Models of Influence

Introduction to Network Science
Carlos Castillo
Topic 25

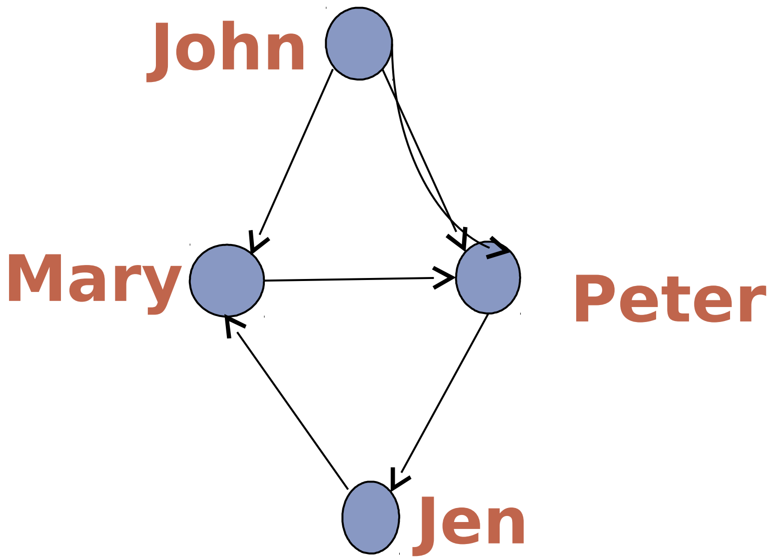
Universitat Pompeu Fabra Barcelona

Sources

- Easley and Kleinberg (2010): Networks, Crowds, and Markets Ch 19
- Carlos Castillo, Wei Chen, Laks V. S. Lakshmanan (2012): Information and Influence Spread in Social Networks, KDD Tutorial.
- Carlos Castillo (2017): Social influence slides

What are our observables?

<u>Graph</u>: users, links/ties



<u>Log</u>: user, action, time

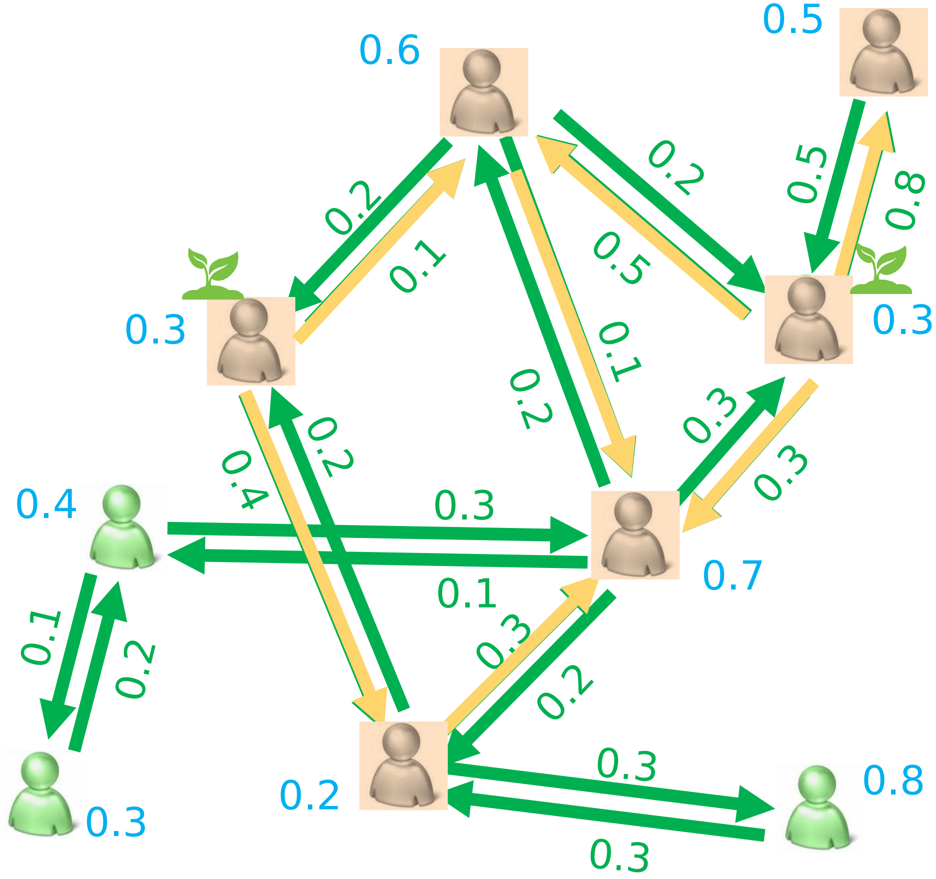
User	Action	Time
John	Rates with 5 stars "*The Artist*"	June 3rd
Peter	Watches "*The Artist*"	June 5th
Jen	…	…

Two main models

- Linear threshold model
- Independent cascade model

Linear threshold model

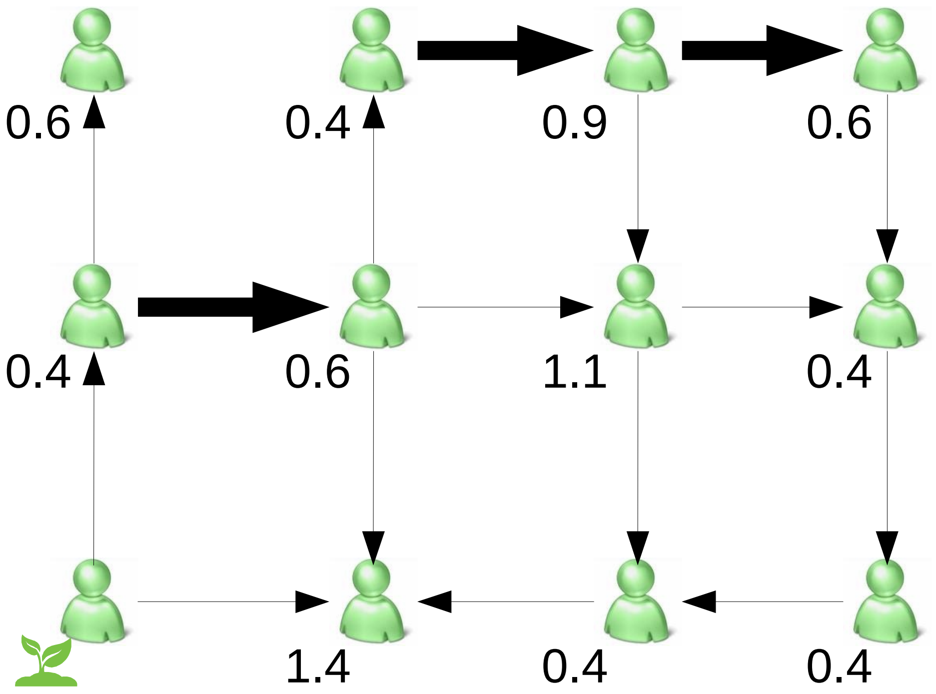
- Nodes have thresholds
- Arcs have weights
- Nodes that receive weighted influence equal or above their threshold become active



[Kempe, Kleinberg and Tardos, KDD 2003]

Exercise

Add your results to this
Google Spreadsheet
(Sheet: Linear Threshold)



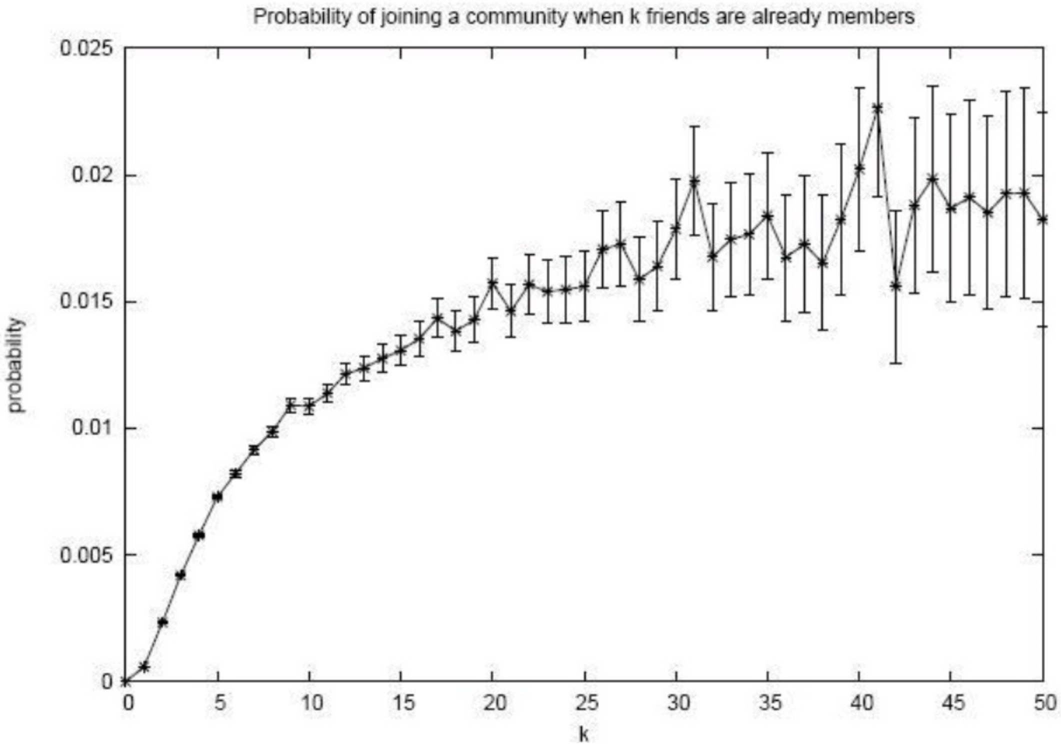
Thick arrows have weight 1.0

Thin arrows have weight 0.5

Execute linear threshold model starting from seed node

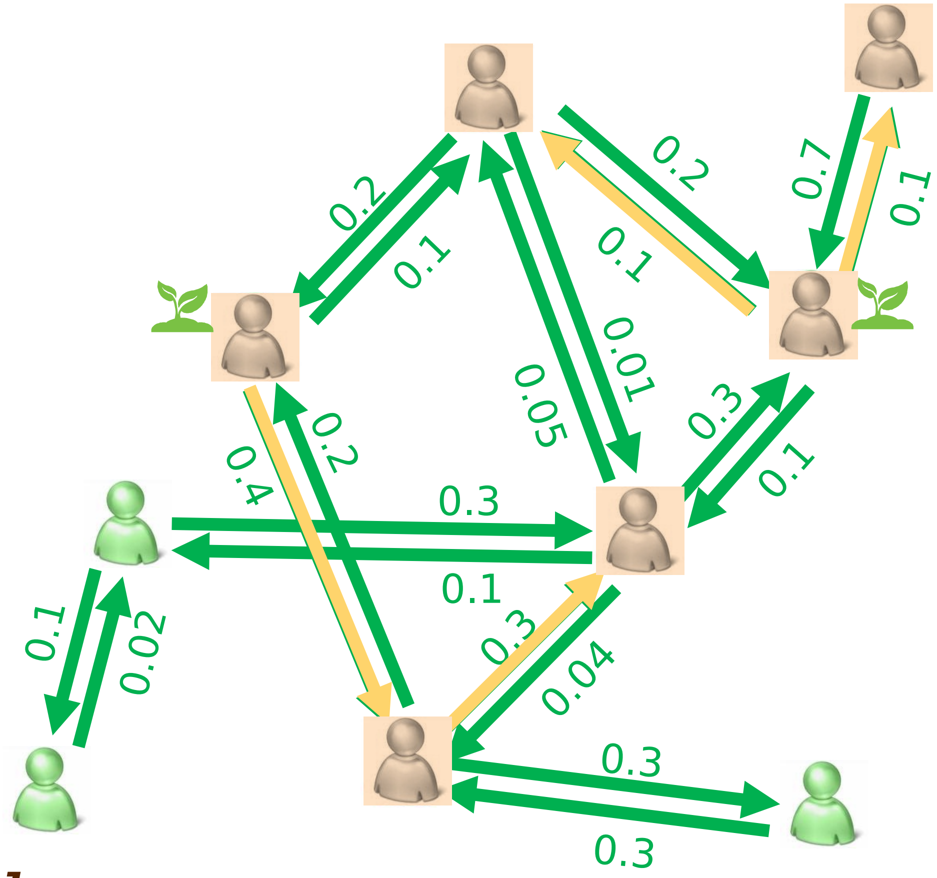
Linear threshold model

Is the linear threshold model compatible with this observation?



Independent cascade model

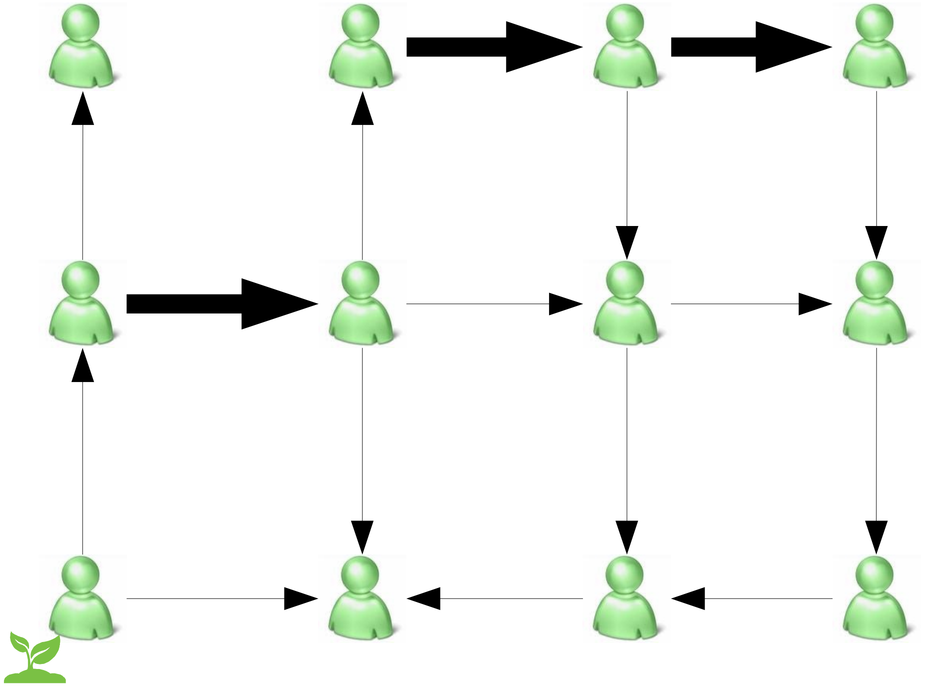
- No thresholds
- Each node, when activating, has one chance of activating each of their neighbors
- Probability of succeeding represented by arc weights



[Kempe, Kleinberg and Tardos, KDD 2003]

Exercise
(you need a coin or 1d4)

Add your results to this
Google Spreadsheet
(Sheet: Independent Cascade)



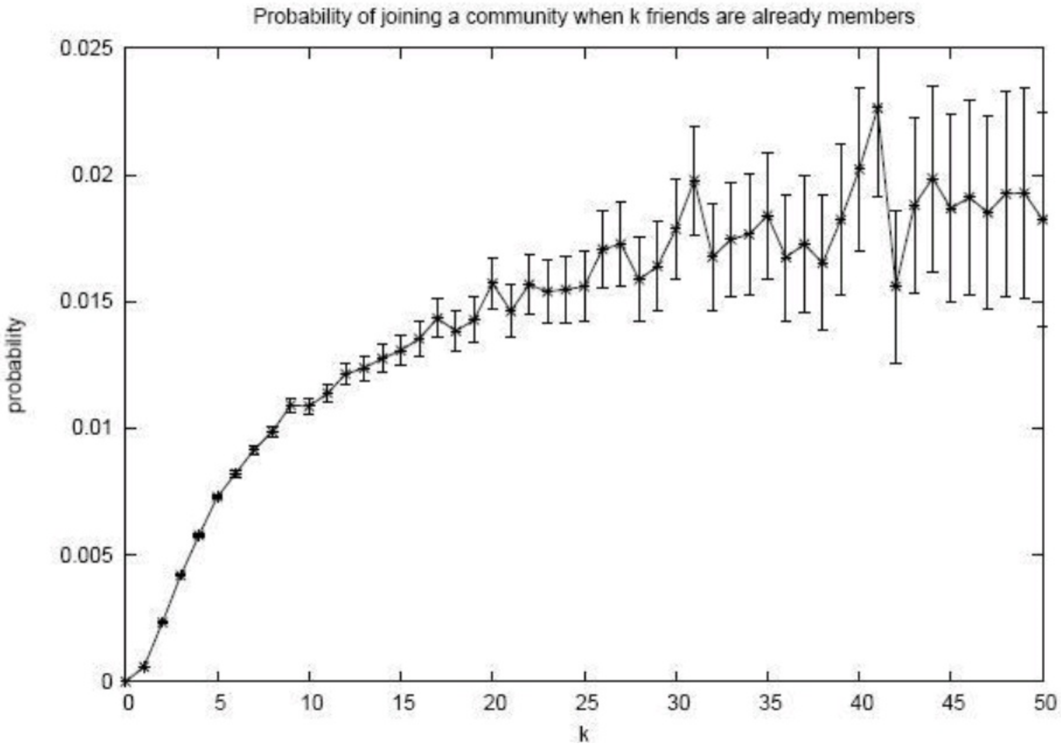
Thick arrows have probability 0.75

Thin arrows have probability 0.5

Execute independent cascade model starting from seed node

Independent cascade model

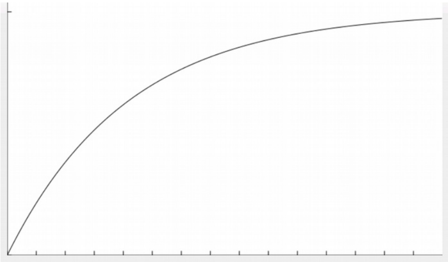
Is the independent cascade model compatible with this observation?



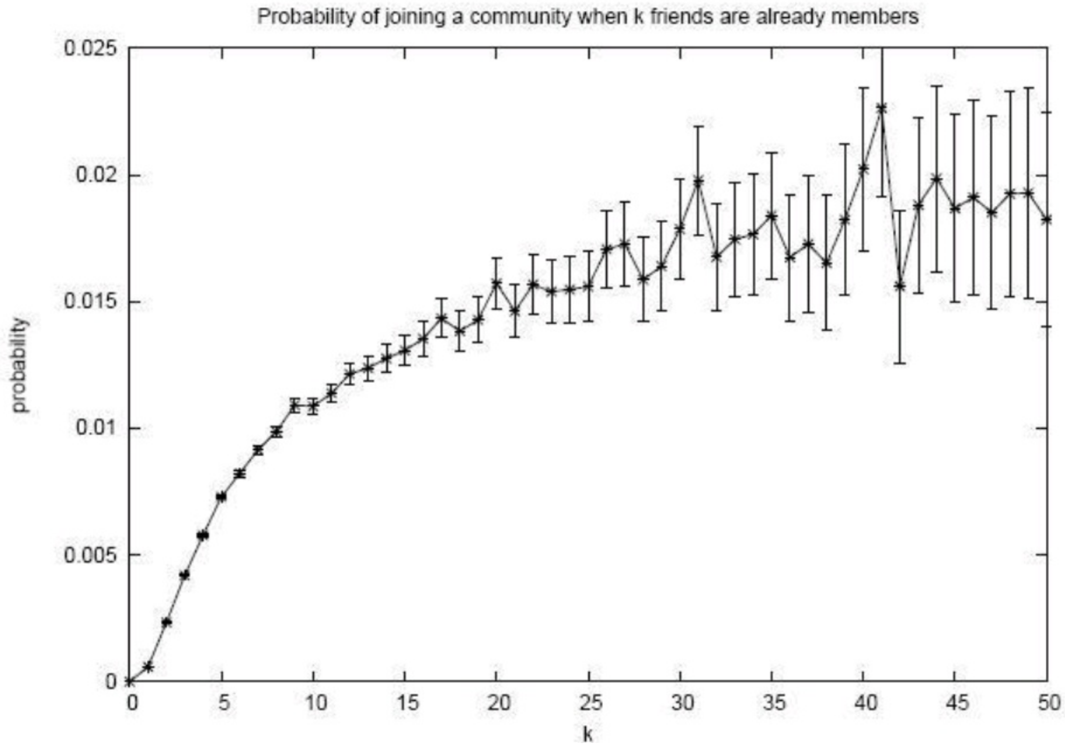
Independent cascade model

Is the independent cascade model compatible with this observation?

Hint:



$1 - q^k$ for $0<q<1$



What are these models assuming?

(List as many assumptions as you can)

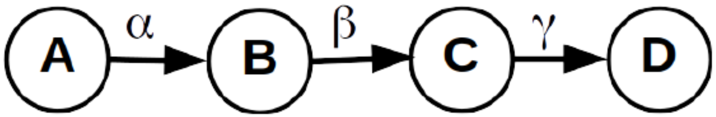
Answer in Nearpod Collaborate
https://nearpod.com/student/
Code to be given during class

Summary

Things to remember

- Influence phenomena exist, they can be modeled, they are hard to create/engineer
- Linear threshold model
- Independent cascade model
- Practice executing these models in small graphs by hand
- Practice writing code implementing them

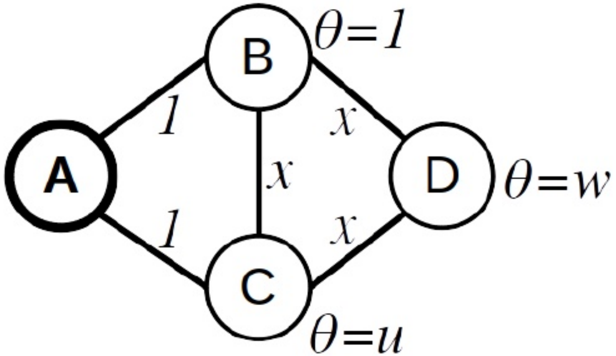
Practice on your own



- 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 $E[X] = P\, x \cdot P(X = x)$, 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.



1. ____≤ x< _____
2. Justification:

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: