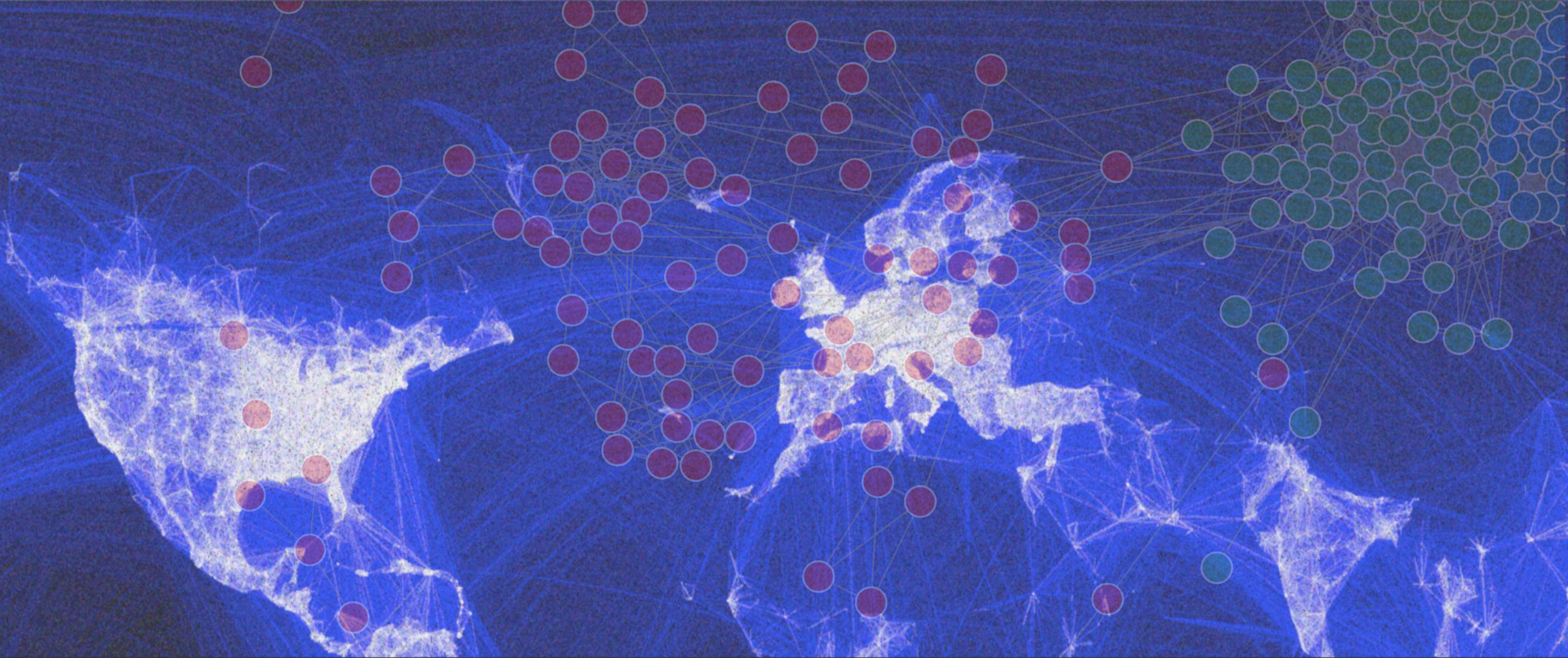


Complex Networks

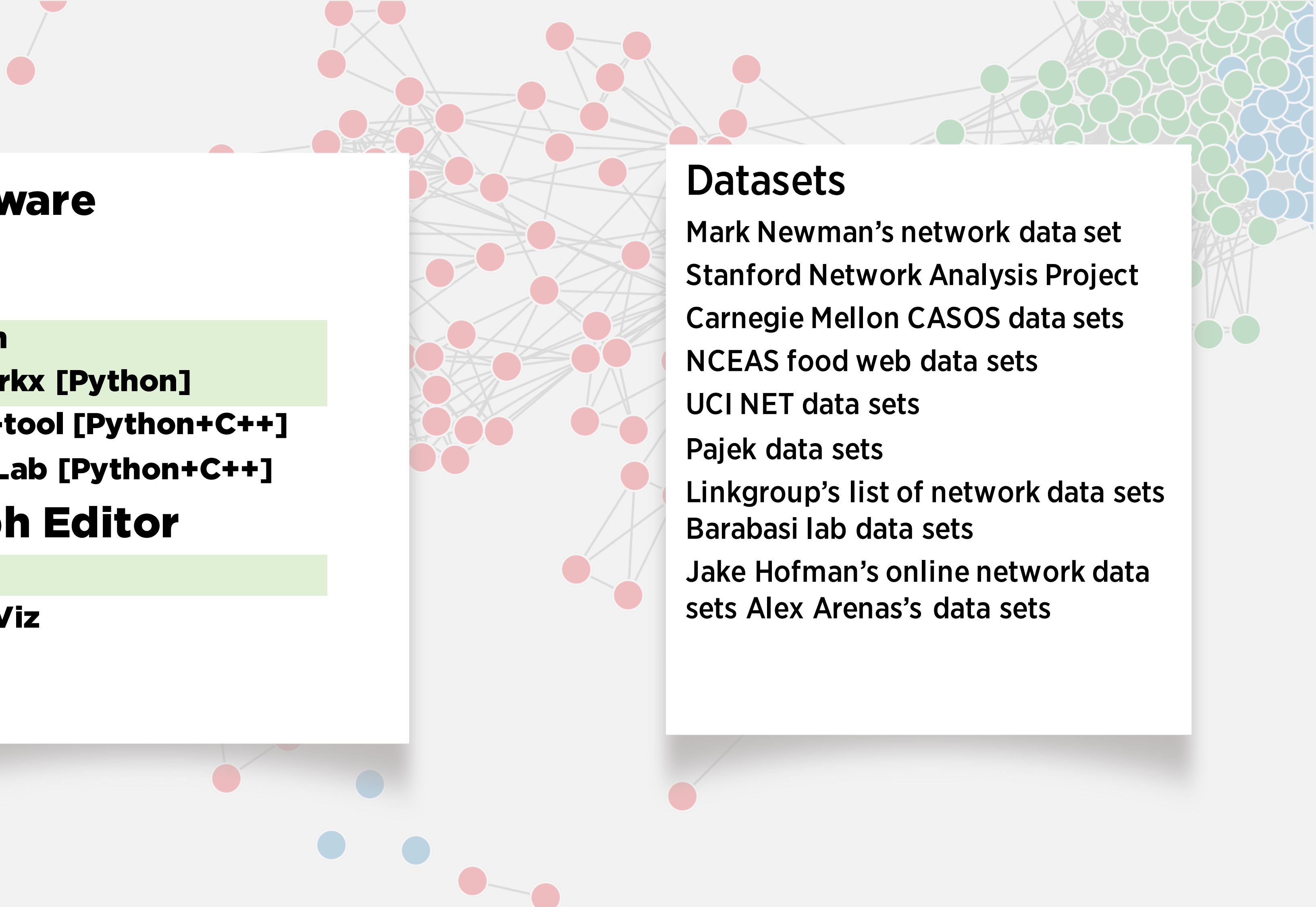
Vincent Gauthier

Associate Professor

@vincentgauthier

http://complex.luxbulb.org

Telecom SudParis-Institut Mines-Telecom/CNRS SAMOVAR



Software

R

Matlab

Python

Networkx [Python]

Graph-tool [Python+C++]

GraphLab [Python+C++]

Graph Editor

Gephi

GraphViz

Pajek

Datasets

Mark Newman’s network data set

Stanford Network Analysis Project

Carnegie Mellon CASOS data sets

NCEAS food web data sets

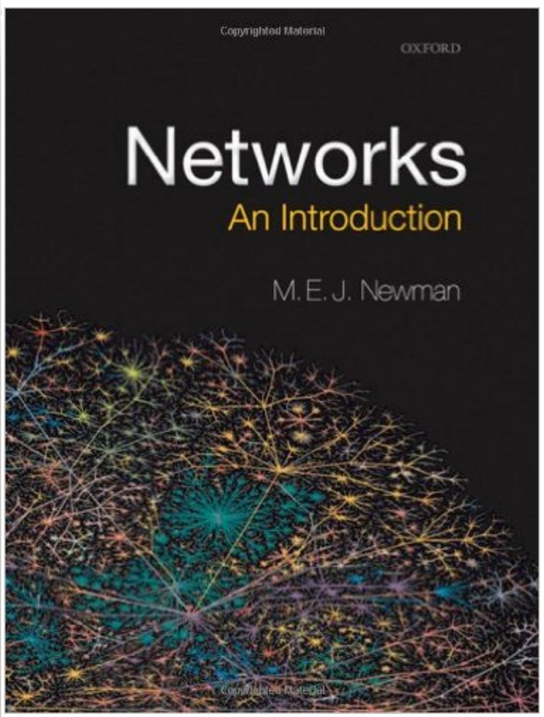
UCI NET data sets

Pajek data sets

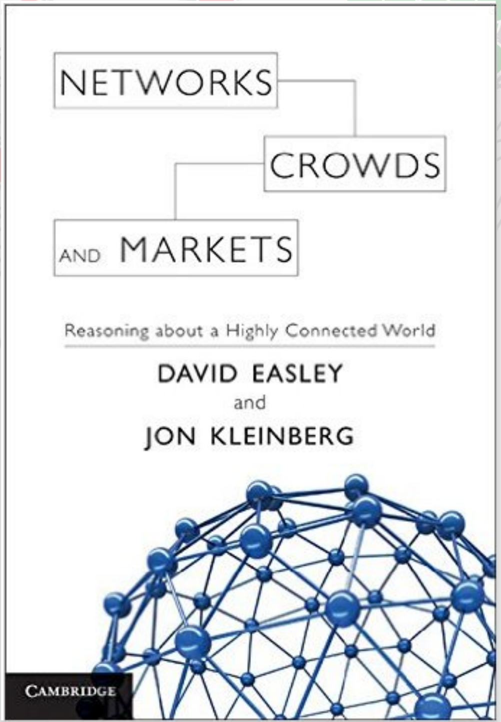
Linkgroup’s list of network data sets

Barabasi lab data sets

Jake Hofman’s online network data sets Alex Arenas’s data sets



Networks An Introduction
Mark Newman, 2010.



Networks, Crowds, and Markets
Reasoning About a Highly Connected World
David Easley, Jon Kleinberg, 2010.

What we learned

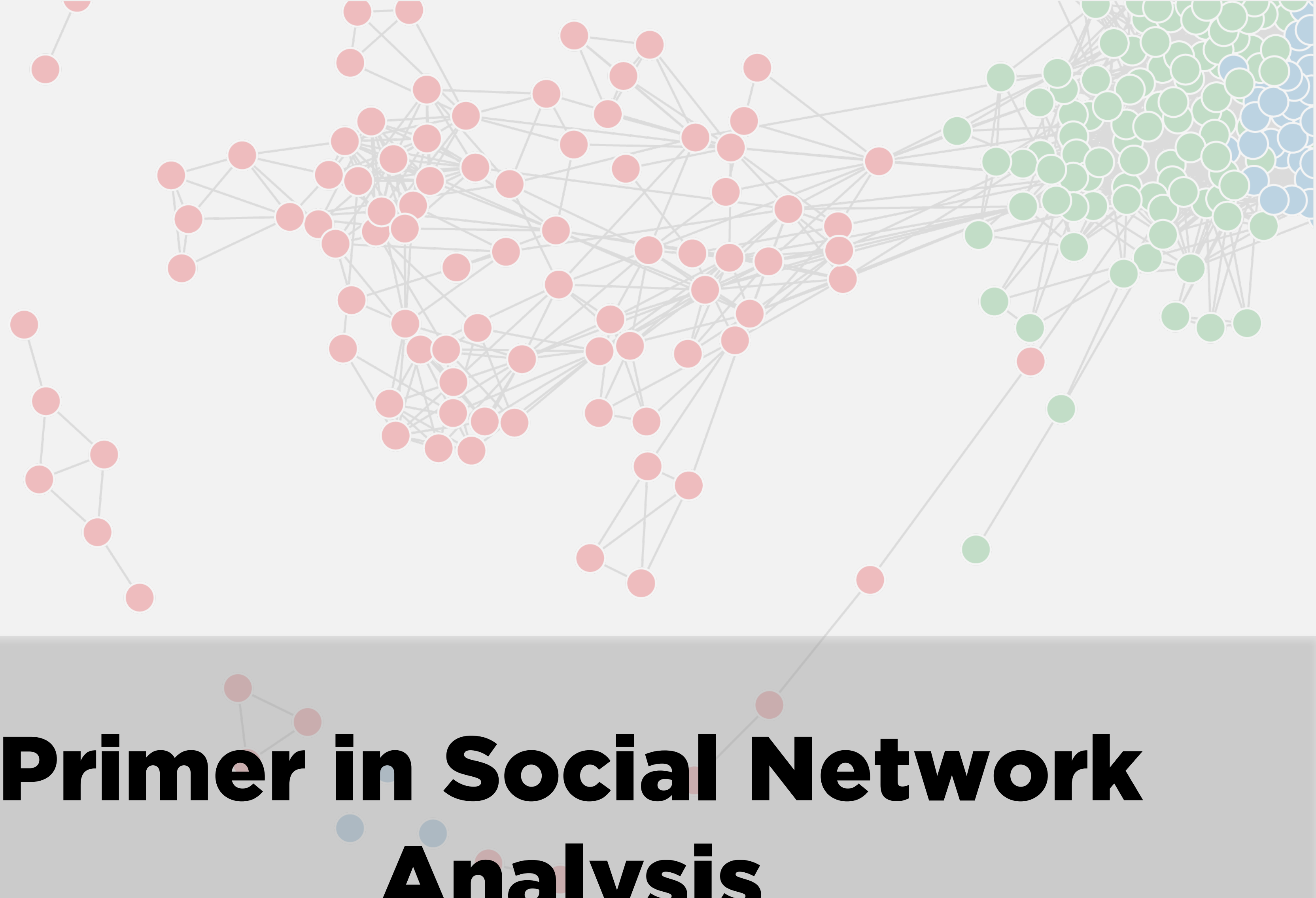
1. We can describe a network with various metrics:
   a) Closeness Centrality (Understanding whom are at the center of the network)
   b) Betweenness Centrality (whom is receiving more in-flux)
   c) Eigen Centrality, PageRank (whom is receiving more trust)
2. Once the average degree exceeds <k>= 1, a giant component should emerge that contains a finite fraction of all nodes
3. Clustering shape the way a network is designed
4. Most of the real networks has small diameter
5. Most of the real networks has a power law degree distribution: there are hubs

1. Definition of a network
2. Basics of graph theory
3. Describing a network
4. Network centrality
5. Eigen centrality, PageRank
6. Small world network (Watts-Strogatz)
7. Network Growth (Barabasi-Albert)
8. Social Network Analysis
9. Diffusion on network

Lecture 1

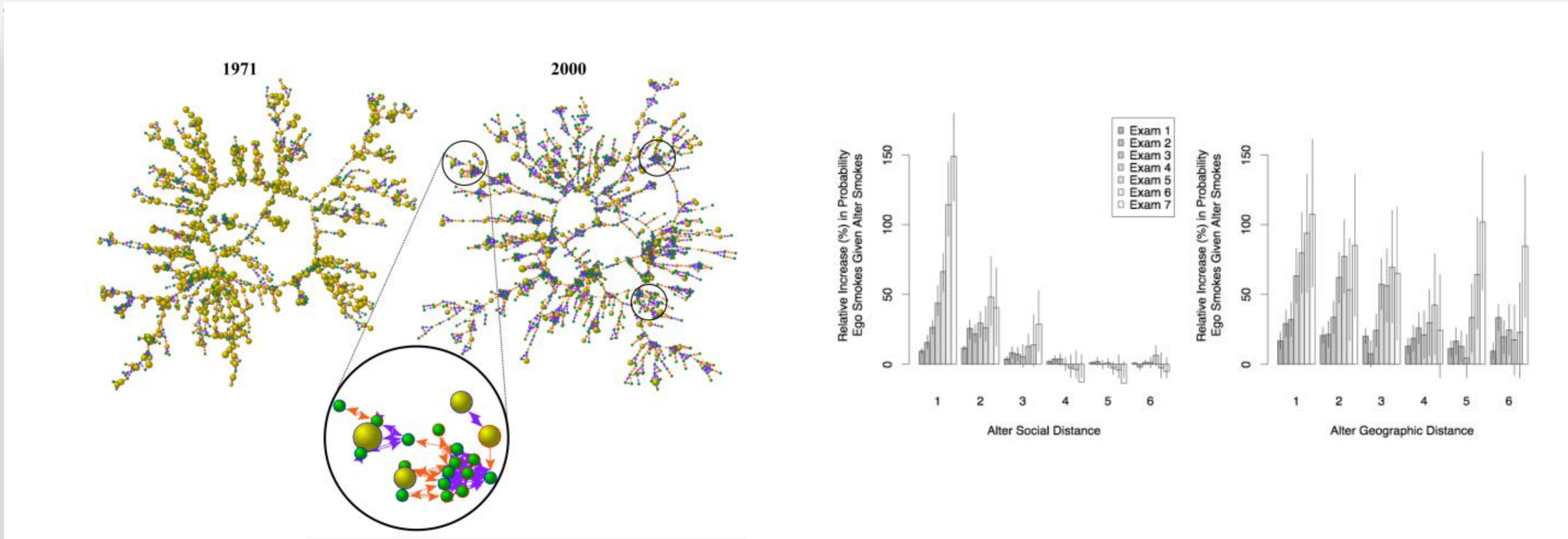
Lecture 2

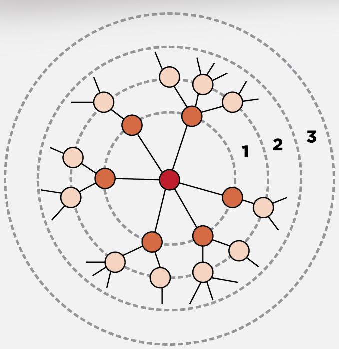
Lecture 3



Primer in Social Network Analysis

Homophily



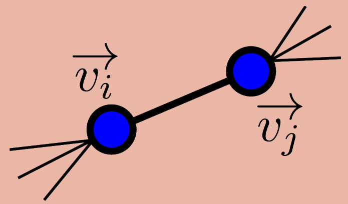


Homophily also known as assortativity:

- Fashion
- Smoking behavior
- Religious beliefs
- Voting
- Residential segregation

The Collective Dynamics of Smoking in a Large Social Network
Nicholas A. Christakis, James H. Fowler
The New England Journal of Medicine, 2008

Assortative mixing

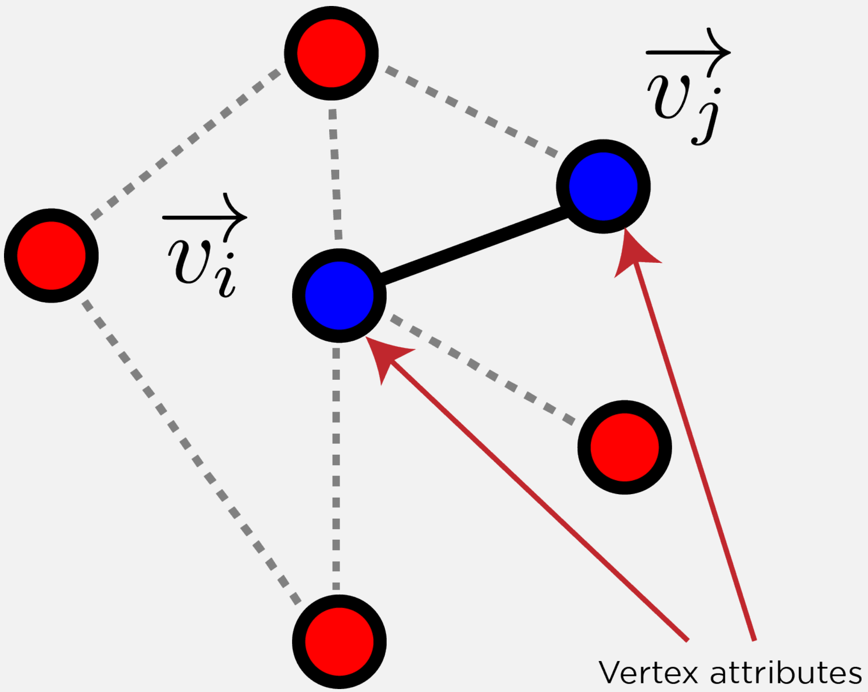


Homophily and assortative mixing:

The assortative coefficient r quantify the homophily

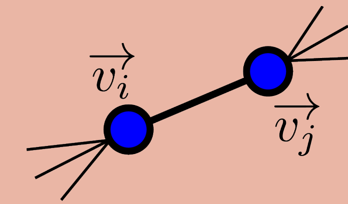
Type of vertex attributes:

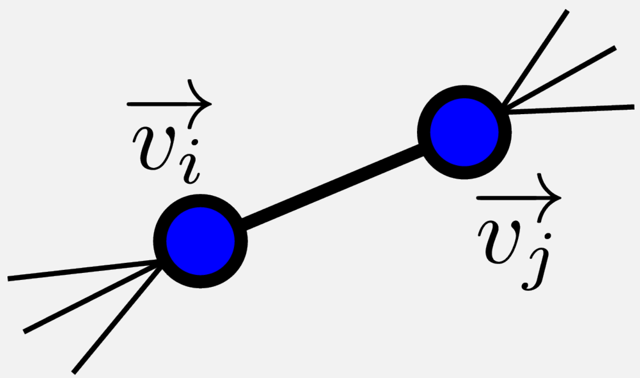
- Scalar value
- Vertex degrees
- Categorical attributes



Mixing patterns in networks
Mark Newman
Phys. Rev. E 67, 2003

Assortative mixing





homophily and assortative mixing

like links with like

scalar attributes:

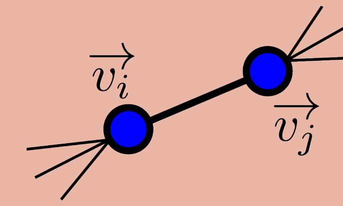
- mean value across ties

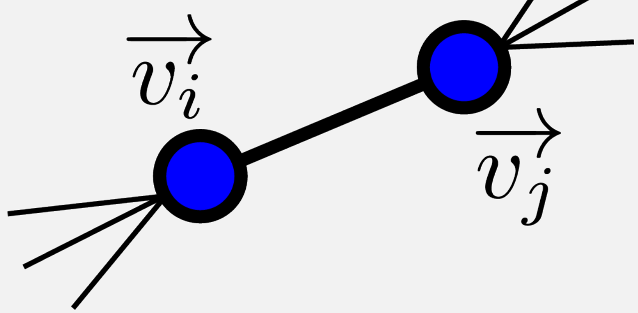
$$\mu = \frac{1}{2m}\sum_i \sum_j A_{ij} v_i$$

$$\boxed{= \frac{1}{2m}\sum_i k_i v_i}$$

Mixing patterns in networks
Mark Newman
Phys. Rev. E **67**, 2003

Assortative mixing





homophily and assortative mixing

like links with like

scalar attributes:

- covariance across ties

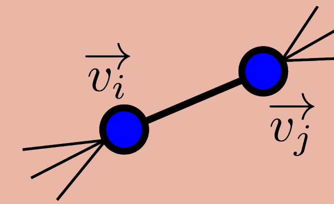
$$\mathrm{cov}(v_i, v_j) = \frac{\sum_{ij} A_{ij}(v_i - \mu)(v_i - \mu)}{\sum_{i,j} A_{ij}}$$

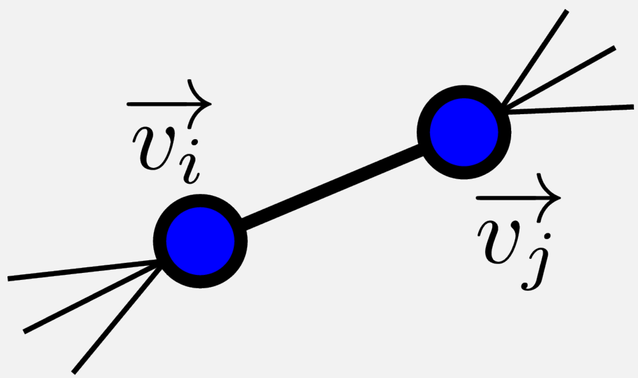
$$\left(\mu = \frac{1}{2m}\sum_i k_i v_i\right) \qquad = \sum_{i,j} A_{ij} v_i v_j - \mu^2$$

$$\boxed{= \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right) v_i v_j}$$

Mixing patterns in networks
Mark Newman
Phys. Rev. E 67, 2003

Assortative mixing





homophily and assortative mixing

like links with like

assortativity coefficient (scalar):

$$r = \frac{\mathrm{cov}(v_i, v_j)}{\mathrm{var}(v_i, v_j)}$$

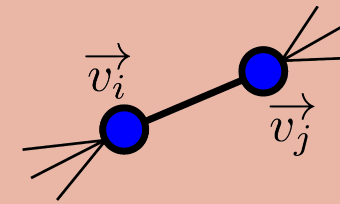
$$\boxed{= \frac{\sum_{i,j}\left(A_{ij} - k_i k_j / 2m\right) v_i v_j}{\sum_{i,j} k_i \delta ij - k_i k_j / 2m}}$$

[This is just a Pearson correlation across edges]

$$-1 < r < 1$$

📕 *Mixing patterns in networks*
Mark Newman
Phys. Rev. E 67, 2003

Assortative mixing



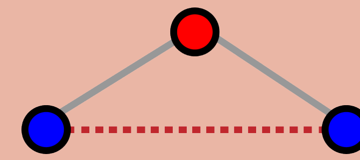
	Network	Type	Size	Degree assortativity	Error
Social	physics coauthorship	undirected	52 909	0.363	0.002
	biology coauthorship	undirected	1 520 251	0.127	0.0004
	mathematics coauthorship	undirected	253 339	0.120	0.002
	film actor collaboration	undirected	449 913	0.208	0.0002
	company directors	undirected	7 673	0.276	0.004
	student relationships	undirected	573	−0.029	0.037
Technological	Email address books	directed	16 881	0.092	0.004
	Power grid	undirected	4 941	−0.003	0.013
	Internet	undirected	10 697	−0.189	0.002
	WWW	directed	269 504	−0.067	0.0002
	software dependencies	directed	3 162	−0.016	0.020
Biological	protein interactions	undirected	2 115	−0.156	0.010
	metabolic network	undirected	765	−0.240	0.007
	neural network	directed	307	−0.226	0.016
	marine food web	directed	134	−0.263	0.037

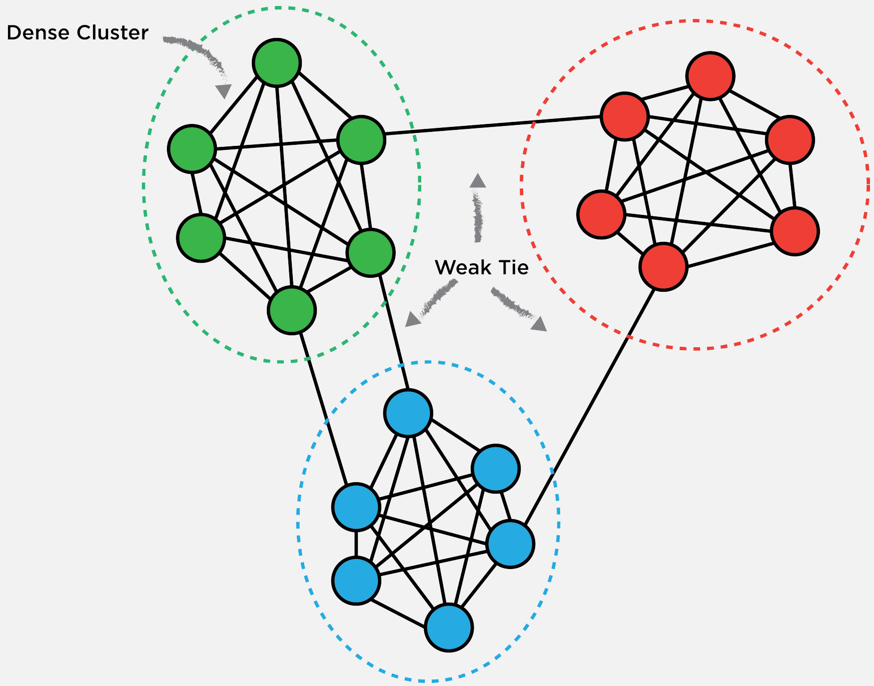
QUIZ: What type of assortativity mixing appear in technological system?

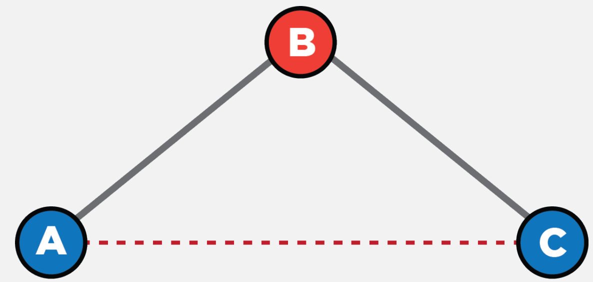
Internet: highly connected router tend to connect with:

a) Poorly connected router
b) Highly connected router

Transitivity





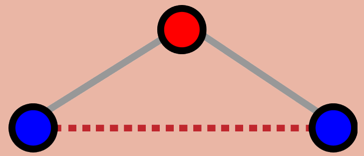


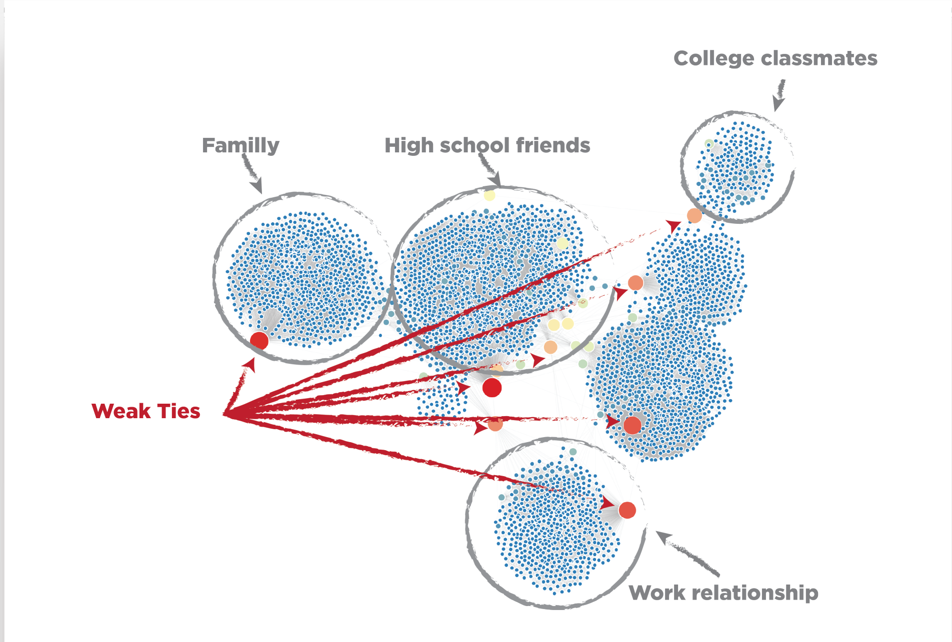
Transitivity: if there is a link between A and B and B and C then in a transitive network A and C will be connected (Triadic closure):

- Strong tie are often transitive
- Transitivity and homophily lead to the formation of clique (fully connect cluster)

- Inside community, triplet tend to close and form a triangle. (i.e my friends are also friends
- Simple assumption, in reality a more complex process take place

Strength of weak ties



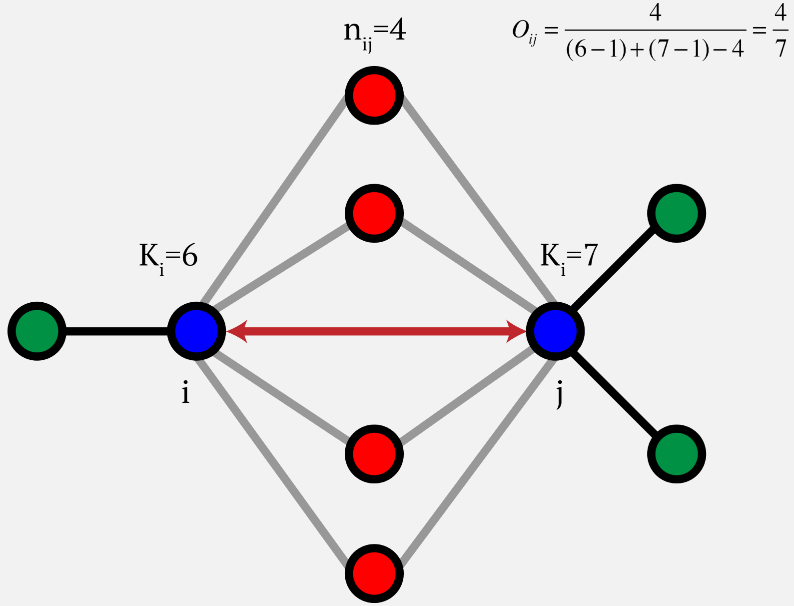


- information spread quickly within communities
- spread slowly across communities
- bridge node enable information to be spread across the entire network
- most new job opportunities from "weak ties"

The Strengh of weak ties
Mark S. Granovetter
American journal of sociology, 1973

Embeddeness: strength of a link

How many common neighbors a pair of node has ?

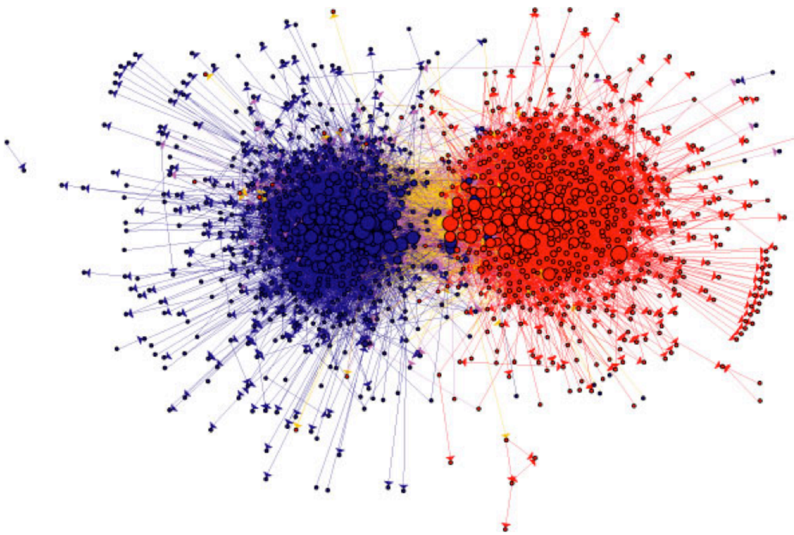


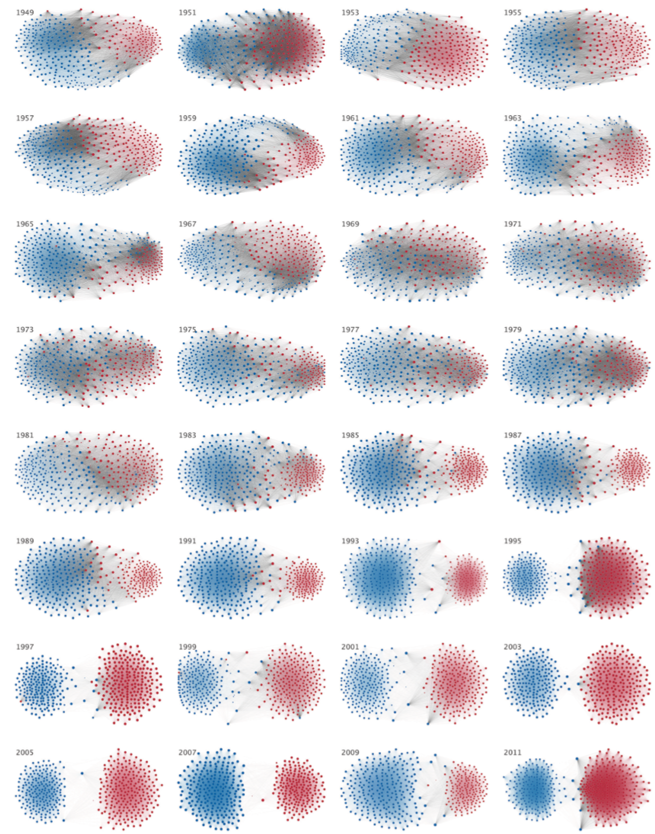
How strong is a link between two people

Structure and tie strengths in mobile communication networks
J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási
PNAS 2007

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}$$

People form group (communities)



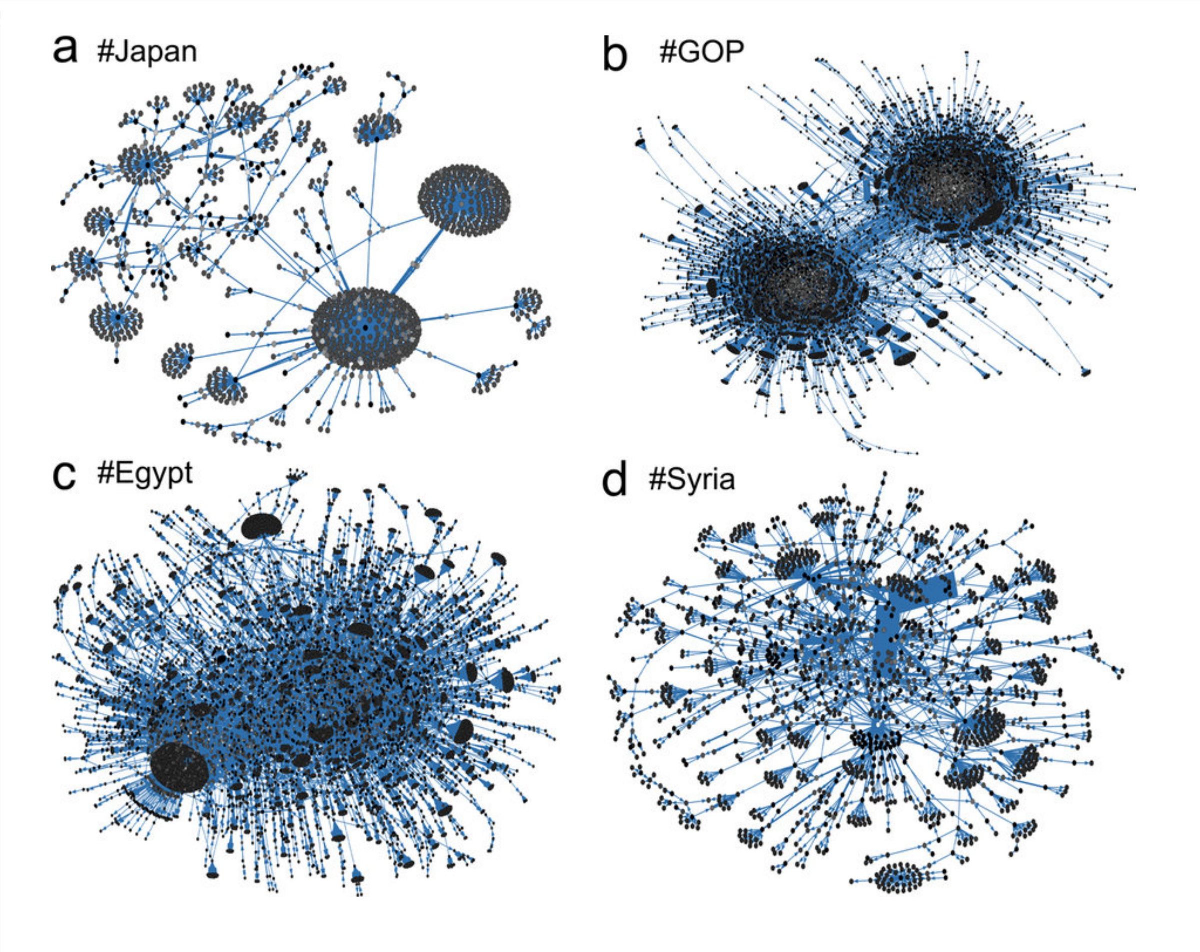


- Homophily: bird of a feather flock together
- In social network node tend to form group (or community)

The political blogosphere and the 2004 US election
L. A. Adamic, N. Glance
LinkKDD '05, 2005

The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives
Clio Andris et al
PlosOne, 2015

Memes Diffusion

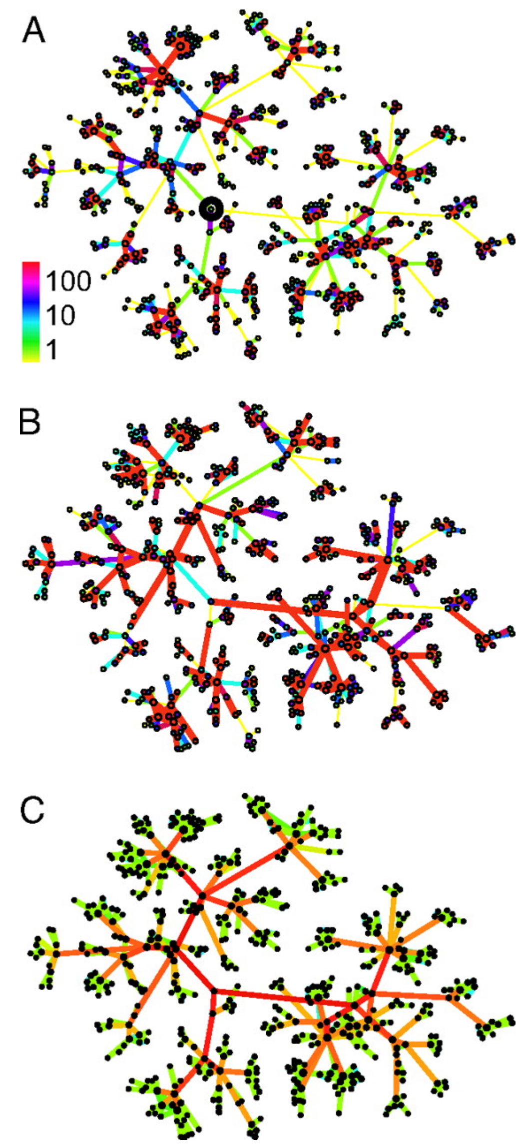


Nodes represent Twitter users, and directed edges represent retweeted posts that carry the meme. The brightness of a node indicates the activity (number of retweets) of a user, and the weight of an edge reflects the number of retweets between two users. (a) The **#Japan** meme shows how news about the March 2011 earthquake propagated. (b) The **#GOP** tag stands for the US Republican Party and as many political memes, displays a strong polarization between people with opposing views. Memes related to the “Arab Spring” and in particular the 2011 uprisings in (c) **#Egypt** and (d) **#Syria** display characteristic hub users and strong connections, respectively.

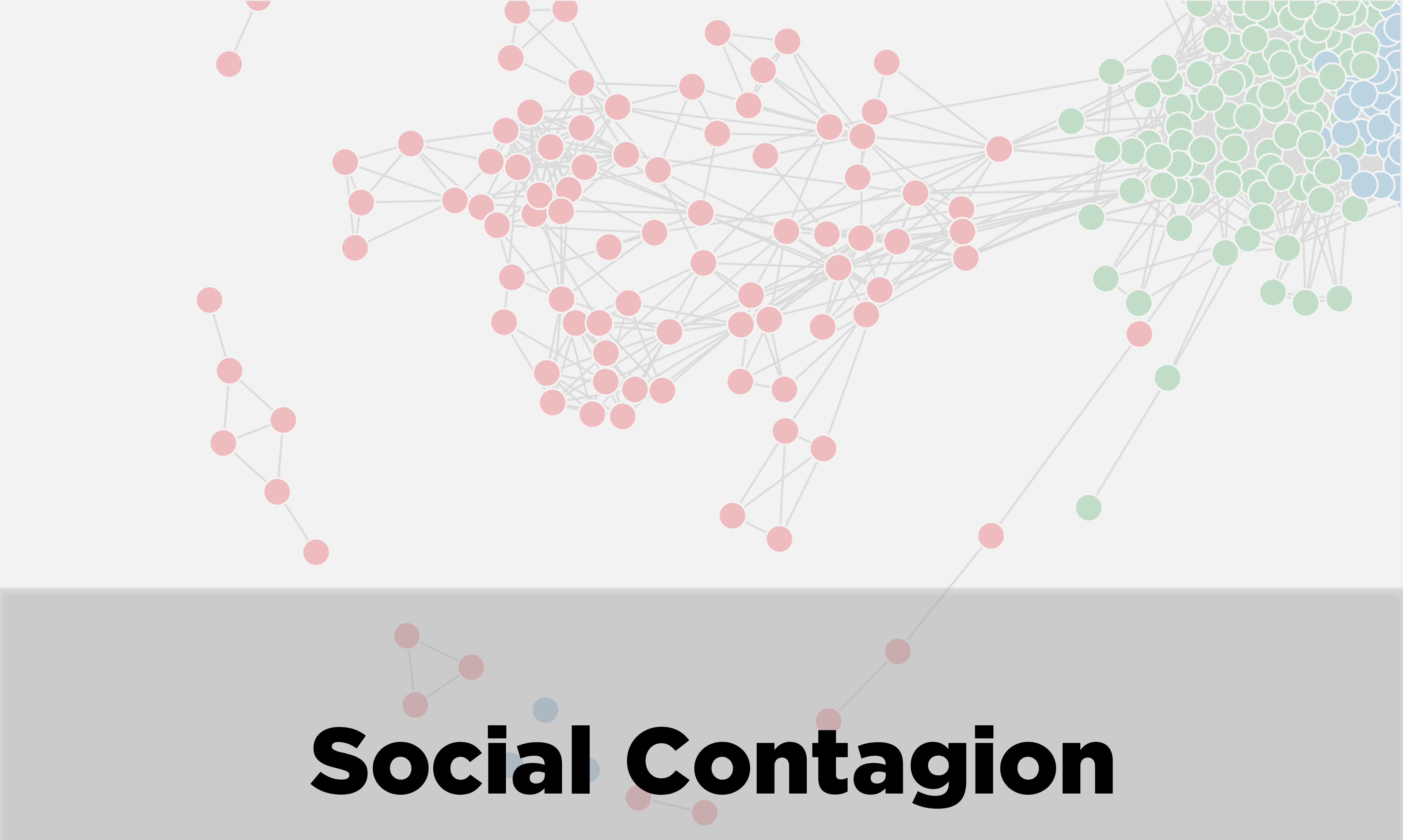
Competition among memes in a world with limited attention
L. Weng, A. Flammini, A. Vespignani & F. Menczer
Scientific Reports 2012

Conclusion

- clustering and embeddedness play a important role especially in social network
- there is an interplay between network diameter and the clustering coefficient
- in social network node tend to form group (or community)
- strong tie and weak tie both play an important role



We need a better understanding of the link between clustering coefficient and the network diameter



Social Contagion

Social contagion

Two focuses for us

- Widespread media influence
- Word-of-mouth influence

We need to understand influence

- Who influences whom? Very hard to measure...
- What kinds of influence response functions are there?
- Are some individuals super influencers?
- The infectious idea of opinion leaders (Katz and Lazarsfeld). Remember the eigen centrality

Social contagion
continued

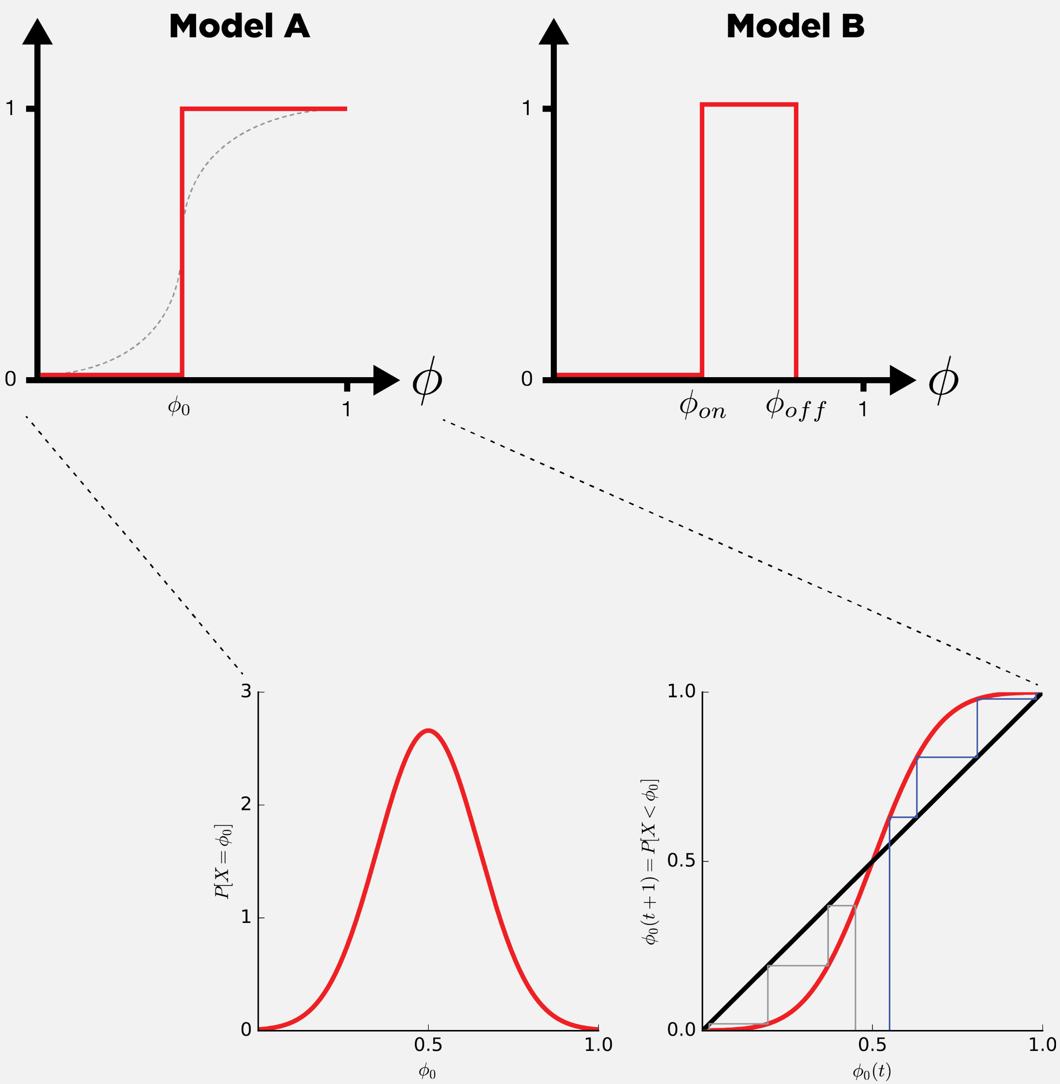
Some important model:

- Tipping models—Schelling (1971)
  - Simulation on checker boards
  - Idea of thresholds
- Thresholds model – Granovetter (1978)
- Herding models—Bikhchandani, Hirschleifer, Welch (1992)
  - Social learning theory, information cascade

Learning from the behavior of others: Conformity, fads, and informational cascades
S. Bikhchandani, D. Hirshleifer, and I. Welch.
J. Econ. Perspect.,, 1998

Threshold models of interpersonal effects in consumer demand
M. S. Granovetter and R. Soong.
J. Econ. Behav. Organ., 1986

Social contagion model



- The threshold is deterministic of stochastic
- ϕ = Fraction of contact ‘active’
- Basic idea: individuals adopt a behavior when a certain fraction of others have adopted
- ‘Others ′may be everyone in a population, an individual’s close friends, any reference group
- Response can be probabilistic or deterministic
- Individual thresholds can vary
- Assumption: order of others′ adoption does not matter… (unrealistic).
- Assumption: level of influence per person is uniform (unrealistic)

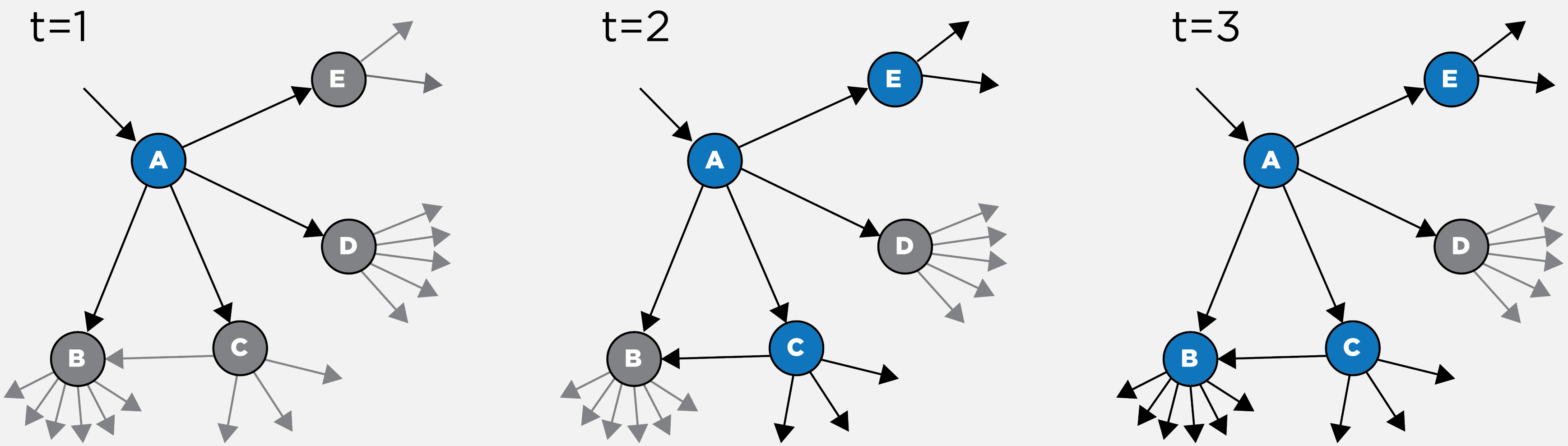
Threshold model on a network

- Interactions between individuals now represented by a network.
- Network is sparse
- An individual i as k_i contacts
- Each individual i has a fixed threshold ϕ_i
- An individual i becomes active when fraction of active contacts $\frac{a_i}{k_i} > \phi_i$

A simple model of global cascades on random networks
Duncan J. Watts
PNAS, 2002

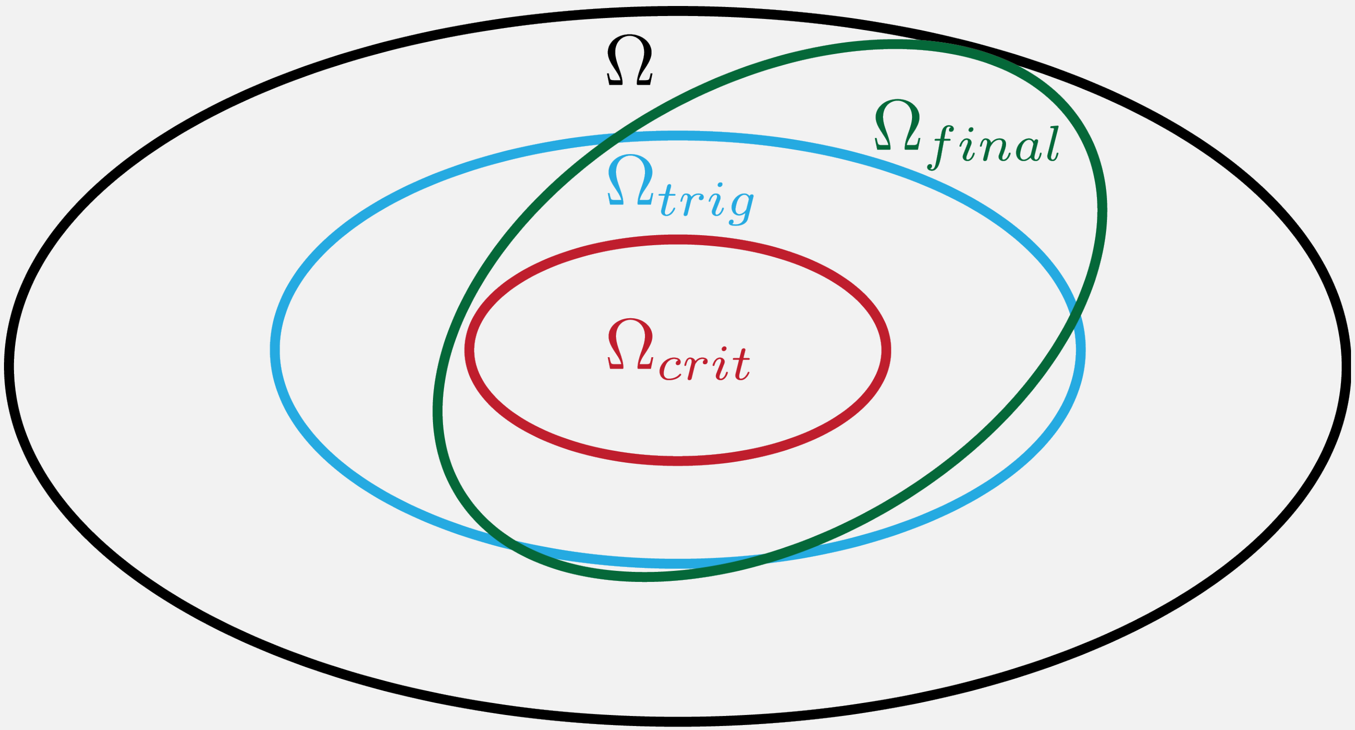
Influentials, networks, and public opinion formation
D. J. Watts and P. S. Dodds
Journal of Consumer Research, 2007.

Threshold model on a network



All nodes have a threshold $\phi_0 = 0.2$

Social contagion model



- Ω_{crit} global vulnerable component
- Ω_{trig} triggering component
- Ω_{final} potential extent of the spread
- Ω entire network

A simple model of global cascades on random networks
Duncan J. Watts
PNAS, 2002.

Snowballing

Cascade condition:

- If one individual is initially activated, what is the probability that an activation will spread over a network?
- What features of a network determine wheher a cascade occur ?

Snowballing:

- An active link is a link connected to an activated node
- If an infected link leads to at least 1 more infected link, then activation spreads
- We need to understand which nodes can be activated when only one of their neighbors becomes active

Snowballing

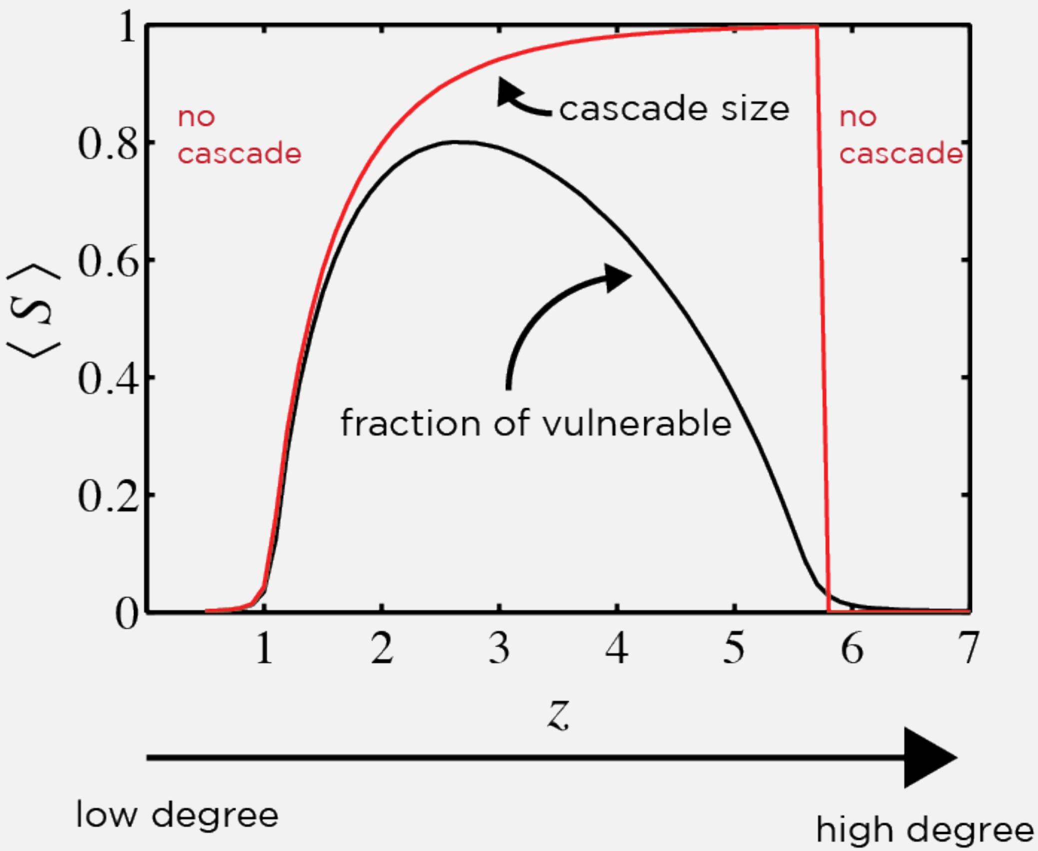
Vulnerable:

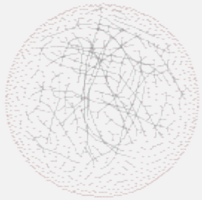
- We call individual who can be activated by just one contact being active vulnerable
- The vulnerability condition for node i

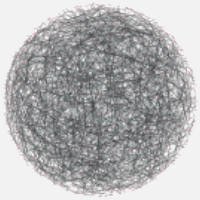
$$1/k_i \geq \phi_0$$

- Which mean #contacts $k_i \leq \lfloor 1/\phi_0 \rfloor$
- For a global cascade on random network, must have a global cluster of vulnerable
- On network :
  - One node $\longrightarrow$ critical mass $\longrightarrow$ everyone

Cascades on random network

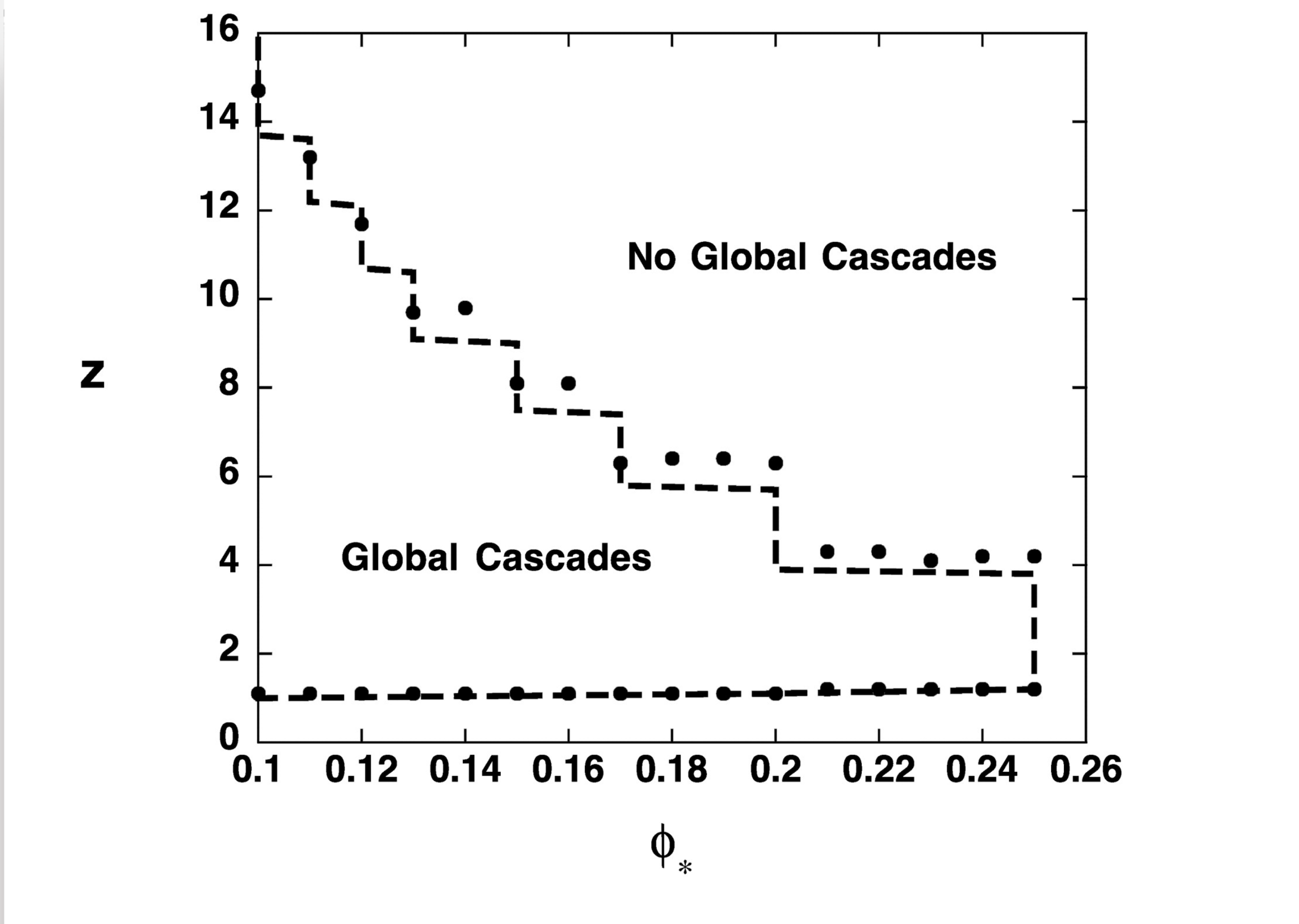






- Cascades occur only if max vulnerable > 0
- System maybe "robust yet fragile"
- Myopic view of the system enable cascades (ignorance enable spreading)

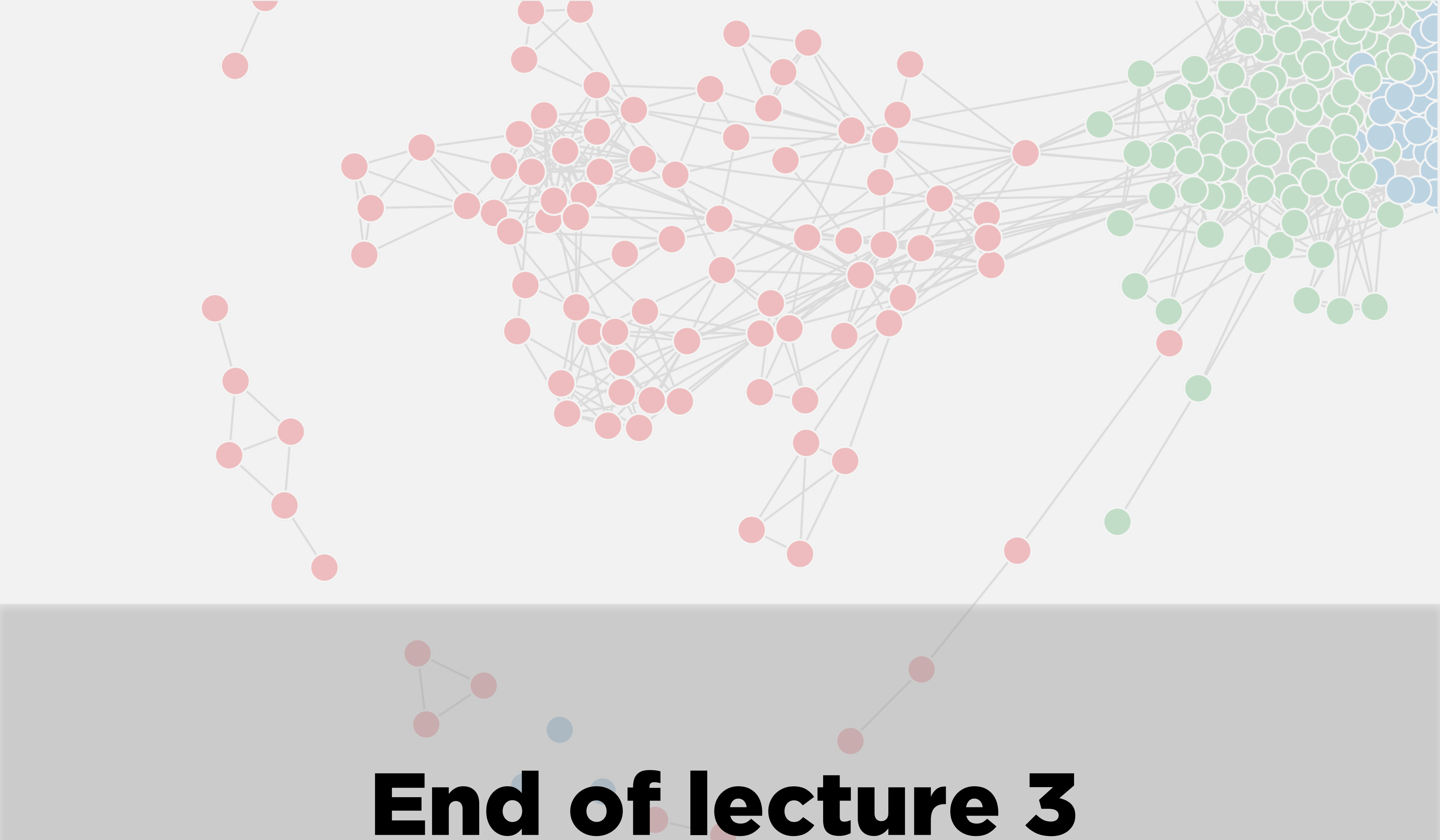
Cascade windows for the threshold model



- Cascade window widens as threshold ϕ decreases
- Lower thresholds enable spreading

A simple model of global cascades on random networks
Duncan J. Watts
PNAS, 2002.

How to start a social movement



End of lecture 3

Credits

- Complex Network Peter Dodds
- Five Lectures on Networks, Aaron Clauset
- Network Science course, Albert-László Barabási

- Networks An Introduction Mark Newman, 2010.
- Networks, Crowds, and Markets, D. Easley, J. Kleinberg, 2010.
- Network Science, Albert-László Barabási, 2016.

Origins of small worlds: efficient network example trade-off between wiring and connectivity

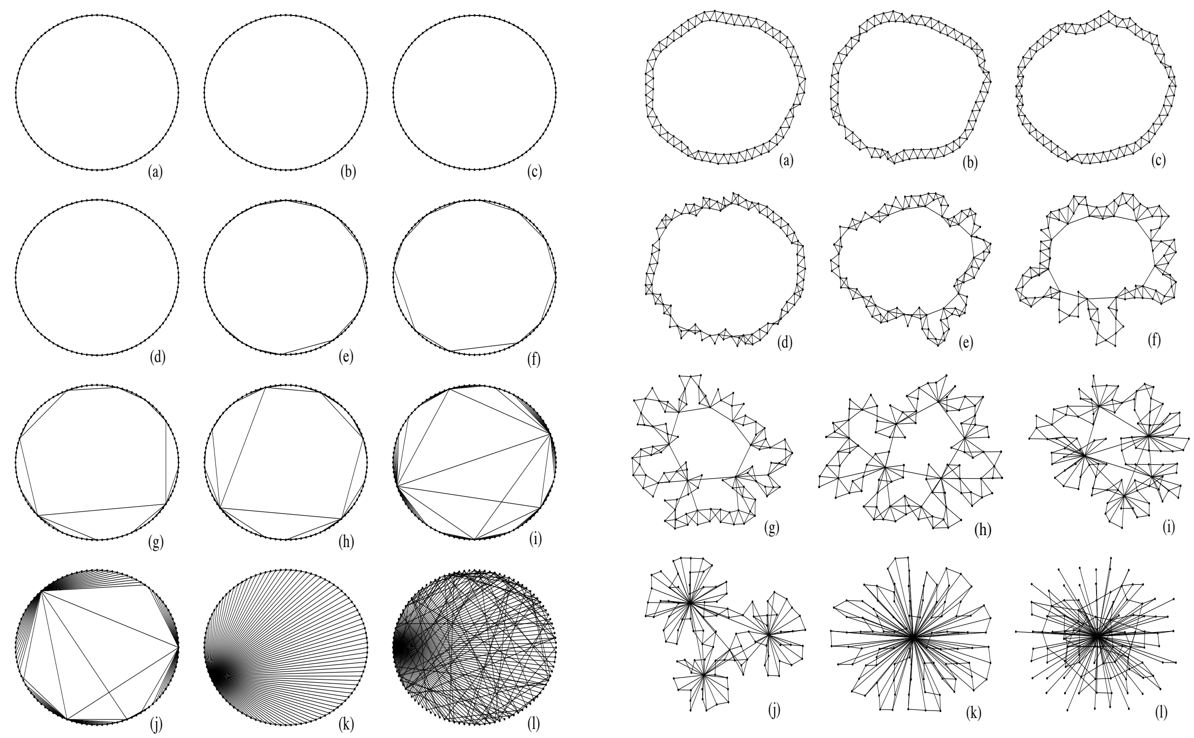
Rewiring according to a optimization:

$$E = \lambda L + (1-\lambda)W$$

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

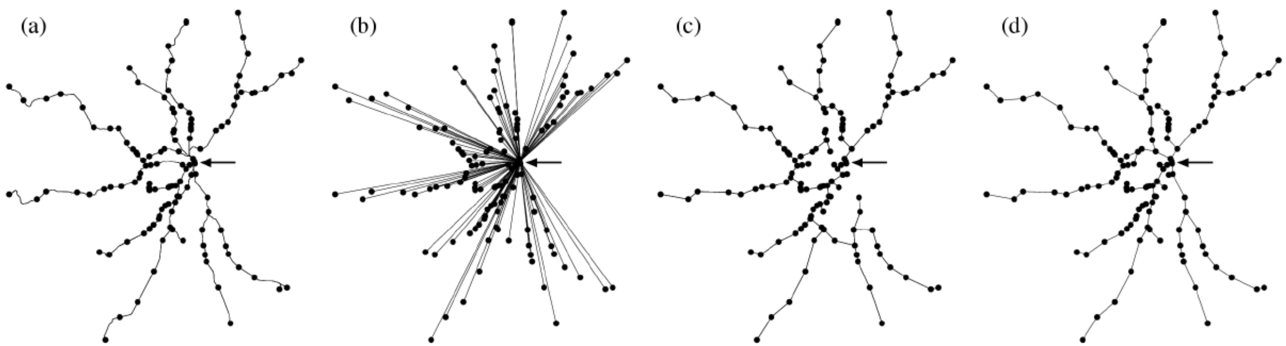
$$W = \sum_{e_{ij}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

- E is the 'energy' cost we are trying to minimize
- L is the average shortest path in 'hops'
- W is the total length of wire used



Small worlds: How and why
Nisha Mathias and Venkatesh Gopal
Phys. Rev. E 63, 021117, 2001.

Optimizing from scratch



(a) Commuter rail network in the Boston area. The arrow marks the assumed root of the network

(b) Star graph

(c) Minimum spanning tree

(d) The model applied to the same set of stations

add edge with smallest weight $W_{ij} = d_{ij} + \beta l_{j0}$ # hops to root node

Euclidian distance between i and j

The Spatial Structure of Networks, M. T. Gastner and M. E.J. Newman

Quiz:

- A network that contains many hubs with far reaching edges is indicative of (check all that apply)
  - high cost of distance traveled
  - low cost of distance traveled
  - high cost of making many hops
  - low cost of making many hops

Price's preferential attachment model

citation networks

Price 1965

- each new paper is generated with m citations (mean)
- new papers cite previous papers with probability proportional to their indegree (citations)
- what about papers without any citations?
  - each paper is considered to have a "default" citation
  - probability of citing a paper with degree k, proportional to k+1

Power law with exponent $\alpha = 2+1/m$