Algorithm 1

Setting

- **Unknown** $(p_h)_h$ — (symmetric) probability of influences
- In each time step $t = 1, \ldots, n$
- Learner picks a node $h$
- Environment **reveals** the set of influenced nodes $S_t$
- Select influential people — find the strategy maximizing $L = \sum_{t=1}^{n} |S_t|$.

**Our solution.**

**Bandit Revelator:** 2-phase algorithm

- **Global exploration phase.**
  - Super-efficient exploration
  - Linear regret needs to be short!
  - Exit $B$-nodes
- **Bandit phase.**
  - Uses a minimax-optimal bandit algorithm
  - GraphMoss is a little brother of MOSS
  - Has a “square root” regret on $B$-nodes
  - $D$-realizes the optimal trade-off!
  - Different from exploration/exploitation tradeoff.

**Guarantees.**

Upper bound on the regret of BARE $\mathbb{E}[R_n] \leq C \min (r_d, n, D, r_s + \sqrt{r_s n D_n})$

**Matching lower bound.**

**Algorithm.**

**BARE - Bandit Revelator**

**Input:** $d$: the number of nodes $n$: time horizon

**Initialization:** $\bar{T}_t = 0$, for $\forall t < d$

$d \leftarrow 1$, $T_0 = 0$, $D_{t+1} = d$, $\bar{N}_{t+1} = d$

Global exploration phase:

while $\bar{t} \leq d - 1$

for $i = 2d + 1, \ldots, n$

$C_{i, t+1} \leftarrow 2N_{t+1} \log (n r d) + 2 r_s \log (n r d) / (t+1)$

$L_{t+1} = \max_{1 \leq h < d} \{ C_{i, t+1} \}$. Sample node $h_t$ and receive $S_{t+1}$.

Update $T_{t+1}$, $\bar{N}_{t+1}$, and $T_{t+1}$.

end while

**Bandit phase.**

Run minimax-optimal bandit algorithm on the $D_{t+1}$-chosen nodes (e.g. Algorithm 1).