Supplementary – Be Greedy: How Chromatic Number meets Regret Minimization in Graph Bandits

A Appendix for Section 2.2

$$
\begin{aligned}
R_T &= \sum_{t=1}^{T}(\mathbf{s}_{i_t}\tilde{\boldsymbol{\alpha}} - \mathbf{s}_{i^*}\tilde{\boldsymbol{\alpha}}) \\
&= \sum_{t=1}^{T}(\mathbf{k}_{i_t}^\top\boldsymbol{\beta} - \mathbf{k}_{i^*}^\top\boldsymbol{\beta}) + \sum_{t=1}^{T}(g_{i_t} - g_{i^*}) \\
&\le \sum_{t=1}^{T}(\mathbf{k}_{i_t}^\top\boldsymbol{\beta} - \mathbf{k}_{i^*}^\top\boldsymbol{\beta}) + \left(T\sum_{t=1}^{T}(g_{i_t} - g_{i^*})^2\right)^{1/2} \\
&\le \sum_{t=1}^{T}(\mathbf{k}_{i_t}^\top\boldsymbol{\beta} - \mathbf{k}_{i^*}^\top\boldsymbol{\beta}) + \left(4T^2\epsilon^2\right)^{1/2} \\
&\le \sum_{t=1}^{T}(\mathbf{k}_{i_t}^\top\boldsymbol{\beta} - \mathbf{k}_{i^*}^\top\boldsymbol{\beta}) + 2\epsilon T
\end{aligned}
$$

the second last inequality is due to Cauchy-Schwarz, and the last one is as $(g_{i_t} - g_{i^*}) \le 2\epsilon,\ \forall t \in [T]$.

B Discussion on Smooth rewards

By definition, any smooth function $\boldsymbol{f} : V \mapsto \mathbb{R}$ over a graph $G(V,E)$ implies $\boldsymbol{f}$ to vary slowly on the neighboring nodes of the graph G; i.e., if $(i,j) \in E$ then $f_i \approx f_j,\ \forall i,j \in V$. The standard way of defining this is by considering $\boldsymbol{f}^\top\mathbf{L}\boldsymbol{f} = \sum_{(i,j)\in E}(f_i - f_j)^2$ to be small, say $\boldsymbol{f}^\top\mathbf{L}\boldsymbol{f} \le B$, for some constant $B \in \mathbb{R}_+$ [29, 16, 34]. Clearly a small value of B implies $|f_i - f_j|$ to be small for any to neighboring nodes, i.e. $(i,j) \in E$.

We first analyze the RKHS view of the above notion of smooth reward functions. As before, let $\mathbf{L} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$ is the SVD of the Laplacian $\mathbf{L}$, where $\mathbf{Q} = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_N] \in \mathbb{R}^{N\times N}$, $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdot, \lambda_n)$ and suppose the singular values $\lambda_i = 0,\ \forall i > d$, for some $d \in [N]$. Now consider the linear space of real-valued vectors,

$$\mathcal{H}(G) = \{\mathbf{g} \in \mathbb{R}^N \mid \mathbf{g}^\top\mathbf{q}_i = 0\ \ \forall i > d\}$$

Note since $\mathbf{L} \in \mathbb{S}_+^N$ is positive semi-definite, the function $\|\cdot\|_{\mathbf{L}} : \mathcal{H}(G) \mapsto \mathbb{R}$, such that $\|\mathbf{g}\|_{\mathbf{L}} = \mathbf{g}^\top\mathbf{L}\mathbf{g}$ defines a valid norm on $\mathcal{H}(G)$. In fact, one can show that $\mathcal{H}(G)$ along with the inner product $\langle\cdot,\cdot\rangle_{\mathbf{L}} : \mathcal{H}(G) \times \mathcal{H}(G) \mapsto \mathbb{R}$, such that $\langle\mathbf{g}_1, \mathbf{g}_2\rangle_{\mathbf{L}} = \mathbf{g}_1^\top\mathbf{L}\mathbf{g}_2,\ \forall \mathbf{g}_1, \mathbf{g}_2 \in \mathcal{H}(G)$, defines a valid RKHS with respect to the reproducing kernel $\mathbf{K} = \mathbf{L}^\dagger$. This can be easily verified from the fact that $\forall \mathbf{g} \in \mathcal{H}(G)$, $\mathbf{L}^\dagger\mathbf{L}\mathbf{g} = \mathbf{g}$, and hence $\langle\mathbf{g}, \mathbf{K}_i\rangle_{\mathbf{L}} = \mathbf{g}^\top\mathbf{L}\mathbf{K}\mathbf{e}_i = (\mathbf{L}\mathbf{K}\mathbf{g})^\top\mathbf{e}_i = (\mathbf{L}^\dagger\mathbf{L}\mathbf{g})^\top\mathbf{e}_i = g_i,\ \forall i \in [N]$.

Thus the smoothness assumption on the reward function $\boldsymbol{f}$, can alternatively be interpreted as $\boldsymbol{f}$ being small in terms of the RKHS norm $\|\cdot\|_{\mathbf{L}}$. The above interpretation gives us the insight of extending the notion of "smoothness" with respect to a general RKHS norm associated to some kernel matrix $\mathbf{K} \in \mathbb{S}_+^N$. More specifically, we choose the kernel matrix $\mathbf{K}$ from the set of orthonormal kernels $\mathcal{K}(G)$ and consider $\boldsymbol{f}$ to be smooth in the corresponding RKHS norm. Note here the Hilbert space of functions $\mathcal{H}(\mathbf{K})$ is given by

$$\mathcal{H}(\mathbf{K}) = \{\mathbf{g} \in \mathbb{R}^N \mid \mathbf{g}^\top\mathbf{q}_i = 0\ \ \forall i > d\}, \tag{6}$$

where same as before, the SVD of $\mathbf{K} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$, $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_N] \in \mathbb{R}^{N\times N}$ being the orthogonal eigenvector matrix of $\mathbf{K}$, $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots \lambda_N)$ be the diagonal matrix containing singular values of $\mathbf{K}$. Clearly $\lambda_i = 0,\ \forall i > d$ implies $r(\mathbf{K}) = d$. Also we define the corresponding inner product $\langle\cdot,\cdot\rangle_{\mathbf{K}} : \mathcal{H}(\mathbf{K}) \times \mathcal{H}(\mathbf{K}) \mapsto \mathbb{R}$, as $\langle\mathbf{g}_1, \mathbf{g}_2\rangle_{\mathbf{K}} =$

$\mathbf{g}_1^\top \mathbf{K}^\dagger \mathbf{g}_2,\ \forall \mathbf{g}_1, \mathbf{g}_2 \in \mathcal{H}(\mathbf{K})$. Then similarly as above, we can show that $\mathcal{H}(\mathbf{K})$ along with $\langle \cdot, \cdot \rangle_\mathbf{K}$ defines a valid RKHS with respect to the reproducing kernel $\mathbf{K}$, as $\forall \mathbf{g} \in \mathcal{H}(\mathbf{K})$, $\langle \mathbf{g}, \mathbf{K}_i \rangle_\mathbf{K} = \mathbf{g}^\top \mathbf{K}^\dagger \mathbf{K} \mathbf{e}_i = g_i,\ \forall i \in [N]$.

The RKHS norm $\|\mathbf{g}\|_\mathbf{K} = \mathbf{g}^\top \mathbf{K}^\dagger \mathbf{g}$ defines a measure of the smoothness of $\mathbf{g}$, with respect to the kernel function $\mathbf{K}$. One way to see this is that $\forall \mathbf{g} \in \mathcal{H}(\mathbf{K})$, $\|g_i - g_j\| = \|\langle \mathbf{g}, (\mathbf{K}(i,\cdot) - \mathbf{K}(j,\cdot)) \rangle\| \le \|\mathbf{g}\|_\mathbf{K} \|\mathbf{K}(i,\cdot) - \mathbf{K}(j,\cdot)\|_\mathbf{K} = \|\mathbf{g}\|_\mathbf{K} |(K_{ii} + K_{jj} - 2K_{ij})|$, where the inequality follows from Cauchy-Schwarz [10]. Note since $\mathbf{K} \in \mathcal{K}(G)$, $K_{ii} = 1,\ \forall i \in [N]$, we have $|(K_{ii} + K_{jj} - 2K_{ij})| \le 4\ \forall i, j \in [N]$. In particular, for two similar nodes i and j, it is expected that $K(i,j) \approx 1$, in which case the quantity $|(K_{ii} + K_{jj} - 2K_{ij})| \approx 0$. Thus to impose a smoothness constraint on $\mathbf{g}$, it is sufficient to upper bound $\|\mathbf{g}\|_\mathbf{K} \le B$, for some fixed $B \in \mathbb{R}, \forall \mathbf{g} \in \mathcal{H}(\mathbf{K})$.

We thus justify our assumption of $\|\boldsymbol{f}\|_\mathbf{K} \le B$ which implies the reward vector $\boldsymbol{f}$ to be a smooth functions over the underlying graph G, with respect to embedding $\mathbf{K}$.

C Proofs for Section 3

For all proofs concerning the algorithm **SupOUCB** we denote for convenience $\mathbf{X}_j = \mathbf{X}_{s_j, t_j+1}$, similarly all notations with subscript j are to be read with subscript $(s_j, t_j + 1)$ i.e., variables at the end of phase j in algorithm **SupOUCB** Let us denote $N_j = t_j - s_j + 1$ i.e., number of rounds in phase j.

Proof of Lemma 4: Clearly, the least square solution of (4) would satisfy $\hat{\boldsymbol{\alpha}}_{s,t}\big(\mathbf{X}_{s,t}\mathbf{X}_{s,t}^\top + \mathbf{I}_N\big) - \mathbf{X}_{s,t}\mathbf{r}_{s,t} = 0$, solving which we get $\hat{\boldsymbol{\alpha}}_{s,t} = (\mathbf{X}_{s,t}\mathbf{X}_{s,t}^\top + \gamma \mathbf{I}_N)^{-1}\mathbf{X}_{s,t}\mathbf{r}_{s,t}$.

Now from Sherman–Morrison–Woodbury (SMW) formula [38] we know that for any non-singular matrix $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$, and any $\mathbf{A}_2, \mathbf{A}_2 \in \mathbb{R}^{n \times m}$,

$$(\mathbf{A}_1 + \mathbf{A}_2\mathbf{A}_3^\top)^{-1} = \mathbf{A}_1^{-1} - \mathbf{A}_1^{-1}\mathbf{A}_2(\mathbf{I}_n + \mathbf{A}_3^\top \mathbf{A}_1^{-1}\mathbf{A}_2)^{-1}\mathbf{A}_3^\top \mathbf{A}_1^{-1}.$$

Using above with $\mathbf{A}_1 = \gamma \mathbf{I}_N$ and $\mathbf{A}_2 = \mathbf{A}_3 = \mathbf{X}_{s,t}$ we get, $(\mathbf{X}_{s,t}\mathbf{X}_{s,t}^\top + \gamma \mathbf{I}_N)^{-1} = \frac{1}{\gamma}\big(\mathbf{I}_N - \mathbf{X}_{s,t}(\gamma \mathbf{I}_{t-s} + \mathbf{X}_{s,t}^\top \mathbf{X}_{s,t})^{-1}\mathbf{X}_{s,t}^\top\big)$. Thus for any $v \in [N]$, we have

$$\begin{aligned}
\hat{\boldsymbol{f}}_{s,t}(v) &= \mathbf{U}_v^\top \hat{\boldsymbol{\alpha}}_{s,t} \\
&= \frac{\mathbf{U}_v^\top}{\gamma}\big(\mathbf{I}_N - \mathbf{X}_{s,t}(\gamma \mathbf{I}_{t-s} + \mathbf{X}_{s,t}^\top \mathbf{X}_{s,t})^{-1}\mathbf{X}_{s,t}^\top\big)\mathbf{X}_{s,t}\mathbf{r}_{s,t} \\
&= \mathbf{U}_v^\top \mathbf{X}_{s,t}(\gamma \mathbf{I}_{t-s} + \mathbf{X}_{s,t}^\top \mathbf{X}_{s,t})^{-1}\mathbf{r}_{s,t} \\
&= \hat{\boldsymbol{k}}_{s,t}^{v}(\hat{\boldsymbol{K}}_{s,t} + \gamma \mathbf{I}_{t-s})^{-1}\mathbf{r}_{s,t},
\end{aligned}$$

last equality follows from the definition of $\hat{\boldsymbol{k}}_{s,t}^{v}$ and $\hat{\boldsymbol{K}}_{s,t}$. □

Proof of Lemma 7:

Let $\mathbf{V}_j = \mathbf{X}_j\mathbf{X}_j^\top + \gamma \mathbf{I}_N$, we have

$$\begin{aligned}
|\hat{\boldsymbol{f}}_j(v) - \boldsymbol{f}(v)| &= |\mathbf{u}_v^\top (\hat{\boldsymbol{\alpha}}_{j-1} - \boldsymbol{\alpha})| \\
&= |\mathbf{u}_v^\top \mathbf{V}_{j-1}^{-1/2}\mathbf{V}_{j-1}^{1/2} (\hat{\boldsymbol{\alpha}}_{j-1} - \boldsymbol{\alpha})| \\
&\le ||\mathbf{V}_{j-1}^{1/2} (\hat{\boldsymbol{\alpha}}_{j-1} - \boldsymbol{\alpha})||\ ||\mathbf{u}_v||_{\mathbf{V}_{j-1}^{-1}}
\end{aligned}$$

Using Lemma 18 which bounds the term $||\mathbf{V}_{j-1}^{1/2} (\hat{\boldsymbol{\alpha}}_{j-1} - \boldsymbol{\alpha})||$ we get

$$|\hat{\boldsymbol{f}}_j(v) - \boldsymbol{f}(v)| \le b\ ||\mathbf{u}_v||_{\mathbf{V}_{j-1}^{-1}} \tag{7}$$

Note that for the first phase, there is no previous phase hence $V_0 = I$ and $\alpha_0 = 0$ are taken. We see that $||\mathbf{u}_v||_{\mathbf{V}_{j-1}^{-1}}$ can be expressed in terms of $\hat{\boldsymbol{K}}_j$ using Sherman-Morrison-Woodbury formula as follows

$$
\begin{aligned}
\mathbf{u}_v^\top \mathbf{V}_{j-1}^{-1}\mathbf{u}_v &= \gamma^{-1}\mathbf{u}_v^\top \left(\mathbf{I}_{N_j} - \mathbf{X}_j(\gamma \mathbf{I}_{N_j} + \hat{\boldsymbol{K}}_j)^{-1}\mathbf{X}_j^\top\right)\mathbf{u}_v \\
&= \gamma^{-1}\left(1 - \left(\hat{\boldsymbol{k}}_j^v\right)^\top \mathbf{M}_j \hat{\boldsymbol{k}}_j^v\right)
\end{aligned}
$$

Using above result in (7) we get the desired result. □

Lemma 18. *Let $0 \le \delta \le 1$, then for any phase j in the algorithm SupOUCB using graph embedding $\mathbf{K}$ the following holds with high probability $1-\delta$*

$$
|| \left(\mathbf{X}_j\mathbf{X}_j^\top + \gamma \mathbf{I}_{N_j}\right)^{1/2}\left(\hat{\boldsymbol{\alpha}}_j^\top - \boldsymbol{\alpha}\right) || \le b
$$

where $b = \left(R\sqrt{c} + B\max(1, \frac{1}{\sqrt{\gamma}})\right)$ *with* $c = r(\mathbf{K}) + 2\sqrt{r(\mathbf{K})\log\frac{1}{\delta}} + 2\log\frac{1}{\delta}$.

Proof of Lemma 18:

Let $\mathbf{e}_j = [\eta s_j, \cdots, \eta_{t_j}]^\top$. Now consider the following term

$$
\begin{aligned}
\left(\hat{\boldsymbol{\alpha}}_j^\top - \boldsymbol{\alpha}\right) &= \mathbf{V}_j^{-1}\mathbf{X}_j\mathbf{r}_j - \boldsymbol{\alpha} \\
&= \mathbf{V}_j^{-1}\mathbf{X}_j\left(\mathbf{X}_j^\top\boldsymbol{\alpha} + \mathbf{e}_j\right) - \boldsymbol{\alpha} \\
&= \mathbf{V}_j^{-1}\left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right) \qquad (8)
\end{aligned}
$$

The last equality is obtained by adding and subtracting the term $\mathbf{V}_j^{-1}\boldsymbol{\alpha}$ from second equality. Now we try to bound the quantity $D(\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}) = \left(\hat{\boldsymbol{\alpha}}_j^\top - \boldsymbol{\alpha}\right)^\top V_j \left(\hat{\boldsymbol{\alpha}}_j^\top - \boldsymbol{\alpha}\right)$, it cab be verified that $D(\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}) = L_j(\boldsymbol{\alpha}) - L_j(\hat{\boldsymbol{\alpha}}_j)$ where $L_j(w) = \sum_{t=s_j}^{t_j}\left(w^\top\mathbf{x}_t - \mathbf{r}_t\right)^2$ is the cumulative squared loss.

$$
\begin{aligned}
D(\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}) &= \left(\hat{\boldsymbol{\alpha}}_j^\top - \boldsymbol{\alpha}\right)^\top \left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right) \text{ (from (8))} \\
&= \left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right)^\top \mathbf{V}_j^{-1}\left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right) \\
&= \frac{1}{\gamma}\left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right)^\top \left(\mathbf{I} - \mathbf{X}_j\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\mathbf{X}_j^\top\right)\left(\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}\right) \\
&= \frac{1}{\gamma}\Big(2\boldsymbol{\alpha}^\top\mathbf{X}_j\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\hat{\boldsymbol{K}}_j\mathbf{e}_j - \mathbf{e}_j^\top\hat{\boldsymbol{K}}_j\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\hat{\boldsymbol{K}}_j\mathbf{e}_j \\
&\qquad + \mathbf{e}_j^\top\hat{\boldsymbol{K}}_j\mathbf{e}_j + \boldsymbol{\alpha}^\top\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\top\mathbf{X}_j\mathbf{e}_j - \boldsymbol{\alpha}^\top\mathbf{X}_j\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\mathbf{X}_j^\top\boldsymbol{\alpha}\Big) \\
&= \mathbf{e}_j^\top\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\hat{\boldsymbol{K}}_j\mathbf{e}_j - 2\boldsymbol{\alpha}^\top\mathbf{X}_j\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\mathbf{e}_j + \frac{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}{\gamma} \\
&\le 2\sqrt{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}\sqrt{\mathbf{e}_j^\top\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-2}\hat{\boldsymbol{K}}_j\mathbf{e}_j} + \mathbf{e}_j^\top\left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\hat{\boldsymbol{K}}_j\mathbf{e}_j \\
&\qquad + \frac{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}{\gamma} \text{ (Cauchy-Schwartz inequality)} \\
&= ||\mathbf{A}^{1/2}\mathbf{e}_j||^2 + \frac{||\boldsymbol{\alpha}||^2}{\gamma} + 2||\boldsymbol{\alpha}||||\mathbf{B}^{1/2}\mathbf{e}_t|| \qquad (9)
\end{aligned}
$$

where $\mathbf{A} = \left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-1}\hat{\boldsymbol{K}}_j$ and $\mathbf{B} = \left(\hat{\boldsymbol{K}}_j + \gamma\mathbf{I}\right)^{-2}\hat{\boldsymbol{K}}_j$. The last inequality follows from positive semi-definite property of $\mathbf{A}$ and $\mathbf{B}$. Let $\lambda_1 \ge \cdots \ge \lambda_{N_j}$ be the eigenvalues of $\hat{\boldsymbol{K}}_j$. Now we state the concentration inequality for sub-Gaussian quadratic forms [12]

Theorem 19. *Let* $\mathbf{x} = [x_1 \cdots x_m]^\top$ *be sub-Gaussian random vector with independent entries each with mean* 0 *and variance at most* R^2 *then*

$$Pr\left[||\mathbf{Cx}||^2 \geq R^2 F(\mathbf{C}, s)\right] \leq \exp(-s)$$

where $Tr\left(\mathbf{C}^\top \mathbf{C}\right) + 2\sqrt{Tr\left((\mathbf{C}^\top \mathbf{C})^2\right) s} + 2||\mathbf{C}^\top \mathbf{C}||s$ *and* $s \leq 0$

Applying the above theorem to $||\mathbf{A}^{1/2}\mathbf{e}_t||$ we get that with high probability $1 - \delta$(setting $\exp(-s) = \delta$)

$$\begin{aligned}
||\mathbf{A}^{1/2}\mathbf{e}_t||^2 &\leq Tr\left(\mathbf{A}\right) + 2\sqrt{Tr\left(\mathbf{A}^2\right)\log\frac{1}{\delta}} + 2||\mathbf{A}||\log\frac{1}{\delta} \\
&= \sum_{i=1}^{N_j} \frac{\lambda_i}{\lambda_i + \gamma} + 2\sqrt{\sum_{i=1}^{N_j} \frac{\lambda_i^2}{(\lambda_i + \gamma)^2}\log\frac{1}{\delta}} \\
&\quad + 2\frac{\lambda_1}{\lambda_1 + \gamma}\log\frac{1}{\delta} \text{ (using definition of } \mathbf{A}) \\
&\leq r(\mathbf{K}) + 2\sqrt{r(\mathbf{K})\log\frac{1}{\delta}} + 2\log\frac{1}{\delta} \; (= c) \qquad (10)
\end{aligned}$$

The last inequality follows from the fact that if $\mathbf{K}_j$ rank is $r(j)$ then except $r(j)$ number of eigenvalues the rest of the eigenvalues will be zero and $r(\mathbf{K}_j) \leq r(\mathbf{K})$. Now again we apply theorem 19 to the term $||\mathbf{B}^{1/2}\mathbf{e}_j||$ we get that with high probability $1 - \delta$

$$\begin{aligned}
||\mathbf{B}^{1/2}\mathbf{e}_j||^2 &\leq Tr\left(\mathbf{B}\right) + 2\sqrt{Tr\left(\mathbf{B}^2\right)\log\frac{1}{\delta}} + 2||\mathbf{B}||\log\frac{1}{\delta} \\
&= \sum_{i=1}^{N_j} \frac{\lambda_i}{(\lambda_i + \gamma)^2} + 2\frac{\lambda_1}{(\lambda_i + \gamma)^2}\log\frac{1}{\delta} \\
&\quad + 2\sqrt{\sum_{i=1}^{N_j} \frac{\lambda_i^2}{(\lambda_i + \gamma)^4}\log\frac{1}{\delta}} \text{ (from definition of B)} \\
&\leq r(\mathbf{K}) + 2\sqrt{r(\mathbf{K})\log\frac{1}{\delta}} + 2\log\frac{1}{\delta} \; (= c) \qquad (11)
\end{aligned}$$

Now we make use of (10) and (11) in the equation (9) to obtain

$$\begin{aligned}
D(\hat{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}) &\leq cR^2 + \frac{||\boldsymbol{\alpha}||^2}{\gamma} + 2BR\sqrt{c} \\
&\leq cR^2 + \frac{B^2}{\gamma} + 2BR\sqrt{c} \text{ (Since } ||\boldsymbol{\alpha}|| \leq B) \\
&\leq cR^2 + B^2\max(1, \frac{1}{\gamma}) + 2BR\sqrt{c}\, \max(1, \frac{1}{\sqrt{\gamma}}) \qquad (12) \\
& \qquad (13)
\end{aligned}$$

The final inequality follows by considering 2 cases when $\gamma \geq 1$ and $\gamma < 1$. The Lemma follows from the last inequality. □

Proof of Theorem 8: Let $\mathbf{x}_*$ be the optimal arm/vertex and $\mathbf{x}_t$ be arm selected by algorithm in round t. Consider the definition of regret

$$
\begin{aligned}
R_T(\mathbf{K}) &= \sum_{t=1}^{T} (\mathbf{x}_* - \mathbf{x}_t)^\top \boldsymbol{\alpha} \\
&= \sum_{j=1}^{J} \sum_{t=s_j}^{t_j} (\mathbf{x}_* - \mathbf{x}_t)^\top \boldsymbol{\alpha} \\
&\leq \sum_{j=1}^{J} N_j \left((\mathbf{x}_* - \mathbf{x}_{t'})^\top \boldsymbol{\alpha}_{j-1} + bZ \right) \qquad (14)
\end{aligned}
$$

where $t' = \operatorname{argmax}_{t \in \{s_j, \cdots, t_j\}} (\mathbf{x}_* - \mathbf{x}_t)^\top \boldsymbol{\alpha}$ and $Z = ||\mathbf{x}_{t'}||_{\mathbf{V}_{j-1}^{-1}} + ||\mathbf{x}_*||_{\mathbf{V}_{j-1}^{-1}}$. The final inequality follows from (7) and holds good for each phase with probability $1 - J\delta$. Note that $v_* \in \mathcal{A}_j$ for every j (under high probability) which follows from Lemma 7. By applying Lemma 18 we get

$$
(\mathbf{x}_* - \mathbf{x}_{t'})^\top \boldsymbol{\alpha}_{j-1} \leq b \left(||\mathbf{x}_{t'}||_{\mathbf{V}_{j-1}^{-1}} + ||\mathbf{x}_*||_{\mathbf{V}_{j-1}^{-1}} \right) \qquad (15)
$$

Now we state 2 lemmas from [29] without proof

Lemma 20. *For all $v \in \mathcal{A}_j$, we have:*

$$
||\mathbf{u}_v||^2_{\mathbf{V}_j^{-1}} \leq \frac{1}{N_{j-1}} \sum_{l=s_{j-1}}^{t_{j-1}} ||\mathbf{x}_l||^2_{\mathbf{V}_l^{-1}}
$$

Lemma 21. *For each phase j, we have:*

$$
\sum_{l=s_j}^{t_j} ||\mathbf{x}_l||^2_{\mathbf{V}_l^{-1}} \leq \log \left(\frac{|\mathbf{V}_j|}{|\gamma \mathbf{I}_N|} \right)
$$

Now by applying Lemma 20, Lemma 21 and (15) for (14) we get

$$
R_T(\mathbf{K}) \leq 4b \sum_{j=1}^{J} N_j \sqrt{\frac{1}{N_{j-1}} \log \left(\frac{\det \mathbf{V}_{j-1}}{\det(\gamma \mathbf{I}_N)} \right)}
$$

We bound the term $\det \mathbf{V}_j$ in terms of the rank of the embedding $r(\mathbf{K})$ in the following Lemma

Lemma 22.

$$
\frac{\det \mathbf{V}_j}{\det(\gamma \mathbf{I}_N)} \leq \left(1 + \frac{N_j}{\gamma} \right)^{r(\mathbf{K})}
$$

Proof. We know that $\hat{\boldsymbol{K}}_j = \mathbf{X}_j^\top \mathbf{X}_j$ and $\mathbf{X}_j^\top \mathbf{X}_j$ have same non-zero eigenvalues. Let $\mathbf{B}_j = \hat{\boldsymbol{K}}_j + \gamma \mathbf{I}_{N_j}$, we see that for setting $t \leq N$, $\det \mathbf{V}_j = \gamma^{N - N_j} \det \mathbf{B}_j$. Moreover $\det \mathbf{B}_j = \gamma^{N_j - r(\hat{\boldsymbol{K}}_j)} \prod_{i=1}^{r(\hat{\boldsymbol{K}}_j)} (\gamma + \lambda_i)$ where λ_i are eigenvalues of $\hat{\boldsymbol{K}}_j$. So using expression for $\det \mathbf{B}_j$ we have

$$
\begin{aligned}
\frac{\det \mathbf{V}_t}{\det(\gamma \mathbf{I}_N)} &= \frac{\gamma^{N - r(\hat{\boldsymbol{K}}_j)} \prod_{i=1}^{r(\hat{\boldsymbol{K}}_j)} (\gamma + \lambda_i)}{\gamma^N} \\
&= \prod_{i=1}^{r(\hat{\boldsymbol{K}}_j)} \left(1 + \frac{\lambda_i}{\gamma} \right) \\
&\leq \left(1 + \frac{N_j}{\gamma} \right)^{r(\mathbf{K})}
\end{aligned}
$$

The final inequality follows from the fact that $r(\hat{\boldsymbol{K}}_t) \leq r(\mathbf{K})$ and $\lambda_i \leq Tr(\hat{\boldsymbol{K}}_t) = N_j$. Hence the result. ∎

By applying Lemma 22 we have

$$
\begin{aligned}
R_T(\mathbf{K}) &\leq 4b \sum_{j=1}^{J} N_j \sqrt{\frac{r(\mathbf{K})}{N_{j-1}} \log(1 + \frac{N_j}{\gamma})} \\
&\leq 4b\sqrt{2} \sum_{j=1}^{J} 2^{j/2} \sqrt{r(\mathbf{K}) \log\left(1 + \frac{T}{\gamma}\right)} \\
&\qquad \text{(using the definition } s_j) \\
&\leq 8b \sqrt{r(\mathbf{K}) T \log\left(1 + \frac{T}{\gamma}\right)}
\end{aligned}
$$

Using the definition of b we see that

$$R_T(\mathbf{K}) = O\left(r(\mathbf{K})\sqrt{T \log T}\right)$$

□

For all proofs concerning the algorithm **OUCB** we denote for convenience $\mathbf{X}_j = \mathbf{X}_{1,j}$, similarly all notations with subscript j are to be read with subscript $(1, j)$ i.e., variables at the end of round $j - 1$ in algorithm **OUCB**.

Proof of Lemma 5: From Theorem 2 in the paper [39] we have with probability $1 - \delta$

$$|\hat{\boldsymbol{f}}_{1,t}(v) - \boldsymbol{f}(v)| \leq R \sqrt{2 \log\left(\frac{\det(\mathbf{V}_t)^{1/2}}{\delta \det(\gamma \mathbf{I}_N)^{1/2}}\right)} + B \tag{16}$$

where $\mathbf{V}_t = \mathbf{X}_t^\top \mathbf{X}_t + \gamma \mathbf{I}_N$. The result follows using Lemma 22 to bound $\frac{\det \mathbf{V}_t}{\det(\gamma \mathbf{I}_N)}$. □

Proof of Theorem 6: Using Lemma 5 and Lemma 22 we get the confidence bound in round t on the reward

$$|\hat{\boldsymbol{f}}_{1,t}(v) - \boldsymbol{f}(v)| \leq R \sqrt{r(\mathbf{K}) \log(1 + \frac{t}{\gamma}) + 2 \log \frac{1}{\delta}} + B, \ (= B_t) \tag{17}$$

Let $\mathbf{x}_* = \text{argmax}_v \, \boldsymbol{\alpha}_*^T \mathbf{u}_v$. Let the regret per round be $r_t = \max_{v \in [N]} \boldsymbol{f}(v) - \hat{\boldsymbol{f}}_{1,t}(v) = \boldsymbol{\alpha}_*^T (\mathbf{x}_* - \mathbf{x}_t)$ where $\mathbf{x}_t$ is the vertex chosen in round t. We now state the following result from the proof of Theorem 1 in [29] without proof which bounds r_t with a high probability of $1 - \delta$

$$r_t \leq 2 B_t ||\mathbf{x}_t||_{\mathbf{V}_t^{-1}} \tag{18}$$

Now the cumulative regret is given by

$$
\begin{aligned}
R_T(\mathbf{K}) = \sum_{t=1}^{T} B_t &\leq \sqrt{T \sum_{t=1}^{T} r_t^2} \\
&\leq 2 B_T \sqrt{T \sum_{t=1}^{T} ||\mathbf{x}_t||^2_{\mathbf{V}_t^{-1}}} \\
&\leq 2 B_T \sqrt{2T \log \det \mathbf{V}_T}
\end{aligned}
$$

Substituting and $|V_T|$ (using Lemma 22), we get

$$R_T(\mathbf{K}) \leq 2 B_T \sqrt{2 r(\mathbf{K}) T \log\left(1 + \frac{T}{\gamma}\right)} \tag{19}$$

Substituting for B_T, we get $R_T = O(r(\mathbf{K})\sqrt{T} \log T)$ □

D Proofs for Section 4

Proof of Corollary 11

Proof. An immediate consequence of Theorem 8 gives

$$R_T = O\left(d^* \sqrt{T \log T}\right),$$

and the first claim follows by recalling that d^* satisfies the sandwich property [32]

$$\alpha(G) \leq d^* \leq \chi(\overline{G}),$$

where recall that $\alpha(G)$ denotes the independence number of G. The second claim is due to [30]. □

Proof of Lemma 12

Proof. Note than in the embedding $\mathbf{U}$, for any two vertices $i, j \in [N]$, if $(i,j) \notin E$, then $\mathbf{U}_i^\top \mathbf{U}_j = 0$. Also $\|\mathbf{U}_i\|_2 = 1$, $\forall i \in [N]$. Thus $\mathbf{U} \in Lab(G)$. Clearly $r(\mathbf{U}) = c$, since its columns consists of only c many standard basis vectors of $\mathbb{R}^N$. Moreover since $\mathbf{K} = \mathbf{U}^\top \mathbf{U}$, we have that $K_{ii} = \mathbf{U}_i^\top \mathbf{U}_i = 1$ and $K_{ij} = \mathbf{U}_i^\top \mathbf{U}_j = 0$, $\forall i, j \in [N]$ such that $(i,j) \notin E$. Thus $\mathbf{K} \in \mathcal{K}(G)$. Suppose if the SVD of $\mathbf{U}$ is given by $\mathbf{U} = \mathbf{P\Sigma Q}^\top$, we know that $\mathbf{K} = \mathbf{U}^\top \mathbf{U} = \mathbf{Q\Sigma}^2 \mathbf{Q}^\top$. Thus $rank(\mathbf{K}) = r(\mathbf{U}^\top \mathbf{U}) = r(\mathbf{U}) = c$. □

Proof of Corollary 15

Proof. If $\mathbf{U}$ is the embedding returned by Algorithm 3, then by Lemma 12, $\mathbf{K} = \mathbf{U}^\top \mathbf{U} \in \mathcal{K}(G)$ and $r(\mathbf{K}) = |C|$. The results now follows from a straightforward application of Theorem 8 with $\mathbf{K}_c = \mathbf{K}$. □

Proof of Corollary 15

Proof. From [19] we know that given any graph G', the number of colors used by the greedy coloring algorithm is at most $d_{max}(G') + 1$. Thus $|C_g(\overline{G})| \leq d_{max}(\overline{G}) + 1$. The claim now follows directly from Corollary 13. □

E Proofs for Section 5

Proof of Theorem 16

Proof. The proof is similar to the proof provided in [16]. Consider a graph with k disjoint cliques G_k. The reward structure r is given as follows:

- Every node in a clique has the same expected reward.
- A clique is chosen randomly i.e., a number I is picked uniformly at random from $\{1, \cdots, k\}$.
- For nodes $v \in G_I$ have the rewards distributed as $\mathcal{N}\left(\frac{1}{2} + \delta, 1\right)$ and the rest of nodes in other cliques $v \notin G_I$ have rewards distributed as $\mathcal{N}\left(\frac{1}{2}, 1\right)$.

Clearly the reward structure for node v has following form $r_v = f_v + \epsilon$ with $\epsilon = \mathcal{N}(0,1)$ and $f_v = \frac{1}{2} + \delta \mathbb{I}\{v \in G_I\}$ is the expected reward.

It is to be shown that expected reward $f \in \mathbb{F}(G)$. This can be seen by constructing orthogonal labeling X using k orthogonal vectors $\{x_1, \cdots, x_k\}$ and assigning x_i label to every node in clique G_i. Since X has rank k there exists some α_* such that $f = X^T \alpha_*$.

Since the actions and rewards are indistinguishable within a clique, we see that graph MAB setting reduces to picking a clique with highest reward. Thus graph MAB setting can be transformed into k-arm MAB setting with each clique representing an arm with normally distributed rewards.

Now we refer Theorem 23, which is similar to Theorem 5.1 in [6] where rewards are generated randomly according to Bernoulli distributions to obtain the lower bound. Alternatively we prove the same thoerem with rewards generated under Gaussian noise; a proof sketch of this theorem was provided by [16] but we provide the entire proof for sake of completeness.

Clearly $k = \chi(\overline{G})$ and with reduction to k-arm bandits and proof of 23 under the described reward structure, we get the theorem. □

Theorem 23. *For any number of actions $k \geq 2$ and for any time horizon T, there exists a distribution over the assignment of rewards such that the expected weak regret of any algorithm (where the expectation is taken with respect to both the randomization over rewards and the algorithms internal randomization) is at least*

$$\frac{1}{28} \min\{T, \sqrt{kT}\}$$

Proof. The proof follows the proof of [6] closely and we borrow the same notation. The rewards are constructed as follows:

- An arm I is picked uniformly at random from $A = \{1, \cdots, k\}$ arms. Arm I is termed 'good' arm.
- For arm the reward is distributed as $\mathcal{N}\left(\frac{1}{2} + \delta, 1\right)$ and the rest of arm have rewards distributed as $\mathcal{N}\left(\frac{1}{2}, 1\right)$.

The expected reward of the best arm is $\left(\frac{1}{2} + \delta\right) T$. Let $P_*\{\cdot\}$ to denote probability with respect to this random choice of rewards, and we also write $P_i\{\cdot\}$ to denote probability conditioned on i being the good arm: $P_i\{\cdot\} = P_*\{\cdot | I = i\}$. Finally, we write $P_{unif}\{\cdot\}$ to denote probability with respect to a uniformly random choice of rewards for all actions (including the good arm). Analogous expectation notation $E_*[\cdot]$, $E_i[\cdot]$, and $E_{unif}[\cdot]$ will also be used.

Let A be the algorithm and $x_{i(t)}(t)$ be the reward obtained by A by choosing arm $i(t)$ in round t. Let $r_t = x_{i(t)}(t)$ be a random variable denoting the reward received at time t, and let $\mathbf{r_t}$ denote the sequence of rewards received up through trial $t : \mathbf{r_t} = (r_1, ..., r_t)$. For shorthand, $\mathbf{r} = \mathbf{r_T}$ is the entire sequence of rewards. As usual, $G_A = \sum_{t=1}^{T} r_T$ denotes the return of the algorithm A, and $G_{max} = \max_{a \in A} \sum_{t=1}^{T} x_a(t)$ is the return of the best action. Let N_i be a random variable denoting the number of times action i is chosen by A.

Lemma 24. *Let $f : \{0, 1\}^T \to [0, M]$ be any function defined on reward sequences $\mathbf{r}$. Then for any action i,*

$$E_i[f(\mathbf{r})] \leq E_{unif}[f(\mathbf{r})] + M\delta\sqrt{2E_{unif}[N(i)]}$$

Proof. For any distributions P and Q, denote the Total Variation distance between P and Q [15] by

$$d_{TV}(P, Q) = \frac{1}{2} \max_{|h| \leq 1} \left| \int_{\Omega} hdP - \int_{\Omega} hdQ \right|$$

where $h : \Omega \to$ is a function such that $h(x) \leq 1$. Let $D_{KL}(P||Q)$ denote the KL-divergence between the distributions P and Q. We have

$$\begin{aligned} E_i[f(\mathbf{r})] - E_{unif}[f(\mathbf{r})] &= \int_{\Omega} f dP_i - \int_{\Omega} f dP_{unif} \\ &\leq M \left| \int_{\Omega} \frac{f}{M} dP_i - \int_{\Omega} \frac{f}{M} dP_{unif} \right| \\ &\leq 2M d_{TV}(P_i, P_{unif}) \end{aligned}$$

The Pinsker's inequality [15] relates d_{TV} to D_{KL}

$$2d_{TV}^2 \leq D_{KL}$$

Using Pinsker's inequality, we have

$$E_i[f(\mathbf{r})] - E_{unif}[f(\mathbf{r})] \leq M\sqrt{2D_{KL}(P_{unif}||P_i)} \tag{20}$$

Using chain rule for relative entropy, we have

$$D_{KL}(P_{unif}||P_i) = \sum_{t=1}^{T} D_{KL}(P_{unif}\{r_t|\mathbf{r_{t-1}}\}||P_i\{r_t|\mathbf{r_{t-1}}\})$$
$$= \sum_{t=1}^{T} \left(P_{unif}\{i(t) \neq i\} D_{KL}\left(\mathcal{N}_1||\mathcal{N}_1\right) + P_{unif}\{i(t) = i\} D_{KL}\left(\mathcal{N}_1||\mathcal{N}_2\right)\right)$$

where $\mathcal{N}_1 = \mathcal{N}\left(\frac{1}{2}, 1\right)$ and $\mathcal{N}_2 = \mathcal{N}\left(\frac{1}{2}+\delta, 1\right)$. Since $D_{KL}\left(\mathcal{N}_1||\mathcal{N}_2\right) = \delta^2$, we have

$$D_{KL}(P_{unif}||P_i) \leq \delta^2 \sum_{t=1}^{T} P_{unif}\{i(t) = i\} = E_{unif}\left[N(i)\right]\delta^2 \tag{21}$$

From (20) and (21) we get the theorem. □

Now we state the following theorem, we skip the proof as the steps are same as those for Theorem A.2 in [6] except that we apply Lemma 24 (instead of Lemma A.1 found in [6]) to N_i.

Theorem 25. *For any algorithm A, and for the distribution on rewards described above, the expected regret of algorithm A is lower bounded by*

$$E_*[G_{max} - G_A] \geq \delta\left(T - \frac{T}{K} - T\delta\sqrt{2\frac{T}{K}}\right)$$

Now using Theorem 25 and setting $\delta = \frac{1}{4}\min\{\sqrt{\frac{K}{T}}, 1\}$ we obtain the result.

□