Information diffusion

Introduction to Network Science

Instructor: Michele Starnini — https://github.com/chatox/networks-science-course

Universitat Pompeu Fabra Barcelona

Content

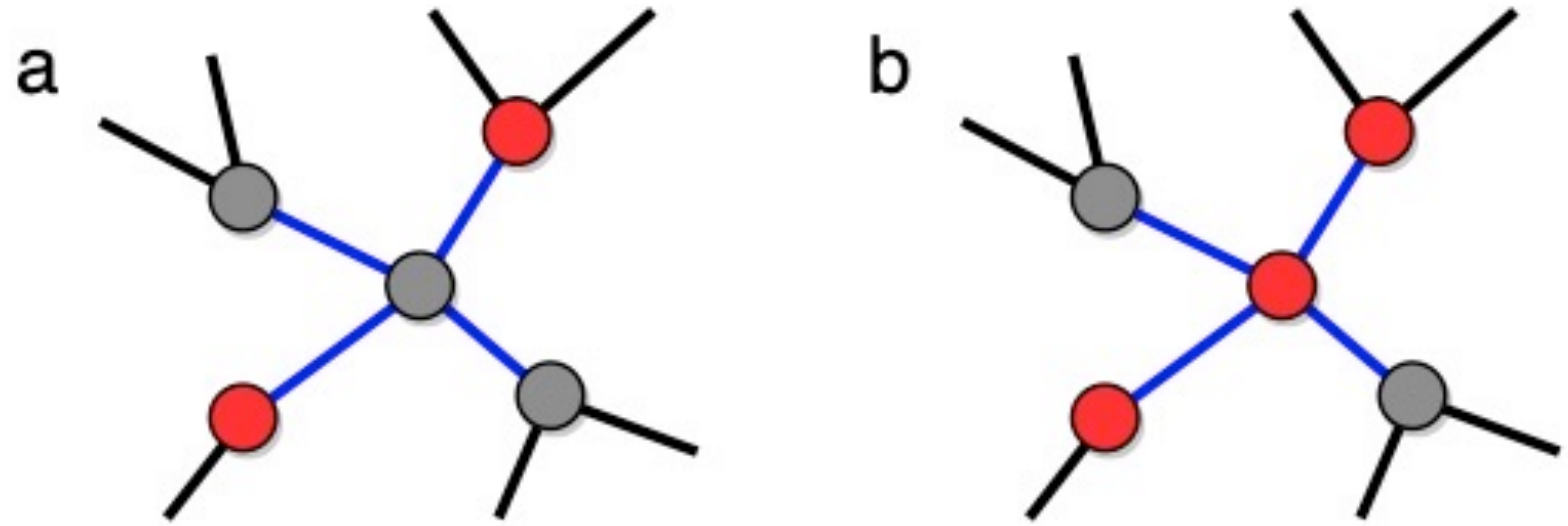
- Threshold models
- Independent cascade models

Information diffusion

- Networks are critical to the diffusion of information:
  - We may buy a new phone because a friend has bought one
  - We may discover a piece of news because friends forward it to us
- **Social contagion:** how individuals' opinions, beliefs, ideas influence or are influenced by their neighbors in social networks

Information diffusion

- Standard setup of models of influence/information spreading:
  - A certain number of nodes (**influencers**) are initially activated (e.g., some individuals adopt an idea, innovation, behavior)
  - Each inactive node is activated (or not) according to some rule that depends on the presence of active neighbors and other factors
- Outcome: **influence cascades**, the activation in sequence of a subset of the nodes in the network
- Options: cascades may involve a handful of nodes or a major proportion of the network (**global cascades**)



Threshold models

Threshold models

- **Principle:** a node can be activated only if the influence exerted on it by its active neighbors exceeds a given value (**threshold**)

- **Linear threshold model:** the influence on a node is defined as a sum over its active neighbors, in which the contribution of each neighbor is given by the weight of the link joining it to the node:

$$I(i) = \sum_{j:\,active} w_{ji}$$

- w_{ji} = weight of the link from *j* to *i*

Threshold models

- **Activation condition:** $$I(i) \geq \theta_i$$

  where θ_i is the threshold of node *i*, indicating its tendency to be influenced

- **On unweighted networks:** $$I(i) = n_i^{on} \geq \theta_i$$

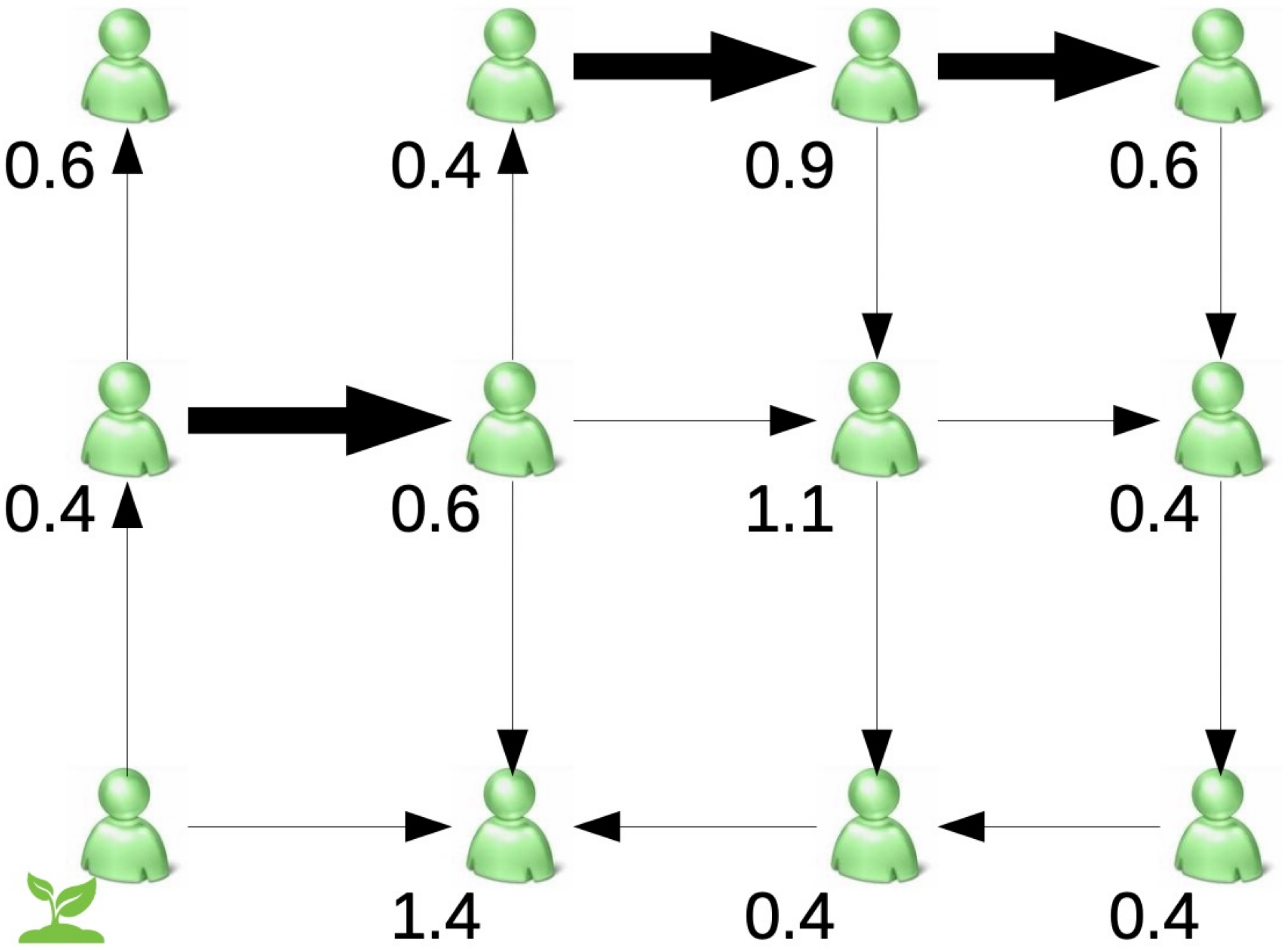
  where n_i^{on} is the number of active neighbors of node *i*

- If all nodes have the same threshold θ: $$I(i) = n_i^{on} \geq \theta$$

Threshold models

- **Model dynamics:**
  - Initially some nodes are activated, for instance a random number or a random proportion
  - All active nodes remain active
  - Each inactive node is activated if the number of active neighbors is at or above its threshold

Exercise: linear threshold model



Thick arrows have weight 1.0

Thin arrows have weight 0.5

Execute linear threshold model starting from seed node

Spreadsheet links: https://upfbarcelona.padlet.org/chato/shyq9m6f2g2dh1bw

Fractional threshold model

- **Principle:** instead of the number of active neighbors, we consider the **fraction**

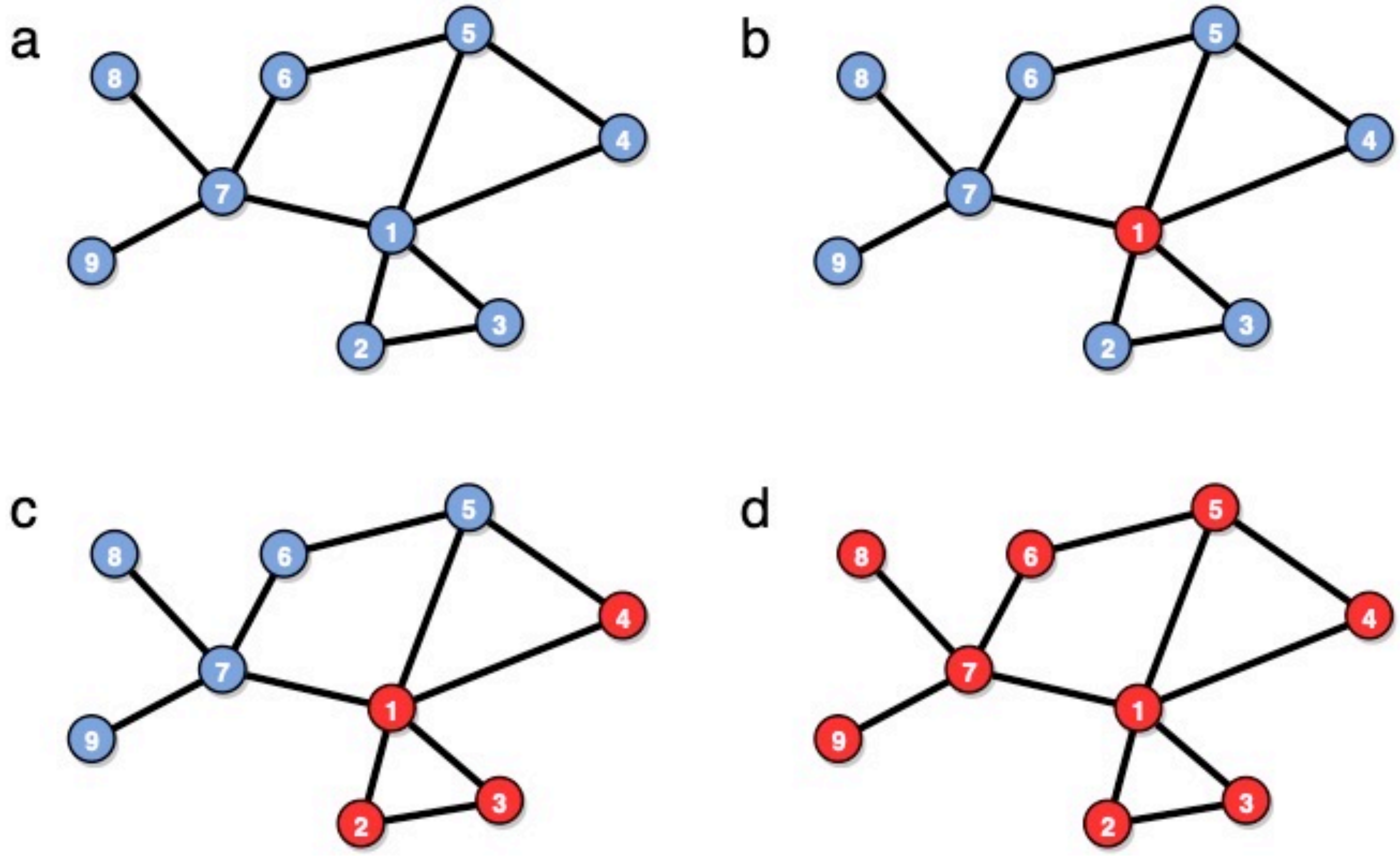
- **Example:** if θ=1/2, at least half of the neighbors must be active!

- **Activation condition:**

$$\frac{n_i^{on}}{k_i} \geq \theta_i$$

where k_i is the degree of node *i*

Fractional threshold model



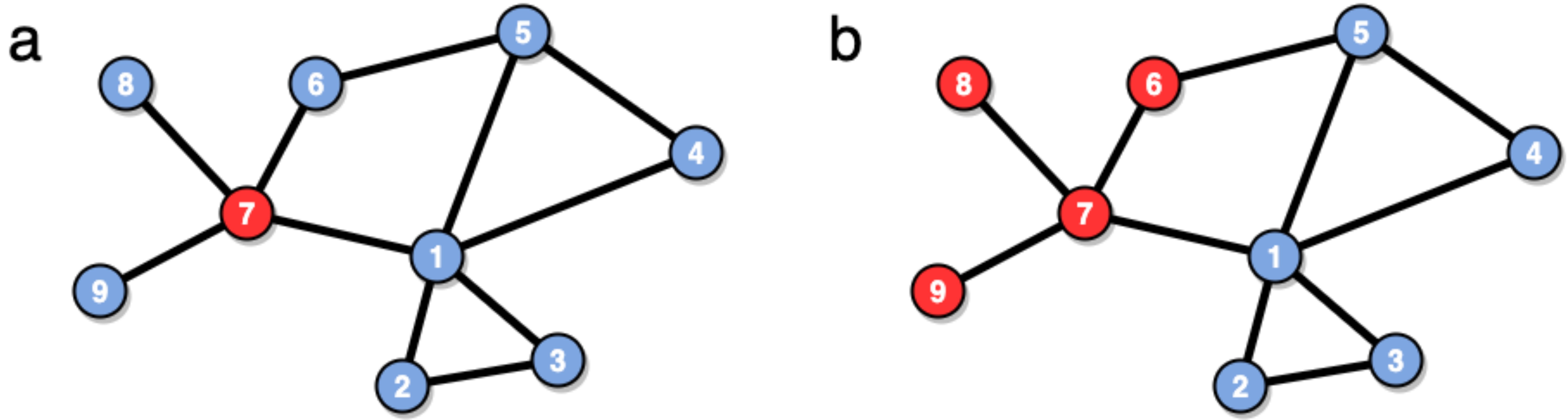
Fractional threshold model

- **Remark:** if the network is sparse, whether or not a cascade is triggered depends on its structure
- Key drivers: **vulnerable nodes**, i.e., the nodes that can be activated by a single active neighbor
- **Condition for a node to be vulnerable:** $k_i \leq \frac{1}{\theta_i}$
- To have global cascades, the number of vulnerable nodes must be sufficiently large!

Fractional threshold model

- Hubs are *usually* very effective influencers
- **Caveat:** being a hub does *not necessarily* imply being a good influencer
- **Importance of active node position:** being in the core of the network makes global cascades more likely than if the node is in the periphery
- **Importance of community structure:** the spread is easier within dense communities than across communities! In fact, to influence another community it is necessary for (some of) its nodes to be activated, which is not easy because they tend to have few neighbors in the activated community

Fractional threshold model



Fractional threshold model

- **Cascade control:** knowing the structure of the network allows us to control the evolution of a cascade. Sometimes small cascades can become large by activating a few carefully chosen nodes
- **Detecting key influencers:** critical for the success of a product or idea!
- **Viral marketing:** social networks are used to promote products

Independent cascade models

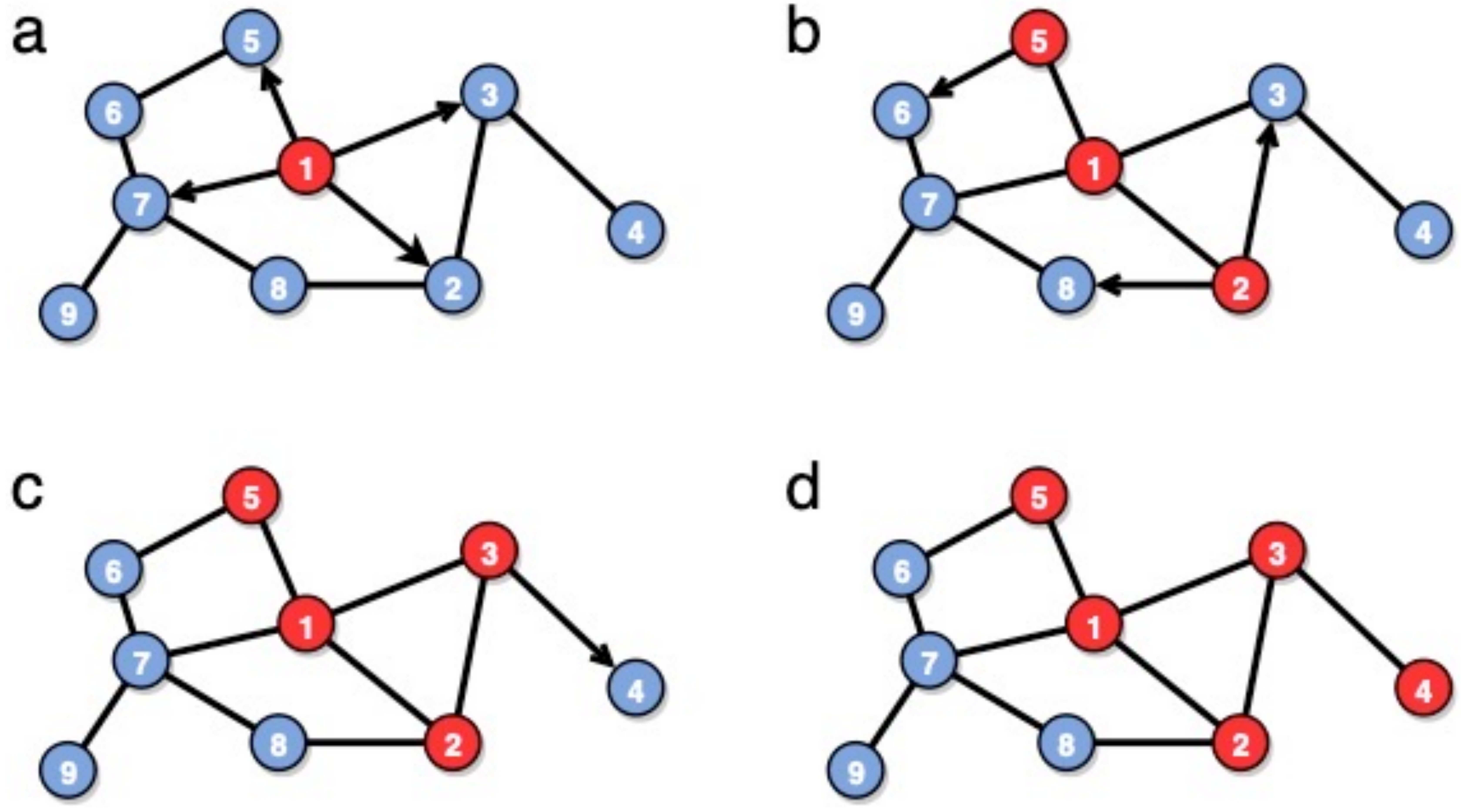
Independent cascade models

- Principle of threshold models: **peer pressure**, the more people try to persuade you, the more likely they will succeed
- **Remark:** social influence often works **one-to-one**, we may be persuaded by a single passionate individual
- **Alternative principle:** each of our contacts has their own influence
- **Independent cascade models** are based on node-node interactions!

Independent cascade models

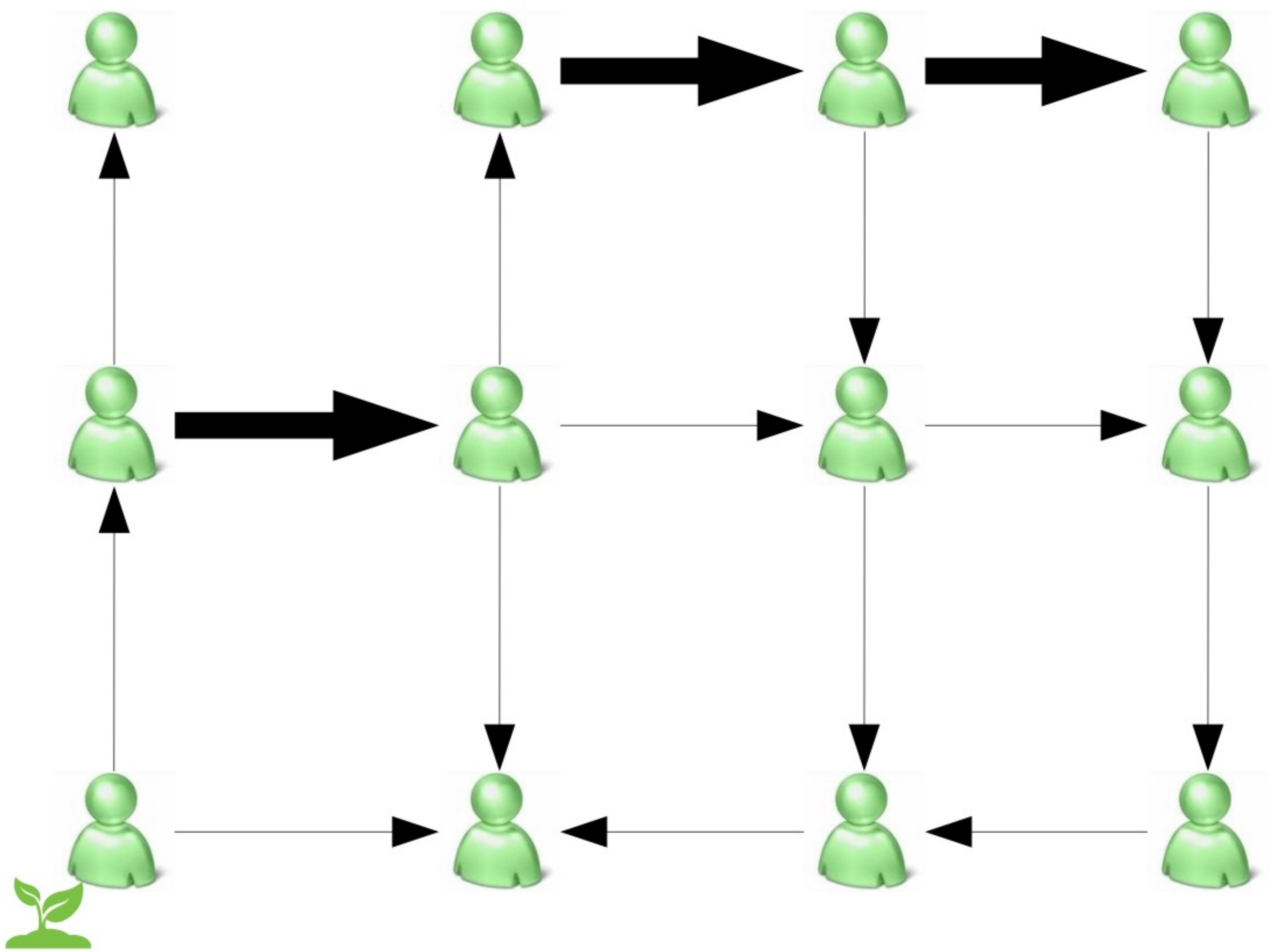
- **Model dynamics:**
  - An active node *i* has a probability p_{ij} to convince its inactive neighbor *j* ($p_{ij} \neq p_{ji}$, in general)
  - All active nodes are considered in sequence: the inactive neighbor *j* of the active node *i* is activated with probability p_{ij}. All inactive neighbors of *i* have one chance to be persuaded by *i*
  - If a node *j* is activated, it has only one chance to activate its inactive neighbors

Independent cascade models



Exercise: independent cascade model

(you need a coin or 1d4)



Thick arrows have probability 0.75

Thin arrows have probability 0.5

Execute independent cascade model starting from seed node

Spreadsheet links: https://upfbarcelona.padlet.org/chato/shyq9m6f2g2dh1bw

Independent cascade models

- **Remark**: the more active neighbors, the more likely a node will be activated
- **Independent cascade versus threshold models**:
  - Threshold models focus on the inactive nodes, independent cascade models on the active ones
  - Threshold models are (usually) **deterministic**: the dynamics depends on whether the threshold condition is satisfied or not
  - Independent cascade models are **probabilistic**: nodes are activated with a given probability —> it is more **difficult to control a cascade!**

Information diffusion

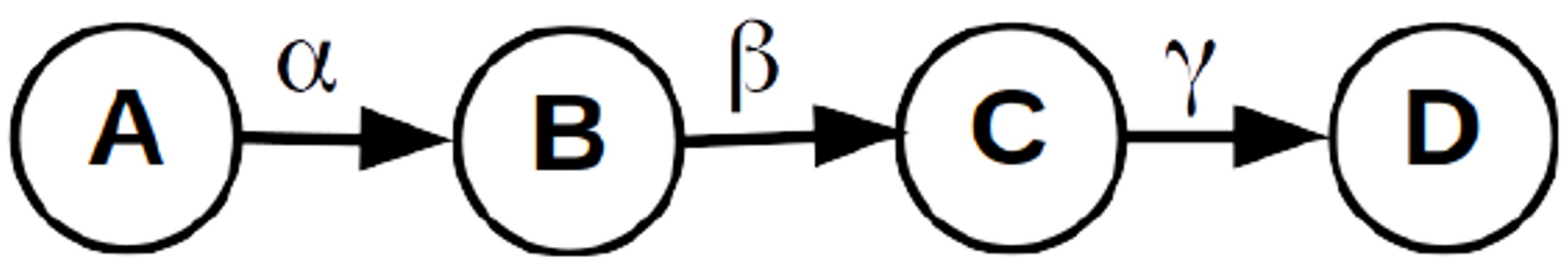
- **Problem**: models are too simple to be realistic
- **Solution**: more sophisticated variants!
- **Example**:
  - Probabilistic version of threshold model, in which the chance of being activated grows with the number of active neighbors (instead of the usual yes/no dynamics)
  - Similar to independent cascade model, except that the active neighbors **do not exert influence independently of each other!**
- **Complex contagion:** each new person exposing us to a new idea or product has greater influence than the previous ones!

Summary

Things to remember

- Linear threshold model
- Fractional threshold model
- Independent cascade model
- Practice executing these models in small graphs by hand

Practice on your own



- Consider the graph on the top-right, including the infection probabilities indicated in the edges: α, β and γ. Let X_i be the expected number of nodes infected under the Independent Cascade Model for an infection starting at node i, including the node initially infected.

- For instance, if an infection starts from node B, the probability that the number of nodes infected is 2 is $P(X_B = 2) = \beta \cdot (1 - \gamma)$. This is because for the infected to be 2 we need the infection from B to C to succeed and the infection from C to D to fail.

- Remember that the expectation of a variable X is $E[X] = x \cdot P(X = x)$, where the summation is done over the possible values x that the variable can take.

- 1. What is $E[X_C]$ as a function of γ?

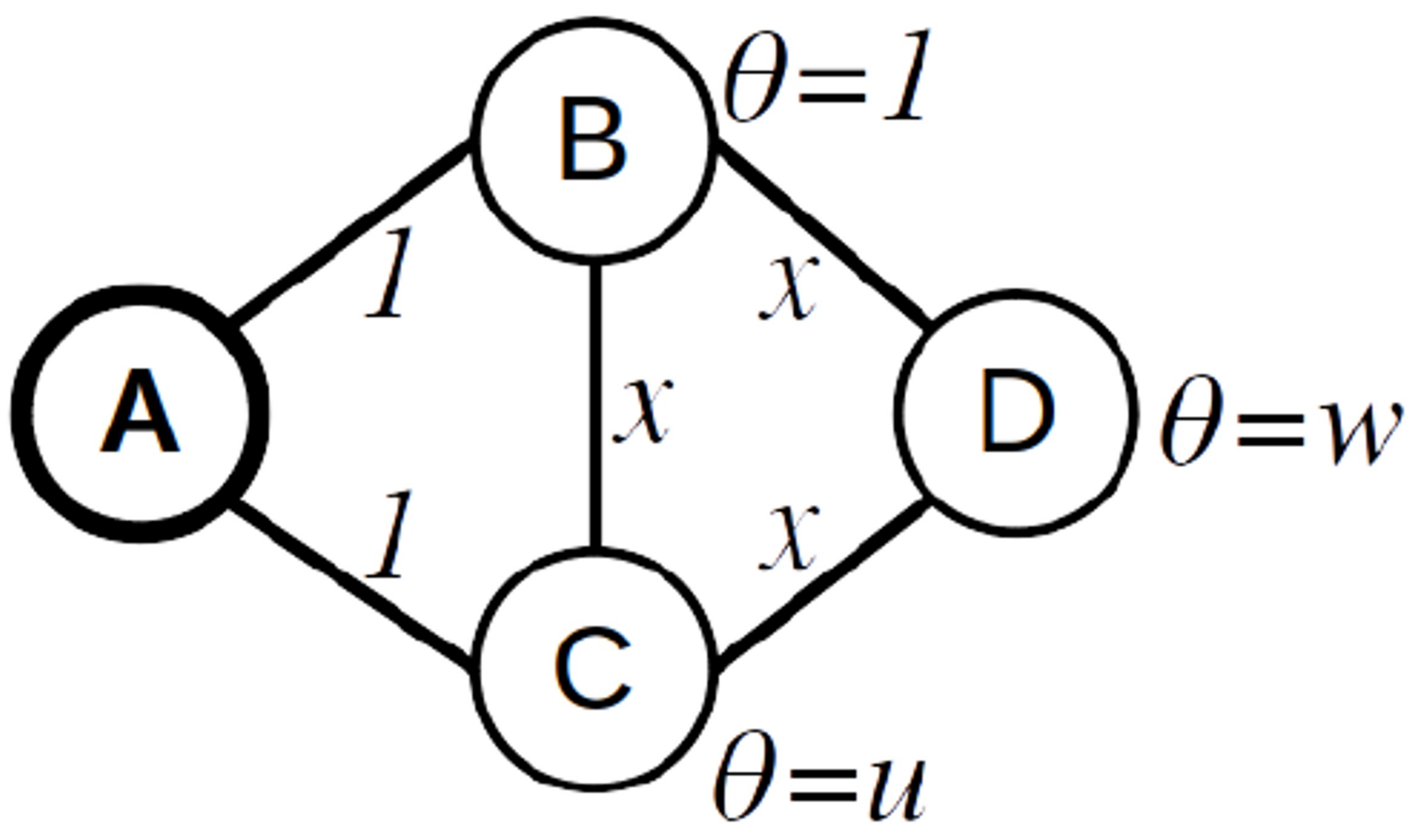
- 2. What is $E[X_A]$ as a function of α, β, γ?

Practice on your own (cont.)

- Consider this graph and the Linear Threshold model executed on it, starting from seed node A.

- The influence weights are written next to the edges, and the thresholds θ are written next to the nodes.

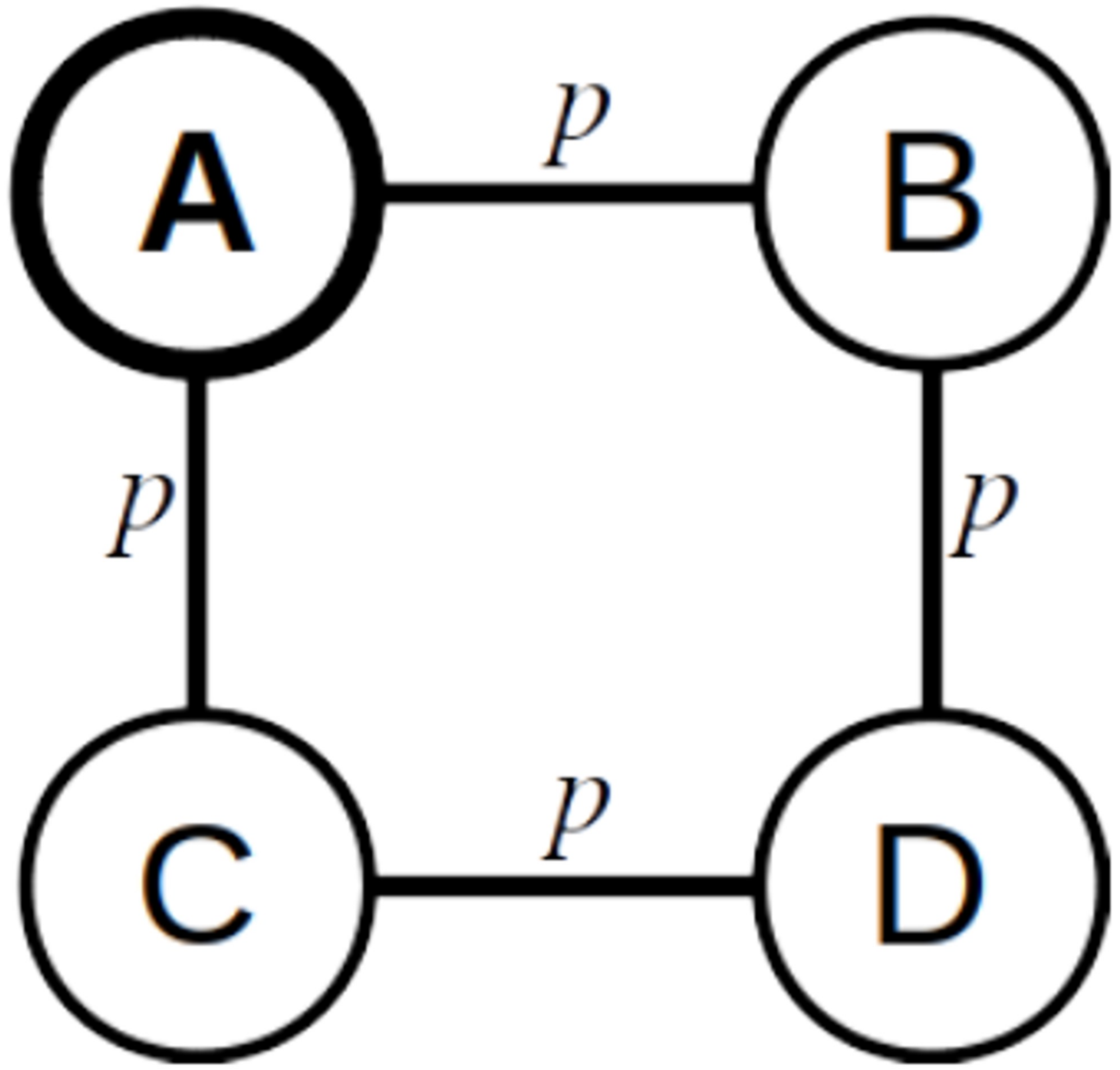
- Indicate what is the range of values of x for node C to be infected, but not node D. Justify briefly your answer.



Practice on your own (cont.)

- Consider the graph on the right an the Independent Cascade model executed on it, starting from seed node A.
- The contagion probability of all edges is *p*
- Indicate what is the probability that at the end of the process:

1. Only node A is infected:
2. Only nodes A, B are infected:
3. Only nodes A, B, C are infected:
4. Only nodes A, B, D are infected:



Sources

- D. Easley and J. Kleinberg (2010). Networks, Crowds, and Markets — Chapter 19
- C. Castillo, W. Chen, L. V. S. Lakshmanan (2012). Information and Influence Spread in Social Networks, KDD Tutorial.
- URLs in the footer of slides