Lecture 13: Infect (3/4)

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

COMS 4995-1: Introduction to Social Networks

Part II: Cascade

Tuesday, October 25th

Epidemic model #2: $S \xrightarrow[\beta]{} I \xrightarrow[1]{} R$

* Thm: Assuming $\beta\rho<1$, $E[\ |Y(\infty)|\ ] \le C\sqrt{N} / (1-\beta\rho)$
  - $\rho(G)$: 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 : $\rho(G) = d$
  - Can be applied to bound unif. random graphs

Outline

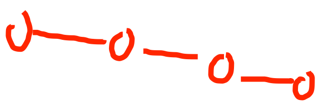
* Continuous epidemics, “logistic model”
* Discrete epidemics, “graph”
  – Adjacency matrix
  – SI, SIR model
  – SIS

* 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)

2^n



Epidemic model #3: S↔I

$X(t) = (X_v(t))_{v \in V}$

$e^{t(\beta\rho-1)} = \frac{1}{n^2}$

- Can we recover fast from an epidemy?

$\beta\rho < 1$ $\quad t \geq \frac{2 \ln n}{1-\beta\rho}$

- 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
  - $C = \sqrt{\#\{\text{initial infected population}\}}$

$P[\tau \geq t] \leq C\sqrt{N} e^{t(\beta\rho-1)}$

$\tau = \inf\{t \geq 0 \mid X(t) = (0,0,\dots,0)\}$

- Corollary: If we have $\beta \cancel{<} \rho$ $\beta\rho<1$
  - E[extinction time] $\leq (1+\ln(n)) / (1-\beta\rho)$ τ

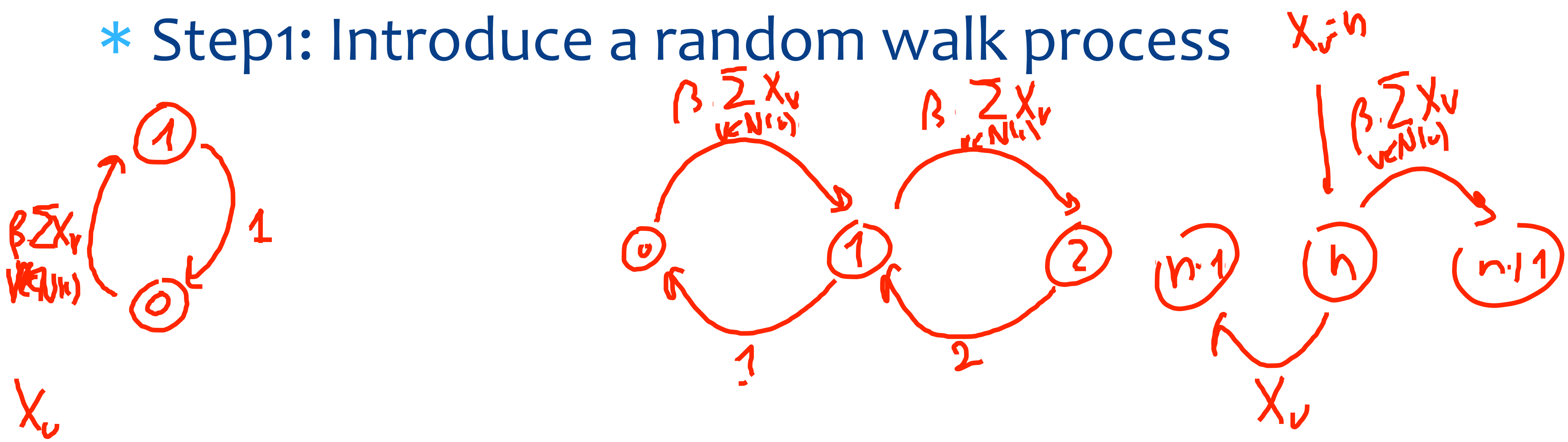
$E[\tau] = \int_0^\infty P[\tau \geq s]\, ds$

- Bottom line: goes to zero very fast if $\beta\rho<1$
  - complete graph: $\rho(G)=n-1$ $\quad \beta < \frac{1}{n-1}$
  - 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



* Intuitively we have $P[X(t) \neq 0] \leq P[Z(t) \neq 0]$
  – This statement can be made precise by coupling

CS@CU

Proof:

$e^z = \sum_{k \geq 0} \frac{z^k}{k!}$ $e^{tA} = \sum_{k \geq 0} \frac{(tA)^k}{k!}$

$(AX)_u = \sum_{v \in V} A_{uv} X_v = \sum_{v \in N(u)} X_v$

$(XA)_u = \sum_{v \in V} X_v A_{vu} = \sum_{v \in N(u)} X_v$

$P[X(t) \neq 0] \leq$

* We have $P[Z(t) \neq 0] \leq \Sigma_v P[Z_v(t) \neq 0] \leq \Sigma_v E[Z_v(t)]$ $\leq$ RHS in thm
  - Using Markov Inequality
* We can rewrite evolution of Z(t)

$A = \begin{cases} A_{uv} = 1 \text{ if } u \sim v \\ 0 \text{ otherwise} \end{cases}$

$\frac{d}{dt} \mathbb{E}[Z_v(t)] = \beta \sum_{v \in N(u)} \mathbb{E}[Z_v(t)] - \mathbb{E}[Z_u(t)]$

$\hat{Z}_u(t) = \mathbb{E}[Z_u(t)]$

$\frac{d}{dt} \hat{Z}_u(t) = \beta \sum_{v \in N(u)} \hat{Z}_v(t) - \hat{Z}_u(t) = (\hat{Z} \cdot \beta A)_u - \hat{Z}_u$

$\frac{d}{dt} \hat{Z}(t) = \hat{Z}(\beta A - I)$

* So that $P[Z(t) \neq 0] \leq ||e_1|| \; ||\exp(t \cdot (\beta A - I)) X(0)||$

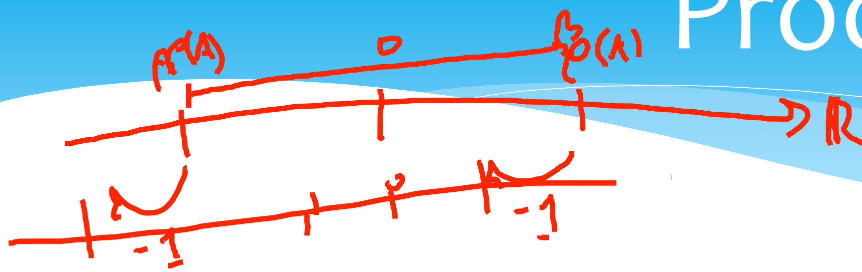
$\frac{dz}{dt} = z \cdot c$: $z = e^{ct}$

$\frac{d\hat{Z}}{dt} = \hat{Z}(\beta A - I) \Leftrightarrow \hat{Z} = \exp((\beta A - I)t) \, Z(0)$

$\frac{dz}{z} = c \, dt$

CS@CU

Proof:



$\hat{z}(0) = \mathbb{E}[z(0)] = X(0)$

* Finally, we can apply the same bounding technique

$$\frac{d\hat{z}}{dt} = \hat{z}\,(\beta A - \mathbb{1}_d) \Rightarrow \hat{z}(t) = \exp\big(t(\beta A - \mathbb{1}_d)\big)\, X(0)$$

$$P[X(t)\neq 0] \le \sum_v \mathbb{E}[z_v(t)] = \langle e_1, \hat{z}_v(t)\rangle =$$

$$\le \|e_1\| \times \|\underbrace{\exp\big(t(\beta A - \mathbb{1}_d)\big)}\| \; \underbrace{\|X(0)\|}_{\sqrt{c}}$$

$\sqrt{N}$ $sp\big(t(\beta A - \mathbb{1}_d)\big) \pm$

$A: \begin{cases} 1 \text{ if } u,v \text{ connected} \\ 0 \text{ otherwise} \end{cases}$

$\rho(A)$

A: $sp(A) = \{\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n\}$

$t(\beta A - \mathbb{1}_d)$: $sp(\;) = \{t(\beta\lambda_1 - 1), t(\beta\lambda_2 - 1)\}$

$\exp(\;)$: $sp(\;) = e^{t(\beta\lambda_1 - 1)}, e^{t(\beta\lambda_2 - 1)}$

$\exp(t(\beta\rho - 1))$ $\exp A = \sum_{k\ge 0} \frac{A^k}{k!} = \sum$ $A^k x_1$ $\frac{\lambda_1^k x_1}{k!}$ e^{λ_1}

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?

* A 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)