Analysis of Social Information Networks

Thursday February 24th, Lecture 6: Epidemics

Outline

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

Model of Epidemics

* Dynamics of population growth
  - reproduction
  - access to resources (food)

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right)$$

  - r: rate of reproduction,
  - K: # indiv. sustainable with available resources

Solution of logistic dynamics

* Lemma: If dynamic system y satisfies dy/dt = c/f(y)
  - Then for any t, we have F(y(t))-F(y(0)) = c · t
  - where F is a primitive function of f

* Assume r=K=1 we have $$\frac{d}{dt}P(t) = P(t)(1 - P(t))$$

  - P satifies lemma with f(y) = 1/y + 1/(1-y)
  - Hence P(t) = P(0) / (P(0) + (1-P(0)) e^{-t})

CS @CU

Epidemic Model #1: S⟶I

* Individual infection is persistent:
  - Assume all infected individuals remain infectious
  - S ⟶ I model (S for "Susceptible", I for "Infected")
* Growth of infected populations (denoted by y):
  - Reproduction rate: infection probability β
  - Resources: non-infected nodes (e.g. n-y)

$$\frac{dy}{dt} = \beta x y = \beta y(n-y) \qquad y(t) = \frac{y(0)n}{y(0) + (n - y(0))e^{-\beta n t}}$$

CS@CU

Epidemic model #2: S↔I

- Assume that infection is temporary and recurrent
  - An infected node goes through an infectious period (equivalently, becomes non-infectious with rate γ)
  - After the period, it is again susceptible
- Evolution of

$$\frac{dy}{dt} = \beta y(n - y) - \gamma\, y$$

  - β>γ : endemics lim y=(1-γ/β)
  - β<γ : epidemics dies lim y = 0

Epidemic model #3: S→I→R

* Assume that infection is temporary and transient
  - An infected node goes through an infectious period (equivalently, becomes non-infectious with rate γ)
  - After the period, it is removed (vaccinated or dead)
* Evolution follows similar equation
  - Infected invididuals disappear: lim y=0
  - Number of removed individuals lim z satisfies

    $n - \lim z = x(0) \exp(- \beta \lim z / \gamma)$

CS@CU

Continuous epidemics: summary

Type	Outcomes
S→I	Everyone infected
S↔I	Depends on infection/recovery rate
S→I→R	No infectious node

Follows differential equation

- Initial conditions:
  Fraction already infected
- Outcomes depend on type
- Cvg exponentially fast

- Mean field limit of discrete
- No topology

Outline

* Continuous epidemics, "logistic model"
* Discrete epidemics, "graph"
* Epidemic algorithms

Discrete epidemics

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

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)
  - What if rates are not constant? (see further analysis)

Epidemic model #2: 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 #2: 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
  - $C = \sqrt{\#\{\text{initial infected population}\}}$
* Bottom line: goes to zero very fast if $\beta\rho<1$
  - complete graph: $\rho(G)=n-1$
  - hypercube: $\rho(G)=\log_2(n)$
  - 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)

CS@CU

Epidemic model #3: S→I→R

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

Epidemic model #3: S→I→R

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

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

CS@CU

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)

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?

CS@CU

Effect of network topology

* 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 $\Phi(P) = \min_{S \subset V : |S| \leq n/2} \frac{\sum_{i \in S; j \in S^c} P_{ij}}{|S|}$

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

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

UP-COMING

* Proof of previous results
* Consequences on spread of epidemics