \mathcal{X}$ is associated to a new arm $\tilde{x} \in \tilde{\mathcal{X}}$ defined as $\tilde{x}^\top = (x^\top, 1)$. Moreover, let us define the matrix $\tilde{\mathbf{M}}^\star \in \mathbb{R}^{(d+1)\times(d+1)}$ as follows:

$$
\tilde{\mathbf{M}}_\star = \begin{pmatrix} \begin{bmatrix} & \mathbf{M}_\star & \end{bmatrix} \begin{bmatrix} \\ \beta_\star \\ \end{bmatrix} \\ \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \end{pmatrix} \ .
$$

One can easily verify that for any edge $(i,j) \in E$ and any time step t, the reward $r_t^{(i,j)}$ can now be written as follows:

$$
r_t^{(i,j)} = \tilde{x}_t^{(i)\top}\tilde{\mathbf{M}}_\star\tilde{x}_t^{(j)} + \eta_t^{(i,j)} \ ,
$$

which leads to the same graphical bilinear bandit setting explained in Section 3, this time in dimension $d+1$ instead of d. Hence, all the previous results hold for this more general graphical bilinear bandit problem, provided any dependence in d is modified to $d+1$.

F Computing $\mu^\star$

In Algorithm 2, we need to find the solution $\mu_\star$ of $\min_{\mu\in\mathcal{S}_\mathcal{X}} h_\mathcal{X}(\mu)$. In fact we need $\mu_\star$ to sample from it. We show that for any set X, the function h_X is convex and we use the Frank-Wolfe algorithm (Frank et al., 1956) to compute $\mu_\star$ and $\lambda_\star$. The convergence of the algorithm has been proven in Damla Ahipasaoglu et al. (2008). Note that one can only compute $\mu_\star$ or $\lambda_\star$ to obtain the other one thanks to C.2.

Proposition F.1. *Let $d > 0$, for any set $X \subset \mathbb{R}^d$, h_X is convex.*

Proof. Let $(\lambda, \lambda') \in \mathcal{S}_X^2$ be two distributions in $\mathcal{S}_X$. If either $\Sigma_X(\lambda)$ or $\Sigma_X(\lambda')$ are not invertible, then for any $t \in [0, 1]$ one has

$$h_X(t\lambda + (1-t)\lambda') \leq t h_X(\lambda) + (1-t)h_X(\lambda') = +\infty \ .$$

Otherwise, for $t \in [0, 1]$, we define the positive definite matrix $\mathbf{Z}(t) \in \mathbb{R}^{d\times d}$ as follows:

$$\mathbf{Z}(t) = t\Sigma_X(\lambda) + (1-t)\Sigma_X(\lambda') \ .$$

Simple linear algebra (Petersen and Pedersen, 2012) yields

$$\frac{\partial \mathbf{Z}(t)^{-1}}{\partial t} = \mathbf{Z}(t)^{-1}\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1} \ .$$

Using this result and the fact that $\partial^2\mathbf{Z}(t)/\partial t^2 = 0$, we obtain

$$\frac{\partial^2 \mathbf{Z}(t)^{-1}}{\partial t^2} = 2\mathbf{Z}(t)^{-1}\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1}\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1} \ .$$

Therefore, for any $x \in X$,

$$\begin{aligned}
\frac{\partial^2 x^\top \mathbf{Z}(t)^{-1} x}{\partial t^2} &= 2x^\top \mathbf{Z}(t)^{-1}\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1}\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1}x \\
&= 2\left(\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1}x\right)^\top \mathbf{Z}(t)^{-1}\left(\frac{\partial \mathbf{Z}(t)}{\partial t}\mathbf{Z}(t)^{-1}x\right) \\
&\geq 0 \ ,
\end{aligned}$$

which shows convexity for any fixed $x \in X$. The final results yields from the fact that h_X is a maximum over convex functions.

□

G Additional experiment and information

We define the set of arms $\mathcal{X} \subset \mathbb{R}^5$ that is made of $|\mathcal{X}| = 100$ node-arms randomly sampled from a multivariate 5-dimensional Gaussian distribution $\mathcal{N}(0, I)$ and then normalized so that $\|x\| = 1$ for all $x \in \mathcal{X}$. In all the figures the results are averaged over 100 random repetitions of the experiments.

We propose to validate our insight and compute the evolution of $\|\mathrm{Var}\left(\mathbf{A}_1\right)\|$ for the three types of graphs (star, complete and circle) and different number of edges. The results are shown in Figure 1. One can notice that we retrieve the $O(m^2)$ dependence of the variance for the star graph, the $O(m\sqrt{m})$ for the complete graph and the linear dependence $O(m)$ for the circle graph.

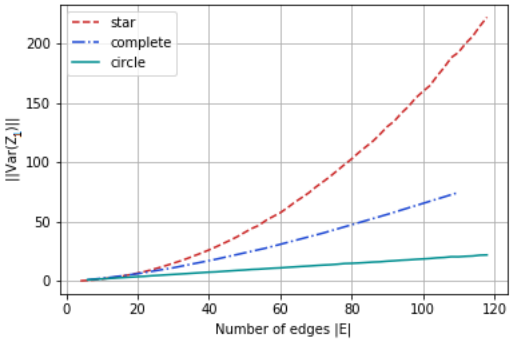

Figure 1: Evolution of the variance according to the number of edges and the type of graph (star, complete, circle), the variance being averaged over 100 repetitions.

Machine used for all the experiments. Intel(R) Xeon(R) CPU E5-2667 v4 @ 3.20GHz - 24 CPUs used.

Bibliography

Damla Ahipasaoglu, S., Sun, P., and Todd, M. J. (2008). Linear convergence of a modified frank–wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimisation Methods and Software*, 23(1):5–19.

Frank, M., Wolfe, P., et al. (1956). An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Petersen, K. B. and Pedersen, M. S. (2012). The matrix cookbook, nov 2012. *URL http://www2. imm. dtu. dk/pubdb/p. php*, 3274:14.

Rizk, G., Colin, I., Thomas, A., and Draief, M. (2019). Refined bounds for randomized experimental design. *NeurIPS Workshop on Machine Learning with Guarantees*.

Soare, M., Lazaric, A., and Munos, R. (2014). Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836.

Tropp, J. A. et al. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230.