Appendix for **Competing Bandits in Matching Markets**

A Proof of Theorem 1

Before we turn to the proof of Theorem 1, we present two useful technical lemmas. Throughout the remainder of this section, we say the ranking $\widehat{r}_{i,t}$ submitted by p_i at time t is *valid* if whenever an arm a_j is ranked higher than $\overline{m}(i)$, i.e. $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$, it follows that $\mu_i(j) > \mu_i(\overline{m}(i))$.

Lemma 1. *If all the agents submit valid rankings to the planner, then the GS-algorithm finds a match m such that $\mu_i(m(i)) \geq \mu_i(\overline{m}(i))$ for all players p_i.*

Proof. First we show that true agent optimal matching $\overline{m}$ is stable according to the rankings submitted by the agents when all those rankings are valid. Let a_j be an arm such that $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$ for an agent p_i. Since $\widehat{r}_{i,t}$ is valid, it means p_i prefers a_j over $\overline{m}(i)$ according to the true preferences also. However, since $\overline{m}$ is stable according to the true preferences, arm a_j must prefer player $\overline{m}^{-1}(j)$ over p_i, where $\overline{m}^{-1}(j)$ is a_j's match according to $\overline{m}$ or the emptyset if a_j does not have a match. Therefore, according to the ranking $\widehat{r}_{i,t}$, p_i has no incentive to deviate to arm a_j because that arm would reject her. Now, since $\overline{m}$ is stable according to the rankings $\widehat{r}_{i,t}$, we know that the GS-algorithm will output a matching which is at least as good as $\overline{m}$ for all agents according to the rankings $\widehat{r}_{i,t}$. Since all the rankings are valid, it follows that the GS-algorithm will output a matching m which is as least as good as $\overline{m}$ according to the true preferences also, i.e., $\mu_i(m(i)) > \mu_i(\overline{m}(i))$. □

Lemma 2. *Consider the agent p_i and let $\overline{\Delta}_{i,j} = \mu_i(\overline{m}(i)) - \mu_i(j)$ and $\overline{\Delta}_{i,\min} = \min_{j:\, \overline{\Delta}_{i,j}>0} \overline{\Delta}_{i,j}$. Then, if p_i follows the Explore-then-Commit platform (see Table 1(a)), we have*

$$\mathbb{P}(\widehat{r}_{i,hK} \text{ is invalid }) \leq Ke^{-\frac{h\overline{\Delta}_{i,\min}^2}{2}}.$$

Proof. Throughout this proof we denote $t = hK$ as a shorthand. In order for the ranking $\widehat{r}_{i,t}$ to not be valid there must exist an arm a_j such that $\mu_i(\overline{m}(i)) > \mu_i(j)$, but $\widehat{r}_{i,j}(t) < \widehat{r}_{i,\overline{m}(i)}(t)$. This can happen only when $\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)$. The probability of this event is equal to

$$\begin{aligned}\mathbb{P}\left(\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)\right) &= \mathbb{P}\left(\widehat{\mu}_{i,\overline{m}(i)}(t) - \mu_i(\overline{m}(i)) - \widehat{\mu}_{i,j}(t) + \mu_i(j) \leq \mu_i(j) - \mu_i(\overline{m}(i))\right) \\ &\leq \mathbb{P}\left(\widehat{\mu}_{i,\overline{m}(i)}(t) - \mu_i(\overline{m}(i)) - \widehat{\mu}_{i,j}(t) + \mu_i(j) \leq \overline{\Delta}_{i,\min}\right).\end{aligned}$$

Since each agent pulls each arm exactly h times during the exploration stage and since the rewards from each arm are 1-sub-Gassian, we know that $\widehat{\mu}_{i,j'}(t) - \mu_i(j') - \widehat{\mu}_{i,j}(t) + \mu_i(j)$ is $\sqrt{2/h}$-sub-Gaussian. Therefore,

$$\mathbb{P}\left(\widehat{\mu}_{i,j}(t) \geq \widehat{\mu}_{i,\overline{m}(i)}(t)\right) \leq e^{-\frac{h\Delta_i^2}{4}}.$$

The conclusion follows by a union bound over all possible arms a_j. □

Proof of Theorem 1. During the exploration stage each player p_i pulls each arm a_j exactly h times. Therefore, the expected agent-optimal stable regret of agent p_i after the first hK time steps is exactly equal to $h\sum_{j=1}^{K}\overline{\Delta}_{i,j}$ (note that $\overline{\Delta}_{i,j}$ might be negative for some values of j). The agent-optimal stable regret p_i from time $hK+1$ to time n is at most $(n-hK)\overline{\Delta}_{i,\max}$. However, from Lemma 1 we know that p_i can incur positive regret only if there exists a player who submits an invalid ranking at time $hK+1$. Lemma 2, together with a union bound over all agents, ensures that the probability there exists a player who submits an invalid ranking is at most $N\exp\left(-\frac{h\Delta^2}{4}\right)$. This completes the proof. □