Supplementary Material to "Online Clustering of Bandits"

Claudio Gentile CLAUDIO.GENTILE@UNINSUBRIA.IT
DiSTA, University of Insubria, Italy

Shuai Li SHUAILI.SLI@GMAIL.COM
DiSTA, University of Insubria, Italy

Giovanni Zappella ZAPPELLA@AMAZON.COM
Amazon Development Center Germany, Germany
(Work done when the author was PhD student at Univeristy of Milan)

Abstract

This supplementary material contains all proofs and technical details omitted from the main text, along with ancillary comments, discussion about related work, and extra experimental results.

1. Proof of Theorem 1

The following sequence of lemmas are of preliminary importance. The first one needs extra variance conditions on the process X generating the context vectors.

We find it convenient to introduce the node counterpart to $\mathrm{TCB}_{j,t-1}(\boldsymbol{x})$, and the cluster counterpart to $\widetilde{\mathrm{TCB}}_{i,t-1}$. Given round t, node $i \in V$, and cluster index $j \in \{1,\ldots,m_t\}$, we let

$$\mathrm{TCB}_{i,t-1}(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^\top M_{i,t-1}^{-1}\boldsymbol{x}}\,\left(\sigma\sqrt{2\,\log\frac{|M_{i,t-1}|}{\delta/2}}+1\right)$$

$$\widetilde{\mathrm{TCB}}_{j,t-1} = \frac{\sigma\sqrt{2d\,\log t + 2\log(2/\delta)}+1}{\sqrt{1+A_\lambda(\bar{T}_{j,t-1}, \delta/(2^{m+1}d))}}\ ,$$

being

$$\bar{T}_{j,t-1} = \sum_{i\in\hat{V}_{j,t}} T_{i,t-1} = |\{s \le t-1\ :\ i_s \in \hat{V}_{j,t}\}|\ ,$$

i.e., the number of past rounds where a node lying in cluster $\hat{V}_{j,t}$ was served. From a notational standpoint, notice the difference[1] between $\widetilde{\mathrm{TCB}}_{i,t-1}$ and $\mathrm{TCB}_{i,t-1}(\boldsymbol{x})$, both referring to a single node $i \in V$, and $\widetilde{\mathrm{TCB}}_{j,t-1}$ and $\mathrm{TCB}_{j,t-1}(\boldsymbol{x})$ which refer to an aggregation (cluster) of nodes j among the available ones at time t.

Lemma 1. *Let, at each round t, context vectors $C_{i_t} = \{\boldsymbol{x}_{t,1},\ldots,\boldsymbol{x}_{t,c_t}\}$ being generated i.i.d. (conditioned on i_t, c_t and all past indices $i_1,\ldots,i_{t-1}$, rewards $a_1,\ldots,a_{t-1}$, and sets $C_{i_1},\ldots,C_{i_{t-1}}$) from a random process X such that $||X|| = 1$, $\mathbb{E}[XX^\top]$ is full rank, with minimal eigenvalue $\lambda > 0$. Let also, for any fixed unit vector $\boldsymbol{z} \in \mathbb{R}^d$, the random variable $(\boldsymbol{z}^\top X)^2$ be (conditionally) sub-Gaussian with variance parameter*[2]

$$\nu^2 = \mathbb{V}_t\big[(\boldsymbol{z}^\top X)^2\,|\,c_t\,\big] \le \frac{\lambda^2}{8\log(4c_t)} \quad \forall t\ .$$

Then

$$\mathrm{TCB}_{i,t}(\boldsymbol{x}) \le \widetilde{\mathrm{TCB}}_{i,t}$$

holds with probability at least $1-\delta/2$, uniformly over $i \in V$, $t = 0,1,2\ldots$, and $\boldsymbol{x} \in \mathbb{R}^d$ such that $||\boldsymbol{x}|| = 1$.

Proof. Fix node $i \in V$ and round t. By the very way the algorithm in Figure 1 is defined, we have

$$M_{i,t} = I + \sum_{s\le t\,:\,i_s=i} \bar{\boldsymbol{x}}_s\bar{\boldsymbol{x}}_s^\top = I + S_{i,t}\ .$$

First, notice that by standard arguments (e.g., (Dekel et al., 2010)) we have

$$\log|M_{i,t}| \le d\,\log(1+T_{i,t}/d) \le d\,\log(1+t)\ .$$

Moreover, denoting by $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ the maximal and the minimal eigenvalue of the matrix at argument we

[1] Also observe that $2nd$ has been replaced by $2^{m+1}d$ inside the log's.

[2] Random variable $(\boldsymbol{z}^\top X)^2$ is conditionally sub-Gaussian with variance parameter $\sigma^2 > 0$ when $\mathbb{E}_t\big[\exp(\gamma\,(\boldsymbol{z}^\top X)^2)|\,c_t\,\big] \le \exp\big(\sigma^2\gamma^2/2\big)$ for all $\gamma \in \mathbb{R}$. The sub-Gaussian assumption can be removed here at the cost of assuming the conditional variance of $(\boldsymbol{z}^\top X)^2$ scales with c_t like $\frac{\lambda^2}{c_t}$, instead of $\frac{\lambda^2}{\log(c_t)}$.

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

have that, for any fixed unit norm $\boldsymbol{x} \in \mathbb{R}^d$,

$$\boldsymbol{x}^\top M_{i,t}^{-1} \boldsymbol{x} \le \lambda_{\max}(M_{i,t}^{-1}) = \frac{1}{1+\lambda_{\min}(S_{i,t})} .$$

Hence, we want to show with probability at least $1-\delta/(2n)$ such that

$$\begin{aligned}\lambda_{\min}(S_{i,t}) \ge \lambda T_{i,t}/4 - 8\log\left(\frac{T_{i,t}+3}{\delta/(2nd)}\right) \\ - 2\sqrt{T_{i,t}\log\left(\frac{T_{i,t}+3}{\delta/(2nd)}\right)}\end{aligned} \tag{1}$$

holds for any fixed node i. To this end, fix a unit norm vector $\boldsymbol{z} \in \mathbb{R}^d$, a round $s \le t$, and consider the variable

$$\begin{aligned}V_s &= \boldsymbol{z}^\top \left(\bar{\boldsymbol{x}}_s \bar{\boldsymbol{x}}_s^\top - \mathbb{E}_s[\bar{\boldsymbol{x}}_s \bar{\boldsymbol{x}}_s^\top \,|\, c_s]\right) \boldsymbol{z} \\ &= (\boldsymbol{z}^\top \bar{\boldsymbol{x}}_s)^2 - \mathbb{E}_s[(\boldsymbol{z}^\top \bar{\boldsymbol{x}}_s)^2 \,|\, c_s] .\end{aligned}$$

The sequence $V_1, V_2, \ldots, V_{T_{i,t}}$ is a martingale difference sequence, with optional skipping, where $T_{i,t}$ is a stopping time.[3] Moreover, the following claim holds.

Claim 1. *Under the assumption of this lemma,*

$$\mathbb{E}_s[(\boldsymbol{z}^\top \bar{\boldsymbol{x}}_s)^2 \,|\, c_s] \ge \lambda/4 .$$

Proof of claim. Let[4] in round s the context vectors be $C_{i_s} = \{\boldsymbol{x}_{s,1}, \ldots, \boldsymbol{x}_{s,c_s}\}$, and consider the corresponding i.i.d. random variables $Z_i = (\boldsymbol{z}^\top \boldsymbol{x}_{s,i})^2 - \mathbb{E}_s[(\boldsymbol{z}^\top \boldsymbol{x}_{s,i})^2 \,|\, c_s]$, $i = 1, \ldots, c_s$. Since by assumption these variables are (zero-mean) sub-Gaussian, we have that (see, e.g., (Massart, 2007)[Ch.2])

$$\mathbb{P}_s\left(Z_i < -a \,|\, c_t\right) \le \mathbb{P}_s\left(|Z_i| > a \,|\, c_t\right) \le 2e^{-a^2/2\nu^2}.$$

holds for any i, where $\mathbb{P}_s(\cdot)$ is the shorthand for the conditional probability

$$\mathbb{P}\left(\,\cdot\,\middle|\,(i_1, C_{i_1}, a_1), \ldots, (i_{s-1}, C_{i_{s-1}}, a_{s-1}), i_s\right) .$$

The above implies

$$\begin{aligned}\mathbb{P}_s\left(\min_{i=1,\ldots,c_s} (\boldsymbol{z}^\top \boldsymbol{x}_{s,i})^2 \ge \lambda - a \,\middle|\, c_t\right) \\ \ge \left(1 - 2e^{-a^2/2\nu^2}\right)^{c_s} .\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}_s[(\boldsymbol{z}^\top \bar{\boldsymbol{x}}_s)^2 \,|\, c_s] &\ge \mathbb{E}_s\left[\min_{i=1,\ldots,c_s} (\boldsymbol{z}^\top \boldsymbol{x}_{s,i})^2 \,\middle|\, c_s\right] \\ &\ge (\lambda - a)\left(1 - 2e^{-a^2/2\nu^2}\right)^{c_s} .\end{aligned}$$

Since this holds for all $a \in \mathbb{R}$, we set $a = \sqrt{2\nu^2 \log(4c_s)}$ to get $\left(1 - 2e^{-a^2/2\nu^2}\right)^{c_s} = (1 - \frac{1}{2c_s})^{c_s} \ge 1/2$ (because $c_s \ge 1$), and $\lambda - a \ge \lambda/2$ (because of the assumption on ν^2). Putting together concludes the proof of the claim. □

We are now in a position to apply a Freedman-like inequality for matrix martingales due to (Oliveira, 2010; Tropp, 2011) to the (matrix) martingale difference sequence

$$\mathbb{E}_1[\bar{\boldsymbol{x}}_1 \bar{\boldsymbol{x}}_1^\top \,|\, c_1] - \bar{\boldsymbol{x}}_1 \bar{\boldsymbol{x}}_1^\top,\ \mathbb{E}_2[\bar{\boldsymbol{x}}_2 \bar{\boldsymbol{x}}_2^\top \,|\, c_2] - \bar{\boldsymbol{x}}_2 \bar{\boldsymbol{x}}_2^\top, \ldots$$

with optional skipping. Setting for brevity $X_s = \bar{\boldsymbol{x}}_s \bar{\boldsymbol{x}}_s^\top$, and

$$W_t = \sum_{s \le t \,:\, i_s = i} \left(\mathbb{E}_s[X_s^2 \,|\, c_s] - \mathbb{E}_s^2[X_s \,|\, c_s]\right) ,$$

Theorem 1.2 in (Tropp, 2011) implies

$$\begin{aligned}\mathbb{P}\Big(\exists t \,:\, \lambda_{\min}\left(S_{i,t}\right) \le T_{i,t}\lambda_{\min}(\mathbb{E}_1[X_1 \,|\, c_1]) - a, ||W_t|| \le \sigma^2\Big) \\ \le d\, e^{-\frac{a^2/2}{\sigma^2 + 2a/3}} .\end{aligned} \tag{2}$$

where $||W_t||$ denotes the operator norm of matrix W_t.

We apply Claim 1, so that $\lambda_{\min}(\mathbb{E}_1[X_1 \,|\, c_1]) \ge \lambda/4$, and proceed as in, e.g., (Cesa-Bianchi & Gentile, 2008). We set for brevity $A(x, \delta) = 2\log\frac{(x+1)(x+3)}{\delta}$, and $f(A, r) = 2A + \sqrt{Ar}$. We can write

$$\begin{aligned}&\mathbb{P}\Big(\exists t \,:\, \lambda_{\min}(S_{i,t}) \le \lambda_{\min}T_{i,t}/4 - f(A(||W_t||, \delta), ||W_t||)\Big) \\ &\le \sum_{r=0}^{\infty} \mathbb{P}\Big(\exists t \,:\, \lambda_{\min}(S_{i,t}) \le \lambda_{\min}T_{i,t}/4 - f(A(r, \delta), r), \\ &\qquad\qquad \lfloor ||W_t|| \rfloor = r\Big) \\ &\le \sum_{r=0}^{\infty} \mathbb{P}\Big(\exists t \,:\, \lambda_{\min}\left(S_{i,t}\right) \le \lambda_{\min}T_{i,t}/4 - f(A(r, \delta), r), \\ &\qquad\qquad ||W_t|| \le r+1\Big) \\ &\le d \sum_{r=0}^{\infty} e^{-\frac{f^2(A(r,\delta),r)/2}{r+1+2f(A(r,\delta),r)/3}} ,\end{aligned}$$

the last inequality deriving from (2). Because $f(A, r)$ satisfies $f^2(A, r) \ge Ar + A + \frac{2}{3}f(A, r)A$, we have that the exponent in the last exponential is at least $A(r, \delta)/2$, implying

$$\sum_{r=0}^{\infty} e^{-A(r,\delta)/2} = \sum_{r=0}^{\infty} \frac{\delta}{(r+1)(r+3)} < \delta$$

which, in turn, yields

$$\begin{aligned}\mathbb{P}\Big(\exists t \,:\, \lambda_{\min}(S_{i,t}) \le T_{i,t}\lambda_{\min}/4 \\ - f(A(||W_t||, \delta/d), ||W_t||)\Big) \\ \le \delta .\end{aligned}$$

[3] More precisely, we are implicitly considering the sequence $\eta_{i,1}V_1, \eta_{i,2}V_2, \ldots, \eta_{i,t}V_t$, where $\eta_{i,s} = 1$ if $i_s = i$, and 0 otherwise, with $T_{i,t} = \sum_{s=1}^{t} \eta_{i,s}$.

[4] This proof is based on standard arguments, and is reported here for the sake of completeness.

Finally, observe that

$$\begin{aligned}||W_t|| &\le \sum_{s\le t\,:\,i_s=i} ||\mathbb{E}_s[X_s^2\,|\,c_s]|| \\ &= \sum_{s\le t\,:\,i_s=i} ||\mathbb{E}_s[X_s\,|\,c_s]|| \\ &\le \sum_{s\le t\,:\,i_s=i} \mathbb{E}_s[||X_s\,|\,c_s||] \\ &\le T_{i,t}\ .\end{aligned}$$

Therefore we conclude

$$\begin{aligned}\mathbb{P}\Big(\forall t\,:\,\lambda_{\min}(S_{i,t}) \ge \lambda_{\min}T_{i,t}/4 - f(A(T_{i,t},\delta/d),T_{i,t})\Big)\\ \ge 1-\delta\ .\end{aligned}$$

Stratifying over $i \in V$, replacing δ by $\delta/(2n)$ in the last inequality, and overapproximating proves the lemma. □

Lemma 2. *Under the same assumptions as in Lemma 1, we have*

$$||\boldsymbol{u}_i - \boldsymbol{w}_{i,t}|| \le \widetilde{\mathrm{TCB}}_{i,t}$$

holds with probability at least $1-\delta$, uniformly over $i \in V$, and $t = 0,1,2,\ldots$.

Proof. From (Abbasi-Yadkori et al., 2011) it follows that

$$|\boldsymbol{u}_i^\top \boldsymbol{x} - \boldsymbol{w}_{i,t}^\top \boldsymbol{x}| \le \mathrm{TCB}_{i,t}(\boldsymbol{x})$$

holds with probability at least $1-\delta/2$, uniformly over $i \in V$, $t = 0,1,2,\ldots$ and $\boldsymbol{x} \in \mathbb{R}^d$. Hence,

$$\begin{aligned}||\boldsymbol{u}_i - \boldsymbol{w}_{i,t}|| &\le \max_{\boldsymbol{x}\in\mathbb{R}^d\,:\,||\boldsymbol{x}||=1} |\boldsymbol{u}_i^\top \boldsymbol{x} - \boldsymbol{w}_{i,t}^\top \boldsymbol{x}| \\ &\le \max_{\boldsymbol{x}\in\mathbb{R}^d\,:\,||\boldsymbol{x}||=1} \mathrm{TCB}_{i,t}(\boldsymbol{x}) \\ &\le \widetilde{\mathrm{TCB}}_{i,t}\ ,\end{aligned}$$

the last inequality holding with probability $\ge 1-\delta/2$ by Lemma 1. This concludes the proof. □

Lemma 3. *Under the same assumptions as in Lemma 1:*

1. *If $||\boldsymbol{u}_i - \boldsymbol{u}_j|| \ge \gamma$ and $\widetilde{\mathrm{TCB}}_{i,t} + \widetilde{\mathrm{TCB}}_{j,t} < \gamma/2$ then*

$$||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| > \widetilde{\mathrm{TCB}}_{i,t} + \widetilde{\mathrm{TCB}}_{j,t}$$

holds with probability at least $1-\delta$, uniformly over $i,j \in V$ and $t = 0,1,2,\ldots$;

2. *if $||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| > \widetilde{\mathrm{TCB}}_{i,t} + \widetilde{\mathrm{TCB}}_{j,t}$ then*

$$||\boldsymbol{u}_i - \boldsymbol{u}_j|| \ge \gamma$$

holds with probability at least $1-\delta$, uniformly over $i,j \in V$ and $t = 0,1,2,\ldots$.

Proof. 1. We have

$$\begin{aligned}\gamma &\le ||\boldsymbol{u}_i - \boldsymbol{u}_j|| \\ &= ||\boldsymbol{u}_i - \boldsymbol{w}_{i,t} + \boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t} + \boldsymbol{w}_{j,t} - \boldsymbol{u}_j|| \\ &\le ||\boldsymbol{u}_i - \boldsymbol{w}_{i,t}|| + ||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| + ||\boldsymbol{w}_{j,t} - \boldsymbol{u}_j|| \\ &\le \widetilde{\mathrm{TCB}}_{i,t} + ||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| + \widetilde{\mathrm{TCB}}_{j,t} \\ &\quad \text{(from Lemma 2)} \\ &\le ||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| + \gamma/2,\end{aligned}$$

i.e., $||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| \ge \gamma/2 > \widetilde{\mathrm{TCB}}_{i,t} + \widetilde{\mathrm{TCB}}_{j,t}$.

2. Similarly, we have

$$\begin{aligned}\widetilde{\mathrm{TCB}}_{i,t} + \widetilde{\mathrm{TCB}}_{j,t} &< ||\boldsymbol{w}_{i,t} - \boldsymbol{w}_{j,t}|| \\ &\le ||\boldsymbol{u}_i - \boldsymbol{w}_{i,t}|| + ||\boldsymbol{u}_i - \boldsymbol{u}_j|| \\ &\quad + ||\boldsymbol{w}_{j,t} - \boldsymbol{u}_j|| \\ &\le \widetilde{\mathrm{TCB}}_{i,t} + ||\boldsymbol{u}_i - \boldsymbol{u}_j|| + \widetilde{\mathrm{TCB}}_{j,t}\ ,\end{aligned}$$

implying $||\boldsymbol{u}_i - \boldsymbol{u}_j|| > 0$. By the well-separatedness assumption, it must be the case that $||\boldsymbol{u}_i - \boldsymbol{u}_j|| \ge \gamma$.

□

From Lemma 3, it follows that if any two nodes i and j belong to different true clusters and the upper confidence bounds $\widetilde{\mathrm{TCB}}_{i,t}$ and $\widetilde{\mathrm{TCB}}_{j,t}$ are both small enough, then it is very likely that edge (i,j) will get deleted by the algorithm (Lemma 3, Item 1). Conversely, if the algorithm deletes an edge (i,j), then it is very likely that the two involved nodes i and j belong to different true clusters (Lemma 3, Item 2). Notice that, we have $E \subseteq E_t$ with high probability for all t. Because the clusters $\hat{V}_{1,t},\ldots,\hat{V}_{m_t,t}$ are induced by the connected components of $G_t = (V,E_t)$, every true cluster V_i must be entirely included (with high probability) in some cluster $\hat{V}_{j,t}$. Said differently, for all rounds t, the partition of V produced by $V_1,\ldots,V_m$ is likely to be a refinement of the one produced by $\hat{V}_{1,t},\ldots,\hat{V}_{m_t,t}$ (in passing, this also shows that, with high probability, $m_t \le m$ for all t). This is a key property to all our analysis. See Figure 2 in the main text for reference.

Lemma 4. *Under the same assumptions as in Lemma 1, if $\widehat{j}_t$ is the index of the current cluster node i_t belongs to, then we have*

$$\mathrm{TCB}_{\widehat{j}_t,t-1}(\boldsymbol{x}) \le \widetilde{\mathrm{TCB}}_{\widehat{j}_t,t-1}$$

holds with probability at least $1-\delta/2$, uniformly over all rounds $t = 1,2,\ldots$, and $\boldsymbol{x} \in \mathbb{R}^d$ such that $||\boldsymbol{x}|| = 1$.

Proof. The proof is the same as the one of Lemma 1, except that at the very end we need to stratify over all possible shapes for cluster $\hat{V}_{\widehat{j}_t,t}$, rather than over the n nodes. Now, since with high probability (Lemma 3), $\hat{V}_{\widehat{j}_t,t}$ is the union of true clusters, the set of all such unions is with the same probability upper bounded by 2^m. □

The next lemma is a generalization of Theorem 1 in (Abbasi-Yadkori et al., 2011), and shows a convergence result for aggregate vector $\bar{\boldsymbol{w}}_{j,t-1}$.

Lemma 5. *Let t be any round, and assume the partition of V produced by true clusters $V_1, \ldots, V_m$ is a refinement of the one produced by the current clusters $\hat{V}_{1,t}, \ldots, \hat{V}_{m_t,t}$. Let $j = \widehat{j}_t$ be the index of the current cluster node i_t belongs to. Let this cluster be the union of true clusters $V_{j_1}, V_{j_2}, \ldots, V_{j_k}$, associated with (distinct) parameter vectors $\boldsymbol{u}_{j_1}, \boldsymbol{u}_{j_2}, \ldots, \boldsymbol{u}_{j_k}$, respectively. Define*

$$\bar{\boldsymbol{u}}_t = \bar{M}_{j,t-1}^{-1}\left(\sum_{\ell=1}^{k}\left(\frac{1}{k}\,I + \sum_{i \in V_{j_\ell}}(M_{i,t-1} - I)\right)\boldsymbol{u}_{j_\ell}\right).$$

Then:

1. *Under the same assumptions as in Lemma 1,*

$$||\bar{\boldsymbol{u}}_t - \bar{\boldsymbol{w}}_{j,t-1}|| \le \sqrt{3m}\,\widetilde{\mathrm{TCB}}_{j,t-1}$$

   holds with probability at least $1-\delta$, uniformly over cluster indices $j = 1, \ldots, m_t$, and rounds $t = 1, 2, \ldots$.

2. *For any fixed $\boldsymbol{u} \in \mathbb{R}^d$ we have*

$$||\bar{\boldsymbol{u}}_t - \boldsymbol{u}|| \le 2\sum_{\ell=1}^{k}||\boldsymbol{u}_{j_\ell} - \boldsymbol{u}|| \le 2\,SD(\boldsymbol{u}).$$

Proof. Let $X_{\ell,t-1}$ be the matrix whose columns are the d-dimensional vectors $\bar{\boldsymbol{x}}_s$, for all $s < t \,:\, i_s \in V_{j_\ell}$, $\boldsymbol{a}_{\ell,t-1}$ be the column vector collecting all payoffs a_s, $s < t \,:\, i_s \in V_{j_\ell}$, and $\boldsymbol{\eta}_{\ell,t-1}$ be the corresponding column vector of noise values. We have

$$\bar{\boldsymbol{w}}_{j,t-1} = \bar{M}_{j,t-1}^{-1}\bar{\boldsymbol{b}}_{j,t-1},$$

with

$$\begin{aligned}\bar{\boldsymbol{b}}_{j,t-1} &= \sum_{\ell=1}^{k} X_{\ell,t-1}\boldsymbol{a}_{\ell,t-1}\\ &= \sum_{\ell=1}^{k} X_{\ell,t-1}\left(X_{\ell,t-1}^\top \boldsymbol{u}_{j_\ell} + \boldsymbol{\eta}_{\ell,t-1}\right)\\ &= \sum_{\ell=1}^{k}\left(\sum_{i \in V_{j_\ell}}(M_{i,t-1} - I)\boldsymbol{u}_{j_\ell} + X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1}\right).\end{aligned}$$

Thus

$$\bar{\boldsymbol{w}}_{j,t-1} - \bar{\boldsymbol{u}}_t = \bar{M}_{j,t-1}^{-1}\left(\sum_{\ell=1}^{k}\left(X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1} - \frac{1}{k}\,\boldsymbol{u}_{j_\ell}\right)\right)$$

and, for any fixed $\boldsymbol{x} \in \mathbb{R}^d \,:\, ||\boldsymbol{x}|| = 1$, we have

$$\begin{aligned}&\left(\bar{\boldsymbol{w}}_{j,t-1}^\top \boldsymbol{x} - \bar{\boldsymbol{u}}_t^\top \boldsymbol{x}\right)^2\\ &= \left(\left(\sum_{\ell=1}^{k}\left(X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1} - \frac{1}{k}\,\boldsymbol{u}_{j_\ell}\right)\right)^\top \bar{M}_{j,t-1}^{-1}\boldsymbol{x}\right)^2\\ &\le \boldsymbol{x}^\top \bar{M}_{j,t-1}^{-1}\boldsymbol{x}\,\left(\sum_{\ell=1}^{k}\left(X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1} - \frac{1}{k}\,\boldsymbol{u}_{j_\ell}\right)\right)^\top \bar{M}_{j,t-1}^{-1}\\ &\quad\times\left(\sum_{\ell=1}^{k}\left(X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1} - \frac{1}{k}\,\boldsymbol{u}_{j_\ell}\right)\right)\\ &\le 2\,\boldsymbol{x}^\top \bar{M}_{j,t-1}^{-1}\boldsymbol{x}\\ &\quad\times\left(\Big(\sum_{\ell=1}^{k} X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1}\Big)^\top \bar{M}_{j,t-1}^{-1}\Big(\sum_{\ell=1}^{k} X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1}\Big)\right.\\ &\qquad\left. + \frac{1}{k^2}\Big(\sum_{\ell=1}^{k}\boldsymbol{u}_{j_\ell}\Big)^\top \bar{M}_{j,t-1}^{-1}\Big(\sum_{\ell=1}^{k}\boldsymbol{u}_{j_\ell}\Big)\right)\end{aligned}$$

(using $(a+b)^2 \le 2a^2 + 2b^2$).

We focus on the two terms inside the big braces. Because $\hat{V}_{j,t}$ is made up of the union of true clusters, we can stratify over the set of all such unions (which are at most 2^m with high probability), and then apply the martingale result in (Abbasi-Yadkori et al., 2011) (Theorem 1 therein), showing that

$$\left(\sum_{\ell=1}^{k} X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1}\right)^\top \bar{M}_{j,t-1}^{-1}\left(\sum_{\ell=1}^{k} X_{\ell,t-1}\,\boldsymbol{\eta}_{\ell,t-1}\right) \le 2\,\sigma^2\left(\log\frac{|\bar{M}_{j,t-1}|}{\delta/2^{m+1}}\right)$$

holds with probability at least $1-\delta/2$. As for the second term, we simply write

$$\frac{1}{k^2}\left(\sum_{\ell=1}^{k}\boldsymbol{u}_{j_\ell}\right)^\top \bar{M}_{j,t-1}^{-1}\left(\sum_{\ell=1}^{k}\boldsymbol{u}_{j_\ell}\right) \le \frac{1}{k^2}\left|\left|\sum_{\ell=1}^{k}\boldsymbol{u}_{j_\ell}\right|\right|^2 \le 1.$$

Putting together and overapproximating we conclude that

$$|\bar{\boldsymbol{w}}_{j,t-1}^\top \boldsymbol{x} - \bar{\boldsymbol{u}}_t^\top \boldsymbol{x}| \le \sqrt{3m}\,\mathrm{TCB}_{j,t-1}(\boldsymbol{x})$$

and, since this holds for all unit-norm $\boldsymbol{x}$, Lemma 4 yields

$$||\bar{\boldsymbol{w}}_{j,t-1} - \bar{\boldsymbol{u}}_t|| \le \sqrt{3m}\,\widetilde{\mathrm{TCB}}_{j,t-1},$$

thereby concluding the proof of part 1.

As for part 2, because

$$\bar{M}_{j,t-1} = I + \sum_{\ell=1}^{k}\sum_{i \in V_{j_\ell}}(M_{i,t-1} - I),$$

we can rewrite $\boldsymbol{u}$ as

$$\boldsymbol{u} = \bar{M}_{j,t-1}^{-1}\left(\boldsymbol{u} + \sum_{\ell=1}^{k}\sum_{i\in V_{j_\ell}}(M_{i,t-1}-I)\boldsymbol{u}\right),$$

so that

$$\bar{\boldsymbol{u}}_t - \boldsymbol{u} = \bar{M}_{j,t-1}^{-1}\Bigg(\frac{1}{k}\sum_{\ell=1}^{k}(\boldsymbol{u}_{j_\ell}-\boldsymbol{u}) + \sum_{\ell=1}^{k}\sum_{i\in V_{j_\ell}}(M_{i,t-1}-I)(\boldsymbol{u}_{j_\ell}-\boldsymbol{u})\Bigg).$$

Hence

$$\begin{aligned}||\bar{\boldsymbol{u}}_t - \boldsymbol{u}|| &\le \frac{1}{k}\left|\left|\bar{M}_{j,t-1}^{-1}\sum_{\ell=1}^{k}(\boldsymbol{u}_{j_\ell}-\boldsymbol{u})\right|\right| \\ &\quad + \sum_{\ell=1}^{k}\left|\left|\bar{M}_{j,t-1}^{-1}\sum_{i\in V_{j_\ell}}(M_{i,t-1}-I)(\boldsymbol{u}_{j_\ell}-\boldsymbol{u})\right|\right| \\ &\le \frac{1}{k}\sum_{\ell=1}^{k}||\boldsymbol{u}_{j_\ell}-\boldsymbol{u})|| + \sum_{\ell=1}^{k}||\boldsymbol{u}_{j_\ell}-\boldsymbol{u}|| \\ &\le 2\sum_{\ell=1}^{k}||\boldsymbol{u}_{j_\ell}-\boldsymbol{u}||,\end{aligned}$$

as claimed. □

The next lemma gives sufficient conditions on $T_{i,t}$ (or on $\bar{T}_{j,t}$) to insure that $\widetilde{\text{TCB}}_{i,t}$ (or $\widetilde{\text{TCB}}_{j,t}$) is small. We state the lemma for $\widetilde{\text{TCB}}_{i,t}$, but the very same statement clearly holds when we replace $\widetilde{\text{TCB}}_{i,t}$ by $\widetilde{\text{TCB}}_{j,t}$, $T_{i,t}$ by $\bar{T}_{j,t}$, and n by 2^m.

Lemma 6. *The following properties hold for upper confidence bound* $\widetilde{\text{TCB}}_{i,t}$*:*

1. $\widetilde{\text{TCB}}_{i,t}$ *is nonincreasing in* $T_{i,t}$*;*
2. *Let* $A = \sigma\sqrt{2d\,\log(1+t) + 2\log(2/\delta)} + 1$*. Then*
$$\widetilde{\text{TCB}}_{i,t} \le \frac{A}{\sqrt{1+\lambda T_{i,t}/8}}$$
when
$$T_{i,t} \ge \frac{2\cdot 32^2}{\lambda^2}\,\log\left(\frac{2nd}{\delta}\right)\,\log\left(\frac{32^2}{\lambda^2}\log\left(\frac{2nd}{\delta}\right)\right);$$
3. *We have*
$$\widetilde{\text{TCB}}_{i,t} \le \gamma/4$$
when
$$T_{i,t} \ge \frac{32}{\lambda}\,\max\left\{\frac{A^2}{\gamma^2}, \frac{64}{\lambda}\,\log\left(\frac{2nd}{\delta}\right) \times \log\left(\frac{32^2}{\lambda^2}\log\left(\frac{2nd}{\delta}\right)\right)\right\}.$$

Proof. The proof follows from simple but annoying calculations, and is therefore omitted. □

We are now ready to combine all previous lemmas into the proof of Theorem 1.

Proof. Let t be a generic round, $\widehat{j}_t$ be the index of the current cluster node i_t belongs to, and j_t be the index of the *true* cluster i_t belongs to. Also, let us define the aggregate vector $\bar{\boldsymbol{w}}_{j_t,t-1}$ as follows :

$$\begin{aligned}\bar{\boldsymbol{w}}_{j_t,t-1} &= \bar{M}_{j_t,t-1}^{-1}\bar{\boldsymbol{b}}_{j_t,t-1}, \\ \bar{M}_{j_t,t-1} &= I + \sum_{i\in V_{j_t}}(M_{i,t-1}-I), \\ \bar{\boldsymbol{b}}_{j_t,t-1} &= \sum_{i\in V_{j_t}}\boldsymbol{b}_{i,t-1}.\end{aligned}$$

Assume Lemma 3 holds, implying that the current cluster $\hat{V}_{\widehat{j}_t,t}$ is the (disjoint) union of true clusters, and define the aggregate vector $\bar{\boldsymbol{u}}_t$ accordingly, as in the statement of Lemma 5. Notice that $\bar{\boldsymbol{w}}_{j_t,t-1}$ is the true cluster counterpart to $\bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}$, that is, $\bar{\boldsymbol{w}}_{j_t,t-1} = \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}$ if $V_{j_t} = \hat{V}_{\widehat{j}_t,t}$. Also, observe that $\bar{\boldsymbol{u}}_t = \boldsymbol{u}_{i_t}$ when $V_{j_t} = \hat{V}_{\widehat{j}_t,t}$. Finally, set for brevity

$$\boldsymbol{x}_t^* = \operatorname*{argmax}_{k=1,\dots,c_t}\boldsymbol{u}_{i_t}^\top\boldsymbol{x}_{t,k}.$$

We can rewrite the time-t regret r_t as follows:

$$\begin{aligned}r_t &= \boldsymbol{u}_{i_t}^\top\boldsymbol{x}_t^* - \boldsymbol{u}_{i_t}^\top\bar{\boldsymbol{x}}_t \\ &= \boldsymbol{u}_{i_t}^\top\boldsymbol{x}_t^* - \bar{\boldsymbol{w}}_{j_t,t-1}^\top\boldsymbol{x}_t^* + \bar{\boldsymbol{w}}_{j_t,t-1}^\top\boldsymbol{x}_t^* - \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}^\top\boldsymbol{x}_t^* \\ &\quad + \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}^\top\boldsymbol{x}_t^* - \bar{\boldsymbol{w}}_{j_t,t-1}^\top\bar{\boldsymbol{x}}_t + \bar{\boldsymbol{w}}_{j_t,t-1}^\top\bar{\boldsymbol{x}}_t - \boldsymbol{u}_{i_t}^\top\bar{\boldsymbol{x}}_t.\end{aligned}$$

Combined with

$$\bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}^\top\boldsymbol{x}_t^* + \text{TCB}_{\widehat{j}_t,t-1}(\boldsymbol{x}_t^*) \le \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}^\top\bar{\boldsymbol{x}}_t + \text{TCB}_{\widehat{j}_t,t-1}(\bar{\boldsymbol{x}}_t),$$

and rearranging gives

$$\begin{aligned}r_t &\le \boldsymbol{u}_{i_t}^\top\boldsymbol{x}_t^* - \bar{\boldsymbol{w}}_{j_t,t-1}^\top\boldsymbol{x}_t^* - \text{TCB}_{\widehat{j}_t,t-1}(\boldsymbol{x}_t^*) && (3)\\ &\quad + \bar{\boldsymbol{w}}_{j_t,t-1}^\top\bar{\boldsymbol{x}}_t - \boldsymbol{u}_{i_t}^\top\bar{\boldsymbol{x}}_t + \text{TCB}_{\widehat{j}_t,t-1}(\bar{\boldsymbol{x}}_t) && (4)\\ &\quad + (\bar{\boldsymbol{w}}_{j_t,t-1} - \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1})^\top(\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t). && (5)\end{aligned}$$

We continue by bounding with high probability the three terms (3), (4), and (5).

As for (3), and (4), we simply observe that Lemma 2 allows[5] us to write

$$\boldsymbol{u}_{i_t}^\top\boldsymbol{x}_t^* - \bar{\boldsymbol{w}}_{j_t,t-1}^\top\boldsymbol{x}_t^* \le ||\boldsymbol{u}_{i_t} - \bar{\boldsymbol{w}}_{j_t,t-1}|| \le \widetilde{\text{TCB}}_{j_t,t-1},$$

[5] This lemma applies here since, by definition, $\bar{\boldsymbol{w}}_{j_t,t-1}$ is built only from payoffs from nodes in V_{j_t}, sharing the common unknown vector $\boldsymbol{u}_{i_t}$.

and

$$\bar{\boldsymbol{w}}_{j_t,t-1}^\top \bar{\boldsymbol{x}}_t - \boldsymbol{u}_{i_t}^\top \bar{\boldsymbol{x}}_t \le ||\boldsymbol{u}_{i_t} - \bar{\boldsymbol{w}}_{j_t,t-1}|| \le \widetilde{\text{TCB}}_{j_t,t-1} .$$

Moreover,

$$\begin{aligned}\text{TCB}_{\widehat{j}_t,t-1}(\bar{\boldsymbol{x}}_t) &\le \widetilde{\text{TCB}}_{\widehat{j}_t,t-1} \\ &\quad \text{(by Lemma 4)} \\ &\le \widetilde{\text{TCB}}_{j_t,t-1} \\ &\quad \text{(by Lemma 3 and the definition of } \widehat{j}_t).\end{aligned}$$

Hence,

$$(3) + (4) \le 3\widetilde{\text{TCB}}_{j_t,t-1} \qquad (6)$$

holds with probability at least $1-2\delta$, uniformly over t.

As for (5), letting $\{\cdot\}$ be the indicator function of the predicate at argument, we can write

$$\begin{aligned}&(\bar{\boldsymbol{w}}_{j_t,t-1} - \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1})^\top(\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t) \\ &= (\bar{\boldsymbol{w}}_{j_t,t-1} - \boldsymbol{u}_{i_t})^\top(\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t) + (\boldsymbol{u}_{i_t} - \bar{\boldsymbol{u}}_t)^\top(\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t) \\ &\quad + (\bar{\boldsymbol{u}}_t - \bar{\boldsymbol{w}}_{\widehat{j}_t,t-1})^\top(\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t) \\ &\le 2\,\widetilde{\text{TCB}}_{j_t,t-1} + 2\,||\boldsymbol{u}_{i_t} - \bar{\boldsymbol{u}}_t|| + 2\sqrt{3m}\,\widetilde{\text{TCB}}_{\widehat{j}_t,t-1} \\ &\quad \text{(using Lemma 2, } ||\boldsymbol{x}_t^* - \bar{\boldsymbol{x}}_t|| \le 2\text{, and Lemma 5, part 1)} \\ &= 2\,\widetilde{\text{TCB}}_{j_t,t-1} + 2\,\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\}\,||\boldsymbol{u}_{i_t} - \bar{\boldsymbol{u}}_t|| \\ &\qquad + 2\sqrt{3m}\,\widetilde{\text{TCB}}_{\widehat{j}_t,t-1} \\ &\le 2(1+\sqrt{3m})\,\widetilde{\text{TCB}}_{j_t,t-1} + 4\,\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\}\,SD(\boldsymbol{u}_{i_t}) \\ &\quad \text{(by Lemma 3, and Lemma 5, part 2)} .\end{aligned}$$

Piecing together we have so far obtained

$$\begin{aligned}r_t &\le (5+2\sqrt{3m})\,\widetilde{\text{TCB}}_{j_t,t-1} \\ &\quad + 4\,\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\}\,SD(\boldsymbol{u}_{i_t}) . \qquad (7)\end{aligned}$$

We continue by bounding $\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\}$. From Lemma 3, we clearly have

$$\begin{aligned}&\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\} \\ &\le \{\exists i \in V_{j_t}, \exists j \notin V_{j_t} \,:\, (i,j) \in E_t\} \\ &\le \Big\{\exists i \in V_{j_t}, \exists j \notin V_{j_t} \,:\, \forall s < t\big((i_s \ne i) \\ &\quad \vee (i_s = i, ||\boldsymbol{w}_{i,s-1} + \boldsymbol{w}_{j,s-1}|| \le \widetilde{\text{TCB}}_{i,s-1} + \widetilde{\text{TCB}}_{j,s-1})\big)\Big\} \\ &\le \{\exists i \in V_{j_t} \,:\, \forall s < t\ i_s \ne i\} \\ &\quad + \Big\{\exists i \in V_{j_t}, \exists j \notin V_{j_t} \,:\, \\ &\qquad \forall s < t\ ||\boldsymbol{w}_{i,s-1} + \boldsymbol{w}_{j,s-1}|| \le \widetilde{\text{TCB}}_{i,s-1} + \widetilde{\text{TCB}}_{j,s-1}\Big\} \\ &\le \{\exists i \in V_{j_t} \,:\, \forall s < t\ i_s \ne i\} \\ &\quad + \{\exists i \in V_{j_t}, \exists j \notin V_{j_t} \,:\, \\ &\qquad \forall s < t\ \widetilde{\text{TCB}}_{i,s-1} + \widetilde{\text{TCB}}_{j,s-1} \ge \gamma/2\} \\ &\le \{\exists i \in V_{j_t} \,:\, \forall s < t\ i_s \ne i\} \\ &\quad + \{\exists i \in V \,:\, \forall s < t\ \widetilde{\text{TCB}}_{i,s-1} \ge \gamma/4\} .\end{aligned}$$

At this point, we apply Lemma 6 to $\widetilde{\text{TCB}}_{i,t}$ with

$$\begin{aligned}A^2 &= \left(\sigma\sqrt{2d\,\log(1+T) + 2\log(2/\delta)} + 1\right)^2 \\ &\le 4\sigma^2(d\,\log(1+T) + \log(2/\delta)) + 2,\end{aligned}$$

and set for brevity

$$\begin{aligned}B &= \frac{32}{\lambda}\,\max\Bigg\{\frac{A^2}{\gamma^2}, \frac{64}{\lambda}\,\log\left(\frac{2nd}{\delta}\right) \\ &\qquad \times \log\left(\frac{32^2}{\lambda^2}\log\left(\frac{2nd}{\delta}\right)\right)\Bigg\} , \\ C &= \frac{2\cdot 32^2}{\lambda^2}\,\log\left(\frac{2^{m+1}d}{\delta}\right)\log\left(\frac{32^2}{\lambda^2}\log\left(\frac{2^{m+1}d}{\delta}\right)\right) .\end{aligned}$$

We can write

$$\begin{aligned}&\{\exists i \in V \,:\, \forall s < t\ \widetilde{\text{TCB}}_{i,s-1} \ge \gamma/4\} \\ &\quad \le \{\exists i \in V \,:\, \widetilde{\text{TCB}}_{i,t-2} \ge \gamma/4\} \\ &\quad \le \{\exists i \in V \,:\, T_{i,t-2} \le B\} .\end{aligned}$$

Moreover,

$$\begin{aligned}&\{\exists i \in V_{j_t} \,:\, \forall s < t\ i_s \ne i\} \\ &\quad \le \{\exists i \in V_{j_t} \setminus \{i_t\} \,:\, T_{i,t-1} = 0\} \\ &\quad \le \{\exists i \in V \,:\, T_{i,t-1} = 0\} .\end{aligned}$$

That is,

$$\begin{aligned}\{V_{j_t} \ne \hat{V}_{\widehat{j}_t,t}\} &\le \{\exists i \in V \,:\, T_{i,t-2} \le B\} \\ &\quad + \{\exists i \in V \,:\, T_{i,t-1} = 0\} .\end{aligned}$$

Further, using again Lemma 6 (applied this time to $\widetilde{\text{TCB}}_{j,t}$) combined with the fact that $\widetilde{\text{TCB}}_{j,t} \le A$ for all j and t, we have

$$\widetilde{\text{TCB}}_{j_t,t-1} = A\,\{\bar{T}_{j_t,t-1} < C\} + \frac{A}{\sqrt{1+\lambda\,\bar{T}_{j_t,t-1}/8}} ,$$

where

$$\bar{T}_{j_t,t-1} = \sum_{i\in V_{j_t}} T_{i,t-1} = |\{s \le t-1 \,:\, i_s \in V_{j_t}\}| .$$

Putting together as in (7), and summing over $t = 1,\ldots,T$, we have shown so far that with probability at least $1-7\delta/2$,

$$\begin{aligned}\sum_{t=1}^T r_t &\le (5+2\sqrt{3m})A\,\sum_{t=1}^T\{\bar{T}_{j_t,t-1} < C\} \\ &\quad + (5+2\sqrt{3m})A\,\sum_{t=1}^T\frac{1}{\sqrt{1+\lambda\,\bar{T}_{j_t,t-1}/8}} \\ &\quad + 4\sum_{t=1}^T SD(\boldsymbol{u}_{i_t})\,\{\exists i \in V \,:\, T_{i,t-2} \le B\} \\ &\quad + 4\sum_{t=1}^T SD(\boldsymbol{u}_{i_t})\,\{\exists i \in V \,:\, T_{i,t-1} = 0\} ,\end{aligned}$$

with $T_{i,t} = 0$ if $t \le 0$.

We continue by upper bounding with high probability the four terms in the right-hand side of the last inequality. First, observe that for any fixed i and t, $T_{i,t}$ is a binomial random variable with parameters t and $1/n$, and $\bar{T}_{j_t,t-1} = \sum_{i \in V_{j_t}} T_{i,t-1}$ which, for fixed i_t, is again binomial with parameters t, and $\frac{v_{j_t}}{n}$, where v_{j_t} is the size of the true cluster i_t falls into. Moreover, for any fixed t, the variables $T_{i,t}$, $i \in V$ are indepedent.

To bound the third term, we use a standard Bernstein inequality twice: first, we apply it to sequences of independent Bernoulli variables, whose sum $T_{i,t-2}$ has average $\mathbb{E}[T_{i,t-2}] = \frac{t-2}{n}$ (for $t \ge 3$), and then to the sequence of variables $SD(\boldsymbol{u}_{i_t})$ whose average $\mathbb{E}[SD(\boldsymbol{u}_{i_t})] = \frac{1}{n}\sum_{i\in V} SD(\boldsymbol{u}_i)$ is over the random choice of i_t.

Setting for brevity

$$D(B) = 2n\left(B + \frac{5}{3}\,\log(Tn/\delta)\right) + 2,$$

where B has been defined before, we can write

$$\begin{aligned}
&\sum_{t=1}^{T} SD(\boldsymbol{u}_{i_t})\ \{\exists i \in V\ :\ T_{i,t-2} \le B\} \\
&\quad = \sum_{t \le D(B)} SD(\boldsymbol{u}_{i_t})\ \{\exists i \in V\ :\ T_{i,t-2} \le B\} \\
&\qquad + \sum_{t > D(B)} SD(\boldsymbol{u}_{i_t})\ \{\exists i \in V\ :\ T_{i,t-2} \le B\} \\
&\quad \le \sum_{t \le D(B)} SD(\boldsymbol{u}_{i_t}) \\
&\qquad + m \sum_{t > D(B)} \{\exists i \in V\ :\ T_{i,t-2} \le B\}\ .
\end{aligned}$$

Then from Bernstein's inequality,

$$\mathbb{P}\left(\exists i \in V\ \exists t > D(B)\ :\ T_{i,t-2} \le B\right) \le \delta\ ,$$

and

$$\mathbb{P}\Biggl(\sum_{t \le D(B)} SD(\boldsymbol{u}_{i_t}) \ge \frac{3}{2}\,D(B)\,\mathbb{E}[SD(\boldsymbol{u}_{i_t})] + \frac{5}{3}\,m\,\log(1/\delta)\Biggr) \le \delta\ .$$

Thus with probability $\ge 1 - 2\delta$

$$\begin{aligned}
&\sum_{t=1}^{T} SD(\boldsymbol{u}_{i_t})\,\{\exists i \in V\ :\ T_{i,t-2} \le B\} \\
&\qquad \le \frac{3}{2}\,D(B)\,\mathbb{E}[SD(\boldsymbol{u}_{i_t})] + \frac{5}{3}\,m\,\log(1/\delta)\ .
\end{aligned}$$

Similarly, to bound the fourth term we have, with probability $\ge 1 - 2\delta$,

$$\begin{aligned}
&\sum_{t=1}^{T} SD(\boldsymbol{u}_{i_t})\,\{\exists i \in V\ :\ T_{i,t-1} = 0\} \\
&\qquad \le \frac{3}{2}\,D(0)\,\mathbb{E}[SD(\boldsymbol{u}_{i_t})] + \frac{5}{3}\,m\,\log(1/\delta)\ .
\end{aligned}$$

Next, we crudely upper bound the first term as

$$\begin{aligned}
&(5{+}2\sqrt{3m})A \sum_{t=1}^{T}\{\bar{T}_{j_t,t-1} < C\} \\
&\qquad \le (5 + 2\sqrt{3m})A \sum_{t=1}^{T}\{T_{i_t,t-1} < C\}\ ,
\end{aligned}$$

and then apply a very similar argument as before to show that with probability $\ge 1 - \delta$,

$$\sum_{t=1}^{T}\{T_{i_t,t-1} < C\} \le n\left(C + \frac{5}{3}\log\left(\frac{T}{\delta}\right)\right) + 1\ .$$

Finally, we are left to bound the second term. The following is a simple property of binomial random variables we be useful.

Claim 2. *Let X be a binomial random variable with parameters n and p, and $\lambda \in (0,1)$ be a constant. Then*

$$\mathbb{E}\left[\frac{1}{\sqrt{1+\lambda X}}\right] \le \begin{cases} \frac{3}{\sqrt{1+\lambda\,n\,p}} & \textit{if } np \ge 10\ ; \\ 1 & \textit{if } np < 10\ . \end{cases}$$

Proof of claim. The second branch of the inequality is clearly trivial, so we focus on the first one under the assumption $np \ge 10$. Let then $\beta \in (0,1)$ be a parameter that will be set later on. We have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{1+\lambda X}}\right] &\le \mathbb{P}(X \le (1-\beta)\,n\,p) \\
&\quad + \frac{1}{\sqrt{1+\lambda\,(1-\beta)\,n\,p}}\,\mathbb{P}(X \ge (1-\beta)\,n\,p) \\
&\le e^{-\beta^2\,n\,p/2} + \frac{1}{\sqrt{1+\lambda\,(1-\beta)\,n\,p}}\ ,
\end{aligned}$$

the last inequality following from the standard Chernoff bounds. Setting $\beta = \sqrt{\frac{\log(1+\lambda\, n\, p)}{n\, p}}$ gives

$$\mathbb{E}\left[\frac{1}{\sqrt{1+\lambda X}}\right] \le \frac{1}{\sqrt{1+\lambda\, n\, p}} + \frac{1}{\sqrt{1+\lambda\left(np - \sqrt{np\log(1+\lambda np)}\right)}}$$
$$\le \frac{1}{\sqrt{1+\lambda\, n\, p}} + \frac{1}{\sqrt{1+\lambda\, n\, p/2}}$$
$$\text{(using } np \ge 10\text{)}$$
$$\le \frac{3}{\sqrt{1+\lambda\, n\, p}}\ ,$$

i.e., the claimed inequality □

Now,

$$\mathbb{E}_{t-1}\left[\frac{1}{\sqrt{1+\lambda \bar{T}_{j_t,t-1}/8}}\right] = \sum_{j=1}^{m} \frac{v_j}{n}\,\frac{1}{\sqrt{1+\lambda \bar{T}_{j,t-1}/8}}\ ,$$

being $\bar{T}_{j,t-1} = |\{s < t\ :\ i_s \in V_j\}|$ a binomial variable with parameters $t-1$ and $\frac{v_j}{n}$, where $v_j = |V_j|$. By the standard Hoeffding-Azuma inequality

$$\sum_{t=1}^{T} \frac{1}{\sqrt{1+\lambda \bar{T}_{j_t,t-1}/8}} \le \sum_{t=1}^{T}\sum_{j=1}^{m} \frac{v_j}{n}\,\frac{1}{\sqrt{1+\lambda \bar{T}_{j,t-1}/8}} + \sqrt{2T\,\log(1/\delta)}$$

holds with probability at least $1-\delta$, In turn, from Bernstein's inequality, we have

$$\mathbb{P}\left(\exists t\,\exists j\ :\ \bar{T}_{j,t-1} \le \frac{t-1}{2n}\,v_j - \frac{5}{3}\,\log(Tm/\delta)\right) \le \delta\ .$$

Therefore, with probability at least $1-2\delta$,

$$\sum_{t=1}^{T} \frac{1}{\sqrt{1+\lambda \bar{T}_{j_t,t-1}/8}}$$
$$\le \sum_{t=1}^{T}\sum_{j=1}^{m} \frac{v_j}{n}\,\frac{1}{\sqrt{1+\frac{\lambda}{8}\,\left(\frac{t-1}{2n}\,v_j - \frac{5}{3}\,\log(Tm/\delta)\right)_+}} + \sqrt{2T\,\log(1/\delta)}$$
$$\le \sum_{j=1}^{m} \frac{v_j}{n}\left(4n\,\frac{5}{3}\,\log(Tm/\delta) + 1 + \sum_{t=1}^{T} \frac{1}{\sqrt{1+\frac{\lambda}{8}\,\frac{t-1}{4n}\,v_j}}\right) + \sqrt{2T\,\log(1/\delta)}$$
$$= 4n\,\frac{5}{3}\,\log(Tm/\delta) + 1 + \sum_{j=1}^{m} \frac{v_j}{n}\,\sum_{t=1}^{T} \frac{1}{\sqrt{1+\frac{\lambda}{8}\,\frac{t-1}{4n}\,v_j}} + \sqrt{2T\,\log(1/\delta)}\ .$$

If we set for brevity $r_j = \frac{\lambda}{8}\,\frac{v_j}{4n}$, $j = 1, \ldots, m$, we have

$$\sum_{t=1}^{T} \frac{1}{\sqrt{1+\frac{\lambda}{8}\,\frac{t-1}{4n}\,v_j}} \le \int_0^T \frac{dx}{\sqrt{1+(x-1)r_j}}$$
$$= \frac{2}{r_j}\left(\sqrt{1+T\,r_j - r_j} - \sqrt{1-r_j}\right)$$
$$\le 2\sqrt{\frac{T}{r_j}}\ ,$$

so that

$$\sum_{t=1}^{T} \frac{1}{\sqrt{1+\lambda \bar{T}_{j_t,t-1}/8}} \le 4n\,\frac{5}{3}\,\log(Tm/\delta) + 1 + \sqrt{2T\,\log(1/\delta)} + 8\sum_{j=1}^{m}\sqrt{\frac{2Tv_j}{\lambda n}}\ .$$

Finally, we put all pieces together. In order for all claims to hold simultaneously with probability at least $1-\delta$, we need to replace δ throughout by $\delta/10.5$. Then we switch to a $\widetilde{O}$-notation, and overapproximate once more to conclude the proof. □

2. Implementation

As we said in the main text, in implementing the algorithm in Figure 1, the reader should keep in mind that it is reasonable to expect n (the number of users) to be quite large, d (the number of features of each item) to be relatively small, and m (the number of true clusters) to be very small compared to n. Then the algorithm can be implemented by storing a least-squares estimator $\boldsymbol{w}_{i,t-1}$ at each node $i \in V$, an aggregate least squares estimator $\bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}$ for each current cluster $\widehat{j}_t \in \{1, \ldots, m_t\}$, and an extra data-structure which is able to perform decremental dynamic connectivity. Fast implementations of such data-structures are those studied by (Thorup, 1997; Kapron et al., 2013) (see also the research thread referenced therein). In particular, in (Thorup, 1997) (Theorem 1.1 therein) it is shown that a randomized construction exists that maintains a spanning forerst which, given an initial undirected graph $G_1 = (V, E_1)$, is able to perform edge deletions and answer connectivity queries of the form "Is node i connected to node j" in expected total time $O\left(\min\{|V|^2, |E_1|\,\log|V|\} + \sqrt{|V|\,|E_1|}\,\log^{2.5}|V|\right)$ for $|E_1|$ deletions. Connectivity queries and deletions can be interleaved, the former being performed in *constant* time. Notice that when we start off from the full graph, we have $|E_1| = O(|V|^2)$, so that the expected amortized time per query becomes constant. On the other hand, if our initial graph has $|E_1| = O(|V|\,\log|V|)$ edges, then the expected amortized time per query is $O(\log^2|V|)$. This becomes $O(\log^{2.5}|V|)$ if the initial graph has $|E_1| = O(|V|)$. In addition, we maintain an n-dimensional vector CLUSTERINDICES containing, for each node $i \in V$, the index j of the current cluster i belongs to.

With these data-structures handy, we can implement our algorithm as follows. After receiving i_t, computing j_t is $O(1)$ (just by accessing CLUSTERINDICES). Then, computing k_t can be done in time $O(d^2)$ (matrix-vector multiplication, executed c_t times, assuming c_t is a constant). Then the algorithm directly updates $\boldsymbol{b}_{i_t,t-1}$ and $\bar{\boldsymbol{b}}_{\widehat{j}_t,t-1}$, as well as the inverses of matrices $M_{i_t,t-1}$ and $\bar{M}_{\widehat{j}_t,t-1}$, which is again $O(d^2)$, using standard formulas for rank-one adjustment of inverse matrices. In order to prepare the ground for the subsequent edge deletion phase, it is convenient that the algorithm also stores at each node i matrix $M_{i,t-1}$ (whose time-t update is again $O(d^2)$).

Let DELETE(i,ℓ) and IS-CONNECTED(i,ℓ) be the two operations delivered by the decremental dynamic connectivity data-structure. Edge deletion at time t corresponds to cycling through all nodes ℓ such that (i_t,ℓ) is an existing edge. The number of such edges is on average equal to the average degree of node i_t, which is $O\left(\frac{|E_1|}{n}\right)$, where $|E_1|$ is the number of edges in the initial graph G_1. Now, if (i_t,ℓ) has to be deleted (each the deletion test being $O(d)$), then we invoke DELETE(i_t,ℓ), and then IS-CONNECTED(i_t,ℓ). If IS-CONNECTED(i_t,ℓ) = "no", this means that the current cluster $\hat{V}_{j_t,t-1}$ has to split into two new clusters as a consequence of the deletion of edge (i_t,ℓ). The set of nodes contained in these two clusters correspond to the two sets

$$\{k \in V \,:\, \text{IS-CONNECTED}(i_t,k) = \text{“}yes\text{”}\},$$
$$\{k \in V \,:\, \text{IS-CONNECTED}(\ell,k) = \text{“}yes\text{”}\}‘,$$

whose expected amortized computation *per node* is $O(1)$ to $O(\log^{2.5} n)$ (depending on the density of the initial graph G_1). We modify the CLUSTERINDICES vector accordingly, but also the aggregate least squares estimators. This is because $\bar{\boldsymbol{w}}_{\widehat{j}_t,t-1}$ (represented through $\bar{M}^{-1}_{\widehat{j}_t,t}$ and $\bar{\boldsymbol{b}}_{\widehat{j}_t,t}$) has to be spread over the two newborn clusters. This operation can be performed by adding up all matrices $M_{i,t}$ and all $\boldsymbol{b}_{i,t}$, over all i belonging to each of the two new clusters (it is at this point that we need to access $M_{i,t}$ for each i), and then inverting the resulting aggregate matrices. This operation takes $O(n\,d^2 + d^3)$. However, as argued in the comments following Lemma 3, with high probability the number of current clusters m_t can never exceed m, so that with the same probability this operation is only performed at most m times throughout the learning process. Hence in T rounds we have an overall (expected) running time

$$O\left(T\,\left(d^2 + \frac{|E_1|}{n}\,d\right) + m\,(n\,d^2 + d^3) + |E_1| + \min\{n^2, |E_1|\,\log n\} + \sqrt{n\,|E_1|}\,\log^{2.5} n\right).$$

Notice that the above is $n \cdot \text{poly}(\log n)$, if so is $|E_1|$. In addition, if T is large compared to n and d, the average running time per round becomes $O(d^2 + d \cdot \text{poly}(\log n))$.

As for memory requirements, we need to store two $d \times d$ matrices and one d-dimensional vector at each node, one $d \times d$ matrix and one d-dimensional vector for each current cluster, vector CLUSTERINDICES, and the data-structures allowing for fast deletion and connectivity tests. Overall, these data-structures do not require more than $O(|E_1|)$ memory to be stored, so that this implementation takes $O(n\,d^2 + m\,d^2 + |E_1|) = O(n\,d^2 + |E_1|)$, where we again relied upon the $m_t \le m$ condition. Again, this is $n \cdot \text{poly}(\log n)$ if so is $|E_1|$.

3. Further Plots

This section contains a more thorough set of comparative plots on the synthetic datasets described in the main text. See Figure 1 and Figure 2.

4. Derivation of the Reference Bounds

We now provide a proof sketch of the reference bounds mentioned in Section 2 of the main text.

Let us start off from the *single user* bound for LINUCB (either ONE or IND) one can extract from (Abbasi-Yadkori et al., 2011). Let $\boldsymbol{u}_j \in \mathbb{R}^d$ be the profile vector of this user. Then, with probability at least $1-\delta$, we have

$$\begin{aligned}\sum_{t=1}^{T} r_t &= O\left(\sqrt{T\left(\sigma^2\,d\,\log T + \sigma^2\,\log\frac{1}{\delta} + ||\boldsymbol{u}_i||^2\right)\,d\,\log T}\right)\\ &= \widetilde{O}\left(\sqrt{T\,(\sigma^2\,d + ||\boldsymbol{u}_j||^2)\,d}\right)\\ &= \widetilde{O}\left((\sigma\,d + \sqrt{d})\sqrt{T}\right),\end{aligned}$$

the last line following from assuming $||\boldsymbol{u}_j|| = 1$.

Then, a straightforward way of turning this bound into a bound for the CLEARVOYANT algorithm that knows all clusters $V_1,\ldots,V_m$ ahead of time and runs one instance of LINUCB per cluster is to sum the regret contributed by each cluster throughout the T rounds. Letting $T_{j,T}$ denote

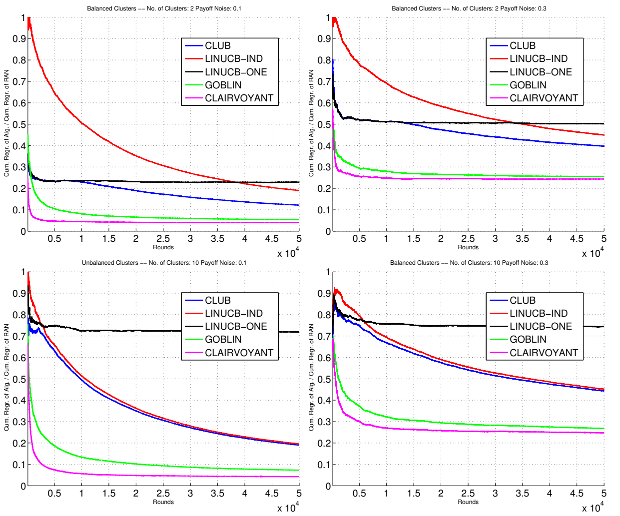

Figure 1. Results on synthetic datasets. Each plot displays the behavior of the ratio of the current cumulative regret of the algorithm ("Alg") to the current cumulative regret of RAN, where where "Alg" is either "CLUB" or "LinUCB-IND" or "LinUCB-ONE" or "GOBLIN"or "CLAIRVOYANT". The cluster sizes are balanced ($z = 0$). From left to right, payoff noise steps from 0.1 to 0.3, and from top to bottom the number of clusters jumps from 2 to 10.

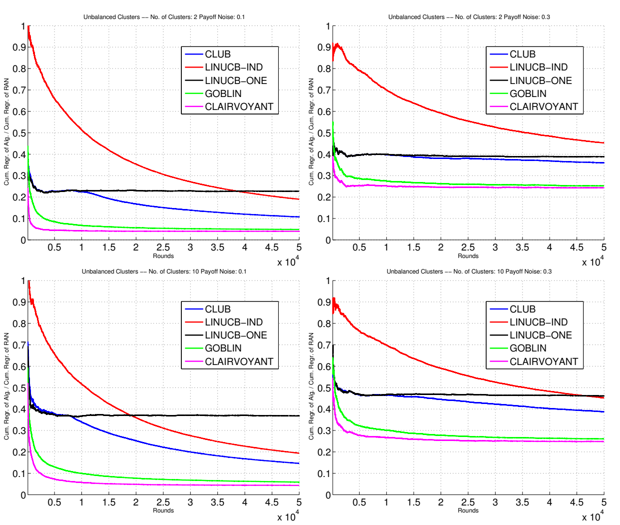

Figure 2. Results on synthetic datasets in the case of unbalanced ($z = 2$) cluster sizes. The rest is the same as in Figure 1.

the set of rounds t such that $i_t \in V_j$, we can write

$$\sum_{t=1}^{T} r_t = \widetilde{O}\left((\sigma\, d + \sqrt{d}) \sum_{j=1}^{m} \sqrt{T_{j,T}} \right) .$$

However, because i_t is drawn uniformly at random over V, we also have $\mathbb{E}[T_{j,T}] = T\frac{|V_j|}{n}$, so that we essentially have with high probability

$$\sum_{t=1}^{T} r_t = \widetilde{O}\left((\sigma\, d + \sqrt{d})\sqrt{T} \left(1 + \sum_{j=1}^{m} \sqrt{\frac{|V_j|}{n}}\right)\right) ,$$

i.e., Eq. (1) in the main text.

5. Further Comments

As we said in Remark 3, a data-dependent variant of the CLUB algorithm can be designed and analyzed which relies on data-dependent clusterability assumptions of the set of users with respect to a set of context vectors. These data-dependent assumptions allow us to work in a fixed design setting for the sequence of context vectors $\boldsymbol{x}_{t,k}$, and remove the sub-Gaussian and full-rank hypotheses regarding $\mathbb{E}[XX^\top]$. To make this more precise, consider an adversary that generates (unit norm) context vectors in a (possibly adaptive) way that *for all* $\boldsymbol{x}$ so generated $|\boldsymbol{u}_j^\top \boldsymbol{x} - \boldsymbol{u}_{j'}^\top \boldsymbol{x}| \geq \gamma$, whenever $j \neq j'$. In words, the adversary's power is restricted in that it cannot generate two distict context vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ such that $|\boldsymbol{u}_j^\top \boldsymbol{x} - \boldsymbol{u}_{j'}^\top \boldsymbol{x}|$ is small and $|\boldsymbol{u}_j^\top \boldsymbol{x}' - \boldsymbol{u}_{j'}^\top \boldsymbol{x}'|$ is large. The two quantities must either be both zero (when $j = j'$) or both bounded away from 0 (when $j \neq j'$). Under this assumption, one can show that a modification to the $\text{TCB}_{i,t}(\boldsymbol{x})$ and $\text{TCB}_{j,t}(\boldsymbol{x})$ functions exists that makes the CLUB algorithm in Figure 1 achieve a cumulative regret bound similar to the one in (5), where the $\sqrt{\frac{1}{\lambda}}$ factor therein is turned back into $\sqrt{d}$, as in the reference bound (1), but with a worse dependence on the geometry of the set of $\boldsymbol{u}_j$, as compared to $\mathbb{E}[SD(\boldsymbol{u}_{i_t})]$. The analysis goes along the very same lines as the one of Theorem 1.

6. Related Work

The most closely related papers are (Djolonga et al., 2013; Azar et al., 2013; Brunskill & Li, 2013; Maillard & Mannor, 2014).

In (Azar et al., 2013), the authors define a transfer learning problem within a stochastic multiarmed bandit setting, where a prior distribution is defined over the set of possible models over the tasks. More similar in spirit to our paper is the recent work (Brunskill & Li, 2013) that relies on clustering Markov Decision Processes based on their model parameter similarity. A paper sharing significant similarities with ours, in terms of both setting and technical tools is the very recent paper (Maillard & Mannor, 2014) that came to our attention at the time of writing ours. In that paper, the authors analyze a noncontextual stochastic bandit problem where model parameters can indeed be clustered in a few (unknown) types, thereby requiring the algorithm to learn the clusters rather than learning the parameters in isolation. Yet, the provided algorithmic solutions are completely different from ours. Finally, in (Djolonga et al., 2013), the authors work under the assumption that users are defined using a context vector, and try to learn a low-rank subspace under the assumption that variation across users is low-rank. The paper combines low-rank matrix recovery with high-dimensional Gaussian Process Bandits, but it gives rise to algorithms which do not seem easy to use in large scale practical scenarios.

7. Ongoing Research

This work could be extended along several directions. First, we may rely on a softer notion of clustering than the one we adopted here: a cluster is made up of nodes where the "within distance" between associated profile vectors is smaller than their "between distance". Yet, this is likely to require prior knowledge of either the distance threshold or the number of underlying clusters, which are assumed to be unknown in this paper. Second, it might be possible to handle partially overlapping clusters. Third, CLUB can clearly be modified so as to cluster nodes through off-the-shelf graph clustering techniques (mincut, spectral clustering, etc.). Clustering via connected components has the twofold advantage of being computationally faster and relatively easy to analyze. In fact, we do not know how to analyze CLUB when combined with alternative clustering techniques, and we suspect that Theorem 1 already delivers the sharpest results (as $T \rightarrow \infty$) when clustering is indeed based on connected components only. Fourth, from a practical standpoint, it would be important to incorporate further side information, like must-link and cannot-link constraints. Fifth, in recommender system practice, it is often relevant to provide recommendations to new users, even in the absence of past information (the so-called "cold start" problem). In fact, there is a way of tackling this problem through the machinery we developed here (the idea is to duplicate the newcomer's node as many times as the current clusters are, and then treat each copy as a separate user). This would potentially allow CLUB to work even in the presence of (almost) idle users. We haven't so far collected any experimental evidence on the effectiveness of this strategy. Sixth, following the comments we made in Remark 3, we are trying to see if the i.i.d. and other statistical assumptions we made in Theorem 1 could be removed.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Proc. NIPS*, 2011.

Azar, M. G., Lazaric, A., and Brunskill, E. Sequential transfer in multi-armed bandit with finite set of models. In *NIPS*, pp. 2220–2228, 2013.

Brunskill, E. and Li, L. Sample complexity of multi-task reinforcement learning. In *UAI*, 2013.

Cesa-Bianchi, N. and Gentile, C. Improved risk tail bounds for on-line algorithms. *IEEE Trans. on Information Theory*, 54(1):386–390, 2008.

Dekel, O., Gentile, C., and Sridharan, K. Robust selective sampling from single and multiple teachers. In *COLT*, pp. 346–358, 2010.

Djolonga, J., Krause, A., and Cevher, V. High-dimensional gaussian process bandits. In *NIPS*, pp. 1025–1033, 2013.

Kapron, Bruce M., King, Valerie, and Mountjoy, Ben. Dynamic graph connectivity in polylogarithmic worst case time. In *Proc. SODA*, pp. 1131–1142, 2013.

Maillard, O. and Mannor, S. Latent bandits. In *ICML*, 2014.

Massart, P. *Concentration Inequalities and Model Selection*. Volume 1896 of Lecture Notes in Mathematics. Springer, Berlin, 2007.

Oliveira, R.I. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2010.

Thorup, M. Decremental dynamic connectivity. In *Proc. SODA*, pp. 305–313, 1997.

Tropp, J. Freedman's inequality for matrix martingales. *arXiv preprint arXiv:1101.3039v1*, 2011.