Introduction to Time in Networks

Michele Berlingerio

Why studying time in networks?

- It’s everywhere
  - Evolving social networks
  - Information diffusion in networks
  - Spread of viruses
  - Performed tasks in workflows
  - Folding of proteins
  - Chemical reactions

  …

Why studying time in networks?

- It’s difficult to model
  - From simple temporal dependencies A -> B ..
  - ..to temporal annotations A(2003) -> B(2004)
  - ..to recurring events A(t) -> B(t+1) -> A(t+2) -> B(t+3) -> .. -> A(t+2k) -> B(t+ 2k+1) -> ..
  - ..to intervals of time A –[0 , 5]->B –[2 , 25] -> C
  - .. to recurring intervals of time …

…

Thus: it’s both interesting and difficult to mine!

Time in networks: possible scenarios

- Action
  - Users perform tasks individually
  - Users perform tasks jointly
  - Users exchange information
  - Web pages get updated
- Evolution
  - New users join the communities
  - Users connect to other users
  - Users quit the communities
  - Global network statistics change
- The two may coexist!
  - Online social networks
  - The Web
  - Internet
  - …

Time in networks: possible analyses

- Analyze global statistics
- Analyze local statistics
- Mine the information propagation
- Mine frequent evolution patterns
- Model an action log with a temporally annotated graph
- …

A few examples

- Information propagation

  Leskovec et al., 2005: The Dynamics of Viral marketing

  - Viral Marketing studied with statistical approaches
  - 4 millions of people in a recommendation network
  - Decreasing influence with repeated interactions
  - But increasing with the number of recommendations

A few examples

- Workflow Mining

  Hwang et. al, 2002: On the discovery of process models from their instances

  - Directed graphs to model workflows
  - Temporal dependencies between tasks
  - Overlapping tasks
  - Disjointed activities

A few examples

- Network evolution

  Sun et al., 2007: Graphscope: parameter-free mining of large time-evolving graphs

  - Discovery of communities in dynamic networks
  - MDL principle
  - Parameter-free framework
  - Clusters, no exact patterns found

Let's focus on Information Diffusion

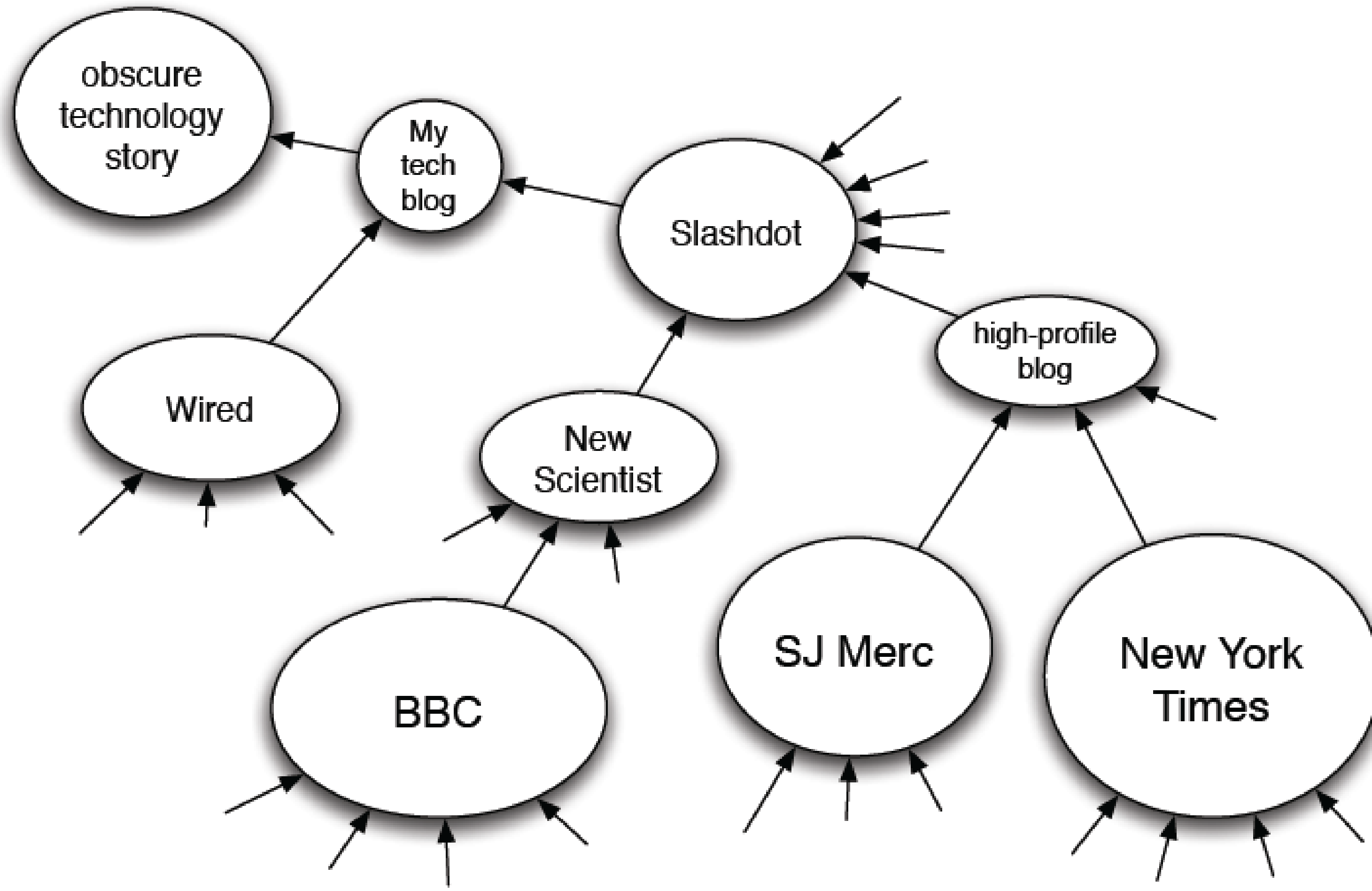
Next slides by Jure Leskovec

Information cascades and Network effects

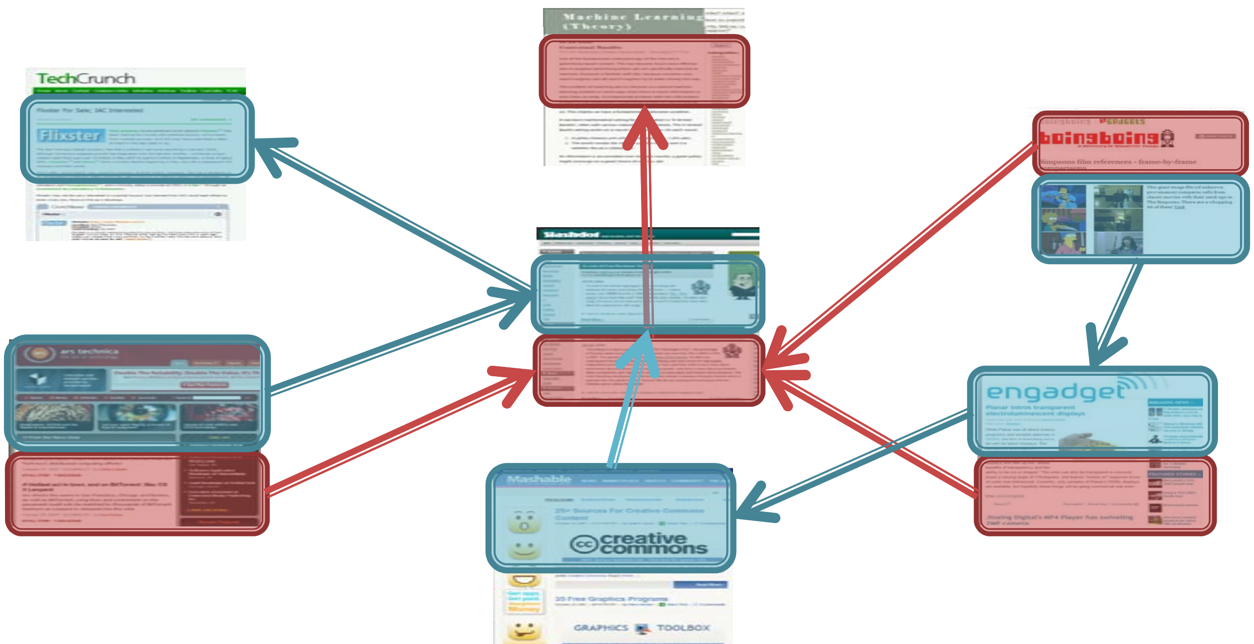
Processes and dynamics

- Spreading through networks:
  - Cascading behavior
  - Diffusion of innovations
  - Epidemics
- Examples:
  - Biological:
    - Diseases via contagion
  - Technological:
    - Cascading failures
    - Spread of information
  - Social:
    - Rumors, news, new technology
    - Viral marketing

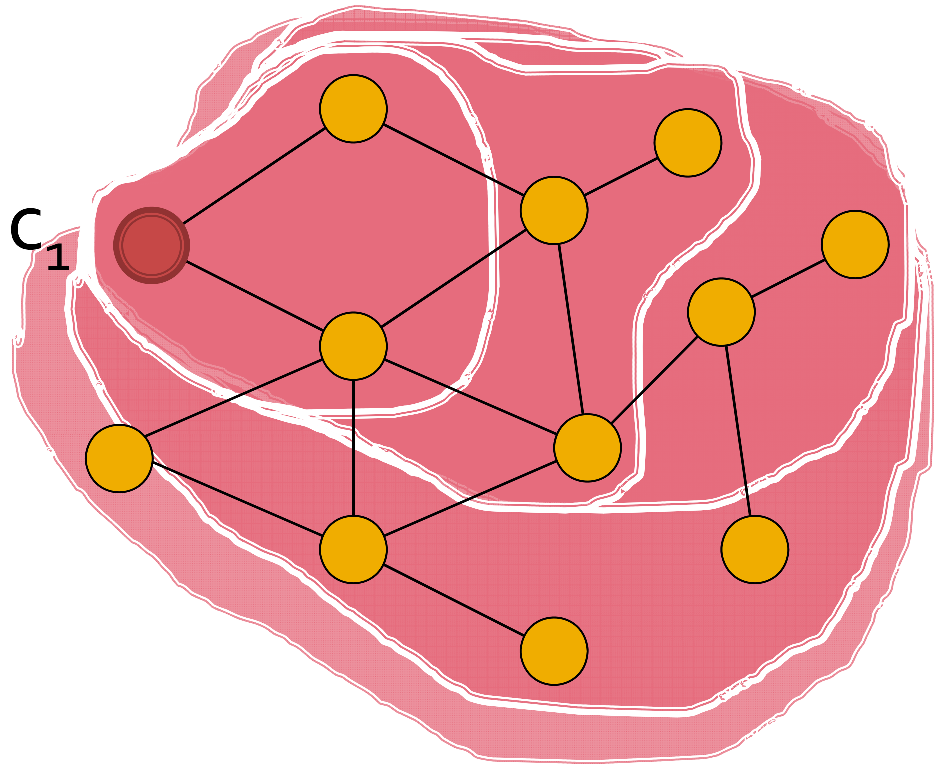
Information diffusion



Information diffusion

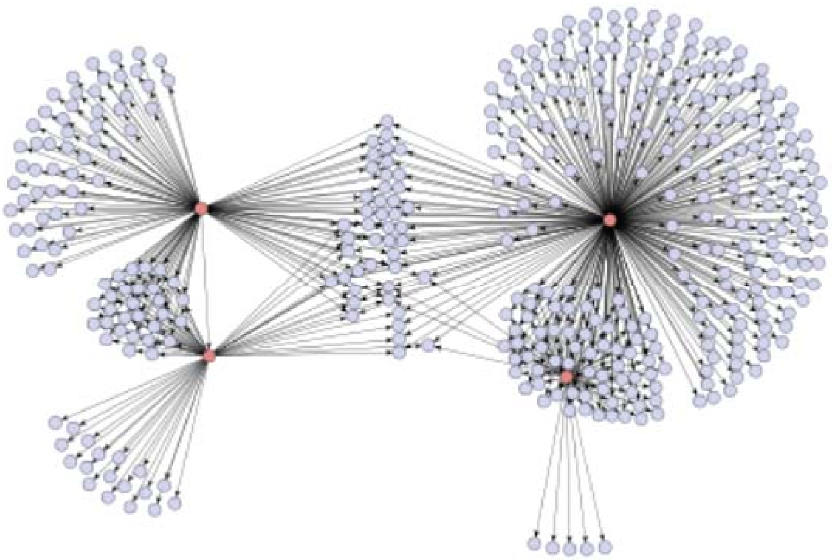


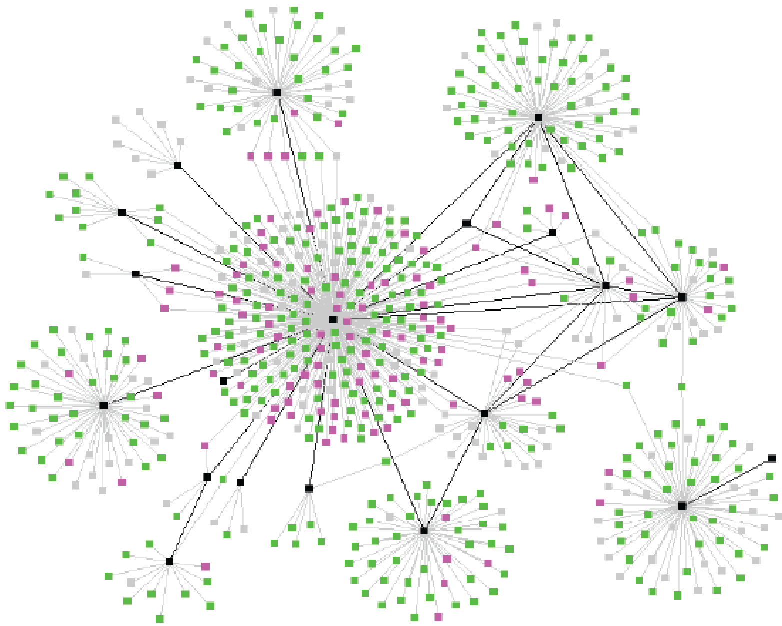
Spread of diseases



Diffusion in Social Networks

- One of the networks is a spread of a disease, the other one is product recommendations
- Which is which? ☺





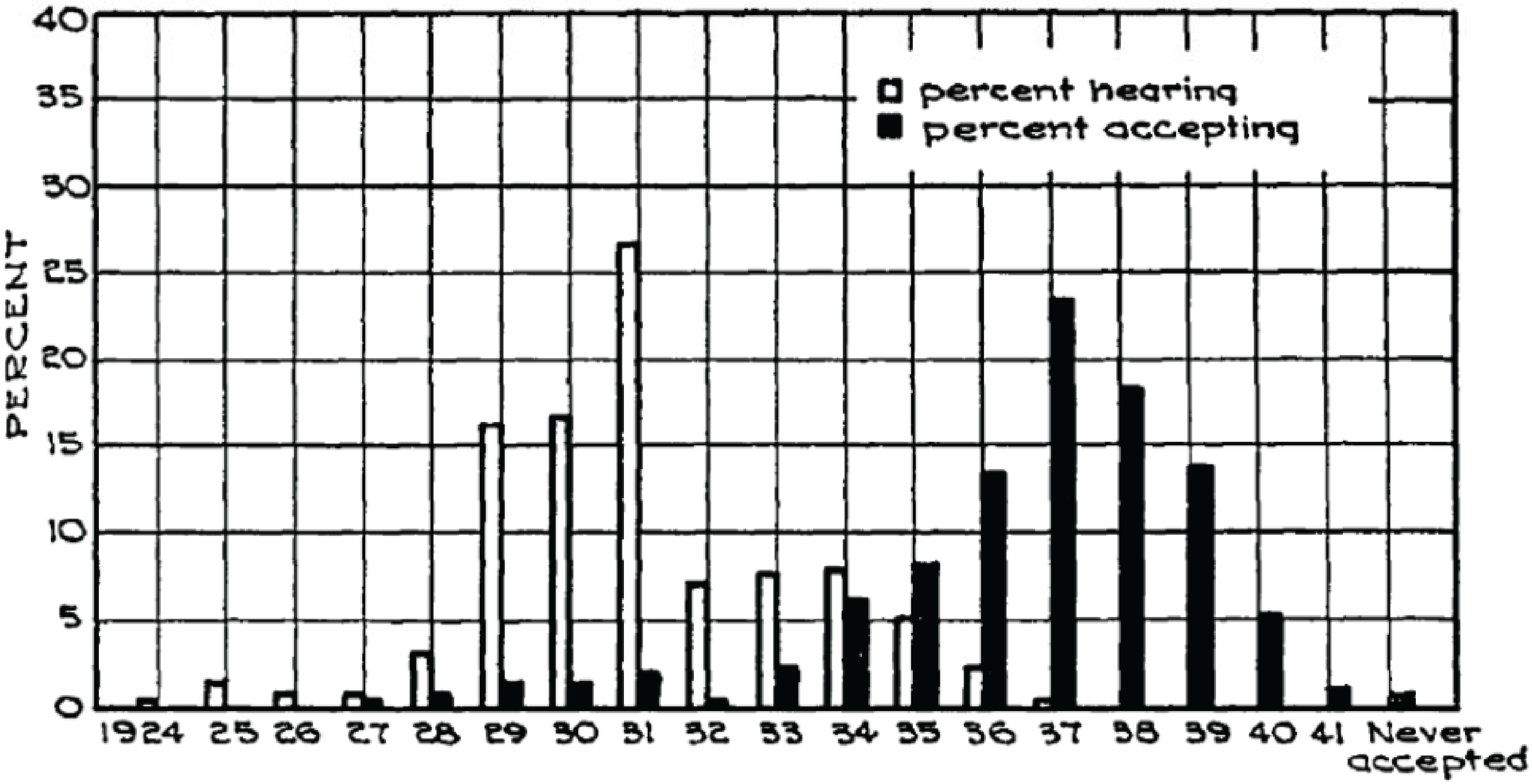
Diffusion in Networks

- A fundamental process in social networks:
  Behaviors that cascade from node to node like an epidemic
  - News, opinions, rumors, fads, urban legends, …
  - Word-of-mouth effects in marketing: rise of new websites, free web based services
  - Virus, disease propagation
  - Change in social priorities: smoking, recycling
  - Saturation news coverage: topic diffusion among bloggers
  - Internet-energized political campaigns
  - Cascading failures in financial markets
  - Localized effects: riots, people walking out of a lecture

Empirical Studies of Diffusion

- Experimental studies of diffusion:
  - Spread of new agricultural practices [Ryan-Gross 1943]
    - Adoption of a new hybrid-corn between the 259 farmers in Iowa
    - Classical study of diffusion
    - Interpersonal network plays important role in adoption → Diffusion is a social process
  - Spread of new medical practices [Coleman et al. 1966]
    - Studied the adoption of a new drug between doctors in Illinois
    - Clinical studies and scientific evaluations were not sufficient to convince the doctors
    - It was the social power of peers that led to adoption

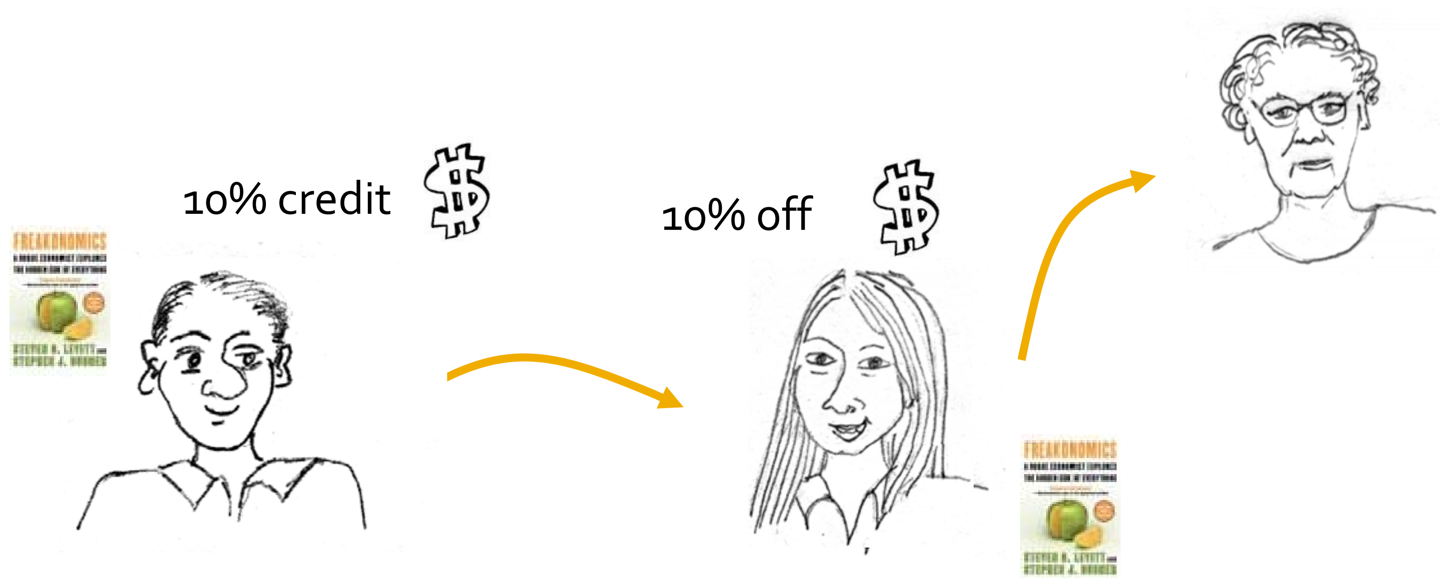
Hybrid Corn [Ryan-Gross 1966]



Diffusion is a social process

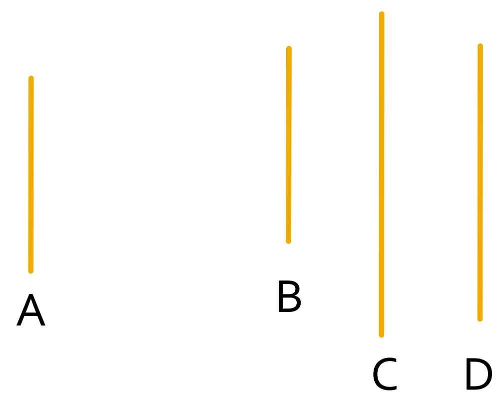
Diffusion in Viral Marketing

- Senders and followers of recommendations receive discounts on products



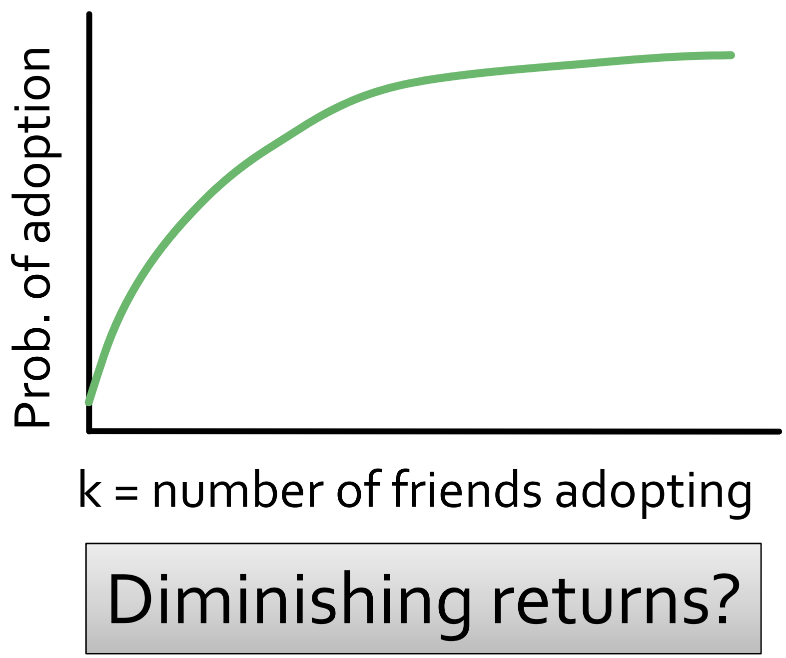
Empirical Studies of Diffusion (2)

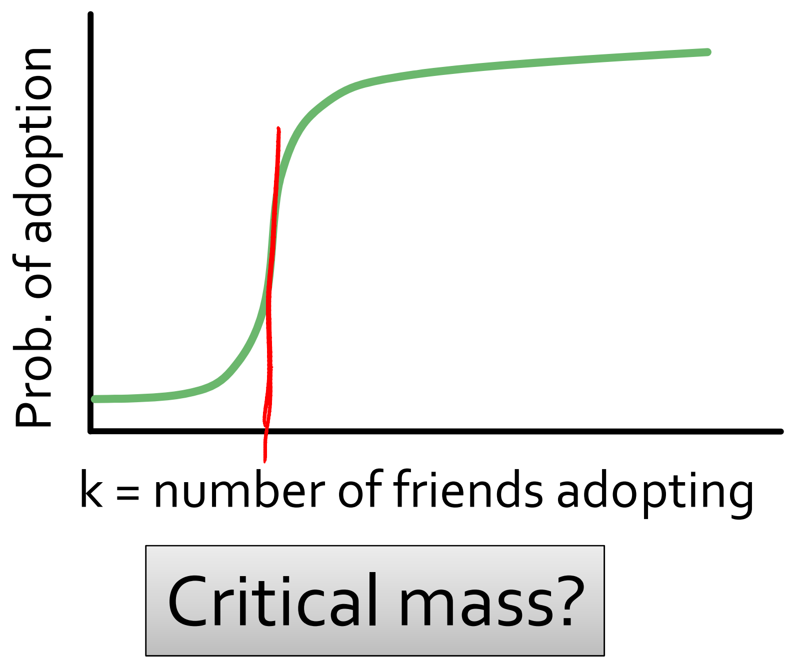
- Diffusion has many (very interesting) flavors:
  - The contagion of obesity [Christakis et al. 2007]
    - If you have an overweight friend your chances of becoming obese increases by 57%
  - Psychological effects of others' opinions, *e.g.*: Which line is closest in length to A? [Asch 1958]



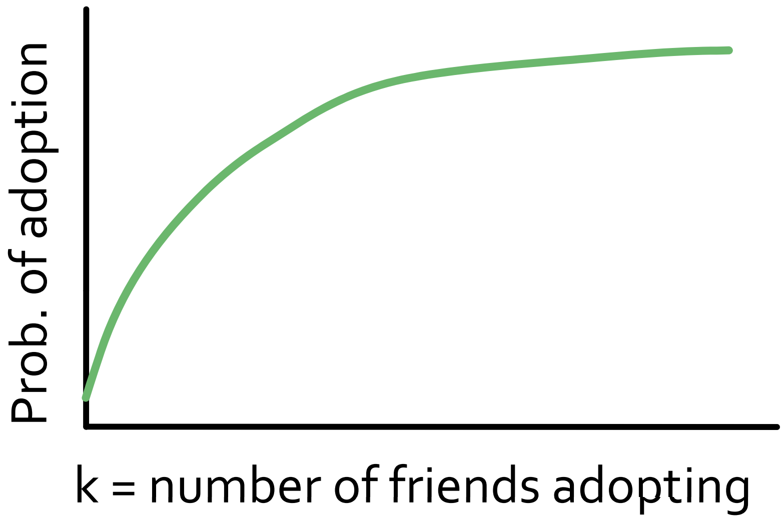
Diffusion Curves (1)

- Basis for models:
  - Probability of adopting new behavior depends on the number of friends who have adopted [Bass '69, Granovetter '78, Shelling '78]
- What's the dependence?

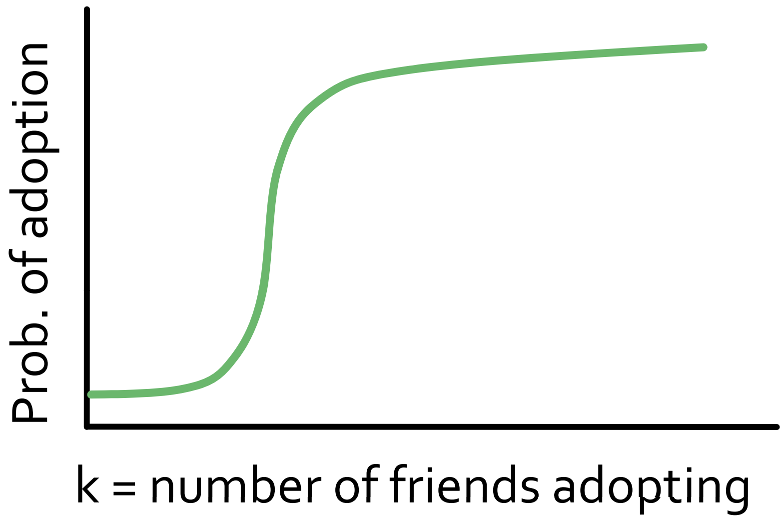




Diffusion Curves (2)



Diminishing returns?



Critical mass?

- Key issue: qualitative shape of diffusion curves
  - Diminishing returns? Critical mass?
  - Distinction has consequences for models of diffusion at population level

How to model diffusion?

- Probabilistic models:
  - Example:
    - "catch" a disease with some prob. from neighbors in the network
- Decision based models:
  - Example:
    - Adopt new behaviors if *k* of your friends do

Models

- Two flavors, two types of questions:
  - A) Probabilistic models: Virus Propagation
    - SIS: Susceptible–Infective–Susceptible (*e.g.*, flu)
    - SIR: Susceptible–Infective–Recovered (*e.g.*, chicken-pox)
    - Question: Will the virus take over the network?
    - Independent contagion model
  - B) Decision based models: Diffusion of Innovation
    - Threshold model
    - Herding behavior
    - Questions:
      - Finding influential nodes
      - Detecting cascades

[Banerjee '92]

Decision based model: Herding

- Influence of actions of others
  - Model where everyone sees everyone else's behavior
- Sequential decision making
  - Picking a restaurant:
    - Consider you are choosing a restaurant in an unfamiliar town
    - Based on Yelp reviews you intend to go to restaurant A
    - But then you arrive there is no one eating at A but the next door restaurant B is nearly full
  - What will you do?
    - Information that you can infer from other's choices may be more powerful than your own

Herding: Structure

- Herding:
  - There is a decision to be made
  - People make the decision sequentially
  - Each person has some private information that helps guide the decision
  - You can’t directly observe private info of others but can see what they do
    - Can make inferences about their private information

Herding: Simple experiment

- Consider an urn with 3 marbles. It can be either:
  - Majority-blue: 2 blue, 1 red, **or**
  - Majority-red: 1 blue, 2 red
- Each person wants to **best guess** whether the urn is majority-blue or majority-red
- Experiment: One by one each person:
  - Draws a marble
  - Privately looks are the color and puts the marble back
  - Publicly guesses whether the urn is majority-red or majority-blue
- You see all the guesses beforehand
- How should you guess?

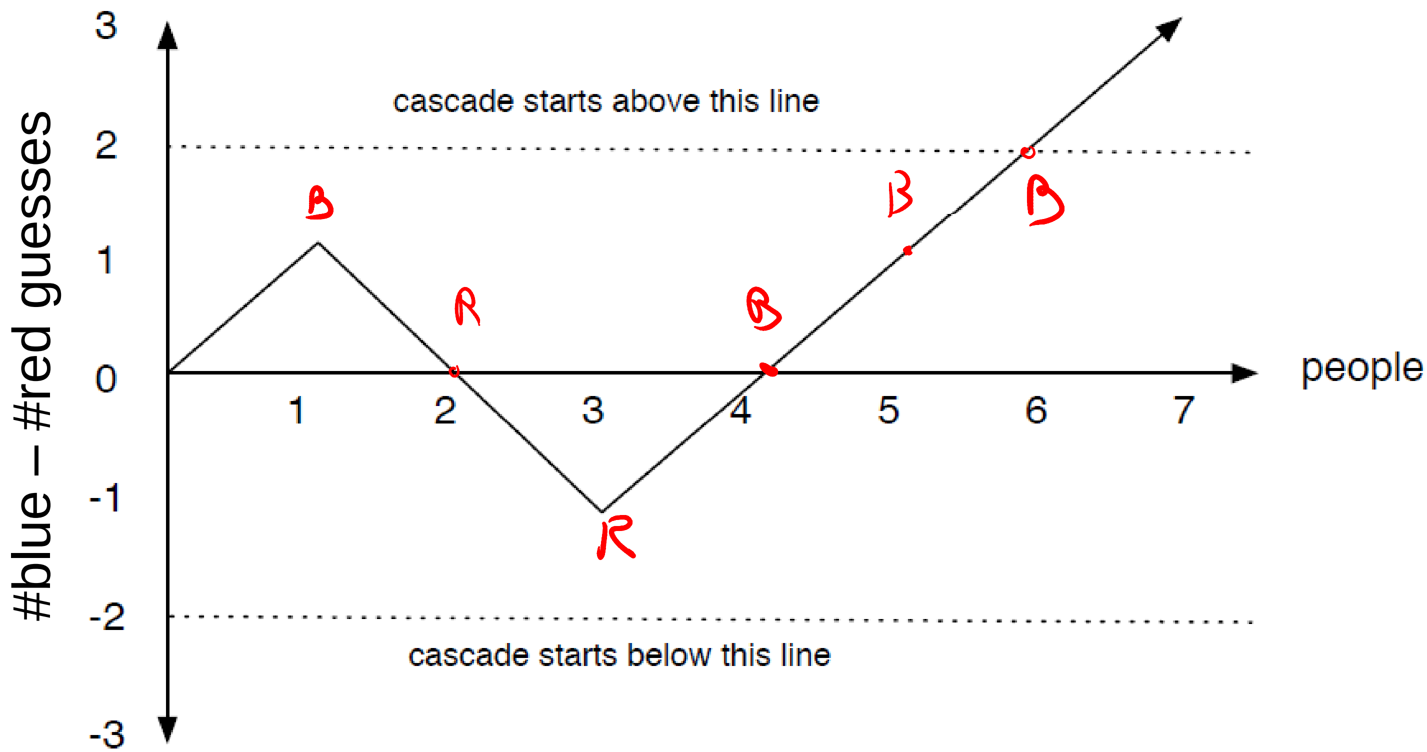
[Banerjee '92]

Herding: What happens?

- What happens:
  - 1st person: Guess the color you draw from the urn
  - 2nd person: Guess the color you draw from the urn
    - if same color as 1st, then go with it
    - If different, break the tie by doing with your own color ← TIE BREAKING
  - 3rd person:
    - If the two before made different guesses, go with your color
    - Else, just go with their guess (regardless of the color you see)
  - 4th person:
    - If the first two guesses were the same, go with it
      - 3rd person's guess conveys no information
- Can model this type of reasoning using the Bayes rule
  - see chapter 16 of Easley-Kleinberg

Herding: What happens?

- Cascade begins when the difference between the number of blue and red guesses reaches 2



Herding: Observations

- Easy to occur given right structural conditions
  - Can lead to bizarre patterns of decisions
- Non-optimal outcomes
  - With prob. $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ first two see the wrong color, from then on the whole population guesses wrong
- Can be very fragile
  - Suppose first two guess blue
  - People 100 and 101 draw red and cheat by showing their marbles
  - Person 102 now has 4 pieces of information, she guesses based on her own color
  - Cascade is broken

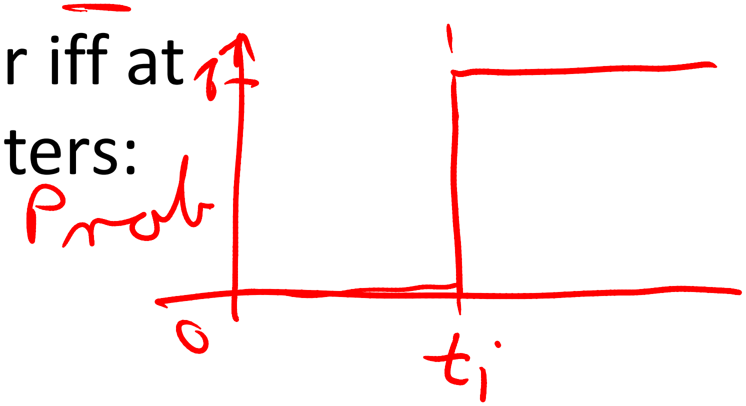
[Granovetter '78]

Decision based models

- Collective action [Granovetter, '78]
  - Model where everyone sees everyone else's behavior
  - Examples:
    - Clapping or getting up and leaving in a theater
    - Keeping your money or not in a stock market
    - Neighborhoods in cities changing ethnic composition
    - Riots, protests, strikes

Collective action: The model

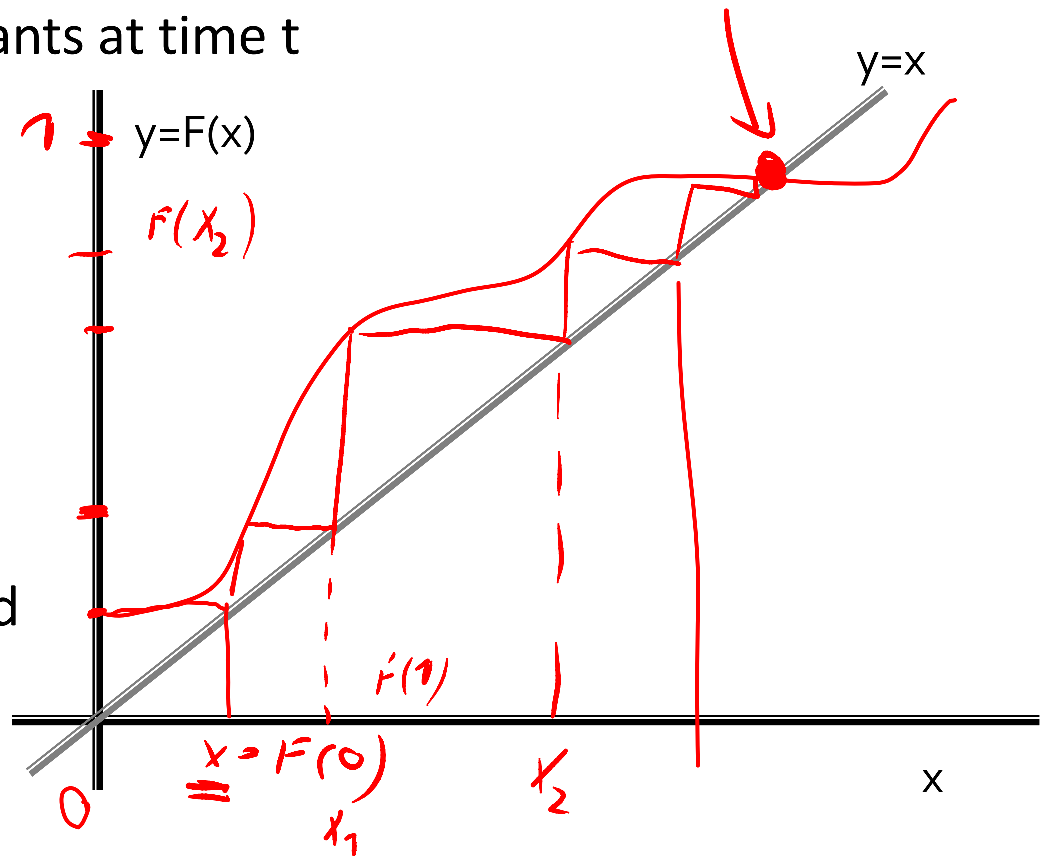
- n people – everyone observes all actions
- Each person i has a threshold t_i
  - Node i will adopt the behavior iff at least t_i other people are adopters:
    - Small t_i: early adopter
    - Large t_i: late adopter



- The population is described by $\{t_1,\ldots,t_n\}$
  - F(x) … fraction of people with threshold $t_i \leq x$

Collective action: Dynamics

- Think of the step-by-step change in number of people adopting the behavior:
  - F(x) … fraction of people with threshold ≤ x
  - s(t) … number of participants at time t
- Easy to simulate:
  - s(0) = 0
  - s(1) = F(0)
  - s(2) = F(s(1)) = F(F(0))
  - s(t+1) = F(s(t)) = $F^{t+1}(0)$
- F(x)=x – stable point
  - There could be other fixed points but starting from 0 we never reach them



Weaknesses of the model

- It does not take into account:
  - No notion of social network – more influential users
  - It matters who the early adopters are, not just how many
  - Models people's awareness of size of participation not just actual number of people participating
  - Modeling thresholds
    - Richer distributions
    - Deriving thresholds from mode basic assumptions
      - game theoretic models

Weaknesses of the model

- It does not take into account:
  - Modeling perceptions of who is adopting the behavior/ who you believe is adopting
  - Non monotone behavior – dropping out if too many people adopt
  - Similarity – thresholds not based only on numbers
  - People get "locked in" to certain choice over a period of time
- Network matters! (next slide)

How should we organize a revolt?

- You live in an oppressive society
- You know of a demonstration against the government planned for tomorrow
- If a lot of people show up, the government will fall
- If only a few people show up, the demonstrators will be arrested and it would have been better had everyone stayed at home

Pluralistic ignorance

- You should do something if you believe you are in the majority!

- Dictator tip: Pluralistic ignorance – erroneous estimates about the prevalence of certain opinions in the population

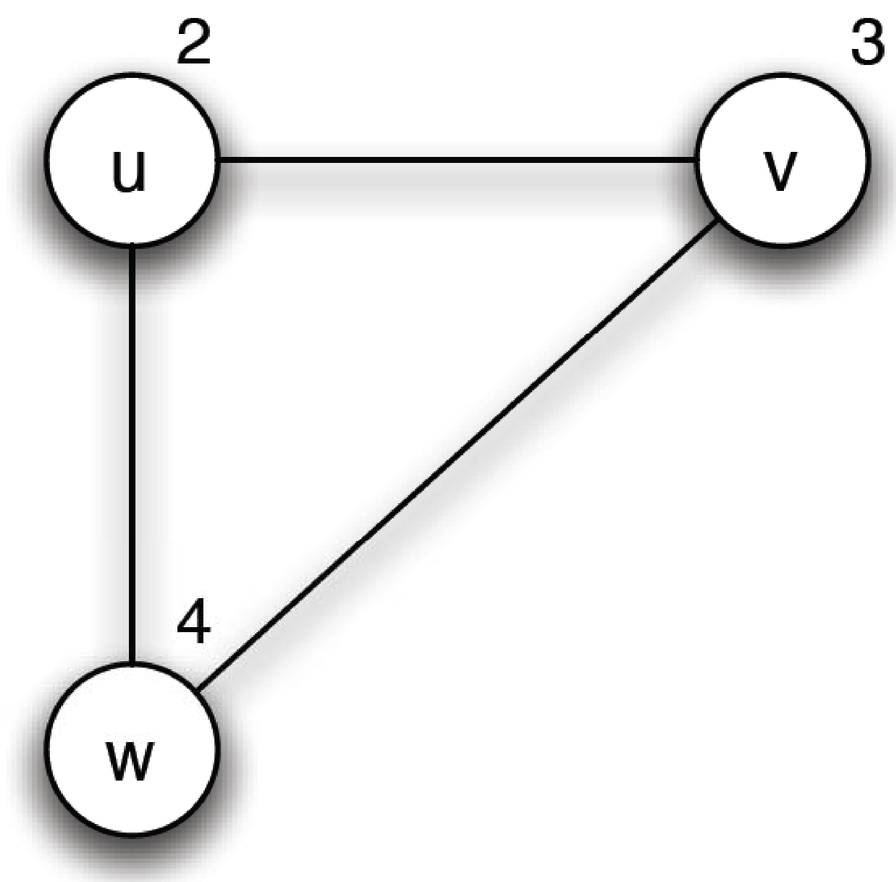
  - Survey conducted in the U.S. in 1970 showed that while a clear minority of white Americans at that point favored racial segregation, significantly more than 50% believed that it was favored by a majority of white Americans in their region of the country

Organizing the revolt: The model

- Personal threshold *k*: "I will show up to the protest if I am sure at least *k* people in total (including myself) will show up"
- Each node in the network knows the thresholds of all their friends

Subtle issues

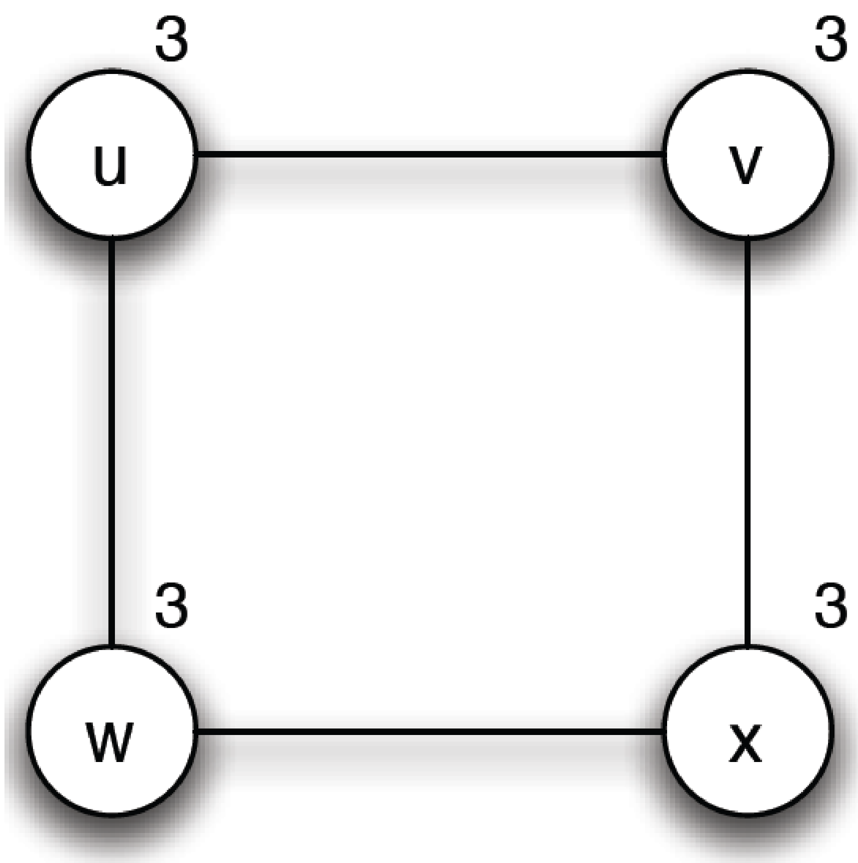
- Will uprising occur?



No!

Subtle issues

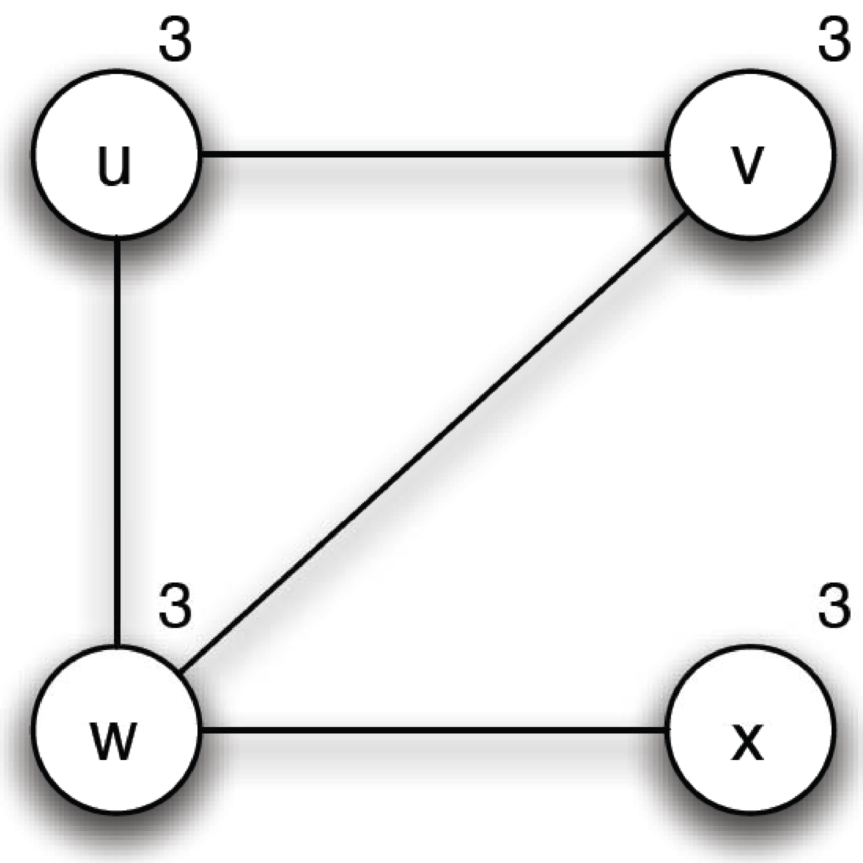
- Will uprising occur?



No!

Subtle issues

- Will uprising occur?



Yes!