CLIQUE

Diffusion Mechanisms in Social Networks

Conor Hayes
DERI, NUI Galway

Information Diffusion in Social Networks

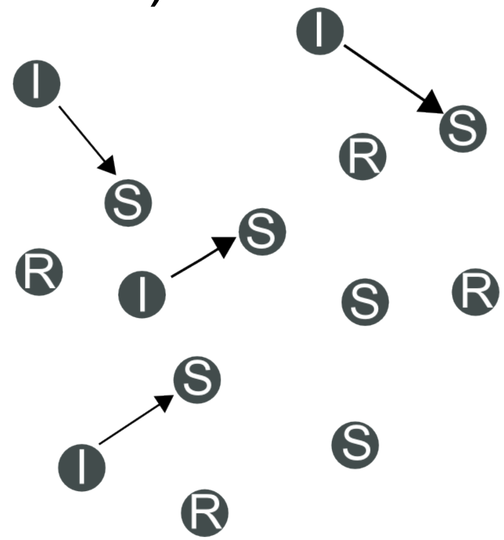
- Earliest graph theory: Connectivity enables information spreading
- Information moves from node to node
    - News, opinion, rumours, fads,
    - Word-of-mouth effects in marketing : Hotmail, GMail
    - Virus, disease propagation
    - Topic diffusion among bloggers
    - Internet- political campaigns
    - Cascading failures in financial markets
    - Localised effects: riots, people walking out of a lecture
- Which processes are involved?
- Which selection criteria?

Outline

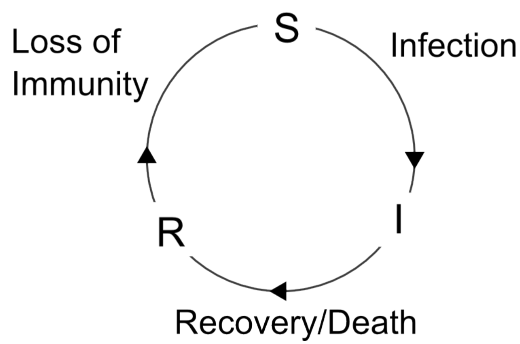
- Part 1: Epidemic model
  - SIRS/SIR
  - SIR on a small world
- Part 2: Diffusion on social networks
  - Threshold model
  - Independent cascade model
  - Watts model
- Part 3: challenges for DERI

Classic SIR model

- Individuals are either **S**usceptible, **I**nfectious or **R**ecovered (Kermack and McKendrick 1927)
  - Population is fixed – no birth, no death, immigration, emigration
  - Duration of infectivity is the same as duration of clinical disease
  - The disease is first introduced by a single person



- Random contact
- Infection depends on
  - Infection rate β
    - β = contact rate * infectiousness
  - Recovery rate γ
  - Ratio of infective to susceptible

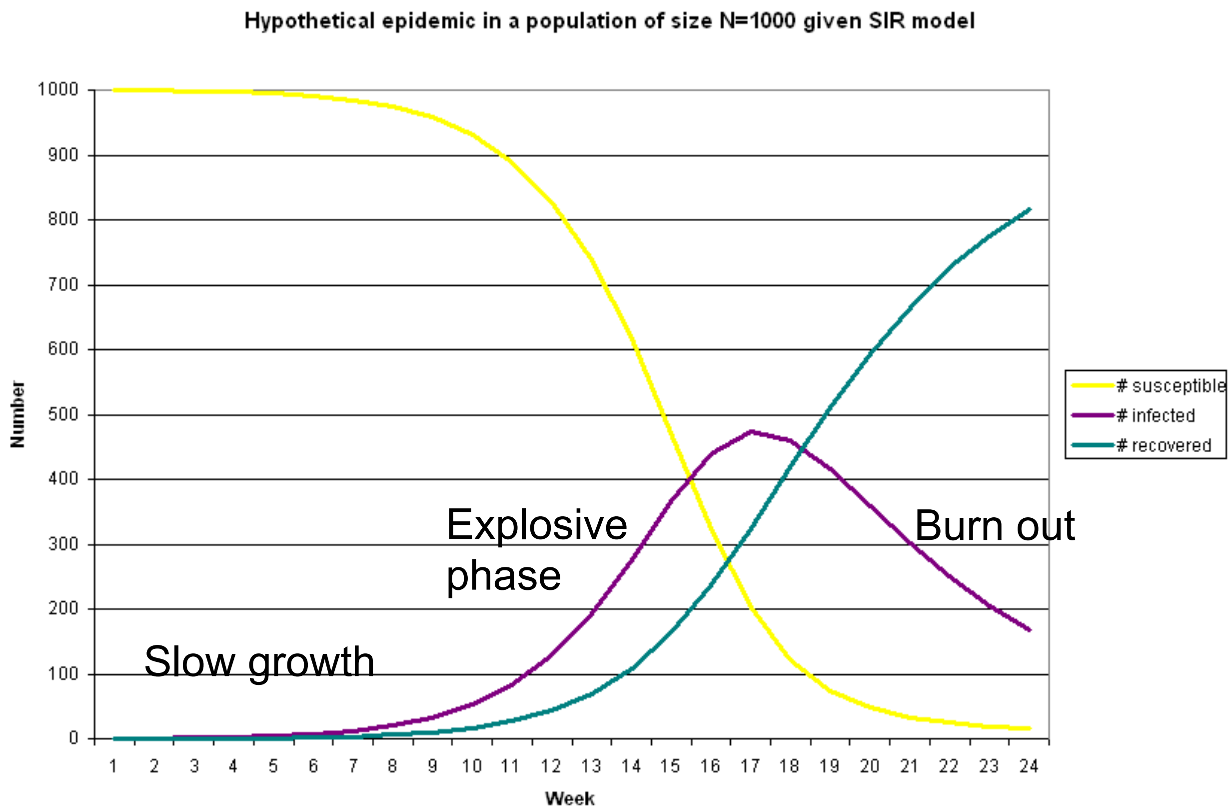


SIR Equations

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$



Epidemic Threshold

$$T = \frac{\beta S_0}{\gamma}$$

= the number of secondary infections caused by a single primary infection

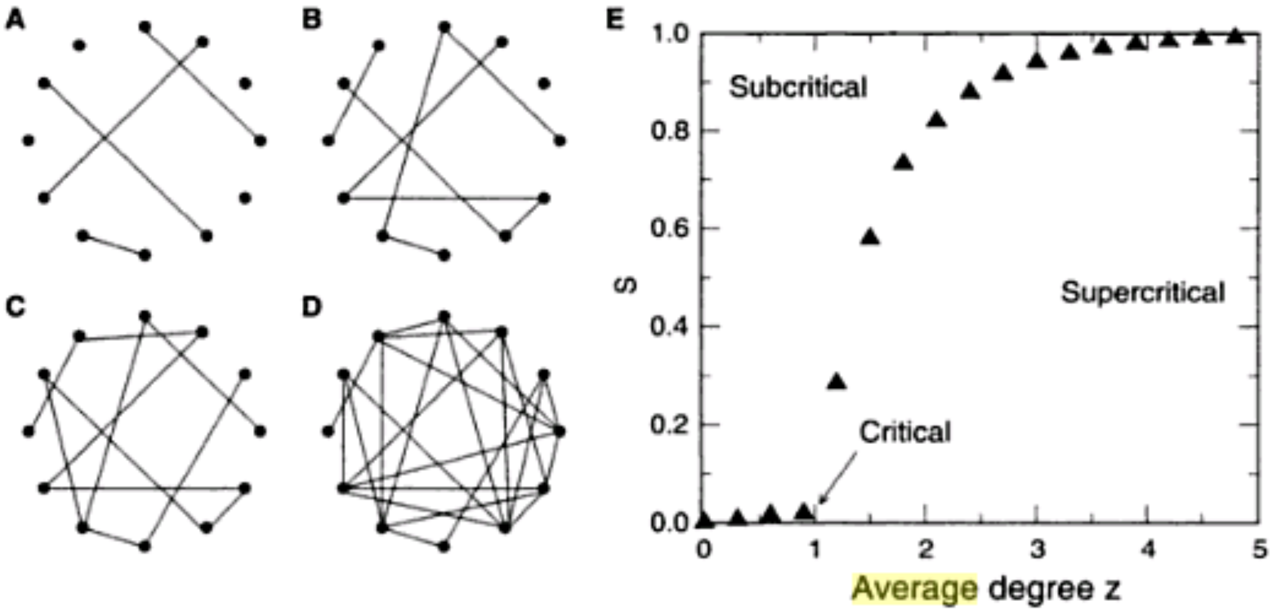
- S_0 is the initial size of the susceptible population
- **T < 1** disease dies out; *dI/dt* < 0.

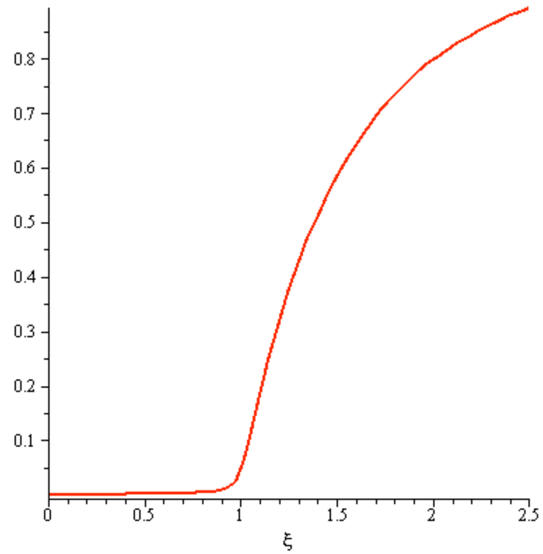
  : I infects less than one ***S*** before recovering or dying.
- **T=1: phase transition**
- **T > 1** disease will spread until full population gets infected; *dI/dt* > 0.

  :I infects more than one ***S***.

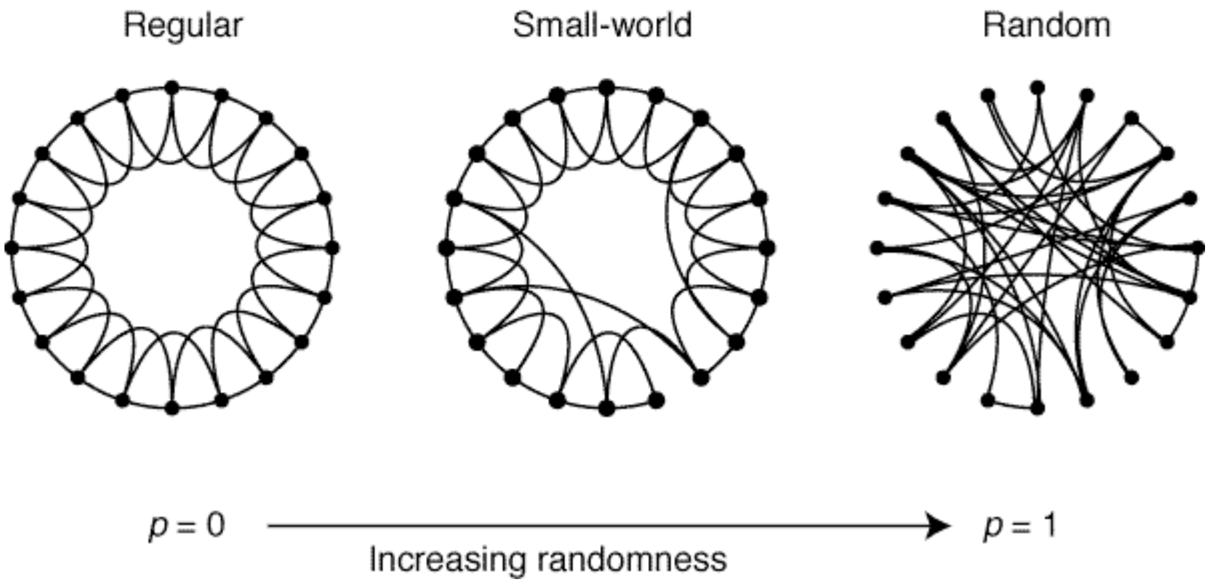
Relation to Random Graph

- Phase transition on random graph
  - when average degree =1
- Analogous to **phase transition at T =1**
  - the number of secondary infections caused by a single primary infection





Watts and Strogatz model



Regular, *p*=0

- High cluster coefficient
- **High path length**

Small world Network 0<*p*<1

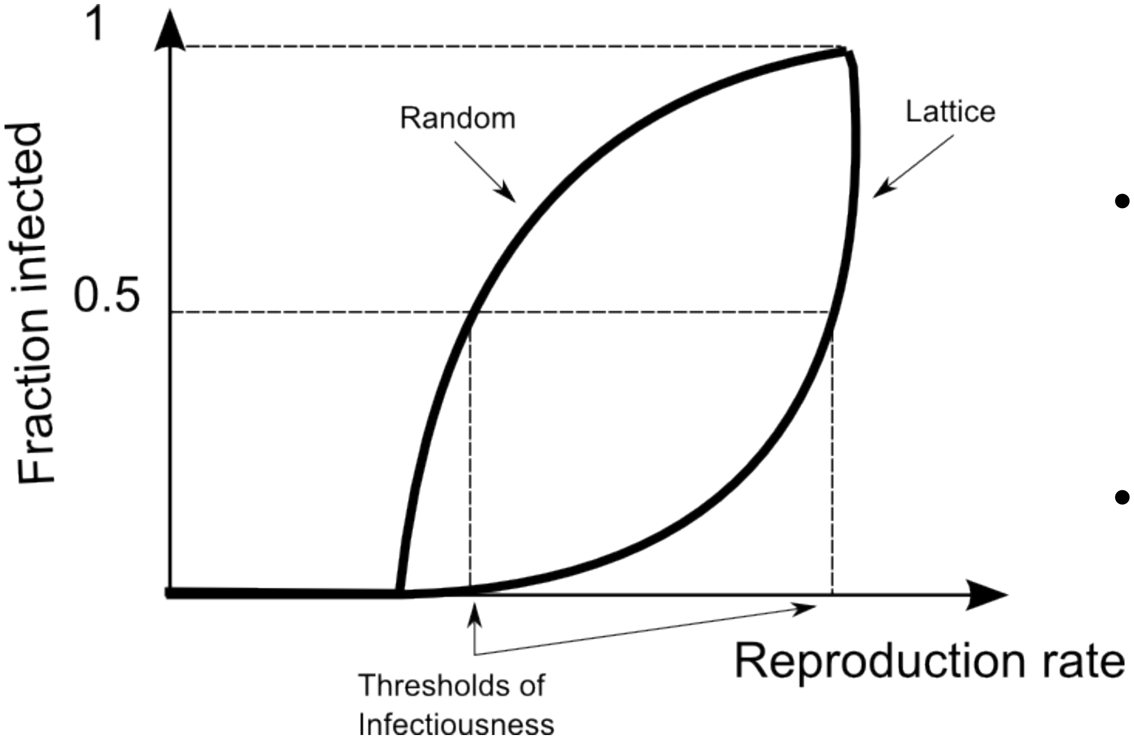
- **High cluster coefficient**
- **Low path length**

Random, *p*=1

- low cluster coefficient
- **low path length**

D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". Nature 393: 440–442.

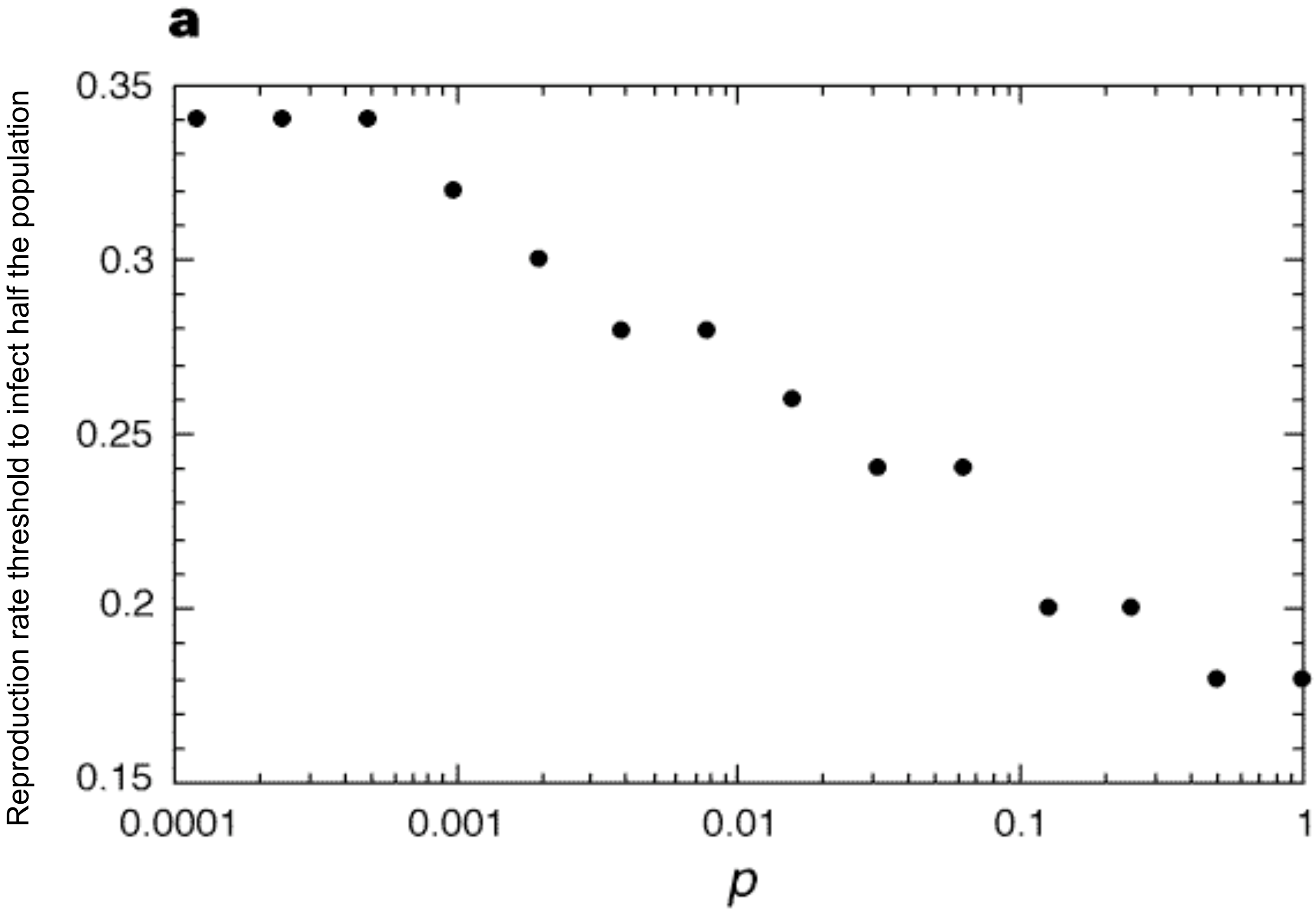
Random vs. Ordered Graph



- **Random graph:** low reproduction rate to infect 50% of population
- **Regular lattice graph**: high reproduction rate to infect 50% of population
- **High cluster coefficient** : diseases spread more slowly when contact is mainly local

Infectious diseases spread much easier on random graphs

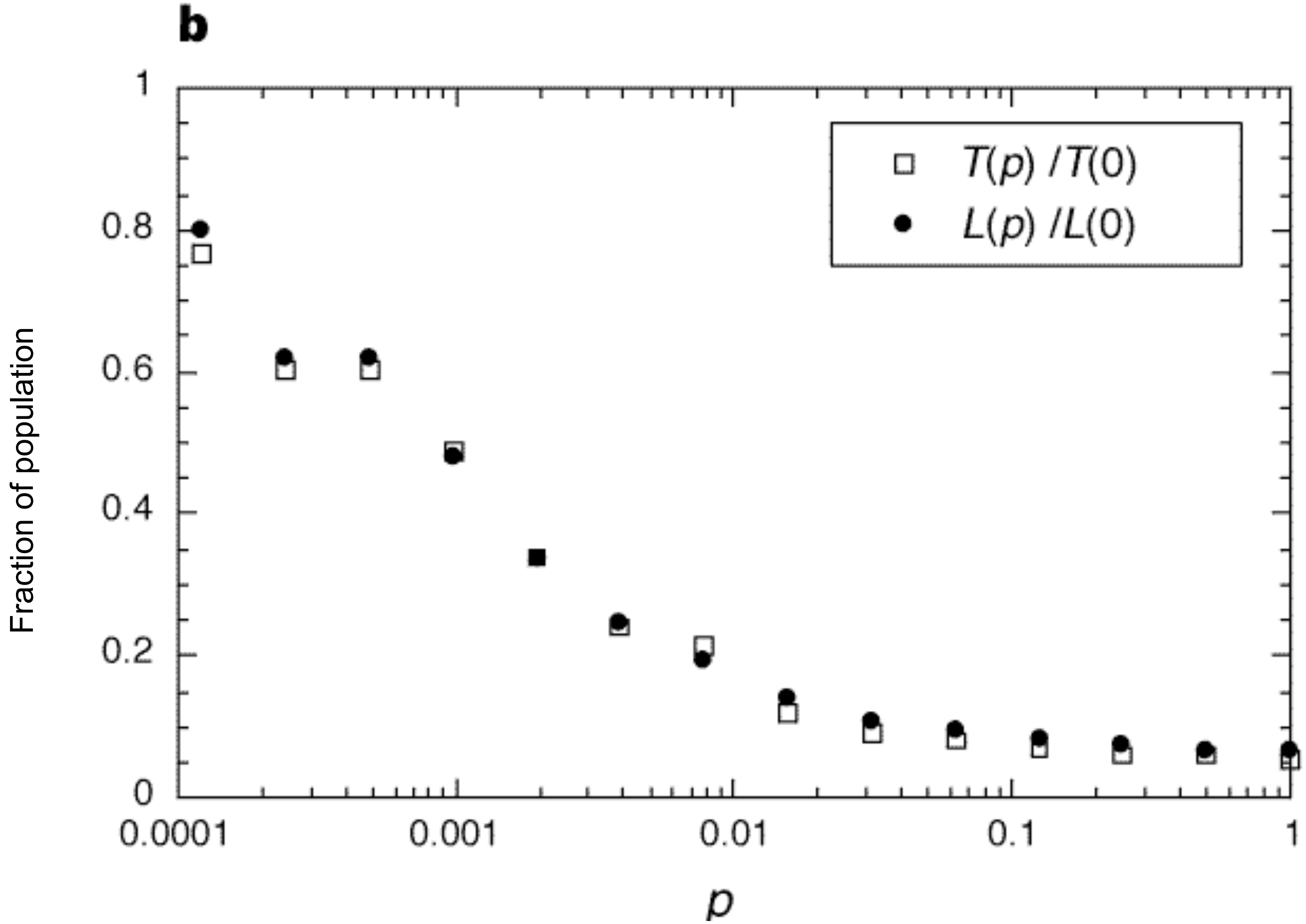
Simulation results



D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". Nature 393: 440–442.

Reproduction rate threshold *decreases* as the probability of links increases

Simulation results 2



- The time to global infection *follows* the characteristic path length
- **Even with a few re-wirings, time to global infection is almost as short as a random graph**

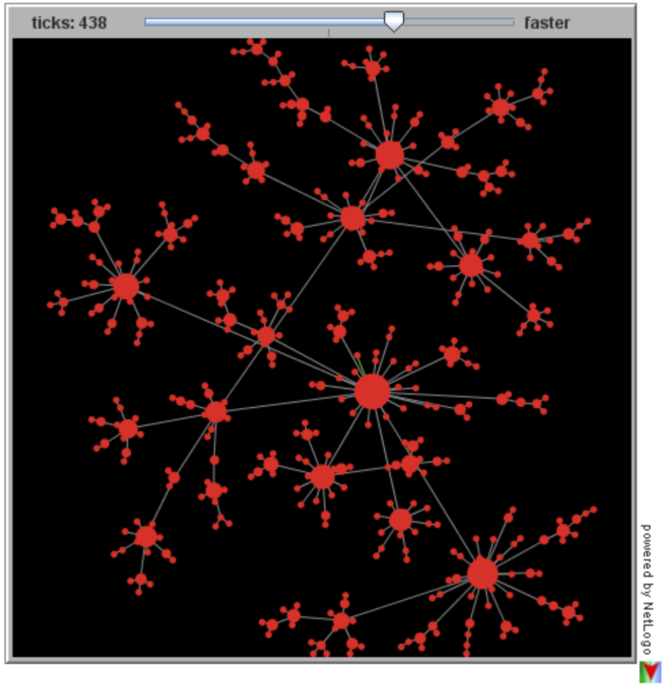
D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". Nature 393: 440–442.

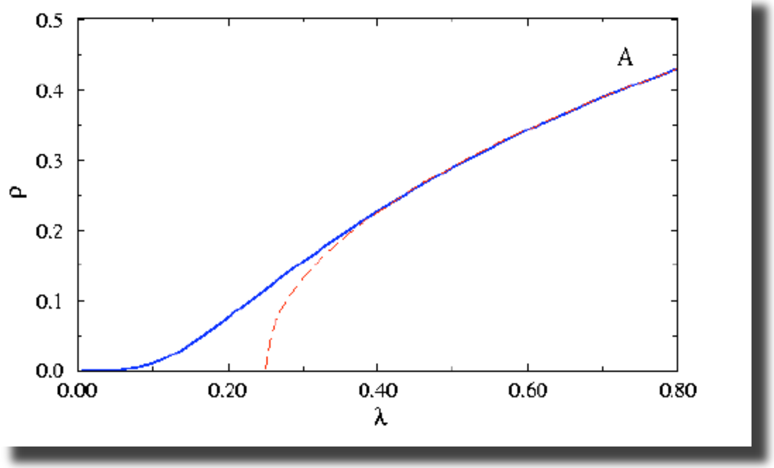
Scale Free Networks

- Theoretically, no clear threshold in PA-generated Scale Free networks
- Infection can spread even when transmission probabilities are tiny
- limiting the max degree can induce a threshold

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- Vaccination of a few super-connected nodes can be sufficient to prevent an epidemic





R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001)

Lloyd, A. L. & May, R. M. 2001 How viruses spread among computers and people. Science 292, 1316–

Networks and epidemic models. M. J Keeling and K. T.D Eames (2005) J R Soc Interface 2, 295-307

SIR epidemics

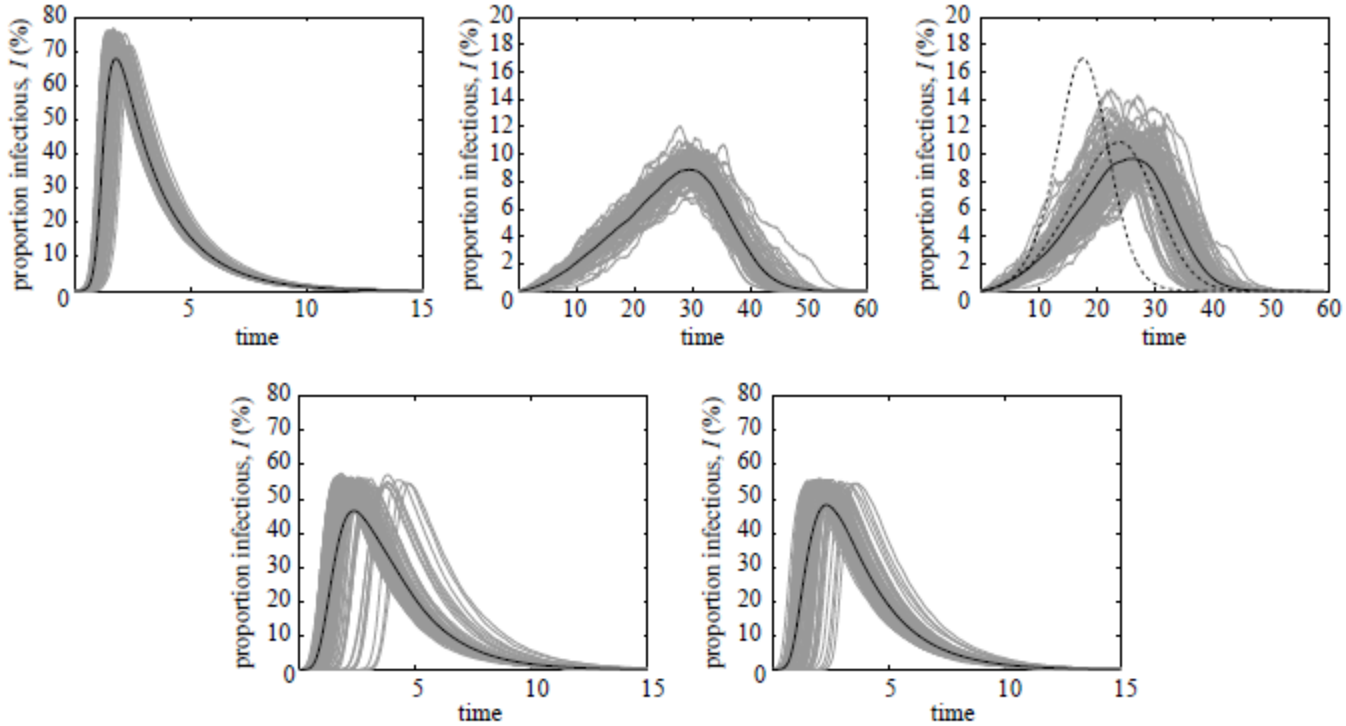


Figure 3. Typical SIR epidemics on the five network types shown in figure 2. These are from left to right: random, lattice, small world (top row), spatial and scale-free (bottom row). Each graph shows 100 epidemic curves (grey) together with the average for all major epidemics (black) for a single example of each network type; therefore, all variability within each graph is a result of the stochastic nature of transmission and not variation in the network. All five networks contain 10 000 individuals, although all individuals are not necessarily interconnected as part of a giant component. For the spatial and scale-free networks, approximately 88 and 74% are part of the giant component and can therefore potentially become infected. For these networks, the proportion of infectious individuals has been rescaled as a fraction of the giant component. In all networks, the average number of contacts per individual is approximately 4, although for the scale-free network, there is considerable heterogeneity with one individual having 85 contacts. For consistency, the small-world network is formed from a two-dimensional lattice (not a one-dimensional circle as shown in figure 2) with 10 additional random 'long-range' contacts. The dashed lines show the effect on the mean epidemic of increasing the number of long-range contacts to 20 and 100. ($\tau=1$, $g=0.5$, $b=d=0$).

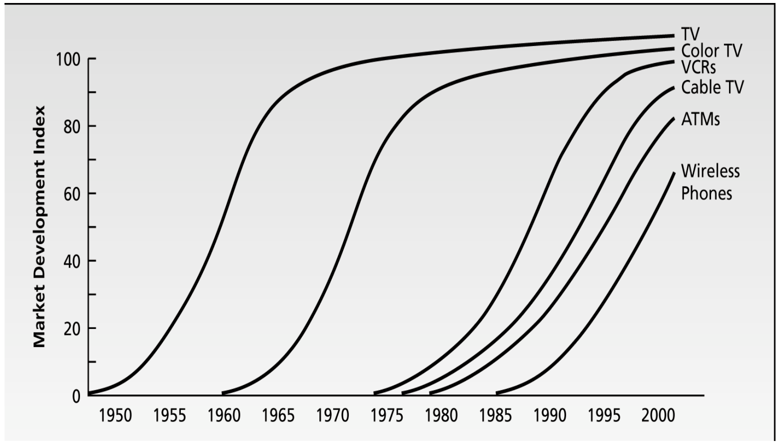
Networks and epidemic models. M. J Keeling and K. T.D Eames (2005) J R Soc Interface 2, 295-307

Part 2: Diffusion in Social Networks

- Clearly related to classic epidemic models
- Spread of infection ~ spread of innovation, opinion, rumour,
- Viral marketing, ‘word-of-mouth’
  - exploiting epidemics rather than vaccinating against them
- Identification of efficient ‘vaccination’ points
  - Against churn?
- Trend detection and prediction

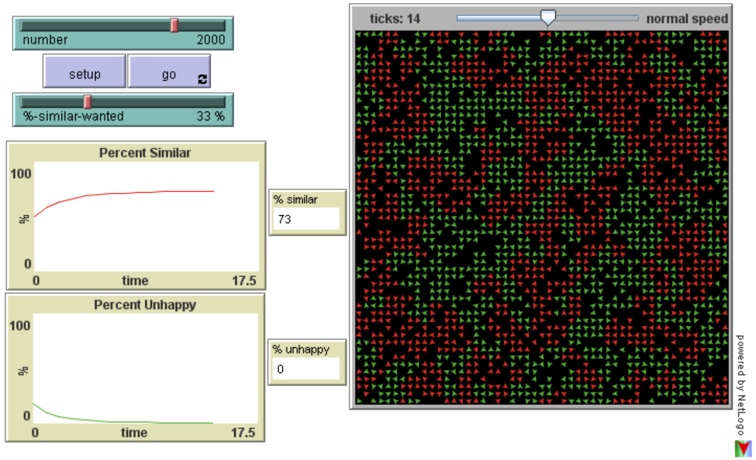
Diffusion of an idea of product

- Underlying hypothesis: Probability of adopting new behaviour depends on the **number of friends who have adopted**
- Bass, Frank (1969). "A new product growth model for consumer durables“: Management Science
  - **Innovators** adopt an innovation independently of other individuals in a social system
  - **Imitators** are influenced in the timing of adoption by the decision of other members of the social system
  - $h(t) = p + qF(t)$



Diffusion of an idea of product

- Underlying hypothesis: Probability of adopting new behaviour depends on the **number of friends who have adopted**
  - Schelling, Thomas C. 1971. "Dynamic Models of Segregation."
  - Why are most neighbourhoods racially and culturally uniform?
  - Actors move depending on ε, the fraction of same coloured neighbours in their neighbour hood



Diffusion of an idea of product

- Underlying hypothesis: Probability of adopting new behaviour depends on the **number of friends who have adopted**

- Granovetter, Mark. (1978) Threshold models of collective behaviour:
- Cost-benefit analysis of ‘joining in’ a social ‘event’: e.g. a riot, adopting birth control, walking out
- Every actor has a threshold
- Threshold = proportion of the group he would have to see before joining in
  – ‘radicals’ ~ 0
  – conservatives 100%

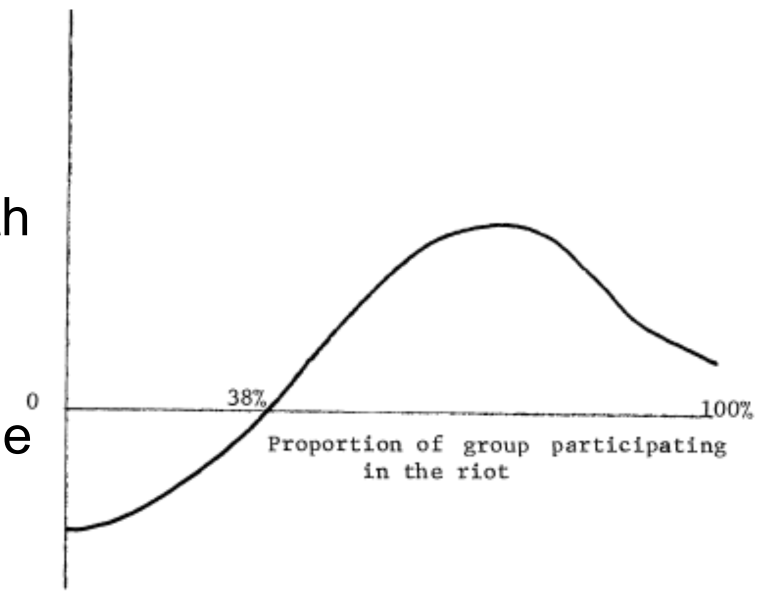
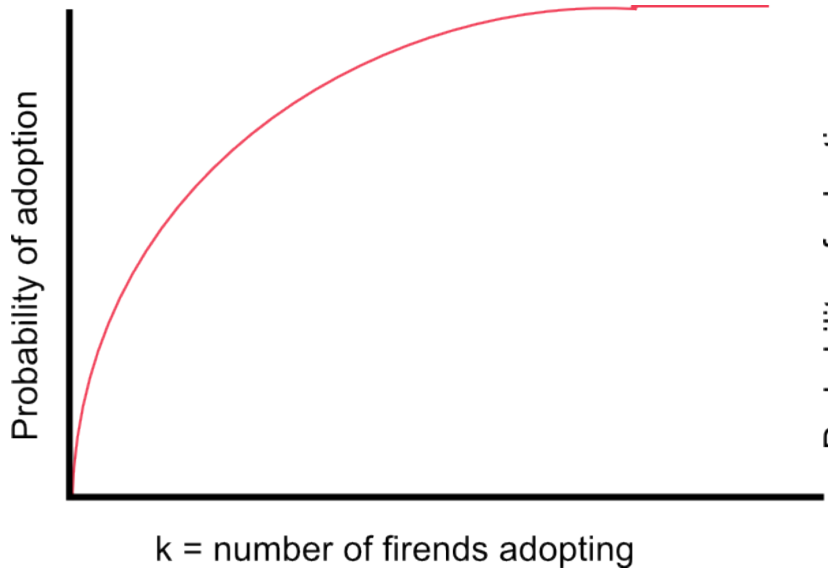
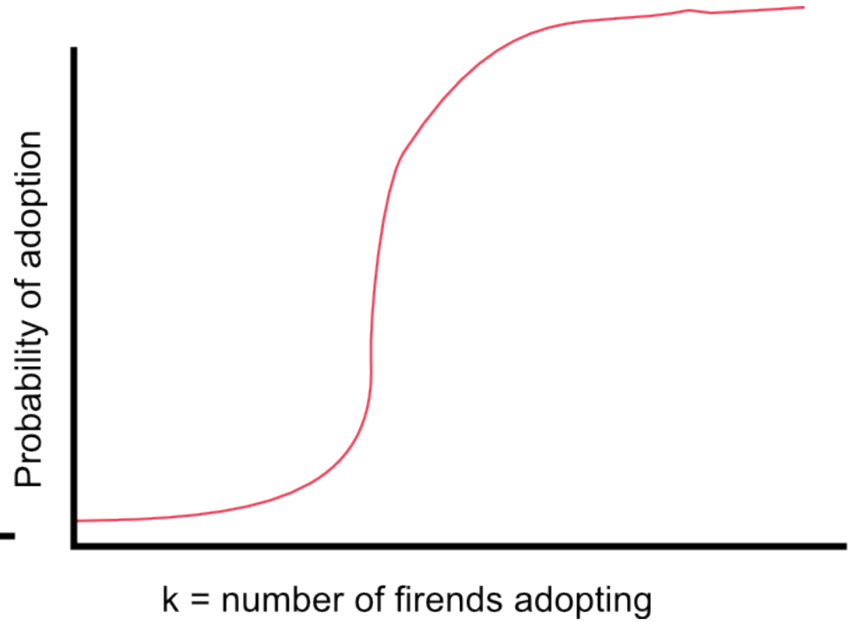


FIG. 3.—Net benefit to an individual, with threshold 38%, of joining a riot, plotted against the proportion of the group participating. (Total benefits minus total costs.)

Diffusion



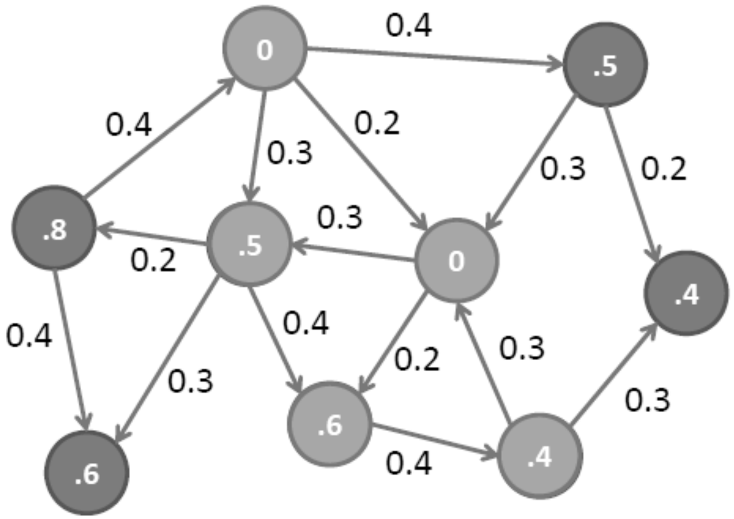


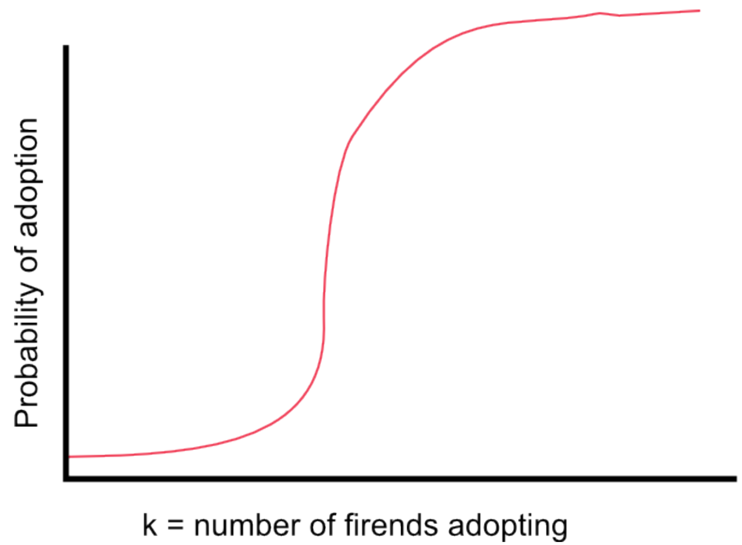
Does the ‘influence curve’ exhibit a **diminishing return** or **critical mass** behaviour

http://www.cs.cmu.edu/~jure/talks/diffusion-nato07.pdf

Threshold Model [Granovetter '78]

- Initially some nodes are active
- Each edge *(u,v)* has weight w_{uv}
- Each node has a threshold t
- Node u activates if $t < \Sigma_{adopted(u)} w_{uv}$
- Effect of a node v on node u is *dependent* on the activation of other nodes.

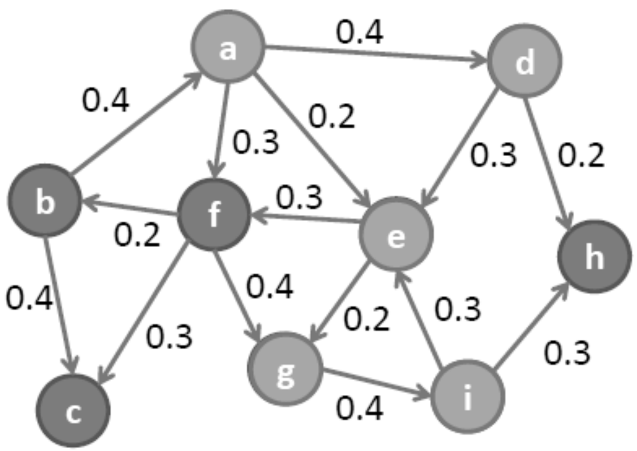


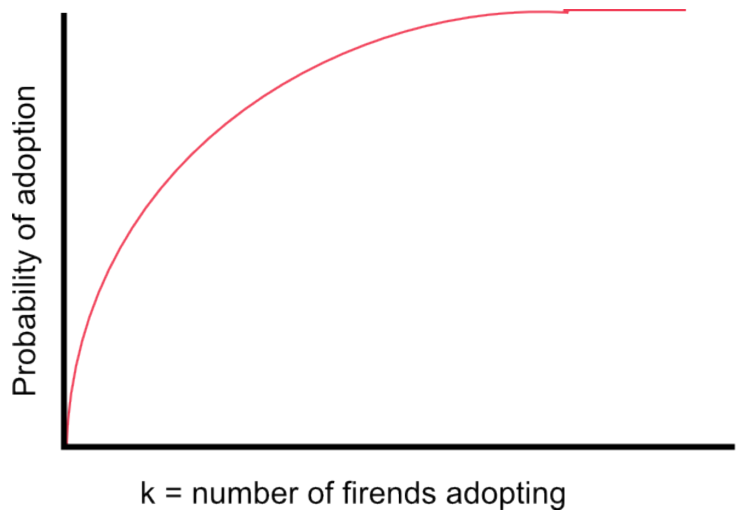


http://www.cs.cmu.edu/~jure/talks/diffusion-nato07.pdf

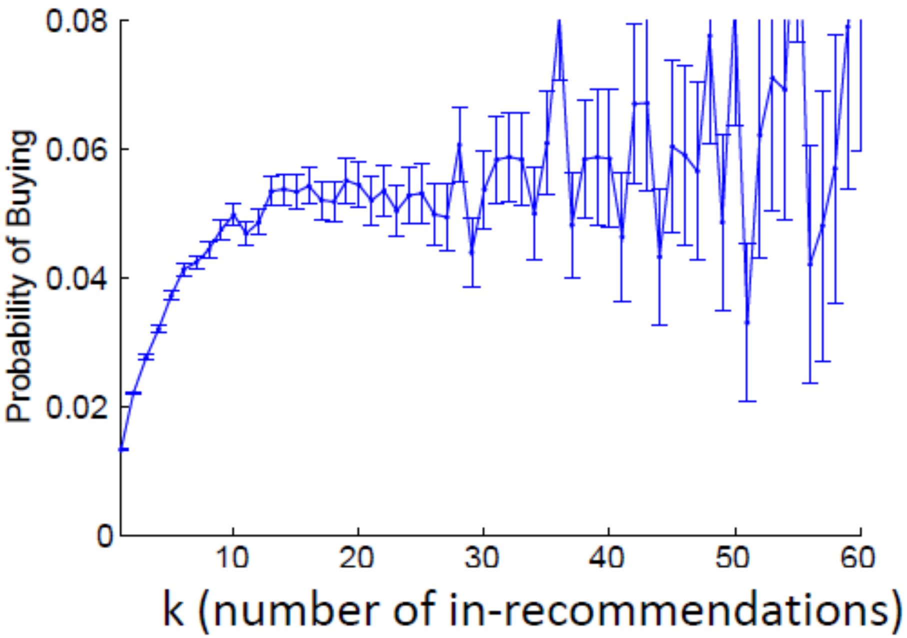
Independent Cascade Model [Goldenberg et al. 2001]

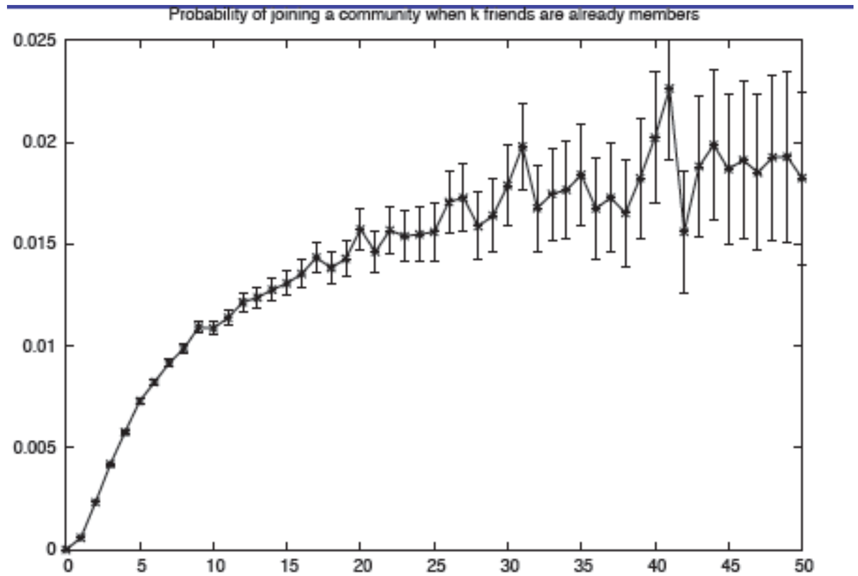
- Initially some nodes are active
- Each edge *(u,v)* has probability (weight) p_{uv}
- Node activates neighbouring node with prob. p_{uv}
- Activations spread through the network
- Effect of a node *v* on node *u* is *independent* of the activation of other nodes





Empirical results

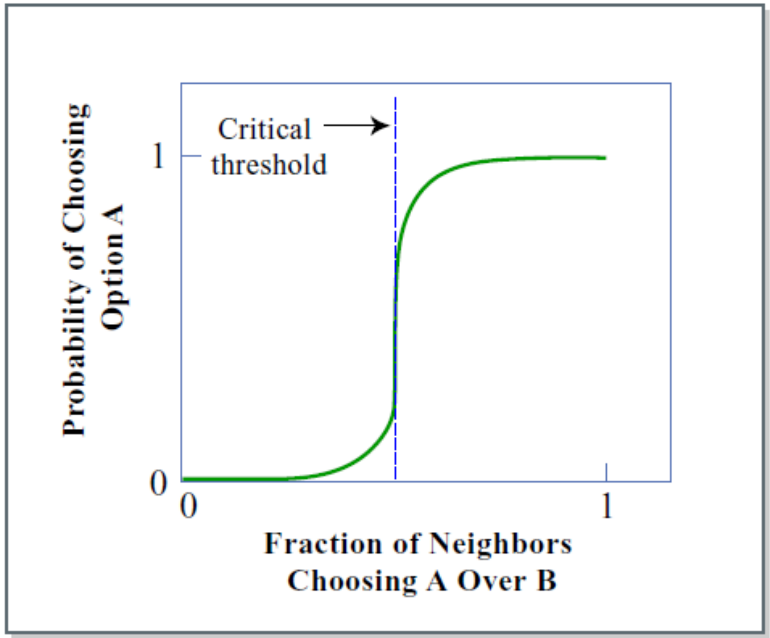


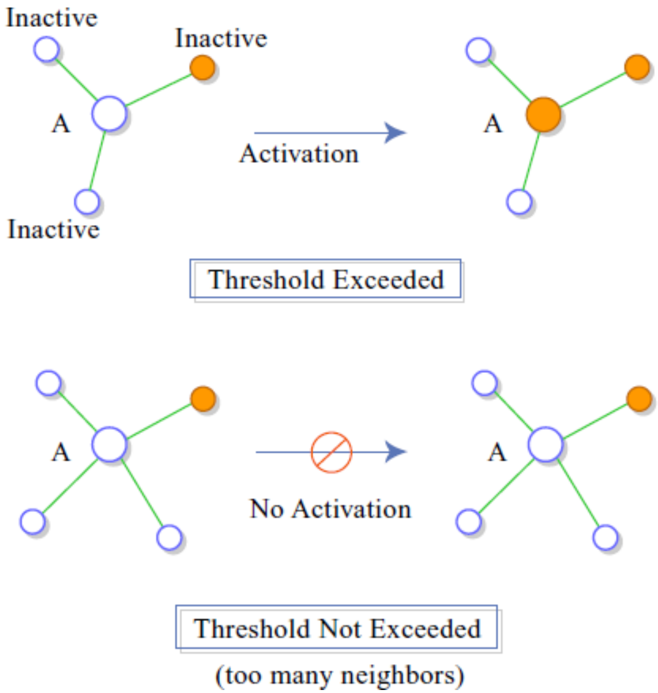


- Mainly diminishing returns
  - Lekovec 2008: probability of purchasing based on incoming recommendations
  - Backstrom et al. 06: probability of joining a conference, community when *k* friends are already members

Watts Model (02,04) 1

- Locally Dependent Cascade Model
- "Binary decisions with externalities"
- Threshold: ***fraction of linked nodes*** making the same decision

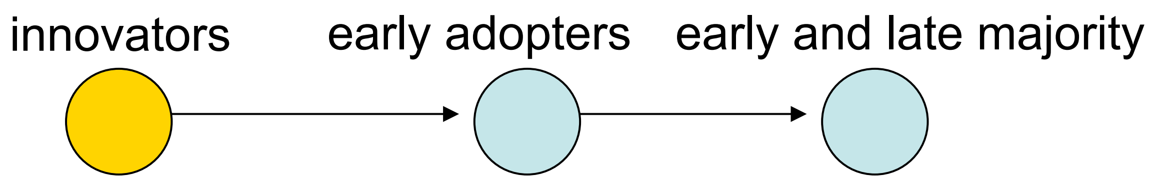




Figures by MIT OCW. After Watts.

Watts Model (02, 04) : key points

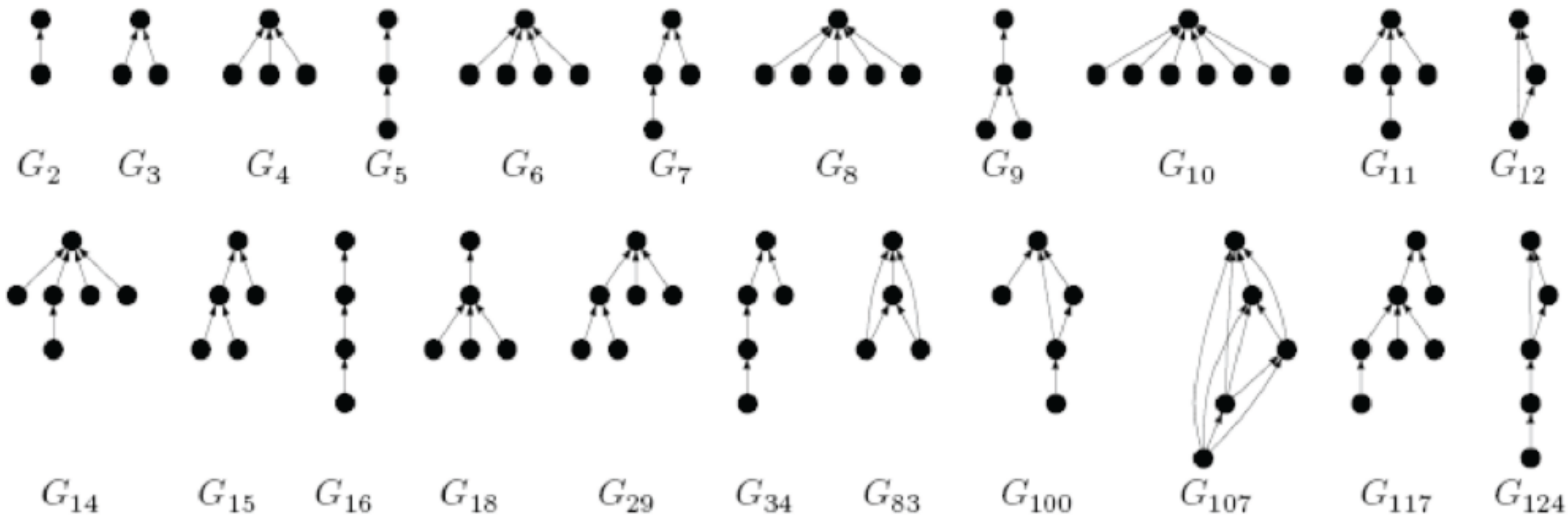
- Local dependencies, fractional thresholds and heterogeneity :



- Initial seed: *innovators*
- Vulnerable nodes = *early adopters*
  - threshold $\leq 1/k$
- Initial Cascade success: depends on *structure* and *connectivity* of early adopters : *largest vulnerable cluster*

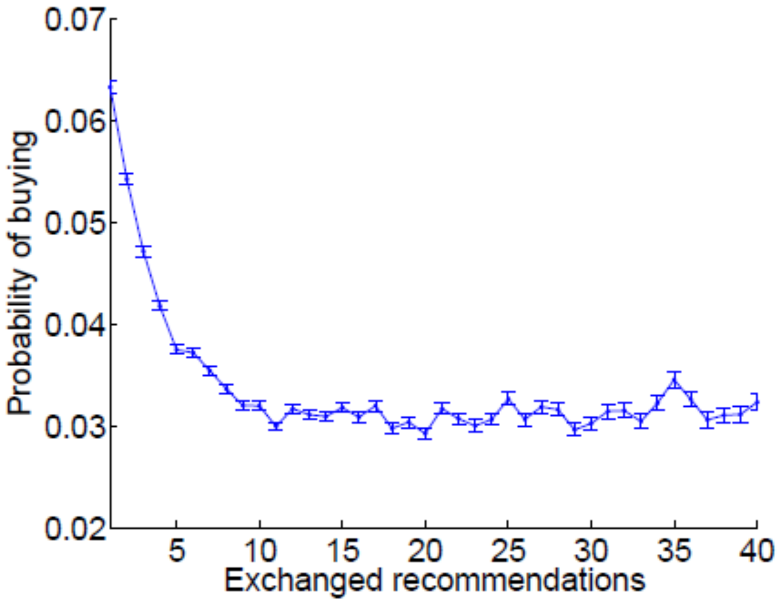
Real World Cascades (Leskovec, Singh, Kleinberg; PAKDD 06)

- Tend to be quite short
- Mainly stars or chains
- Power law distribution: frequency vs. depth



Real World Cascades (Leskovec 2008)

- Domain: Recommendation Network
- Certain product types are likely to be more contagious: more expensive types
- Limits to the influence of high degree nodes
- **Prob. of adopting quickly reaches saturation at a constant and relatively low probability**



Challenges for DERI

- Topic diffusion processes in large Web data sets
  - Can we detect cascade behaviour other than by explicit links
- Churn in user communities
  - Unusual type of cascade
- Diffusion behaviour in Scientific communities

References

- Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." Proc. Roy. Soc. Lond. A 115, 700-721, 1927.
- Epidemic Spreading in Scale-Free Networks, R. Pastor-Satorras and A. Vespignani Phys. Rev. Lett. 86 (2001).
- The Epidemic Threshold in Scale-Free Graphs, N. Berger, C. Borgs, J. Chayes, and A. Saberi, Symposium on Discrete Algorithms (SODA), (2005).
- T. Schelling. Micromotives and Macrobehavior. Norton, 1978.