COMP4641 Lab8

Review of Lectures 8-11: Cascading, Contagion, Influence and Signed Net

Diffusion Models

To model diffusion: **decision based models** (**cascading**, a node makes decisions based on observations of its neighbor's decisions) / **probabilistic models** (**contagion**, a node gets infection by its infected neighbors in a probabilistic manner)

Cascading

- payoff matrix:

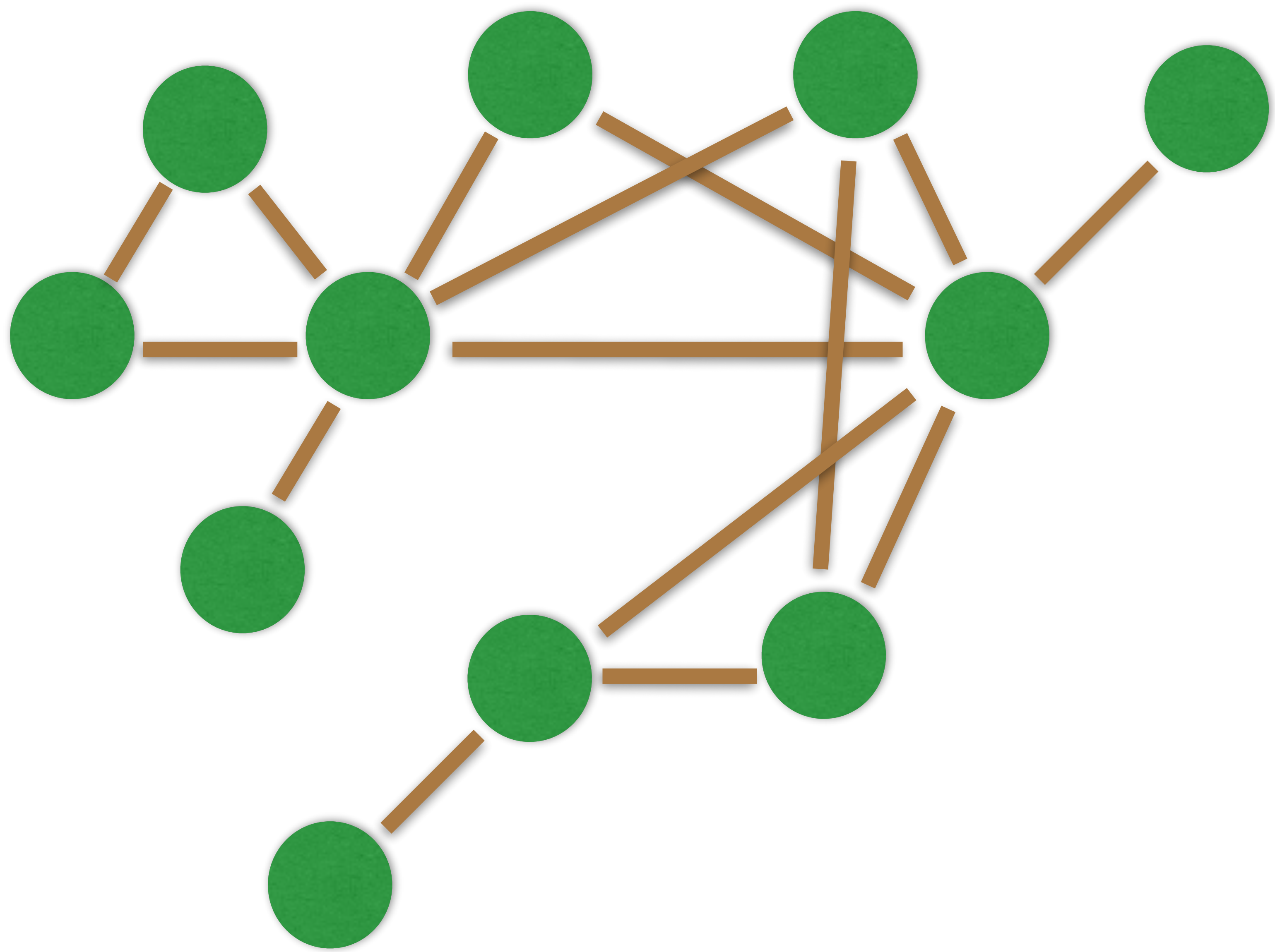
payoff	A	B
A	a	0
B	0	b

object of decision making is to maximize payoff

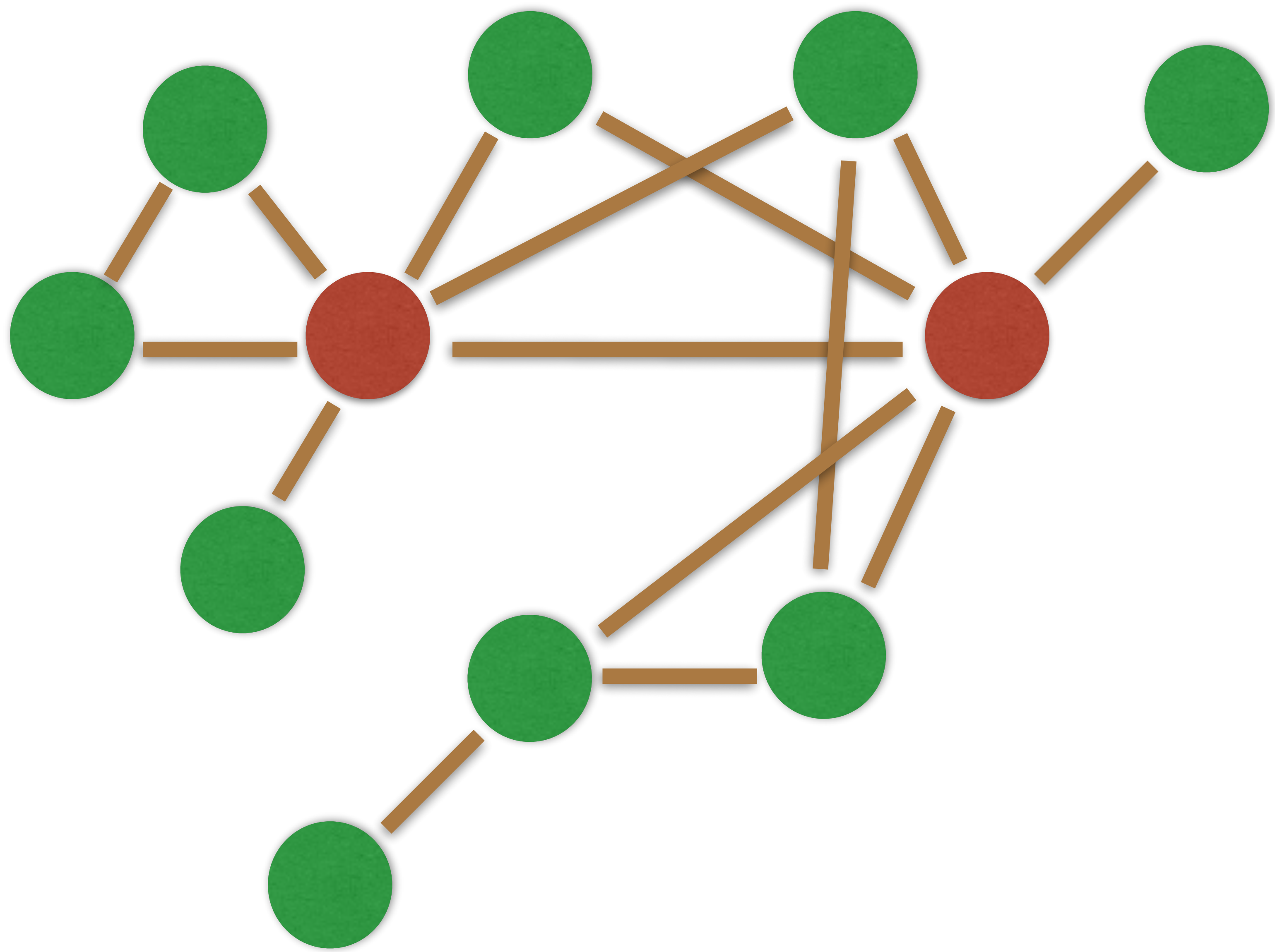
- threshold: for a node which has *d* neighbors, proportion *p* of them are *A* nodes, proportion *q=1-p* of them are *B* nodes, the payoff it has if choosing to be *A* node is $a \cdot p \cdot d$, the payoff it has if choosing to be *B* node is $b \cdot q \cdot d$, suppose the graph starts with all *B* nodes, then at a certain moment of cascading, the node will switch to be *A* node if

$$a \cdot p \cdot d > b \cdot q \cdot d \Rightarrow p > \frac{b}{a+b}$$

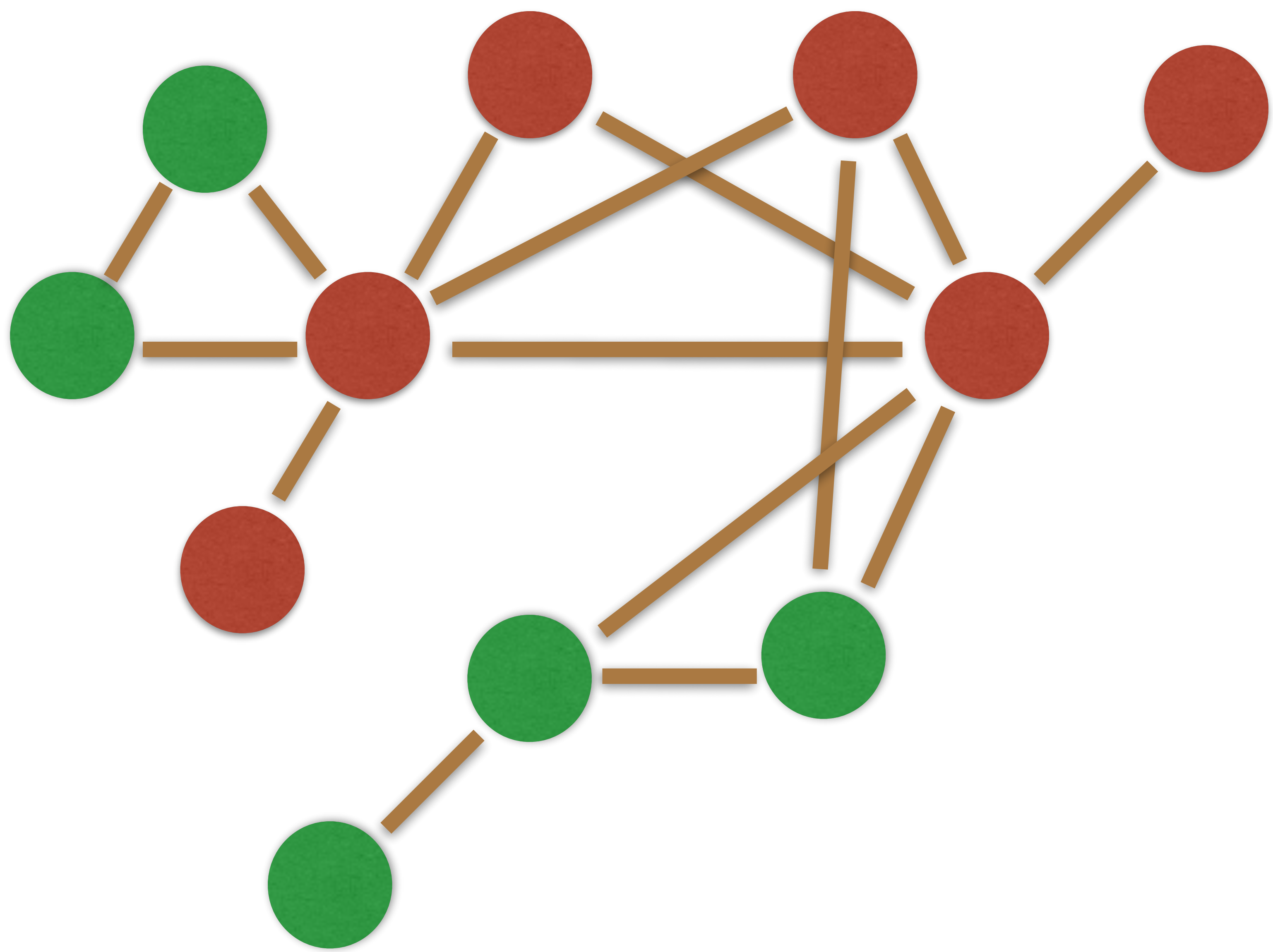
Cascading Example



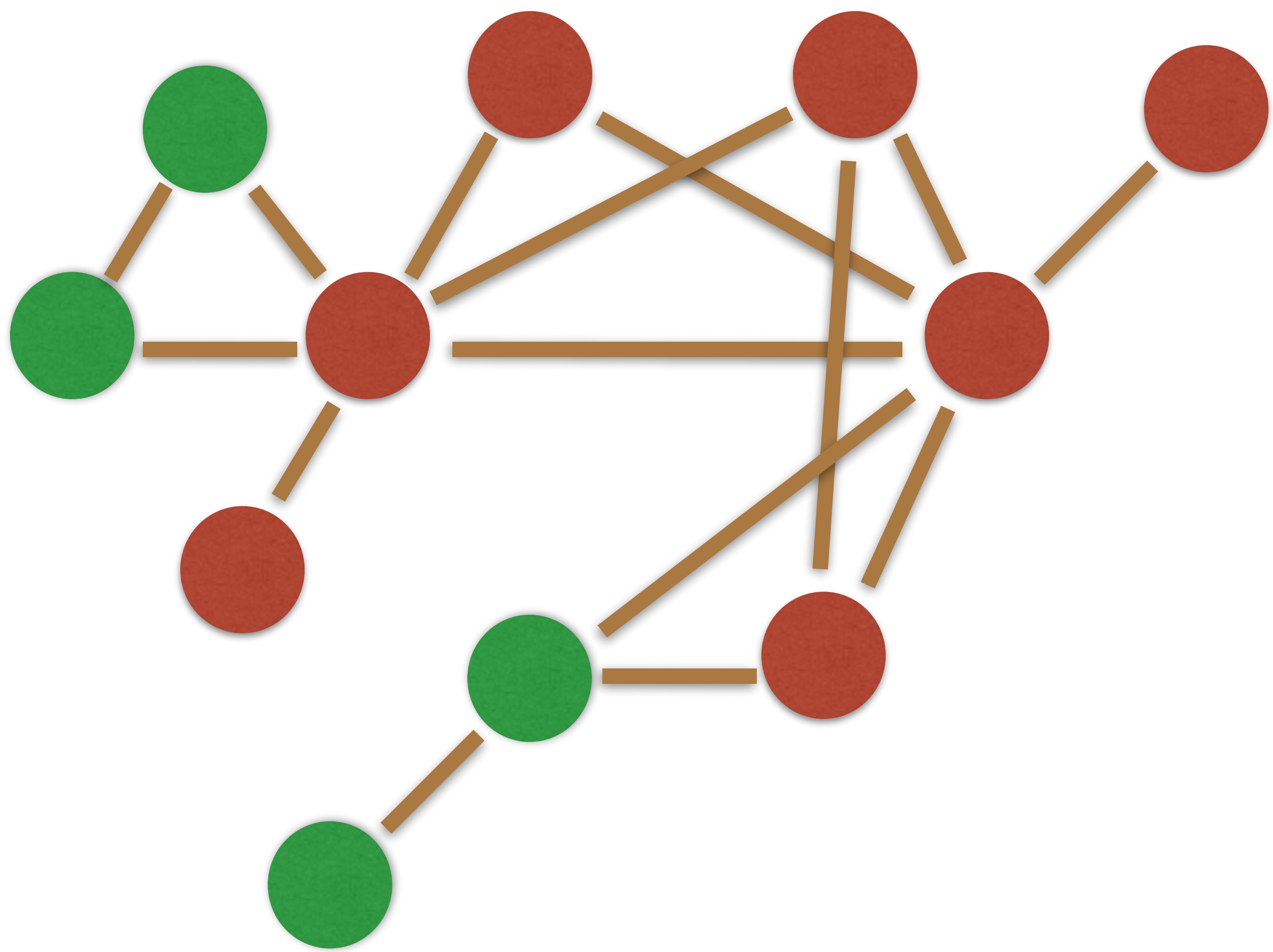
Cascading Example



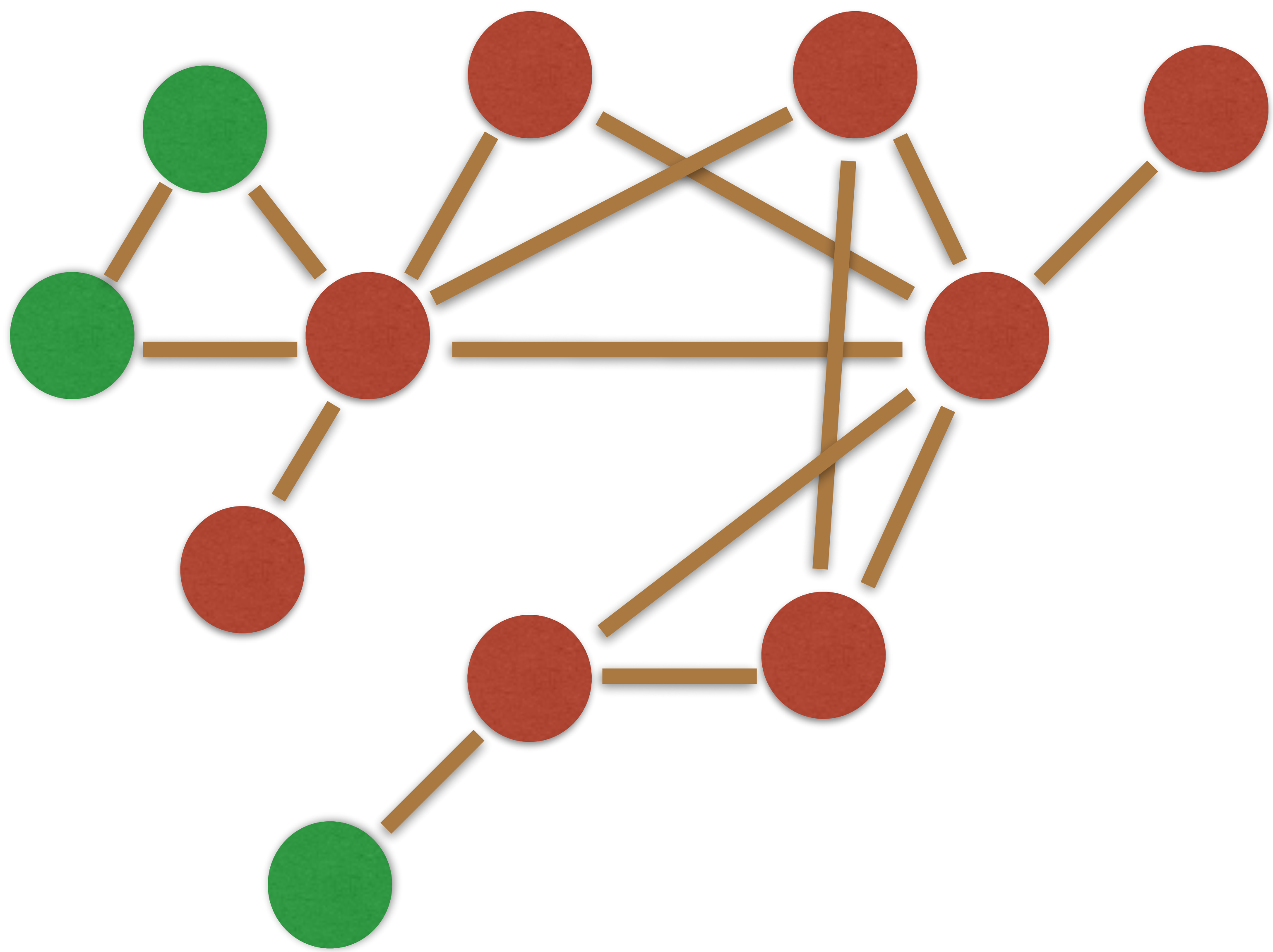
Cascading Example



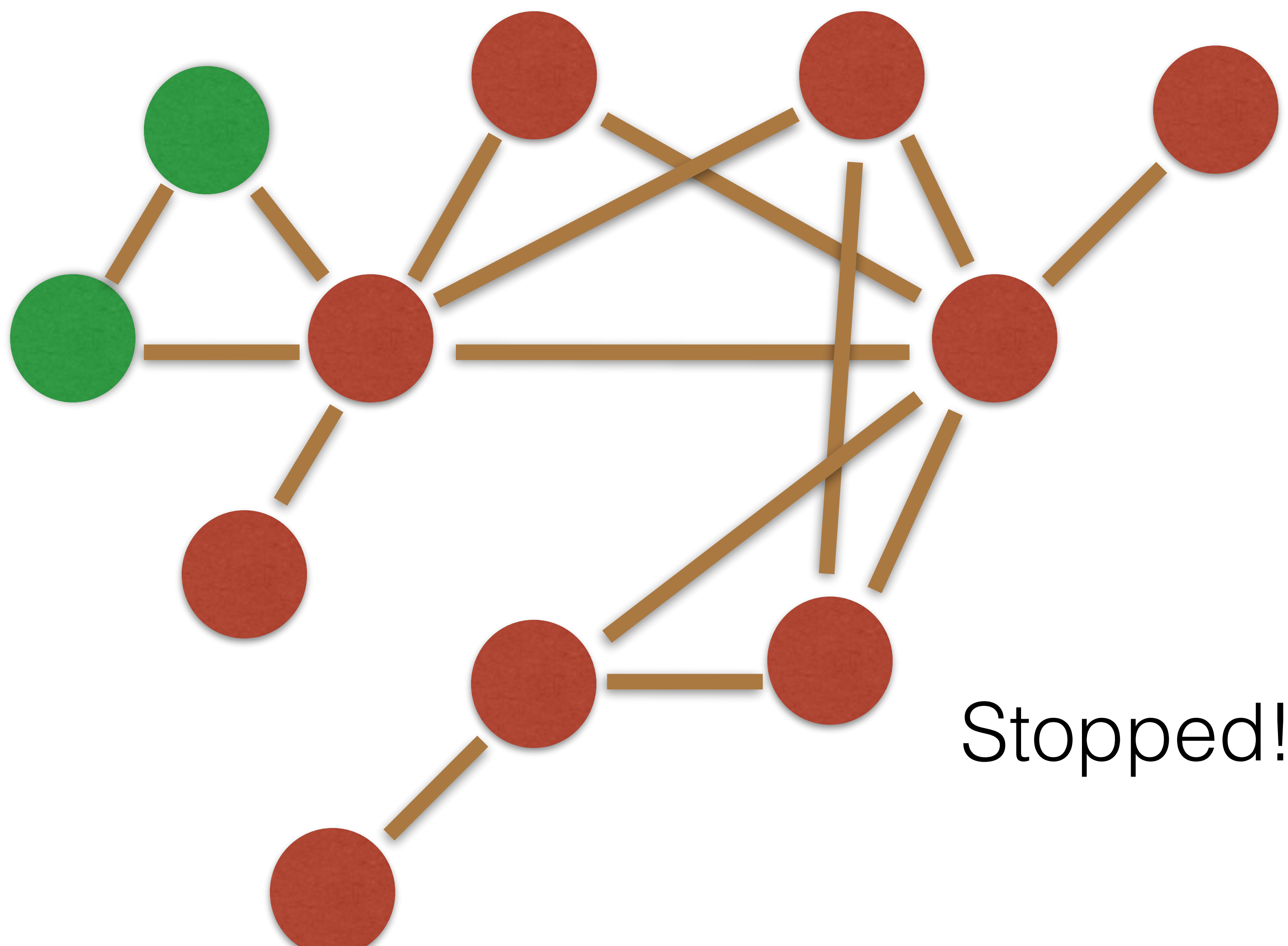
Cascading Example



Cascading Example



Cascading Example



Stopped! Why?

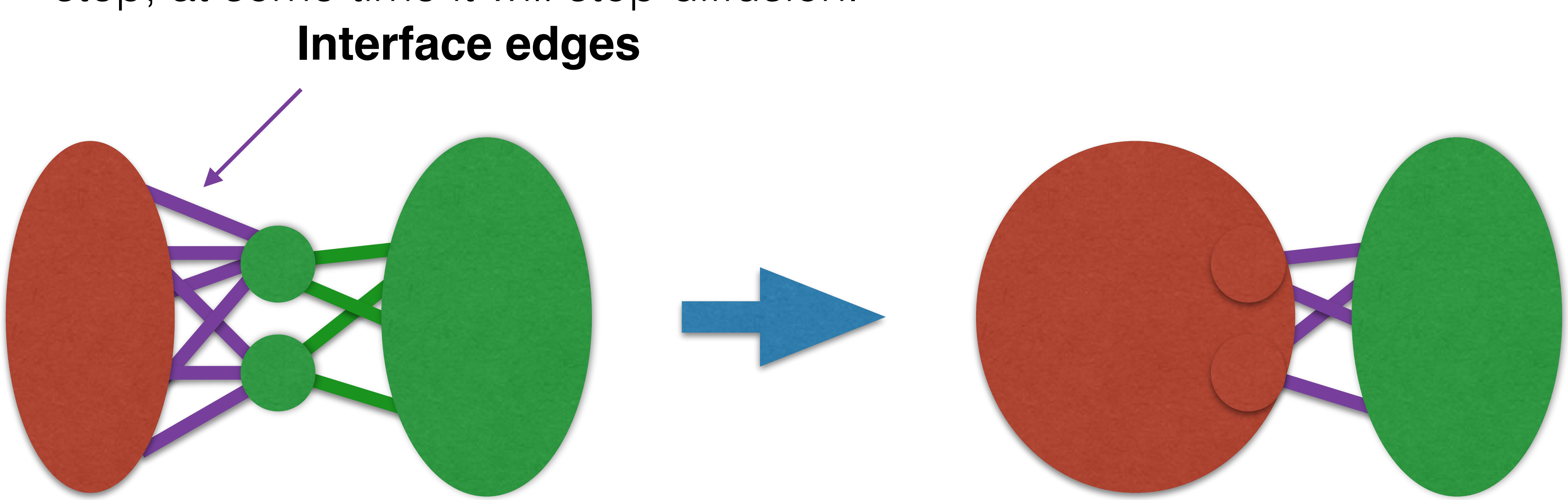
Monotonic Spreading

Reason: at time t, a B node has many A neighboring nodes with proportion p larger than threshold named as q, then it switches from B to A, after time t, the number of its A neighboring nodes can only increase or maintain the same, so it will never switch back to B

Cascade Capacity

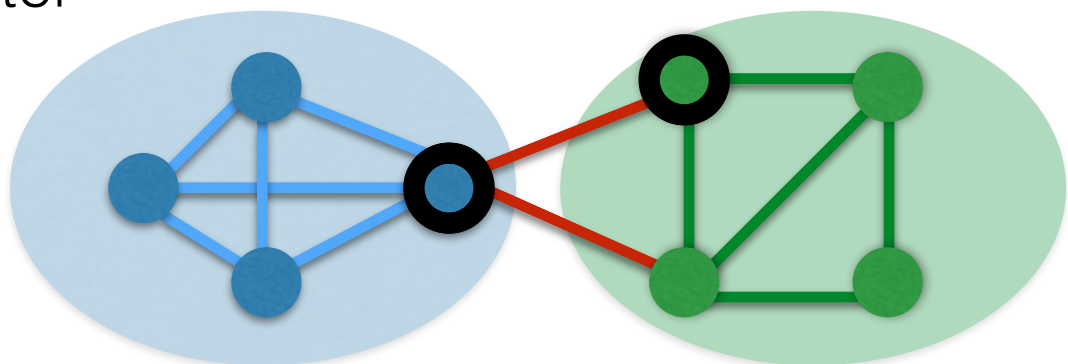
- Def: the largest q (threshold) for which some initial finite set S can cause a cascade. Why largest? because the smaller the threshold, the easier some set will cause a cascade, we want to look at the hardest case.

- capacity $\leq \frac{1}{2}$, why? If q is greater than 1/2, for it to be able to keep cascading, then for each node that will switch in the following step, the proportion of its interface edges must be greater than 1/2, so the total number of interface edges for the graph will decrease in each cascading step, at some time it will stop diffusion.



When Cascades Stop?

- cluster of density ρ : for each node in the cluster, it has at least ρ fraction of edges inside the cluster

$\rho = \frac{3}{5}$  $\rho = \frac{2}{3}$

- stopping condition: say S are the initial adopters, then

there is a cluster in $G \backslash S$ with density $\rho > 1 - q$

$$\Updownarrow$$

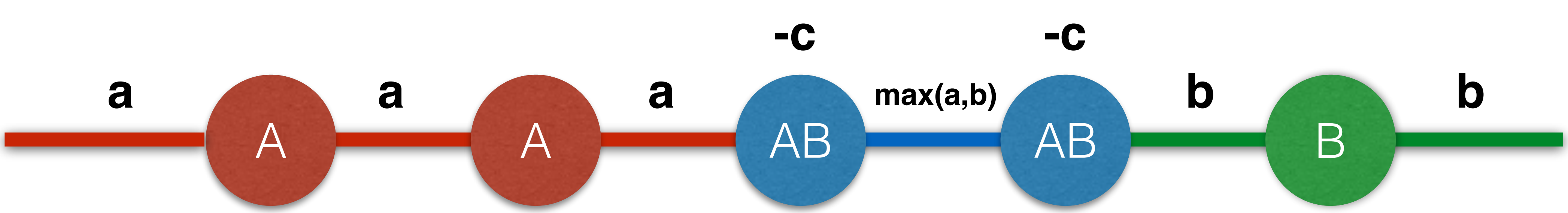
with threshold q, S cannot cause a cascade

Compatibility

- payoff matrix

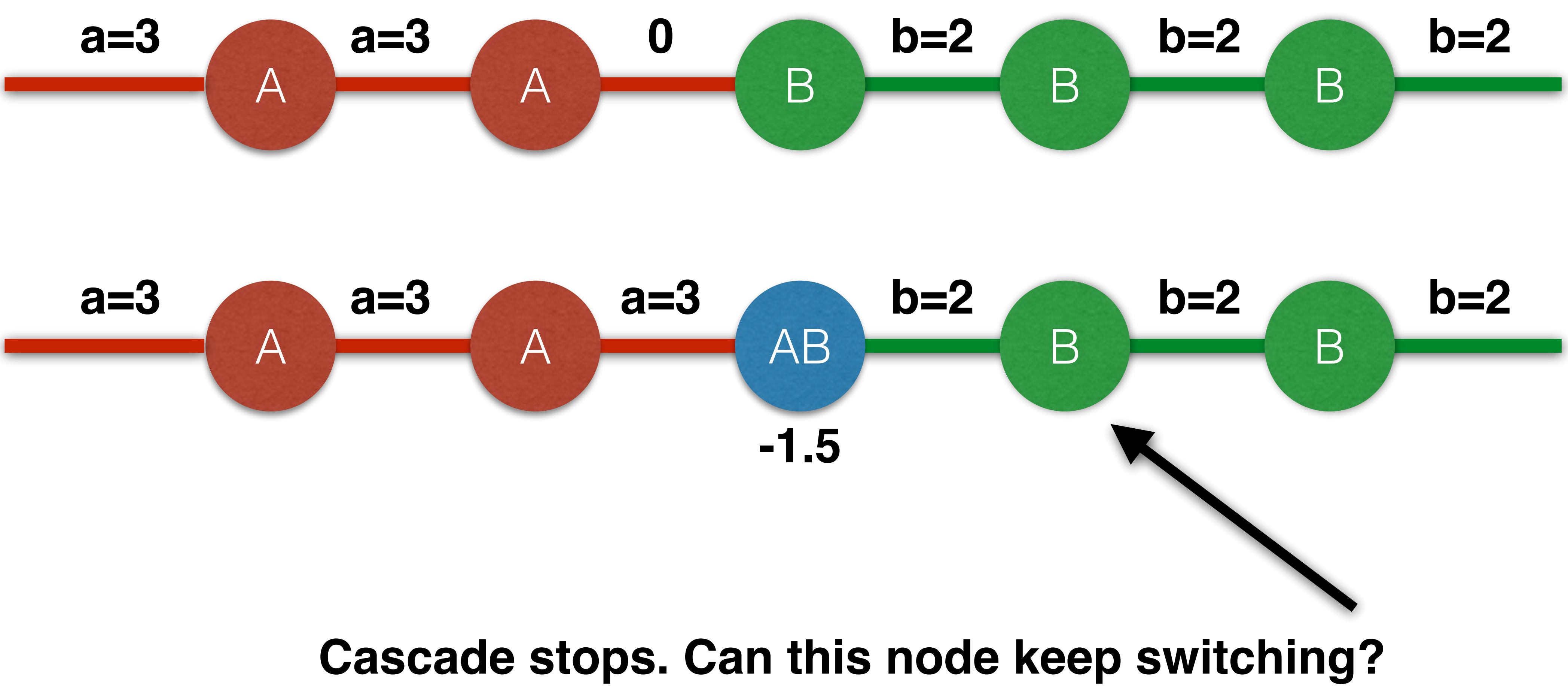
payoff	A	B	AB
A	a	0	a
B	0	b	b
AB	a	b	max(a,b)

, other than these edge payoff, also some cost *c* for the *AB* node itself



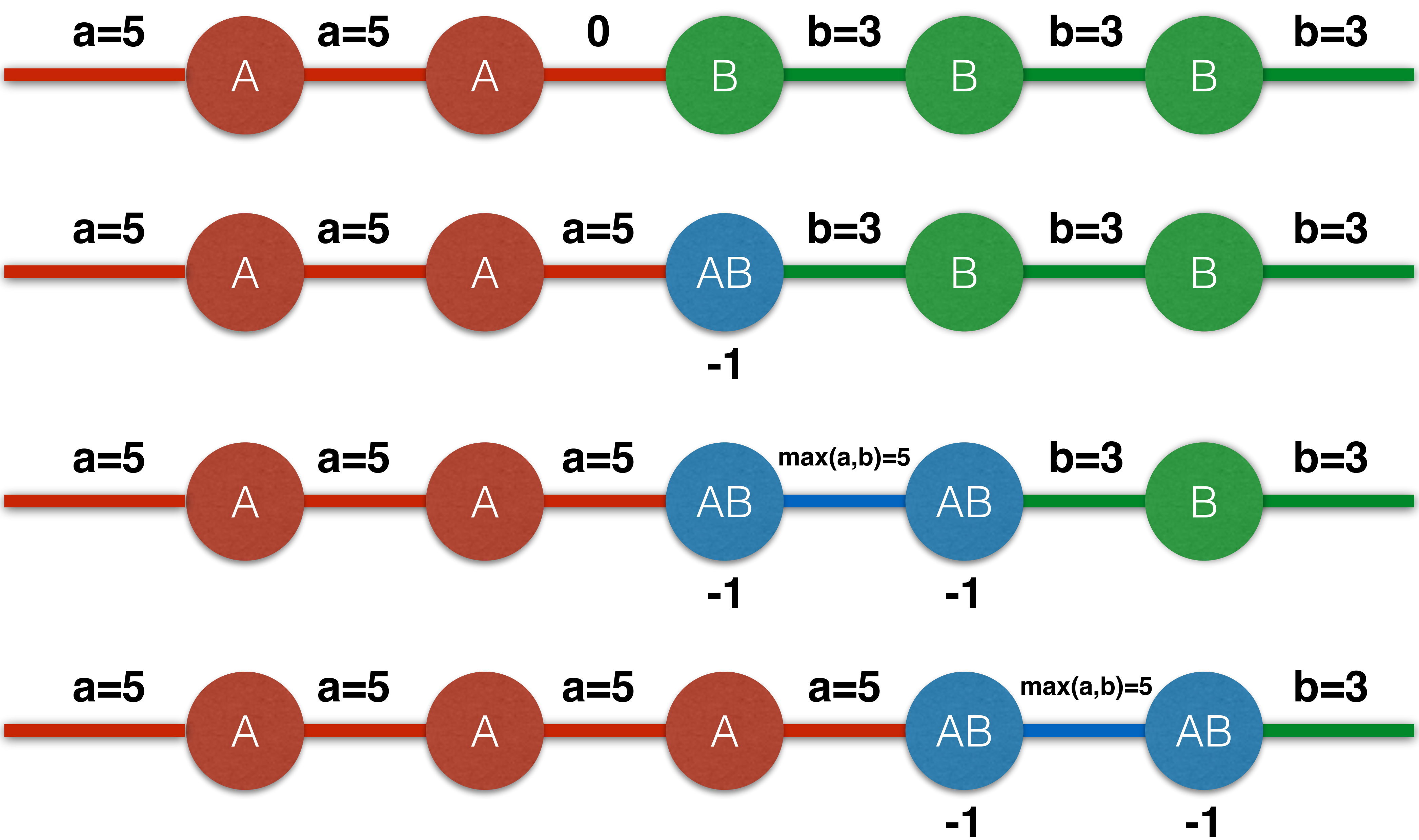
Infinite Path Examples

- suppose *a=3, b=2, c=1.5*

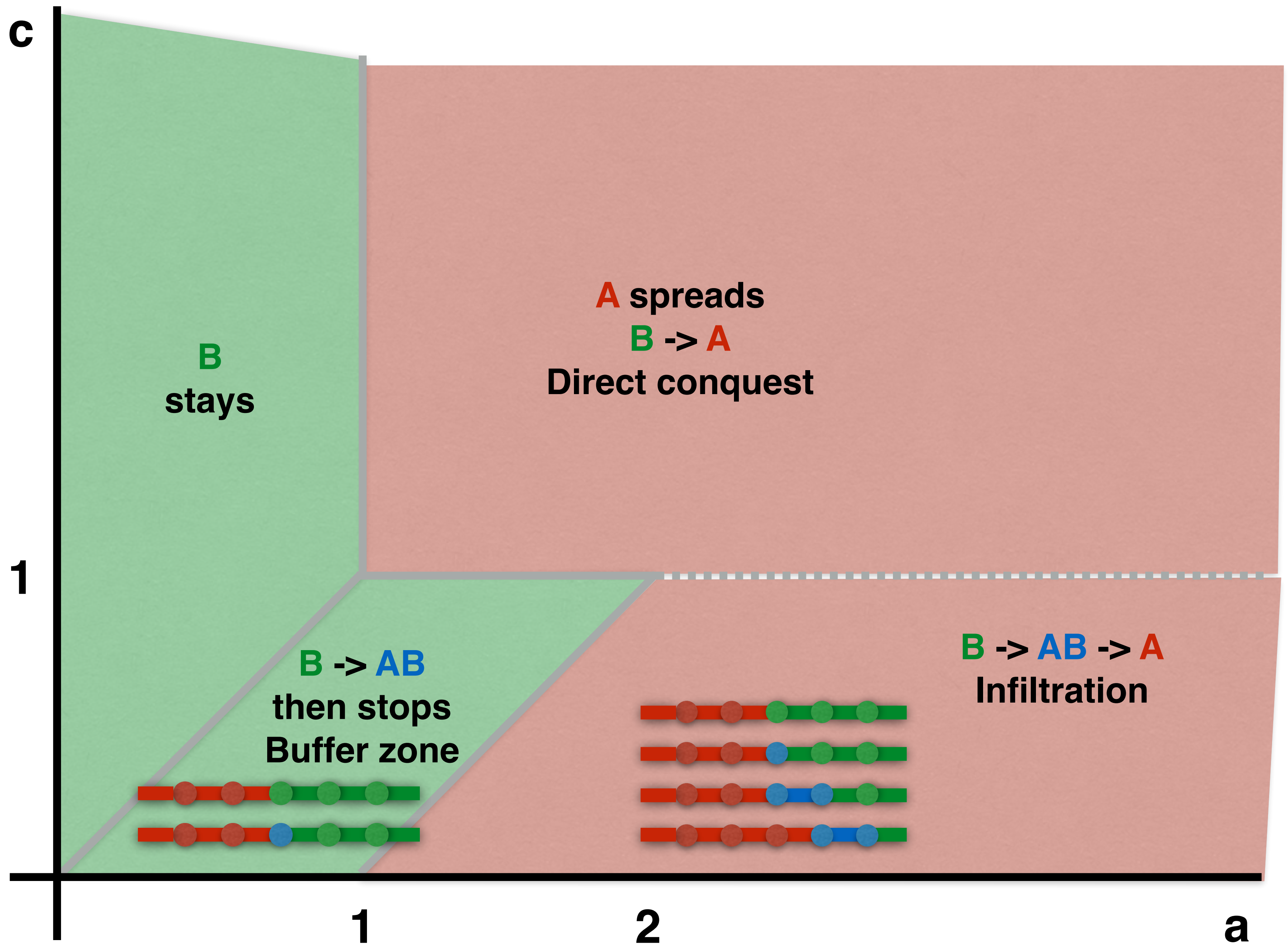


Infinite Path Examples

- suppost *a=5, b=3, c=1*



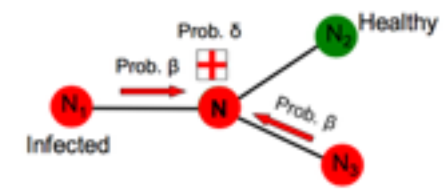
Cascading Phase Graph



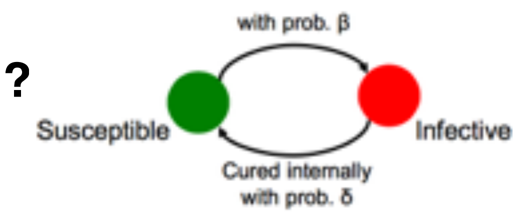
b=1, start with all *B* nodes, how cascading moves on for an infinite path graph?

Contagion - Disease Models

- **meaning of S, I and R?** **susceptible, infected and recovered**



- **what is SIR model, SIS model?**

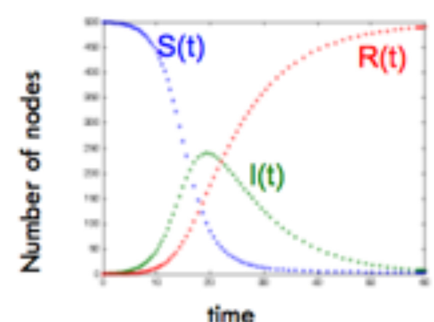


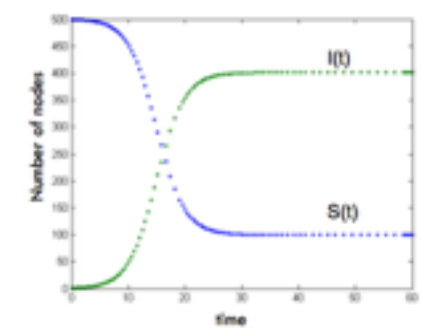
- **what is the meaning of their model dynamics?**

$$\frac{dS}{dt} = -\beta S I$$
$$\frac{dI}{dt} = \beta S I - \delta I$$
$$\frac{dR}{dt} = \delta I$$

$$\frac{dS}{dt} = -\beta S I + \delta I$$
$$\frac{dI}{dt} = \beta S I - \delta I$$

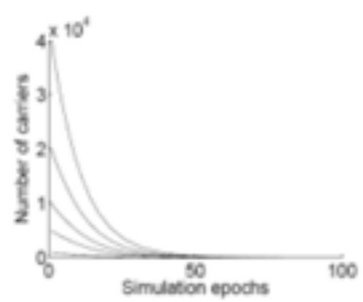
- **what does S(t), I(t) and R(t) curves looks like?**



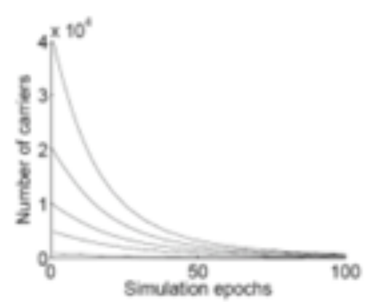


- **what is epidemic threshold in SIS model?** **If virus strength $\frac{\beta}{\delta} < \tau = \frac{1}{\lambda_{1,A}}$, the epidemic eventually dies out.**

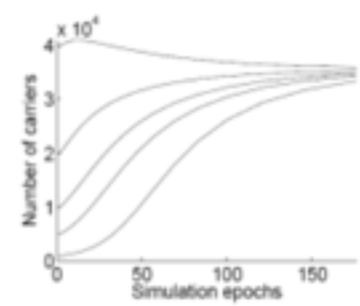
- **does it matter how many people are initially infected?**



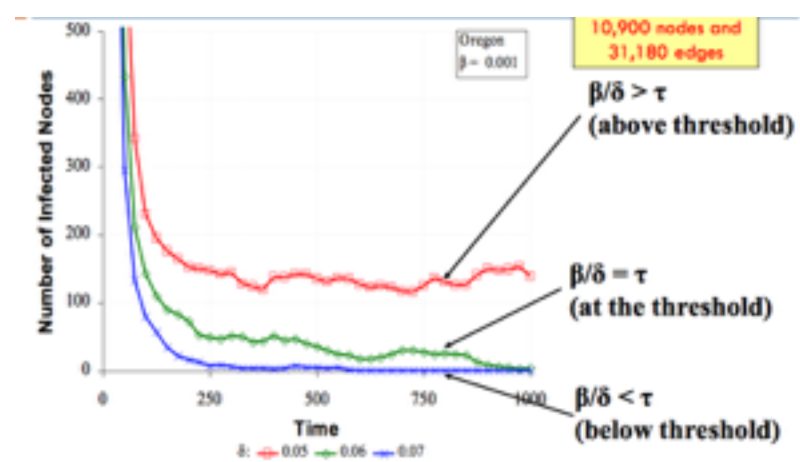
(a) Below the threshold, s=0.912



(b) At the threshold, s=1.003



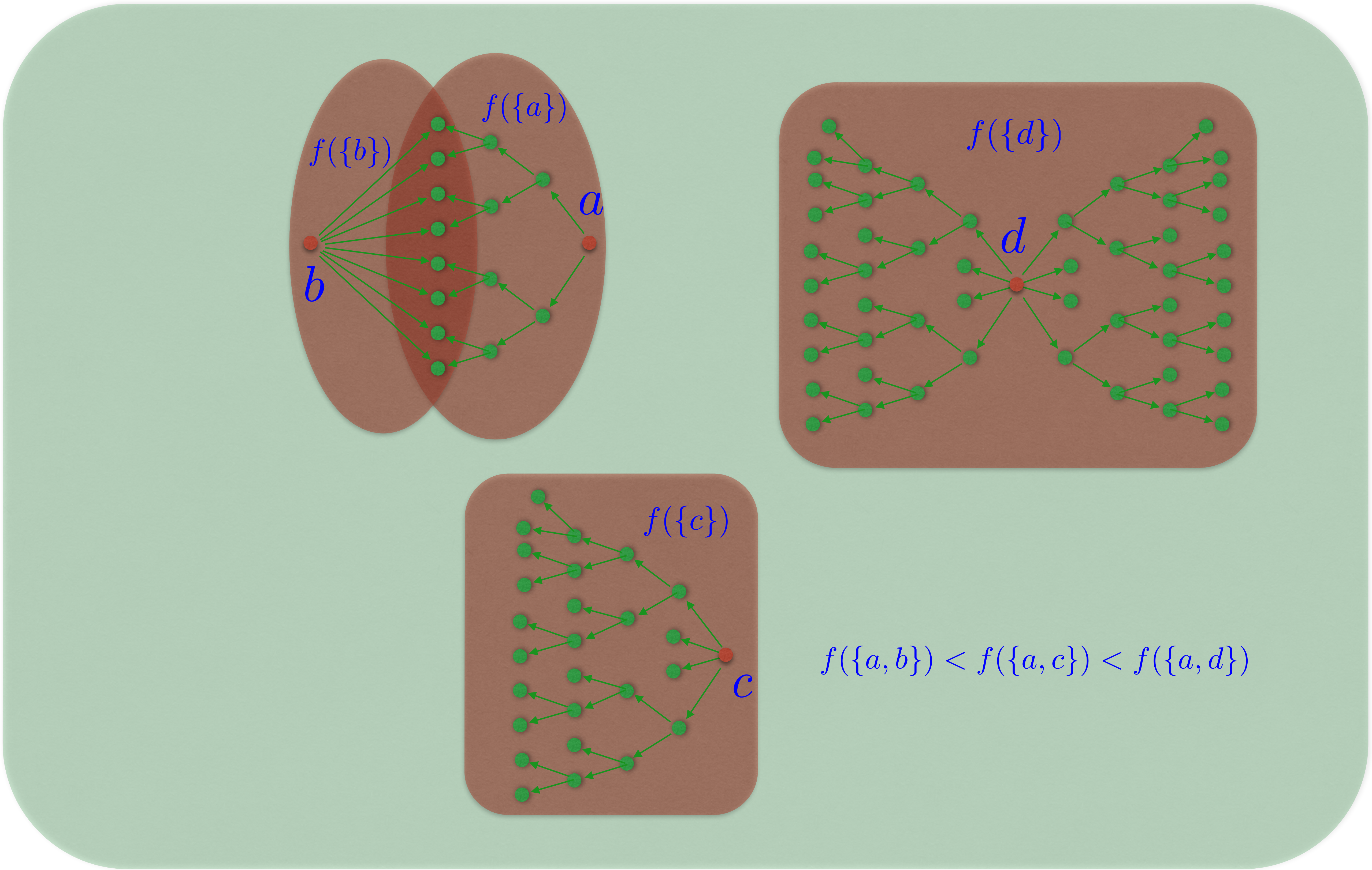
(c) Above the threshold, s=1.1



Influence Maximization

- independent cascade model: directed graph; set S as initial set with active nodes; each edge (v,w) has a probability p_{vw} that if any node is active, it get one chance to make the other node active with that probability

- $f(s)$ is the expected size of final active set



$f(\{a\})$
$f(\{b\})$
a
b
$f(\{d\})$
d
$f(\{c\})$
c
$f(\{a,b\}) < f(\{a,c\}) < f(\{a,d\})$

Most Influential Subset

- Given size of S is k ($|S|=k$), influential maximization is to find the most influential subset of size k:

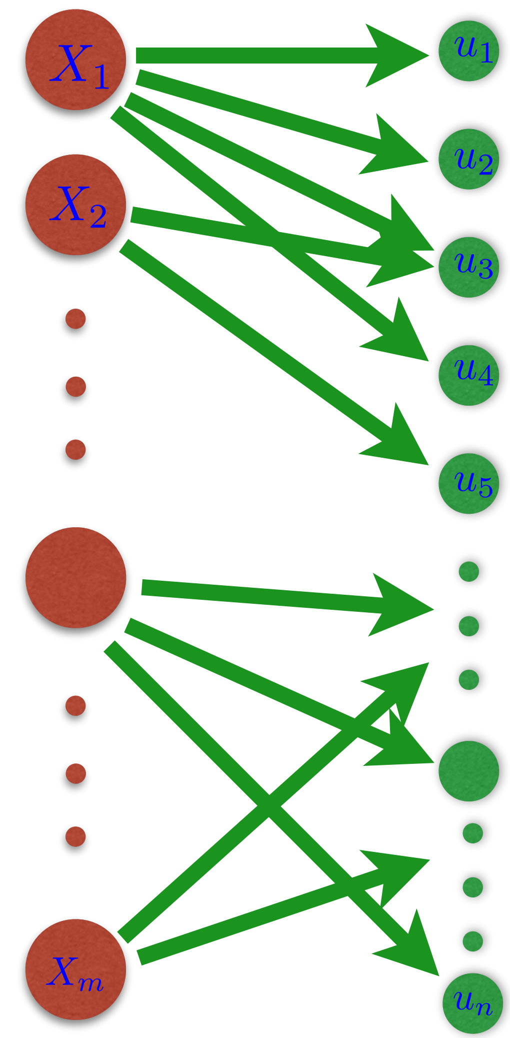
$$\max_{|S|=k} f(S)$$

- NP-completeness: A vertex / set cover problem can be reduced to a most influential subset problem

Vertex / Set Cover instance | **Most Influential Subset instance**

$$X_1 = \{u_1, u_2, u_3, u_4\}$$
$$X_2 = \{u_3, u_5\}$$
$$\vdots$$
$$\vdots$$
$$X_m = \{\cdots\}$$

reduction



There exists size k set cover $\xleftrightarrow{\text{iff}}$ **There exists a set S of size k with f(s)=k+n**

Observation: optimal solution does not contain u nodes

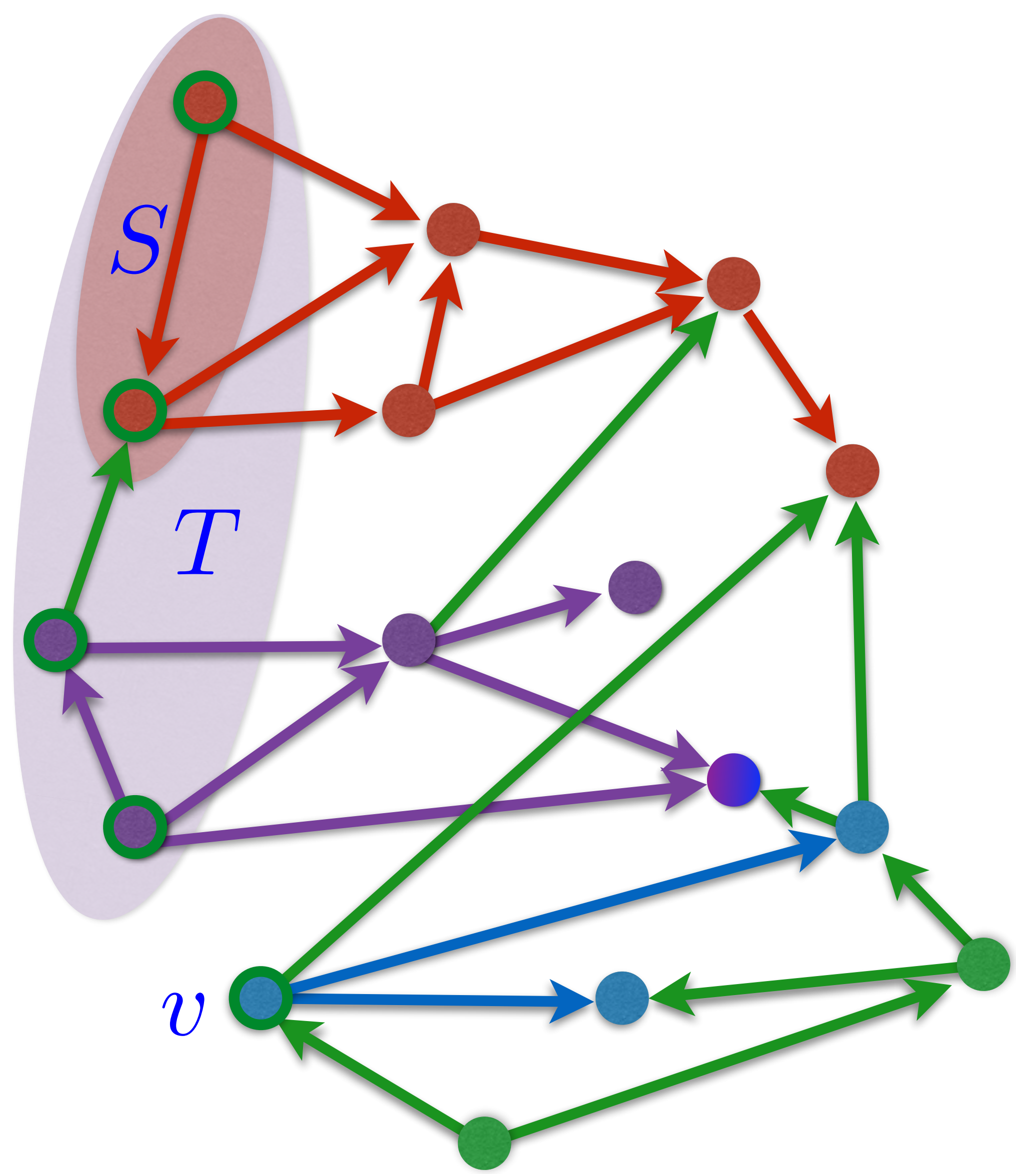
Greedy Algorithm

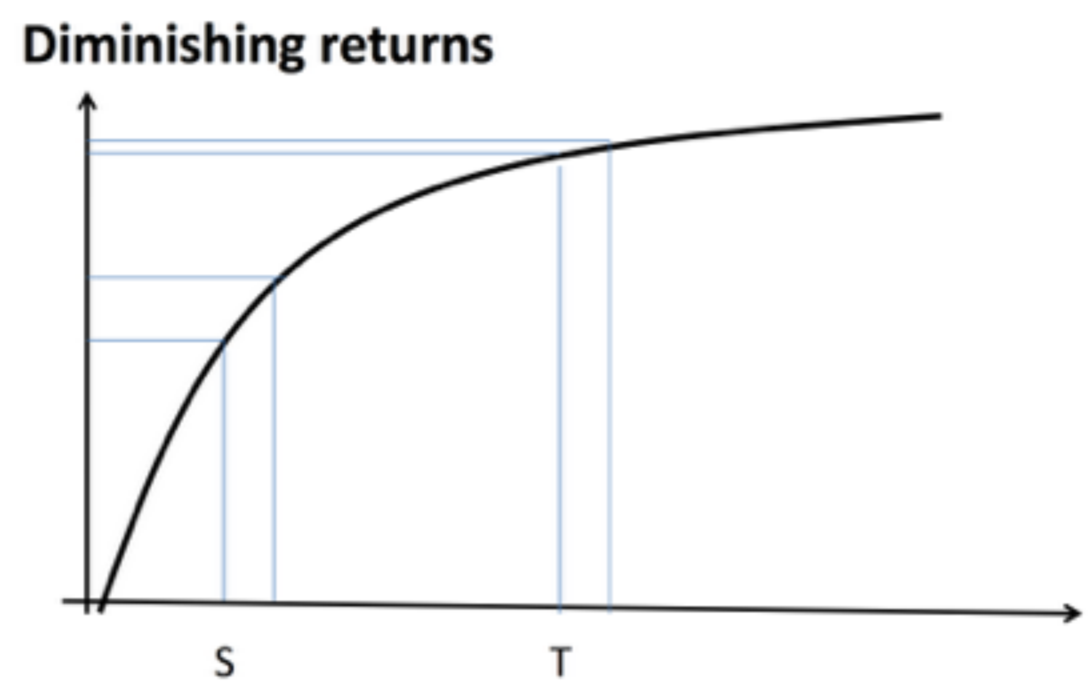
- start with $S_0 = \{\}$
- for $i = 1, \cdots, k$
  - take node u that $\max_u f(S_{i-1} \cup u)$
  - let $S_i = S_{i-1} \cup u$

f is a set function: $2^{|U|} \rightarrow R$

$$\left.\begin{array}{ll} \textbf{nondecreasing:} & f(S \cup \{v\}) \geq f(S) \\ \textbf{submodular:} & f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T), \forall S \subset T \end{array}\right\} \Rightarrow f(S) \geq (1 - \frac{1}{e}) * OPT$$

Submodular Function

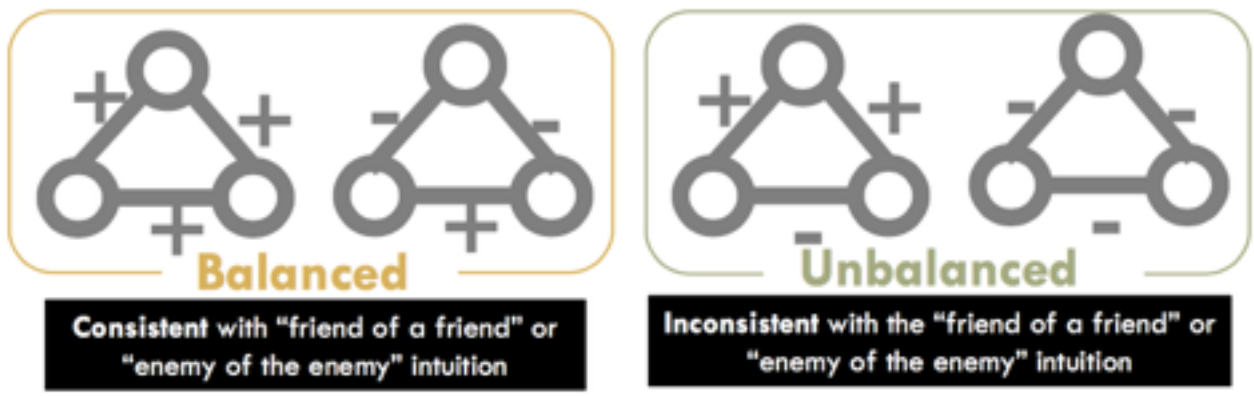




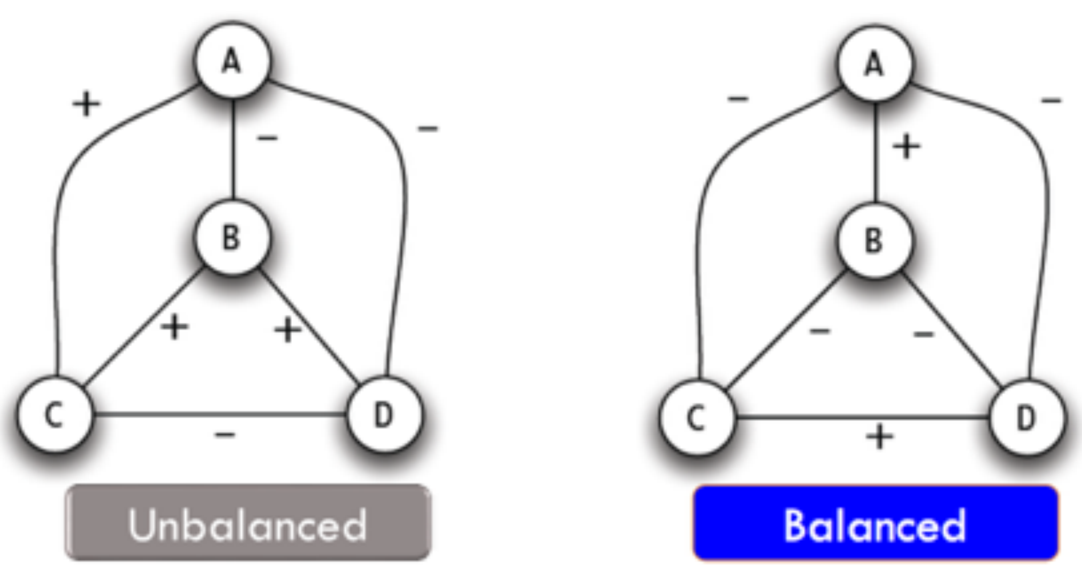
$$f(S) = 6, f(S \cup \{v\}) = 10, f(S \cup \{v\}) - f(S) = 4$$
$$f(T) = 11, f(T \cup \{v\}) = 14, f(T \cup \{v\}) - f(T) = 3$$

Signed Network: Balance

- intuition

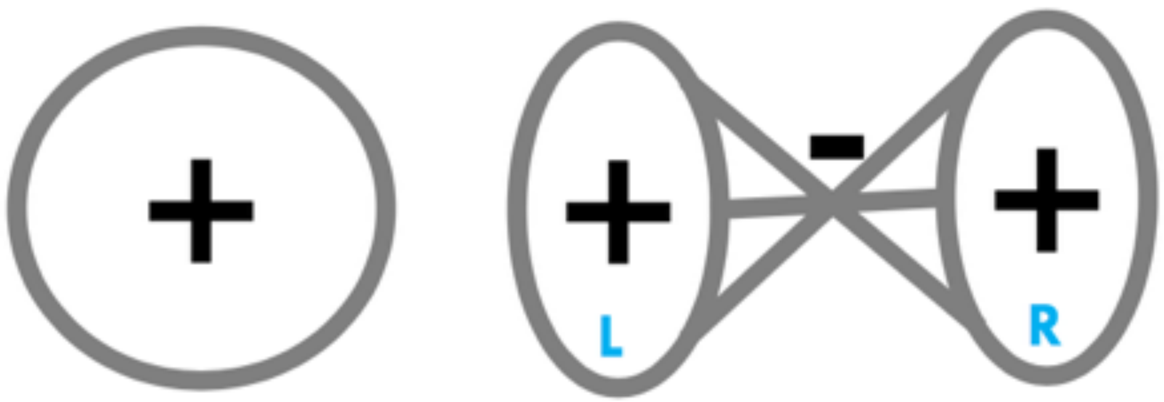


- Graph is balanced if every triangle is balanced

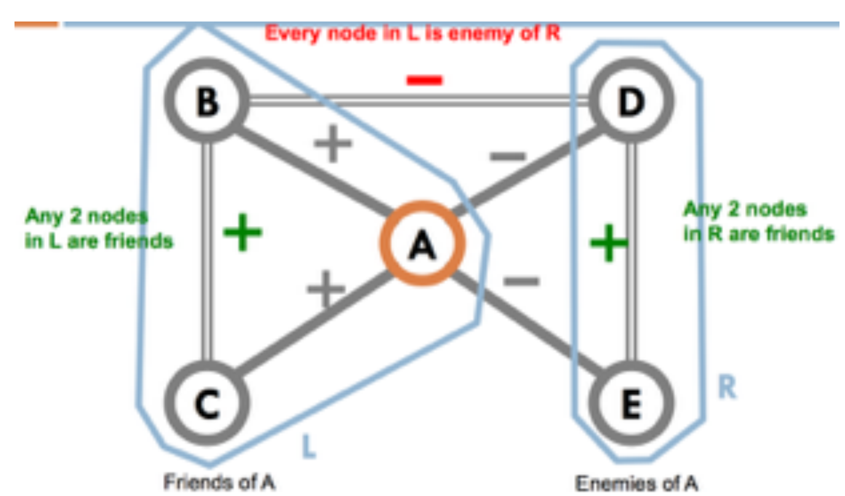


Balance Leads to Coalitions

- local balance implies global coalitions



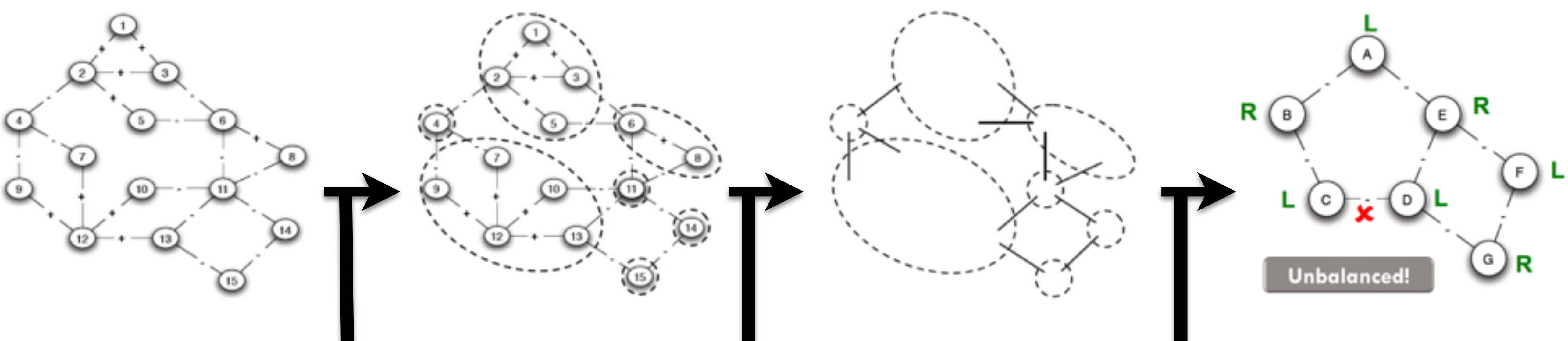
- why?



Is it Balanced?

- graph is balanced if and only if it contains no circle with an odd number of negative edges

- how to check?



Find connected components on + edges, if we find a component contains a - edge, then unbalanced

For each component create a super-node, connect 2 super-nodes if there is a - edge between them

Using BFS assign each node a side, if there is neighbor which is already assigned the same side, then unbalanced