# Lecture 20: Infect (2/2)

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

COMS 4995-2: Introduction to Social Networks

Thursday, December 8th

# Epidemic model #2: $S \rightarrow I \rightarrow R$

- Thm: Assuming $\beta\rho<1$, $E[\,|Y(\infty)|\,] \leq C\sqrt{N} / (1-\beta\rho)$
  - $\rho(G)$: largest eigenvalue of G's adjacency matrix
  - $C = \sqrt{\#\{\text{initial infected population}\}}$
- If $\beta\rho<1$ and $C=o(\sqrt{N})$, negligible fraction removed

- Examples:
  - G is d-regular (same degree): $\rho(G) = d$
  - Can be applied to bound unif. random graphs

# Proof: recap of step1

* By counting all chains of infection from v to u

$$\mathbb{E}[|Y(\infty)|] = \sum_{u\in V} P[Y_u(\infty)=1] \leq \sum_{u\in V}\sum_{v\in v} X_v(0)\cdot \sum_{t\geq 0}\beta^t A^t_{v,u}$$

  – Because $A^t_{v,u}$ is # sequences $v=u_0,u_1,\ldots,u_t=u$
  – The chance that each sequence succeeds is $\beta^t$
  – And probability of union event ≤ sum of probability

# Proof: rewriting

$$\mathbb{E}[|Y(\infty)|] = \sum_{u\in V} P[Y_u(\infty)=1] \leq \sum_{u\in V}\sum_{v\in v} X_v(0)\cdot \sum_{t\geq 0} \beta^t A^t_{v,u}$$

- Rewrite as $\mathbb{E}[|Y(\infty)|] \leq \left\langle e_1, \sum_{t\geq 0} (\beta A)^t X(0) \right\rangle$
  - Same using vector/matrix notation
  - Where $\langle x, y\rangle = \sum_{i=1}^{N} x_i \cdot y_i$ and $e_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

# Some definition on norms

* We introduce norms
  - vectors $||v||_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$ Note it implies $\langle x, y \rangle \leq ||x||_2 \cdot ||y||_2$
  - and matrices: $||A||_2 = \max_{x \in \mathbb{R}^n} \frac{||A.x||_2}{||x||_2}$
  - This implies $||A \cdot x||_2 \leq ||A||_2 \cdot ||x||_2$
  - This also implies $||A|| = \rho(A)$

# Proof: completing

$$\mathbb{E}[||Y(\infty)||] \leq \left\langle e_1, \sum_{t\geq 0} (\beta A)^t X(0) \right\rangle$$

* We deduce $\mathbb{E}[||Y(\infty)||] \leq ||e_1|| \times ||\sum_{t\geq 0} (\beta A)^t|| \times ||X(0)||$
* Note that $\sum_{t\geq 0} (\beta A)^t = (\mathrm{Id} - \beta A)^{-1}$
  – First, the series converge as $\beta\rho(A) < 1$
  – Second, we can verify it is the inverse of

# Conclusion

- We hence have

$$\mathbb{E}[||Y(\infty)||] \leq ||e_1|| \times ||(\mathrm{Id} - \beta A)^{-1}|| \times ||X(0)||$$

  - From there, we can conclude the theorem
  - E[ |Y(∞)| ] ≤ C √N / (1-βρ) where C=√initial inf. pop.

# Epidemic model #2: S→I→R

* Thm: Assuming $\beta\rho<1$, $E[\ |Y(\infty)|\ ] \leq C\sqrt{N} / (1-\beta\rho)$
  - $\rho(G)$: largest eigenvalue of G's adjacency matrix
  - $C = \sqrt{\#\{\text{initial infected population}\}}$
* Examples of application
  - G is d-regular (same degree)?
  - G is a complete graph?
  - G is a star network?
  - G a uniform random graph?

# Outline

* Continuous epidemics, "logistic model"
* Discrete epidemics, "graph"
  - Adjacency matrix
  - SI, SIR model
  - SIS
* Epidemic algorithms

# Epidemic model #3: S↔I

* Nodes follow neighbor contamination / recovery
  - Node $u \in V$ infectious ($X_u = 1$) or susceptible ($X_u = 0$)
  - Node u becomes infected with rate $\beta \cdot \sum_{v \in N(u)} X_v$
  - Node u recovers with rate $\gamma=1$
* In a finite graph, all nodes eventually recover
  - Because ($X_u = 0 \ \forall u \in V$) is the only absorbing state
  - Different on infinite graphs (e.g. lattices, trees)

# Epidemic model #3: S↔I

* Can we recover fast from an epidemy?
* Thm: $P[X(t) \neq (0,..,0)] \leq C \sqrt{N} \exp(t \cdot (\beta\rho - 1))$
  - $\rho(G)$: largest eigenvalue of G's adjacency matrix
  - $C = \sqrt{\#\{\text{initial infected population}\}}$
* Corollary: If $\beta\rho<1$, choosing $t=\ln(n)/(1-\beta\rho)$ we can prove
  - $E[\text{extinction time}] \leq (1+\ln(n)) / (1-\beta\rho)$
* Bottom line: goes to zero very fast if $\beta\rho<1$
  - complete graph: $\rho(G)=n-1$
  - uniform random graph: $\rho(G) \approx (n-1)p$ (if $np = \omega(\log n)$)

The effect of network topology on the spread of epidemics,
A Ganesh, L Massoulié, D Towsley, IEEE Infocom (2005)

# Proof:

* Step1: Introduce a random walk process $Z_u(t)$

* Intuitively we have $P[X(t) \neq 0] \leq P[Z(t) \neq 0]$
  - This statement can be made precise by coupling
  - Note: $P[Z(t) \neq 0] \leq \Sigma_v P[Z_v(t) \neq 0] \leq \Sigma_v E[Z_v(t)]$

# Proof:

* How does Z(t) evolve?

$$\frac{d}{dt} Z_u(t) = \sum_v \beta A_{u,v} Z_v(t) - Z_u(t)$$

  – This is a linear evolution!
* In expectation, it is $\frac{d}{dt}\mathbb{E}[Z_u(t)] = \sum_v \beta A_{u,v} \mathbb{E}[Z_v(t)] - \mathbb{E}[Z_u(t)]$
  – This is a linear deterministic evolution (N dimension)
  – Which is $\mathbb{E}[Z(t)] = e^{t(\beta A - \mathrm{Id})} Z(0) = e^{t(\beta A - \mathrm{Id})} X(0)$
* So that $P[Z(t) \neq 0] \leq ||e_1|| \, ||\exp(t \cdot (\beta A - I))\, X(0)||$

# Proof:

* Finally, we can apply the same bounding technique
  – $P[X(t) \neq 0] \leq ||e_1|| \, ||\exp(t \cdot (\beta A - I)) \, X(0) \, ||$

# Discrete epidemics: summary

| Type | Outcomes |
|---|---|
| S→I | Everyone infected |
| S↔I | No infectious nodes |
| S→I→R | No infectious node |

Follow processes of infection

- Initial conditions:
  small set infected nodes

Outcomes generally trivial

- Speed or span depend on graph topology
  (e.g. spectral analysis)

# 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)

# 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 ω(n) messages, and Ω(n loglog(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?

# Effect of network topology

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

# 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

# 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

* 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



CS@CU

# 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