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
* 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

Epidemic algorithms for replicated database maintenance,
A Demers et. al, ACM PODC. (1987)
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)
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?
Main result: If P irreducible, symmetric

- Let
  \( T_{\text{spr}}^{\text{one}}(\varepsilon) = \sup_{v \in V} \inf \{ t : \Pr(S(t) \neq V | S(0) = \{v\}) \leq \varepsilon \} \)

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

- Where
  \[
  \Phi(P) = \min_{S \subset V : |S| \leq n/2} \frac{\sum_{i \in S, j \in S^c} P_{ij}}{|S|} \]
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 $T_{\text{one spr}}(\varepsilon)$ is $O(\log n)$, which is optimal
- Ring: $P_{u,u+1} = 1/4, P_{u,u-1} = 1/4, P_{u,u} = 1/2; \Phi(P) \propto 1/n$
  $T_{\text{one spr}}(\varepsilon) = O(n \log n)$, optimal uses at least $n$ steps
- $\alpha$ _expander, d regular: $P_{u,v} = 1/2d, P_{u,u} = 1/2; \Phi(P) = \alpha/2d$
  $T_{\text{one spr}}(\varepsilon) = O(\log n)$, which is optimal
* 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: 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 ($X \geq 0 \Rightarrow P[X \geq a] \leq E[X]/a$)
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
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

\[ S(t), S(t)^C \]

\[ \mathbb{E} \left[ |S(t)^C - 1|S(t+1)^C \mid S(t) \right] \geq \sum_{i \in S(t)} \mathbb{P}_{j_i} \geq |S(t)| \cdot \phi(t) \]

\[ \mathbb{E} \left[ |S(t+1)| - \mathbb{E}|S(t+1)| \right] \geq (\Phi(\phi)) \cdot \mathbb{E}|S(t)| \cdot \phi(t) \]

\[ \mathbb{E}[S(t+1)] \leq \mathbb{E}[S(t)] \cdot \Phi(1-\phi(t)) \leq \mathbb{E}[S(t)] \cdot e^{-\Phi(t)} \]
Proof

\[ |S(0)| \leq \frac{n}{2} \]

\[ \mathbb{E}[|S(\epsilon)|] \leq e^{-\exp(-\varphi(p)) \cdot \epsilon} \cdot \frac{n}{2}. \]

\[ \mathbb{P}[|S(\epsilon)| \neq 0] \text{ small} \]

\[ \mathbb{P}[|S(\epsilon)| > 0] = \mathbb{P}[|S_{\epsilon}(\epsilon)| \geq 1] \leq \frac{n}{2} \exp(-\varphi(p) \cdot \epsilon) \]