# Appendix A Technical Assumptions in Section 3.1

We make two technical assumptions on the input parameters.

**(A.1)** The communication matrix $P$ is irreducible. Namely, for any two $i, j \in \{1, \cdots, N\}$, with $i \neq j$, there exists $2 \leq l \leq N$ and $k_1, \cdots, k_l \in \{1, \cdots, N\}$, with $k_1 = i$ and $k_l = j$ such that the product $P(k_1, k_2) \cdots, P(k_{l-1}, k_l) > 0$ is strictly positive.

**(A.2)** The communication budget $(B_t)_{t \in \mathbb{N}}$ and $\varepsilon > 0$ is such that for all $D > 0$, there exists $t_0(D)$ such that for all $t \geq t_0(D)$, $B_t \geq D \log(t)$ (i.e., $B_t = \Omega(\log(t))$). Furthermore, we shall assume a convexity condition, i.e., for every $x, y \in \mathbb{N}$ and $\lambda \in [0, 1]$, $A_{\lfloor \lambda x + (1-\lambda) y \rfloor} \leq \lambda A_x + (1-\lambda) A_y$, where the sequence $(A_x)_{x \in \mathbb{N}}$ is given in Equation (1). Furthermore, $\sum_{l \geq 2} \frac{A_{2l}}{A_{l-1}^3} < \infty$.

Assumption **A.1** states that the graph of communication among agents is connected. Observe that if **A.1** is not satisfied, then there exists at-least a pair of agents that can never exchange information among each other, making the setup degenerate. Assumption **A.2** implies that, any agent over a time interval of $T$ arm-pulls, can engage in information-pulls, at-least $\Omega(\log(T))$ times. The convergence of the series in **A.2** also hold true for all 'natural' examples, such as exponential and polynomial. For instance, the series is convergent if for all large $l$, either $B_l = \lceil \frac{1}{D} \log^\beta(l) \rceil$ or $B_l = \lceil l^{1/(D+1)} \rceil$, for all $D > 0$ and $\beta > 1$. Thus, conditions **A.1** and **A.2** do not impact any practical insights we can draw from our results.

# Appendix B Discussion on Theorem 1

In order to get some intuition from the Theorem, we consider a special case. Recall from Equation (1), that $A_x$ is the time slot when any agent pulls information for the $x$ th time. Thus, if for some $\beta > 1$, the communication budget $B_t = \lfloor t^{1/\beta} \rfloor$, then for all small $\varepsilon$ and all large $x$, the sequence $A_x = \lceil x^\beta \rceil$ . In other words, if communication budget scales polynomially (but sub-linearly) with time, then $A_x$ is also polynomial, but super linear. Similarly, if the gossip matrix corresponded to the complete graph, i.e., $P(i, j) = 1/N$, for all $i \neq j$ and $A_x = x^\beta$, we will show in the sequel (Corollary 18), that there exists an universal constant $C > 0$ such that $\mathbb{E}[A_{2\tau_{spr}^{(P)}}] \leq (C \log(N))^\beta$. Thus, we have the following corollary.

**Corollary 5.** *Suppose the communication budget satisfies $B_t = \lfloor t^{1/\beta} \rfloor$, for all $t \geq 1$, for some $\beta > 1$. Let $\varepsilon > 0$ be sufficiently small. Then the communication sequence $(A_x)_{x \in \mathbb{N}}$ in Equation (1) with $\varepsilon < \beta - 1$ is such that $A_x = \lceil x^\beta \rceil$, for all large $x$. If the gossip matrix connecting the agents corresponded to the complete graph, i.e., $P(i, j) = 1/N$, for all $i \neq j$, then under the conditions of Theorem 1, the regret of any agent $i \in \{1, \cdots, N\}$ at time $T \in \mathbb{N}$ satisfies*

$$\mathbb{E}[R_T^{(i)}] \leq \underbrace{\left( \sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j} \right) 4\alpha \ln(T) + \frac{K}{4}}_{\textit{Collaborative UCB Regret}} + \underbrace{\frac{4}{2\alpha - 3} \frac{\pi^2}{6} 3^\beta + 4 \max \left( K^{\frac{3}{(2\alpha - 6)}}, \left( 16\alpha \frac{2 + \lceil \frac{K}{N} \rceil}{\Delta_2^2} \right)^{\frac{\beta}{\beta - 1}} \right) + (C \log(N))^\beta}_{\textit{Cost of Infrequent Pairwise Communications}},$$

*where $C$ is an universal constant given in Corollary 18.*

The proof is provided in Appendix K. The terms denoting cost of pairwise communications correspond to the average amount of time any agent must wait before the best arm is in the playing set of that agent. This cost can be decomposed into the sum of two dominant terms. The term of order $\left( \frac{\lceil \frac{K}{N} \rceil}{\Delta_2} \right)^{\frac{2\beta}{\beta - 1}}$ is the expected number of samples needed to identify the best arm by any agent. The term $(\log(N))^\beta$ is the amount of time taken by a pure gossip process to spread a message (the best arm in our case) to all agents, if the communication budget is given by $B_t = \lfloor t^{1/\beta} \rfloor$.

# Appendix C Proof of Theorem 1

In order to give the proof, we first set some notations and definitions. We make explicit a probability space construction from (Lattimore and Szepesvári, 2018), that makes the proof simpler. We assume that there is a sequence of independent $\{0, 1\}$ valued random variables $(Y_j^{(i)}(t))_{i \in [N], j \in [K], t \geq 0}$, where for every $j \in [K]$, the collection $(Y_j^{(i)}(t))_{t\geq 0, i\in[N]}$ is an i.i.d. Bernoulli random variable of mean $\mu_j$. The interpretation being that if an agent $i$ pulls arm $j$ for the $l$th time, it will receive reward $Y_j^{(i)}(l)$. Additionally, we also have on the probability space a sequence of independent $[N]$ valued random variables $(Z_j^{(i)})_{j\geq 0, i\in[N]}$, where for each $i\in[N]$, the sequence $(Z_j^{(i)})_{j\geq 0}$ is iid distributed as $P(i,\cdot)$. The interpretation is that when agent $i$ wishes to receive a recommendation at the end of phase $j$, it will do so from agent $Z_j^{(i)}$.

### C.1 Definitions and Notations

In order to analyze the algorithm, we set some definitions. Let $\mathcal{B}_j^{(i)}$ to be the best arm in $S_j^{(i)}$, i.e., $\mu_{\mathcal{B}_j^{(i)}} = \max_{l\in S_j^{(i)}} \mu_l$. Observe that since the set $S_j^{(i)}$ is random, $\mathcal{B}_j^{(i)}$ is also a random variable. For every agent $i\in[N]$ and phase $j\geq 0$, we denote by $\widehat{\mathcal{O}}_j^{(i)} \in S_j^{(i)}$ to be that arm, that agent $i$ played the most in phase $j$. Note, from the algorithm, if any agent $i^{'}$ pulled an arm from agent $i$ at the end of phase $j$ for a recommendation, it would have received arm $\widehat{\mathcal{O}}_j^{(i)}$.

Fix an agent $i\in[N]$ and phase $j\geq 0$. Let $\mathcal{S}^{(i)}$ be a collection of all subsets $S\subset[K]$ of cardinality $|S| = \lceil \frac{K}{N}\rceil + 2$, such that $1\in S, \widehat{S}^{(i)} \subset S$. For any $S\in\mathcal{S}^{(i)}$, index the elements in $S$ as $\{l_1,\cdots,l_{\lceil \frac{K}{N}\rceil+2}\}$ in increasing order of arm-ids. Let $a_1,\cdots a_{\lceil \frac{K}{N}\rceil+2} \in \mathbb{N}^{\lceil \frac{K}{N}\rceil+2}$ be such that $\sum_{m=0}^{\lceil \frac{K}{N}\rceil+2} a_m \geq 0$. For every agent $i\in[N]$, phase $j\geq 0$ and $(a_1,\cdots,a_{\lceil \frac{K}{N}\rceil+2}) \in \mathbb{N}^{\lceil \frac{K}{N}\rceil+2}$, denote by the event $\xi_j^{(i)}(S;a_1,\cdots,a_{\lceil \frac{K}{N}\rceil})$ as

$$\xi_j^{(i)}(S;a_1,\cdots,a_{\lceil \frac{K}{N}\rceil}) := \left\{ S_j^{(i)} = S, T_{l_1}(A_{j-1}) = a_1, \cdots, T_{l_{\lceil \frac{K}{N}\rceil+2}(A_{j-1})} = a_{\lceil \frac{K}{N}\rceil+2}, \widehat{\mathcal{O}}_j^{(i)} \neq 1 \right\}.$$

Denote by $\Xi_j^{(i)}$ as the union of all such events, i.e.,

$$\Xi_j^{(i)} := \bigcup_{S\in\mathcal{S}^{(i)}} \left( \bigcup_{\left(a_1,\cdots a_{\lceil \frac{K}{N}\rceil+2}\right)\in\mathbb{N}^{\lceil \frac{K}{N}\rceil+2}} \xi_j^{(i)}(S;a_1,\cdots,a_{\lceil \frac{K}{N}\rceil}) \right),$$

and by $\chi_j^{(i)}$ its indicator random variable, i.e.,

$$\chi_j^{(i)} = \mathbf{1}_{\Xi_j^{(i)}}. \tag{6}$$

In words, the event $\chi_j^{(i)}$ is the indicator variable indicating whether agent $i$ does not recommend the best arm at the end of phase $j$, under *some sample path*, i.e., we take an union over all possible set of playing arms that contain arm 1 (i.e., set $\mathcal{S}^{(i)}$) and all possible number of plays of the various arms in $S$ until the beginning of phase $j$ (i.e., the set of histories in $\mathcal{A}_j$). In Lemma 6, we provide an upper bound to this quantity. Notice from the construction that for each agent $i\in[N]$ and phase $j\geq 0$, the random variable $\chi_j^{(i)}$ is measurable with respect to the reward sequence $(Y_j^{(i)}(t))_{j\in[K], t\in[0,A_j]}$. Also, trivially by definition, observe that $\chi_j^{(i)} \geq \mathbf{1}_{\widehat{\mathcal{O}}_j^{(i)}\neq 1, 1\in S_j^{(i)}}$ almost-surely. This is so since $\chi_j^{(i)}$ is an union bound over all possible realizations of the communication sequence and reward sequence of other agents, while $\mathbf{1}_{\widehat{\mathcal{O}}_j^{(i)}\neq 1, 1\in S_j^{(i)}}$ considers a particular realization of the communication and rewards of other agents.

We now define certain random times that will be useful in the analysis.

$$\widehat{\tau}_{stab}^{(i)} = \inf\{j^{'} \geq j^* : \forall j \geq j^{'}, \chi_j^{(i)} = 0\},$$
$$\widehat{\tau}_{stab} = \max_{i\in[N]} \widehat{\tau}_{stab}^{(i)},$$

$$\begin{aligned}\widehat{\tau}^{(i)}_{spr} &= \inf\{j \geq \widehat{\tau}_{stab} : 1 \in S^{(i)}_j\} - \widehat{\tau}_{stab},\\ \widehat{\tau}_{spr} &= \max_{i \in \{1,\cdots,N\}} \widehat{\tau}^{(i)}_{spr},\\ \tau &= \widehat{\tau}_{stab} + \widehat{\tau}_{spr}.\end{aligned}$$

In words, $\widehat{\tau}^{(i)}_{stab}$ is the earliest phase such that, for all subsequent phases, if agent $i$ has the best arm, then it will recommend the best arm. The time $\widehat{\tau}^{(i)}_{spr}$ is the number of phases it takes after $\widehat{\tau}_{stab}$ for agent $i$ to have arm 1 in its playing set. The following proposition follows from the definition of the random times.

**Proposition 1.** *For all agents $i \in \{1,\cdots,N\}$, we have almost-surely,*

$$\begin{aligned}\bigcap_{j \geq \tau} S^{(i)}_j &= S^{(i)}_{\tau},\\ \widehat{\mathcal{O}}^{(i)}_l &= 1 \ \forall l \geq \tau, \ \forall i \in \{1,\cdots,N\}.\end{aligned}$$

*Proof.* Fix any agent $i \in [N]$ and any phase $j \geq \tau$. Since $\tau \geq \widehat{\tau}^{(i)}_{stab}$, we have for all $j \geq \tau$,

$$\chi^{(i)}_j = 0. \tag{7}$$

Furthermore, from the definition of $\chi^{(i)}_j$, we know that

$$\chi^{(i)}_j \geq \mathbf{1}_{1 \in S^{(i)}_j, \widehat{\mathcal{O}}^{(i)}_j \neq 1}, \tag{8}$$

almost-surely. However, as $\tau \geq \widehat{\tau}^{(i)}_{spr} + \widehat{\tau}_{stab}$, we know that

$$1 \in S^{(i)}_j. \tag{9}$$

Thus, from Equations (7), (8) and (9), we have that $\widehat{\mathcal{O}}^{(i)}_j = 1$. Since $j \geq \tau$ was arbitrary, we have that for all $j \geq \tau$, $\widehat{\mathcal{O}}^{(i)}_j = 1$. Since agent $i \in [N]$ was arbitrary, we have that for all agents $i \in [N]$ and all phases $j \geq \tau$, we have $\widehat{\mathcal{O}}^{(i)}_j$=1. From the Algorithm, we know that any agent will change its set of arms only if the recommendation it receives is not present in the playing set (see line 8 of Algorithm 1). The preceding argument says that is not the case and hence for all agents $i \in [N]$, $\bigcap_{j \geq \tau} S^{(i)}_j = S^{(i)}_{\tau}$. □

In other words, after phase $\tau$, the system is *frozen*, i.e., the set of arms of all agents remain fixed for *all time in the future.* Moreover, all agents will only recommend the best arm going forward from this phase. We will show in the sequel that $\mathbb{E}[A_\tau] < \infty$ for all settings of the algorithm and hence the system freezes after only almost-surely finitely many changes in the set of arms played by the different agents.

### C.2 Intermediate Propositions

**Proposition 2.** *The regret of any agent $i \in \{1,\cdots,N\}$ after playing for $T$ steps is bounded by*

$$\mathbb{E}[R^{(i)}_T] \leq \mathbb{E}[A_\tau] + \frac{K}{4} + 4\alpha \ln(T) \left( \sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j} \right).$$

*Proof.* From the definition of regret, we can write,

$$R^{(i)}_T = \sum_{l=1}^{T} (\mu_1 - \mu_{I^{(i)}_l}),$$

$$
\begin{aligned}
&= \sum_{l=1}^{T}\sum_{j=2}^{K}\Delta_j \mathbf{1}_{I_l^{(i)}=j},\\
&\leq A_\tau + \sum_{l=A_\tau+1}^{T}\sum_{j=2}^{K}\Delta_j \mathbf{1}_{I_l^{(i)}=j},\\
&= A_\tau + \sum_{j=2}^{K}\Delta_j \sum_{l=A_\tau+1}^{T}\mathbf{1}_{I_l^{(i)}=j}.
\end{aligned}
$$

Thus, taking expectations on both sides, we get that

$$
\mathbb{E}[R_T^{(i)}] \leq \mathbb{E}[A_\tau] + \sum_{j=2}^{K}\Delta_j \sum_{l=A_\tau+1}^{T}\mathbb{P}[I_l^{(i)}=j, j\in S_\tau^{(i)}]. \tag{10}
$$

We can break up the summation on the RHS as follows. Fix an arm $j \in \{2,\cdots,K\}$ and evaluate the sum

$$
\begin{aligned}
\sum_{l=A_\tau+1}^{T}\mathbb{P}[I_l^{(i)}=j, j\in S_\tau^{(i)}] &= \sum_{l=A_\tau+1}^{T}\mathbb{P}\left[I_l^{(i)}=j, T_j^{(i)}(l)\leq \frac{4\alpha\ln(T)}{\Delta_j^2}, j\in S_\tau^{(i)}\right] + \qquad (11)\\
&\sum_{l=A_\tau+1}^{T}\mathbb{P}\left[I_l^{(i)}=j, T_j^{(i)}(l)\geq \frac{4\alpha\ln(T)}{\Delta_j^2}, j\in S_\tau^{(i)}\right],\\
&\leq \frac{4\alpha\ln(T)}{\Delta_j^2}\mathbb{P}[j\in S_\tau^{(i)}] + \sum_{l=A_\tau+1}^{T}\mathbb{P}\left[I_l^{(i)}=j, T_j^{(i)}(l)\geq \frac{4\alpha\ln(T)}{\Delta_j^2}\right],\\
&\leq \frac{4\alpha\ln(T)}{\Delta_j^2}\mathbb{P}[j\in S_\tau^{(i)}] + \sum_{l=3}^{\infty}2l^{2(1-\alpha)}, \qquad (12)
\end{aligned}
$$

where in the last line we substitute the classical estimate from (Auer et al., 2002). We can use this estimate, as we know that both the best arm, i.e., arm indexed 1 and the sub-optimal arm indexed $j$ are in the set $S_\tau^{(i)}$ and hence the agent can potentially play those arms. Now plugging Equation (12) into Equation (10), we get that

$$
\begin{aligned}
\mathbb{E}[R_T^{(i)}] &\leq \mathbb{E}[A_\tau] + \sum_{j=2}^{K}\Delta_j\left(\frac{4\alpha\ln(T)}{\Delta_j^2}\mathbb{P}[j\in S_\tau^{(i)}] + \sum_{l=3}^{\infty}2l^{2(1-\alpha)}\right),\\
&\overset{(a)}{\leq} \mathbb{E}[A_\tau] + \frac{4\alpha\ln(T)}{\Delta}\sum_{j=2}^{K}\mathbb{P}[j\in S_\tau^{(i)}] + \sum_{j=2}^{K}\frac{\Delta_j}{4},\\
&\overset{(b)}{\leq} \mathbb{E}[A_\tau] + \frac{4\alpha\ln(T)}{\Delta}\left(\left\lceil\frac{K}{N}\right\rceil + 2\right) + \frac{K}{4}.
\end{aligned}
$$

In step $(a)$, we use the bound that $\Delta_j \geq \Delta$, for all $j\in\{2,\cdots,K\}$ and the fact that for $\alpha > 3$, we have $\sum_{l=3}^{\infty}2l^{2(1-\alpha)} \leq 1/8$. In step $(b)$, we use the crucial identity that for any agent $i\in\{1,\cdots,N\}$ and any phase $\psi$ either deterministic or random, we have almost-surely,

$$
\sum_{j=1}^{K}\mathbf{1}_{j\in S_\psi^{(i)}} = \left\lceil\frac{K}{N}\right\rceil + 2.
$$

Taking expectations on both sides yields the result. If one were more precise in step $(a)$, then it is possible to establish that

$$
\mathbb{E}[R_T^{(i)}] \leq \mathbb{E}[A_\tau] + 4\alpha\ln(T)\sum_{j=2}^{K}\frac{1}{\Delta_j}\mathbb{P}[j\in S_\tau^{(i)}] + \sum_{j=2}^{K}\frac{\Delta_j}{4},
$$

$$\leq \mathbb{E}[A_\tau] + 4\alpha \ln(T) \left( \sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j} \right) + \sum_{j=2}^{K} \frac{\Delta_j}{4}.$$

This will then yield the proof. □

**Proposition 3.** *For all $N \in \mathbb{N}$, $\Delta \in (0,1]$, $\alpha > 3$ and $M > 0$,*

$$\mathbb{E}[A_\tau] \leq A_{j^*} + \frac{2}{2\alpha - 3} \sum_{l \geq \frac{j^*}{2} - 1} \frac{A_{2l+1}}{A_{l-1}^3} + \mathbb{E}[A_{2\widehat{\tau}_{spr}}],$$

*where $j^*$ is defined in Theorem 1.*

*Proof.* Recall the fact that for any $\mathbb{N}$ valued random variable $X$, its expectation can be written as a sum of its tail probabilities, i.e., $\mathbb{E}[X] = \sum_{t \geq 1} \mathbb{P}[X \geq t]$. We use this fact to bound the expected value of $\mathbb{E}[A_\tau]$ as

$$\begin{aligned}
\mathbb{E}[A_\tau] &= \sum_{t \geq 1} \mathbb{P}[A_\tau \geq t], \\
&\overset{(a)}{\leq} \sum_{t \geq 1} \mathbb{P}[\tau \geq A^{-1}(t)], \\
&= \sum_{t \geq 1} \mathbb{P}[\widehat{\tau}_{stab} + \widehat{\tau}_{spr} \geq A^{-1}(t)], \\
&\leq \sum_{t \geq 1} \mathbb{P}\left[\widehat{\tau}_{stab} \geq \frac{1}{2}(A^{-1}(t))\right] + \sum_{t \geq 1} \mathbb{P}\left[\widehat{\tau}_{spr} \geq \frac{1}{2}\left(A^{-1}(t)\right)\right], \\
&\leq A_{j^*} + \sum_{t \geq A_{j^*}+1} \mathbb{P}\left[\widehat{\tau}_{stab} \geq \frac{1}{2}\left(A^{-1}(t)\right)\right] + \mathbb{E}[A_{2\widehat{\tau}_{spr}}].
\end{aligned}$$

Step $(a)$ follows from the definition of $A^{-1}(\cdot)$ given in Theorem 1. The estimate for $\mathbb{E}[A(2\widehat{\tau}_{spr})]$ follows by noticing that this random variable can be coupled to the spreading time for a classical rumor spreading model, which we do so in the sequel in Proposition 4. The first summation can be bounded by using estimates from Lemma 6. We do so by applying a union bound over all agents and phases as follows. Fix some $x \geq j^*/2$ in the following calculations.

$$\begin{aligned}
\mathbb{P}[\widehat{\tau}_{stab} \geq x] &= \mathbb{P}\left[\bigcup_{i=1}^{N} \widehat{\tau}_{stab}^{(i)} \geq x\right], \\
&\leq \sum_{i=1}^{N} \mathbb{P}[\widehat{\tau}_{stab}^{(i)} \geq x], \\
&= \sum_{i=1}^{N} \mathbb{P}\left[\bigcup_{l=x}^{\infty} \chi_l^{(i)} = 1\right], \\
&\leq \sum_{i=1}^{N} \sum_{l \geq x} \mathbb{P}\left[\chi_l^{(i)} = 1\right], \\
&\overset{(a)}{\leq} \sum_{i=1}^{N} \sum_{l \geq x} \frac{2}{2\alpha - 3} \binom{K}{2} \left( \left\lceil \frac{K}{N} \right\rceil + 1 \right) A_{l-1}^{-(2\alpha-3)}, \\
&= \sum_{l \geq x} \frac{2}{2\alpha - 3} N \binom{K}{2} \left( \left\lceil \frac{K}{N} \right\rceil + 1 \right) A_{l-1}^{-(2\alpha-3)}, \\
&\overset{(b)}{\leq} \frac{2}{2\alpha - 3} \sum_{l \geq x} A_{l-1}^{-3},
\end{aligned}$$

In the above calculations, we use the bound from Lemma 6 in step $(a)$ as $x \geq j^*/2$. In step $(b)$, we use $N\binom{K}{2}\left(\left\lceil\frac{K}{N}\right\rceil+1\right) \leq \left(A_{\frac{j^*}{2}-1}\right)^{2\alpha-6}$, which follows from the definition of $j^*$ given in Theorem 1. Thus, we can obtain the following.

$$\begin{aligned}
\sum_{t \geq A_{j^*}+1} \mathbb{P}\left[\widehat{\tau}_{stab} \geq \frac{1}{2}\left(A^{-1}(t)\right)\right] &\leq \sum_{t \geq A_{j^*}+1}\left(\frac{2}{2\alpha-3}\right)\sum_{l \geq \frac{1}{2}A^{-1}(t)} A_{l-1}^{-3}, \\
&\leq \left(\frac{2}{2\alpha-3}\right)\sum_{t \geq A_{j^*}+1}\sum_{l \geq \frac{1}{2}A^{-1}(t)} A_{l-1}^{-3}, \\
&\overset{(c)}{\leq} \left(\frac{2}{2\alpha-3}\right)\sum_{l \geq \frac{1}{2}A^{-1}(A_{j^*}+1)}\sum_{t=A_{j^*}+1}^{A_{2l}} A_{l-1}^{-3}, \\
&\leq \left(\frac{2}{2\alpha-3}\right)\sum_{l \geq \frac{j^*}{2}} \frac{A_{2l}}{A_{l-1}^{3}} < \infty.
\end{aligned}$$

Step $(c)$ follows by swapping the order of summations. The condition **A.2** in Section 3.1 satisfied by the sequence $(A_j)_{j \in \mathbb{N}}$ ensures that the last summation is finite.

□

**Proposition 4.** *The random variable $\widehat{\tau}_{spr}$ is stochastically dominated by $\tau_{spr}^{(P)}$.*

*Proof.* We construct a coupling of the spreading process induced by our algorithm and a PULL based rumor spreading on $P$. We construct the coupling as follows. First we sample the reward vectors $(Y_j^{(i)}(t))_{i \in [N], j \in [K], t \geq 0}$. Then we can construct the random variable $\widehat{\tau}_{stab}$, which is a measurable function of the reward vectors. We then sample the communication random variables of our algorithm $(Z_j^{(i)})_{i \in [N], j \geq 0}$. We then construct a PULL based communication protocol with the random variables $(Z_j^{(i)})_{i \in [N], j \geq \widehat{\tau}_{stab}}$. Since $\widehat{\tau}_{stab}$ is independent of $(Z_j^{(i)})_{i \in [N], j \geq 0}$, the sequence of $(Z_{j-\widehat{\tau}_{stab}}^{(i)})_{i \in [N], j \geq \widehat{\tau}_{stab}}$ is identically distributed as $(Z_j^{(i)})_{i \in [N], j \geq 0}$.

Now, for the stochastic domination, consider the case where in the PULL based system, which starts at phase (time) $\widehat{\tau}_{stab}$, only agent 1 has the rumor (best-arm). By definition of $\widehat{\tau}_{stab}$, any agent that contacts another agent possesing the rumor (best-arm), is also aware of the rumor (best-arm). The stochastic domination is concluded as at phase $\widehat{\tau}_{stab}$, many agents may be aware of the rumor (best-arm) in our algorithm, while in the rumor spreading process, only agent 1 is aware of the rumor at phase $\widehat{\tau}$.

□

### Proof of Theorem 1

*Proof.* We can conclude Theorem 1 by plugging in the estimates from Propositions 3 and 4 into Proposition 2. □

## Appendix D Analysis of the UCB Error Estimates

**Lemma 6.** *For any agent $i \in [N]$ and phase $j$ such that $\frac{A_j - A_{j-1}}{2+\lceil\frac{K}{n}\rceil} \geq 1 + \frac{4\alpha\log(A_j)}{\Delta^2}$, we have*

$$\mathbb{E}[\chi_j^{(i)}] \leq \frac{2}{2\alpha-3}\binom{K}{2}\left(\left\lceil\frac{K}{N}\right\rceil+1\right)\left(\frac{1}{A_{j-1}^{2\alpha-3}}\right),$$

*where $\chi_j^{(i)}$ is defined in Equation (6).*

*Proof.* As the algorithm recommends the most played arm in a phase, the arm that is recommended (i.e., $\mathcal{O}_j^{(i)}$) must be payed by agent $i$ at-least $\frac{A_j - A_{j-1}}{|S_j^{(i)}|}$ times in phase $j$. This follows from an elementary pigeon hole argument. Let $\mathcal{S}^{(i)}$ be the collection of all subsets $S \subset \{1, \cdots, K\}$ such that $\widehat{S}^{(i)} \subset S$ and $1 \in S$. Let $\mathcal{A}_j$ be a collection of all $\mathbb{N}$ valued tuples $(a_1, \cdots a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}$ s.t. $\sum_{m=0}^{\lceil \frac{K}{N} \rceil + 2} a_m = A_{j-1}$. We shall however, consider all possible histories, i.e., $\mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}$.

$$
\begin{aligned}
\mathbb{E}[\chi_j^{(i)}] &\overset{(a)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \mathbb{P}\left[ \bigcup_{(a_1, \cdots a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}} \chi_j^{(i)}(S; a_1, \cdots, a_{\lceil \frac{K}{N} \rceil}) \right], \\
&\overset{(b)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \mathbb{P}\left[ \bigcup_{(a_1, \cdots a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}} \bigcup_{l \in S, l \neq 1} T_l^{(i)}(A_j) - T_l^{(i)}(A_{j-1}) \geq \frac{A_j - A_{j-1}}{|S|} \right], \\
&\overset{(c)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \sum_{t = A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}}^{A_j} \mathbb{P}\left[ \bigcup_{(a_1, \cdots a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}} \bigcup_{l \in S, l \neq 1} T_l^{(i)}(t-1) - T_l^{(i)}(A_{j-1}) = \frac{A_j - A_{j-1}}{|S_j^{(i)}|} - 1, I_t^{(i)} = l \right], \\
&\overset{(d)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \sum_{t = A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}}^{A_j} \sum_{l \in S, l \neq 1} \mathbb{P}\left[ \bigcup_{(a_1, \cdots a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}} T_l^{(i)}(t-1) - T_l^{(i)}(A_{j-1}) = \frac{A_j - A_{j-1}}{|S_j^{(i)}|} - 1, I_t^{(i)} = l \right], \\
&\overset{(e)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \sum_{l \in S, l \neq 1} \sum_{t = A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}}^{A_j} \mathbb{P}\left[ T_l^{(i)}(t-1) \geq \frac{A_j - A_{j-1}}{|S_j^{(i)}|} - 1, \mathrm{UCB}_l^{(i)}(t) \geq \mathrm{UCB}_1^{(i)}(t) \right], \qquad (13) \\
&\overset{(f)}{\leq} \sum_{S \in \mathcal{S}^{(i)}} \sum_{l \in S, l \neq 1} \sum_{t = A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}}^{A_j} 2t^{2(1-\alpha)}. \qquad (14)
\end{aligned}
$$

Step $(a)$ follows from an union bound over $\mathcal{S}^{(i)}$. In step $(b)$ we use the fact that if an arm $l$ has to be the most played, then it must be played at-least $\frac{A_j - A_{j-1}}{|S|}$ times. In step $(c)$, we search over times, when the number of times arm $l$ has been played exceeds $\frac{A_j - A_{j-1}}{|S|}$ exactly. In step $(d)$, we use an union bound over $S$. In step $(e)$, for any arm $l \in S$, $\mathrm{UCB}_l^{(i)}(t) = \widehat{\mu}_l^{(i)}(t-1) + \sqrt{\frac{\alpha \ln(t)}{T_l^{(i)}(t-1)}}$. In step $(e)$, we ask that arm $l$ and 1 has been played at-least 0 or more times in the past before time $t$ and that the UCB index of arm $l$ at agent $i$ at time $t$, exceed that of the index of the best arm. In step $(f)$, we plug in the classical estimate from (Auer et al., 2002). This bound is applicable in our case as $1 \in S$ and the arm gap between the best and the second best arm in $S$ is at-least $\Delta$. Furthermore, the condition in the lemma $\frac{A_j - A_{j-1}}{2 + \lceil \frac{K}{n} \rceil} \geq 1 + \frac{4\alpha \log(A_j)}{\Delta^2}$ implies that for all $t \in \left[ A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}, A_j \right]$, the conditions in the bound in (Auer et al., 2002) is satisfied and is hence applicable. Notice that $|\mathcal{S}^{(i)}| \leq \binom{K}{2}$. Thus, switching the order of summation and simplifying Equation 14, we get

$$
\begin{aligned}
\mathbb{E}[\chi_j^{(i)}] &\leq \binom{K}{2} \left( \left\lceil \frac{K}{N} \right\rceil + 1 \right) \sum_{t = A_{j-1} + \frac{A_j - A_{j-1}}{|S_j^{(i)}|}}^{A_j} 2t^{2(1-\alpha)} \\
&\leq 2 \binom{K}{2} \left( \left\lceil \frac{K}{N} \right\rceil + 1 \right) \int_{A_{j-1}}^{A_j} u^{2(1-\alpha)} \mathrm{d}u
\end{aligned}
$$

**Algorithm 3** Asynch GosInE Algorithm (at Agent $i$)

```
Input: Communication Budget $(B_t)_{t\in\mathbb{N}}$, UCB Parameter $\alpha$, Slack $\delta$, $\varepsilon > 0$
Initialization: $\widehat{S}^{(i)}, S_0^{(i)}$ according to Equations (2) and (3) respectively.
$j \leftarrow 0$
$A_j = \max\left(\min\{t \geq 0, B_t \geq j\}, \lceil (1+j)^{1+\varepsilon} \rceil\right)$    ▷ Reparametrize the communication budget
$\mathcal{P}_j^{(i)} \sim \mathrm{Unif}[(A_j - A_{j-1}), (1+\delta)(A_j - A_{j-1})]$    ▷ Uniformly distributed phase length
for Time $t \in \mathbb{N}$ do
    Pull - $\arg\max_{l \in S_i^{(j)}} \left( \widehat{\mu}_l^{(i)}(t-1) + \sqrt{\frac{\alpha \ln(t)}{T_l^{(i)}(t-1)}} \right)$
    if $t == \sum_{y=0}^{j} \mathcal{P}_y^{(i)}$ then
        $\mathcal{O}_j^{(i)} \leftarrow$ GET-ARM-PREV$(i, t)$
        if $\mathcal{O}_j^{(i)} \notin S_i^{(j)}$ then
            $U_{j+1}^{(i)} \leftarrow \arg\max_{l \in \{U_j^{(i)}, L_j^{(i)}\}} \left( T_l\left(\sum_{y=0}^{j} \mathcal{P}_y^{(i)}\right) - T_l\left(\sum_{y=0}^{j-1} \mathcal{P}_y^{(i)}\right)\right)$    ▷ Most played arm in current phase
            $L_{j+1}^{(i)} \leftarrow \mathcal{O}_j^{(i)}$
            $S_{j+1}^{(i)} \leftarrow \widehat{S}^{(i)} \cup L_{j+1}^{(i)} \cup U_{j+1}^{(i)}$    ▷ Update set of playing arms
        else
            $S_{j+1}^{(i)} \leftarrow S_j^{(i)}$.
        $j \leftarrow j + 1$
        $A_j = \max\left(\min\{t \geq 0, B_t \geq j\}, \lceil (1+j)^{1+\varepsilon} \rceil\right)$    ▷ Reparametrize the communication budget
        $\mathcal{P}_j^{(i)} \sim \mathrm{Unif}[(A_j - A_{j-1}), (1+\delta)(A_j - A_{j-1})]$    ▷ Update next phase length
```

$$\leq \frac{2}{2\alpha - 3}\binom{K}{2}\left(\left\lceil \frac{K}{N} \right\rceil + 1\right)\left(\frac{1}{A_{j-1}^{2\alpha-3}} - \frac{1}{A_j^{2\alpha-3}}\right).$$

□

Similarly, we also have a bound on the error probability in the case of random phase length system in the following lemma.

**Lemma 7.** *For any agent $i \in [N]$ and every $j \in \mathbb{N}$ such that $\frac{A_j - A_{j-1}}{2 + \lceil \frac{K}{n} \rceil} \geq 1 + \frac{4\alpha \log(A_j)}{\Delta^2}$, we have*

$$\mathbb{E}[\chi_j^{(i)} \mathbf{1}_{j \geq H^*}] \leq \frac{2}{2\alpha - 3}\binom{K}{2}\left(\left\lceil \frac{K}{N} \right\rceil + 1\right)\left(\frac{1}{A_{j-1}^{2\alpha-3}}\right),$$

*where $\chi_j^{(i)}$ is defined in Equation (16).*

*Proof.* The proof is identical to that in Lemma 6 upto Equation 13, where the upper limit of summation is $(1+\delta)A_j$ in the asynchronous communication scenario. Continuing with the rest of the calculation, identical to that in Lemma 6 yields the result. □

## Appendix E Asynchronous GosInE Algorithm

We give the pseudo code of the Asynchronous GosInE in Algorithm 3.

## Appendix F Poisson Asynchronous Algorithm - Buildup to Proof of Theorem 2

In order to prove Theorem 2, we will state a more general algorithm in the sequel in Algorithm 5 and prove a performance bound on it in Theorem 8. We shall then subsequently prove Theorem 8 in Appendix G and as a corollary of the proof, deduce Theorem 2 in Appendix H.

**Algorithm 4** Asynch Arm Recommendation

```
procedure GET-ARM-PREV($(i,t)$)                                   ▷ Input an agent $i$ and time $t$
    $m \sim P(i,\cdot)$                                            ▷ Sample another agent
    $j \leftarrow \inf\{r \geq 0 : \sum_{y=0}^{r} \mathcal{P}_y^{(m)} \geq t\}$   ▷ Phase of agent $m$ at time $t$
    $\mathcal{Y}_{j-1}^{(m)} \leftarrow \sum_{y=0}^{j-1} \mathcal{P}_y^{(m)}$, $\mathcal{Y}_{j-2}^{(m)} \leftarrow \sum_{y=0}^{j-2} \mathcal{P}_y^{(m)}$
return $\arg\max_{l \in S_m^{(j-1)}} \left(T_l^{(m)}(\mathcal{Y}_{j-1}^{(m)}) - T_l^{(m)}(\mathcal{Y}_{j-2}^{(m)})\right)$   ▷ Most played arm in phase $j-1$
```

**Algorithm 5** Distributed Poisson Asynchronous MAB Regret Minimization (at Agent $i$)

```
Input Parameters: Communication Budget $(B_t)_{t\in\mathbb{N}}$, UCB Parameter $\alpha$, Slack $\delta, \varepsilon > 0$
Initialization: $\widehat{S}^{(i)}, S_i^{(0)}$ according to Equations (2) and (3).
$j \leftarrow 0$
$A_j = \max\left(\inf\{t \geq 0, B_t \geq j\}, (1+j)^{1+\varepsilon}\right)$      ▷ Reparametrize the commuication budget
$\mathcal{P}_j \sim \text{Poisson}[(A_j - A_{j-1}), (1+\delta)(A_j - A_{j-1})]$    ▷ Poisson distributed phase length
for Time $t \in \mathbb{N}$ do
    Pull arm - $\arg\max_{l \in S_i^{(j)}} \left(\widehat{\mu}_l^{(i)}(t-1) + \sqrt{\frac{\alpha \ln(t)}{T_l^{(i)}(t-1)}}\right)$
    if $t == \sum_{y=0}^{j} \mathcal{P}_y$ then
        $O_i^{(j)} \leftarrow$ GET-ARM-PREV($i$)                         ▷ Given in Algorithm 4
        if $O_i^{(j)} \not\in S_i^{(j)}$ then
            $U_{j+1}^{(i)} \leftarrow \arg\max_{l \in \{U_j^{(i)}, L_j^{(i)}\}} (T_l(A_j) - T_l(A_{j-1}))$   ▷ The most played arm
            $L_{j+1}^{(i)} \leftarrow O_j^{(i)}$                              ▷ Update the set of playing arms
            $S_{j+1}^{(i)} \leftarrow \widehat{S}^{(i)} \cup L_{j+1}^{(i)} \cup U_{j+1}^{(i)}$
        else
            $S_{j+1}^{(i)} \leftarrow S_j^{(i)}$.
        $j \leftarrow j+1$
        $A_j = \max\left(\inf\{t \geq 0, B_t \geq j\}, (1+j)^{1+\varepsilon}\right)$   ▷ Reparametrize the commuication budget
        $\mathcal{P}_j \sim \text{Poisson}[(A_j - A_{j-1}), (1+\delta)(A_j - A_{j-1})]$
```

This algorithm does not fit our framework exactly, as the communication budget is not necessarily met. In particular, this algorithm only ensures that *with high probability*, the number of information pulls by agents in the first $t$ time slots is within the prescribed budget $B_t$. Thus, we present this algorithm in the Appendix and not as a solution to the multi-agent MAB problem. In order to prove this result, we will need a further assumption on the input parameters.

**(A.3) -** The communication budget $(B_t)_{t\in\mathbb{N}}$ and $\varepsilon > 0$ is such that $\exists\ \kappa > 0$, $\sum_{x\geq 1} A_{A_x} e^{-\kappa(A_x - A_{x-1})} < \infty$, where $(A_x)_{x\in\mathbb{N}}$ is given in Equation (1).

**Theorem 8.** *Suppose in a system of $N \geq 2$ agents connected by a communication matrix $P$ satisfying assumption* **A.1** *and $K \geq 2$ arms, each agent runs Algorithm 5, with input parameters $(B_t)_{t\in\mathbb{N}}$, and the UCB parameter $\alpha > 3$ and $\varepsilon > 0$ satisfying assumptions* **A.2** *and* **A.3** *and $\delta > 0$ such that $\exists D > 0$ with $c(\delta) \geq \frac{5}{4}D$ and $(3 + 2\delta + \ln(4+2\delta)) \geq \frac{5}{4}D^{-1}$, where $c(\delta) = \min\left(\frac{\delta}{2} + \ln\left(1+\frac{\delta}{2}\right), (1+\delta)\ln\left(\frac{2+2\delta}{2+\delta}\right) - \frac{\delta}{2}\right)$. Then the regret of any agent $i \in [N]$, after any time $T \in \mathbb{N}$ is bounded by*

$$\mathbb{E}[R_T^{(i)}] \leq \underbrace{\left(\sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j}\right) 4\alpha \ln(T) + \frac{K}{4}}_{Collaborative\ UCB\ Regret} + \underbrace{(1+\delta)\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}] + \widehat{g}_1((A_x)_{x\in\mathbb{N}}, \delta) + N\widehat{g}_2((A_x)_{x\in\mathbb{N}}, \delta)}_{Cost\ of\ Asynchronous\ Infrequent\ Pairwise\ Communications},$$

*where*

$$\widehat{g}_1((A_x)_{x\in\mathbb{N}},\delta) = 2(1+\delta)\left(A_{2\lceil 2+\delta\rceil j^*} + \left(\frac{2}{2\alpha-3}\right)\sum_{l\geq 3}\frac{A_{2l}}{A_{l-1}^3} + 2\sum_{x\geq\lceil\frac{j^*}{2}\rceil}(A_{\lceil 4\lceil 2+\delta\rceil x\rceil})^{(2(2\alpha-6)+2)}e^{-c(\delta)(A_{x+1}-A_x)}\right),$$

*and*

$$\widehat{g}_2((A_x)_{x\in\mathbb{N}},\delta) = 2\left(A_{x_0}^2 e^{-c(\delta)(A_{x_0}-A_{x_0-1})} + \sum_{x\geq 1}A_x e^{-c(\delta)A_x^{\omega}}\right) + \frac{1}{c(\delta)} +$$
$$(1+\delta)\left(2\sum_{x\geq 1}A_{\lceil A_x\rceil}e^{-c(\delta)(A_x-A_{x-1})} + N\sum_{t\geq 1}e^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}\right),$$

*where $j^*$ is given in Theorem 1, and $x_0\in\mathbb{N}$ is from Assumption* **A.2** *in Section 3.1.*

The proof of this theorem is carried out in Appendix G.

## Appendix G Proof of Theorem 8

For every agent $i\in[N]$ and phase $j\geq 0$, we shall denote by $\mathcal{P}_j^{(i)}\in\mathbb{N}$ to be the number of times agent $i$ pulls an arm in phase $j$. Notice from the conditions on the input parameter $(p_j)_{j\in\mathbb{N}}$ that the following property is satisfied -

$$\begin{gathered}\sum_{j\geq 0}\mathbb{P}[\mathcal{P}_j\leq A_j-A_{j-1}]<\infty,\\ \sum_{j\geq 0}\mathbb{P}[\mathcal{P}_j\geq(1+\delta)(A_j-A_{j-1})]<\infty.\end{gathered}\tag{15}$$

To make things simpler, we shall consider the following probability space. As before, it contains the reward and communication random variables $(Y_j^{(i)}(t))_{i\in[N],j\in[K],t\geq 0}$ and $(Z_j^{(i)})_{j\geq 0,i\in[N]}$. For every $j\in[K]$, the collection $(Y_j^{(i)}(t))_{t\geq 0,i\in[N]}$ is an i.i.d. Bernoulli random variable of mean $\mu_j$. The interpretation being that if an agent $i$ pulls arm $j$ for the $l$th time, it will receive reward $Y_j^{(i)}(l)$. Similarly, for each $i\in[N]$, the sequence $(Z_j^{(i)})_{j\geq 0}$ is iid distributed as $P(i,\cdot)$. The interpretation is that when agent $i$ wishes to receive a recommendation at the end of phase $j$, it will do so from agent $Z_j^{(i)}$. In addition, we also assume that the probability space consists of another independent sequence $(\mathcal{P}_j^{(i)})_{i\in[N],j\geq 0}$, where for each $i\in[N]$ and $j\geq 0$, the random variable $\mathcal{P}_j^{(i)}$ is independent of everything else and distributed as a Poisson random variable with mean $\left(1+\frac{\delta}{2}\right)(A_{j+1}-A_j)$.

### G.1 Definition and Notations

To proceed with the analysis, define by a $\mathbb{N}$ valued random variable $H^*$ as

$$H^* = \inf\left\{j^{'}\geq 0:\forall i\in[1,N],\forall j\geq j^{'},\mathcal{P}_j^{(i)}\in[A_j-A_{j-1},(1+\delta)(A_j-A_{j-1})]\right\},$$

Equations (15) imply from Borel Cantelli lemma that $H^*<\infty$ almost-surely. We will need another random variable $\Gamma\in\mathbb{N}$, which is defined as

$$\Gamma = \sup\left\{t\geq 0:\exists i\in[N],\sum_{j=0}^{H^*}\mathcal{P}_j^{(i)}\geq t\right\}.$$

In words, $\Gamma$ represents the time when the last agent shift to phase $H^*$. Similar to that done in the proof of Theorem 1, we define a sequence of indicator random variables $(\chi_j^{(i)})_{j\geq 0,i\in[N]}$ as follows. The definition is identical to the one used in the proof of Theorem 1, which we reproduce here for completeness. Fix some agent $i \in [N]$ and phase $j \geq 0$. Denote by $\mathcal{S}^{(i)}$ as the collection of all subsets $S$ of $[K]$ with cardinality $\lceil \frac{K}{N} \rceil + 2$, such that $\widehat{S}^{(i)} \subset S$ and $1 \in S$. Clearly, $|\mathcal{S}^{(i)}| \leq \binom{K}{2}$. Denote by the tuples $(a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}$ such that $\sum_{m=0}^{\lceil \frac{K}{N} \rceil + 2} a_m \geq 0$. For any set $S \in \mathcal{S}^{(i)}$ and tuples $(a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathcal{A}_j$, denote by the event $\xi_j^{(i)}(S, a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2})$ as

$$\xi_j^{(i)}(S, a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2}) := \left\{ S_j^{(i)} = S, T_{l_1}(A_{j-1}) = a_1, \cdots, T_{l_{\lceil \frac{K}{N} \rceil + 2}}(A_{j-1}) = a_{\lceil \frac{K}{N} \rceil + 2}, \widehat{\mathcal{O}}_j^{(i)} \neq 1 \right\}.$$

Denote by $\Xi_j^{(i)}$ as the union of all such events above, i.e.,

$$\Xi_j^{(i)} := \bigcup_{S \in \mathcal{S}^{(i)}} \left( \bigcup_{(a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2}) \in \mathbb{N}^{\lceil \frac{K}{N} \rceil + 2}} \xi_j^{(i)}(S, a_1, \cdots, a_{\lceil \frac{K}{N} \rceil + 2}) \right).$$

Denote by $\chi_j^{(i)}$ as the indicator random variable, i.e.,

$$\chi_j^{(i)} = \mathbf{1}_{\Xi_j^{(i)}}. \tag{16}$$

Observe that, as before, for all agents $i \in [N]$ and phases $j \geq 0$, the random variable $\xi_j^{(i)}$ is measurable with respect to the reward sequence $(Y_l^{(i)})_{l \leq A_j}$. Furthermore, we have the almost-sure inequality that

$$\xi_j^{(i)} \geq \mathbf{1}_{1 \in S_j^{(i)}, \widehat{\mathcal{O}}_j^{(i)} \neq 1, j \geq H^*}.$$

This follows from the same reasoning as in Theorem 1 as $\xi_j^{(i)}$ considers all possible sample paths for communication while $\mathbf{1}_{1 \in S_j^{(i)}, \widehat{\mathcal{O}}_j^{(i)} \neq 1, j \geq H^*}$ is for a particular sample path of communications among agents. Notice that since the phase lengths are random, we can only reason about the sample path for agent phases larger than or equal to $H^*$.

Similar to before, we define the random variables $\widehat{\tau}_{stab}^{(i)}, \widehat{\tau}_{stab}, \widehat{\tau}_{spr}^{(i)}$ and $\widehat{\tau}_{spr}$. Denote by $\tau = \widehat{\tau}_{stab} + \widehat{\tau}_{spr}$. These definitions from the Proof of Theorem 1 are reproduced here for completeness.

$$\begin{aligned}
\widehat{\tau}_{stab}^{(i)} &= \inf\{j^{'} \geq j^* : \forall j \geq j^{'}, \chi_j^{(i)} = 0\}, \\
\widehat{\tau}_{stab} &= \max_{i \in [N]} \widehat{\tau}_{stab}^{(i)}, \\
\widehat{\tau}_{spr}^{(i)} &= \inf\{j \geq \widehat{\tau}_{stab} : 1 \in S_j^{(i)}\} - \widehat{\tau}_{stab}, \\
\widehat{\tau}_{spr} &= \max_{i \in \{1, \cdots, N\}} \widehat{\tau}_{spr}^{(i)}, \\
\tau &= \widehat{\tau}_{stab} + \widehat{\tau}_{spr}.
\end{aligned}$$

From the definitions, the statement and proof of Proposition 1 holds verbatim for the present algorithm as well. We will need two additional definitions to help state our result. Denote by $T_{stab} \in \mathbb{N}$ to be the first time when all agents pull arms and are in phase $\widehat{\tau}_{stab}$ or larger, i.e.,

$$T_{stab} = \sup \left\{ t \geq \Gamma : \exists i \in [N], \sum_{j=0}^{\widehat{\tau}_{stab}} \mathcal{P}_j^{(i)} \geq t \right\}.$$

Similarly, define $H$ to be the maximum over all agents phases at time $T_{stab}$, i.e.,

$$H = \sup \left\{ j \geq 0 : \exists i \in [N], \sum_{l=0}^{j} \mathcal{P}_l^{(i)} \leq T_{stab} \right\}.$$

Similarly, denote by $\mathcal{T}$ as the first time when all agents pull arms in phase $\tau$ or larger, i.e.,

$$\mathcal{T} = \sup \left\{ t \geq 0 : \exists i \in [N], \sum_{j=0}^{\tau} \mathcal{P}_j^{(i)} \geq t \right\}$$

### G.2 Structural Results

In this section, we give inequalities relating the random variables defined in the previous section, that will be helpful in proving Theorem 2.

**Lemma 9.**

$$\mathbb{E}[\mathcal{T}] \leq \mathbb{E}[T_{stab}] + (1+\delta)(\mathbb{E}[A_{H+(2\tau_{spr}^{(P)}-1)\lfloor 2+\delta \rfloor+1} - A_H])$$

*where the random variable $\tau_{spr}^{(P)}$ is independent of $H$.*

*Proof.* The proof consists of three steps. First, we will construct a coupling with a standard PULL based rumor process on the communication matrix $P$ such that $H$ and $\tau_{spr}^{(P)}$ are independent. Then we shall argue a stochastic domination and for the constructed coupling show that, almost-surely, we have

$$\mathcal{T} \leq T_{stab} + (1+\delta)(A_{H+(\tau_{spr}^{(P)}-1)\lfloor 2+\delta \rfloor+1} - A_H) \tag{17}$$

where $\tau_{spr}^{(P)}$ is independent of $H$ and $\leq_{st}$ represents stochastic domination. This will then conclude the proof by taking expectations on both sides.

**(1) Coupling Construction** - We proceed with the coupling as follows. We assume that our probability space consists of the random variables $(Y_j^{(i)})_{i\in[N],j\in[K],l\geq 0}, (\mathcal{P}_j^{(i)})_{j\geq 0,i\in[N]}, (Z_j^{(i)})_{j\geq 0,i\in[N]}$ and $(\widehat{Z}_j^{(i)})_{j\geq 0,i\in[N]}$. The sequence $(Y_j^{(i)})_{i\in[N],j\in[K],l\geq 0}$ is independent of everything else and is used to construct the observed rewards of agents. The sequence $(\mathcal{P}_j^{(i)})_{j\geq 0,i\in[N]}$ is independent of everything else and denotes the phase length random variables of agents as before. The sequence $(\widehat{Z}_j^{(i)})_{j\geq 0,i\in[N]}$ denotes a standard PULL based rumor spreading process on $P$, independent of everything else. In other words, for each agent $i\in[N]$, the sequence $(\widehat{Z}_j^{(i)})_{j\geq 0}$ is i.i.d., with each element distributed according to the distribution $P(i,\cdot)$. Thus, they represent the sequence of callers called by agent $i$ in the PULL based rumor process. The random communication sequence $(Z_j^{(i)})_{j\geq 0,i\in[N]}$ will be constructed such that it is independent of $(Y_j^{(i)})_{i\in[N],j\in[K],l\geq 0}, (\mathcal{P}_j^{(i)})_{j\geq 0,i\in[N]}$, and equal in distribution to $(\widehat{Z}_j^{(i)})_{j\geq 0,i\in[N]}$ such that the stochastic domination in Equation (17) holds.

To do so, we will recursively define a sequence of random times $(t_i)_{i\geq 0}$ which are measurable with respect to the agent rewards and phases, i.e., for all $i\geq 0$, $t_i \in \sigma((Y_j^{(i)}(l))_{i\in[N],j\in[K],l\geq 0}, (\mathcal{P}_j^{(i)})_{j\geq 0,i\in[N]})$. Let $t_0 = T_{stab}$. We know that $T_{stab}$ is measurable only with respect to the reward random variables $(Y_j^{(i)}(l))_{i\in[N],j\in[K],l\geq 0}$ and the phase lengths of the agents $(\mathcal{P}_j^{(i)})_{j\geq 0,i\in[N]}$. For all $i\geq 1$, let $t_i$ be the first time after $t_{i-1}$, such that all agents have changed phase at-least once in the time interval $[t_{i-1}, t_i]$. More formally, we have

$$t_i = \inf\left\{x > t_{i-1} : \forall i\in[N], \exists j\in\mathbb{N} \text{ s.t.} \sum_{l=0}^{j}\mathcal{P}_l^{(i)} \geq t_{i-1}, \sum_{l=0}^{j+1}\mathcal{P}_l^{(i)} \leq x\right\}.$$

We construct another sequence of random variables $(j_x^{(i)})_{x\geq 0,i\in[N]}$, where for every agent $i\in[N]$ and $x\geq 0$, $j_x^{(i)}$ is the first phase change of agent $i$ in the time interval $[t_x, t_{x+1})$ of our algorithm, i.e.,

$$j_x^{(i)} = \inf\left\{j\geq 0 : \sum_{l=0}^{j-1}\mathcal{P}_l^{(i)} < t_x, \sum_{l=0}^{j}\mathcal{P}_l^{(i)} \geq t_x\right\},$$

where $\sum_{l=0}^{-1} = 0$. By construction observe that for all agents $i\in[N]$ and all $x\geq 0$, the random variable $j_x^{(i)}$ is measurable with respect to the rewards and phase lengths.

Equipped with these definitions, we construct the communication random variables of our algorithm $(Z_j^{(i)})_{j\geq 0,i\in[N]}$ as follows. For every agent $i\in[N]$ and $x\geq 0$, we let

$$Z_{j_{2x}^{(i)}}^{(i)} = \widehat{Z}_x^{(i)}.$$

For an agent $i \in [N]$, and any phase $j \notin \{j_x^{(i)}, x \geq 0\}$, we let $Z_j^{(i)}$ be i.i.d., from $P(i, \cdot)$.

We only look at alternate intervals $[t_0, t_1], [t_2, t_3]$ and so on because in our algorithm, an agent recommends the most played arm in the *previous phase*. Thus, if an agent becomes aware of the best arm in interval say $[t_0, t_1]$, then it will definitely recommend it in phase $[t_2, t_3]$, if asked, as since $t_2 \geq \Gamma$, agent will recommend the best arm, and moreover at-least one phase elapses after the agent receives the best arm.

**(2) Stochastic Domination** - We now conclude about the stochastic domination as follows. In the algorithm, we will only consider even time intervals $[t_0, t_1], [t_2, t_3]$ and so on, where an agent becomes newly aware of the best arm. This is so since our recommendation algorithm only recommends the best arm in the previous phase. At time $t_0$, exactly one agent knows the rumor in the PULL rumor spreading process while potentially more agents may be aware of the rumor (best-arm) in the algorithm. Furthermore, we consider that there is exactly one communication request in the rumor spreading process per even time-interval, (i.e., in $[t_0, t_1), [t_2, t_3)$ and so on), while potentially many more can occur in our algorithm. Thus, we have the following almost-sure bound under the afore mentioned coupling,

$$\mathcal{T} \leq t_{2\tau_{spr}^{(P)}}. \tag{18}$$

**(3) Deterministic Bounds on** $(t_x)_{x \geq 0}$- If we further establish that for all $x \geq 0$, almost-surely, we have

$$t_x \leq T_{stab} + (1+\delta)(A_{H+(x-1)\lfloor 2+\delta \rfloor + 1} - A_H), \tag{19}$$

then we can conclude the proof from Equations (18) and (19). To establish Equation (19), first observe that $T_{stab} \geq \Gamma$ almost-surely. Thus, if any agent $i \in [N]$ will be in any phase $j$, for at-least $A_j - A_{j-1}$ number of arm-pulls and for at-most $(1+\delta)(A_j - A_{j-1})$ number of arm-pulls. Thus, by definition at time $t_0$, we know that no agent is in phase $H+1$ or beyond. Thus, at time $t_0 + (1+\delta)(A_{H+1} - A_H)$, we know that all agents would have changed phase at-least once after $t_0$. Thus, $t_1 \leq t_0 + (1+\delta)(A_{H+1} - A_H)$ almost-surely.

We now make the above into an induction argument. For the base case, suppose that at time $t_0$, all agents are within phase $H$ – which is true by definition. For all $0 \leq x^{'} \leq x$, assume the induction hypothesis that

$$t_{x'+1} \leq t_{x'} + (1+\delta)(A_{H+x'\lfloor 2+\delta \rfloor + 1} - A_{H+x'\lfloor 2+\delta \rfloor}),$$

and that all agents at time $t_{x'+1}$ are at phase $H + (x^{'}+1)\lfloor 2+\delta \rfloor$ or lower. Since $t_x \geq \Gamma$, we know that in the time interval $[t_x, t_x + (1+\delta)(A_{H+x\lfloor 2+\delta \rfloor+1} - A_{H+x\lfloor 2+\delta \rfloor})]$, all agents would have changed phase at-least once. Thus, $t_{x+1} \leq t_x + (1+\delta)(A_{H+x\lfloor 2+\delta \rfloor+1} - A_{H+x\lfloor 2+\delta \rfloor})$. It now remains to conclude that all agents will be in phase $H + (x+1)\lfloor 2+\delta \rfloor$ or lower at time $t_x + (1+\delta)(A_{H+x\lfloor 2+\delta \rfloor+1} - A_{H+x\lfloor 2+\delta \rfloor})$. Notice that the maximum phase any agent can be in at time $t_x + (1+\delta)(A_{H+x\lfloor 2+\delta \rfloor+1} - A_{H+x\lfloor 2+\delta \rfloor})$, given that it was in a phase $H + x\lfloor 2+\delta \rfloor$ or lower at time $t_x$ is bounded above by Proposition 6 as $H + x\lfloor 2+\delta \rfloor + \lfloor 2+\delta \rfloor$. This then concludes the induction step and hence we have for all $x \geq 0$, almost-surely, by a simple telescoping sum

$$\begin{aligned} t_x &\leq t_0 + (1+\delta)\sum_{l=0}^{x-1}(A_{H+l\lfloor 2+\delta \rfloor+1} - A_{H+l\lfloor 2+\delta \rfloor}), \\ &\leq t_0 + (1+\delta)(A_{H+(x-1)\lfloor 2+\delta \rfloor+1} - A_H). \end{aligned}$$

□

**Lemma 10.** *For any agent $i \in [N]$, the regret after it has pulled arms for $T$ times is bounded by*

$$\mathbb{E}[R_T^{(i)}] \leq \mathbb{E}[\mathcal{T}] + \frac{K}{4} + 4\alpha \left( \sum_{j=1}^{\lceil \frac{K}{N} \rceil + 1} \frac{1}{\Delta_j} \right) \ln(T).$$

*Proof.* The proof of this Lemma follows similarly to that of Lemma 2. We can write the regret of any agent $i \in [N]$ as follows -

$$\begin{aligned} R_T^{(i)} &= \sum_{t=1}^{T} \mu_1 - \mu_{I_t^{(i)}}, \\ &\leq \mathcal{T} + \sum_{t=\mathcal{T}+1}^{T} \mu_1 - \mu_{I_t^{(i)}}, \\ &= \mathcal{T} + \sum_{t=\mathcal{T}+1+1}^{T} \sum_{l=2}^{K} \Delta_l \mathbf{1}_{I_t^{(i)}=l}, \\ &\overset{(a)}{=} \mathcal{T} + \sum_{l=2}^{K} \Delta_l \sum_{t=\mathcal{T}+1}^{T} \mathbf{1}_{I_t^{(i)}=l} \mathbf{1}_{l \in S_\tau^{(i)}} \end{aligned}$$

In step $(a)$, we use Proposition 1 that at time $\mathcal{T}$, all agents are in a phase that is at-least $\tau$. Furthermore, from Proposition 1 (recall that the statement and proof of Proposition 1 holds verbatim for the present case also) implies almost-surely that, for all $j \geq \tau$, and all $i \in [N]$, $S_j^{(i)} = S_\tau^{(i)}$. Taking expectations on the last display yields

$$\mathbb{E}[R_T^{(i)}] \leq \mathbb{E}[\mathcal{T}] + \sum_{l=2}^{K} \Delta_l \sum_{t=\mathcal{T}+1}^{T} \mathbb{P}[I_t^{(i)} = l, l \in S_\tau^{(i)}]$$

Using the same techniques as in the proof of Proposition 2, i.e., following all steps from Equation 12 onwards, one obtains

$$\sum_{l=2}^{K} \Delta_l \sum_{t=\mathcal{T}+1}^{T} \mathbb{P}[I_t^{(i)} = l, l \in S_\tau^{(i)}] \leq \left( \sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j} \right) 4\alpha \ln(T) + \frac{K}{4}$$

□

**Lemma 11.**

$$\mathbb{E}[\mathcal{T}] \leq \mathbb{E}[\Gamma] + (1+\delta)\mathbb{E}[A_{\widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{4\Gamma}] + (1+\delta)\mathbb{E}[A_{4\lceil 1+\delta \rceil \widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}].$$

*Proof.* From Lemma 9, we know that

$$\begin{aligned} \mathbb{E}[\mathcal{T}] &\leq \mathbb{E}[T_{stab}] + (1+\delta)\mathbb{E}[(A_{H + (\tau_{spr}^{(P)} - 1)\lfloor 2+\delta \rfloor + 1} - A_H)], \\ &\overset{(a)}{\leq} \mathbb{E}[T_{stab}] + (1+\delta)(\mathbb{E}[A_{2H}] + \mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}]), \\ &\overset{(b)}{\leq} \mathbb{E}[\Gamma] + (1+\delta)\mathbb{E}[A_{\widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{2\Gamma + 2\lceil 1+\delta \rceil \widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}], \\ &\overset{(c)}{\leq} \mathbb{E}[\Gamma] + (1+\delta)\mathbb{E}[A_{\widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{4\Gamma}] + (1+\delta)\mathbb{E}[A_{4\lceil 1+\delta \rceil \widehat{\tau}_{stab}}] + (1+\delta)\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}]. \end{aligned}$$

Steps $(a)$ and $(c)$ follow from the elementary fact that for any two random variables $X$ and $Y$ and any invertible function $f(\cdot)$, $\mathbb{E}[f(X+Y)] \leq \mathbb{E}[f(2X)] + \mathbb{E}[f(2Y)]$. Step $(b)$ follows from Lemma 12.

□

**Lemma 12.**

$$\mathbb{E}[T_{stab}] \leq \mathbb{E}[\Gamma] + (1+\delta)\mathbb{E}[A_{\widehat{\tau}_{stab}}].$$

*Proof.* The first inequality follows as the time taken to reach $T_{stab}$ is upper bounded by the time it takes all agents to reach phase $\widehat{\tau}_{stab}$ after time $\Gamma$. However, by definition we know that all agents last in any phase $j$ after time $\Gamma$ for at-most $(1+\delta)(A_{j+1}-A_j)$ arm-pulls. The upper bound is concluded by noticing that an agent can be in a phase no smaller than 0 at time $\Gamma$ and subsequently it takes an agent a maximum of $(1+\delta)A_{\widehat{\tau}_{stab}}$ time to reach phase $\widehat{\tau}_{stab}$. □

**Lemma 13.** *Almost-surely, we have*

$$H \leq \Gamma + \lceil 1+\delta \rceil \widehat{\tau}_{stab}.$$

*Proof.* Notice that at time $\Gamma$, the maximum phase any agent can be in is $\Gamma$. This follows from the trivial upper bound, where in each time step, an agent increases its phase by one in each time slot. After time $\Gamma$, we know by definition, that any agent plays arms at-least $A_j - A_{j-1}$ times and at-most $(1+\delta)(A_j - A_{j-1})$ in phase $j$. Thus, the total number of phase changes an agent will have in the time interval $[\Gamma+1, T_{stab}]$ is at-most $A^{-1}(T_{stab}-\Gamma)$. Thus, we get

$$\begin{aligned} H &\leq \Gamma + A^{-1}(T_{stab}-\Gamma), \\ &\overset{(a)}{\leq} \Gamma + A^{-1}((1+\delta)A_{\widehat{\tau}_{stab}}), \\ &\overset{(b)}{\leq} \Gamma + A^{-1}(A_{\lceil 1+\delta \rceil \widehat{\tau}_{stab}}), \\ &\leq \Gamma + \lceil 1+\delta \rceil \widehat{\tau}_{stab}. \end{aligned}$$

Step $(a)$ follows from Lemma 12, step $(b)$ follows from convexity of $(A_x)_{x\geq 1}$ and the last inequality follows from the definition of $A^{-1}(\cdot)$.

□

### G.3 Quantitative Results

In this section, we compute quantitative bounds in terms of the algorithm' input parameters.

**Proposition 5.** *For all $x \geq 2$ and $\delta > 0$,*

$$\mathbb{P}[H^* > l] \leq 2N \sum_{x \geq l} e^{-c(\delta)(A_x - A_{x-1})},$$

*where* $c(\delta) = \min\left(\frac{\delta}{2} + \ln\left(1+\frac{\delta}{2}\right), (1+\delta)\ln\left(\frac{2+2\delta}{2+\delta}\right) - \frac{\delta}{2}\right)$.

*Proof.* From the definition of $H^*$, we have

$$\begin{aligned} \mathbb{P}[H^* \geq l] &= \mathbb{P}\left[\bigcup_{i=1}^{N}\bigcup_{x\geq l} \text{Poisson}\left(\left(1+\frac{\delta}{2}\right)(A_x - A_{x-1})\right) \notin [(A_x - A_{x-1}), (1+\delta)(A_x - A_{x-1})]\right], \\ &\leq N \sum_{x\geq l} \mathbb{P}\left[\text{Poisson}\left(\left(1+\frac{\delta}{2}\right)(A_x - A_{x-1})\right) \notin [(A_x - A_{x-1}), (1+\delta)(A_x - A_{x-1})]\right], \\ &\leq N \sum_{x\geq l} (2e^{-2c(\delta)(A_x - A_{x-1})}). \end{aligned}$$

In the last inequality, we use the classical large-deviation estimate for a Poisson random variable. □

**Lemma 14.**

$$\mathbb{E}[\Gamma] \leq 2N\left(A_{x_0}^2 e^{-c(\delta)(A_{x_0} - A_{x_0 - 1})} + \sum_{x\geq 1} A_x e^{-c(\delta)A_{x-1}^{\omega}}\right) + \frac{N}{c(\delta)},$$

*where $c(\delta)$ is given in Proposition 5 and $x_0$ is from Assumption* **A.2** *in Section 3.1.*

*Proof.* We start by computing the tail probability $\mathbb{P}[\Gamma > t]$. The key observation to do so is the following inequality. For every $L \geq 0$, we have

$$\mathbb{P}[\Gamma \geq t] \leq \mathbb{P}[H^* \geq L] + \mathbb{P}\left[\bigcup_{i=1}^{N} \sum_{j=0}^{L} \mathcal{P}_j^{(i)} \leq t\right].$$

We will then compute $\mathbb{E}[\Gamma]$ by choosing $L = A^{-1}(t)$. We shall compute each of these terms separately.

$$\begin{aligned}
\mathbb{P}[H^* \geq A^-(t)] &= \mathbb{P}\left[\bigcup_{i=1}^{N} \bigcup_{x \geq A^{-1}(t)} \mathcal{P}_x^{(i)} \notin [(A_x - A_{x-1}), (1+\delta)(A_x - A_{x-1})]\right], \\
&\leq N \sum_{x \geq A^{-1}(t)} 2e^{-c(\delta)(A_x - A_{x-1})},
\end{aligned}$$

where the second inequality follows from Proposition 5. Similarly, standard large deviation estimates for Poisson random variables (observe that for all $L$,$\sum_{j=0}^{L} \mathcal{P}_j^{(i)}$ is Poisson distributed with mean $L$ ) and union bound gives

$$\mathbb{P}\left[\bigcup_{i=1}^{N} \sum_{j=0}^{L} \mathcal{P}_j^{(i)} \leq t\right] \leq N e^{-c(\delta)t}.$$

Thus, we can bound $\mathbb{E}[\Gamma]$ as

$$\begin{aligned}
\mathbb{E}[\Gamma] &\leq \sum_{t \geq 1} \mathbb{P}[\Gamma \geq t], \\
&\leq 2N \sum_{t \geq 1} \sum_{x \geq A^{-1}(t)} e^{-c(\delta)(A_x - A_{x-1})} + N \sum_{t \geq 1} e^{-c(\delta)t}, \\
&\overset{(a)}{\leq} 2N \sum_{x \geq 1} \sum_{t=1}^{A_x} e^{-c(\delta)(A_x - A_{x-1})} + N \sum_{t \geq 1} e^{-c(\delta)t}, \\
&\overset{(b)}{\leq} 2N \left( A_{x_0}^2 e^{-c(\delta)(A_{x_0} - A_{x_0 - 1})} + \sum_{x \geq 1} A_x e^{-c(\delta) A_{x-1}^{\omega}} \right) + N \int_{t \geq 0} e^{-c(\delta)t} dt, \\
&= 2N \left( A_{x_0}^2 e^{-c(\delta)(A_{x_0} - A_{x_0 - 1})} + \sum_{x \geq 1} A_x e^{-c(\delta) A_{x-1}^{\omega}} \right) + \frac{N}{c(\delta)}.
\end{aligned}$$

Step $(a)$ follows from changing the order of summation (which is licit as all terms are positive) and step $(b)$ follows from the assumption **A.2** in Section 3.1. Standard results from analysis gives that the series in the last display is finite as $c(\delta) > 0$ and $A_x \leq A_{2x} \leq A_{x-1}^3$, where the second inequality follows from Assumption **A.2** in Section 3.1. □

**Lemma 15.** *For all $\delta > 0$ such that $c(\delta) > \frac{5}{4}D$ and $(3 + 2\delta + \ln(4+2\delta)) \geq \frac{5}{4}D^{-1}$, where $c(\delta)$ is given in Proposition 5 and $D$ is in Assumption* **A.2** *in Section 3.1,*

$$\mathbb{E}[A_{4\Gamma}] \leq 2N \sum_{x \geq 1} A_{A_x} e^{-c(\delta)(A_x - A_{x-1})} + N \sum_{t \geq 1} e^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}.$$

*Proof.* Observe that $\mathbb{E}[A_{4\Gamma}] \leq \sum_{t \geq 1} \mathbb{P}\left[\Gamma \geq \frac{1}{4}A^{-1}(t)\right]$. We use similar ideas as in Lemma 14 to bound the tail probability. Recall that for any $t \geq 1$ and any $L \geq 1$, the following bound holds

$$\mathbb{P}[\Gamma \geq t] \leq \mathbb{P}[H^* \geq L] + \mathbb{P}\left[\bigcup_{i=1}^{N} \sum_{j=0}^{L} \mathcal{P}_j^{(i)} \leq t\right],$$

$$\leq \mathbb{P}[H^* \geq L] + N\mathbb{P}\left[\text{Poisson}\left(\left(1+\frac{\delta}{2}\right)A_L\right) \leq t\right].$$

In this proof, we shall use $L = A^{-1}\left(A^{-1}(t)\right)$. Thus,

$$\begin{aligned}
\mathbb{P}\left[\Gamma \geq \frac{1}{4}A^{-1}(t)\right] &\leq \mathbb{P}\left[H^* \geq A^{-1}\left(A^{-1}(t)\right)\right] + N\mathbb{P}\left[\text{Poisson}\left(\left(1+\frac{\delta}{2}\right)A_{A^{-1}(A^{-1}(t))}\right) \leq \frac{1}{4}A^{-1}(t)\right], \\
&= \mathbb{P}\left[H^* \geq A^{-1}\left(A^{-1}(t)\right)\right] + N\mathbb{P}\left[\text{Poisson}\left(\left(1+\frac{\delta}{2}\right)A^{-1}(t)\right) \leq \frac{1}{4}A^{-1}(t)\right], \\
&\leq 2N \sum_{x \geq A^{-1}\left(\frac{1}{M}A^{-1}(t)\right)} e^{-c(\delta)(A_x - A_{x-1})} + Ne^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}.
\end{aligned}$$

The last display follows from Proposition 5 and standard Poisson random variable Chernoff bound. Thus, we can bound $\mathbb{E}[A_{4\Gamma}]$ as

$$\begin{aligned}
\mathbb{E}[A_{4\Gamma}] &\leq \sum_{t\geq 1}\mathbb{P}\left[\Gamma \geq \frac{1}{4}A^{-1}(t)\right], \\
&\leq 2N\sum_{t\geq 1}\sum_{x\geq A^{-1}(A^{-1}(t))} e^{-c(\delta)(A_x - A_{x-1})} + N\sum_{t\geq 1} e^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}, \\
&\overset{(a)}{=} 2N\sum_{x\geq 1}\sum_{t=1}^{A(A_x)} e^{-c(\delta)(A_x - A_{x-1})} + N\sum_{t\geq 1} e^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}, \\
&= 2N\sum_{x\geq 1} A_{A_x} e^{-c(\delta)(A_x - A_{x-1})} + N\sum_{t\geq 1} e^{-(3+2\delta+\ln(4+2\delta))A^{-1}(t)}.
\end{aligned}$$

We will choose $\delta$ sufficiently large so that both the series are convergent. This is possible as the maps $\delta \longrightarrow c(\delta)$ and $\delta \longrightarrow (3+2\delta+\ln(4+2\delta))$ are non-decreasing and $\lim_{\delta\to\infty} c(\delta) = \lim_{\delta\to\infty}(3+2\delta+\ln(4+2\delta)) = \infty$. Observe that since $A_x \leq e^{Dx}$, for all large $x$, we have $A^{-1}(t) \geq \frac{1}{D}\ln(t)$. Thus, if $c(\delta) \geq \frac{5}{4}D$ and $(3+2\delta+\ln(4+2\delta)) > D^{-1}$, both series are convergent. □

**Lemma 16.** *For any $C \geq 2$,*

$$\mathbb{E}[A_{C\widehat{\tau}_{stab}}] \leq A_{\lceil \frac{C}{2} \rceil j^*} + \left(\frac{2}{2\alpha - 3}\right)\sum_{l\geq 3}\frac{A_{2l}}{A_{l-1}^3} + 2\sum_{x\geq \lceil \frac{j^*}{2}\rceil}(A_{\lceil Cx\rceil})^{(2(2\alpha-6)+2)}e^{-c(\delta)(A_x - A_{x-1})},$$

*where $c(\delta)$ is given in Proposition 5 and $j^*$ is given in Theorem 1.*

*Proof.* We start with the definition of expectation and repeatedly applying union bound yields,

$$\begin{aligned}
\mathbb{E}[A_{C\widehat{\tau}_{stab}}] &= \sum_{t\geq 1}\mathbb{P}[A_{C\widehat{\tau}_{stab}} \geq t], \\
&\leq \sum_{t\geq 1}\mathbb{P}\left[\widehat{\tau}_{stab} \geq \frac{1}{C}A^{-1}(t)\right], \\
&\leq A_{\lceil \frac{C}{2}\rceil j^*} + \sum_{t\geq A_{\lceil \frac{C}{2}\rceil j^*}+1}\mathbb{P}\left[\widehat{\tau}_{stab} \geq \frac{1}{C}A^{-1}(t)\right], \\
&\leq A_{\lceil \frac{C}{2}\rceil j^*} + \sum_{t\geq A_{\lceil \frac{C}{2}\rceil j^*}+1}\mathbb{P}\left[\bigcup_{i=1}^{N}\bigcup_{l\geq \frac{1}{C}A^{-1}(t)}\chi_l^{(i)} = 0\right], \\
&\leq A_{\lceil \frac{C}{2}\rceil j^*} + \sum_{t\geq A_{\lceil \frac{C}{2}\rceil j^*}+1} N\mathbb{P}\left[\bigcup_{l\geq \frac{1}{C}A^{-1}(t)}\chi_l^{(i)} = 0\right],
\end{aligned}$$

$$
\begin{aligned}
&\leq A_{\lceil \frac{C}{2} \rceil j^*} + \sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} N \sum_{l \geq \frac{1}{C} A^{-1}(t)} \left( \mathbb{P}[\chi_l^{(i)} = 0, l \geq H^*] + \mathbb{P}[\chi_l^{(i)} = 0, l < H^*] \right), \\
&\leq A_{\lceil \frac{C}{2} \rceil j^*} + \sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} N \sum_{l \geq \frac{1}{C} A^{-1}(t)} \left( \mathbb{P}[\chi_l^{(i)} = 0, l \geq H^*] + \mathbb{P}[l < H^*] \right), \\
&= A_{\lceil \frac{C}{2} \rceil j^*} + \sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} N \sum_{l \geq \frac{1}{C} A^{-1}(t)} \mathbb{P}[\chi_l^{(i)} = 0, l \geq H^*] + \sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} \sum_{l \geq \frac{1}{C} A^{-1}(t)} N\mathbb{P}[l < H^*], \\
&\overset{(a)}{\leq} A_{\lceil \frac{C}{2} \rceil j^*} + \left( \frac{2}{2\alpha - 3} \right) \sum_{l \geq 3} \frac{A_{2l}}{A_{l-1}^3} + \sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} \sum_{l \geq \frac{1}{C} A^{-1}(t)} N\mathbb{P}[l < H^*].
\end{aligned}
$$

Step $(a)$ follows as $C \geq 2$, and hence, the first summation follows from identical calculations as carried out in Proposition 3. This is so as the bound in Lemma 6 and in Lemma 7 are identical. Thus, the first series is upper bounded by $\left( \frac{2}{2\alpha-3} \right) \sum_{l \geq 3} \frac{A_{2l}}{A_{l-1}^3}$. We shall now estimate the second series.

$$
\begin{aligned}
\sum_{t \geq A_{\lceil \frac{C}{2} \rceil j^*}+1} \sum_{l \geq \frac{1}{C} A^{-1}(t)} N\mathbb{P}[l < H^*] &\leq \sum_{l \geq \lceil \frac{j^*}{2} \rceil} \sum_{t = A_{\lceil \frac{C}{2} \rceil j^*}+1}^{A_{\lceil Cl \rceil}} N\mathbb{P}[l < H^*], \\
&\leq \sum_{l \geq \lceil \frac{j^*}{2} \rceil} N A_{\lceil Cl \rceil} \mathbb{P}[l < H^*], \\
&\overset{(b)}{\leq} \sum_{l \geq \lceil \frac{j^*}{2} \rceil} N^2 A_{\lceil Cl \rceil} \sum_{x \geq l} 2e^{-c(\delta)(A_x - A_{x-1})}, \\
&= N^2 \sum_{x \geq \lceil \frac{j^*}{2} \rceil} \sum_{l = \lceil \frac{j^*}{2} \rceil}^{x} A_{\lceil Cl \rceil} 2e^{-c(\delta)(A_x - A_{x-1})}, \\
&\leq N^2 \sum_{x \geq \lceil \frac{j^*}{2} \rceil} x A_{\lceil Cx \rceil} e^{-c(\delta) A_{x-1}^{\omega}}, \\
&\overset{(c)}{\leq} \sum_{x \geq \lceil \frac{j^*}{2} \rceil} (A_{\lceil Cx \rceil})^{2(2\alpha-6)+2} 2e^{-c(\delta)(A_x - A_{x-1})}, \\
&\overset{(d)}{<} \infty.
\end{aligned}
$$

Step $(b)$ follows from Proposition 5 and in step $(c)$, we use the fact that for all $x \geq \frac{j^*}{2}$, we have $N \leq A_x^{2\alpha-6}$. Step $(d)$ follows from Assumption **A.2** that for all sufficiently large $l$, $A_{2l} \leq A_l^3$, which on iterating yields that for all large $x$ and any $C \geq 2$, we have $A_{Cx} \leq A_x^{3^{\lceil \log_2(C) \rceil}}$. Thus, we have the following chain of inequalities.

$$
\begin{aligned}
\sum_{x \geq \lceil \frac{j^*}{2} \rceil} (A_{\lceil Cx \rceil})^{2(2\alpha-6)+2} 2e^{-c(\delta)(A_x - A_{x-1})} &\leq \sum_{x \geq \lceil \frac{j^*}{2} \rceil} (A_x)^{3^{\lceil \log_2(C) \rceil}(2(2\alpha-6)+2)} 2e^{-c(\delta)(A_x - A_{x-1})}, \\
&\overset{(e)}{\leq} D_1 \sum_{x \geq 2} A_x^{D_2} e^{-c A_{x-1}^{\omega}} < \infty.
\end{aligned}
$$

for some $D_1, D_2, \omega^{'} > 0$. Step $(e)$ follows as we can replace the tail terms of the series with $A_x - A_{x-1} \geq A_{x-1}^{\omega}$ from Assumption **A.2**. The finiteness of the series in $(e)$ is a standard fact from real analysis and can be proven for instance through a Taylor series approximation of the exponential function. □

### G.4 Proof of Theorem 8

The proof of Theorem 2 is concluded by using estimates in Lemmas 14 and 16 into Lemmas 11 and 10.

## Appendix H Proof of Theorem 2

The proof follows identical steps as that of Theorem 8, with the exception that $H^* = \Gamma = 0$ almost-surely. More precisely, substituting these two facts in Lemma 11 and re-using all the remaining structural and quantitative results from the proof of Theorem 8 will yield the desired result.

## Appendix I Auxillary Results

**Proposition 6.** *For each $\delta > 0$, $y \geq 0$ and convex sequence $(A_j)_{j\geq 0}$, we have*

$$\sup\{j + x \geq 0 : \exists j \leq y+1, (A_{j+x} - A_j) \leq (1+\delta)(A_{y+1} - A_y)\} \leq y + \lfloor 2+\delta \rfloor$$

*Proof.* Let $x \geq 0$ and $j \leq y+1$ be such that

$$(A_{j+x} + (1+\delta)A_y) \leq (1+\delta)A_{y+1} + A_j. \tag{20}$$

Now since $j \leq y+1$, and $(A_l)_{l\geq 0}$ is non-decreasing, the above inequality implies

$$\frac{A_{j+x} + (1+\delta)A_y}{2+\delta} \leq A_{y+1}.$$

Now, let $j + x = y + \lfloor 2+\delta \rfloor + k$, for some $k \geq 0$. From convexity of $(A_j)_{j\geq 1}$, we have

$$A_{y+\frac{\lfloor 2+\delta \rfloor + k}{2+\delta}} \leq \frac{A_{j+x} + (1+\delta)A_y}{2+\delta} \leq A_{y+1}$$

But for all $k \geq 1$, we have $A_{y+\frac{\lfloor 2+\delta \rfloor + k}{2+\delta}} > A_{y+1}$ and hence $k = 0$ is the only possibility such that Equation (20) holds. □

## Appendix J Proof of Theorem 3

*Proof.* In order to prove the bound, we shall consider a system of *full interaction among agents*, where there are no constraints on communications. In this system, each agent after pulling an arm and observing a reward, communicates this information (the arm pulled and reward observed) to central *board*. Thus, at the beginning of each time-step, every agent has access to the entire system history (arms pulled and rewards obtained) up-to the previous time step, by which to base the current time step's action (arm pull) on. As all agents have access to the same history at the beginning of a time step, the optimal strategy to minimize per agent regret is one where in each time step, all agents play the same arm. Hence, this system is equivalent to a single *leader* playing arms, such that on playing any arm at any time, the leader observes $N$ i.i.d. reward samples from the chosen arm, each corresponding to the obtained reward by the agents. From henceforth, we mean by the full interaction setting, as one wherein a single leader agent pull an arm at each time step, and observes $N$ i.i.d. reward samples from the chosen arm.

By construction, a lower bound for regret incurred by the leader agent in the full interaction setting forms a lower bound on the per-agent regret in our model with communication constraints. This is so, since the leader agent in full interaction setting can 'simulate' any feasible policy of any agent $i \in [N]$ with communication constraints among agents. Notice that each time the leader agent in the full-interaction setting plays an arm, it receives $N$ i.i.d. samples of rewards, corresponding to the reward on that arm obtained by the $N$ agents. We will consider an alternate system where a *fictitious leader agent* plays for $NT$ time steps, where at each time, the fictitious agent is playing arms, as a measurable function of its observed history. From standard results, (for eg. (Lai and Robbins, 1985)), the total regret of the fictitious agent, after $NT$ arm-pulls satisfies

$$\liminf_{T\to\infty} \frac{\mathbb{E}[R_{NT}^{(\text{fictitious})}]}{\ln(NT)} \geq \left( \sum_{j=1}^{K-1} \frac{\Delta_j}{\mathrm{KL}(\mu_j, \mu_1)} \right), \tag{21}$$

Now, we shall argue that the preceding display implies the desired lower bound on per-agent regret in the full interaction setting. Fix some $a \in \{0, \cdots, N-1\}$. Denote, by the regret incurred by the fictitious agent at time steps $a, N+a, \cdots, N(T-1)+a$ as $\mathcal{R}_a^{(f)}$. Clearly $\sum_{a=1}^{N} \mathcal{R}_a^{(f)} = \mathbb{E}[R_{NT}^{(\text{fictitious})}]$.

Denote by $\mathbf{\Pi}_{agent}$ to be the set of consistent policies for the agents in the full-interaction setting and by $\mathbf{\Pi}_{fictitious}$ as the set of all consistent policies for the fictitious agent. Denote by the set of policies $\widetilde{\mathbf{\Pi}}_{fictitious} \subset \mathbf{\Pi}_{fictitious}$, as those policies for the fictitious agents, where for any policy $\pi \in \widetilde{\mathbf{\Pi}}_{fictitious}$, the arms played at time instants $N, 2N, \cdots, NT$, belong to $\mathbf{\Pi}_{agent}$. Furthermore, for all $a \in \{1, \cdots, T\}$, and all $b \in \{1, \cdots, N-1\}$, and all $\pi \in \widetilde{\mathbf{\Pi}}_{fictitious}$, the arm chosen by $\pi$ at time instant $aN$ is the same as the arm chosen at time-instant $aN+b$. In other words, the the set of policies $\widetilde{\mathbf{\Pi}}_{fictitious}$ are the ones that any agent under the full interaction setting of our model can play. This definitions now give us for any $a \in \{0, \cdots, N-1\}$

$$\begin{aligned}
\inf_{\pi \in \mathbf{\Pi}_{agent}} \mathbb{E}[R_T^{(i)}] &= \inf_{\pi \in \widetilde{\mathbf{\Pi}}_{fictitious}} \mathcal{R}_a^{(f)}, \\
&= \inf_{\pi \in \widetilde{\mathbf{\Pi}}_{fictitious}} \frac{1}{N} \sum_{a=1}^{N} \mathcal{R}_a^{(f)}, \\
&= \inf_{\pi \in \widetilde{\mathbf{\Pi}}_{fictitious}} \frac{1}{N} \mathbb{E}[R_{NT}^{(\text{fictitious})}], \\
&\geq \inf_{\pi \in \mathbf{\Pi}_{fictitious}} \frac{1}{N} \mathbb{E}[R_{NT}^{(\text{fictitious})}].
\end{aligned}$$

The first equality follows as under any policy in $\widetilde{\mathbf{\Pi}}_{fictitious}$, the arms played by the fictitious agent only chooses potentially new arms to play at instants $N, 2N, \cdots$. Now, using Equation (21), we get from the previous display, that for any policy $\pi \in \mathbf{\Pi}_{agent}$,

$$\liminf_{T \to \infty} \frac{\mathbb{E}[R_T^{(i)}]}{\ln(NT)} \geq \left( \frac{1}{N} \sum_{j \geq 1} \frac{\Delta_j}{\text{KL}(\mu_j, \mu_1)} \right).$$

□

# Appendix K  Proof of Corollary 5

In order to prove the corollary, we first establish that $A_x \leq 2x^\beta$, for all small $\varepsilon$ in Equation (1). Notice from Equation (1) that for all $x \in \mathbb{N}$, we have

$$\begin{aligned}
A_x &= \max\left(\min\{t \in \mathbb{N} : B_t \geq x\}, \lceil (1+x)^{1+\varepsilon} \rceil\right), \\
&= \max\left(\min\{t \in \mathbb{N} : B_t \geq x\}, \lceil (1+x)^{1+\varepsilon} \rceil\right), \\
&\leq \max\left(x^\beta, (1+x)^{1+\varepsilon}\right), \\
&\leq \max(2x^\beta, 2x^{1+\varepsilon}), \\
&= 2x^\beta,
\end{aligned}$$

where the last equality follows since $\varepsilon < \beta - 1$. Furthermore, for all $x \geq x_0$ where $\varepsilon < \beta \frac{\ln(x_0)}{\ln(x_0+1)} - 1$, we have $A_x = x^\beta$. Such a $x_0$ exists since $\beta - 1 > 0$. Moreover, from definition of $A_x$, we have $A_x \geq x^\beta$, for all $x$.

Recall that $g((A_x)_{x \in \mathbb{N}}) = A_{j^*} + \frac{2}{2\alpha - 3} \sum_{l \geq \frac{j^*}{2} - 1} \frac{A_{2l+1}}{A_{l-1}^3}$. We first bound the series term in as follows

$$\sum_{l \geq \frac{j^*}{2} - 1} \frac{A_{2l+1}}{A_{l-1}^3} \overset{(a)}{\leq} \sum_{l \geq 2} 2 \frac{(2l+1)^\beta}{(l-1)^{3\beta}}, \tag{22}$$

$$\begin{aligned}
&\leq 2\sum_{l\geq 2} 3^{\beta}\frac{1}{(l-1)^{2\beta}},\\
&\leq 2\frac{\pi^2}{6}3^{\beta}. \qquad (23)
\end{aligned}$$

We now bound $j^*$ in this case. Recall that

$$\begin{aligned}
j^* &= 2\max\left(A^{-1}\left(\left(N\binom{K}{2}\left(\left\lceil\frac{K}{N}\right\rceil+1\right)\right)^{\frac{1}{(2\alpha-6)}}\right)+1, \min\left\{j\in\mathbb{N}: \frac{A_j-A_{j-1}}{2+\lceil\frac{K}{N}\rceil}\geq 1+\frac{4\alpha\log(A_j)}{\Delta_2^2}\right\}\right),\\
&\leq 2\max\left(K^{\frac{3}{\beta(2\alpha-6)}}+1, \min\left\{j\in\mathbb{N}: \frac{j^\beta-(2(j-1))^\beta}{2+\lceil\frac{K}{N}\rceil}\geq 1+\frac{4\alpha\log(j^\beta)}{\Delta_2^2}\right\}\right),\\
&\leq 2\max\left(K^{\frac{3}{\beta(2\alpha-6)}}, \min\left\{j\in\mathbb{N}: \frac{j^\beta-(2(j-1))^\beta}{2+\lceil\frac{K}{N}\rceil}\geq \frac{8\alpha\log(j^\beta)}{\Delta_2^2}\right\}\right),\\
&\leq 2\max\left(K^{\frac{3}{\beta(2\alpha-6)}}, \left(16\alpha\frac{2+\lceil\frac{K}{N}\rceil}{\Delta_2^2}\right)^{\frac{1}{\beta-1}}\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
A_{j^*} &\leq 2(j^*)^\beta,\\
&\leq 4\max\left(K^{\frac{3}{(2\alpha-6)}}, \left(16\alpha\frac{2+\lceil\frac{K}{N}\rceil}{\Delta_2^2}\right)^{\frac{\beta}{\beta-1}}\right). \qquad (24)
\end{aligned}$$

. Thus from Equations (23) and (24), we get that

$$g((A_x)_{x\in\mathbb{N}}) \leq \frac{4}{2\alpha-3}\frac{\pi^2}{6}3^\beta + 4\max\left(K^{\frac{3}{(2\alpha-6)}}, \left(16\alpha\frac{2+\lceil\frac{K}{N}\rceil}{\Delta_2^2}\right)^{\frac{\beta}{\beta-1}}\right).$$

The proof is completed thanks to the formula in Corollary 17.

## Appendix L Impact of Gossip Matrix $P$

**Corollary 17.** *Suppose $N \geq 2$ agents are connected by a d-regular graph with adjacency matrix $\boldsymbol{A}_G$ having conductance $\phi$ and the gossip matrix $P = d^{-1}\boldsymbol{A}_G$. If the agents are using Algorithm 1 with parameters satisfying assumptions in Theorem 1, then for any $i \in [N]$ and $T \in \mathbb{N}$*

$$\mathbb{E}[R_T^{(i)}] \leq \underbrace{4\alpha\ln(T)\left(\sum_{j=2}^{\lceil\frac{K}{N}\rceil+2}\frac{1}{\Delta_j}\right)+\frac{K}{4}}_{\textit{Collaborative UCB Regret}} + \underbrace{A_{2C\frac{\log(N)}{\phi}} + g\left((A_x)_{x\in\mathbb{N}}\right) + A_{j^*} + 1}_{\textit{Cost of Pairwise Communications}},$$

*where $g(\cdot)$ is from Theorem 1, and $C > 0$ is an universal constant stated in Lemma 19 in the Appendix. Similarly, if all agents run Algorithm 3 with assumptions as in Theorem 2, then*

$$\mathbb{E}[R_T^{(i)}] \leq \underbrace{4\alpha\ln(T)\left(\sum_{j=2}^{\lceil\frac{K}{N}\rceil+2}\frac{1}{\Delta_j}\right)+\frac{K}{4}}_{\textit{Collaborative UCB Regret}} + \underbrace{(1+\delta)A_{2\lfloor 2+\delta\rfloor C\frac{\log(N)}{\phi}} + \widehat{g}\left((A_x)_{x\in\mathbb{N}},\delta\right) + 1}_{\textit{Cost of Pairwise Communications}},$$

*where $\widehat{g}(\cdot)$ is given in Theorem 2.*

Before providing the proof, we provide a special case of above to be able to derive some intuition.

**Corollary 18.** *Suppose $N \geq 2$ agents are connected by a $d$-regular graph with adjacency matrix $\boldsymbol{A}_G$ having conductance $\phi$ and the gossip matrix $P = d^{-1}\boldsymbol{A}_G$. Suppose the communication budget scales as $B_t = \lfloor t^{1/\beta} \rfloor$, for all $t \geq 1$, where $\beta > 1$ is arbitrary. If the agents are using Algorithm 1 with parameters satisfying assumptions in Theorem 1, then for any $i \in [N]$ and $T \in \mathbb{N}$*

$$\mathbb{E}[R_T^{(i)}] \leq \underbrace{4\alpha \ln(T) \left( \sum_{j=2}^{\lceil \frac{K}{N} \rceil + 2} \frac{1}{\Delta_j} \right) + \frac{K}{4}}_{\textit{Collaborative UCB Regret}} + \underbrace{\left( 2C \frac{\log(N)}{\phi} \right)^{\beta}}_{\textit{Impact of Gossip Matrix}} + \underbrace{2 \frac{3^{\beta}}{2\alpha - 3} \frac{\pi^2}{6} + (j^*)^{\beta} + 1}_{\textit{Constant Independent of } P},$$

*where $j^*$ is a constant independent of the gossip matrix $P$, depending only on $N, K$ and $\Delta_2$ (given in Theorem 1).*

We now prove Corollary 17.

*Proof.* The proof follows if we establish that $\mathbb{E}[A_{2\tau_{spr}^{(P)}}] \leq A_{\frac{2C\log(N)}{\phi}} + 1$ and $\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}] \leq A_{\frac{2\lfloor 2+\delta \rfloor C \log(N)}{\phi}} + 1$. We can bound them using the main result from (Chierichetti et al., 2010), restated as Lemma 19 in the sequel. That lemma in particular gives that, one can compute $\mathbb{E}[A_{2\tau_{spr}^{(P)}}]$ as follows.

$$\begin{aligned}
\mathbb{E}[A_{2\tau_{spr}^{(P)}}] &\leq A_{\frac{2C\log(N)}{\phi}} + \sum_{t \geq A_{\frac{2C\log(N)}{\phi}}} \mathbb{P}[A_{2\tau_{spr}^{(P)}} \geq t], \\
&\leq A_{\frac{2C\log(N)}{\phi}} + \sum_{l \geq 1} \mathbb{P}\left[A_{2\tau_{spr}^{(P)}} \geq A_{\frac{2Cl\log(N)}{\phi}}\right] A_{\frac{2Cl\log(N)}{\phi}}, \\
&\leq A_{\frac{2C\log(N)}{\phi}} + \sum_{l \geq 1} \mathbb{P}\left[2\tau_{spr}^{(P)} \geq \frac{2Cl\log(N)}{\phi}\right] A_{\frac{2Cl\log(N)}{\phi}}, \\
&\overset{(a)}{\leq} A_{\frac{2C\log(N)}{\phi}} + \sum_{l \geq 1} e^{-4l\log(N)} A_{\frac{2Cl\log(N)}{\phi}}, \\
&\overset{(b)}{\leq} A_{\frac{2C\log(N)}{\phi}} + \sum_{l \geq 1} e^{-2l\log(N)}, \\
&\overset{(c)}{\leq} A_{\frac{2C\log(N)}{\phi}} + 1.
\end{aligned}$$

In step $(a)$, we use the estimate from Lemma 19. In step $(b)$, we use the additional assumption in the corollary that $A_l \leq e^{Dl}$, for all $D > 0$. Thus, we can choose $D \leq \frac{\phi}{C}$ to arrive at the conclusion in step $(b)$. In step $(c)$, we use $N \geq 2$ to bound the geometric series. Similar computation will yield the bound on $\mathbb{E}[A_{2\lfloor 2+\delta \rfloor \tau_{spr}^{(P)}}]$. □

**Lemma 19.** *There exists an universal constant $C > 0$, such that for every $d \geq 2$ regular graph on $N$ vertices with conductance $\phi$, the spreading time of the standard PULL process completes in time $\tau_{spr}^{(P)}$ which satisfies for all $l \in \mathbb{N}$,*

$$\mathbb{P}\left[\tau_{spr}^{(P)} \geq Cl\frac{\log(N)}{\phi}\right] \leq N^{-4l}.$$

*Proof.* The main result (Lemma 6) of (Chierichetti et al., 2010) gives that there exists a constant $C > 0$, such that for all $d$-regular graphs with conductance $\phi$, the spreading time satisfies

$$\mathbb{P}\left[\tau_{spr}^{(P)} \geq C\frac{\log(N)}{\phi}\right] \leq N^{-4}.$$

Now, given any $l \in \mathbb{N}$, we can now divide the time into intervals $\left[0, C\frac{\log(N)}{\phi}\right], \left[C\frac{\log(N)}{\phi}, 2C\frac{\log(N)}{\phi}\right], \cdots$ $, \left[C(l-1)\frac{\log(N)}{\phi}, Cl\frac{\log(N)}{\phi}\right]$. For the event $\left\{\tau_{spr}^{(P)} \geq Cl\frac{\log(N)}{\phi}\right\}$ to occur, we need the spreading to be not finished in each of the $l$ intervals. However, at the beginning of each interval, we know that at-least one node is informed of the rumor. Thus, the probability, that the rumor spreading does not complete in a single interval is at-most $N^{-4}$, which follows from monotonicity, where we can bound by saying that exactly one worst-case node is aware of the rumor. As the sequence of callers is independent across intervals, the probability that rumor spreading fails in all $l$ intervals is then at-most $N^{-4l}$. □

## Appendix M Regret Communication Tradeoff

**Corollary 20.** *Suppose, Algorithm 1 is run with $K$ arms and $N$ agents connected by a gossip matrix $P$, with two different communication schedules $(A_x^{(1)})_{x\in\mathbb{N}}$ and $(A_x^{(2)})_{x\in\mathbb{N}}$, such that $\lim_{x\to\infty}\frac{A_x^{(1)}}{A_x^{(2)}} = 0$. Then there exist positive constants $N_0, K_0 \in \mathbb{N}$ (depending on the two communication sequences), such that for all $N \geq N_0$ and $K \geq K_0$, and $P$, the cost of communications in the regret bound in Equation (4) is ordered as*

$$g((A_x^{(1)})) + \mathbb{E}[A^{(1)}_{2\tau_{spr}^{(P)}}] \geq g((A_x^{(2)})) + \mathbb{E}[A^{(2)}_{2\tau_{spr}^{(P)}}].$$

*Proof.* Consider a fixed $(A_x^{(1)})_{x\in\mathbb{N}}$ and $(A_x^{(2)})_{x\in\mathbb{N}}$, such that $\lim_{x\to\infty}\frac{A_x^{(1)}}{A_x^{(2)}} = 0$. The ordering on $E[(A^{(P)}_{2\tau_{spr}})^{(1)}] \leq E[(A^{(P)}_{2\tau_{spr}})^{(2)}]$ follows trivially as $P$ is fixed for the two cases. It suffices to show that there exist positive constants $N_0$ and $K_0$ (depending on $(A_x^{(1)})_{x\in\mathbb{N}}$ and $(A_x^{(2)})_{x\in\mathbb{N}})$), such that for all $N \geq N_0$ and $K \geq K_0$, $g(A_x^{(1)}) \leq g(A_x^{(2)})$. If $N$ or $K$ is sufficiently large, then $(j^*)^{(i)} = 2(A^{-1})^{(i)}\left(\left(N\binom{K}{2}\left(\left\lceil\frac{K}{N}\right\rceil+1\right)\right)^{\frac{1}{(2\alpha-6)}}\right)$, for $i \in \{1,2\}$. Notice that

$$\begin{aligned} g((A_x^{(2)})) - g((A_x^{(1)})) &= A^{(2)}_{(j^*)^{(2)}} - A^{(1)}_{(j^*)^{(1)}} + \left(\frac{2}{2\alpha-3}\left(\sum_{l\geq\frac{(j^*)^{(2)}}{2}-1}\frac{A^{(2)}_{2l+1}}{(A^{(2)}_{l-1})^3}\right) - \frac{2}{2\alpha-3}\left(\sum_{l\geq\frac{(j^*)^{(1)}}{2}-1}\frac{A^{(1)}_{2l+1}}{(A^{(1)}_{l-1})^3}\right)\right), \\ &\geq A^{(2)}_{(j^*)^{(2)}} - A^{(1)}_{(j^*)^{(1)}} - \frac{2}{2\alpha-3}\left(\sum_{l\geq 1}\frac{A^{(1)}_{2l+1}}{(A^{(1)}_{l-1})^3}\right). \end{aligned} \tag{25}$$

Notice that $A^{(2)}_{(j^*)^{(2)}} - A^{(1)}_{(j^*)^{(1)}} > 0$ and scaling (is monotone non-decreasing) with $N$ and $K$. In other words, for fixed $K$, $\lim_{N\to\infty}(A^{(2)}_{(j^*)^{(2)}} - A^{(1)}_{(j^*)^{(1)}}) = \infty$ and for fixed $N$, $\lim_{K\to\infty}(A^{(2)}_{(j^*)^{(2)}} - A^{(1)}_{(j^*)^{(1)}}) = \infty$. This follows as $(A_x^{(i)})_{x\geq 1}$ is super-linear for $i \in \{1,2\}$ and $\lim_{x\to\infty}\frac{A_x^{(1)}}{A_x^{(2)}} = 0$. From the hypothesis that the two communication sequences satisfy assumption **A.2**, we have that $\frac{2}{2\alpha-3}\left(\sum_{l\geq 1}\frac{A^{(1)}_{2l+1}}{(A^{(1)}_{l-1})^3}\right) < \infty$ and independent of $N$ and $K$. Thus, for all large $N$ or $K$, Equation (25), simplifies to $g(A_x^{(2)}) - g(A_x^{(1)}) > 0$. □

## Appendix N An Algorithm without using agent ids

The initialization in Line 2 of Algorithms 1 and 3 relied on each agent knowing its identity. However, in many settings, it may be desirable to have algorithms that do not depend on the agent's identity. We outline here a randomized initialization procedure in Line 2 to convert Algorithms 1 and 3 to one without using agent ids. Fix some $\gamma \in (0,1)$. We replace Line 2 in Algorithms 1 and 3 with a randomization, where each agent $i \in [N]$ chooses independently of other agents, a uniformly random subset of size $\left\lceil \ln\left(\frac{1}{\gamma}\right)\frac{K}{N}\right\rceil + 2$ from the set of $K$ arms as $S_0^{(i)}$. Each agent $i$, then subsequently chooses a random subset of size $\left\lceil \ln\left(\frac{1}{\gamma}\right)\frac{K}{N}\right\rceil$ uniformly at random from $S_0^{(i)}$ as its 'sticky set' $\widehat{S}^{(i)}$. The rest of the algorithms from Line 3 will be identical. One can then immediately see that the regret guarantees stated in Theorems 1 and 2 hold verbatim for this modification, with probability at-least $1-\gamma$, where the probability is over the initial random assignment of the sets $\widehat{S}^{(i)}$ to agents. More precisely, with probability at-least $1-\gamma$, the above random initialization ensures that there exists an agent $i \in [N]$, such that the best arm $1 \in \widehat{S}^{(i)}$. On this event, the regret guarantees along with the same proof of Theorems 1 and 2 hold.