# Lecture 14: Infect (4/4)

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

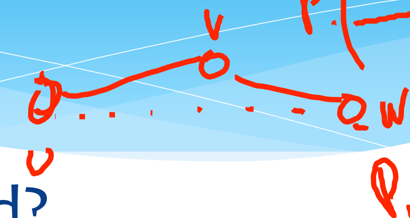
COMS 4995-1: Introduction to Social Networks
Part II: Cascade
Thursday, October 27th

# Outline

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

# Effect of network topology

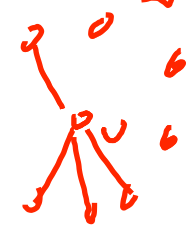
$P_{uw}$ $P_{uv}$ $P_{uw}$ $P.(\ )$




* What if communication is constrained?
  - Draw a graph between gossiping nodes G=(V,E) — A
  - 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

$\sum_{v \in V} P_{uv} = 1$

$P \leq A + \mathbb{1}$ $\quad \sum_{v \in V} A_{uv} = deg(u)$



* Main questions:
  - Which P ensures fast gossip dissemination?
  - How does gossip dissemination compares to optimal?



CS@CU

# Effect of network topology

$S(t)$: {nodes with info at time $t$}.

* Main result: If P irreducible, symmetric
  - Let $$T^{\text{one}}_{\text{spr}}(\varepsilon) = \sup_{v \in V} \inf \{t \colon \Pr(S(t) \neq V \mid S(0) = \{v\}) \leq \varepsilon\}$$
  - We have $$T^{\text{one}}_{\text{spr}}(\varepsilon) = O\left(\frac{\log n + \log \varepsilon^{-1}}{\Phi(P)}\right)$$
  - Where $$\Phi(P) = \min_{S \subset V : |S| \leq n/2} \frac{\sum_{i \in S; j \in S^c} P_{ij}}{|S|}$$

$\{v\}$

CS@CU

# 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

Tmsepp.-o/K+
I., (ii.).-a laying
i-c-55<s'
a-'-I m/55/i<n-
. o-t-J-o1-'-17
.-'W.
= 'tz
E-P,,÷ii
hi,'s,-51/n/i

# 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

* 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 ($X \geq 0 \Rightarrow P[X \geq a] \leq E[X]/a$)



CS@CU

# Proof

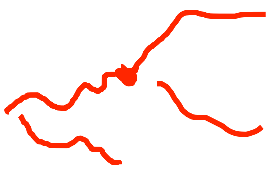
$S(t)$ = all nodes with info $\qquad S(t)^c$ = all nodes without info

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

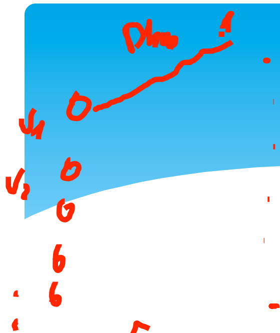
$$|S(0)^c| \leq n/2$$

$$\phi(P) = \min_S \frac{\sum_{i \in S,\ j \notin S} P_{ij}}{|S|}$$

$$\mathbb{E}[|S(t)|^c] - \mathbb{E}[|S(t+1)^c|] \geq \mathbb{E}[|S(t)^c|]\,\phi(P) \leq \frac{\sum_{i \in S(t),\ j \in S(t)^c} P_{ij}}{|S(t)^c|}$$



# Proof



$$\mathbb{E}[|S(t)^c|] \le \frac{n}{2}\exp(-t\cdot\varphi) \qquad T^{an}_{spr} \le \frac{\log n + \log(\frac{1}{\varepsilon})}{\varphi}$$

$$T^{an}_{sp}: \; P[\,S(t) \ne V \mid S(0) = \{v\}] \le \varepsilon.$$

$$P[|S(t)^c| > 0] = P[|S(t)^c| \ge 1] \le \mathbb{E}[|S(t)^c|] \le \frac{n}{2}e^{-t\varphi}$$

$$t = (2\ln n + \ln(\tfrac{1}{\varepsilon}))\frac{1}{\varphi} \qquad \frac{n}{2}\exp(t\varphi) = \frac{n}{2}\exp(-(2\ln n + \ln(\tfrac{1}{\varepsilon}))) = \frac{1}{n}\frac{\varepsilon}{2}$$



CS@CU