Algorithmic and Economic Aspects of Networks

Nicole Immorlica

Diffusion through Networks

How do fads develop? How do diseases spread? What makes peaceful people riot? Why is English an international language?

Bass Model

People are either innovators or immitators, based on random stimulae and interactions.

they innovate at rate p,

and immitate at rate q

Bass Model

Let F(t) be fraction of agents who have adopted behavior by time t. Then,

$$F(t) = F(t\text{-}1) + p\cdot(1 - F(t\text{-}1)) + q\cdot F(t\text{-}1)\cdot(1 - F(t\text{-}1))$$

Additional innovators

Additional immitators

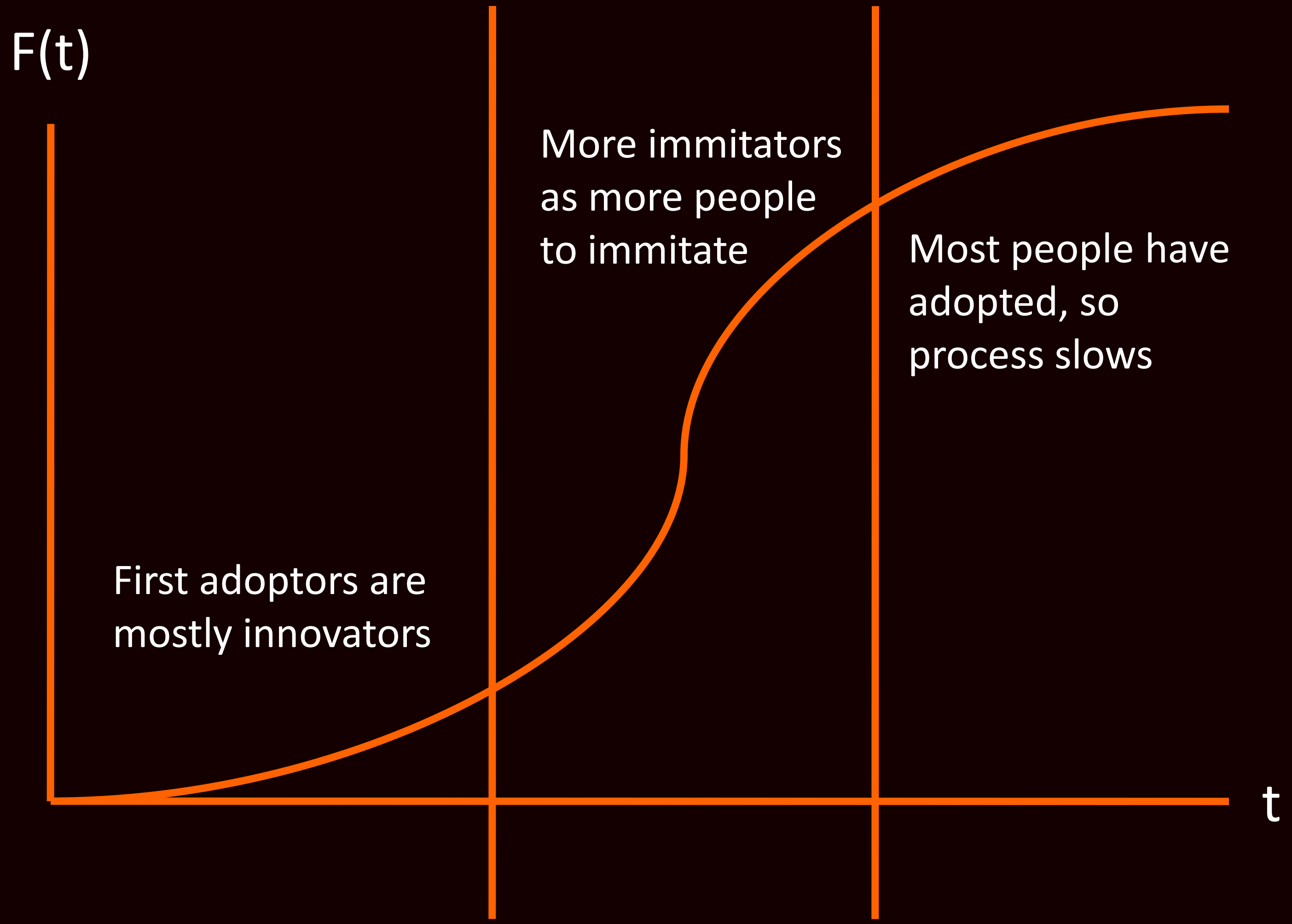
Bass Model

By continuous-time approximation, see

$$F(t) = [1 - \exp(-(p+q)t)] / [1 + (q/p)\exp(-(p+q)/t)]$$

Ratio of immitators to innovators

Bass Model


F(t)
More immitators
as more people
to immitate
Most people have
adopted, so
process slows
First adoptors are
mostly innovators
t

Extending Bass

Impose network structure.

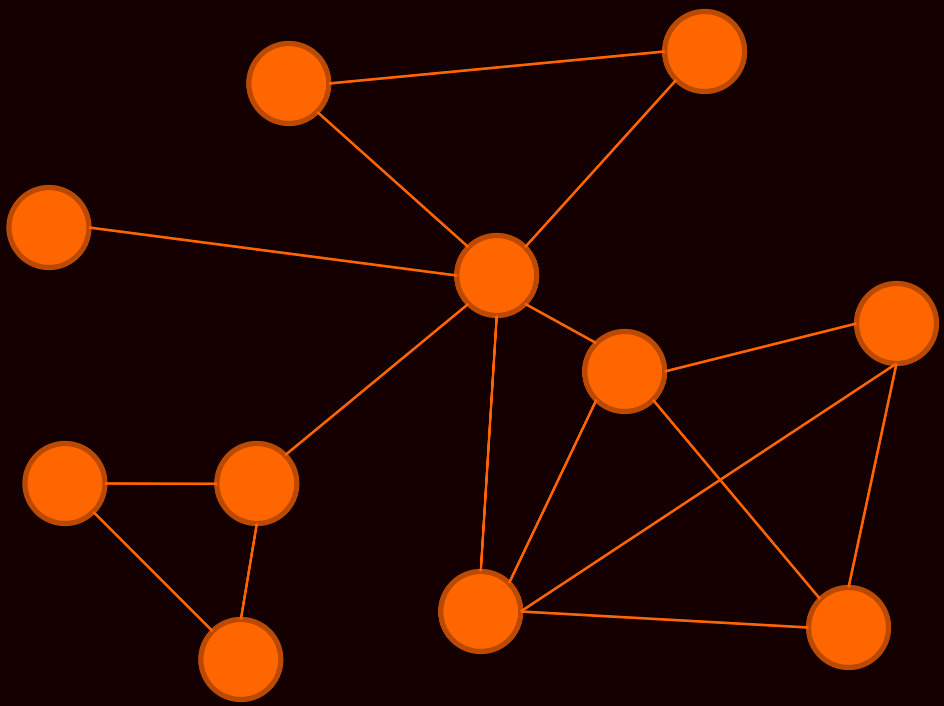
… like finding giant components

… SIR model – susceptible, infect, recover

Allow people to change their minds.

… SIS model – susceptible, infect, susceptible

SIS Model



SIS Model

Most significant factor is varying degrees.

Simplify model:

- People meet *randomly* as in Bass
- Different people have different # of meetings

SIS Model

People have degrees that govern amount of interaction.

$P(d)$ = fraction of people with degree d

$P(d) / \sum_d P(d)$ = probability of interacting with a person of degree d.

SIS Model

Let ρ(d) be fraction of people with degree d who are infected. Then prob. of meeting infected person is:

$$\Theta = \sum_{d} \left(\frac{P(d)}{\sum_d P(d)} \right) \times \rho(d)$$

SIS Model

If α is transmission rate, and β is recovery rate, then fraction of nodes of deg d who get infected is:

$$[(1 - \rho(d)) \cdot d] \times (\Theta \cdot \alpha)$$

and fraction of nodes that recover is:

$$\rho(d) \times \beta$$

SIS Model

Questions:

1. How high should infection rate be compared to recovery rate for disease to live?
2. In steady state, how many people infected?
3. How does this relate to network structure or degree distribution?

SIS Model

In steady state, fraction of infected equals fraction of recovered

$$(1 - \rho(d))d\alpha\Theta = \beta\rho(d)$$

or

$$\rho(d) = \lambda\Theta d \,/\, (\lambda\Theta d + 1)$$

where λ is ratio of α to β.

SIS Model

We know

1. Fraction of population that is infected

$$\Theta = \sum \left(\frac{P(d)}{\sum_d P(d)} \right) \times \rho(d)$$

2. Steady state equation

$$\rho(d) = \lambda\Theta d \,/\, (\lambda\Theta d + 1)$$

SIS Model

Solve for Θ,

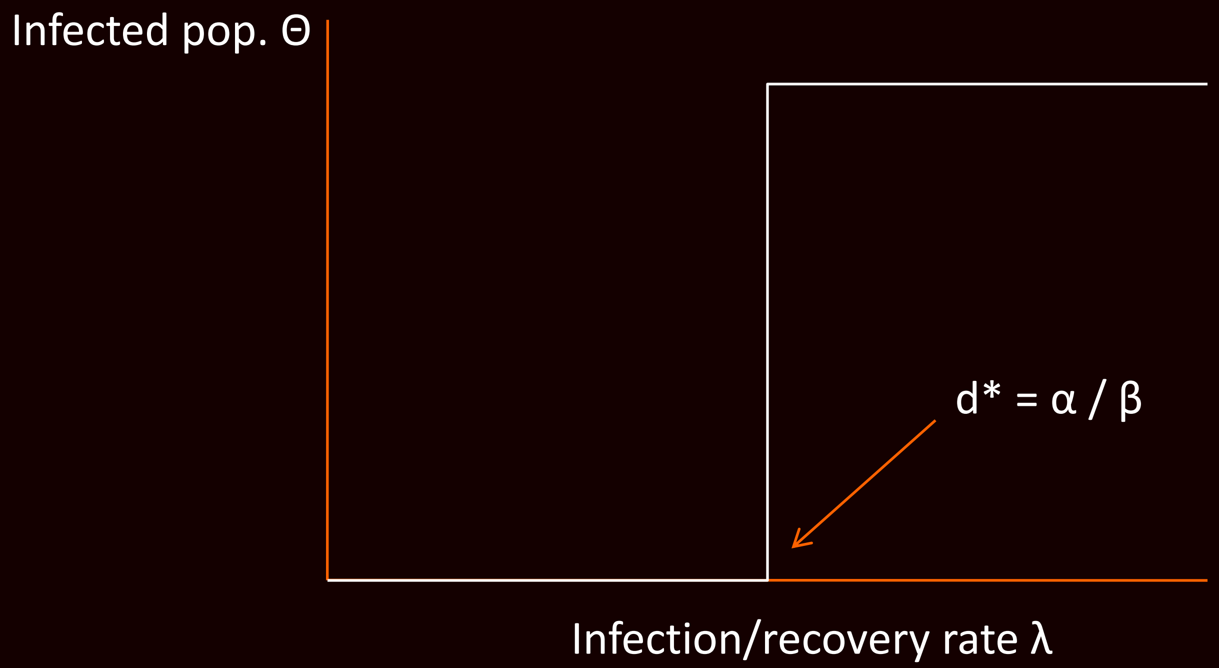
$$\Theta = \sum \left(\frac{P(d) \times \lambda\Theta d}{\sum_d P(d) \times (\lambda\Theta d + 1)} \right)$$

When all degrees are regular, say d*?

When degrees follow a power law $P(d) = d^{-2}$?

SIS Model

Regular degrees, Θ = 0 or Θ = 1 – 1/λd*



SIS Model

Power law degrees, see board

Example: Corrupted Blood

Lesson

Mean-preserving spreads in degree distributions (e.g., power-law vs Poission) lead to lower thresholds for infection.

Questions

How does immunization help? Which nodes should we immunize? How about quarantines? How sensitive is the model to variations in network structure or initial infection sets? What if disease is malicious (e.g., computer viruses)? How can it spread effectively, or spread to a particular person?

Search and Navigation

How to find a particular node in a network?

- do a random walk (perhaps biased by network characteristics)
- do a greedy walk based on similarity of neighbors to target

Finding Target Randomly

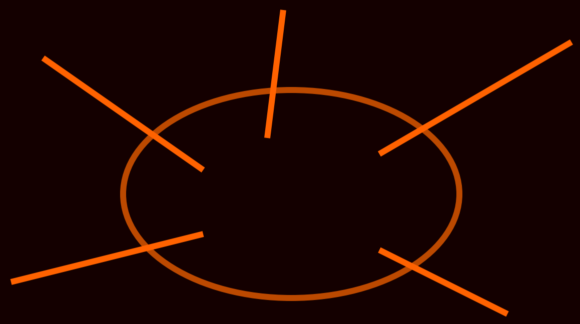
Procedure 1: At each step, visit a new node uniformly at random until target is found.

Theorem. Expected # of steps = $(n+1) / 2$.

Finding Target Randomly

Procedure 2: At each step, visit a new node that is a random neighbor of current node.

Theorem. Expected # of steps is a function of the *expansion* of the network.



Finding Target Randomly

Variations: walk towards neighbors with

- high degree
- high centrality
- least # of common neighbors

Homophily

Suppose people have observable characteristics and tend to befriend people who are similar to themselves.

- geography
- socio-economic status
- profession

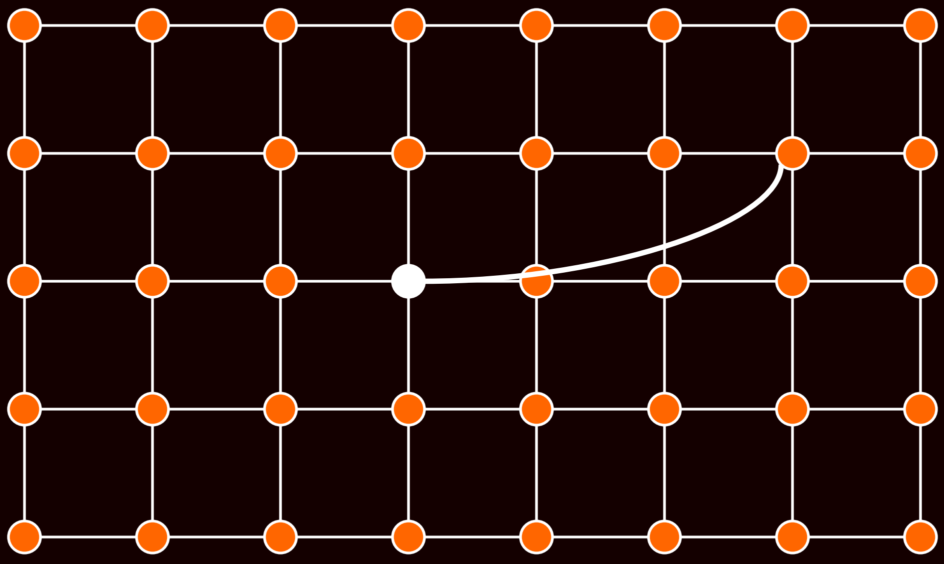
Networks with Homophily

Rewiring model (Watts-Strogatz)

- People have a predictable structure of local links reflecting homophily
- And a few random long-range links

Rewiring Model

1. Start with a grid (or other regular graph)
2. For each node, create one (or k in general) random long-range link



Rewiring Model

1. Exhibits small-world phenomenon (short paths exist)

2. Furthermore, people can find them with a decentralized algorithm for appropriate distribution [Kleinberg 2000]
   – Explains Milgrom experiment

Decentralized Search

Choose long-range links from distribution which favors close nodes

Tradeoff:

- \+ Gives navigational clues
- – Increases path length

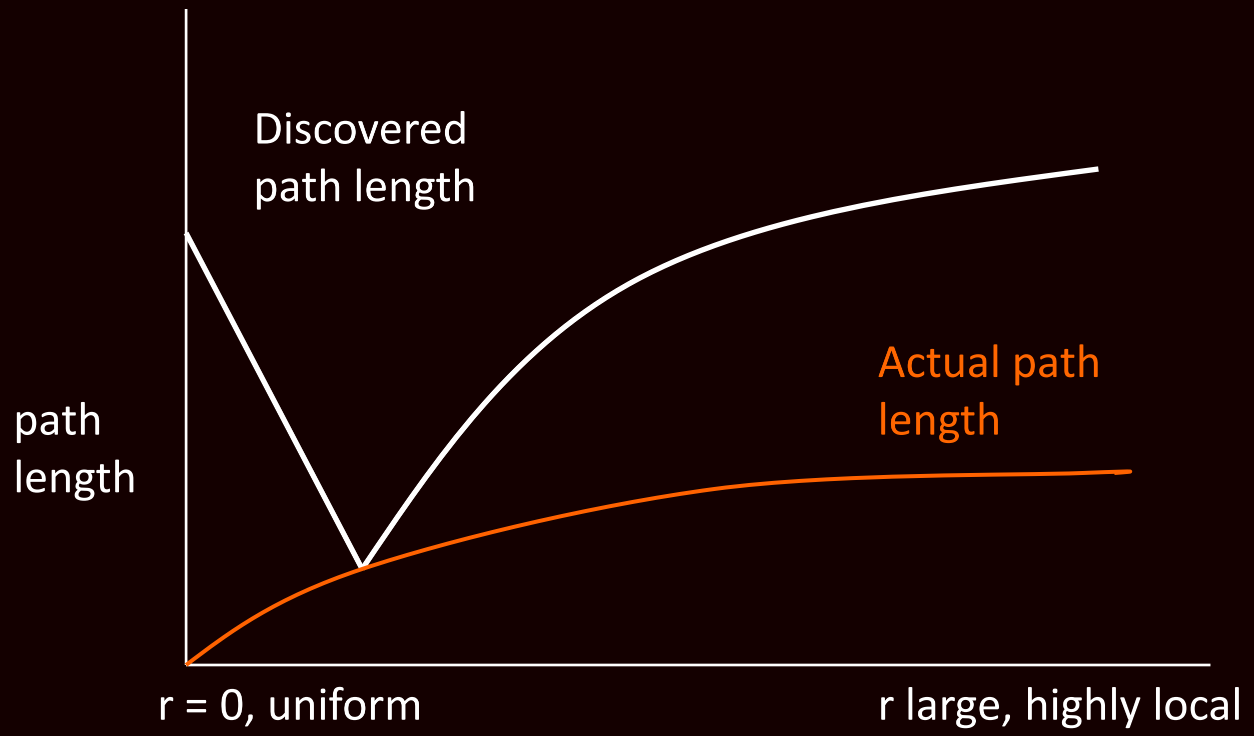
Result. There is a unique optimal distribution where decentralized search finds short paths.

Decentralized Search Model

- n x n grid
- d(u,v) = grid distance between u and v
- Each node u has directed edge to exactly one node v, it’s long-range contact

$$\Pr[\text{u connects to v}] = d(u, v)^{-r}$$

Tradeoff



Discovered
path length
Actual path
length
path
length
r = 0, uniform
r large, highly local

Decentralized Algorithm

Node s must send message m to node t.

At any moment, current message holder u must pass m to neighbor given:

- Set of local contacts of all nodes (grid structure)
- Location on grid of destination t
- Location and long-range contacts of nodes that have seen m

Delivery Time

Definition: The expected delivery time is expectation over choice of long-range contacts and uniformly random s and t of number of steps to deliver m.

Delivery Time

	$0 \leq r < 2$	$r = 2$	$r > 2$
Expected Delivery Time	$\Omega(n^{(2-r)/3})$	$O(\log^2 n)$	$\Omega(n^{(r-2)/(r-1)})$

Algorithm

In each step, current message holder u passes m to his or her neighbor v which is closest (in grid distance) to destination t.

Proof Sketch

- Define phases based on how close m is to t
  - Alg is in phase j if $2^j \leq d(m,t) < 2^{j+1}$
- Prove we don't spend too much time in any one phase
  - Exp time in phase j is c log n for all j
- Conclude since at most log n phases, expected delivery time is $O(\log^2 n)$
  - Follows from linearity of expectation

Proof

- Let $B_j = \{v : d(v,t) \leq 2^j\}$, i.e., the nodes outside phase j

- Then the probability we leave phase j is

$$|B_j|.\Pr[u\text{'s contact is in } B_j]$$

  – Compute prob. long-range contact of u is in B_j
  – Compute cardinality of B_j

Probability of long-range contact

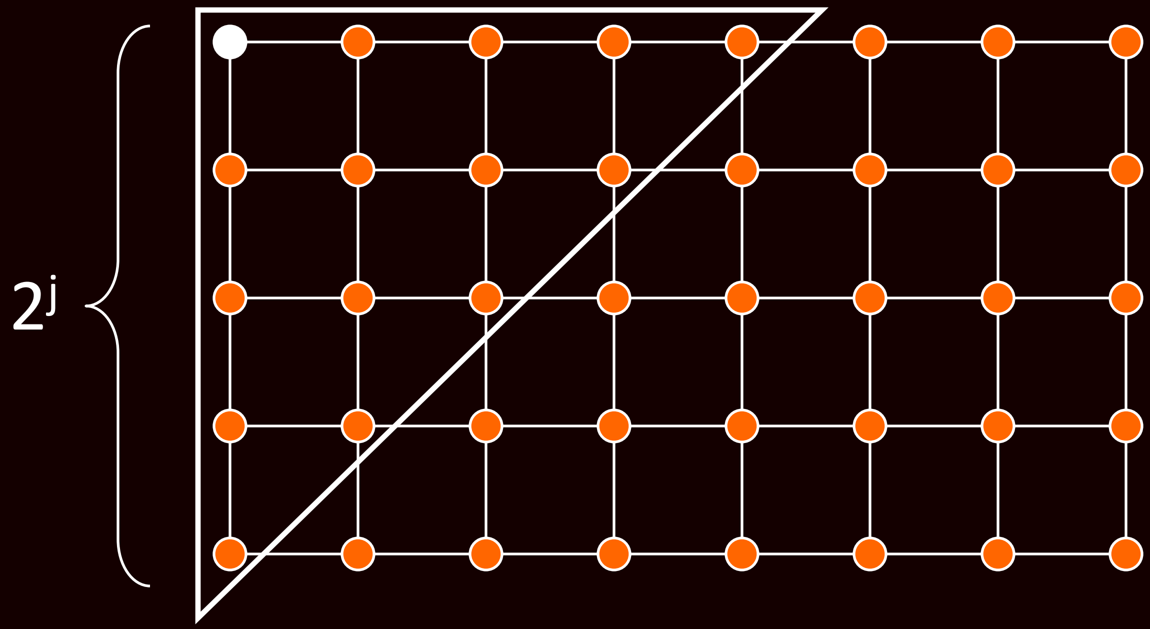
- Recall long-range contact of v is u with prob

$$\frac{d(u,v)^{-2}}{\sum_{v \neq u} d(u,v)^{-2}}$$

- Bound denominator
  - There are 4k nodes at distance k
  - Hence, $\sum_{v \neq u} d(u,v)^{-2} \leq \sum^{2n}_{k=1} (1/k^2)(4k) = O(\log n)$

Cardinality of B_j

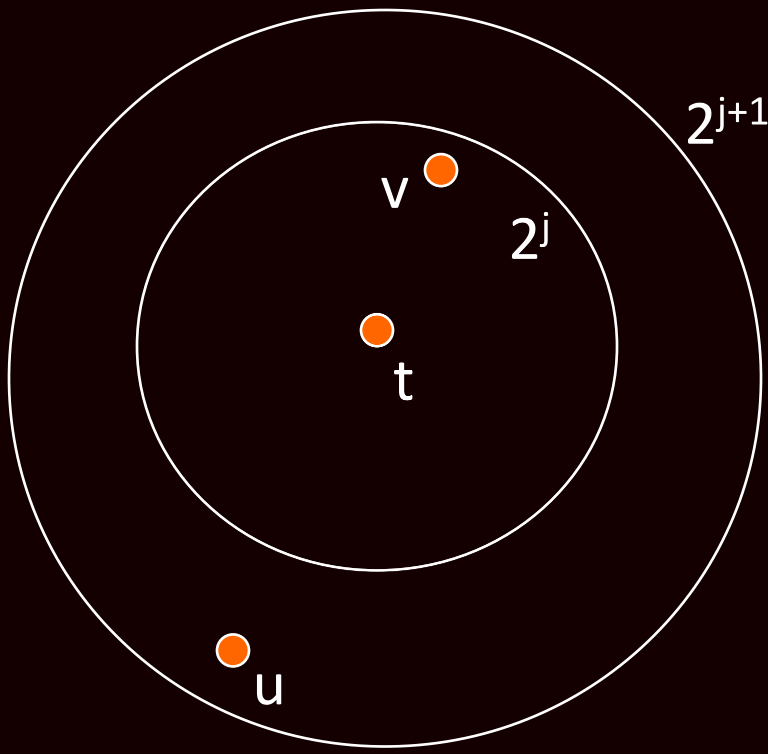
- Number of nodes at distance at most 2^j



- Hence $|B_j| \geq ½ (2^j)(2^j) = 2^{2j-1}$

Probability leave phase j

- Note d(u,v) for $v \in B_j$ is at most $2^j + 2^{j+1}$



Probability leave phase j

$Pr[\ .\ ] \quad = |B_j|.Pr[u\text{'s contact is in } B_j]$

$= |B_j|\ d(u,B_j)^{-2} / \sum_v d(u,v)^{-2}$

$\geq (2^{2j-1})(2^j + 2^{j+1})^{-2} / O(\log n)$

$\geq O(1/\log n)$

Expected # steps in phase j

Let X_j = # steps in phase j

$$E[X_j] = \sum_t \Pr[X_j \geq t] \leq \sum_t (1 - O(1/\log n))^{t-1} = O(\log n)$$

Since # phases is also O(log n), we see exp. delivery time is $O(\log^2 n)$ as claimed.

Other Diffusion Questions

Rumor spreading

- push model
- pull model
- intentional structures

Joint computations: dist. sensors computing

- average temperature in a field

Project Brainstorming

- Theoretical analysis of initial infection set for disease propagation
- Wikipedia study
- Music consumption, trends in modern technologies
- Online forums
- Mine social networks for degree distributions and other network parameters
- Hotelling model in higher dimensions
- Study disease prop in online games
- Economies in online games
- Finding poa/pos bounds in co-authorship networks
- Extending Jason's paper to single pricing, particular sequences, etc.

Assignment:

- Readings:
  - Social and Economic Networks, Chapter 7
  - Stoica, Morris, Liben-Nowell, Karger, Kaashoek, Dabek, Balakrishnan, *Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications*, IEEE/ACM Transactions on Networking, Vol. 11, No. 1, pp. 17-32, February 2003.
- Reaction to Chord paper
- Project proposals due 12/2/2009
- Presentation volunteer? Erik.