Lecture 17: Infect (4/4)

How do epidemic and gossip reach people? (i.e., how computer viruses spread?)

COMS 4995-1: Introduction to Social Networks

Thursday, November 13th

Outline

* Continuous epidemics, “logistic model”
* Discrete epidemics, “graph”
* Epidemic algorithms

Epidemic Algorithms

- Replicated database maintenance
  - Different versions, many locations
  - How to handle communication? failures ?
- 1987 “Epidemic alg., rumor spreading, gossip”
  - Do not maintain fixed communication topology
  - Contact a node unif., spread if one node has a copy
- How many rounds S_n before rumor spreads to all
  - $S_n = (1+1/\ln(2)) \log(n) + O(1)$ in probability

On spreading a rumor, B. Pittel, SIAM J. Appl. Math. (1987)

Epidemic algorithms for replicated database maintenance,
A Demers et. al, ACM PODC. (1987)

CS@CU

How gossip compares to optimal?

* What about using simply a fixed binary tree:
  - Also takes time O(log(n)), using O(n) messages
  - Seems optimal in both ways, but prone to failure
* Gossip:
  - Time O(log(n)) (optimal) and O(n log n) messages
  - In fact, unif. gossip requires at least $\omega(n)$ messages, and $\Omega(n \log\log(n))$ if no addresses are kept (the latter can be attained)

Randomized rumor spreading,
R Karp and C Schindelhauer and S Shenker and B Vocking, FOCS. (2000)

CS@CU

Effect of network topology

- What if communication is constrained?
  - Draw a graph between gossiping nodes G=(V,E)
  - A node u can contact v only if (u,v) is an edge in E
  - Let $P_{u,v}$ be the communication matrix between nodes
    - (u,v) not in E implies $P_{u,v} = 0$
- Main questions:
  - Which P ensures fast gossip dissemination?
  - How does gossip dissemination compares to optimal?

push

pull

Effect of network topology

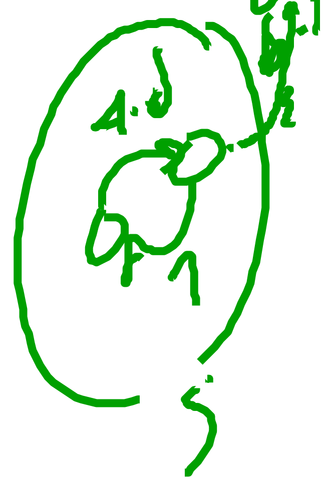
$S(t)$: { nodes that contains information at time t } $S(0) := \{v\}$

* Main result: If P irreducible, symmetric
  - Let $$T^{\text{one}}_{\text{spr}}(\varepsilon) = \sup_{v \in V} \inf \{ t : \Pr(S(t) \neq V \mid S(0) = \{v\}) \leq \varepsilon \}$$

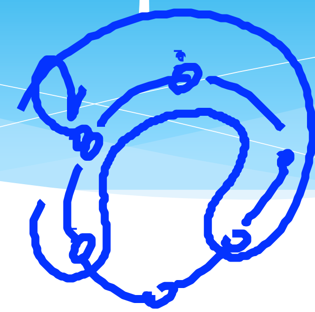
    $T^{\text{one}}_{\text{spr}} \leq 10$

  - We have $$T^{\text{one}}_{\text{spr}}(\varepsilon) = O\left(\frac{\log n + \log \varepsilon^{-1}}{\Phi(P)} \right)$$

  - Where $$\Phi(P) = \min_{S \subset V : |S| \leq n/2} \frac{\sum_{i \in S; j \in S^c} P_{ij}}{|S|}$$



How gossip compares to optimal?



* Depending on graph topology
  - Let $\varepsilon = \Omega(1/n^a)$ for a given a>0
  - Complete graph: $P_{u,v} = 1/n$; $\Phi(P) = 1/2$

    Already seen that $T^{one}_{spr}(\varepsilon)$ is $O(\log n)$, which is optimal
  - Ring: $P_{u,u+1} = 1/4$, $P_{u,u-1} = 1/4$, $P_{u,u} = 1/2$; $\Phi(P) \propto 1/n$

    $T^{one}_{spr}(\varepsilon) = O(n \log n)$, optimal uses at least n steps
  - α_expander, d regular: $P_{u,v} = 1/2d$, $P_{u,u} = 1/2$; $\Phi(P) = \alpha/2d$
    $T^{one}_{spr}(\varepsilon) = O(\log n)$, which is optimal

CS@CU

Proof

* Two phases:
  1. From $S(t) = \{v\}$ to $L-1$ Push
  2. From $L = \inf\{ t \mid \#S(t) > n/2 \}$ to $\#S(t) = n$ Pull.
* Ingredients of the proof: Phase 2
  a. Assume L is attained and hence $\#S(L)>n/2$
  b. Study evolution of conditional expectation $E[\#S(t+1) - \#S(t) \mid S(t)]$
  c. Uses Markov inequality ($X\geq 0 \Rightarrow P[X\geq a] \leq E[X]/a$)

Proof

* Two phases:
  1. From $S(t) = \{v\}$ to $L-1$
  2. From $L = \inf\{ t \mid \#S(t) > n/2 \}$ to $\#S(t) = n$
* Ingredients of the proof:
  a. Study evolution of conditional expectation $E[\#S(t+1) - \#S(t) \mid S(t)]$
  b. Uses Markov inequality ($X \geq 0 \Rightarrow P[X \geq a] \leq E[X]/a$)
  c. For phase 1, need to rewrite as super-martingale

Epidemic algorithm: Summary

* Not far from SI epidemic spread
  - With emphasis on communications constraints
* Key property: graph conductance
* Many extensions:
  - Send a message from each node
  - Send a stream of messages
  - Compute average value

Proof

$S(t),\quad S(t)^c$

$$\mathbb{E}\left[\, |S(t)^c| - |S(t+1)^c| \;\middle|\; S(t) \right] \geq \sum_{\substack{j \in S(t)^c \\ i \in S(t)}} P_{ji} \geq |S(t)^c|\,\phi(P)$$

$$\mathbb{E}[|S(t)^c|] - \mathbb{E}[|S(t+1)^c|] \geq \phi(P)\, \mathbb{E}[|S(t)^c|]$$

$$\mathbb{E}[|S(t+1)^c|] \leq \mathbb{E}[|S(t)^c|] \times (1-\phi(P))$$

$$\leq \mathbb{E}[|S(t)^c|]\, e^{-\phi(P)}$$

$1 - x \leq e^{-x}$

Proof

$$O\left(\frac{\log n + \log(\frac{1}{\varepsilon})}{\phi(P)}\right)$$

$$|S(0)^c| \le \frac{n}{2}$$

$$\mathbb{E}[|S(t)^c|] \le e^{-\phi(P)\cdot t} \cdot \frac{n}{2}.$$

$$P[|S(t)^c| \neq 0] \text{ small.}$$

$$P[|S(t)|^c > 0] = P[|S(t)^c| \ge 1] \le \frac{n}{2}\exp(-\phi(P)\cdot t)$$