Diffusion and Cascading Behavior in Networks

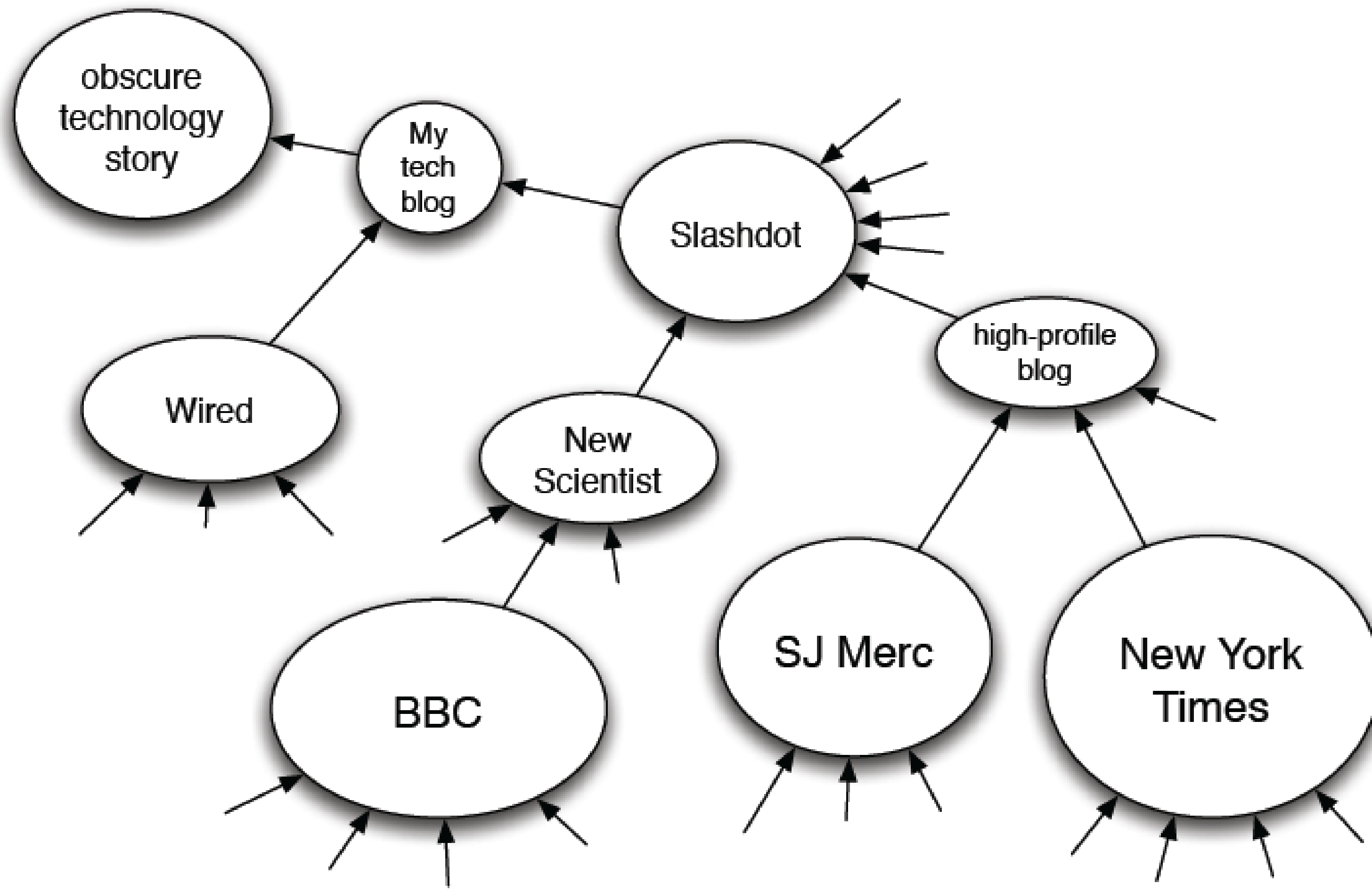
CS224W: Social and Information Network Analysis
Jure Leskovec, Stanford University
http://cs224w.stanford.edu

Spreading Through Networks

- **Spreading through networks:**
  - Cascading behavior
  - Diffusion of innovations
  - Network effects
  - Epidemics
- **Behaviors that cascade from node to node like an epidemic**

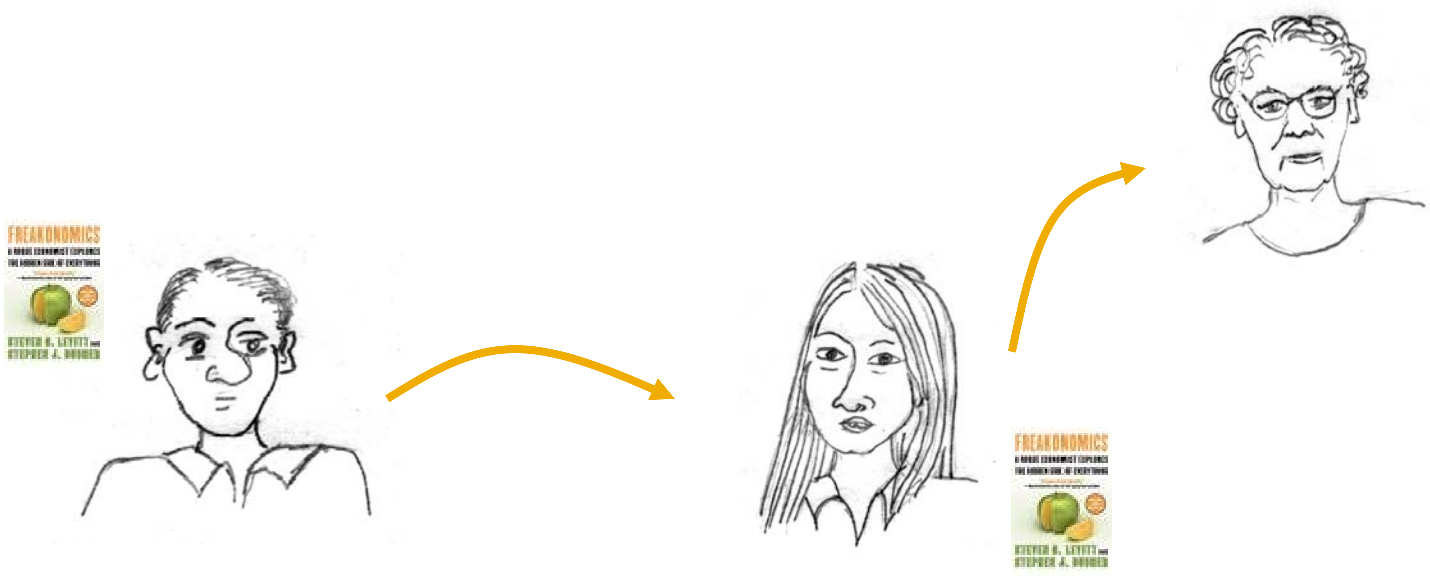
- **Examples:**
  - **Biological:**
    - Diseases via contagion
  - **Technological:**
    - Cascading failures
    - Spread of information
  - **Social:**
    - Rumors, news, new technology
    - Viral marketing

Information Diffusion

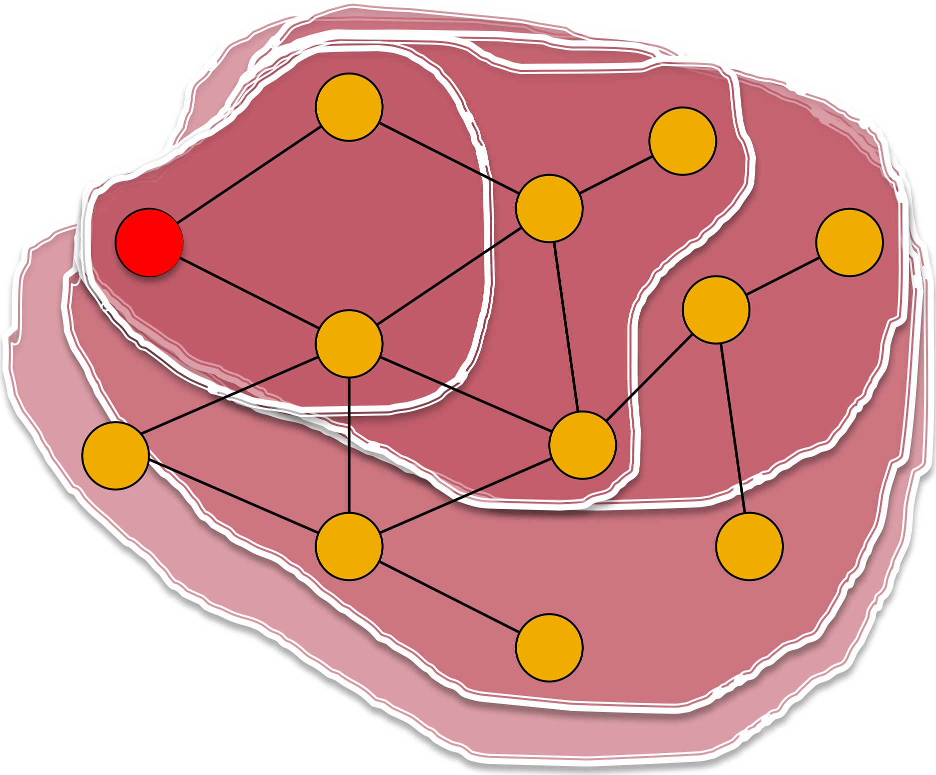


Diffusion in Viral Marketing

- **Product adoption:**
  - **Senders and followers of recommendations**

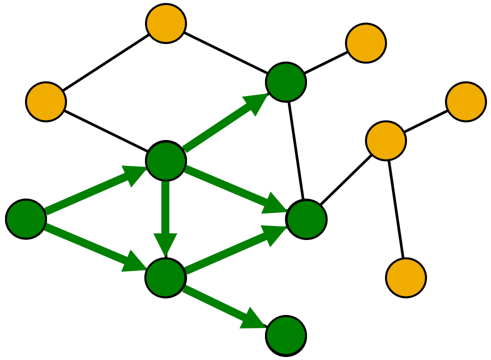


Spread of Diseases

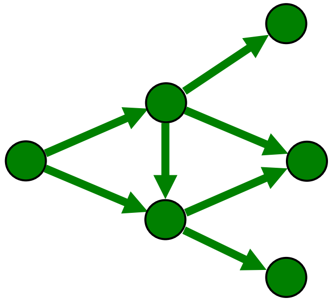


Network Cascades

- **Behavior/contagion spreads over the edges of the network**
- It creates a propagation tree, i.e., **cascade**



Network



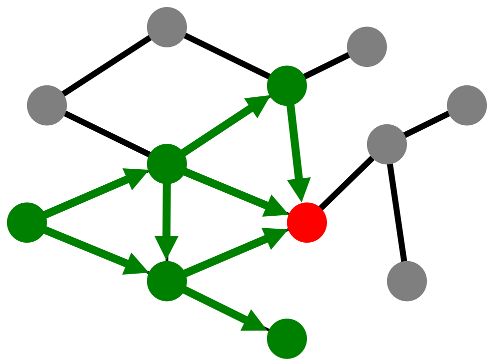
Cascade
(propagation graph)

Terminology:

- Stuff that spreads: Contagion
- “Infection” event: Adoption, infection, activation
- We have: Infected/active nodes, adoptors

How to Model Diffusion?

- **Probabilistic models:**
  - Models of influence or disease spreading
    - An infected node tries to "push" the contagion to an uninfected node
  - **Example:**
    - You "catch" a disease with some prob. from each active neighbor in the network



- **Decision based models:**
  - Models of product adoption, decision making
    - A node observes decisions of its neighbors and makes its own decision
  - **Example:**
    - You join demonstrations if *k* of your friends do so too

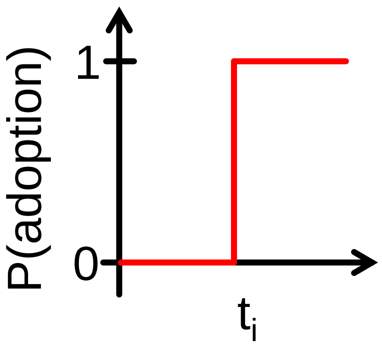
Decision Based Model of Diffusion

Decision Based Models

- **Collective Action** [Granovetter, '78]
  - Model where everyone sees everyone else's behavior
  - **Examples:**
    - Clapping or getting up and leaving in a theater
    - Keeping your money or not in a stock market
    - Neighborhoods in cities changing ethnic composition
    - Riots, protests, strikes

Collective Action: The Model

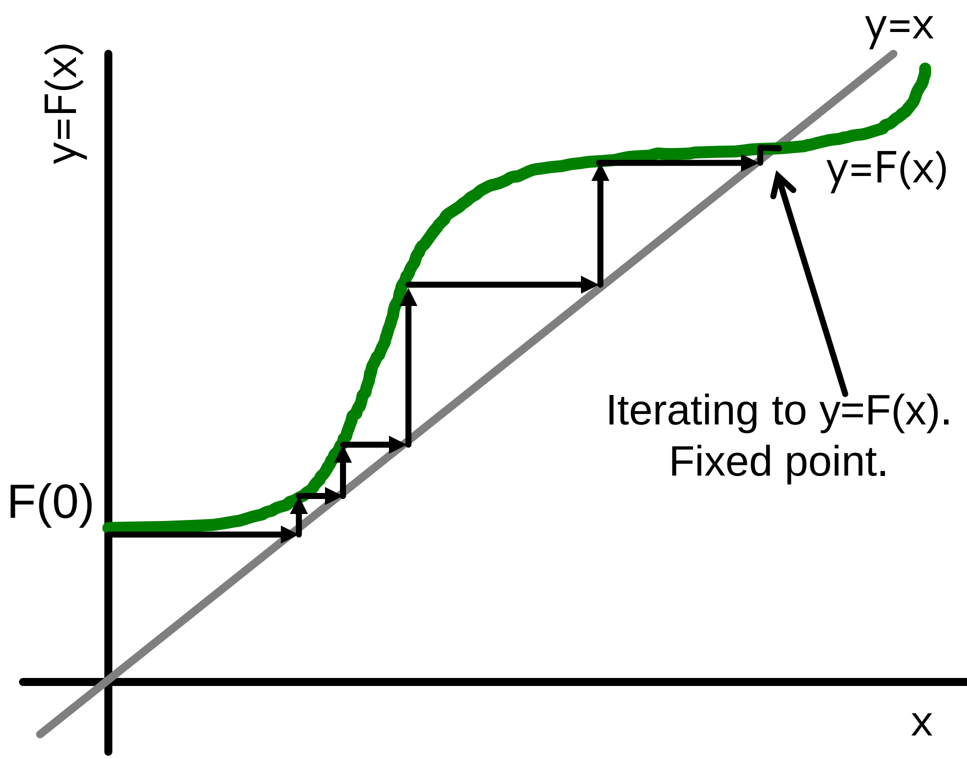
- **n people – everyone observes all actions**
- Each person *i* has a threshold t_i
  - Node *i* will adopt the behavior iff at least t_i other people are adopters:
    - **Small t_i**: early adopter
    - **Large t_i**: late adopter



- **The population is described by $\{t_1,\ldots,t_n\}$**
  - **F(x)** … fraction of people with threshold $t_i \leq x$

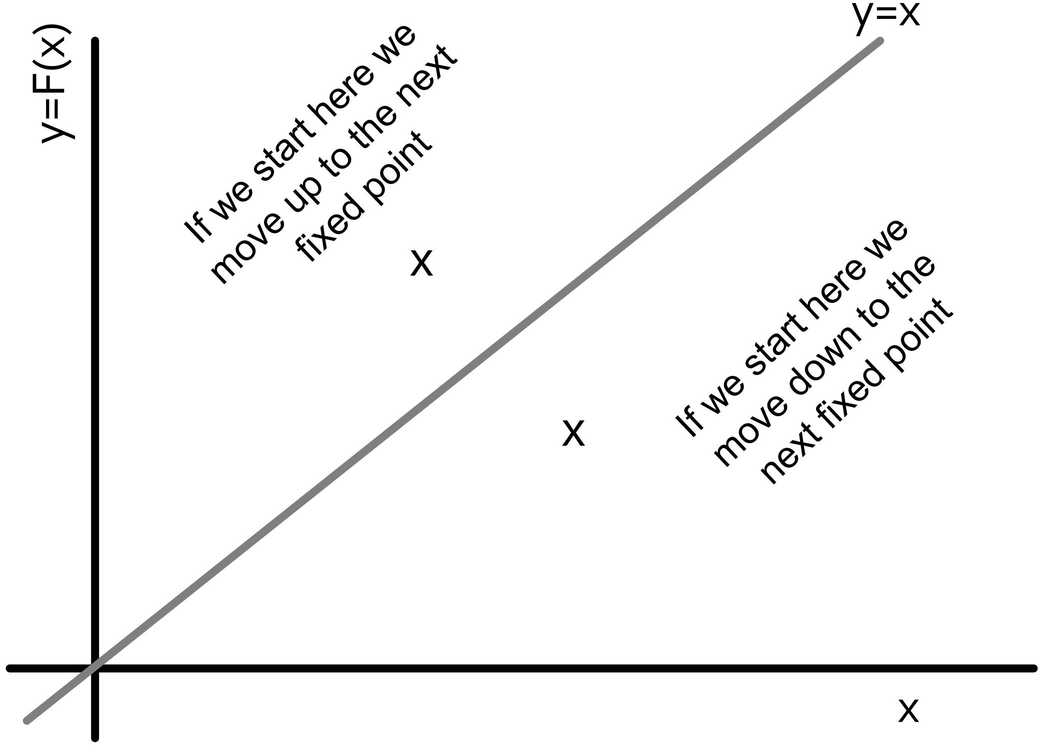
Collective action: Dynamics

- Think of the step-by-step change in number of people adopting the behavior:
  - **F(x)** … fraction of people with threshold $\leq$ x
  - **s(t)** … number of participants at time t
- Easy to simulate:
  - $s(0) = 0$
  - $s(1) = F(0)$
  - $s(2) = F(s(1)) = F(F(0))$
  - $s(t+1) = F(s(t)) = F^{t+1}(0)$
- Fixed point: F(x)=x
  - There could be other fixed points but starting from 0 we never reach them

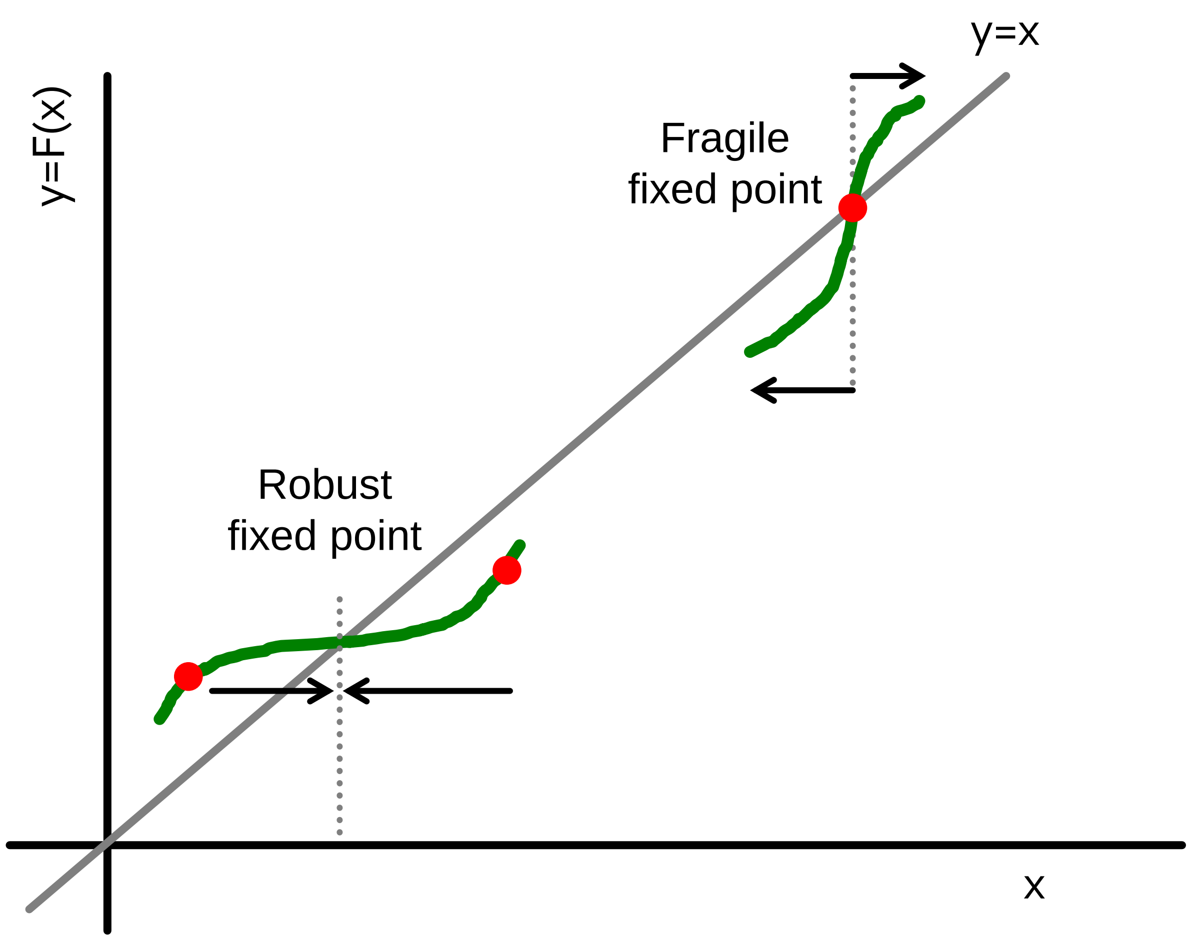


Starting Elsewhere

- **What if we start the process somewhere else?**
  - We move up/down to the next fixed point
  - How is market going to change?



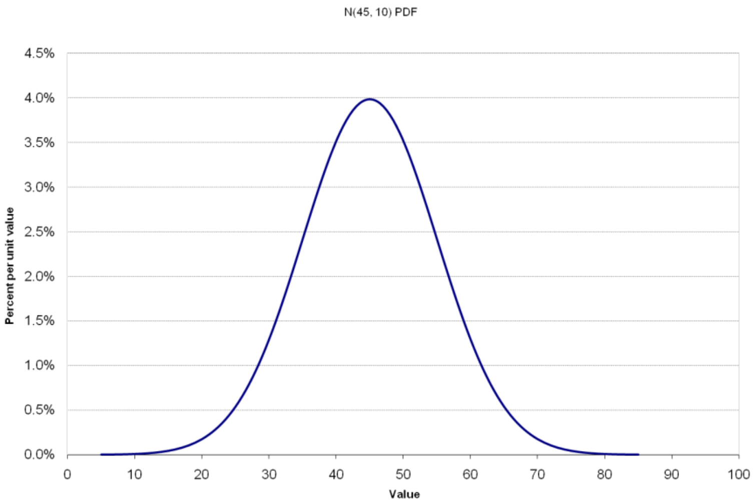
Fragile vs. Robust Fixed Point



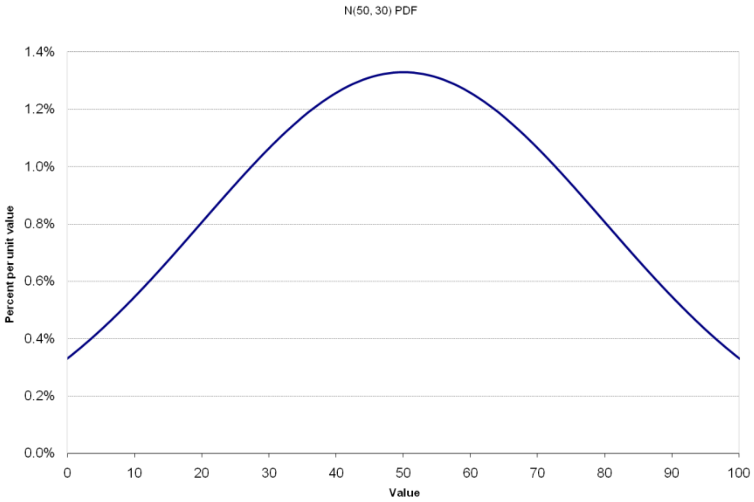
Discontinuous transition

- Each threshold t_i is drawn independently from some distribution $F(x) = Pr[thresh \leq x]$
  - **Suppose:** Normal with μ=n/2, variance σ

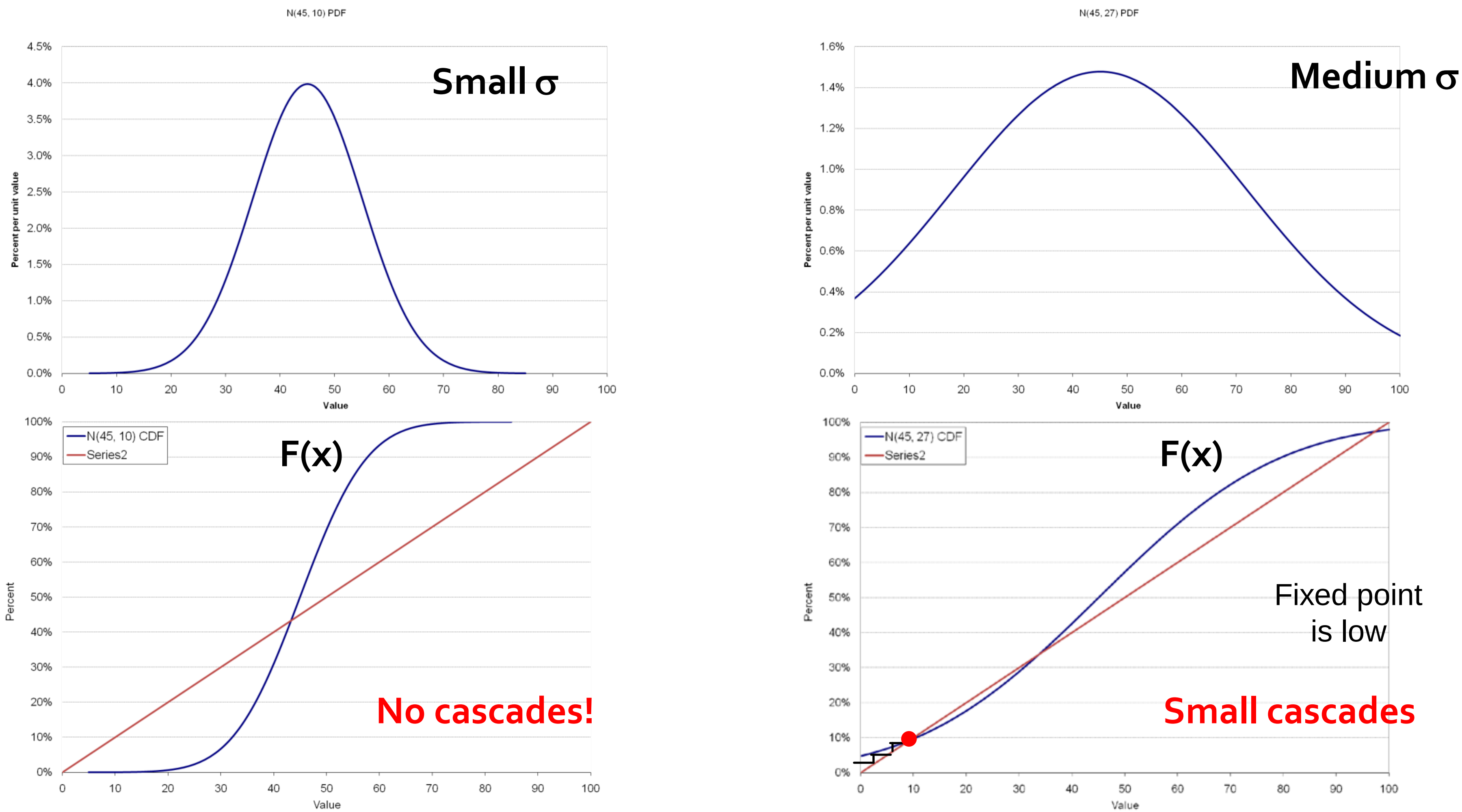
Small σ:



Large σ:

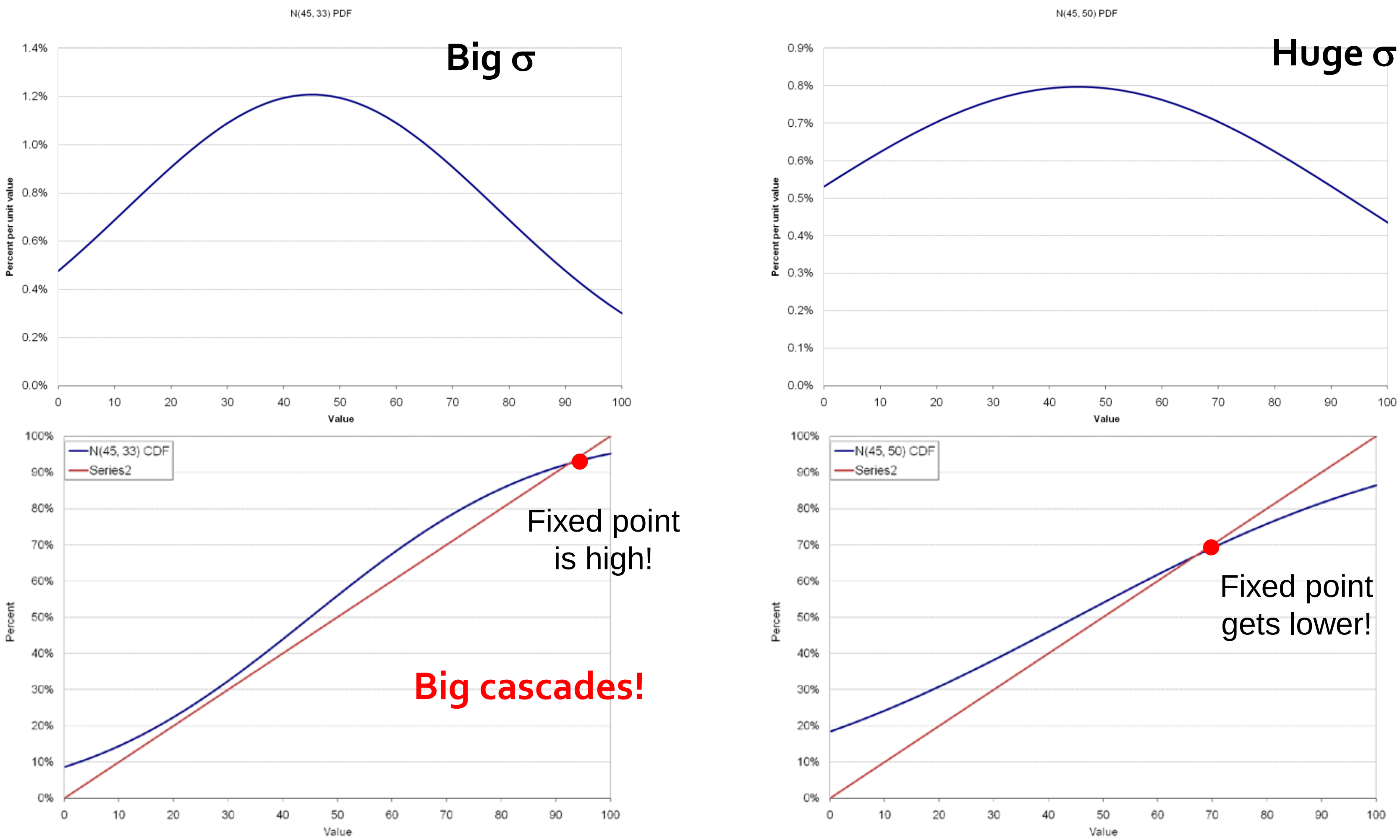


Discontinuous transition



Bigger variance let's you build a bridge from early adopters to mainstream

Discontinuous transition



But if we increase the variance even more we move the higher fixed point lover

Weaknesses of the model

- **It does not take into account:**
  - No notion of social network – more influential users
  - It matters who the early adopters are, not just how many
  - Models people's awareness of size of participation not just actual number of people participating
  - **Modeling thresholds**
    - Richer distributions
    - Deriving thresholds from more basic assumptions
      - game theoretic models

Weaknesses of the model

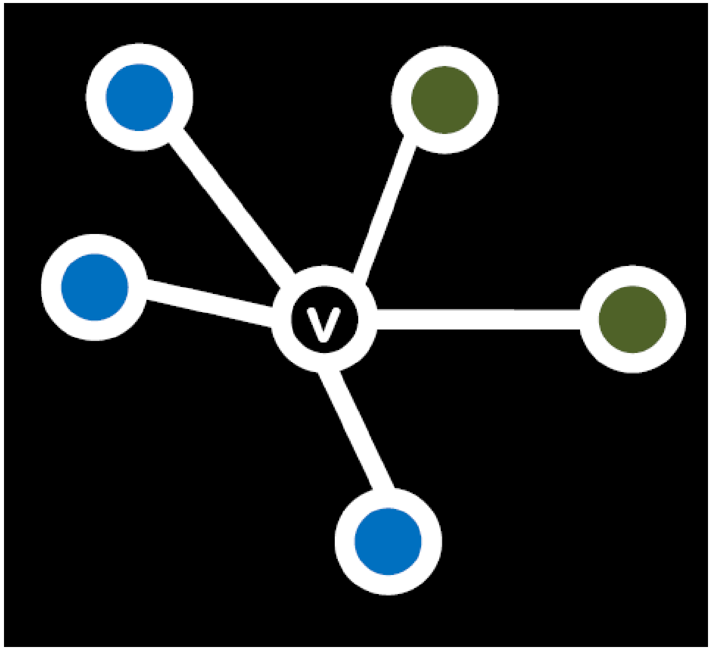
- It does not take into account:
  - Modeling perceptions of who is adopting the behavior/ who you believe is adopting
  - Non monotone behavior – dropping out if too many people adopt
  - Similarity – thresholds not based only on numbers
  - People get "locked in" to certain choice over a period of time
- **Network matters!** (next slide)

Game Theoretic Model of Cascades

[Morris 2000]

Game Theoretic Model of Cascades

- **Based on 2 player coordination game**
  - 2 players – each chooses technology A or B
  - Each person can only adopt **one** "behavior", **A** or **B**
  - You gain more payoff if your friend has adopted the **same** behavior as you

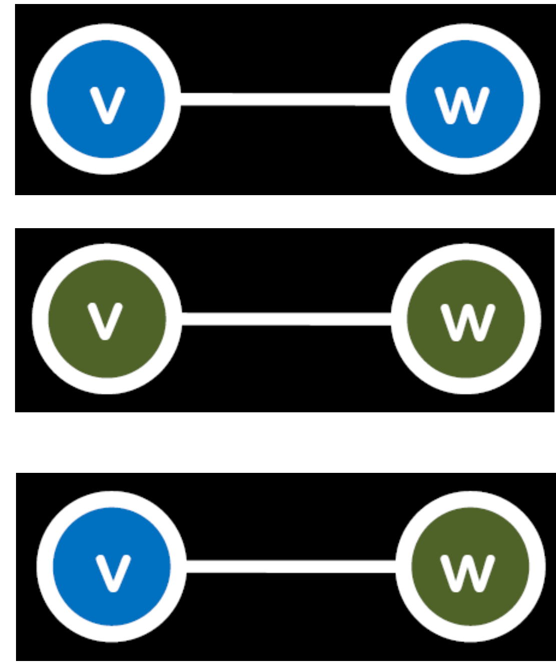


Local view of the network of node v

Example: BlueRay vs. HD DVD

The Model for Two Nodes

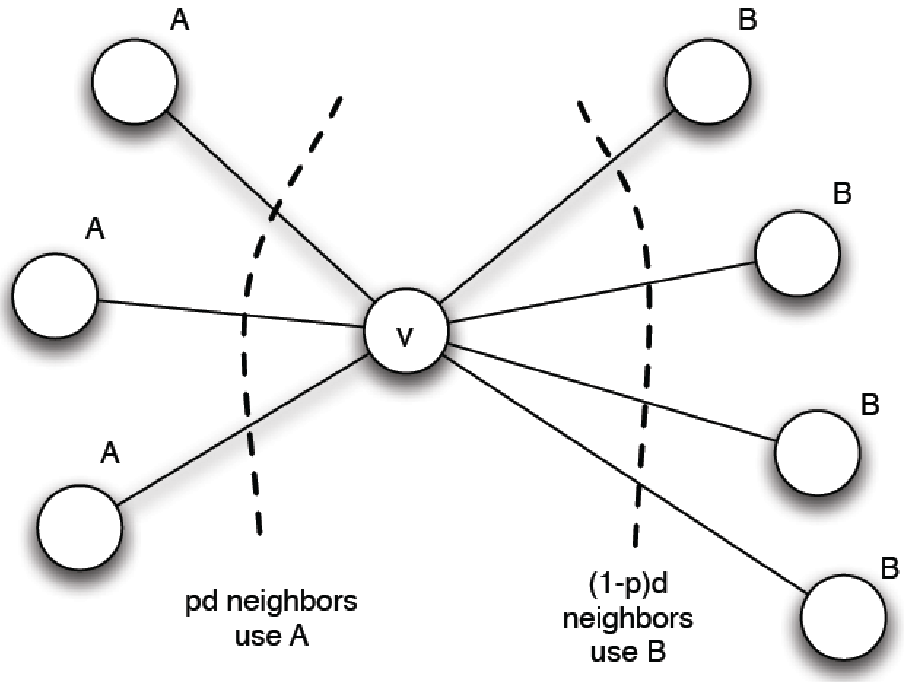
- ***Payoff matrix:***
  - If both v and w adopt behavior A, they each get payoff *a>0*
  - If *v* and *w* adopt behavior *B*, they reach get payoff *b>0*
  - If *v* and *w* adopt the opposite behaviors, they each get *0*



- **In some large network:**
  - Each node *v* is playing a copy of the game with each of its neighbors
  - **Payoff**: sum of node payoffs per game

v \ w	A	B
A	a, a	$0, 0$
B	$0, 0$	b, b

Calculation of Node v



Threshold:

v choses *A* if *p*>*q*

$$q = \frac{b}{a+b}$$

- Let v have *d* neighbors
- Assume fraction *p* of *v*'s neighbors adopt *A*
  - $Payoff_v = a \cdot p \cdot d$ if *v* chooses A
    $= b \cdot (1-p) \cdot d$ if *v* chooses B
- **Thus:** ***v* chooses A if:** $a \cdot p \cdot d > b \cdot (1-p) \cdot d$

Example Scenario

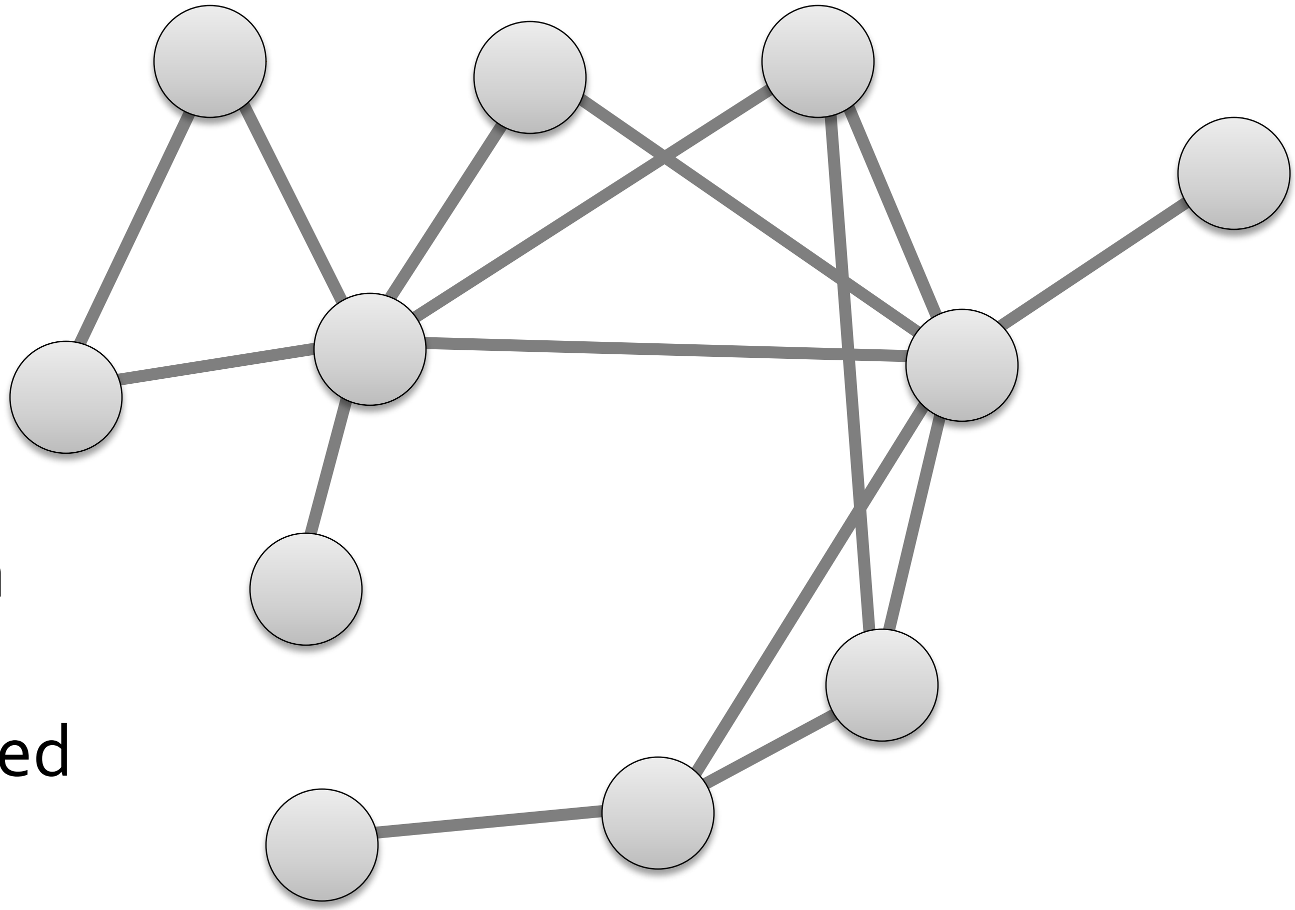
- **Scenario:**
  Graph where everyone starts with B.
  Small set S of early adopters of A
  - Hard wire S – they keep using A no matter what payoffs tell them to do

- Payoffs are set in such a way that nodes say:
  If at least 50% of my friends are red I'll be red

  (this means: $a = b + \varepsilon$)

Example Scenario

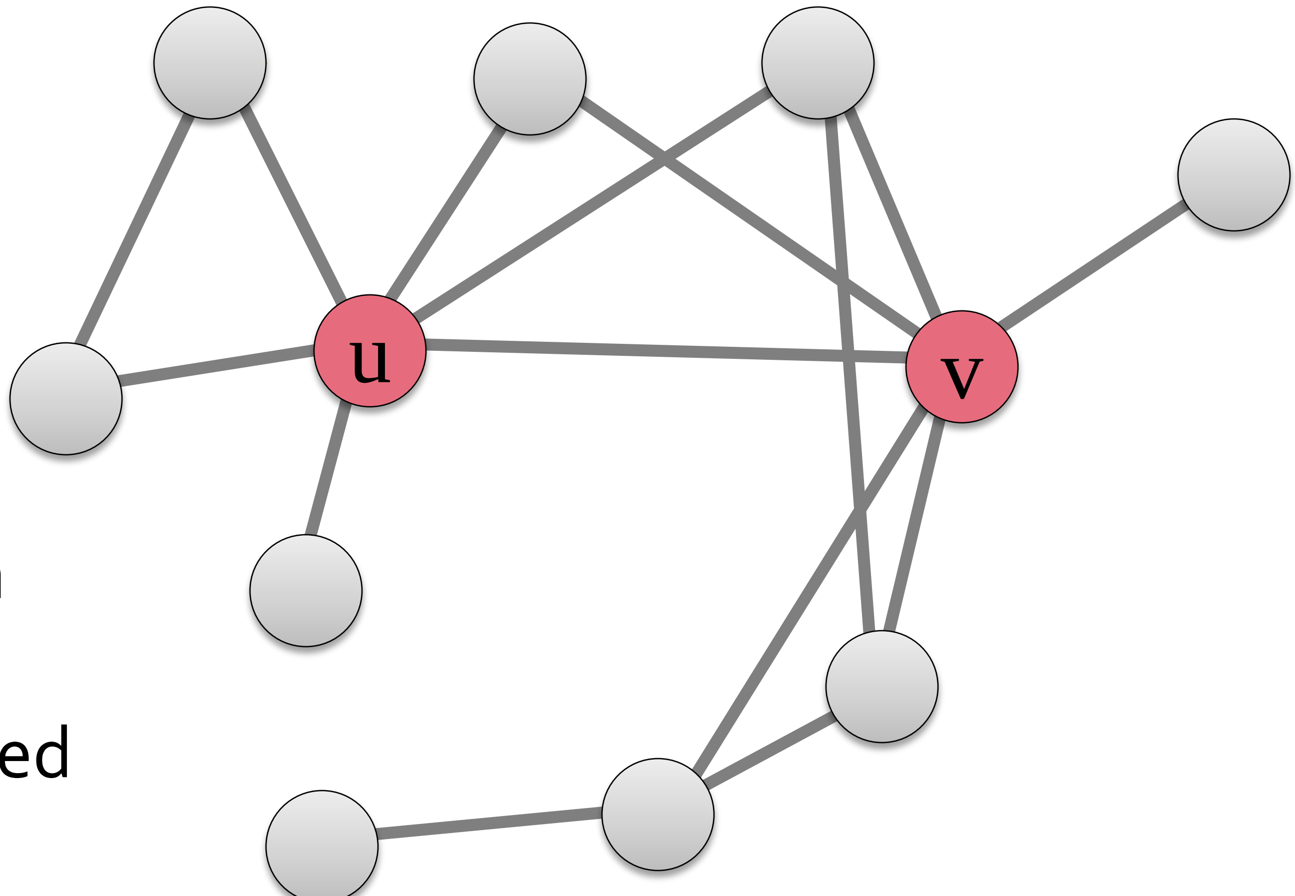
$S = \{u, v\}$



If more than 50% of my friends are red I'll be red

Example Scenario

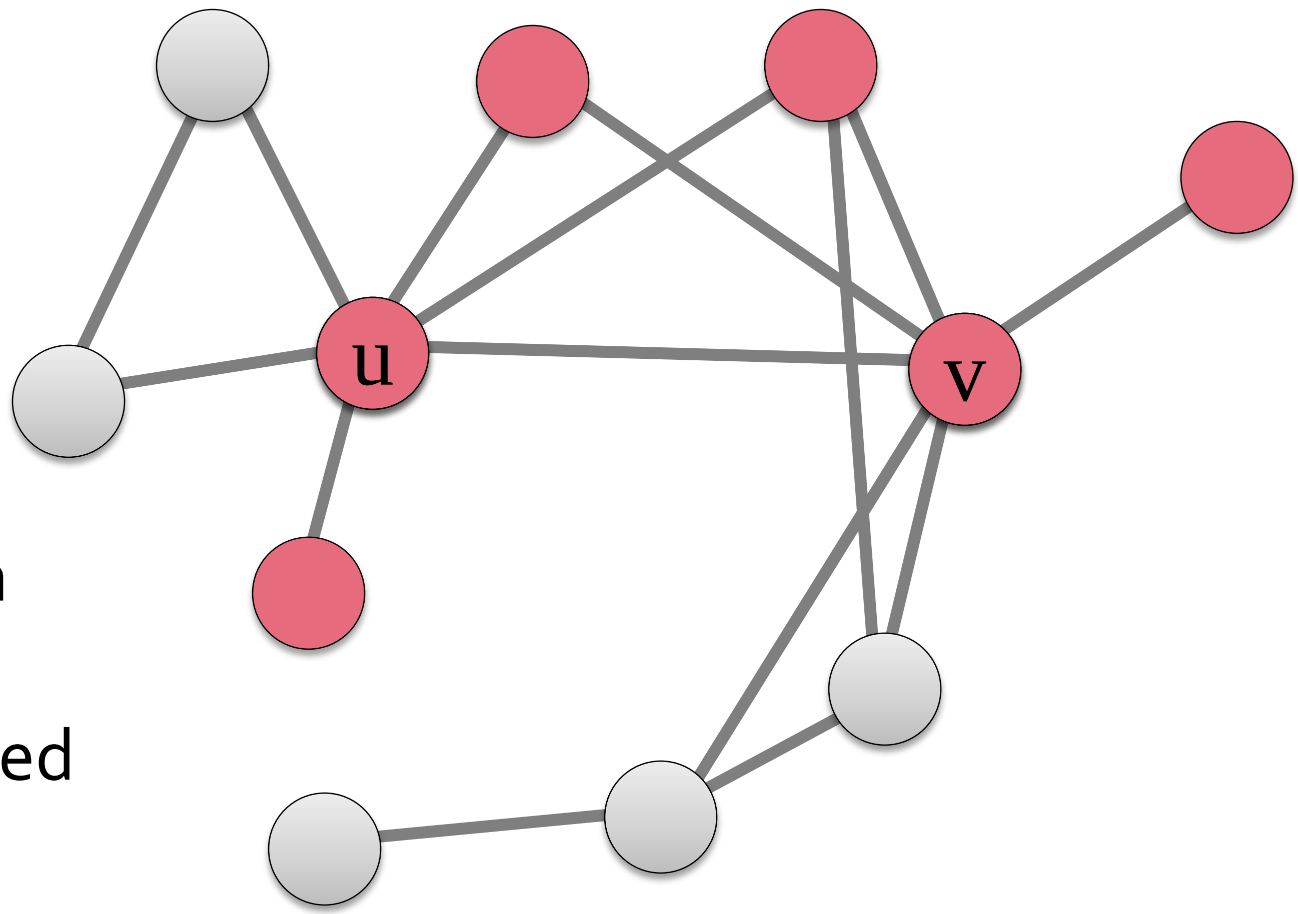
$S = \{u, v\}$



If more than 50% of my friends are red I'll be red

Example Scenario

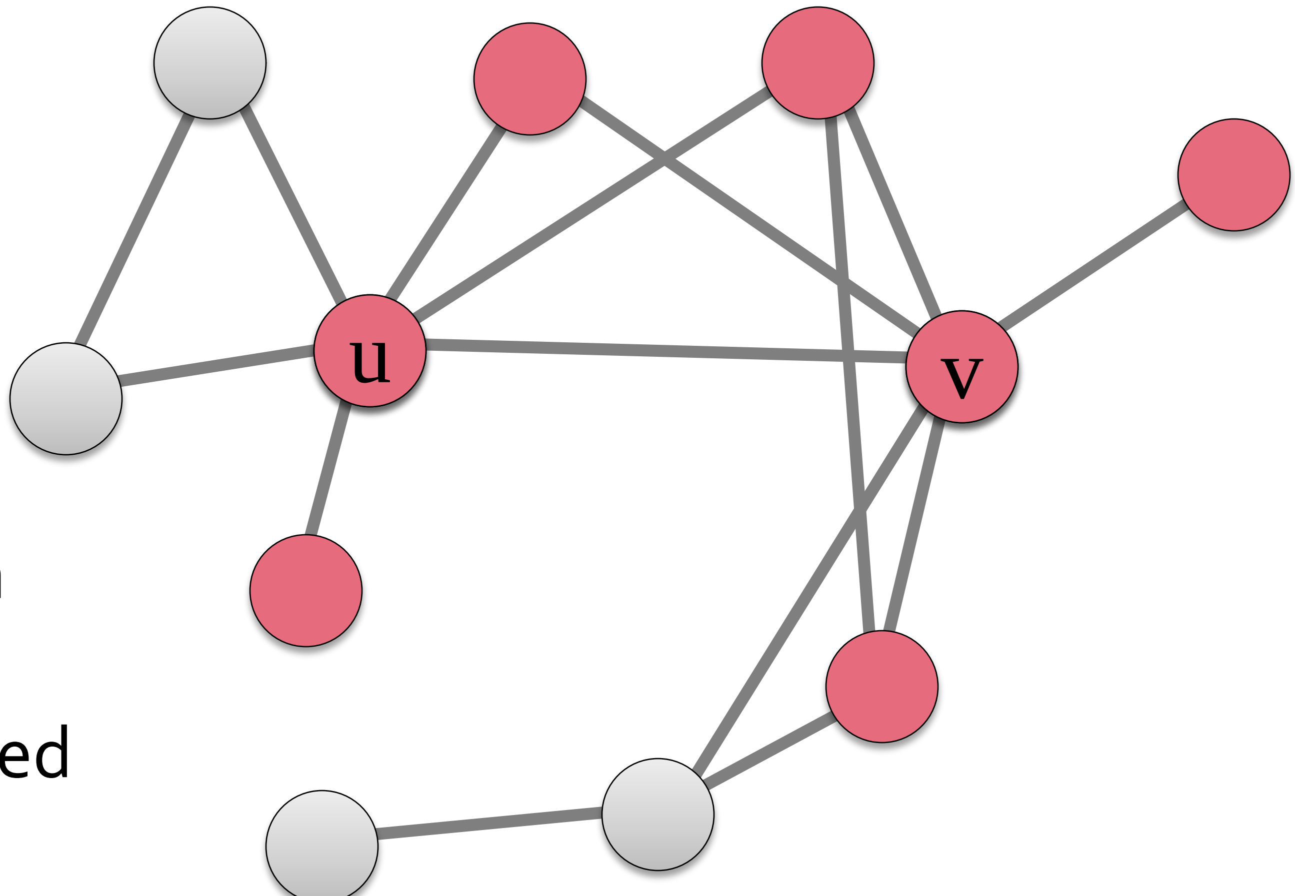
$S = \{u, v\}$



If more than 50% of my friends are red I'll be red

Example Scenario

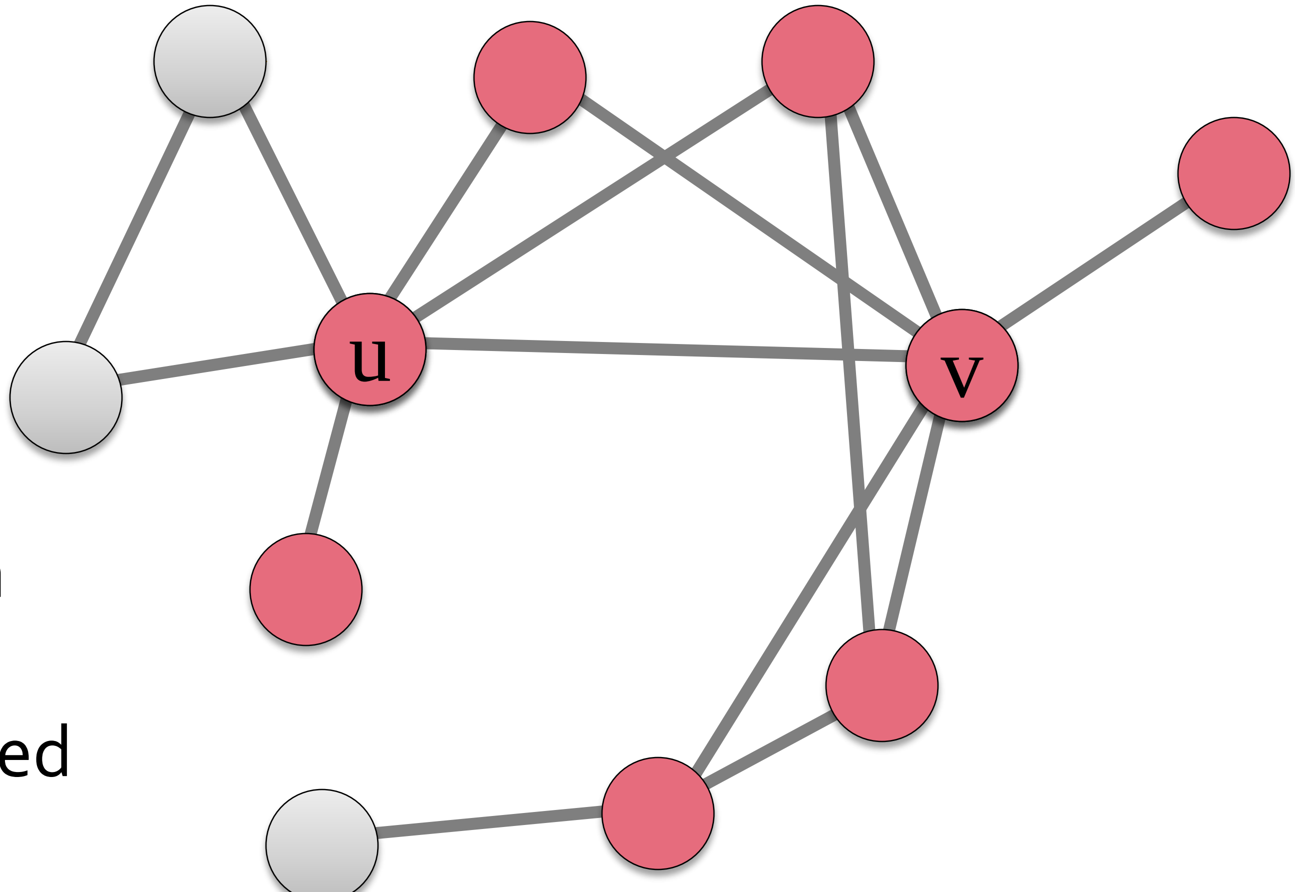
$$S = \{u, v\}$$



If more than 50% of my friends are red I'll be red

Example Scenario

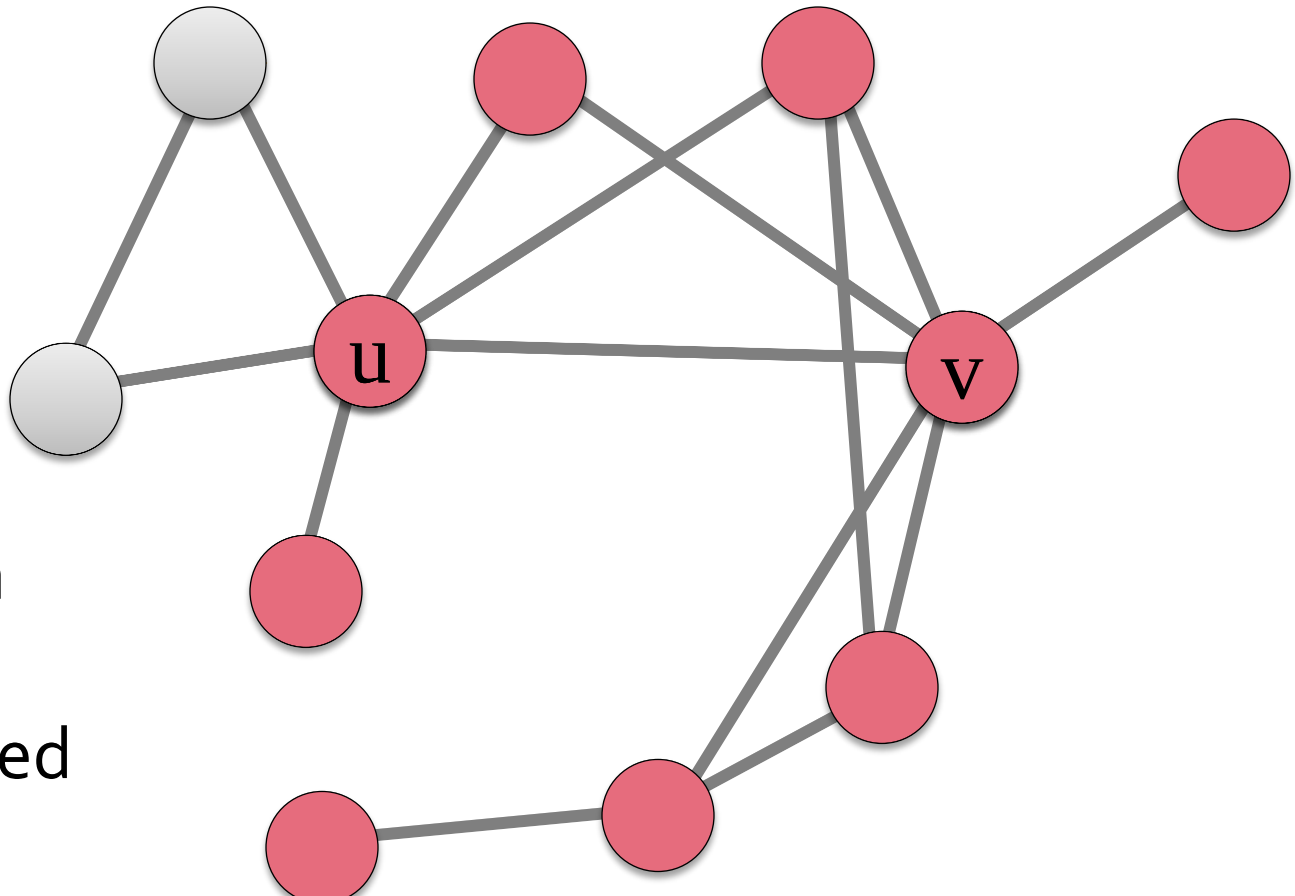
$S = \{u, v\}$



If more than 50% of my friends are red I'll be red

Example Scenario

$S = \{u, v\}$



If more than 50% of my friends are red I'll be red

Monotonic Spreading

- **Observation:**
  - **The use of A spreads monotonically** (Nodes only switch from B to A, but never back to B)
- **Why?** Proof sketch:
  - **Nodes keep switching from B to A:** $B \rightarrow A$
  - Now, suppose some node switched back from $A \rightarrow B$, consider the **first** node v to do so (say at time t)
  - Earlier at time t' ($t' < t$) the same node v switched $B \rightarrow A$
  - So at time t' v was above threshold for A
  - But up to time t no node switched back to B, so node v could only had more neighbors who used A at time t compared to t'. **There was no reason for *v* to switch.**

!! Contradiction !!

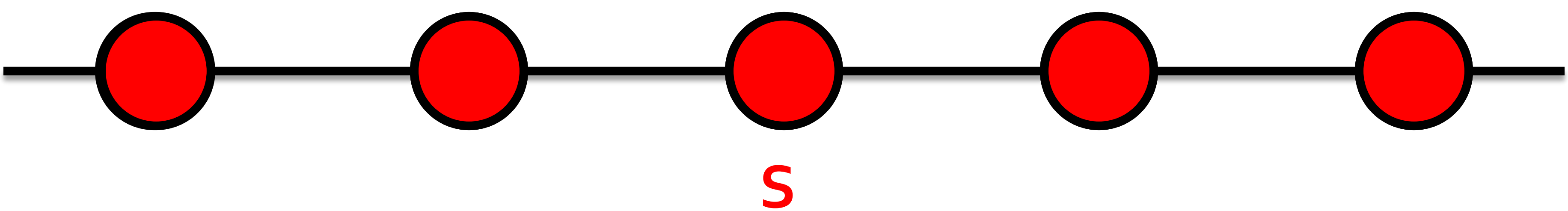
Infinite Graphs

v choses *A* if *p>q*

$$q = \frac{b}{a+b}$$

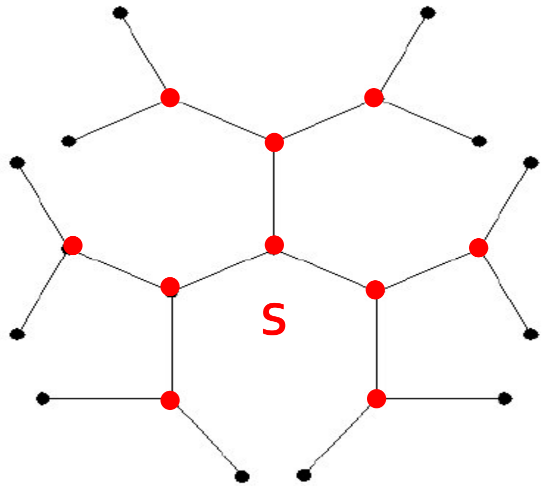
- Consider infinite graph *G*
  - (but each node has finite number of neighbors)
- We say that a finite set ***S*** causes a cascade in *G* with threshold *q* if, when *S* adopts *A*, eventually every node adopts *A*
- Example: **Path**

If q<1/2 then cascade occurs



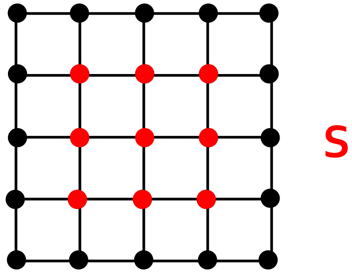
Infinite Graphs

- **Infinite Tree:**



If q<1/3 then cascade occurs

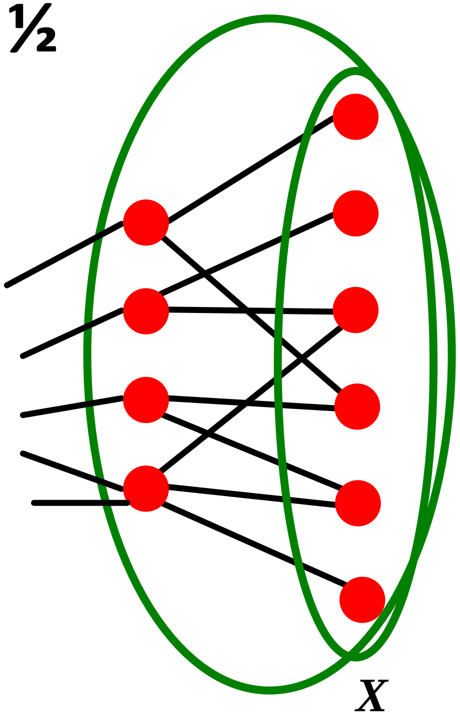
- **Infinite Grid:**



If q<1/4 then cascade occurs

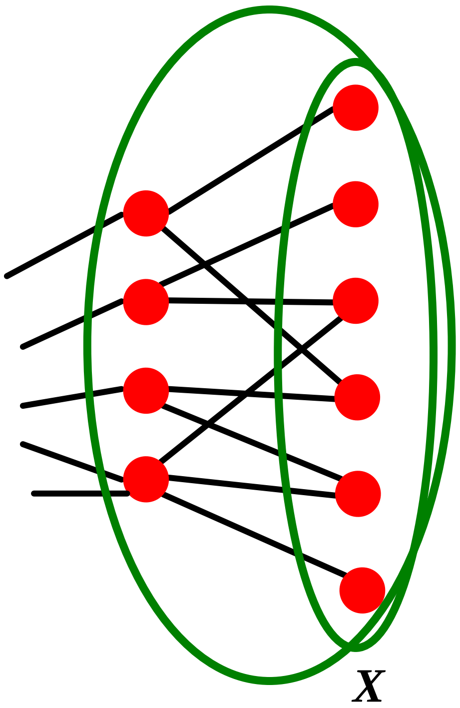
Cascade Capacity

- Def:
  - The **cascade capacity** of a graph *G* is the **largest *q*** for which some **finite set S** can cause a cascade
- Fact:
  - There is no *G* where cascade capacity **> ½**
- **Proof idea:**
  - Suppose such G exists: q>½, finite S causes cascade
  - **Show contradiction:** Argue that nodes stop switching after a finite # of steps



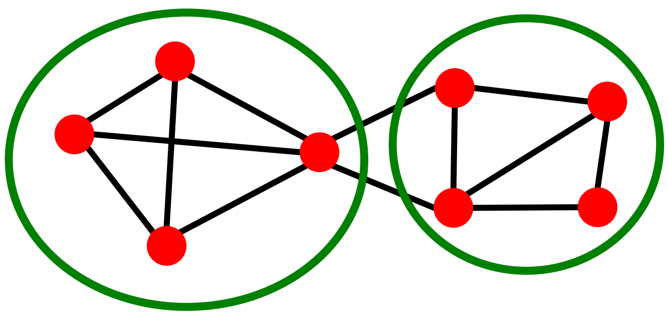
Cascade Capacity

- **Fact:** There is no *G* where cascade capacity **> ½**
- **Proof sketch:**
  - Suppose such G exists: q>½, finite S causes cascade
  - **Contradiction:** Switching stops after a finite # of steps
    - Define "potential energy"
    - Argue that it starts finite (non-negative) and strictly decreases at every step
  - "Energy": = $|d^{out}(X)|$
      - $|d^{out}(X)|$:= # of outgoing edges of active set X
    - The only nodes that switch have a strict majority of its neighbors in *S*
    - $|d^{out}(X)|$ strictly decreases
    - It can do so only a finite number of steps



Stopping Cascades

- **What prevents cascades from spreading?**
- Def: **Cluster of density *ρ*** is a **set of nodes *C*** where each node in the set has at least ***ρ*** fraction of edges in *C*.



$\rho=3/5$ $\rho=2/3$

Stopping Cascades

- Let S be an initial set of adopters of A
- All nodes apply threshold *q* to decide whether to switch to A
- **Two facts:**
  - 1) If $G\backslash S$ contains a cluster of density $>(1-q)$ then *S* can not cause a cascade
  - 2) If S fails to create a cascade, then there is a cluster of density >(1-q) in G\S

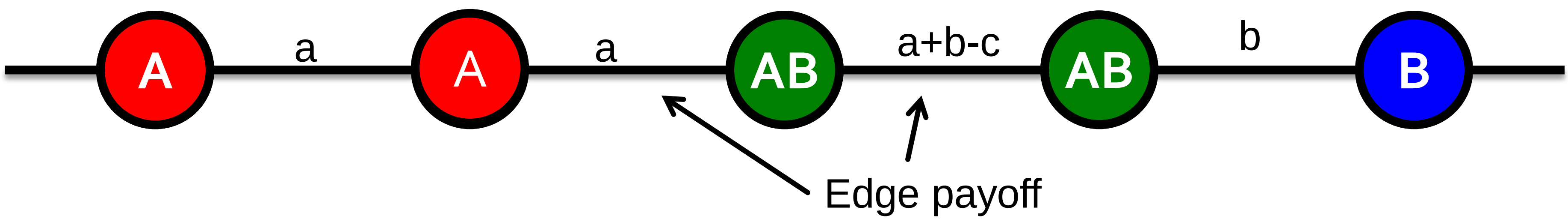
Extending the model: Allow people to adopt A and B

Cascades & Compatibility

- **So far:**
  - Behaviors *A* and *B* compete
  - Can only get utility from neighbors of same behavior: *A-A* get a, *B-B* get *b*, *A-B* get *0*
- **Let's add extra strategy "*A-B*"**
  - *AB-A*: gets *a*
  - *AB-B*: gets *b*
  - *AB-AB*: gets max(*a*, *b*)
  - Also: Some cost *c* for the effort of maintaining both strategies (summed over all interactions)

Cascades & Compatibility: Model

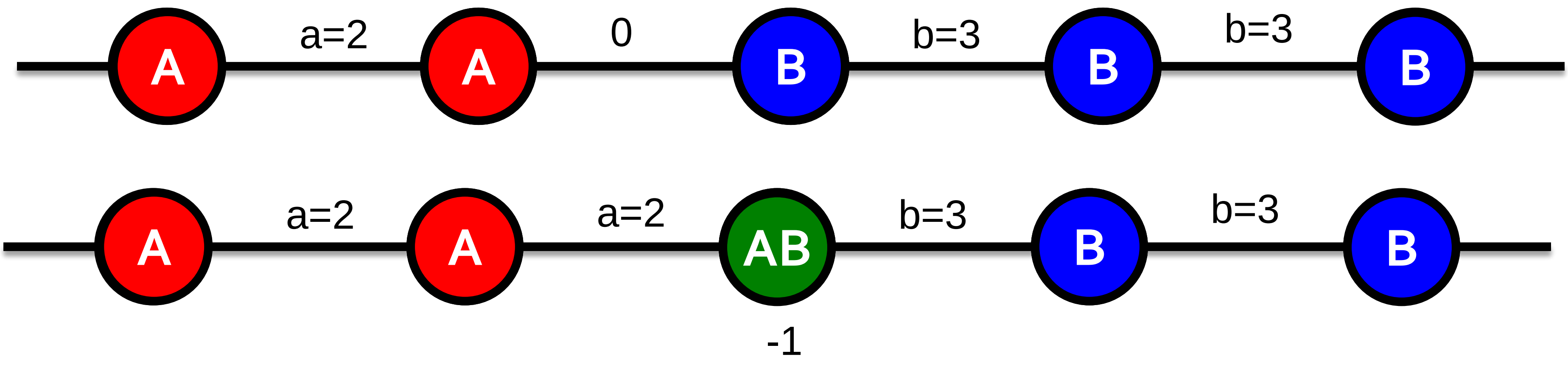
- Every node in an infinite network starts with *B*
- Then a finite set *S* initially adopts *A*
- Run the model for *t=1,2,3,…*
  - Each node selects behavior that will optimize payoff (given what its neighbors did in at time *t-1*)



- **How will nodes switch from *B* to *A* or *AB*?**

Example

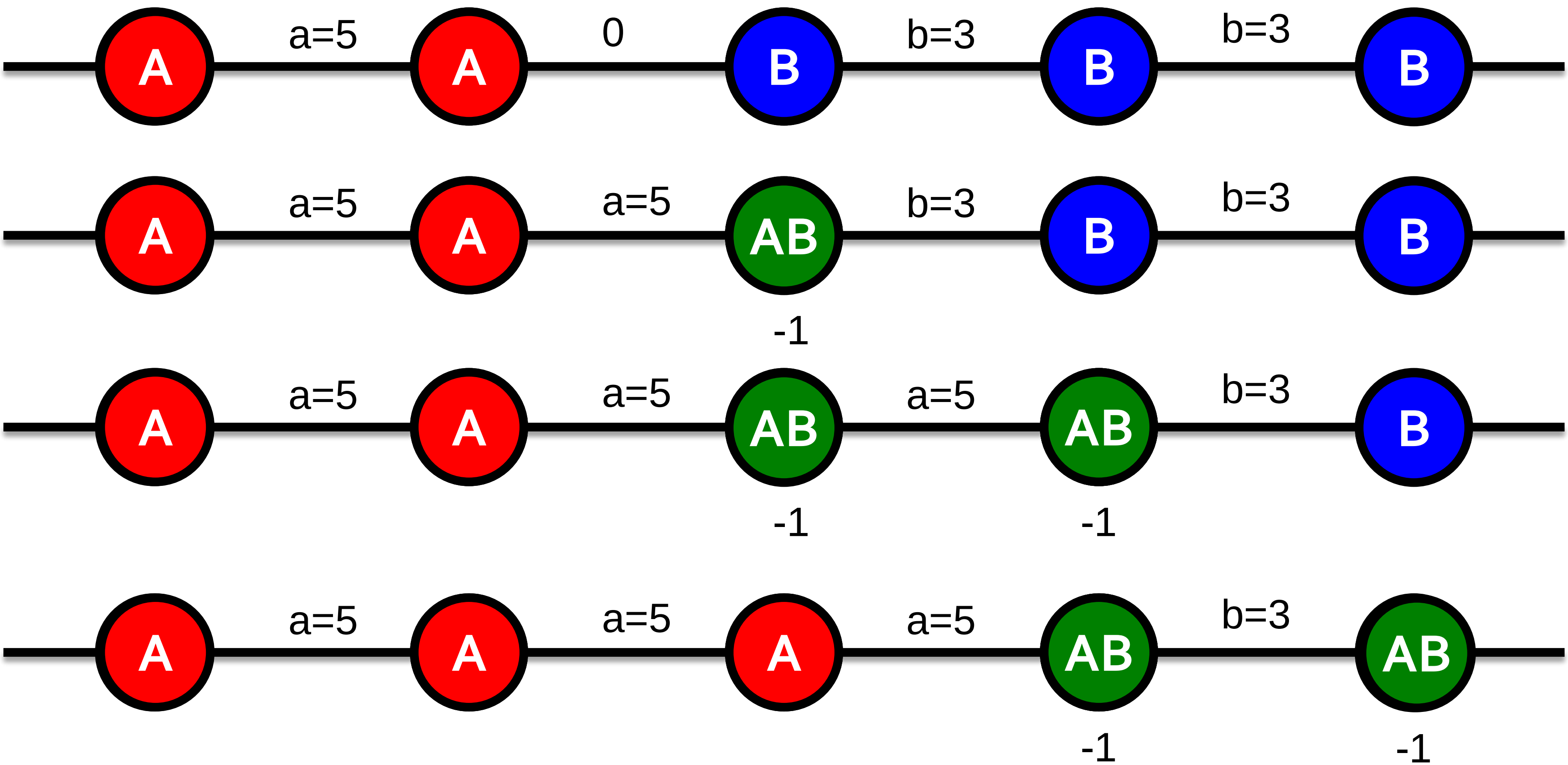
- **Path:** Start with all Bs, **a>b** (A is better)
- **One node switches to A – what happens?**
  - **With just A, B:** A spreads if $b \leq a$
  - **With A, B, AB:** **Does A spread?**
- Assume a=2, b=3, c=1



Cascade stops

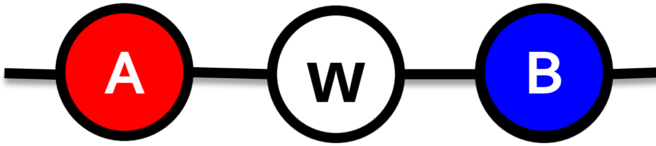
Example

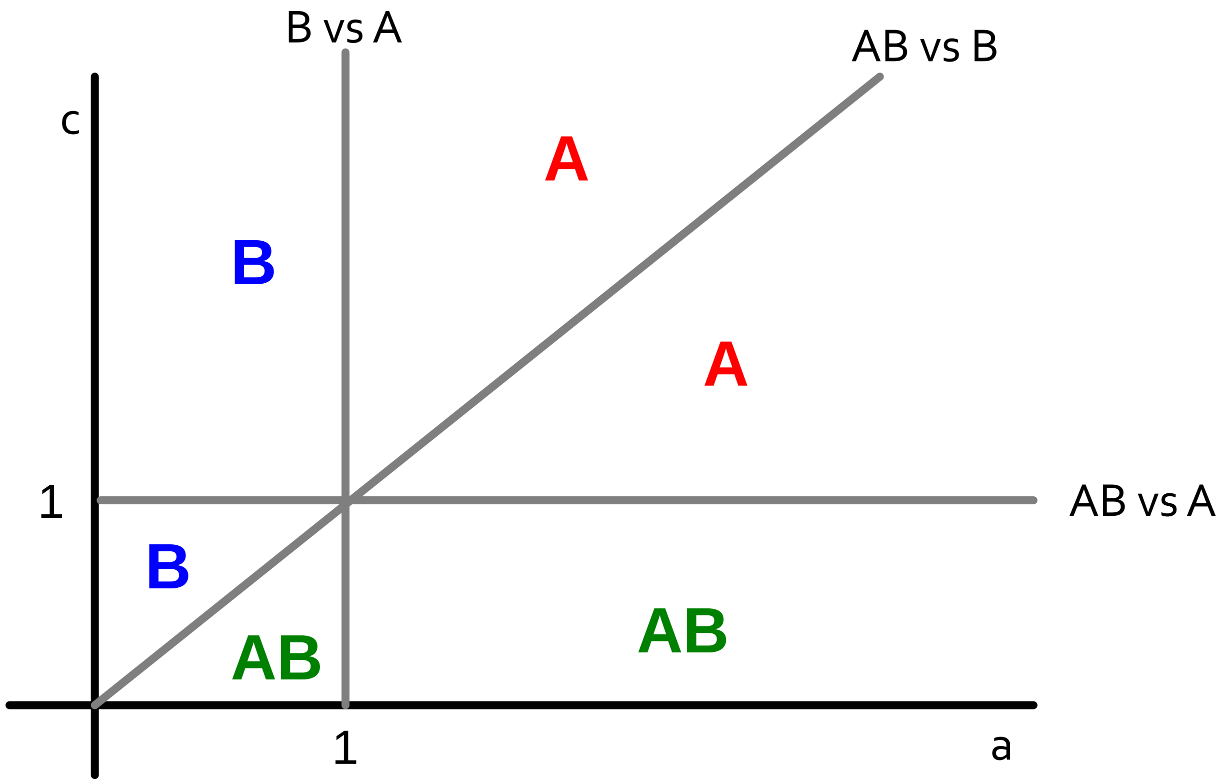
- Let a=5, b=3, c=1



For what pairs (c,a) does A spread?

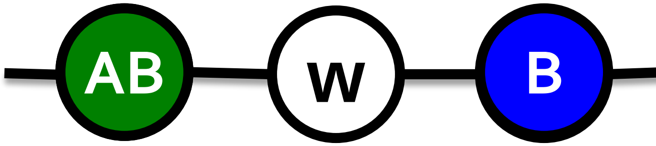
- **Infinite path, start with all Bs**
- **Payoffs:** A:a, B:1, AB:a+1-c
- What does node w in **A-w-B** do?

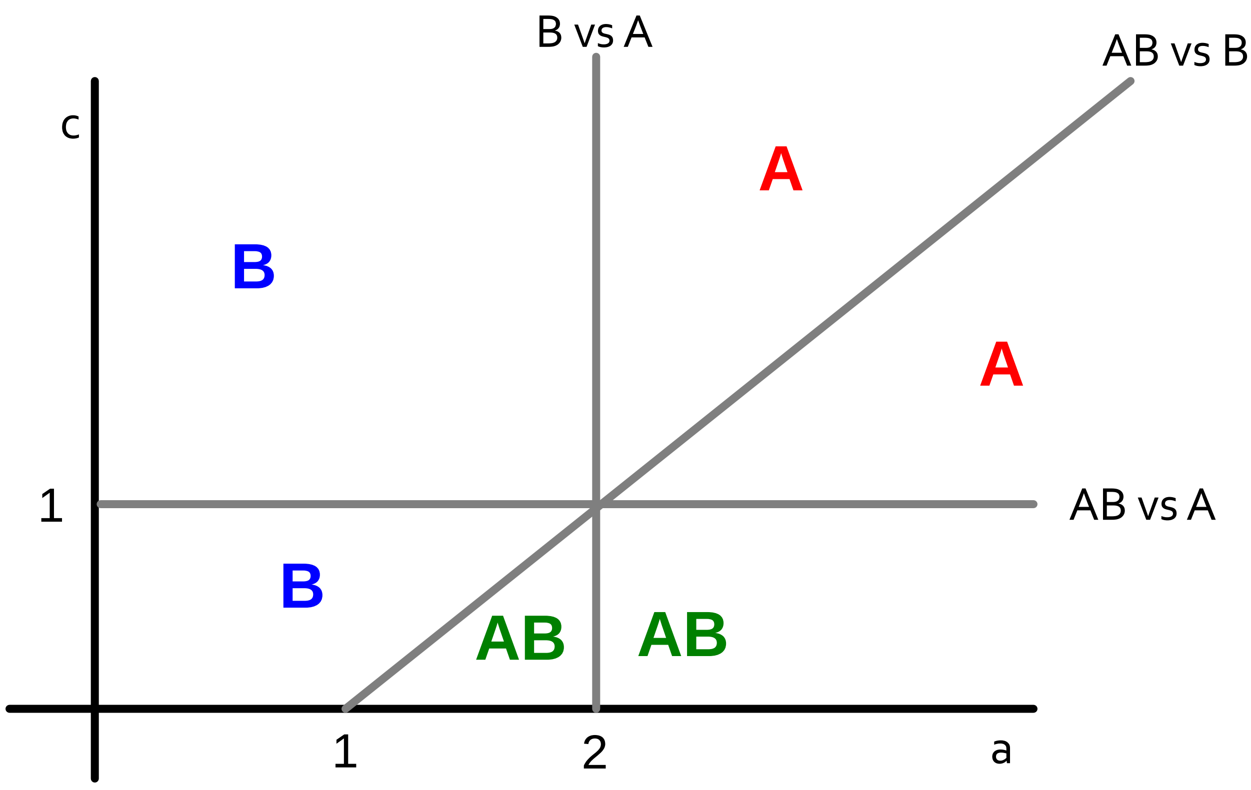




For what pairs (c,a) does A spread?

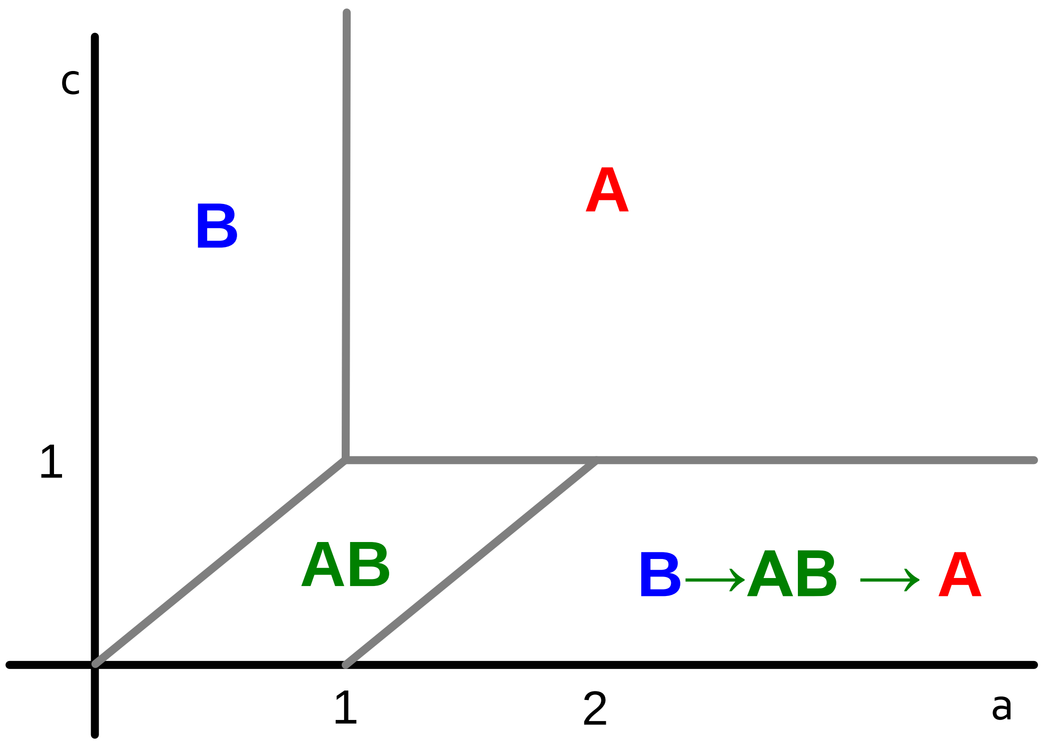
- **Payoffs:** A:a, B:2, AB:a+2-c
- **Notice: now also AB spreads**
- What does node w in **AB-w-B** do?





For what pairs (c,a) does A spread?

- Joining the two pictures:



Lesson

- You manufacture default B and new/better A comes along:
  - Infiltration: If you make B too compatible then people will take on both and then drop the worse one (B)
  - Direct conquest: If A makes itself not compatible – people on the border must choose. They pick the better one (A)
  - Buffer zone: If you choose an optimal level then you keep a static "buffer" between A and B