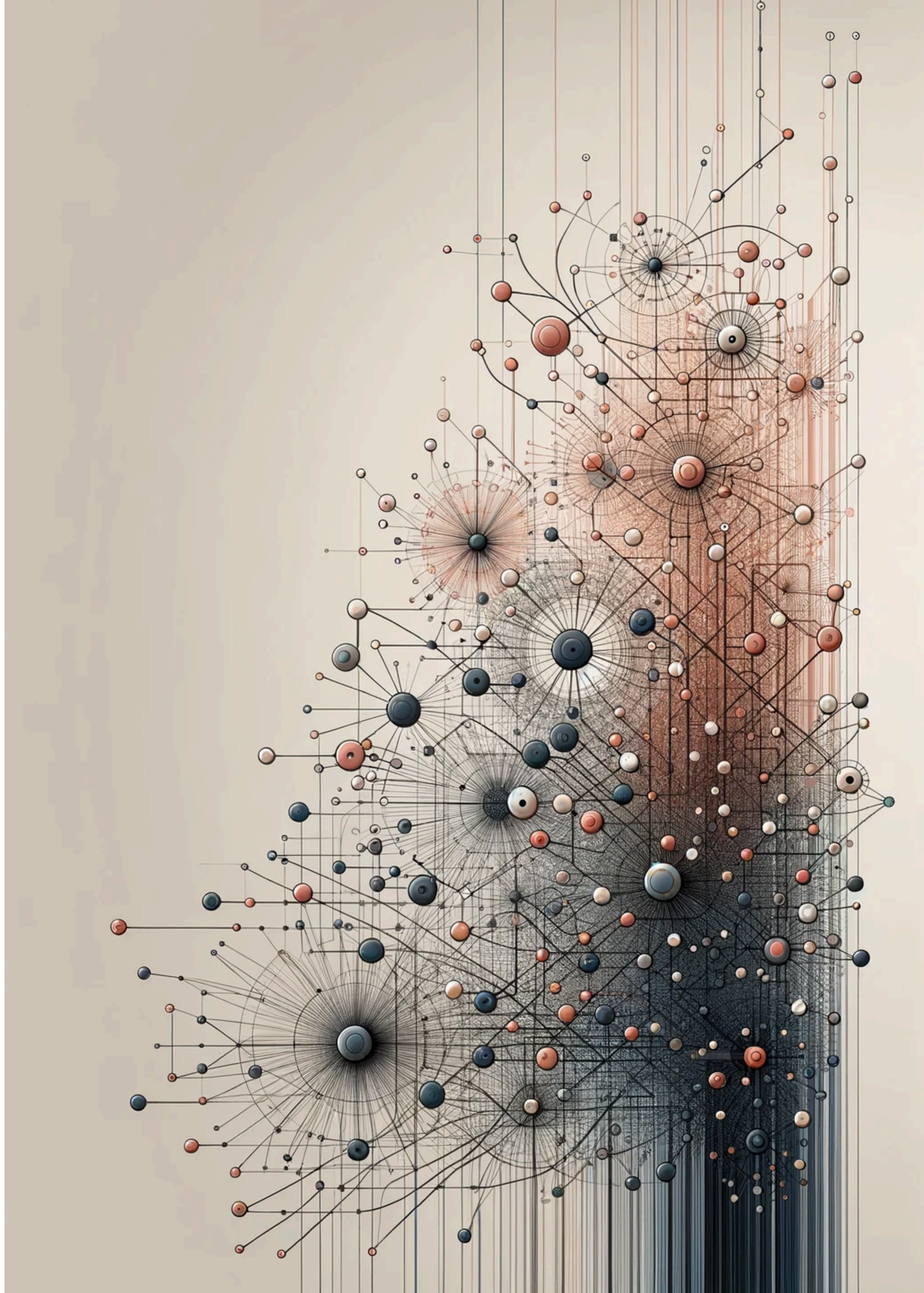
UNIVERSITÀ DI TORINO

# Analisi e Visualizzazione delle Reti Complesse

## NS16 - Cascading behaviors in networks

Prof. Rossano Schifanella



UNIVERSITÀ DI TORINO

# Agenda

- Diffusion in Networks
- Modeling Diffusion through a Network
- Cascades and Clusters
- Diffusion, Thresholds, and the Role of Weak Ties
- Extensions of the Basic Cascade Model
- Knowledge, Thresholds, and Collective Action
- The Cascade Capacity

# The diffusion of innovations

- Let's consider how new behaviors, practices, opinions, conventions, and technologies spread from person to person through a social network, as people influence their friends to adopt new ideas
- diffusion of innovation: long list of classic studies in the middle of the 20th century
    - focus on informational and network effects
- **direct benefit effects** to understand how technologies spread: incentives to adopt telephone, fax, emails based on friends that already adopted those technologies
- we need to take into account **network effects** on a **local network level**
- Objective of this lecture: **formulating a model for the spread of an innovation through a social network**

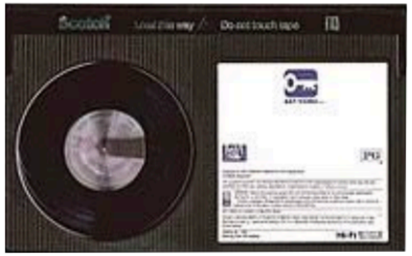
# A diffusion of a new behavior

- Assumption: individuals make decisions based on the choices of their neighbors
  - focus on links
- In this lecture, let's focus on **direct-benefit effects** instead of informational effects
- Natural model introduced by Stephen Morris in 2000
- Reading material:
  - Stephen Morris. Contagion. Review of Economic Studies, 67:57–78, 2000.

# Examples

VHS

vs



Betamax

Blu-ray Disc ™

vs

HD DVD ™

# A networked coordination game

- It is natural to use a **coordination game**
  - each node has a choice between two possible behaviors, A and B
  - players have an incentive to adopt the same behavior

| | | w | |
|---|---|---|---|
| | | A | B |
| v | A | a, a | 0, 0 |
| | B | 0, 0 | b, b |

# A networked coordination game

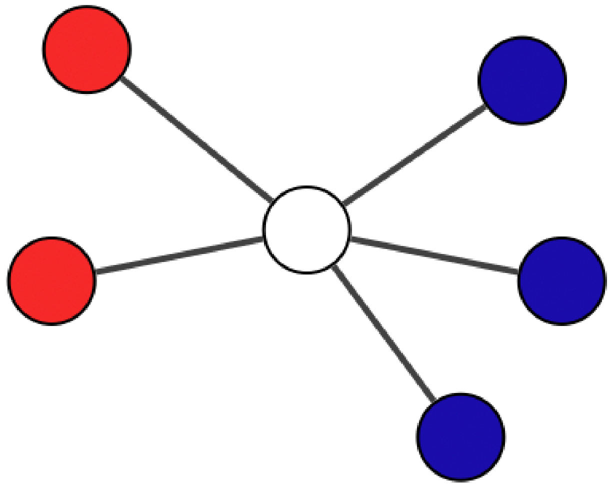
$p$: fraction of neighbors adopting $A$

$1-p$: fraction of neighbors adopting $B$

$d$: is the number of neighbors

the node chooses $A$ if

$$pda \geq (1-p)db \Rightarrow p \geq \frac{b}{a+b} = q$$



# Threshold rule

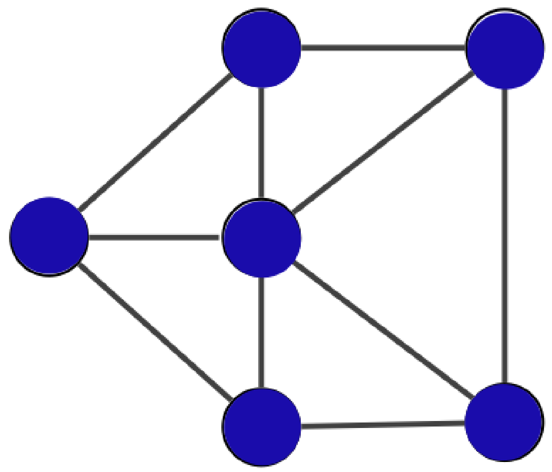
- In order to adopt $A$ or $B$, we just need to check if $p \geq q$.
- Very simple - and myopic - model of individual decision making
- It is a research question to think about richer models that incorporate more long-range considerations

# Cascading behavior

- Two obvious equilibria in the network:
  - one in which everyone adopts $A$
  - another in which everyone adopts $B$
- We want to understand how easy is to get one of these equilibria
- Also, we want to understand if other intermediate equilibria exist and how they look like
- **Assumptions**:
  - everyone is using $B$ at the beginning
  - $S$: small set of initial adopters of $A$
- Will the spread of A make everyone to switch to the new technology, or will the spread stop?
- **Answer:** it depends on the network structure, the choice of nodes in $S$, and the value of $q$.

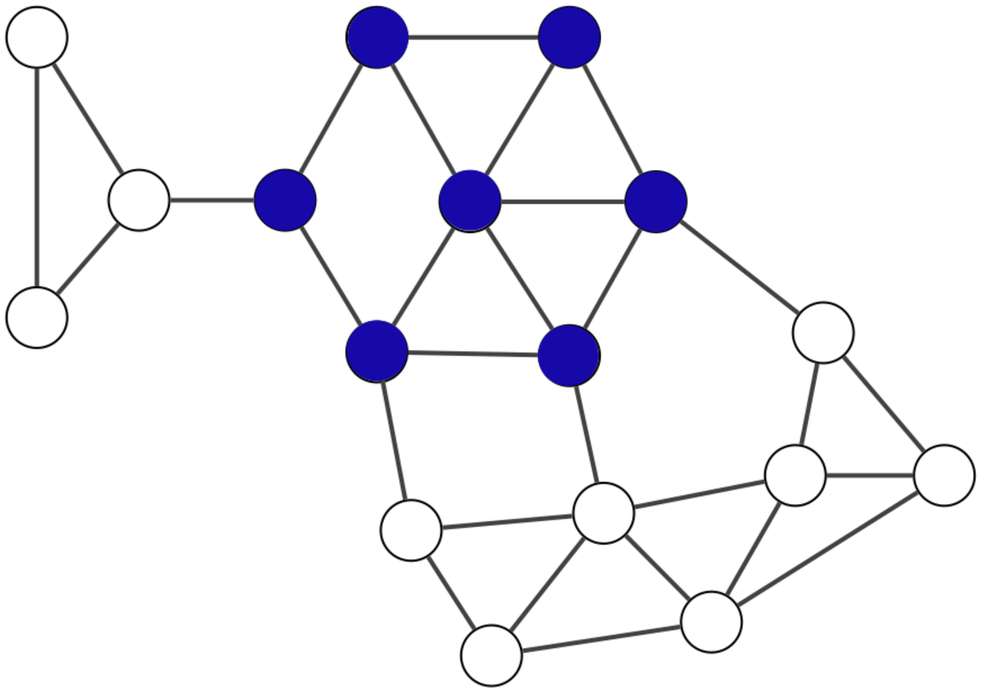
# Example

- $a = 3, b = 2 \Rightarrow q = \frac{2}{5}$
- $S = \{u, v\}$
- Chain reaction: complete cascade



# Example

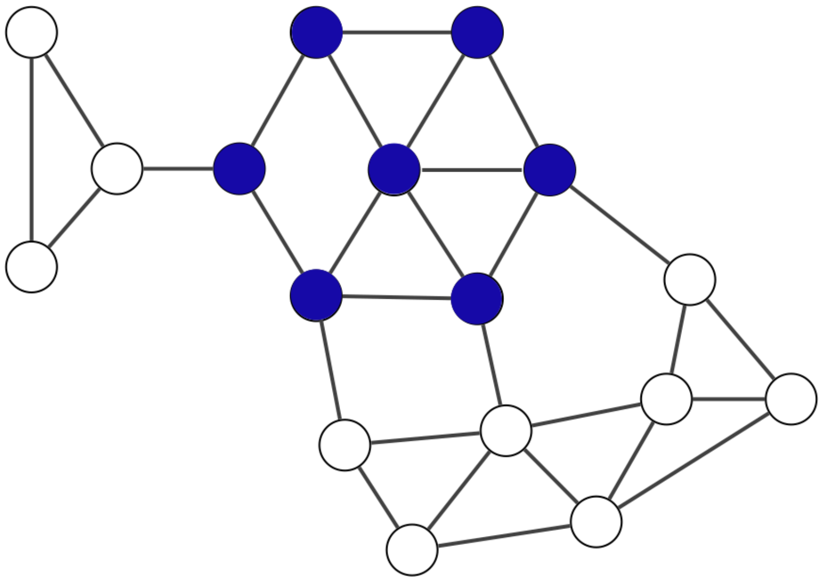
- $a = 3, b = 2 \Rightarrow q = \frac{2}{5}$
- $S = \{u, v\}$
- The diffusion of A stops here: partial cascade



- Consider a set of initial adopters who start with a new behavior $A$, while every other node starts with behavior $B$.
- Nodes then repeatedly evaluate the decision to switch from $B$ to $A$ using a threshold of $q$.
- 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 $q$

## Monotonically spreading

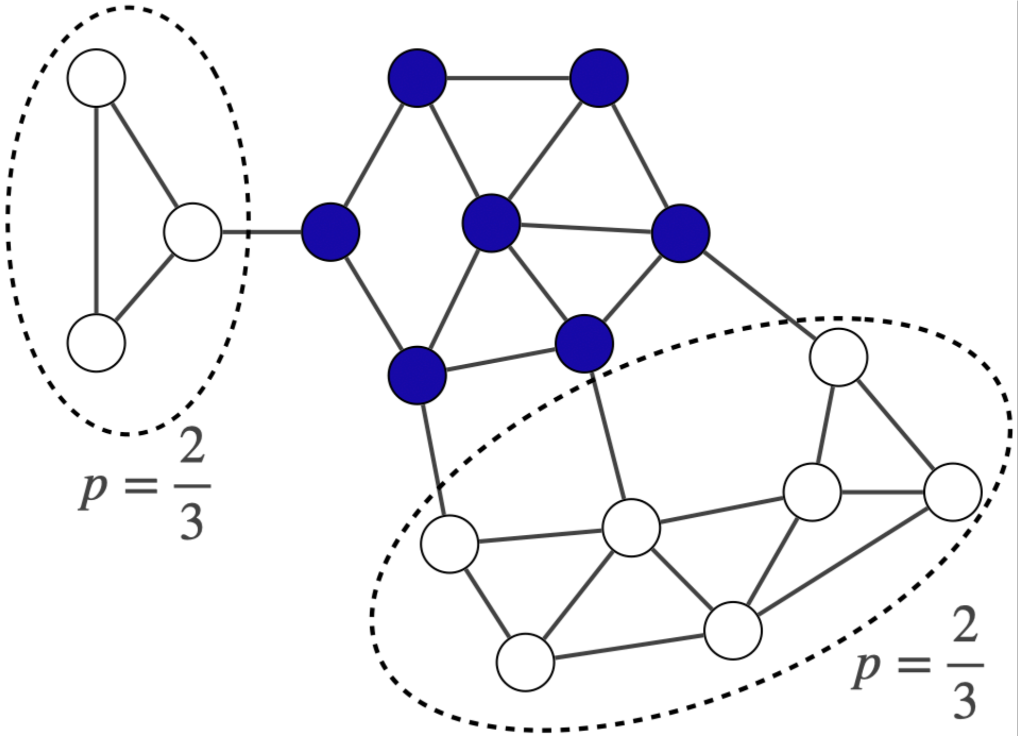
- claim: Adoption of $A$ spreads monotonically
  - Nodes switch $B \to A$, but never back to $B$
- proof (by contradiction)
  - suppose some node switched back from $A \to B$, consider the first node (not in $S$) to do so (at time $t$)
  - earlier at time $t'$ ($t' < t$) the same node switched $B \to A$
  - So at time $t'$ the node was above threshold $q$
  - But up to time $t$ no node switched back to $B$, so the node could only have more neighbors who used $A$ at time $t$ compared to $t'$
  - there was no reason for that node to switch at the first place $\implies$ contradiction!



UNIVERSITÀ DI TORINO

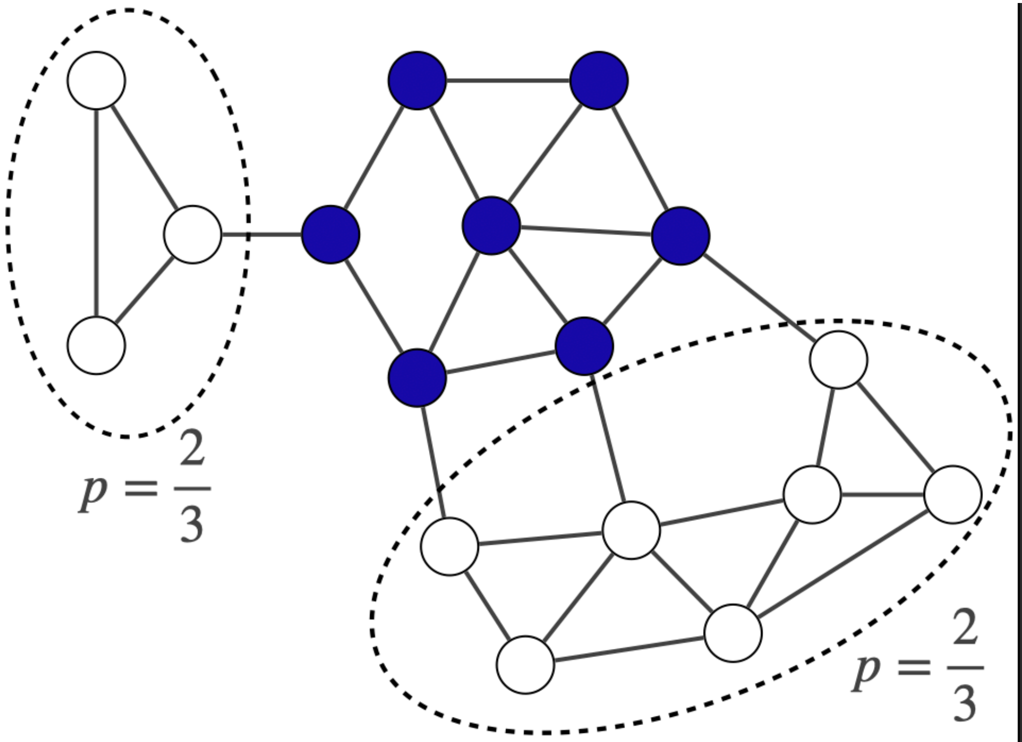
# Stopping cascades

- What prevents cascades from spreading?
    - **Homophily** can serve as a barrier to diffusion: it is hard for innovation to arrive from outside densely connected communities
- Let's try to quantify this intuition:
    - Definition: *cluster of density $p$ is a set of nodes $C$ where **each node** in the set has at least $p$ fraction of edges in $C$*



# Internal cohesion

- Each node in a cluster has a prescribed fraction of its friends residing in the cluster: **cohesion**
- Nodes in the same clusters do not necessarily have much in common
    - Any network is a cluster of density $p = 1$.
    - The union of two clusters of density $p$ is still a cluster of density $p$.
- In fact, clusters in networks can exist simultaneously at different scales.

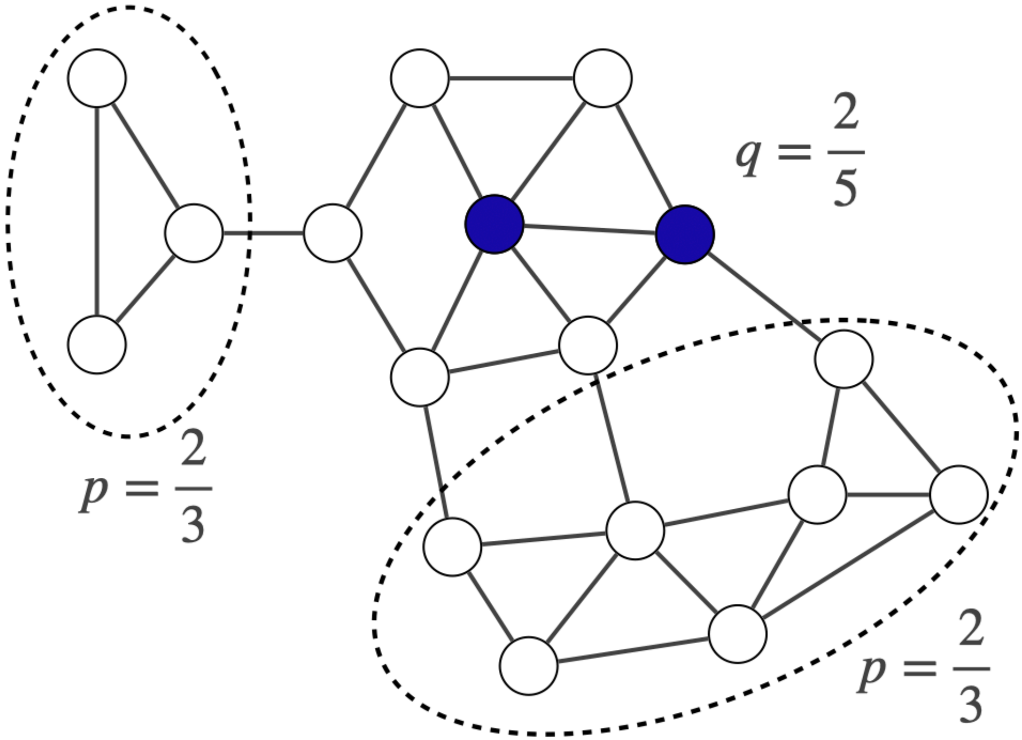


# Relationship between clusters and cascades

- A cascade stops **iff** it runs into a dense cluster
- i.e., **clusters are natural obstacles to cascades**
- Claim: Let $S$ be the set of initial adopters of $A$, with a threshold of $q$

  (i) if the remaining network contains a cluster of density $> (1-q)$, then $S$ cannot cause a complete cascade

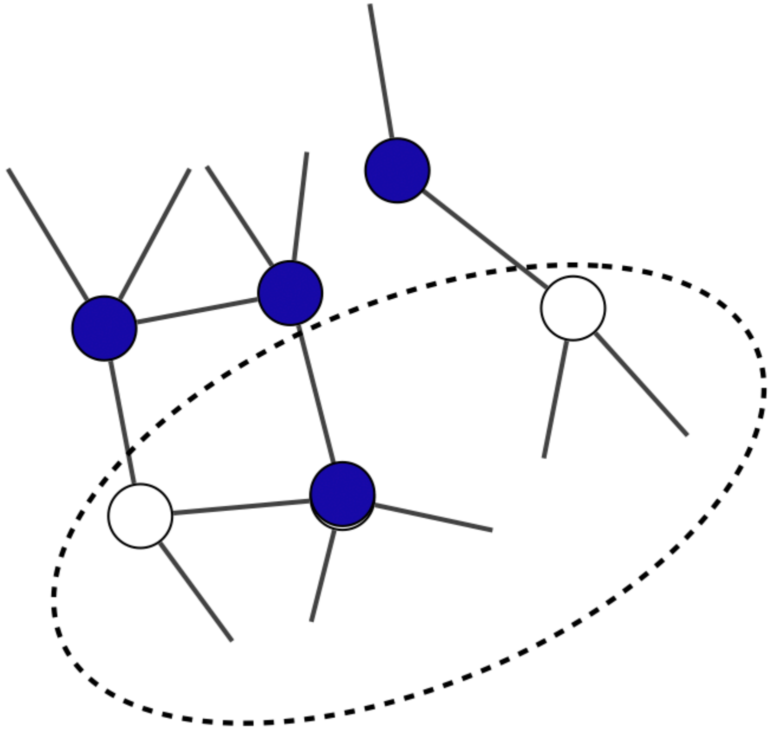
  (ii) if $S$ fails to cause a cascade, then there is a cluster of density $> (1-q)$ in the remaining network



# Clusters implies no cascade

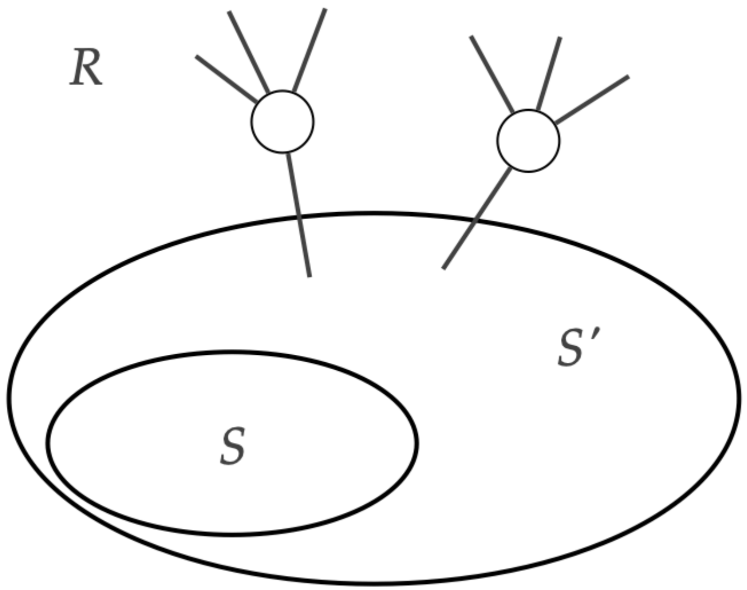
Proof by contradiction:

- some node inside the cluster of density $> (1-q)$ will adopt $A$ at time $t$ (the first to adopt in its cluster);
- this means that the decision was made at time $t-1$, when no other node in the cluster adopted $A$;
- since the cluster has density $> (1-q)$, less than a $q$ fraction of nodes are outside the cluster;
- nodes that could have adopted $A$ were outside the cluster at time $(t-1)$
- it is impossible that a least $$ neighbors of the node adopted $A$: contradiction



# Clusters implies no cascade

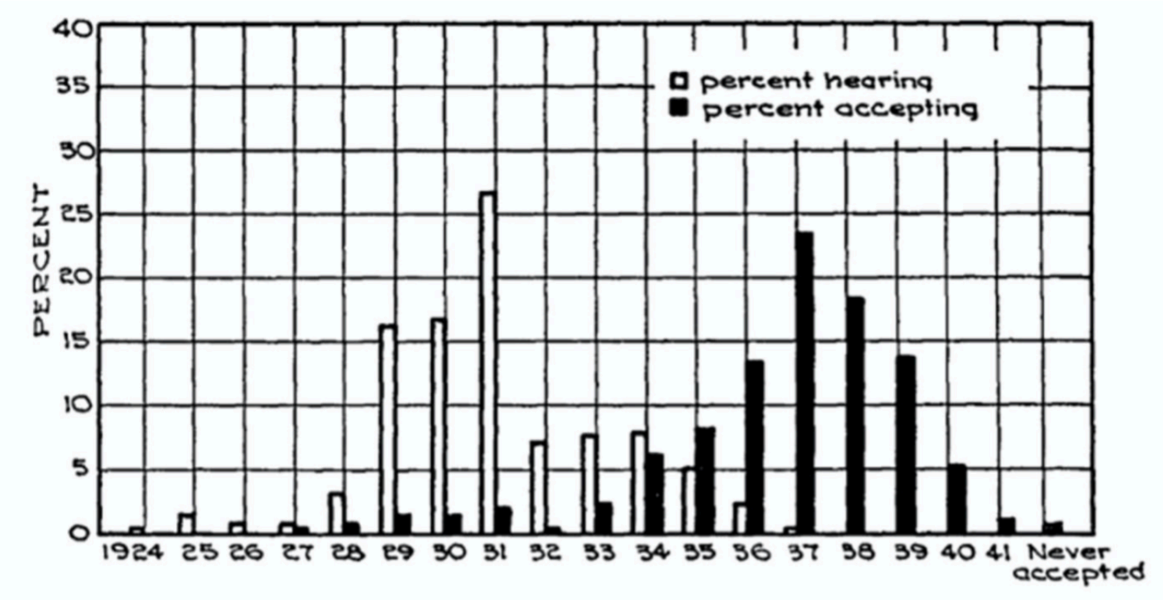
- Let $S$ be the set of initial adopters of $A$
- The spreading process stops: let $S'$ be the maximum set of nodes that switched to $A$
- Let R be the set of nodes still using B at the end of the process
- Pick one node in $R$: it does not want to switch to $A$
  - the fraction of its friends using $A$ is $< q$
  - the fraction of its friends belonging to $R$ is $> (1-q)$
  - this holds for every node in $R$
  - $R$ is a cluster of density $> (1-q)$



# Diffusion, Thresholds, and the Role of Weak Ties

# Viral marketing

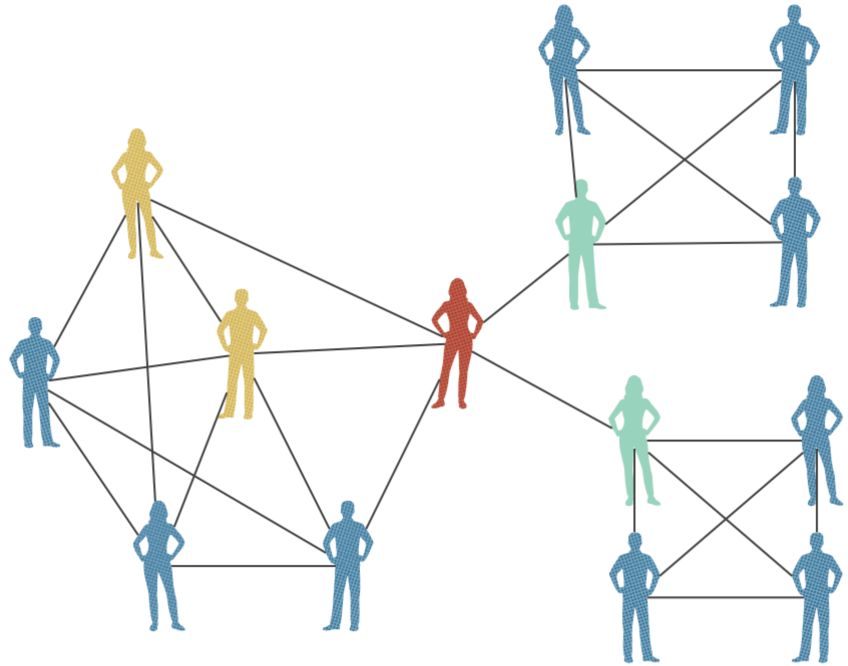
- Lesson learned: tightly-knit communities in the network can work to stop the spread of an innovation
  - We have coexistence of both behaviors (or also competing products)
- In case of coexistence, if a firm can rise the quality of product $A$, than the spread of the new technology can produce a complete cascade
- Alternative strategy (when the firm cannot rise the quality of the product): trying to **convince a small number of key people** in the part of the network still using $B$
  - how to choose key people to trigger a viral marketing strategy?
  - their position in the underlying network can be very relevant - let's explore this now




- In the Ryan-Gross study (1943) on the adoption of hybrid seed corn, it is possible to spot the difference between learning about a new idea and actually deciding to adopt it
- It is like nodes at the border in a cluster: they have been exposed to the idea, but still they decide not to adopt it.
- Reading material: Bryce Ryan and Neal C. Gross. The diffusion of hybrid seed corn in two Iowa commu- nities. Rural Sociology, 8:15–24, 1943.

# The role of weak ties

- Threshold models highlight some important implications of the **strength of weak ties** theory
- They receive very fresh ideas from other communities; not enough for adoption and spread (try with $q = \frac{1}{2}$)
- Bridges and weak ties are great for spreading rumors or jokes across the network, but not for diffusion of innovation or social mobilization
- Strong ties can have more significant role for others in the community to take actions
- Reading material: Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. American Journal of Sociology, 113:702–734, 2007.



# Simple vs Complex Contagion

- **Simple contagion processes:** The activation of a node depends if the threshold condition is satisfied or not;
    - chances do not play a role in the **(linear) threshold model**: it is based on the concept of peer pressure
    - In the **independent cascade models**: we focus on one-to-one social influence and one node can have a chance of activating another node
- **Complex contagion processes:** Each new node that tries to activate a neighbor has greater influence than previous ones

UNIVERSITÀ DI TORINO

# Health behavior and artificial communities

- Centola studied the spread of a health behavior through a network-embedded population by creating an Internet-based health community, comprising 1,528 participants recruited from health-interest World Wide Web sites.
- Each participant created an on-line profile, including an avatar, a username, and a set of health interests. They were then matched with other participants in the study – referred to as **health buddies** – as members of an on-line health community.
- Participants made decisions about whether or not to adopt a health behavior based on the adoption patterns of their health buddies. Arriving participants were randomly assigned to one of two experimental conditions – a clustered lattice network and a random network - which were distinguished only by the topological structure of the social networks.
- Reading material: D. Centola, The Spread of Behavior in an Online Social Network Experiment, Science 03 Sep 2010: 1194-1197

# Results of Centola's experiment

**Simple contagion:** a single contact with an “infected” individual can be sufficient to transmit the behavior. Under "the strength of weak ties" hypothesis, random networks with small world topologies will spread a social behavior farther and more quickly than a network in which ties are highly clustered

Centola investigated the effects of network structure on diffusion by studying the spread of health behavior through artificially structured online communities

- **Complex contagion:** the competing hypothesis argues that when behaviors require social reinforcement, a network with more clustering may be more advantageous, even if the network as a whole has a larger diameter.

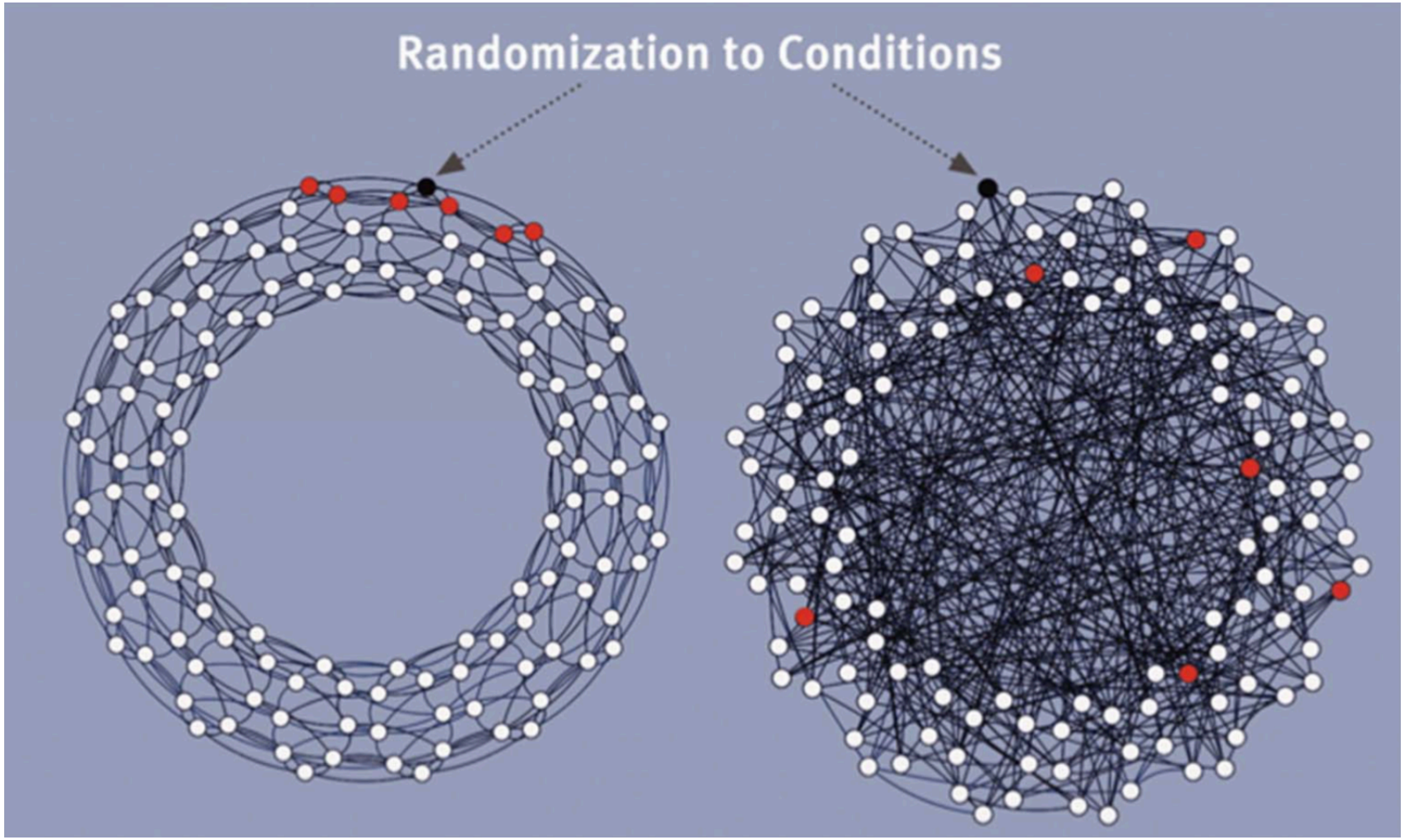
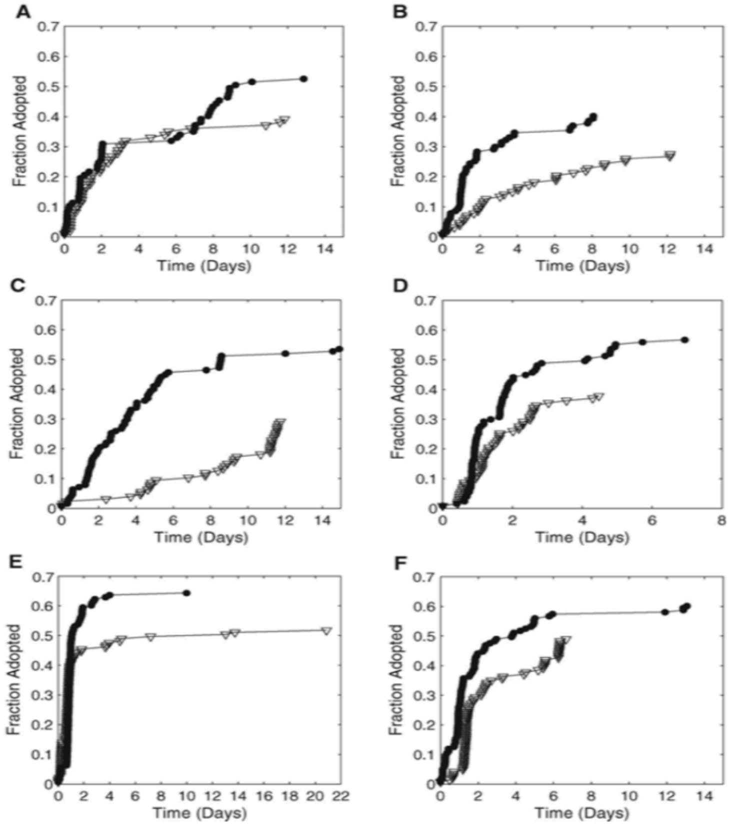
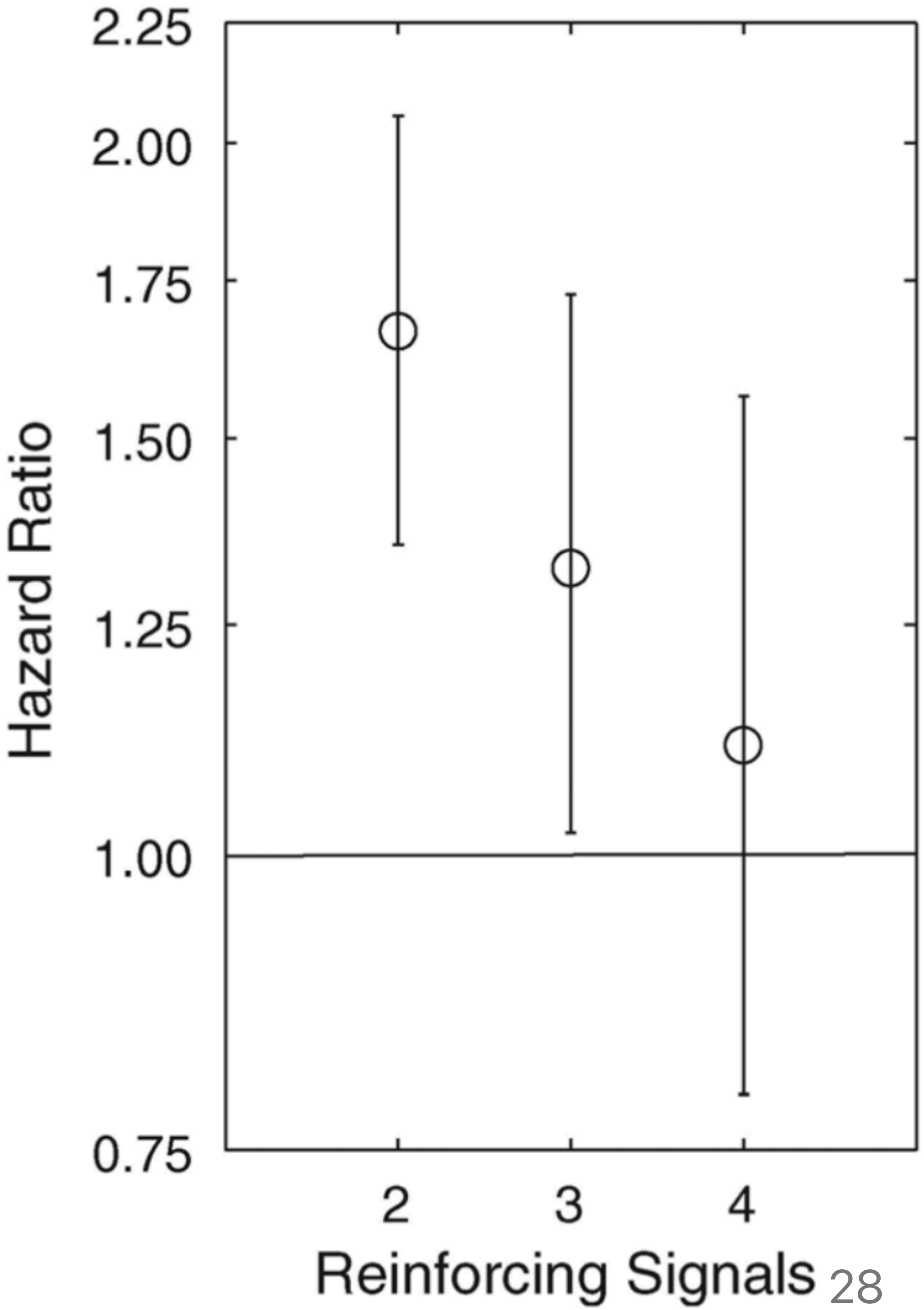

Randomization to Conditions

**Fig. 2 Time series showing the adoption of a health behavior spreading through clustered-lattice (solid black circles) and random (open triangles) social networks.**



- The hazard ratio $g$ indicates that the probability of adoption increases by a factor of $g$ for additional signals compared to the baseline hazard of adoption from receiving a single signal ($g = 1$)

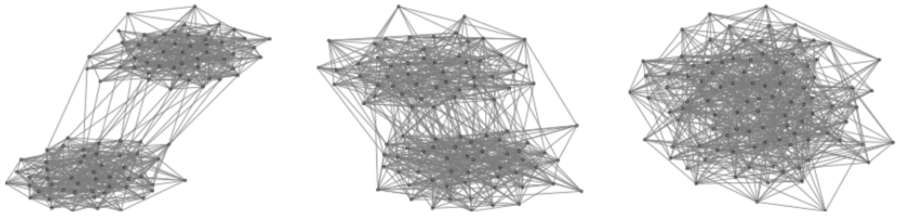


# Findings

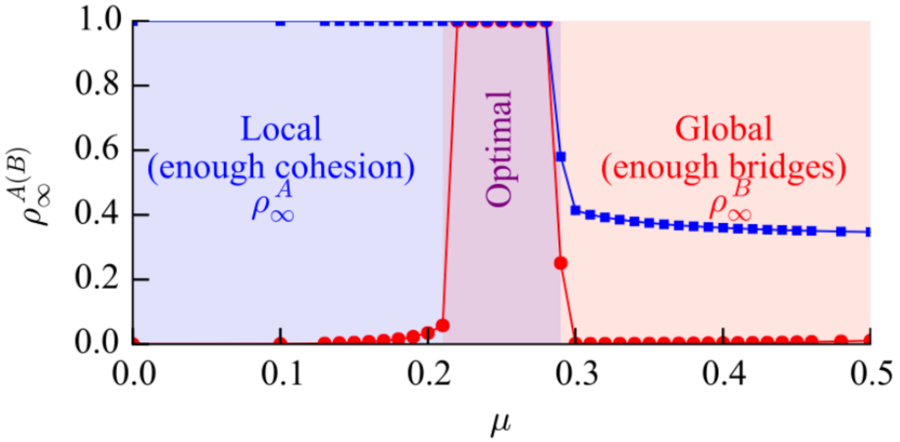
- Individual adoption improved by reinforcing signals that come from clustered social ties
- This individual-level effect translates into a system-level phenomenon whereby large scale diffusion can reach more people, and spread more quickly, in clustered networks than in random networks.
- While locally clustered ties may be redundant for simple contagions, like information or disease, they can be highly efficient for promoting behavioral diffusion.

Centola conjectures that public health interventions aimed at the spread of new health behaviors (e.g., improved diet, regular exercise, condom use, or needle-exchange) may do better to target clustered residential networks than the casual contact networks across which disease may spread very quickly – particularly if the behaviors to be diffused are highly complex, e.g., because they are costly, difficult to do, or contravene existing norms.

# Finding the optimal clustering for spreading



(a)$\mu$=0.03 (b)$\mu$=0.12 (c)$\mu$=0.3




- The tradeoff between intra- and inter-community spreading.
  - Stronger communities (small $\mu$) facilitate spreading within the originating community (local) while weak communities (large $\mu$) provide bridges that allow spreading between communities (global). There is a range of values that allow both (optimal).
  - $\rho^A_\infty$ ($\rho^B_\infty$) represents the final density of active nodes in community $A$ ($B$)
  - Reading Material: Azadeh Nematzadeh, Emilio Ferrara, Alessandro Flammini, and Yong-Yeol Ahn, Optimal Network Modularity for Information Diffusion Phys. Rev. Lett. 113, 088701

# Extensions of the Basic Cascade Model

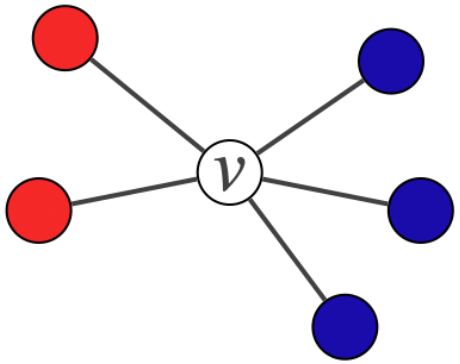
## Heterogeneous thresholds

- Let's suppose each person gives values to $A$ and $B$ subjectively

**w**

| **v** | | **A** | **B** |
|---|---|---|---|
| | **A** | $a_v, a_w$ | $0, 0$ |
| | **B** | $0, 0$ | $b_v, b_w$ |

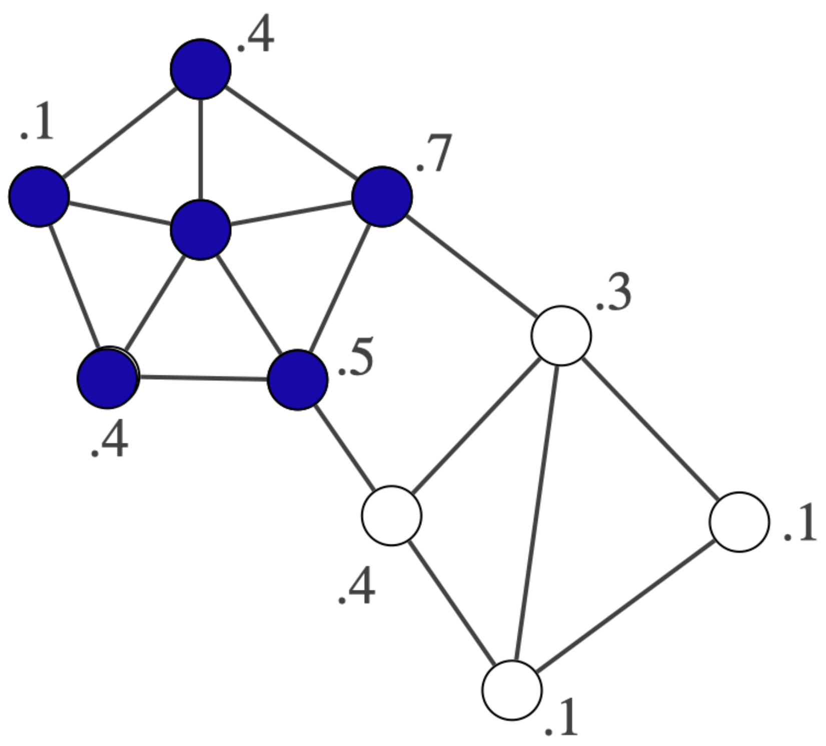
# Heterogeneous thresholds




- $p$ fraction of neighbors adopting $A$
- $1 - p$ fraction of neighbors adopting $B$
- $d$ is the number of neighbors
- The node chooses $A$ if

$$pda_v \geq (1-p)db_v \Rightarrow p \geq \frac{b_v}{a_v + b_v} = q_v$$

- Watts and Dodds: we need to take into account not just the power of influential nodes, but also the extent to which these influential nodes have access to easily **influenceable** people.
- Reformulating the notion of **blocking clusters**: set of nodes for which each node $v$ has a fraction $> (1 - q_v)$ of its friends inside the set.
- The notion of density becomes **heterogeneous** as well: each node has a different requirement for the fraction of friends it needs to have in the cluster.
- Reading material: Duncan J. Watts and Peter S. Dodds. Networks, influence, and public opinion formation. Journal of Consumer Research, 34(4):441–458, 2007.



# Knowledge, Thresholds, and Collective Action

## Integrating network effects

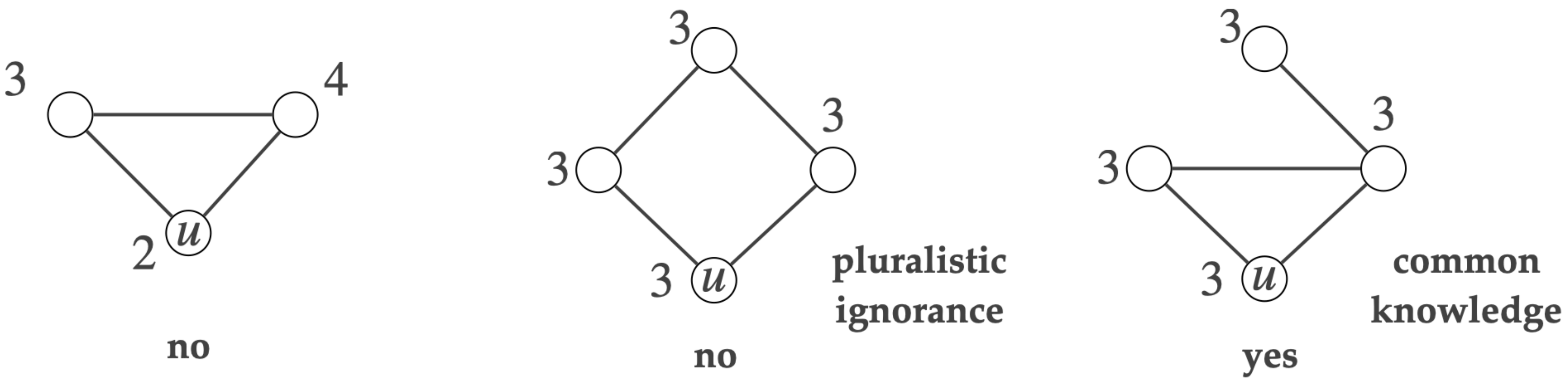
- In many domains you need to integrate network effects at both the **population level** and the **local network level**
  - collective actions and pluralistic ignorance
  - a model for the effect of knowledge on collective actions
  - common knoledge and social institutions

# Collective actions and pluralistic ignorance

- **Objective:** you want to organize a revolt under a repressive regime. Which is the probability to succeed?
- **Collective actions:** benefits only if enough people participate - but you can talk to a very limited number of persons you trust
- **Pluralistic ignorance:** when people have wildly erroneous estimates about the prevalence of certain opinions in the population at large

# A model for the effect of knowledge on collective actions

- Each individual has a personal threshold of $k$: "I will show up for the protest if I am sure that at least $k$ people in total will show up"
- Assumption: every node knows the thresholds of all its neighbors in the network.
- Is it safe for u to join the protest in the following configurations?



# Common knowledge and social institutions

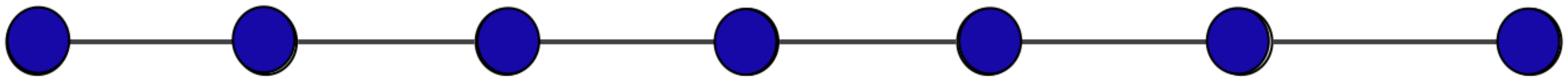
- Some social institutions have the purpose of helping people to achieve common knowledge
- Widely-publicized speech, article in a high circulation newspaper, and so on: people receive the message AND that a lot of other people are receiving the message
- Freedom of the press and freedom of assembly are important social institutions!
- Marketing: commercials during a very popular event (as the famous Apple Macintosh commercial during the 1984 Super Bowl)

# The Cascade Capacity

## Infinite Networks

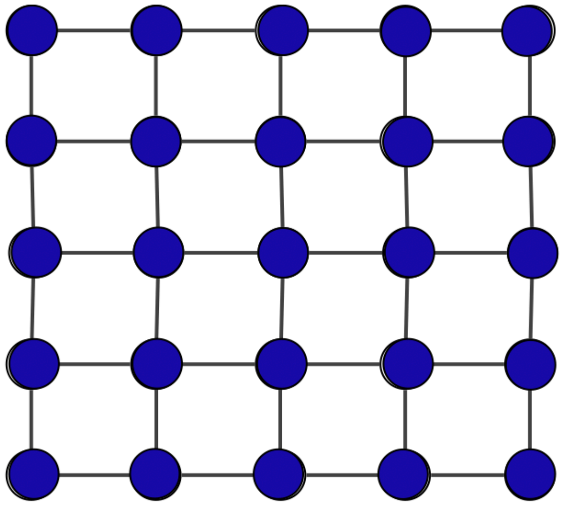
- Understanding how different network structures are more or less prone to cascades
- **Cascade capacity:** the **maximum threshold** at which a small set of initial adopters can cause a complete cascade
- Let's consider **infinite graphs**
  - nodes have finite number of neighbors
- Let's say that a finite set $S$ of initial adopters of $A$ causes a cascade in $G$ with threshold $q$ **if eventually every node in $G$ adopts $A$**
- Example: Infinite path

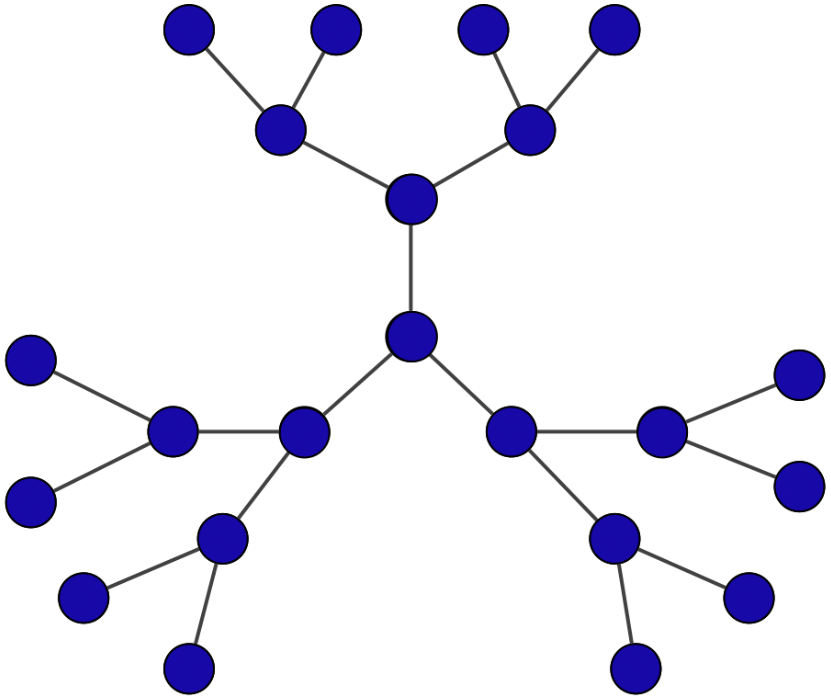
$$q \leq \frac{1}{2}$$



**Infinite grid** $q \leq \frac{1}{4}$

**Infinite tree** $q \leq \frac{1}{3}$





# How large can the cascade capacity be?

Claim: there is no network which the cascade capacity exceeds $\frac{1}{2}$

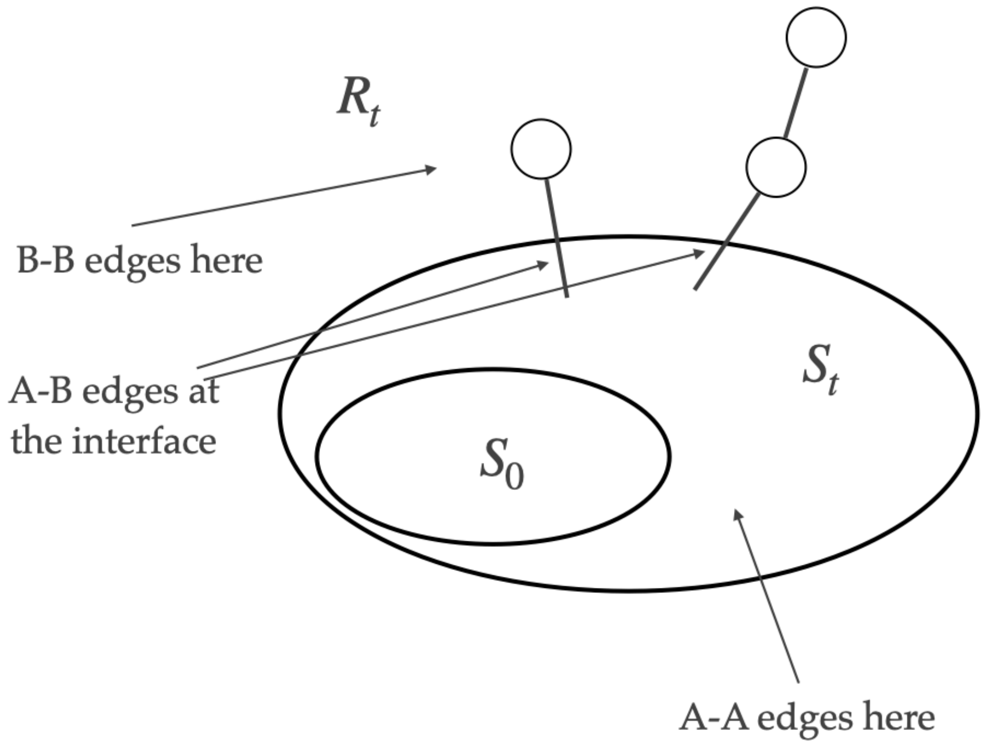
Proof by contradiction:

Let's suppose such $G$ exists with $q > \frac{1}{2}$

We want to find a contradiction, arguing that nodes stop switching from $B \rightarrow A$ after a finite number of steps.

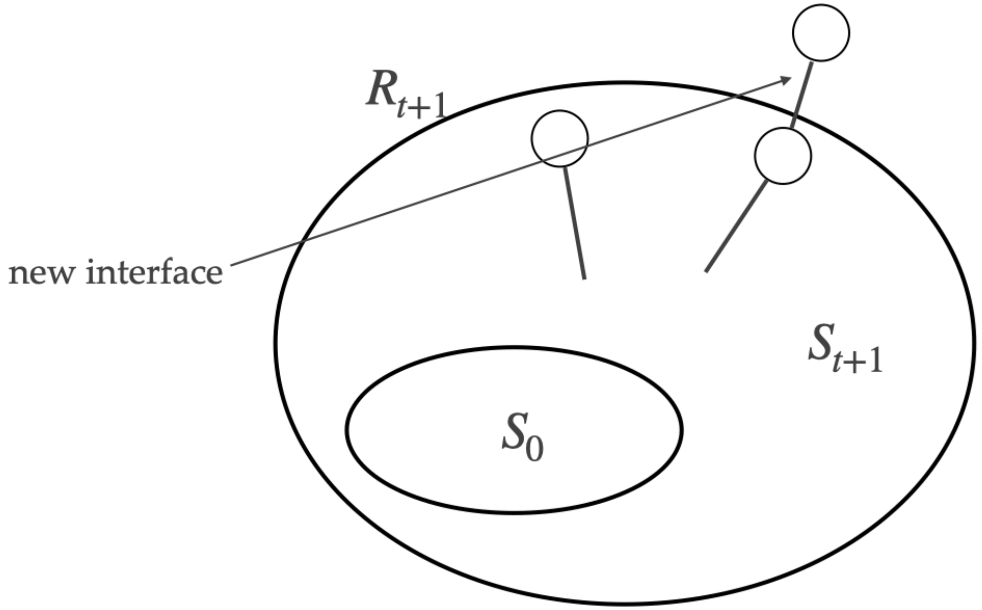
# Focus on the interface

- $S_0$: finite set of initial adopters of $A$
- $S_t$: set of adopters of $A$, potentially larger than $S_{t-1}$
    - size of the **interface** at time $t$: $I_t$
- We show that $I_t < I_{t+1}$
- Hence, if the size of the interface strictly decreases, the diffusion process will terminate after $I_0$ steps



# Size of the interface

- at time $t+1$ some nodes in $R$ switches from $B \to A$
- Focus on a node that switches from $B \to A$
    - it has edges to $B$ nodes
    - and edges with $A$ nodes
    - recall that $q > \frac{1}{2}$, so this node has more edges to nodes in $S$ than to nodes in $R$
    - more edges leaving the interface than edges joining it
    - $\Rightarrow I_{t+1} < I_t$



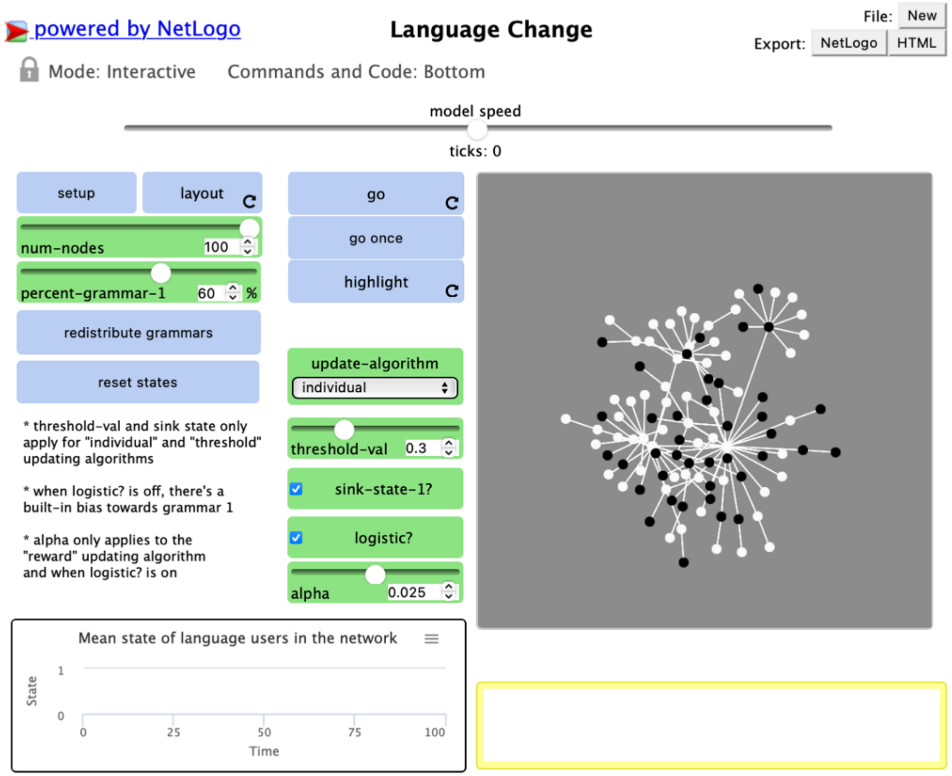
# Observations

- When $q > \frac{1}{2}$ no finite set of nodes can cause a complete cascade in any network
- This threshold corresponds intuitively to a "lower quality" technology: a person will switch to it only if more than 50% of their friends already have.

# Another extension of the model

- Extending the model: allow people to **adopt both $A$ and $B$** (compatibility)
- It is referred as the **bilingual option**
- Although the model is amazingly simple, we have very surprising and complex findings
  - compatibility can be a strategy of a firm to enter in a market and progressively cut out the competitor's product

# Language Change (netlogo model)




- [Play with the model]

# Reading material

[ns2] **Chapter 19 (19.1-19.7)** Cascading behavior in Networks

# Q&A