Lecture 12: Infect (2/4)

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 20th

Outline

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

* Epidemic algorithms

Discrete epidemics

* Infection only spreads along edges of a given graph
  - to account for connections, closeness among nodes
* Initial conditions: one or several nodes infected
* Challenges
  - Can a fraction be infected in a large graph?
  - What is the speed of evolution of epidemics?
  - How does it depend on the properties of the graph?
  - What if some individuals are immune?

Adjacency matrix

* A graph $G=(V,E)$ with N nodes
* Adjacency matrix A is N x N matrix
  - A_{uv}=1 if (u,v) in E, 0 otherwise

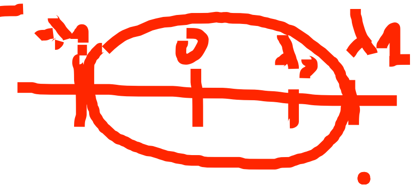
* Prop1: for any k, $(A^k)_{uv}$ is #{ paths u->v of length k}

$$(A^2)_{uv} = (A \cdot A)_{uv} = \sum_{w \in V} A_{uw} \cdot A_{wv}$$

Adjacency matrix

$\forall u,v \quad A_{uv} = A_{vu}$ $\qquad$ $\forall u,v \quad A_{u,v} \geq 0$

* Prop2: A is symmetric, with non-negative entries



* THM 1 (Perron Frobenius):
  - Eigenvalues of A are all real $\lambda_1 \geq \lambda_2 \geq \ldots$
  - λ_1 is the largest in absolute value ($\lambda_1 \geq |\lambda_i|$ for any i)
  - λ_1 has a non-negative eigenvector x_1

  Generally λ_1 is called the spectral radius of A, $\rho(A)$
* THM 2 (Stronger): if A irreducible (G connected)
  - All entries of x_1 are positive, λ_1 has no multiplicity $\lambda_1 > \lambda_2$
  - No other eigen vectors has only non-negative entries

Outline

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

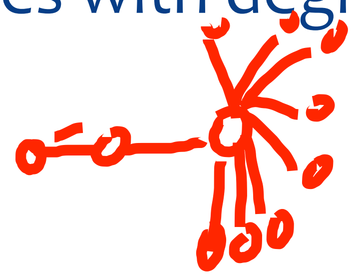
* Epidemic algorithms

Epidemic Model #1: S→I

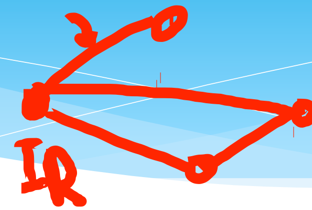
- Model 1: broadcast
  - Node infected at time t infects all its neighbors in t+1
  - Within time D=diam(G), all nodes are infected

- Model 2: gossip
  - Node infects each neighbor with a given rate β
  - Eventually all nodes are infected within O(D/β) time
  - What if rates reduces with degree ? (see later lecture)

Assuming G is connected



Epidemic model #2: S→I→R



* Model 1: single infection attempts
  – Infected node infect neighbors with probability β_e
  – Many names: “Independent cascade model”, Reed-Frost epidemics, SIR with single slot
* Model 2: Random infectious period (normalized)
  – Similar (probability to spread is β) but dependencies!
* Eventually: no infectious nodes, fraction removed

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

- What is the size of the removed fraction?
- 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 $C=o(\sqrt{N})$, remove only negligible fraction

- Proof based on expectation not on independence
  - Similar results hold for model 2

CS @CU

Proof

* X_v (resp. Y_v) = 1 iff v is infected (resp, inf or recov.)
* First, we bound $Y_v(t)$ using adjacency matrix

Proof

$\|X(0)\|$

$X(0) = (X_{u_1}(0), X_{u_2}(0), \dots)$

$e_1 = (1, 1, \dots, 1)$

* Second, we use a bounding norm technique

$$E[|Y(\infty)|] \leq \sum_{t=0}^{\infty} \sum_{v} 1 \sum_{u} \beta^t A^t_{uv} X_u(0)$$

$$\leq \sum_{v} 1 \cdot \left(\sum_{t=0}^{\infty} \sum_{u} \beta^t A^t_{uv} X_u(0) \right)$$

$$\leq \left\langle e_1^T, \left(\sum_{t=0}^{\infty} \beta^t A^t \right) X(0) \right\rangle$$

$$\leq \underbrace{\|e_1^T\|}_{\sqrt{n}} \times \underbrace{\left\| \sum_{t=0}^{\infty} \beta^t A^t \right\|}_{\leq \frac{1}{1-\rho\beta}} \cdot \underbrace{\|X(0)\|}_{\sqrt{|X(0)|}}$$

$\|v\|_2 = \sqrt{\sum v_1^2 + v_2^2 + \dots + v_n^2}$

$\|A\|_2 = \max_{x \in \mathbb{R}^n} \frac{\|Ax\|_2}{\|x\|_2}$

$\langle u, v \rangle \leq \|u\| \, \|v\|$

$A =$