Flooding with Absorption: An Efficient Protocol for Heterogeneous Bandits over Complex Networks

Junghyun Lee, Laura Schmid, Se-Young Yun

Kim Jaechul Graduate School of AI, KAIST

Primer on Multi-Armed Bandits (MABs)

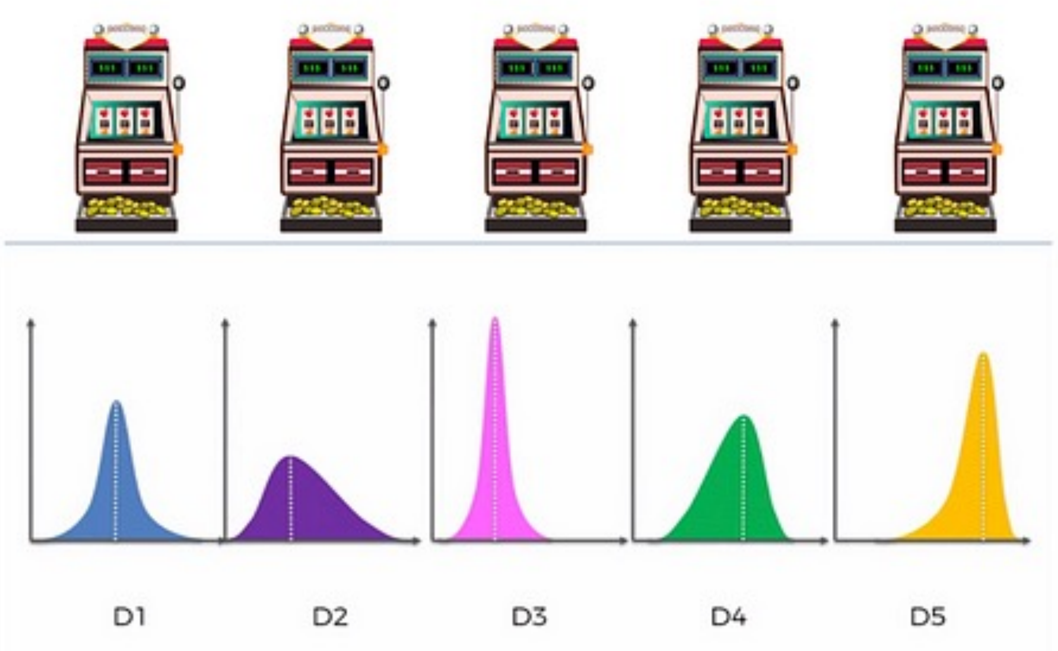
Problem Setting, Applications, UCB Algorithm

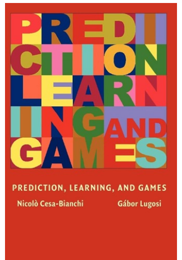
Multi-Armed Bandits (MAB)

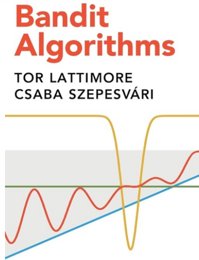
- **<u>Scenario.</u>** There is a collection of slot machines (***arm-set*** $\mathcal{K}$), each with an unknown ***reward*** distribution.
  1. Learner pulls a machine $a_t \in \mathcal{K}$
  2. Observe a reward $r_t \sim \mathcal{D}_{a_t}$

- **<u>Goal.</u>** What is the *optimal strategy* of pulling the arms that optimizes our cumulative reward?

Exploration vs. Exploitation!







Multi-Armed Bandits (MAB)

- For each $a \in \mathcal{K}$, let $\{r_{a,t}\}_{t\in[T]}$ be the random variables for the reward obtained as if one pulls the arm a at time t
  - $\mu_a = \mathbb{E}[r_{a,t}]$: average reward by pulling arm a
- **Regret:**

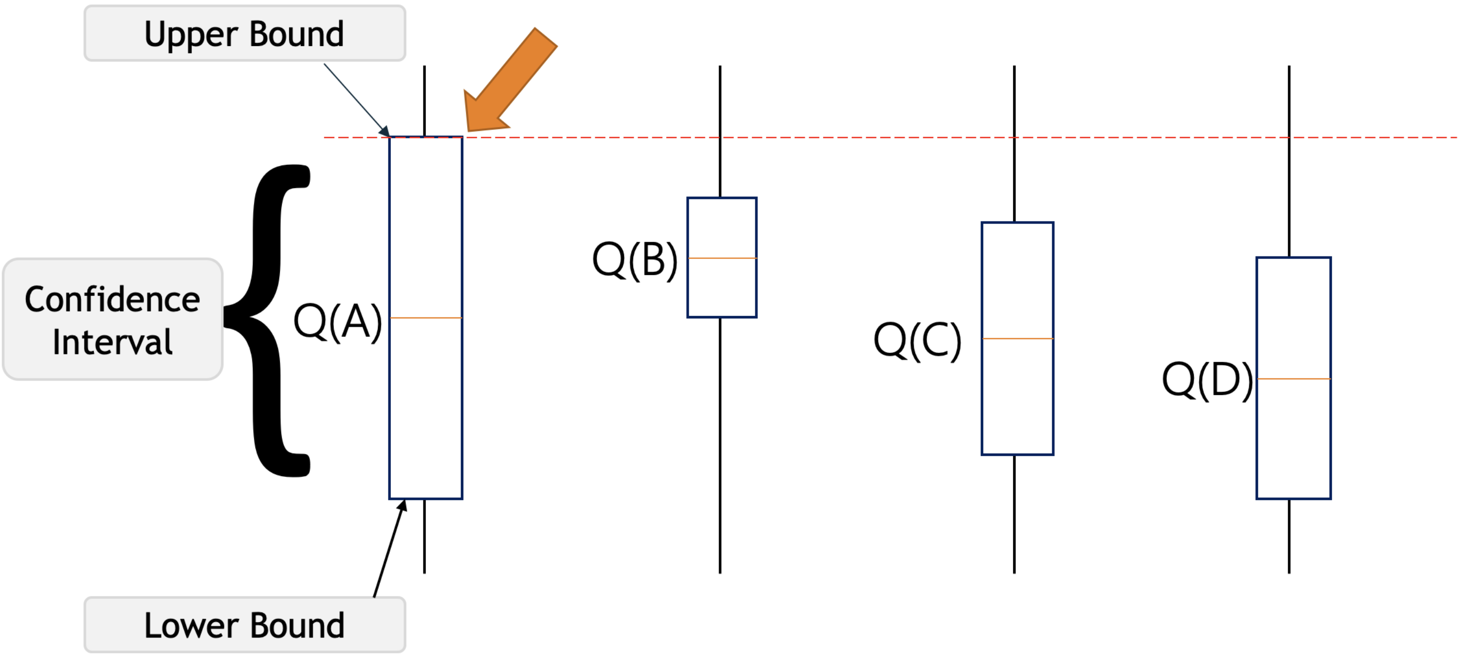
$$\mathrm{Reg}(T) \coloneqq \max_{a\in\mathcal{K}} \mathbb{E}\left[\sum_{t=1}^{T}\left(r_{a,t} - r_{a_t,t}\right)\right]$$
$$= \sum_{a\in\mathcal{K}\setminus\{a_\star\}} \Delta_a N_a(T)$$

  - $a_\star$: best arm, i.e., $\mu_\star > \mu_a$ for any ***suboptimal arm*** $a \in \mathcal{K}\setminus\{a_\star\}$
  - $\Delta_a \coloneqq \mu_\star - \mu_a$: suboptimality gap
  - $N_a(T) \coloneqq \sum_{t=1}^{T} 1[a_t = a]$

Upper-Confidence Bound (UCB)

- **UCB Algorithm** [Auer et al., 2002]:

$$a_{t+1} = \underset{a \in \mathcal{K}}{\operatorname{argmax}} \frac{1}{t} \sum_{\tau=1}^{t} r_\tau 1[a_\tau = a] + \sqrt{\frac{2\alpha \log t}{N_a(t)}}$$



Upper-Confidence Bound (UCB)

- **Regret of UCB:**

$$\mathrm{Reg}(T) \leq C \sum_{a \in \mathcal{K} \setminus \{a_\star\}} \frac{\log T}{\Delta_a}$$

- Note how the regret scales ***logarithmically*** in T
- Small Δ_a means that the regret is larger, i.e., Δ_a quantifies the ***difficulty*** of the given bandit instance!

- This is **optimal**, i.e., a matching lower bound exists [Lai & Robbins, 1952]

Multi-Agent MABs

Motivation, Prior Works, **Our Setting**

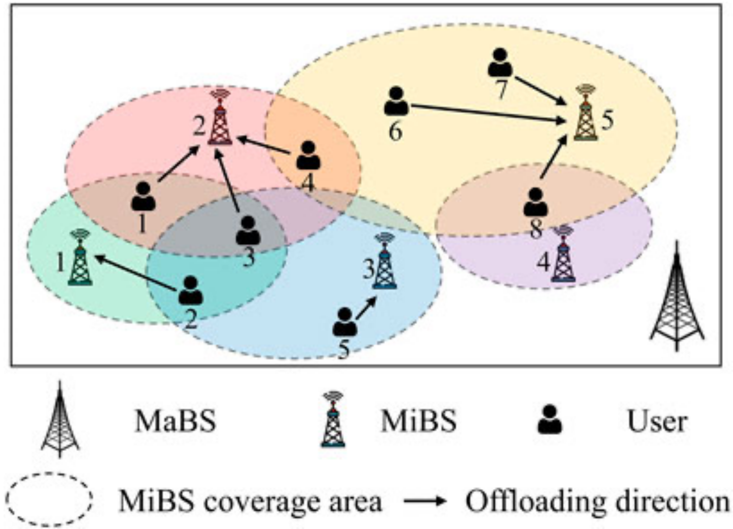
Collaborative, Multi-Agent MABs

- Oftentimes, we must consider multiple agents, each with a bandit instance, to cooperate with one another!
  - ex. online advertisement, wireless channel allocation,

- **Collaboration:** Sharing information (e.g., reward, pulled arm index) to facilitate learning of myself *and* others!

→in UCB, each agent has additional *side information* from its hop neighbors that facilitates its own exploration!!

e.g., $N_a^v(t) \rightarrow M_a^v(t) = \sum_{w \in \mathcal{N}_{\mathcal{G}}(v)} N_a^w(t)$



Collaborative, Multi-Agent MABs

- Let $\mathcal{V}$ be the set of agents, and $\mathrm{Reg}^v(T)$ be the agent v's regret

- **<u>Group regret</u>:**

$$Reg(T) := \sum_{v \in \mathcal{V}} \mathrm{Reg}^v(T)$$

- Without collaboration,

$$Reg(T) \le NC \sum_{a \in \mathcal{K} \setminus \{a_\star\}} \frac{\log T}{\Delta_a}$$

  - Linear in the #agents N!

We want to reduce the dependency on N via collaboration!

Prior Works on Multi-Agent MAB

Homogeneous, Networked

- Every agent shares the *same* bandit instance
  & Agents are on a *network*
  [Kolla et al., 2018; Madhushani et al., 2020]

Heterogeneous, Fully-Connected

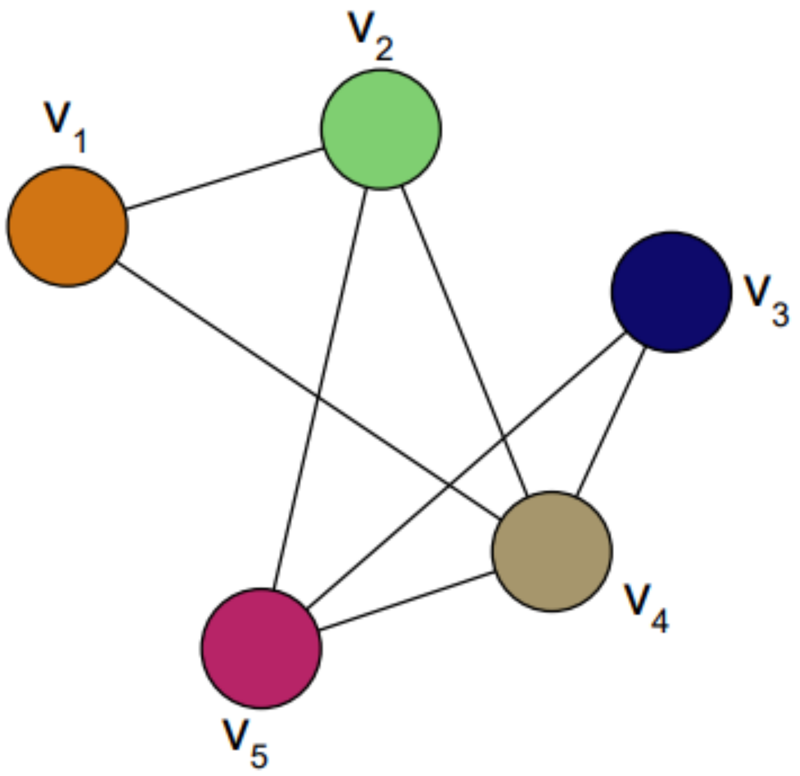
- Every agent has its *own* bandit instance
  & Agents are *fully-connected*
  [Yang et al., 2022]

WHAT IF the two settings are combined?

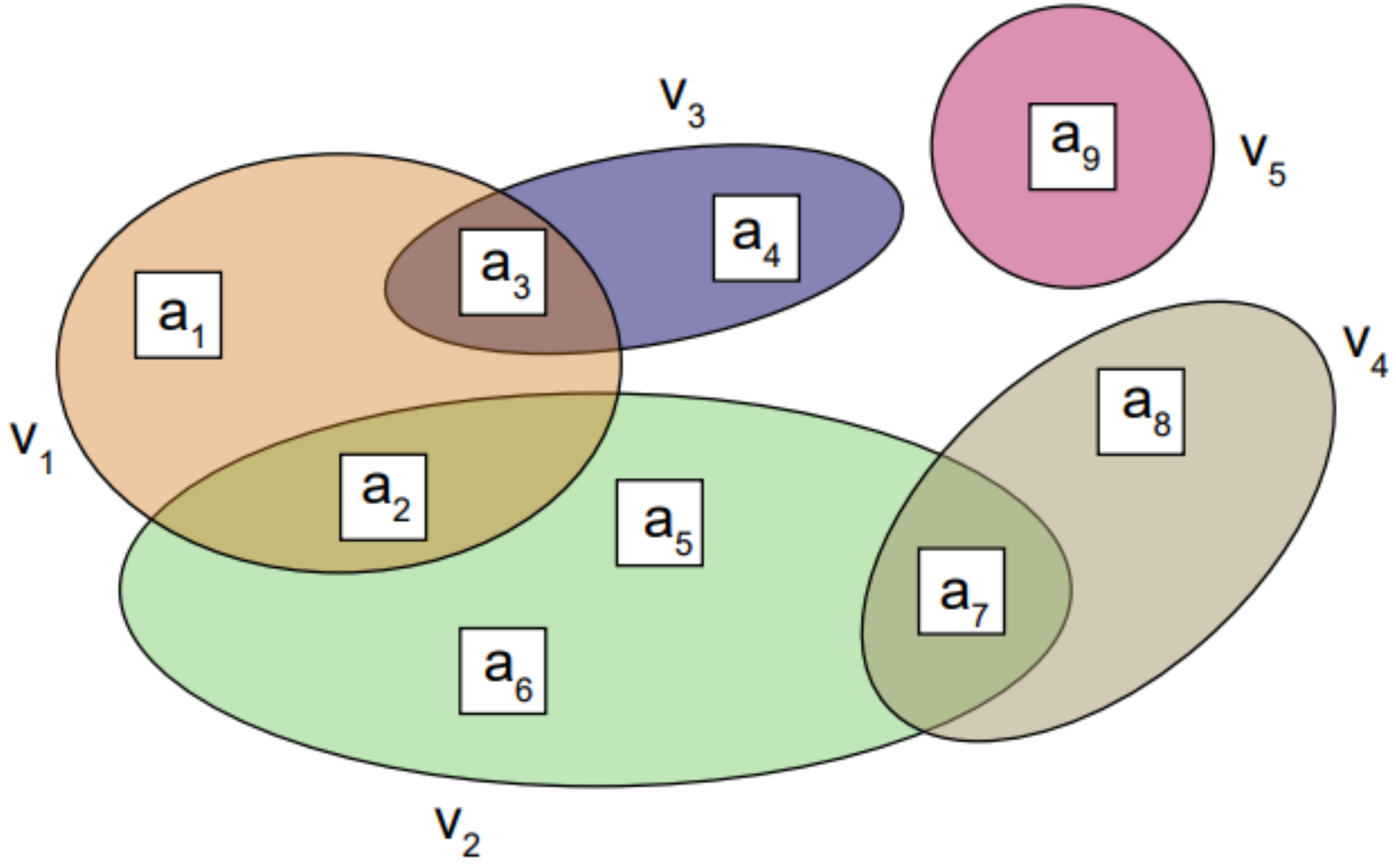
AND, the choice of network protocol, despite its importance, has not been widely studied in this literature!

Our Setting: Collaborative, *Heterogeneous* Networked MABs

Communication network



Arm distribution



Our Setting: Collaborative, *Heterogeneous* Networked MABs

- Two heterogeneities from **network structure** and **arm heterogeneity**
  - Well-connected *very hard* agent vs. Poorly-connected *very easy* agent
  - Who will learn first? Who will be more helpful in exploration of neighbors?

→ Unique challenge in the regret analysis!

- If the network is **complex**, then how to manage communication complexity to not overshadow regret improvement?

→Prompts the need to consider efficient *network protocol design* in this particular setting!

Network Protocols

Instantaneous Reward Sharing (IRS), Flooding, **Flooding with Absorption (FwA)**

Protocol #1. Instantaneous Reward Sharing

- Each message is sent to only its neighbors, and the messages get discarded
- The easiest and the most naïve protocol.

This has good (low) communication complexity

BUT,

- agnostic to the heterogeneity of network *and* bandit instances
- high good group regret

Protocol #2. Flooding

- We use a sequence number-controlled flooding (SNCF) variant
  - to avoid echoing, loops, and potential broadcast storm
- Each message is *passed along* to all its neighbors til the time-to-live (TTL) $\gamma > 1$

(see our paper for the precise regret analysis of UCB + Flooding, which is new!)

This results in the best (lowest) group regret.

BUT,

- agnostic to the heterogeneity of network *and* bandit instances
- always incurs high communication complexity

Our Protocol. Flooding with Absorption (FwA)

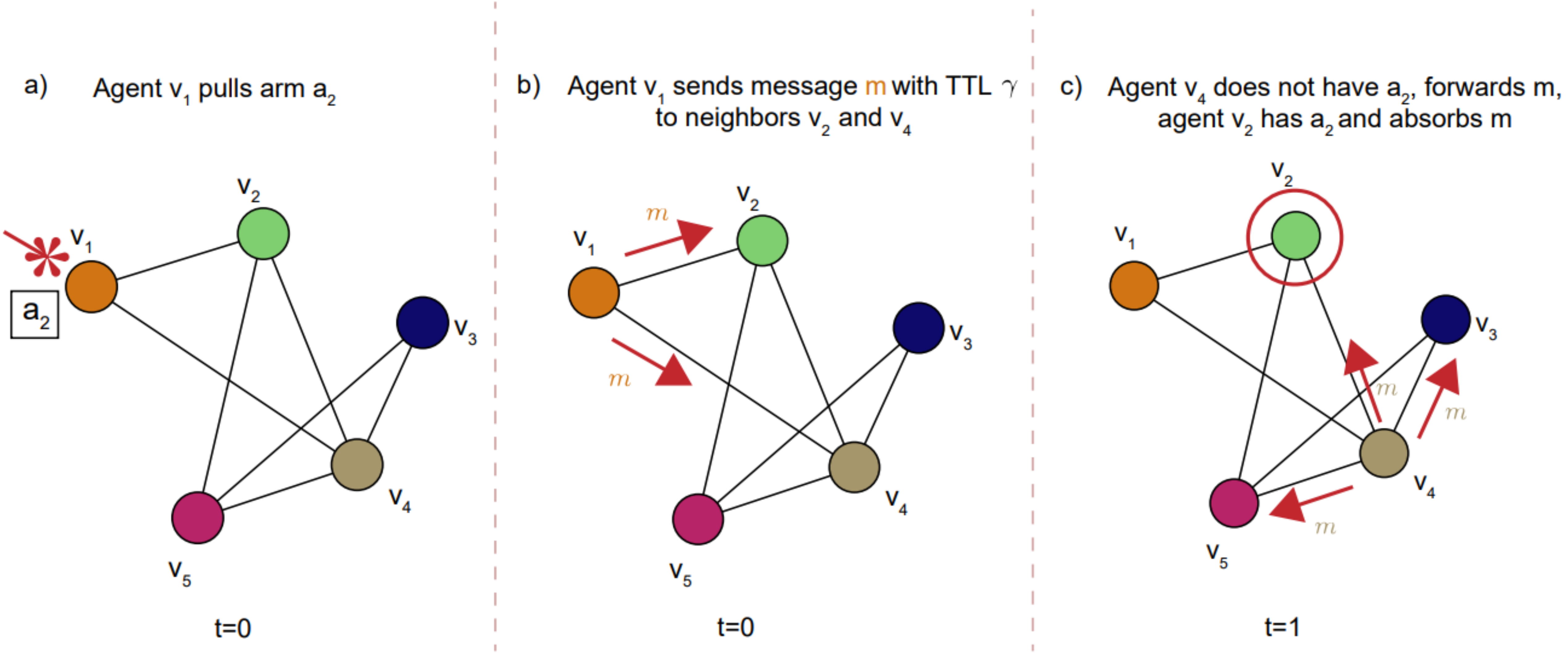
1. Agent pulls one of their arms with highest UCB.

2. Agent creates and sends message containing arm index *a* and received reward to all neighbors

3. *Neighbors with arm a absorb the message*, otherwise forward it unless time-to-live (TTL) expires

- **Prevent routing loops:** hash-based sequence number controlled flooding
- No knowledge of the network topology required!

Our Protocol. Flooding with Absorption (FwA)



Our Protocol. Flooding with Absorption (FwA)

Some advantages:

- Interpolating IRS and Flooding
  - In *dense* region, FwA ~ IRS; in *sparse* region, FwA ~ Flooding
- Comparable regrets guarantees (see our paper)
- Communication Efficiency
- No tuning beyond TTL, i.e., implementation is network-agnostic!

Remark. FwA is similar to, but is quite different from replication-based epidemic- and other controlled flooding and P2P systems.

Experiments

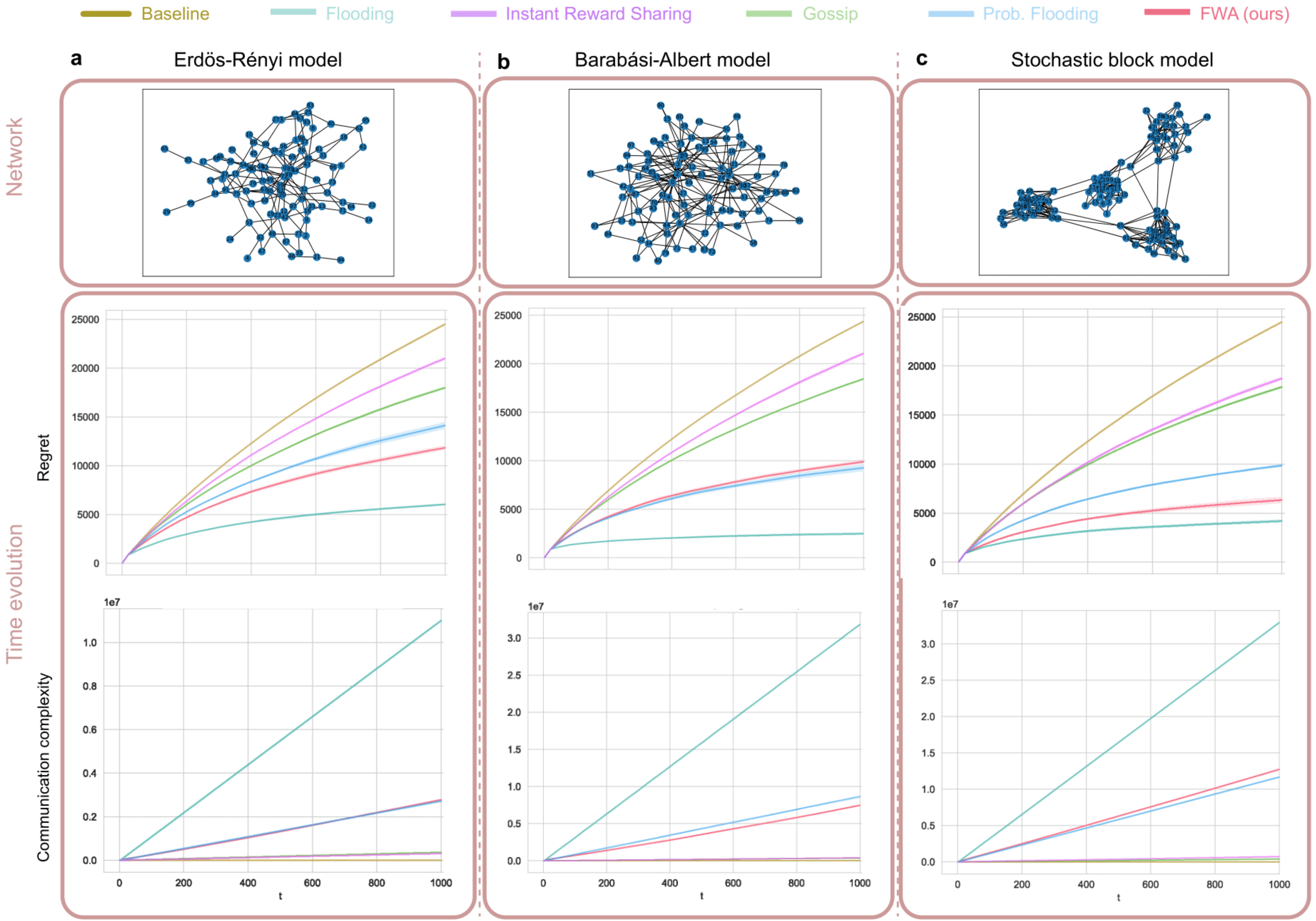
Baseline Comparison, Link Congestion, Dynamic Networks

Baseline Comparison (Static Network)

- We compare *group regret* and *communication complexity*
- Total of six network protocols:
  - No collaboration
  - Flooding
  - Probabilistic Flooding
  - Instantaneous Reward Sharing
  - Gossiping
  - **Flooding with Absorption (FwA)**

What we want:

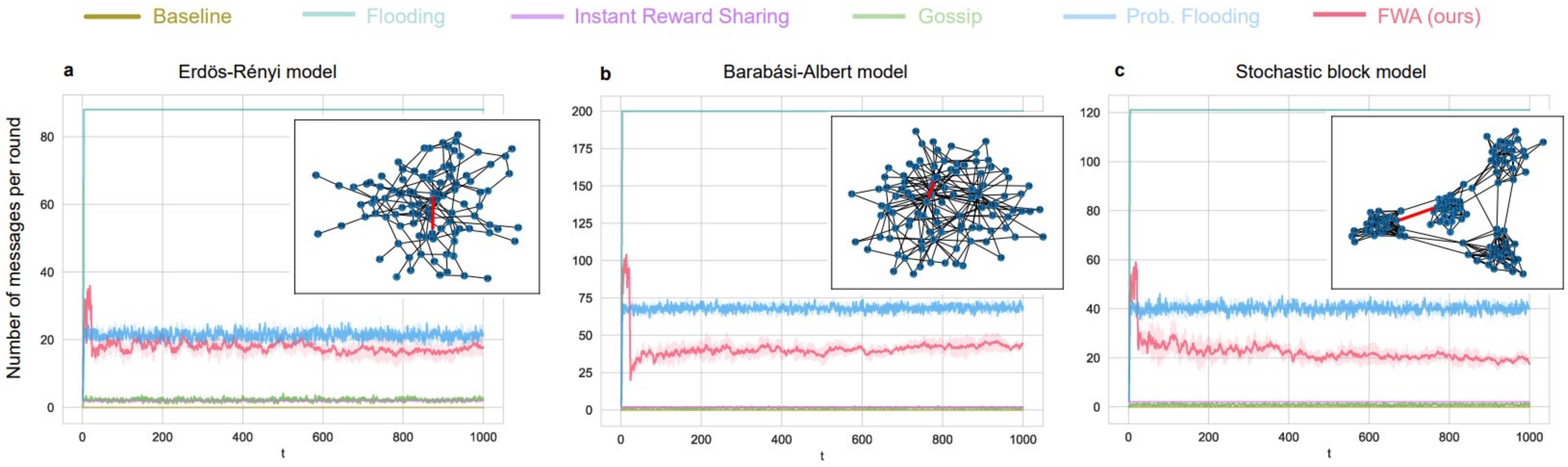
***Similar* group regret to Flooding, *Less* communication complexity**


Baseline
Flooding
Instant Reward Sharing
Gossip
Prob. Flooding
FWA (ours)
a
Erdös-Rényi model
b
Barabási-Albert model
c
Stochastic block model
Network
Time evolution
Regret
Communication complexity
1e7
t

Link Congestion

- We now compare the *#messages passed through a bottleneck edge*

FwA alleviates link congestion, compared to Flooding!



Baseline Comparison (Dynamic Network)

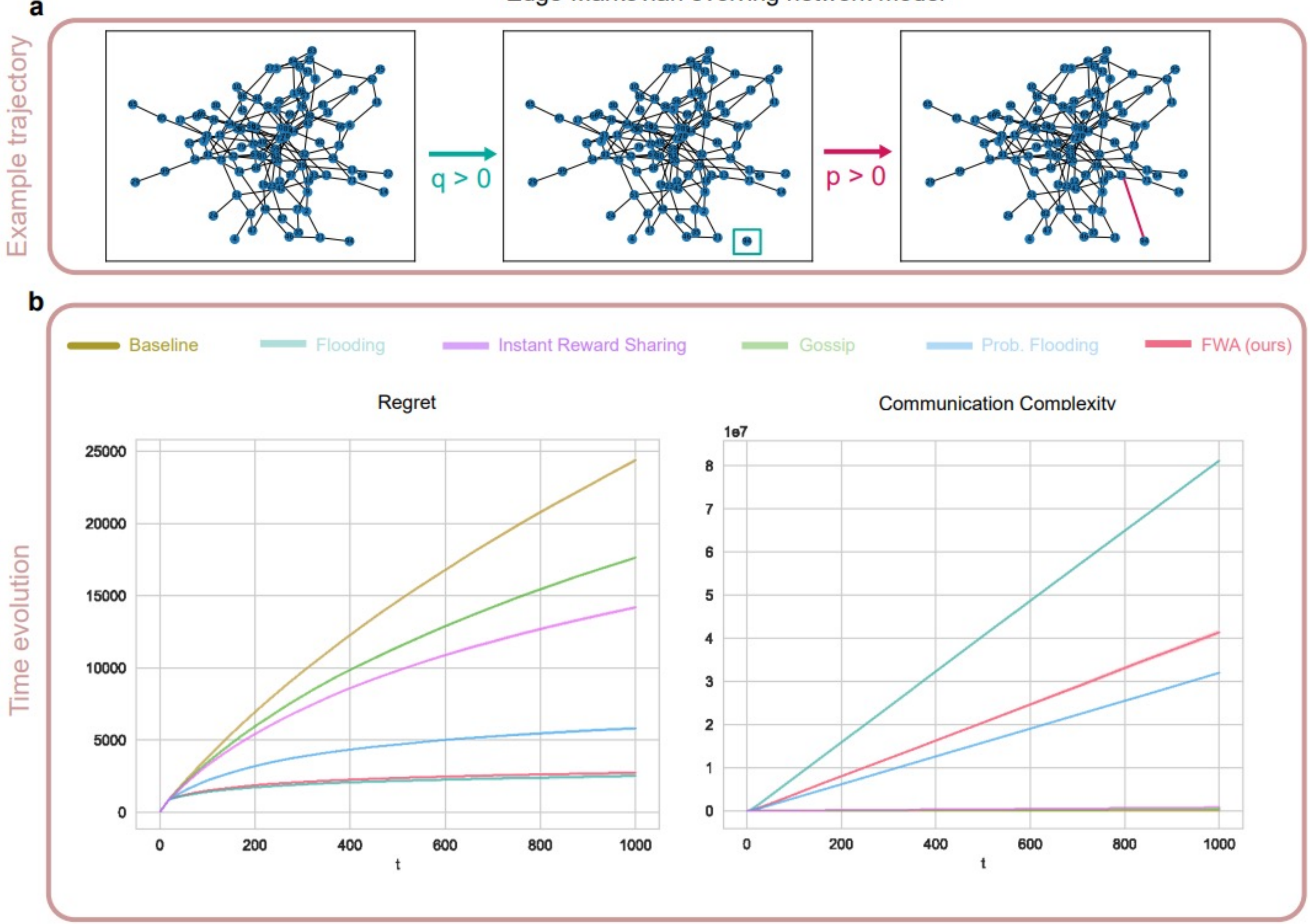
- Same setting, except the network is now *time-varying*
- We consider *edge-Markovian* model:

$$\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0) \rightarrow \; \mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1) \rightarrow \; \ldots$$

$$\mathbb{P}[e \in \mathcal{E}_t | e \notin \mathcal{E}_{t-1}] = p, \qquad \mathbb{P}[e \notin \mathcal{E}_t | e \in \mathcal{E}_{t-1}] = q$$

- We expect **Flooding with Absorption** to perform better, as it implicitly "adapts" to the given network structure!

Edge-Markovian evolving network model



Conclusion

- New setting: Collaborative, *Heterogeneous* Networked MABs
- New network protocol: **Flooding with Absorption (FwA)**
- Extensive experiments showing the efficacy of our FwA

Future Works.

- Network-dependent regret/communication lower bound
- Provably optimal network protocol for networked, heterogeneous bandits?

Thank you for your attention!



{jh_lee00, laura.schmid, yunseyoung}@kaist.ac.kr

Full paper (arXiv)

GitHub link