computational social media

lecture 3: tweeting

part 4

daniel gatica-perez

this lecture

a human-centric view of twitter

1. introduction
2. twitter users & uses
3. understanding large-scale human behavior
4. inferring real-world events & trends
5. spreading information in the real world

spreading information in the real world

1. who talks to whom on twitter
2. **cascading behavior in networks**
3. structural virality of online diffusion
4. twitter and the news

2. cascading behavior in networks

Materials are taken from:
D. Easley and J. Kleinberg. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, 2010. Chapter 19, http://www.cs.cornell.edu/home/kleinber/networks-book/

diffusion of innovations:
how do behaviors and practices spread among people through a social network?

informational effects

* someone adopts a practice
* neighbors observe this person's decisions
* this gives indirect information that motivate neighbors to adopt it too
* examples: doctors adopting tetracycline
  farmers adopting hybrid seed corn

direct-benefit effects

* people adopts a practice given the direct incentives
* examples: telephone, email (benefit: communicating with others who already adopted the practice)

common elements

* innovations tend to arrive from outsiders
* often difficult for innovations to be adopted in tight communities

modeling network diffusion

key concept : given your neighbors, your benefits of adopting a behavior increase as more of your neighbors adopt it too

networked coordination game

* assume nodes *v* and *w* are linked by an edge
* assume 2 possible behaviors *A* and *B*
* game: 2 players (*v* ,*w*) and 2 strategies (adopting *A* or *B*)

- if v and w both adopt behavior A, they each get a payoff of $a > 0$;
- if they both adopt B, they each get a payoff of $b > 0$; and
- if they adopt opposite behaviors, they each get a payoff of 0.

We can write this in terms of a payoff matrix,

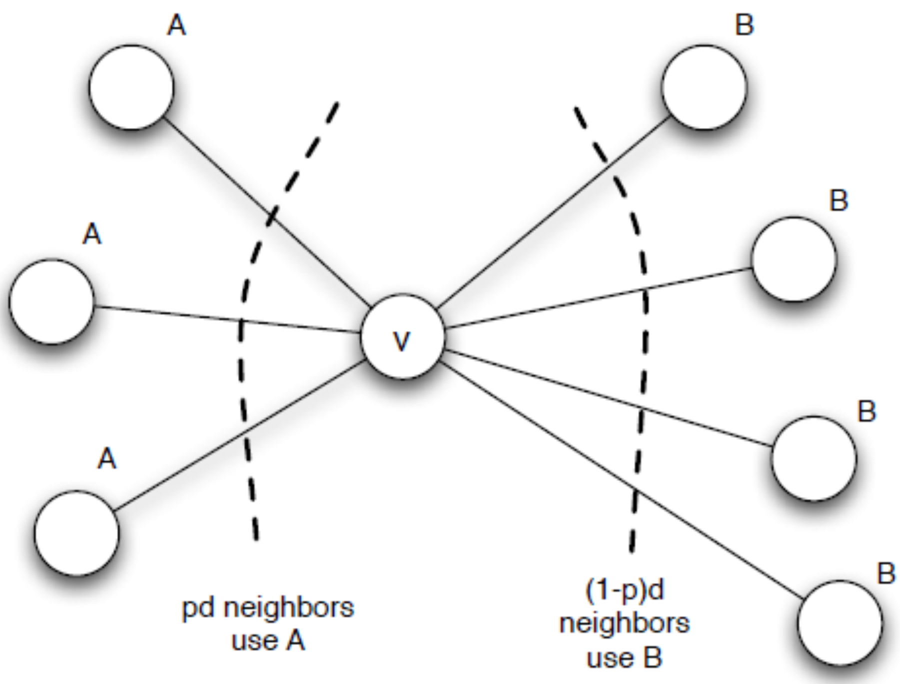
		w: A	w: B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

Figure 19.1: A-B Coordination Game

modeling network diffusion (2)

Assume node *v* has *d* neighbors, out of whom a **fraction** *p* has adopted *A* and the rest *(1-p)* has adopted *B*

What should node *v* do to maximize **payoff** if some neighbors adopt *A* and others adopt *B*?



If *v* chooses *A:*
payoff (v) = pda

If *v* chooses *B:*
payoff (v) = (1-p)db

A is the better choice if

$$pda \geq (1-p)db,$$

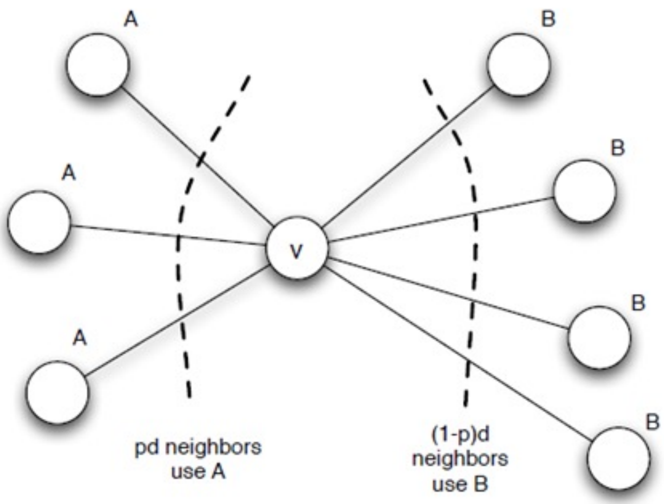
$$p \geq \frac{b}{a+b}.$$

modeling network diffusion (3)

A is the better choice if

$$pda \geq (1-p)db,$$

$$p \geq \frac{b}{a+b}.$$



Strategy: if at least a fraction q = *b/(a+b)* of your neighbors follow behavior *A*, you should too

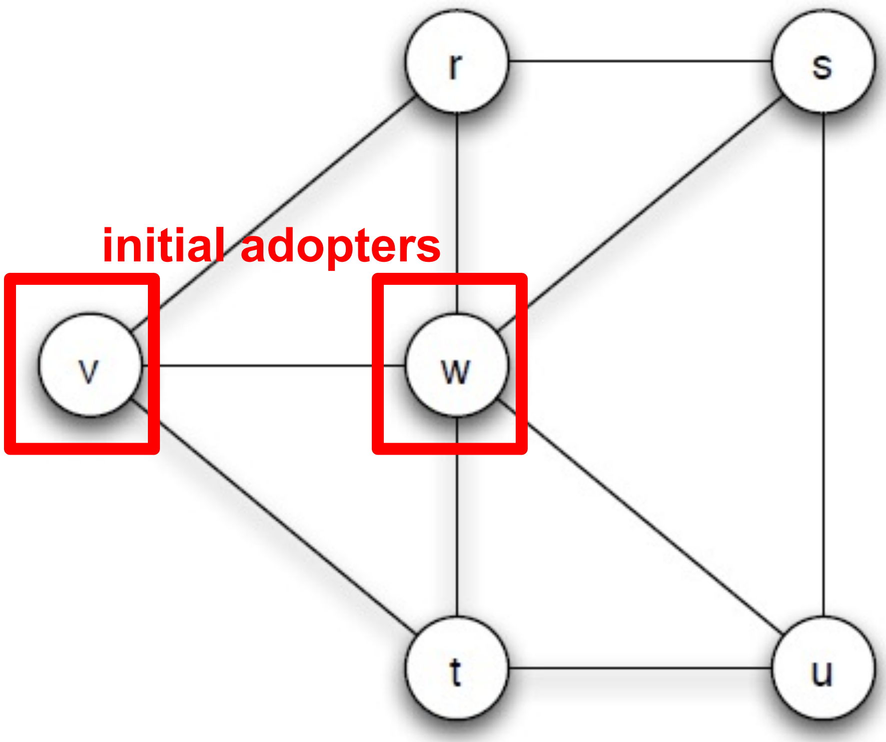
q is called a **threshold**

If q is small (*b/(a+b)* <<1), then payoff *a* is relatively big compared to payoff *b*, and A is the most attractive behavior

If q is large (*b/(a+b)* ~1), then payoff *a* is relatively small compared to payoff *b*, and B is the most attractive behavior

cascading behavior

Two obvious equilibria: everyone adopts A or everyone adopts B
How to tip the network to go from one equilibrium to the other one?
What intermediate equilibria can exist?

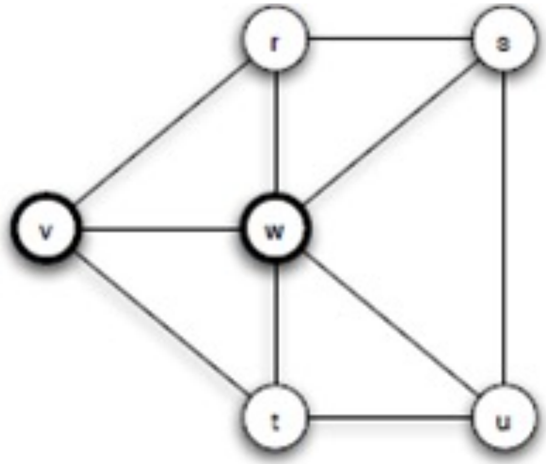


Assume B as default behavior and a small set of **initial adopters** who switched to behavior A (outside of the game rules)

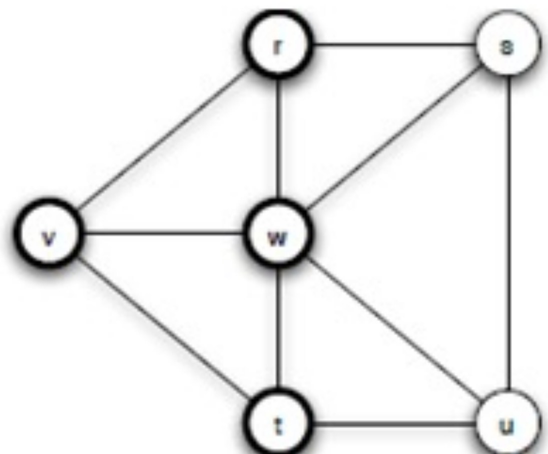
Given this, some neighbors might also decide to switch to A
Who will do it?
When will the process stop?

Dynamics of the process:
* Network structure
* Choice of initial adopters
* Threshold value q

cascading behavior (2)



Two nodes are the initial adopters



After one step, two more nodes have adopted

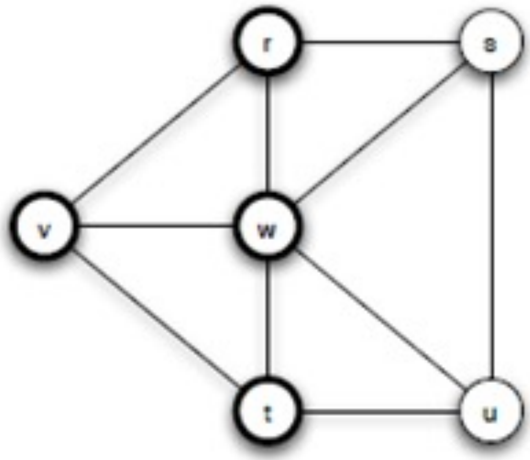
Time runs in steps
At each step, each node uses threshold rule to decide whether to switch from B to A

Process stops when
(1) all nodes switch to A, or
(2) no node switches any more

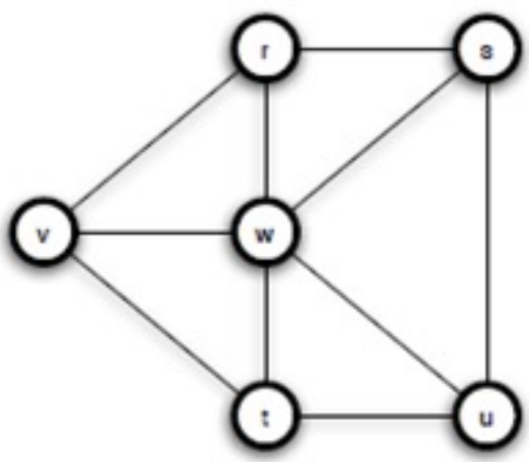
Example:
 * a=3; b=2; q = 2/(2+3)=2/5
 * node will switch from B to A if at least 40% of its neighbors use A

Step 1
* N(r): {s,v,w}
    2/3 neighbors use A: switch
* N(s): {r,u,w}
    1/3 neighbors use A: no-switch
* N(t): {u,v,w}
    2/3 neighbors use A: switch
* N(u): {s,t,w}
    1/3 neighbors use A: no-switch

cascading behavior (3)



After one step, two more nodes have adopted



After a second step, everyone has adopted

Reminder
* node will switch from B to A if at least 40% of its neighbors use A

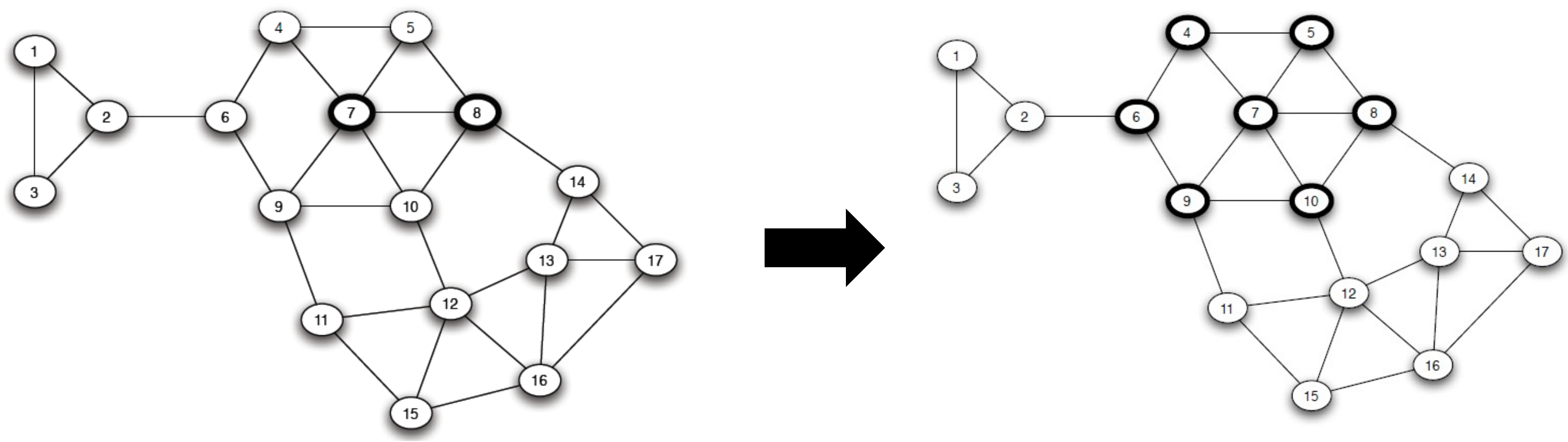
Step 2
* N(s): {r,u,w}
  2/3 neighbors use A: switch
* N(u): {s,t,w}
  2/3 neighbors use A: switch

END: all nodes switched to A

It can be shown that once a node has switched to A, it will NOT switch back to B

cascade of adoptions

Cascade of adoptions of A: chain reaction of switches to A

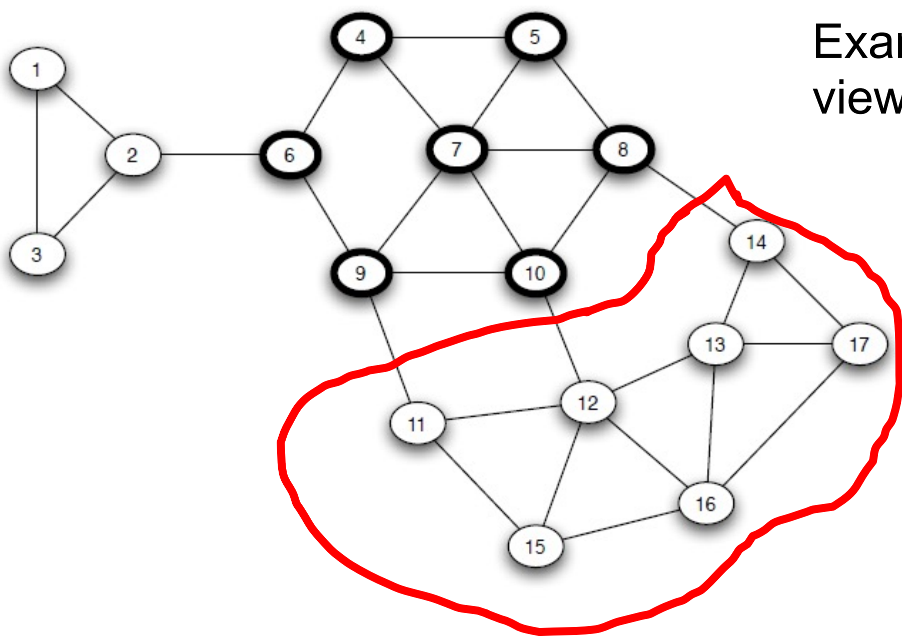


"Consider a set of initial adopters who start with a new behavior A, while every other node starts with behavior B. Then, nodes repeatedly evaluate the decision to switch from B to A using a threshold q. If the resulting cascade of adoptions of A eventually causes every node to switch from B to A, then we say that the set of initial adopters causes a **complete cascade** at threshold q."

cascading behavior & "viral marketing"

Tightly-knit communities can work to limit the spread of an innovation

Example: different dominant political views between adjacent communities



How to spread the cascade?

1. increase payoff to switch to A (i.e., lower threshold q)

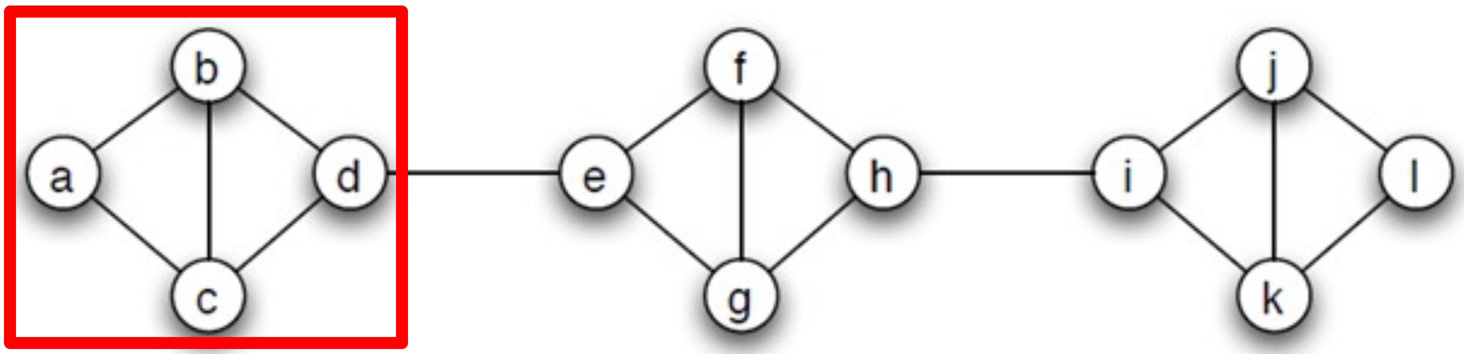
2. convince key people using B to switch to A, choosing them based on their network position to get the cascade moving

Figure 19.5: Starting with nodes 7 and 8 as the initial adopters, the new behavior *A* spreads to some but not all of the remaining nodes.

cascades & clusters

intuition: cascades can **get stuck** when trying to break into close communities.
homophily can be a barrier to diffusion of innovations from outsiders

cluster of density p: set of nodes such that each node in the set has at least a p fraction of its neighbors in the set



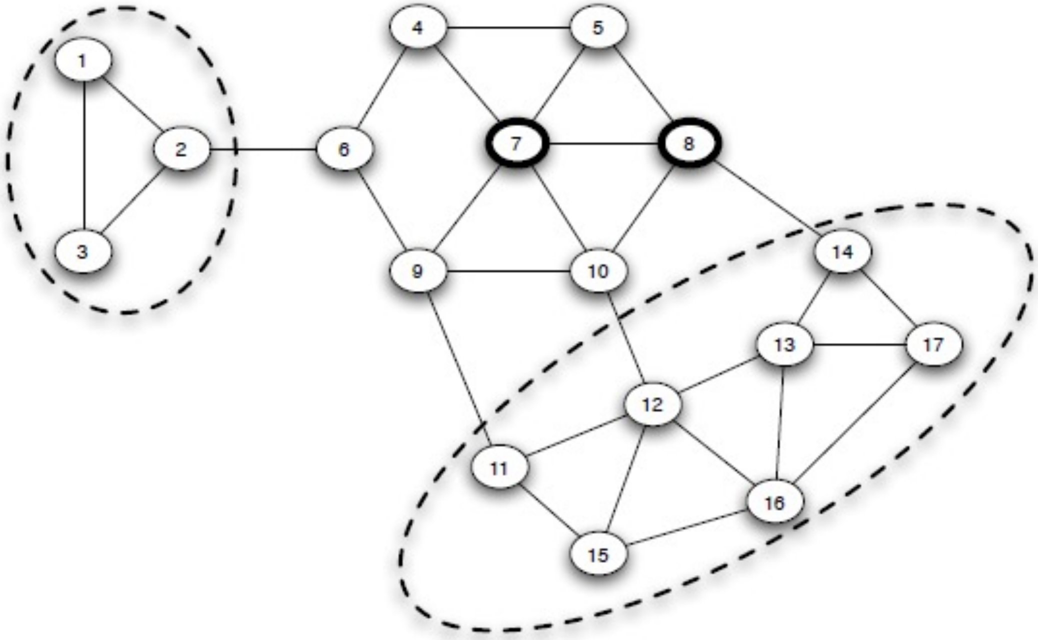
cluster X = {a,b,c,d}:

N(a) = {b,c} p(a,X) = 2/2
N(b) = {a,c,d} p(b,X) = 3/3
N(c) = {a,b,d} p(c,X) = 3/3
N(d) = {b,c,e} p(d,X) = 2/3

density(X) = 2/3

clusters are the natural obstacles to cascades

the cluster structure of the network tells about the success of a cascade

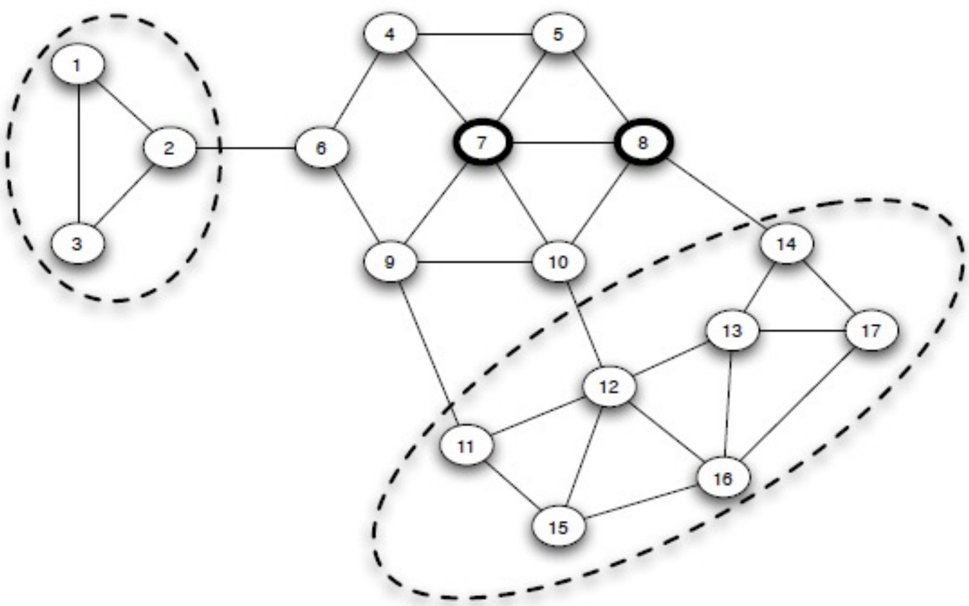


Two clusters of density 2/3 in the network

cascades stop when they run into dense clusters

this is the only thing that cause cascades to stop

clusters are the natural obstacles to cascades (2)



Two clusters of density 2/3 in the network

For a set of initial adopters of A and threshold q to adopt A:

(1) If the remaining network contains a cluster of density greater than (1-q), the set of initial adopters will not cause a complete cascade

(2) Whenever a set of initial adopters does not cause a complete cascade, the remaining network must contain a cluster of density greater than (1-q)

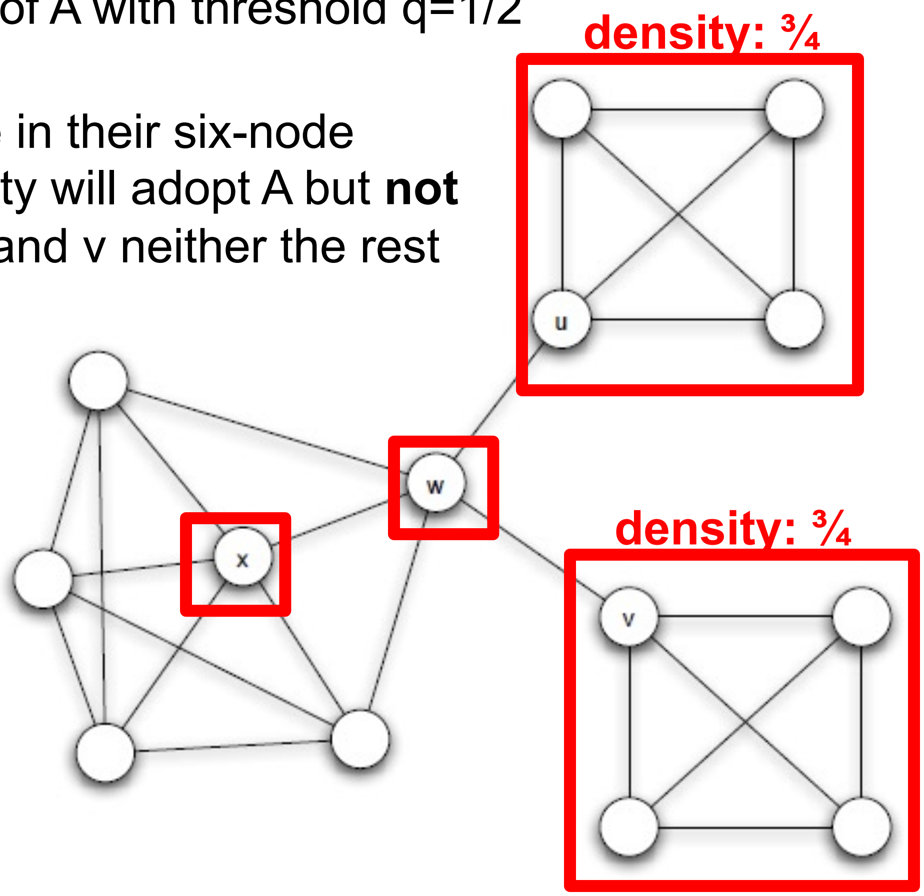
extreme cases:

* very large payoff *a*: q tends to 0; (1-q) tends to 1
* very small payoff *a*: q tends to 1; (1-q) tends to 0

diffusion, thresholds, and weak ties

assume nodes x and w to be initial adopters of A with threshold q=1/2

everyone in their six-node community will adopt A but **not** nodes u and v neither the rest



local bridges can convey ***awareness*** of new info, but are weak at transmitting behaviors that are **costly to *adopt*** (for which you need a **higher threshold** of neighbors doing it before you do it too)

nodes u & v have ***informational*** advantages over their own communities. but for behaviors with higher thresholds they will still want to **align with their community**

what to remember

cascading behavior in networks

diffusion of innovations: how behaviors & practices spread

interplay between local interactions & network structure

- initial adopters, adoption thresholds
- clusters are the natural obstacles of cascades

diffusion, thresholds, weak ties

- awareness of new info is different than costly-to-adopt behaviors (for which a higher threshold of neighbors adopting the behavior is needed)

questions?

daniel.gatica-perez@epfl.ch