Weighted networks (continued)

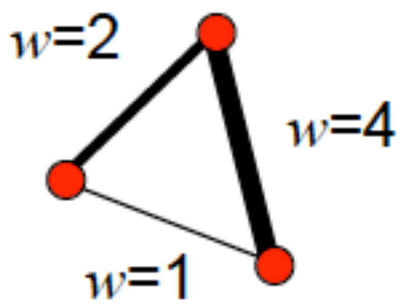
Weighted Networks: Weight-Topology Correlations

- Instead of simply attempting to generalize existing measures, it might be better to focus on **correlations between weights and topology**
- Question: What role do edges of different weight play in the network?
- E.g. transport networks: edge weights are high where capacity is needed, i.e. where betweenness centrality is high
- There are different scales to this problem: the global, macroscopic scale of overall connectivity and the mesoscopic scale of clusters and modules

Weighted Subgraphs: Intensity

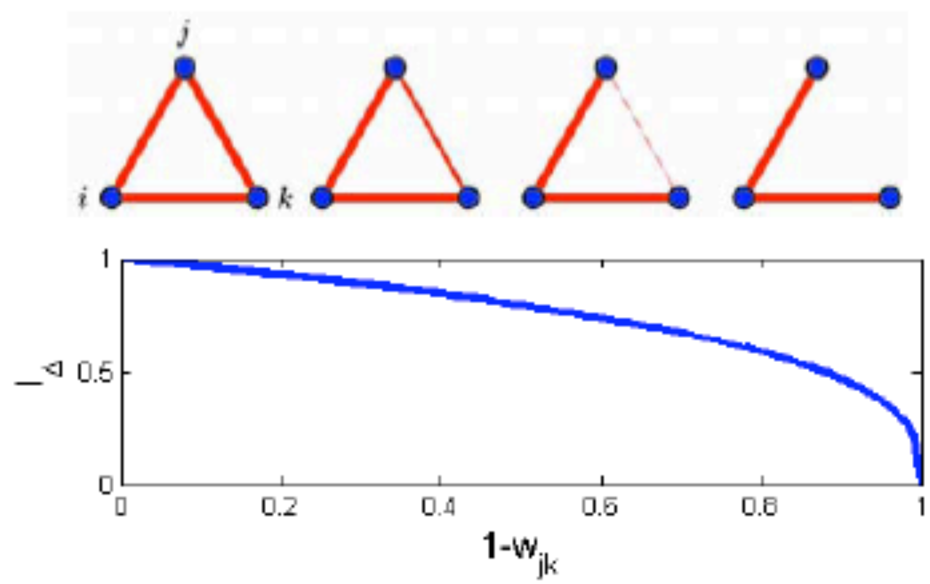
- The **intensity** I of a subgraph g with nodes v_g and links l_g, such that $|l_g|$ is the number of links in g, is defined as

$$I(g) = \left(\prod_{ij \in l_g} w_{ij} \right)^{1/|l_g|}$$



$$I(\Delta) = (1 \times 2 \times 4)^{1/3} = 2$$

- Measures the "weight" of a subgraph
- Becomes low if any of the weights is low



[Onnela, Saramäki et al, Phys Rev E **71**, 065103 (2005)]

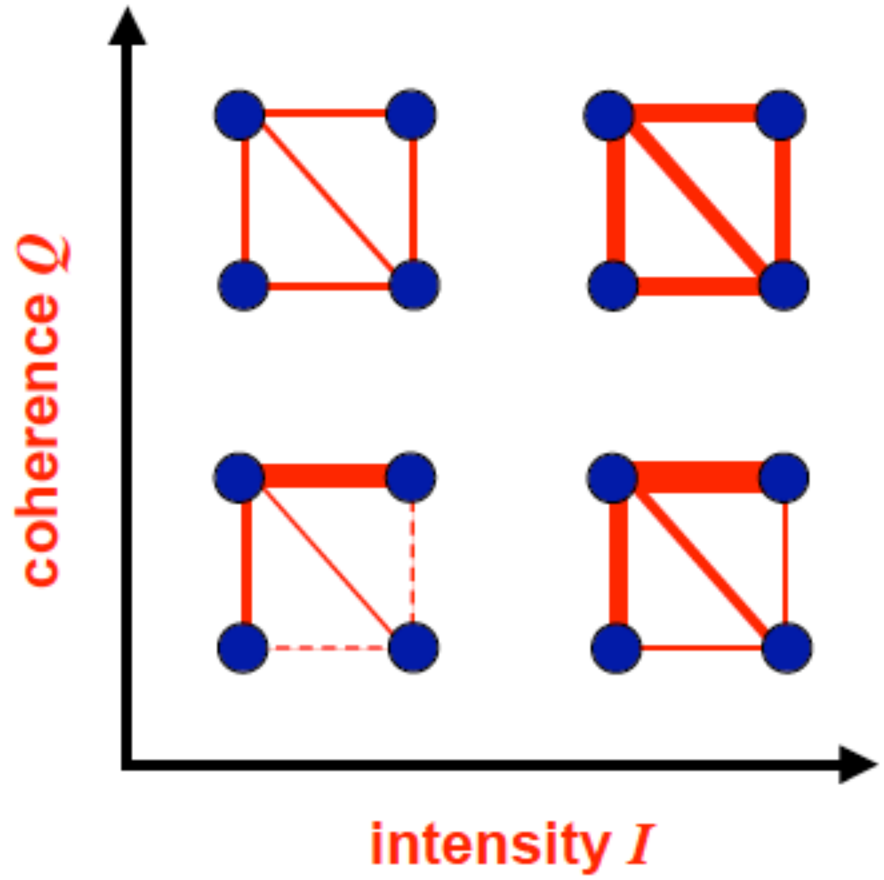
Weighted Subgraphs: Coherence

- The **coherence Q** of a subgraph g with nodes v_g and links l_g, such that $|l_g|$ is the number of links in g, is defined as

$$Q(g) = \frac{I(g)}{\frac{1}{|l_g|}\sum_{ij \in g} w_{ij}}$$

- Measures how equal ("coherent") the weights are

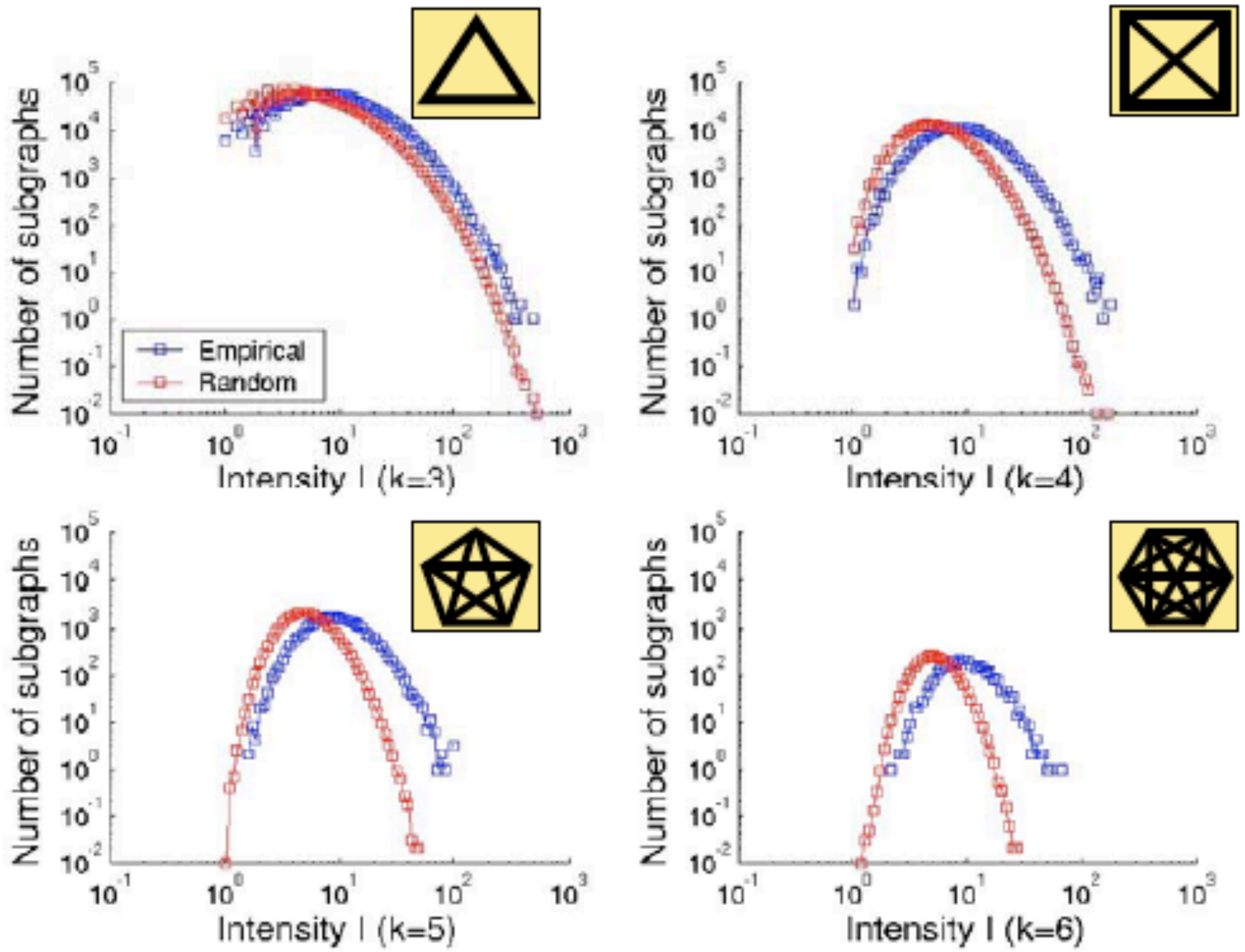
- If all weights $w_{ij} = w$, $Q(g)=1$



[Onnela, Saramäki et al, Phys Rev E **71**, 065103 (2005)]

Example: Intensity Distributions

- Data from a large social network inferred from mobile telephony call records
- Weight = total duration of calls between i and j in 18 weeks
- **Blue = original network, red = reference ensemble average**
- Reference ensemble: same network, weights permuted

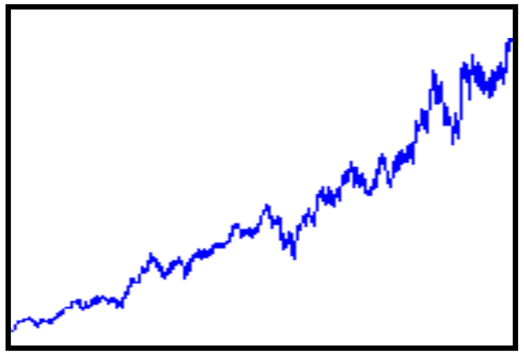


Onnela et al, New Journal of Physics 9, 179 (2007)

Weighted networks from time series
Example: stock networks

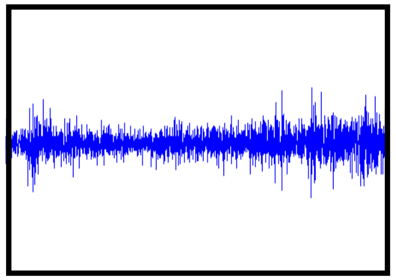
1. Starting point: time series of prices for N stocks

$$P_i(t)$$

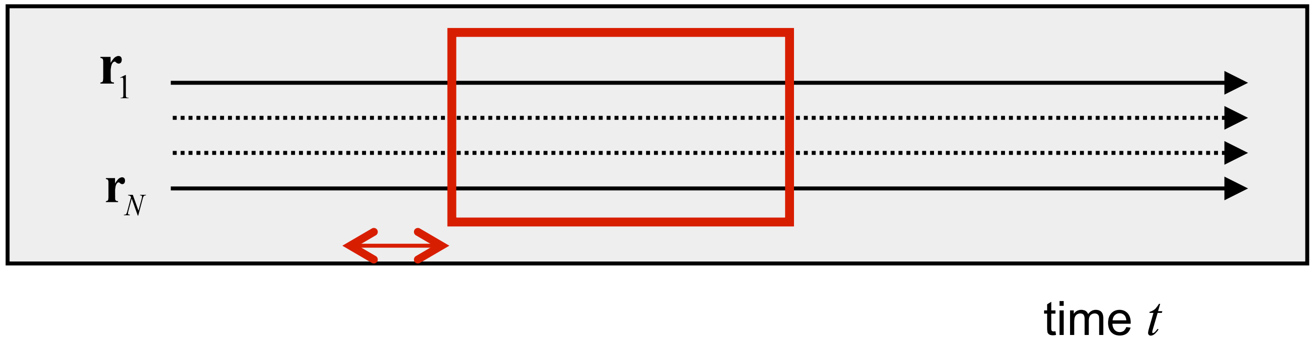


2. Calculate log-returns (usually done for stock prices, not feasible for all time series)

$$r_i(t) = \log \frac{P_i(t)}{P_i(t-1)}$$



3. Divide data into windows, or use the whole data length



(Note: time series length in data points must be >> N)

Weighted networks from time series
Example: stock networks

4. Calculate correlation coefficients between time series within your windows →

$$\rho_{ij}^{\tau} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2}\sqrt{\langle r_j^2 \rangle - \langle r_j \rangle^2}}$$

$$-1 \leq \rho_{ij}^{\tau} \leq 1$$

5. If negative coefficients exist, take absolute values

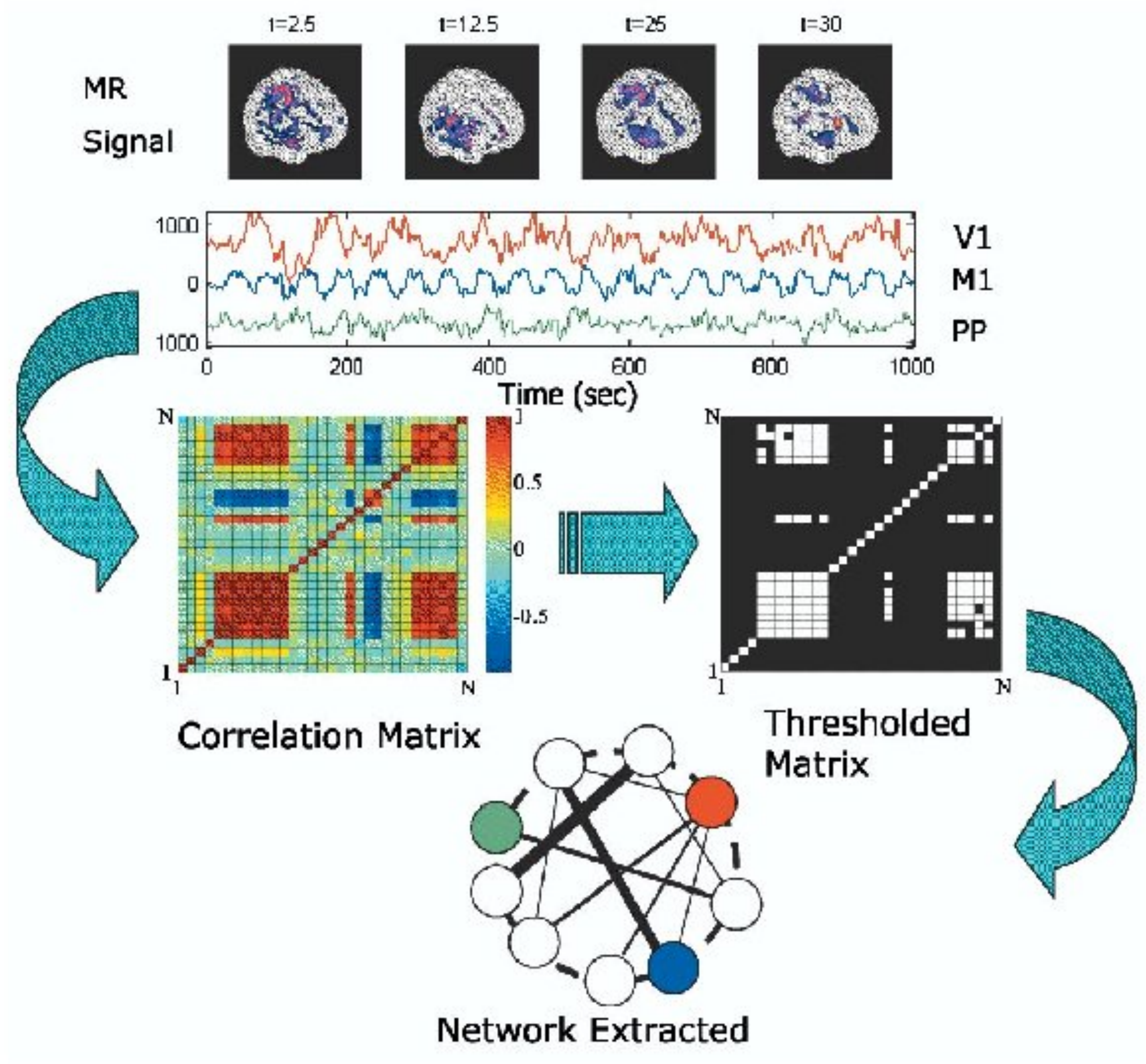
6. This correlation matrix is your weight matrix W_{ij}!

7. Treat as a (full) weighted network - calculate MST, use thresholding, or similar

Weighted networks from time series
Example: brain functional networks

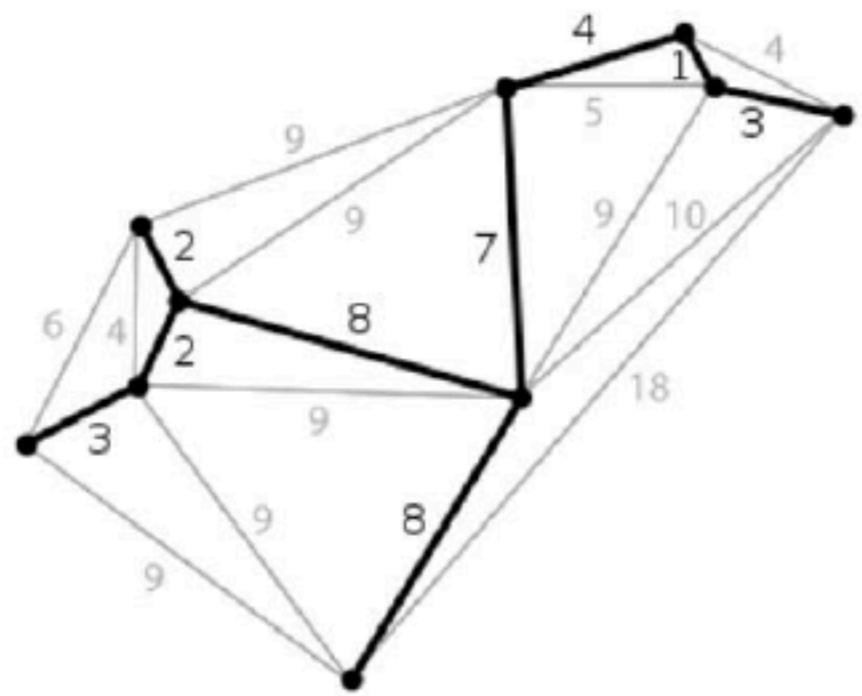
• **Scale-free functional brain networks,** Eguiluz et al, PRL **94**, 018102 (2005):

• Method:
  - use FMRI time series on activity of small voxels
  - construct correlation matrix
  - leave only highest correlations
  - construct network



Maximal/Minimal Spanning Trees

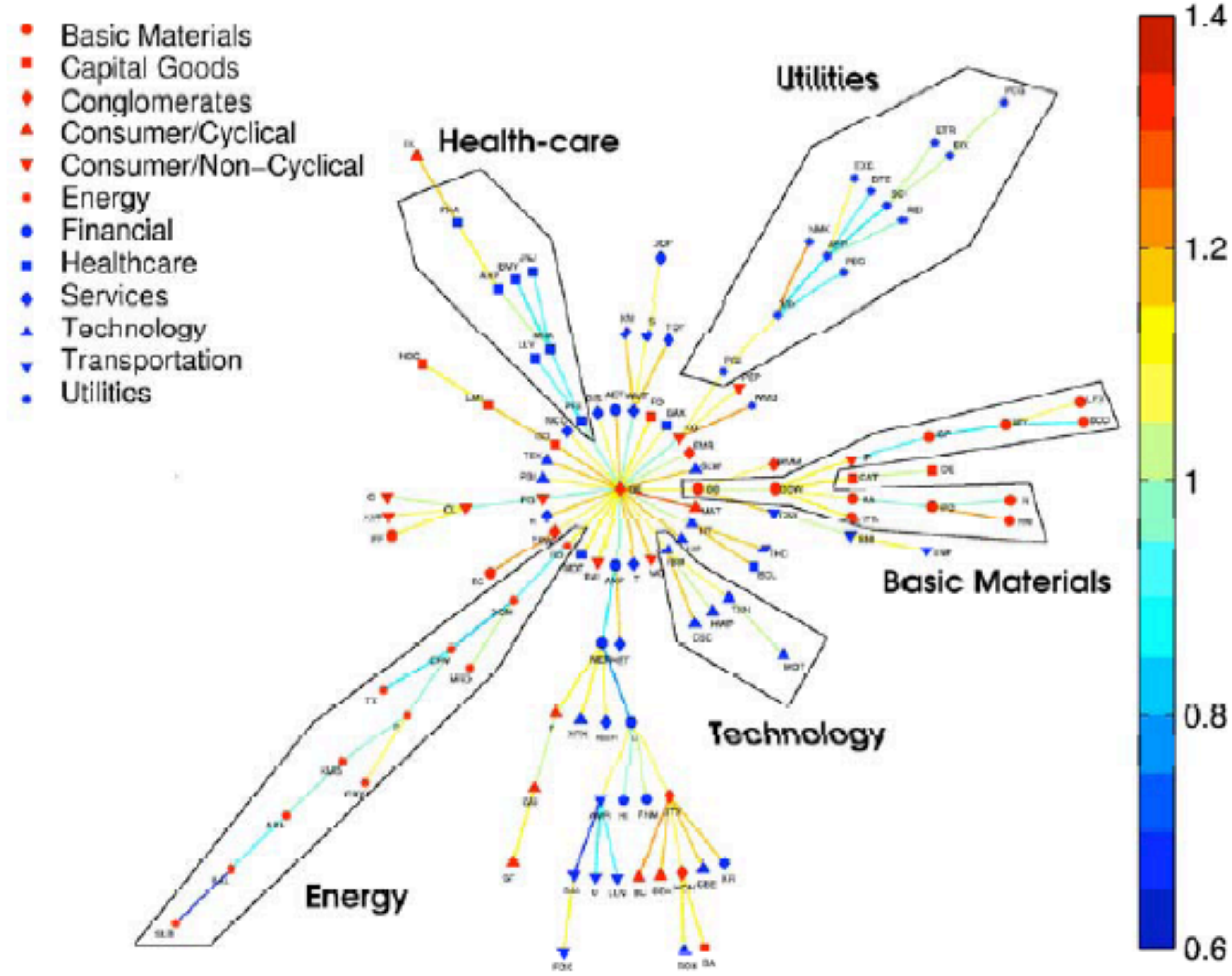
- Idea: to distill the "essential skeleton" of a weighted network
- Works as well for full matrices; can be used to transform a weight matrix into a tree
- Definition: a tree containing all the nodes in the network, such that the sum of weights of edges in the tree is minimized/ maximized



Minimal/Maximal Spanning Trees: Interpretation

- Branches of MST's reflect clusters or "modules" in data
- One of the first proposed methods to analyze such structure
- But be careful: MSTs are very sensitive to noise & much information is discarded!

MST for correlation matrix of NYSE stocks

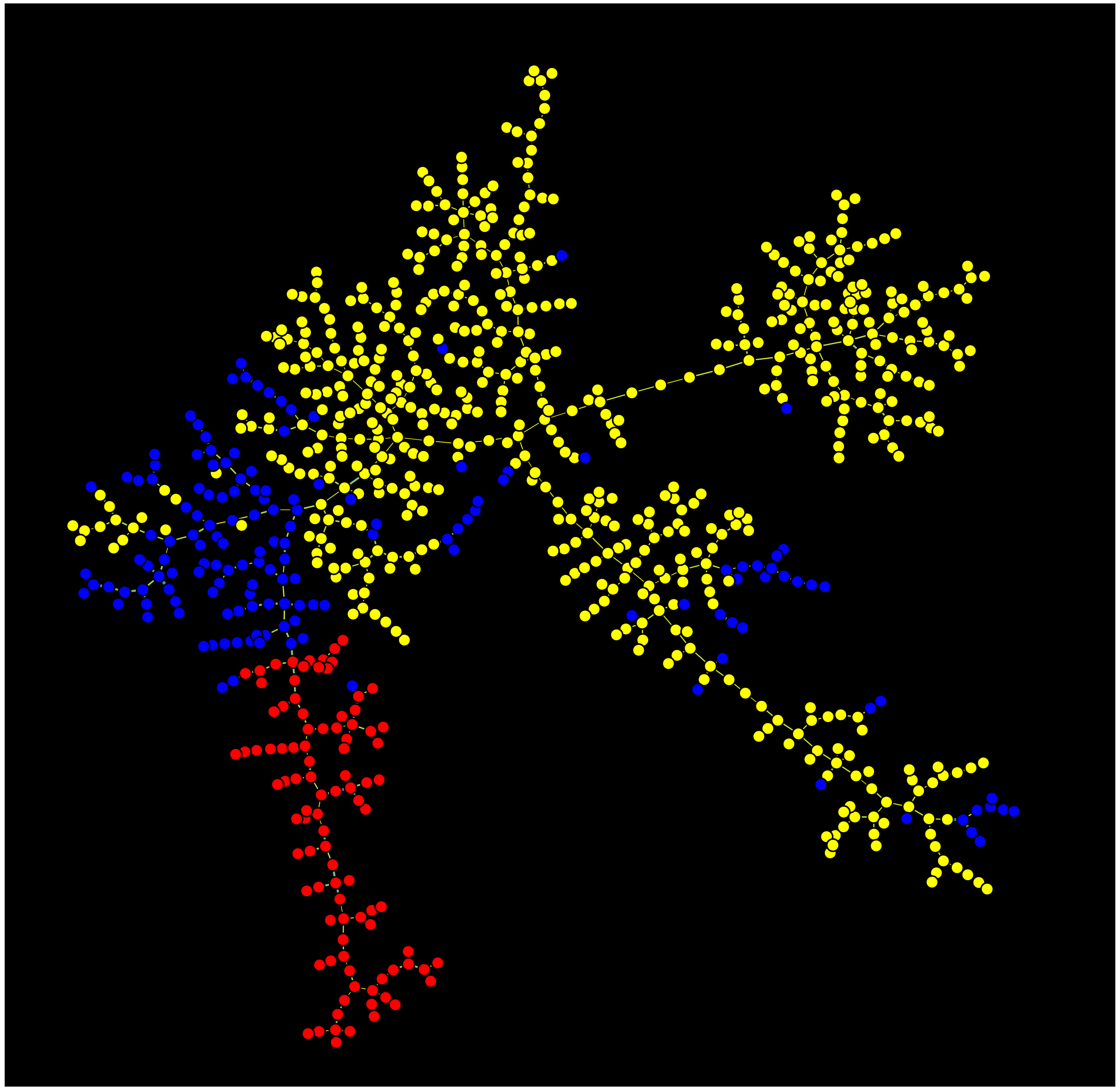


Onnela et al. Phys. Rev. E 68, 056110 (2003)

Maximal/minimal spanning trees

- In practice calculated using Kruskal's algorithm

1. Sort original network's edges by decreasing weight
2. Generate a network G with all original nodes but no links
3. Assign each node to its own component
4. Go through edges one by one, such that
   4.1. If endpoint nodes in different components, add edge to network G and merge components
   4.2. If endpoint nodes in same component, do nothing
5. Stop when N-1 links have been added



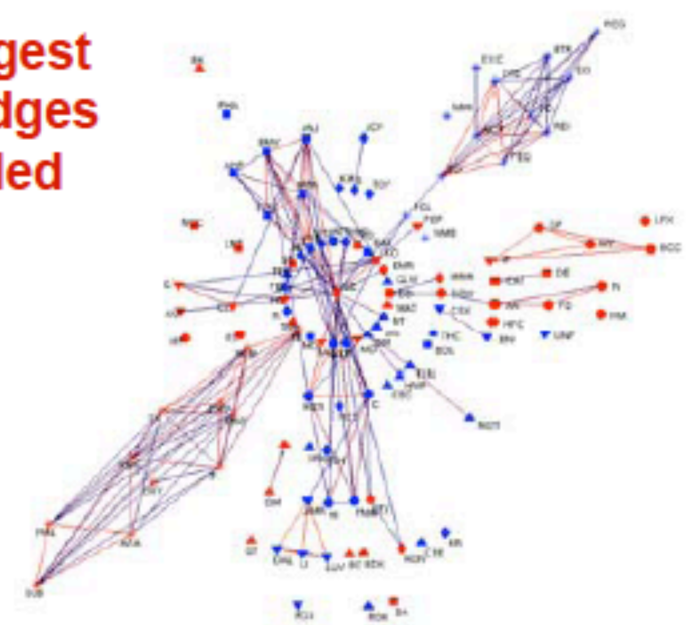
MST of genetic distance matrix of approx 800 specimens of Mediterranean marine plants, colors = geographic area

Matrices to Networks: Thresholding

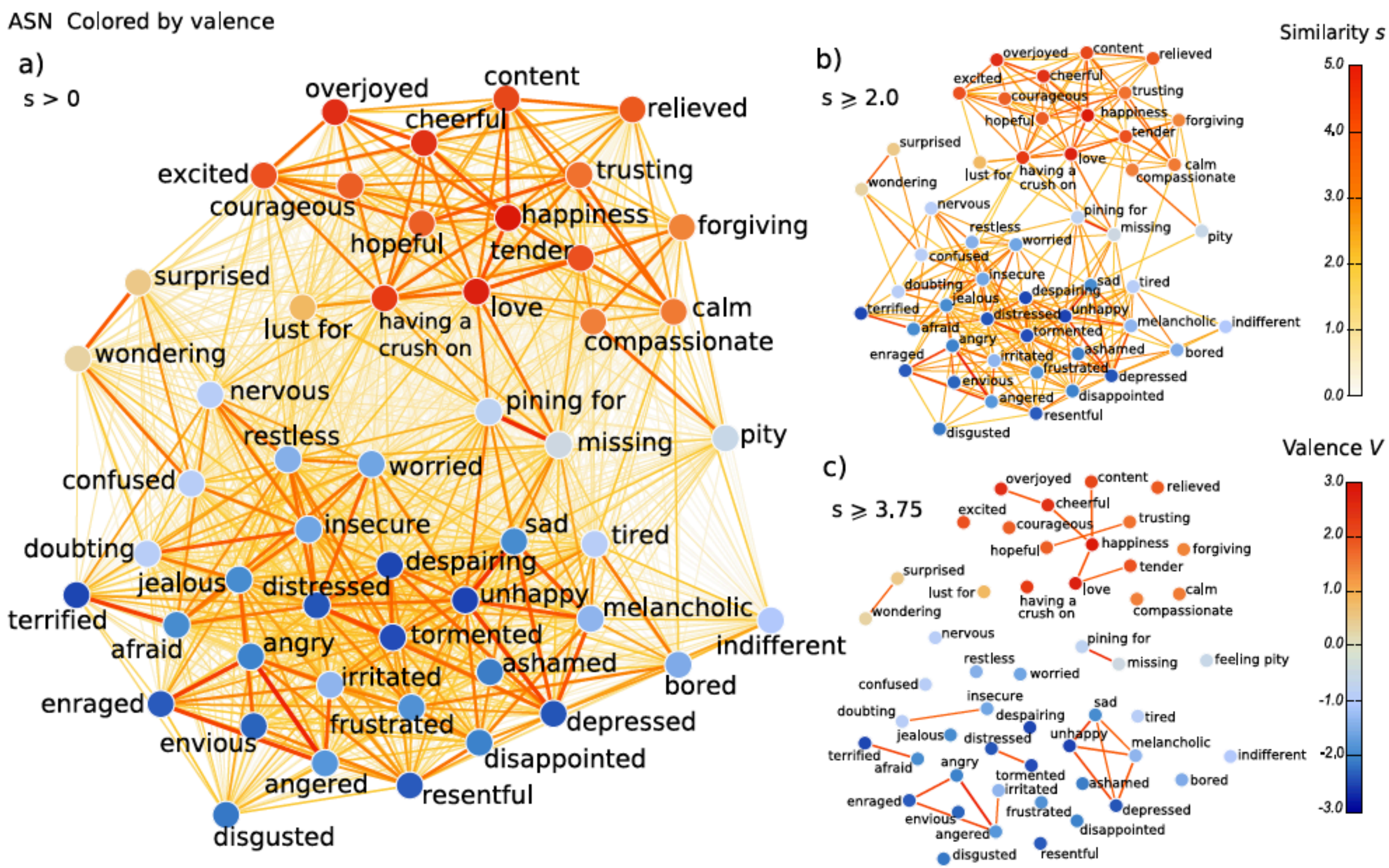
- One can also transform full matrices to networks by **thresholding** them
- Just remove all links (weight matrix elements) below the weight of your choosing
- Again, makes clusters/modules/ communities visible
- In any case, more information is retained than with MST's

Strongest 50 edges included

Strongest 180 edges included



Thresholding: interview-based word association network



Social Networks

Social Networks Studies - Motivation

The social sciences point of view

- Man is a “social animal” - to the point that even our intellectual capability has likely evolved for being able to succeed in the social system of a small “tribe”
- Nowadays, (almost) everyone is part of an enormous social network spanning the entire world
- Micro-level social interactions (e.g. friendships) give rise to larger macro-level structures (social networks)
- Social networks are the “lattices” where information is transmitted, culture is formed, etc...

The network science point of view

- Social networks are self-organizing structures - no-one is designing or controlling them
- Individuals “see” only their immediate network neighbourhood
- Nevertheless there is clear emergent higher-level structure
- Are there simple rules or mechanisms which give rise to this structure?
- How does the structure affect various processes?

Other social animals

Wolf, *Canis lupus*

- Packs of 6-8 wolves
- Packs consist of offspring of the alpha couple

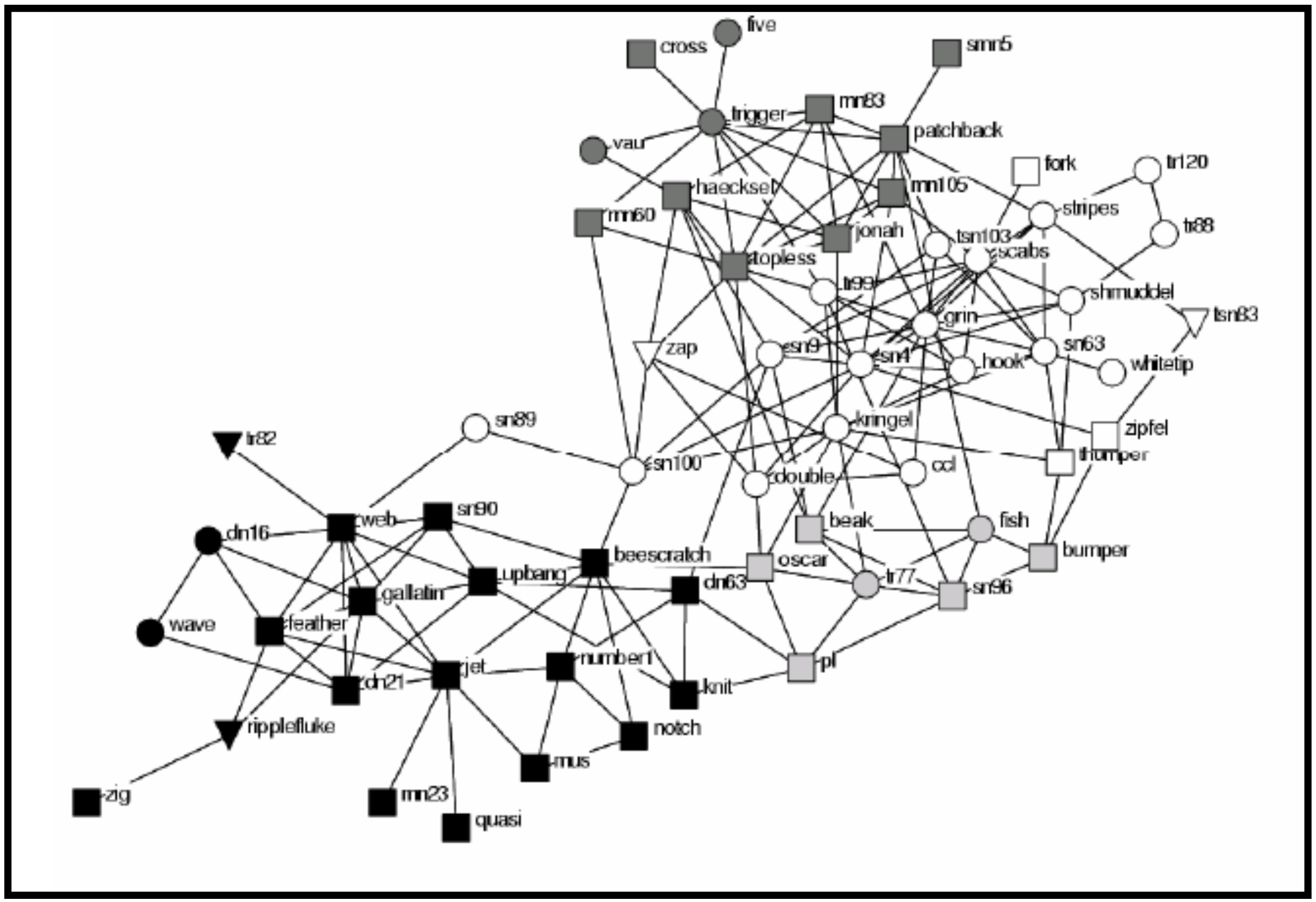
Bottle nose dolphin, *Tursiopsis truncatus*

- Herds of ~100 dolphins
- Herd consists of smaller social groups

Chimpanzee, *Pan troglodytes*

- Around 100 members in a community
- The community consists of temporary, fluid groups of ~10 chimpanzees
- Strong hierarchy!

A dolphin social network

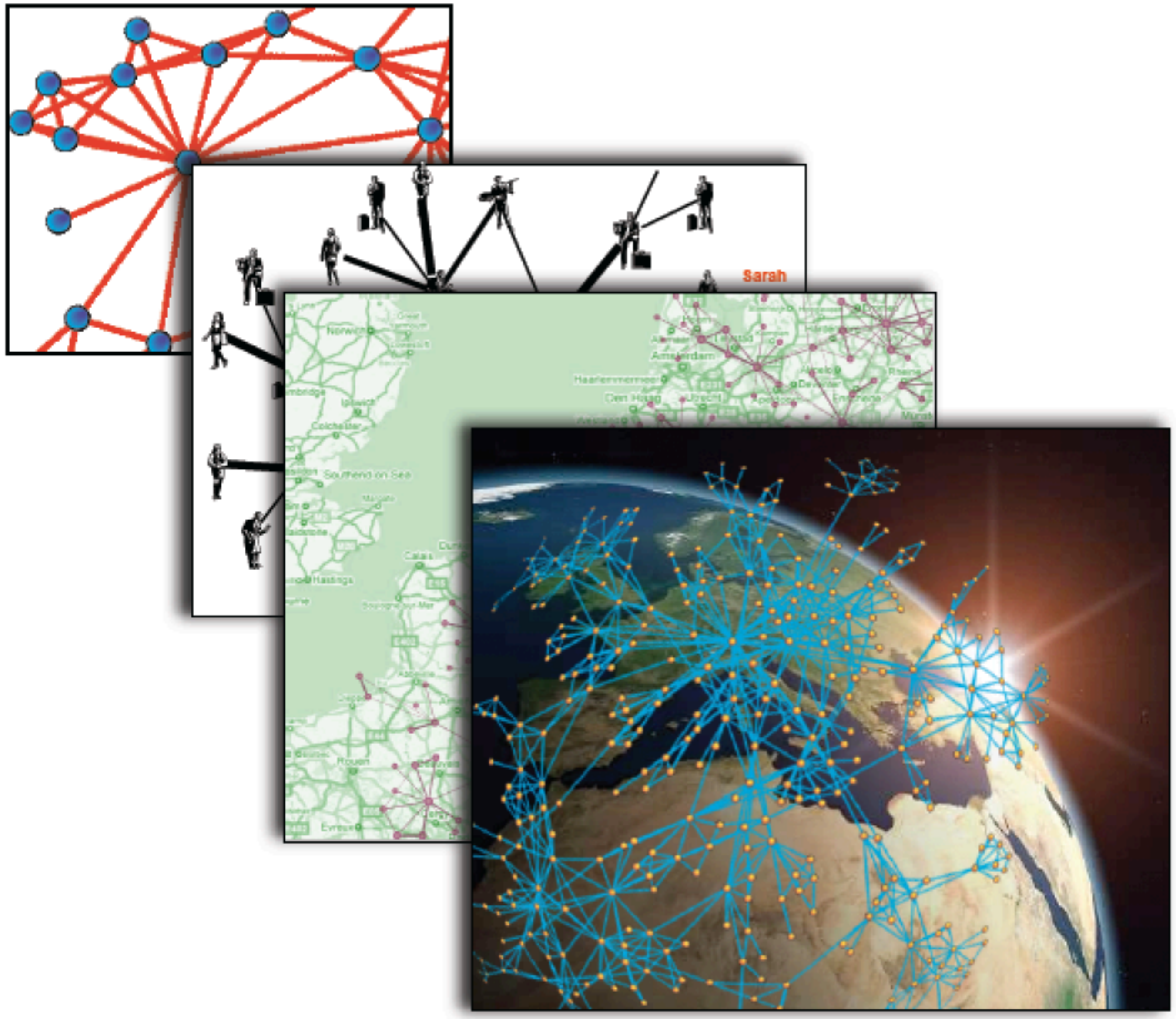


David Lusseau and M. E. J. Newman, *Proc. R. Soc. London B* **271**, S477-S481 (2004).

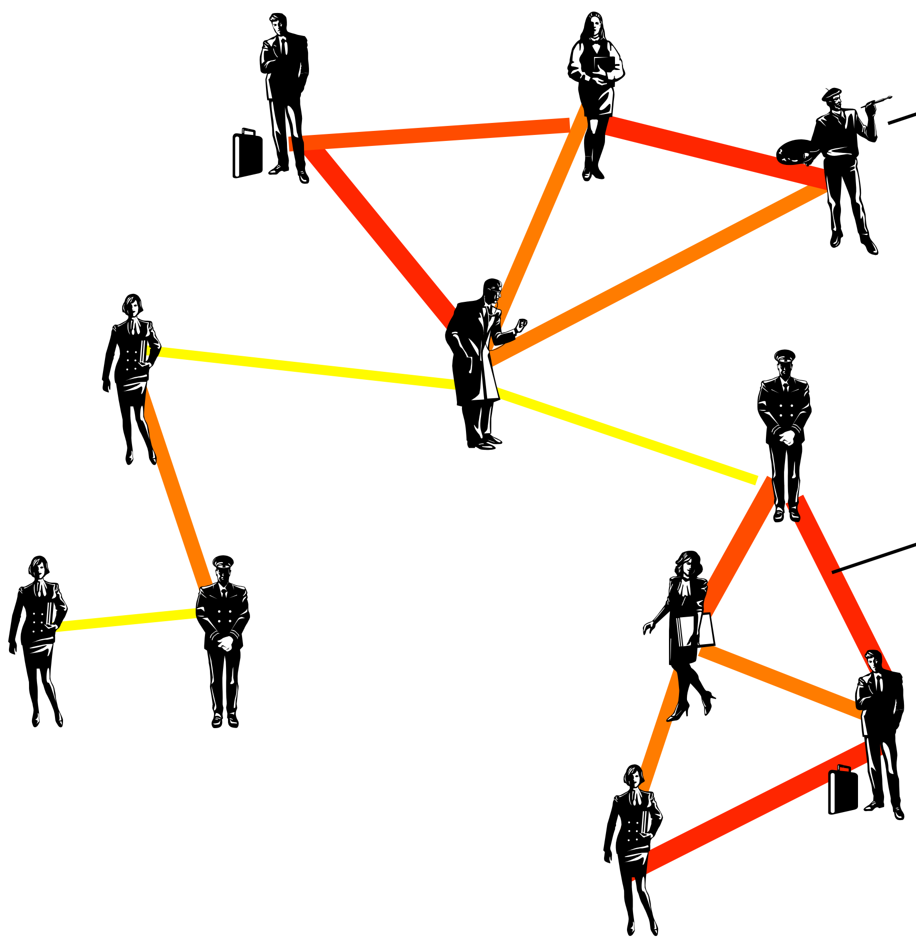
From friendships to societies

- Social networks consist of several (overlapping and fuzzy) levels

People
⇩
Friendships
⇩
Circles of friendship
⇩
Social groups
⇩
Communities
⇩
Society



Social ties as networks



Nodes = individuals

Edges = social ties/ interactions

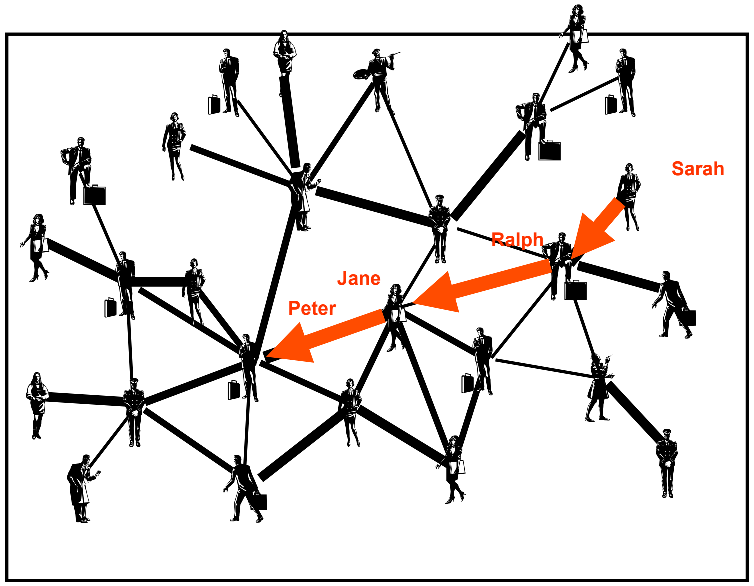
- How to define? "**A *knows* B?**"
- Or based on real interactions (**A has called/emailed/met B)**

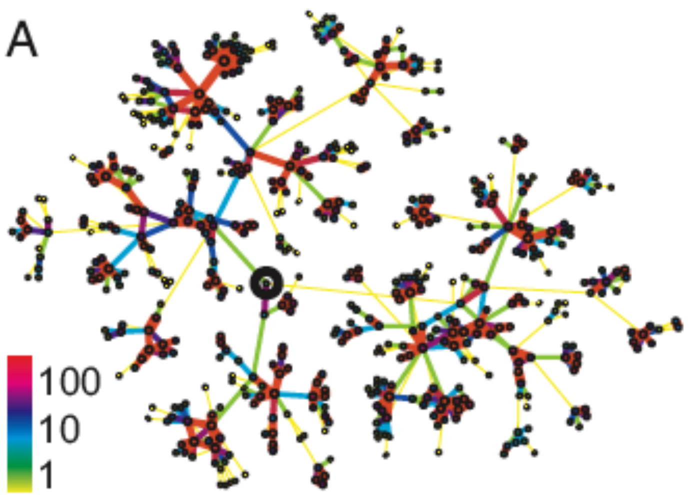
•Strengths of ties can be accounted for with edge weights

•However, there may also be different kinds of ties (family, friendship, professional, etc)

Social networks: known properties

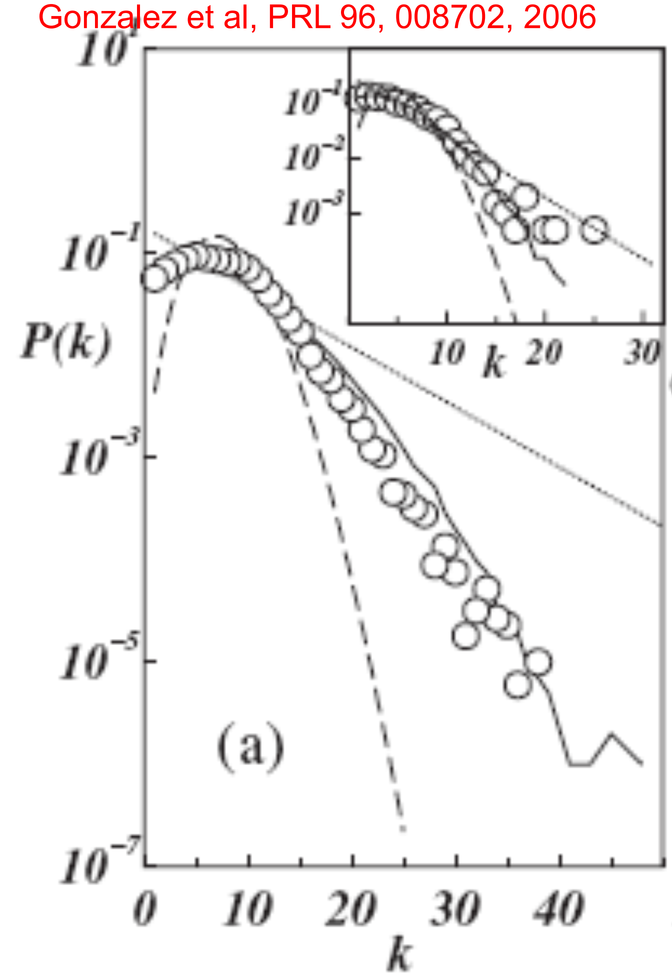
- Short path lengths ("6 degrees", "small world")
- High clustering
- Assortativity: highly connected people friends with similar people
- Contain groups/cliques/communities/clusters





Degree distribution: no universal form

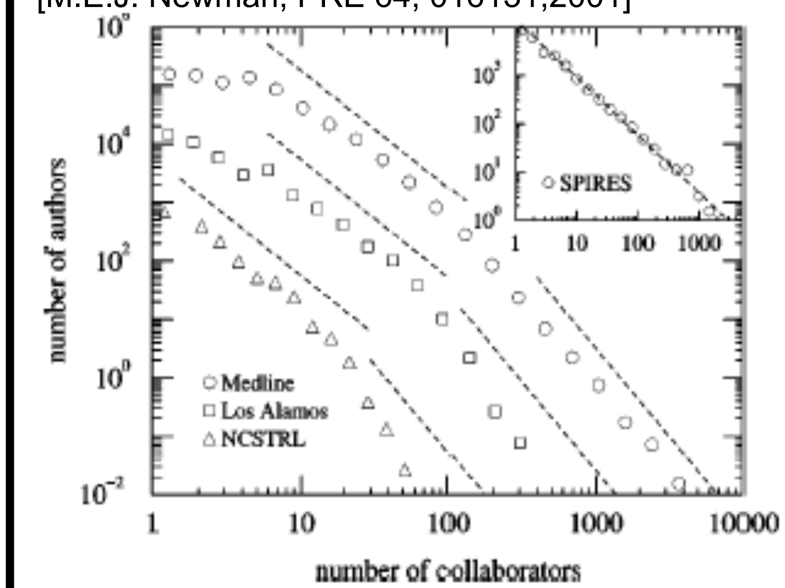
Gonzalez et al, PRL 96, 008702, 2006



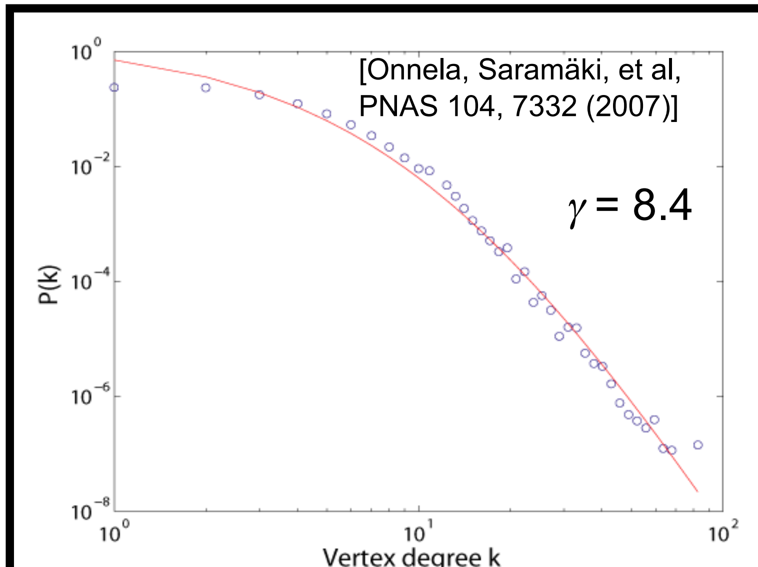
U.S. schoolchildren:

p(k) ~ exp(-k)

[M.E.J. Newman, PRE 64, 016131,2001]



Collaboration networks
- Broad, might be a power law (or two power laws)



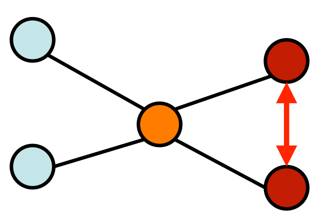
Mobile telephone call network
- tail looks like a power law but with a very large exponent

- Distributions generally broad
- Shape of distribution depends on network
- If power law, exponent has to be large (no-one can have 10000 friends)

Simple social network models

Davidsen et al, Phys Rev Lett 88, 128701 (1999)

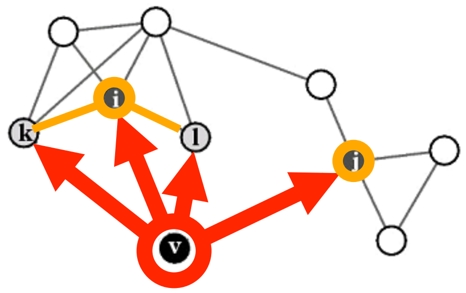
- First create a network of N nodes without any edges
- Repeat the following:
  - Pick a random node. If it has less than 2 neighbours, connect it to a random node. If it has 2 or more neighbours, randomly pick two of these and connect them.
  - With probability p remove the node and create a new one with a single random link.



- Correct: **clustered network, short pathlength, broad degree distribution**
- Incorrect: **no assortativity, no communities/groups**

Toivonen et al, Physica A 371, 88 (2005)

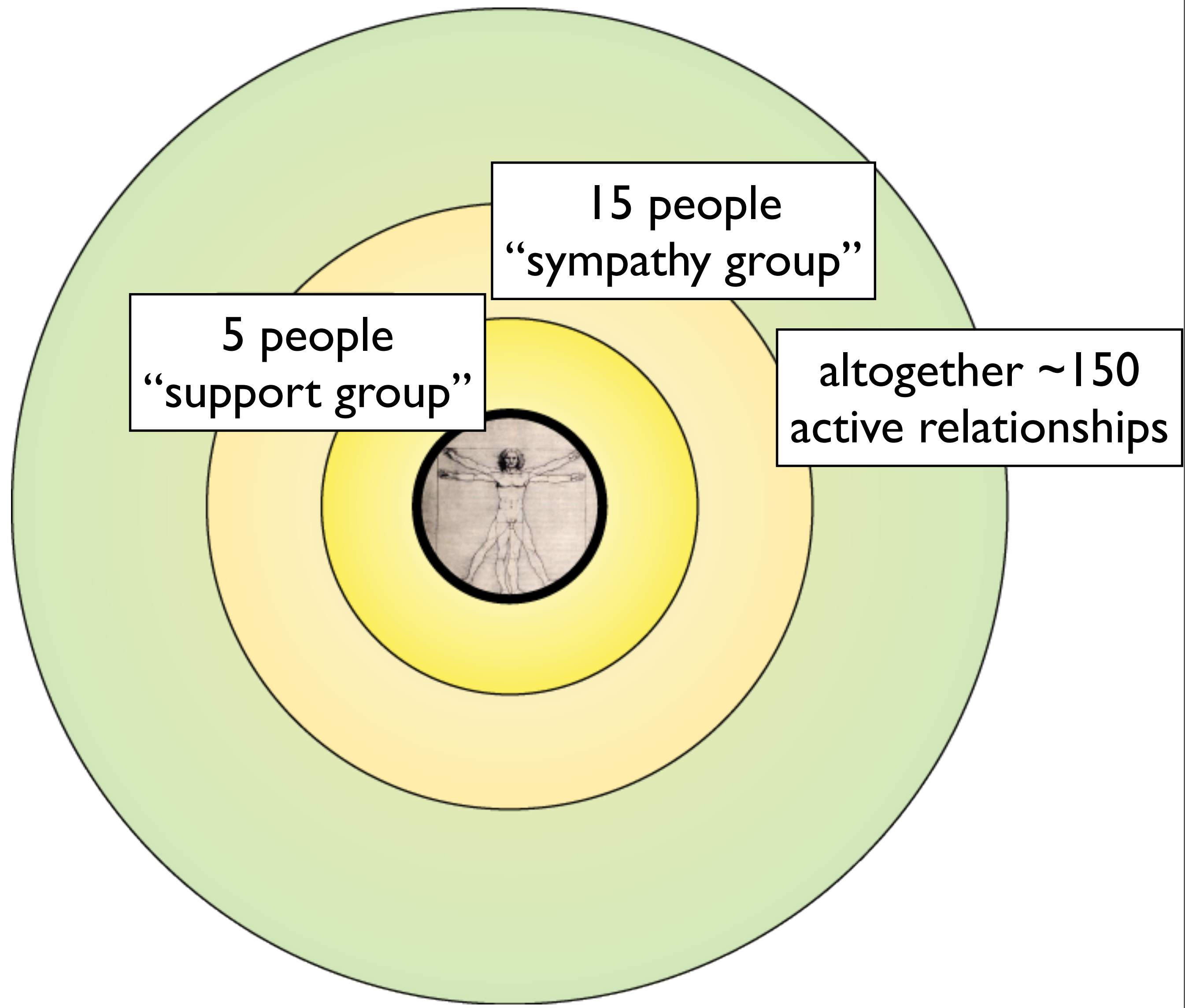
- Growth model
- Create a small initial seed, then repeat the following:
  - Randomly pick on the average m nodes as the “initial contacts”.
  - Pick on the average n of their neighbours.
  - Connect the new node to initial contacts and the chosen neighbours.



- Correct: **clustering, path lengths, degree distribution, assortativity, groups exist**
- Incorrect: **Group structure does not correspond to reality, hubs sit between groups**

The Dunbar Number

- "Egocentric" social networks, i.e. personal networks, are layered
- Robin Dunbar's theory:
  - Core group of ~5 people ("support group")
  - "Sympathy group" of ~15 people
  - Max ~150 active relationships - the "Dunbar Number"
- Evolutionary explanation: our cognitive capabilities do not allow for more
- Can technology increase this number? E.g. Orkut, Facebook, ?

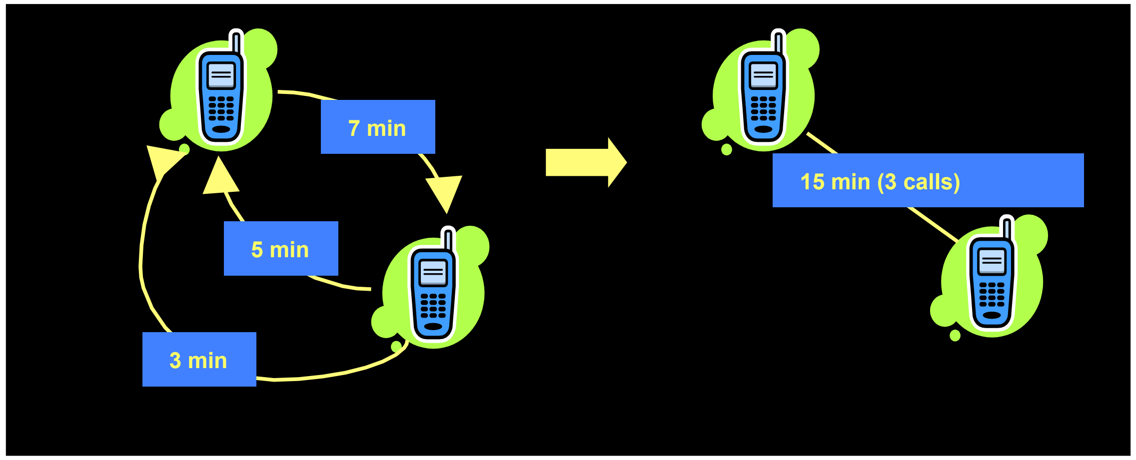


Social network analysis: mobile telephone call records

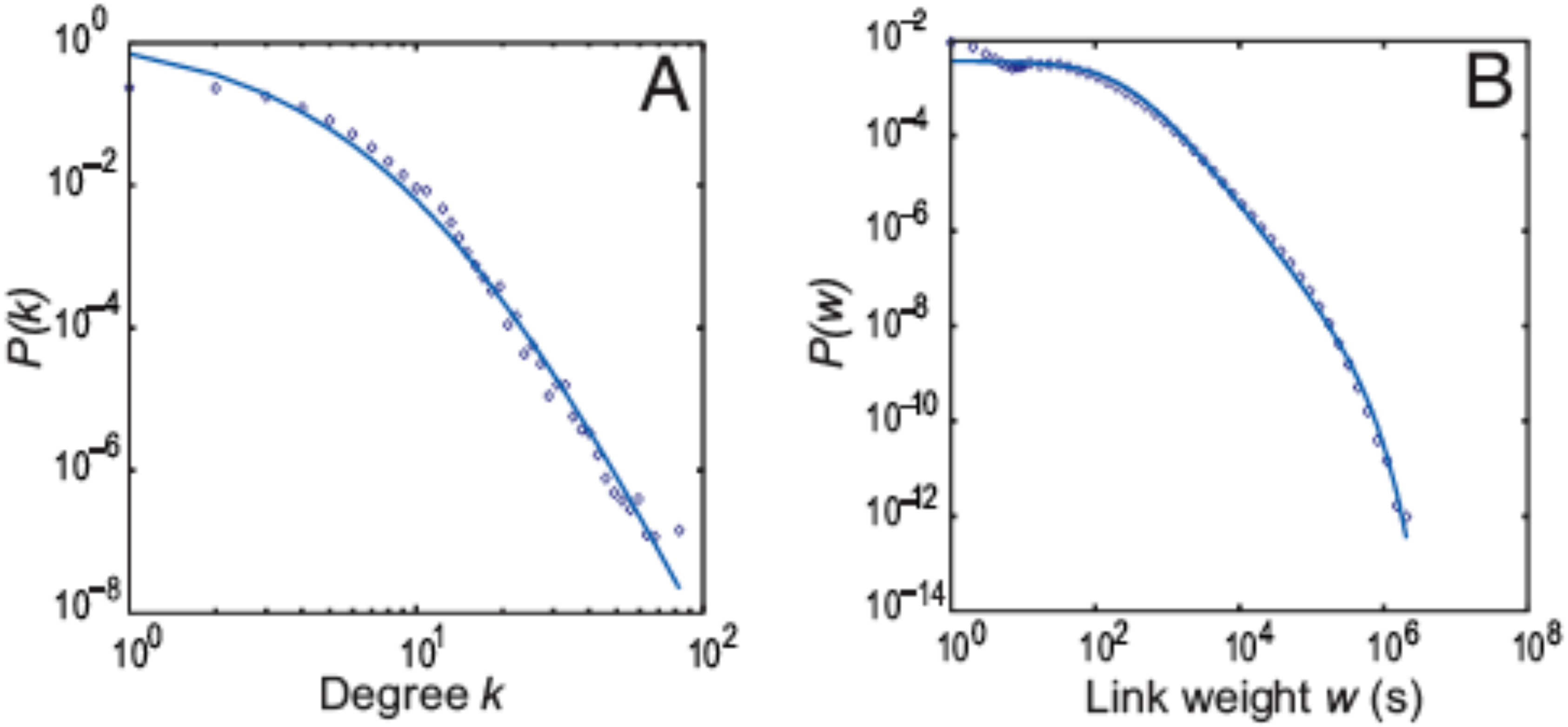
- Research published in
  - J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Proc. Natl. Acad. Sci. USA* **104**, 7332 (2007)
  - J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. Argollo de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, New Journal of Physics **9**, 179 (2007)
- Target: to understand the structure, weight-topology-correlations and their consequences in a very large social network

Social network analysis: mobile telephone call records

- Data: call records (caller, recipient) for 18 weeks for 7 million people within one operator's customer base
- Reciprocity filtering: we require that A has called B AND B has called A at least once
- After this, ~4 million people left in the network
- Edge weights: total call minutes between two persons within 18 weeks



Basic statistics



degree distribution: steep tail which looks like a power law with an exponent -8.4

weight distribution also broad

The Weak Ties Hypothesis

- M. Granovetter, Am. J. Sociol. 78, 1360-1380, 1973.
- "The strength of a tie is a (probably linear) combination of the amount of time,the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie."
- The weak ties hypothesis:
  The relative overlap of two individual's friendship networks varies directly with the strength of their tie to one another.
- The cohesive power of weak ties: important in e.g. obtaining new information

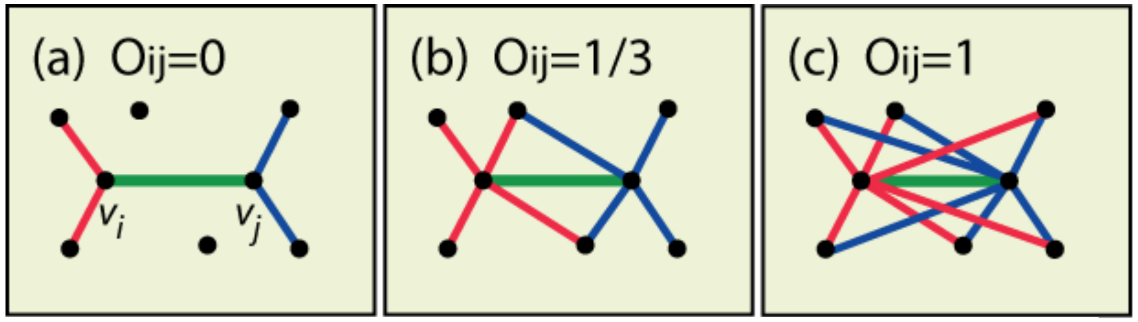
Weak ties in real data

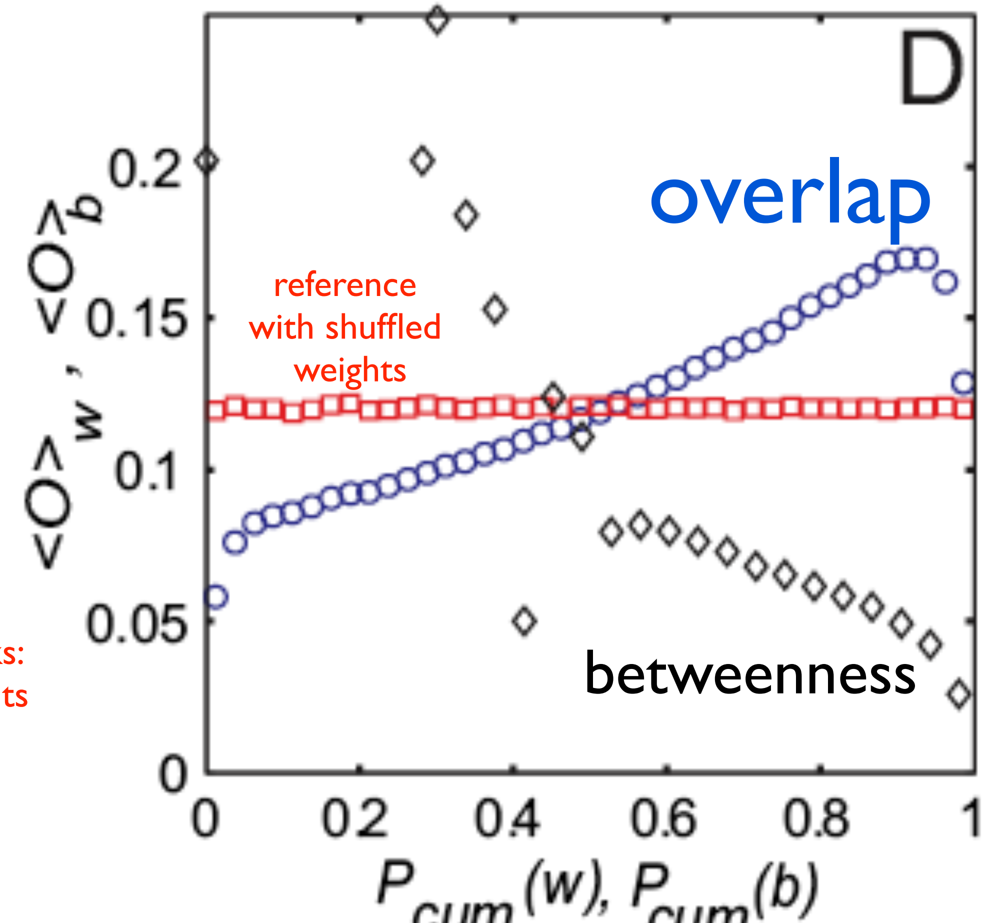
- Weak ties hypothesis: Relative overlap of friends varies with the tie strength

- Define **overlap** O_{ij} of edge (i,j) as the fraction of common neighbours

- Observation in the mobile telephone call network: Average overlap increases as a function of (cumulative) link weights

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





(null hypotheses / reference systems for weighted networks: keep topology, shuffle all weights s.t. $w_{ij}=w_{lm}$, $w_{lm}=w_{ij}$ etc)

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Proc. Natl. Acad. Sci. USA* **104**, 7332 (2007)

Role of weak ties on the network level

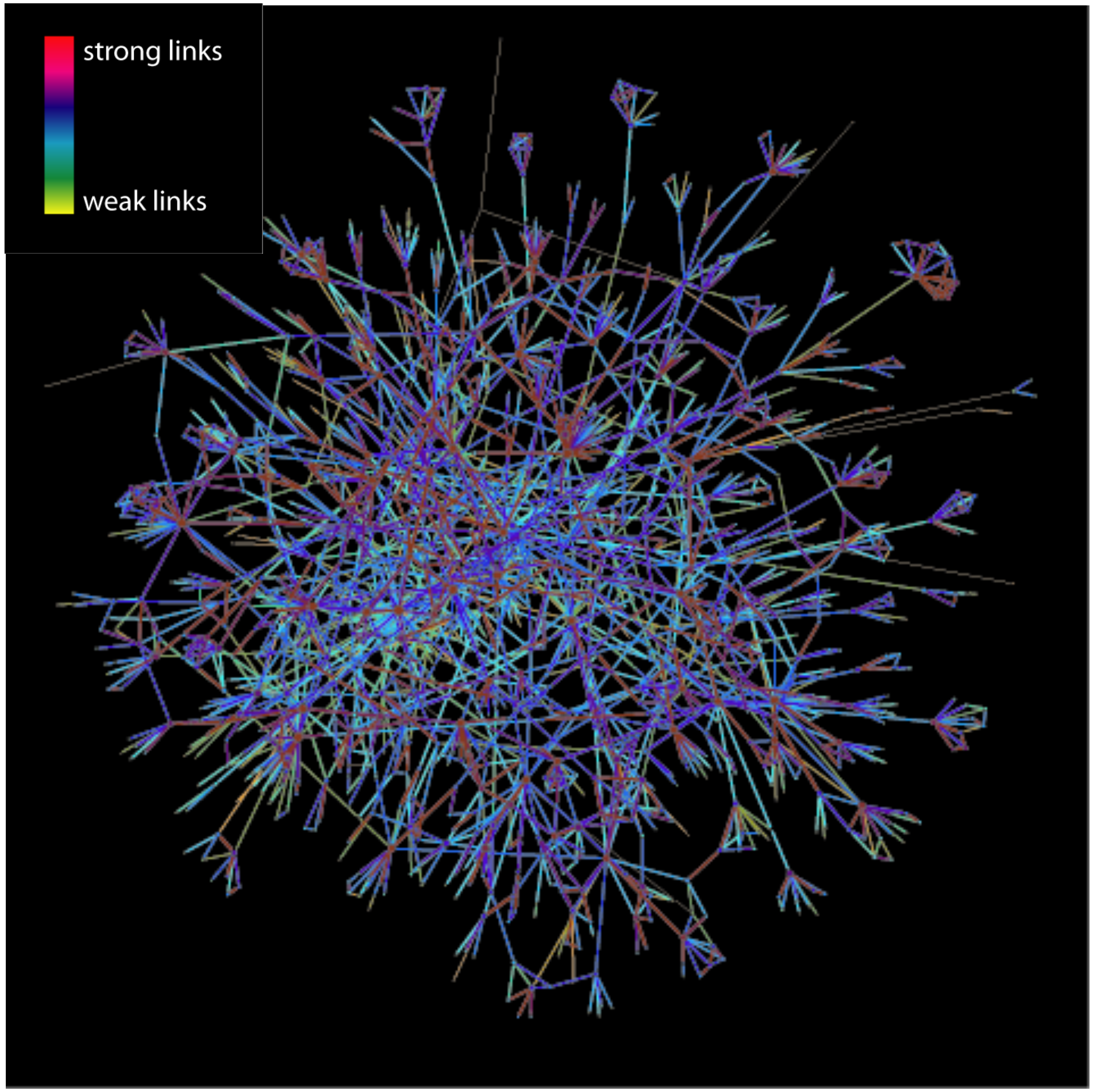
- Probe the global role of links of different weight and local topology

- Thresholding & percolation analysis: Remove links based on their weight (weak to strong, or strong to weak)

- Control parameter f is the fraction of removed links
    - Initial network (f=0); isolated nodes (f=1)

Role of weak ties on the network level

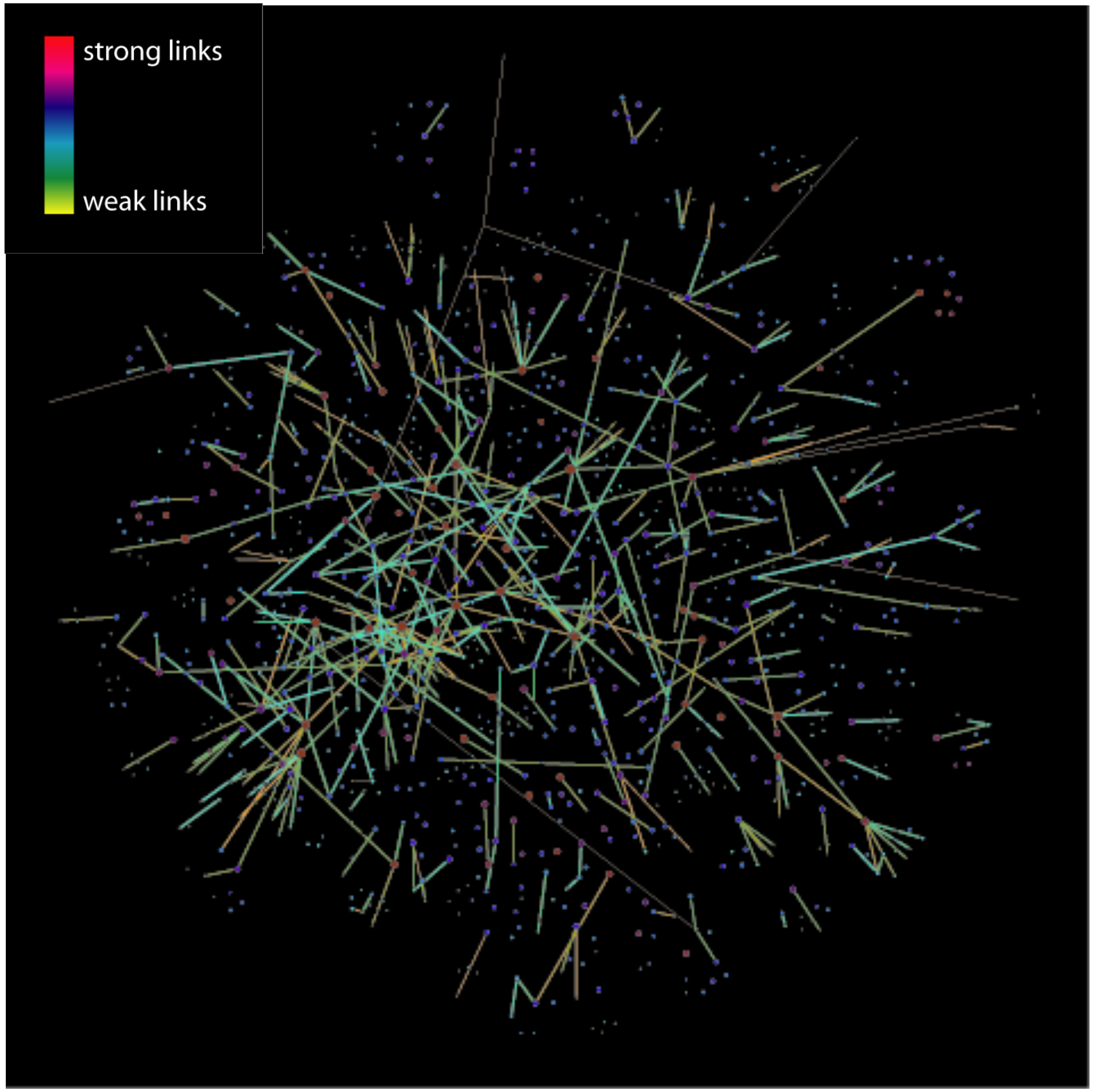
Initial connected network (f=0), small sample
⇒ All links are intact, i.e. the network is in its initial stage



Role of weak ties on the network level

Decreasing weight thresholded network (f=0.8)

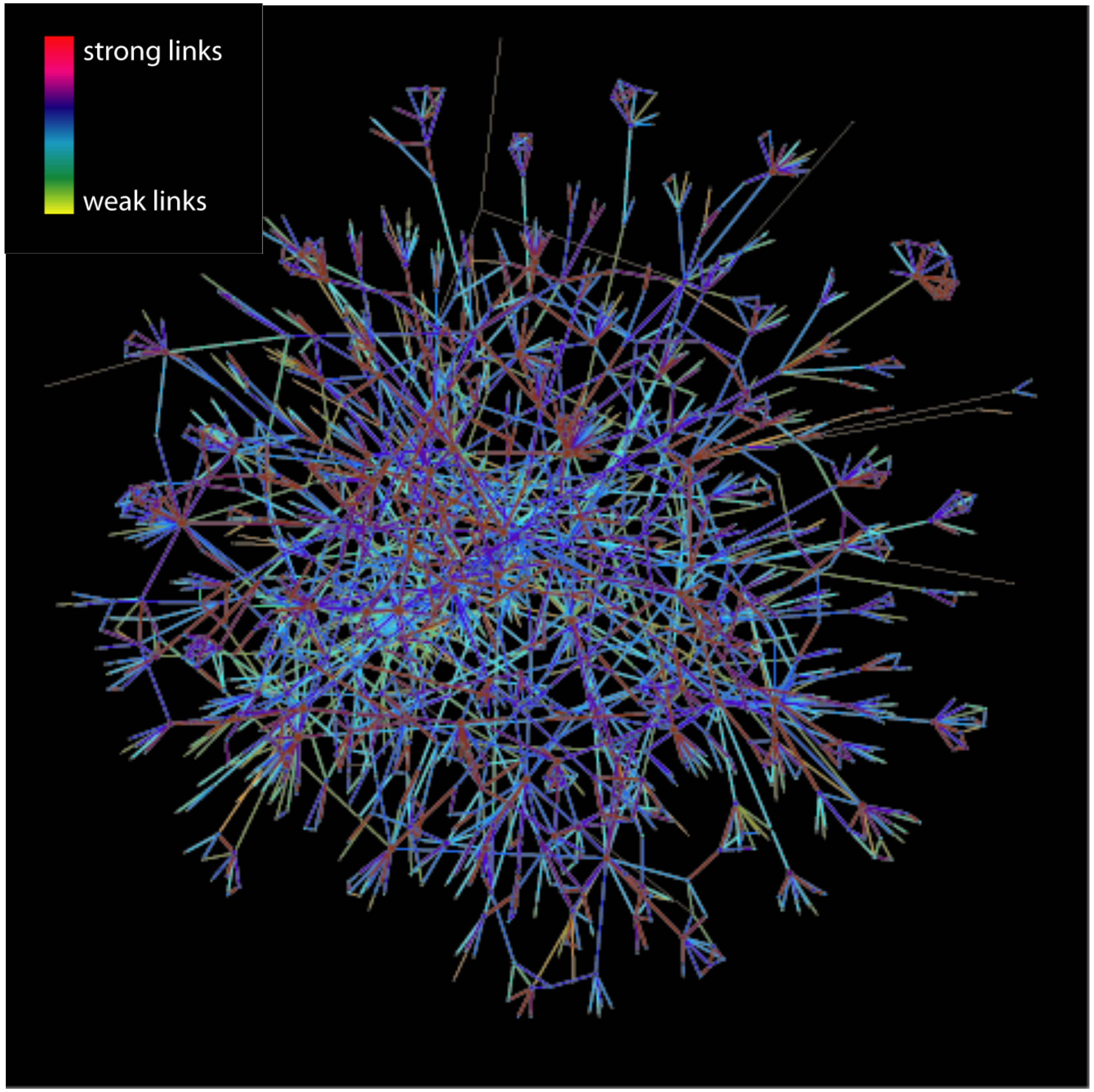
⇒ 80% of the strongest links removed, weakest 20% remain



Role of weak ties on the network level

Initial connected network (f=0), small sample
⇒ All links are intact, i.e. the network is in its initial

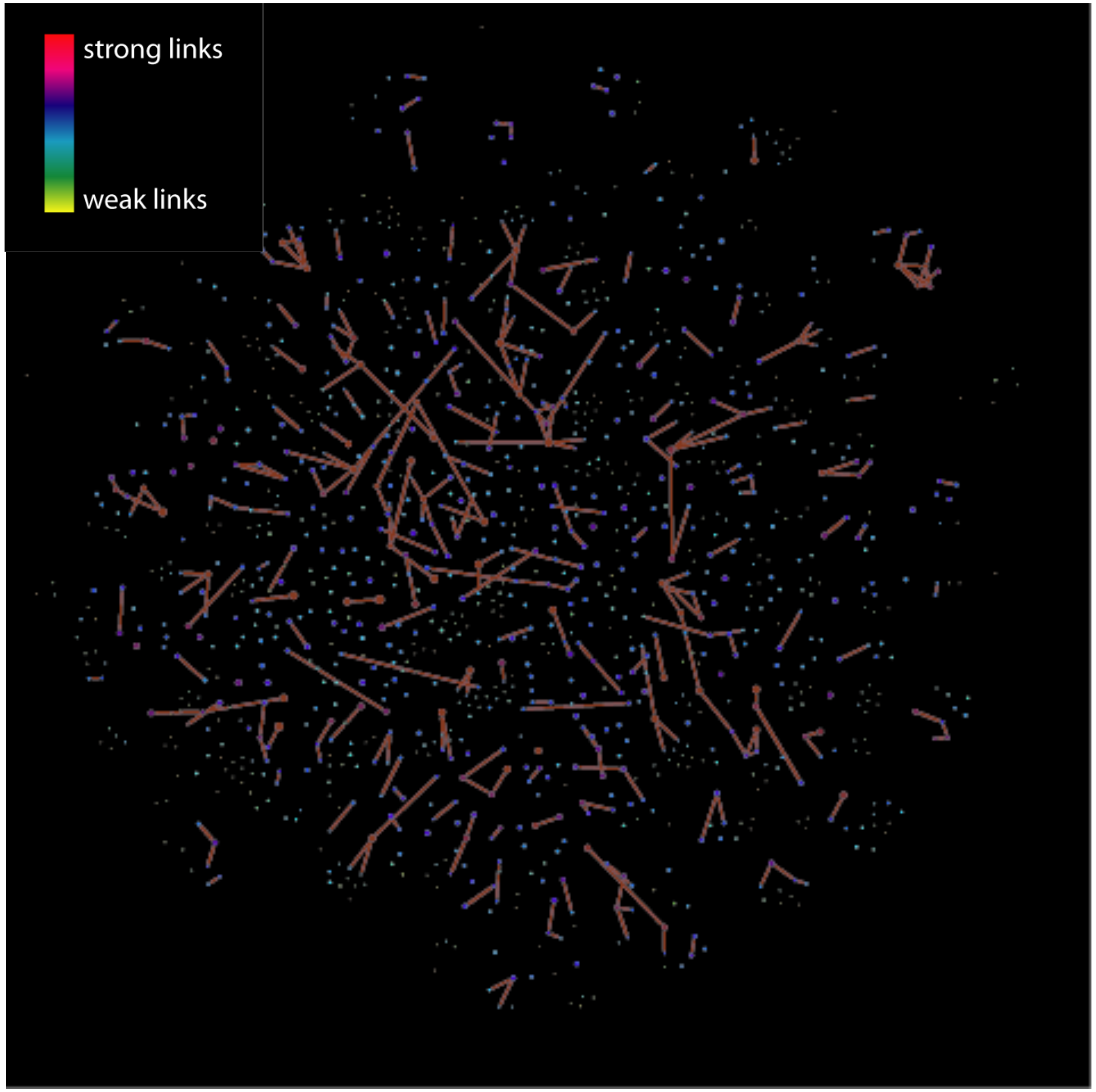
stage



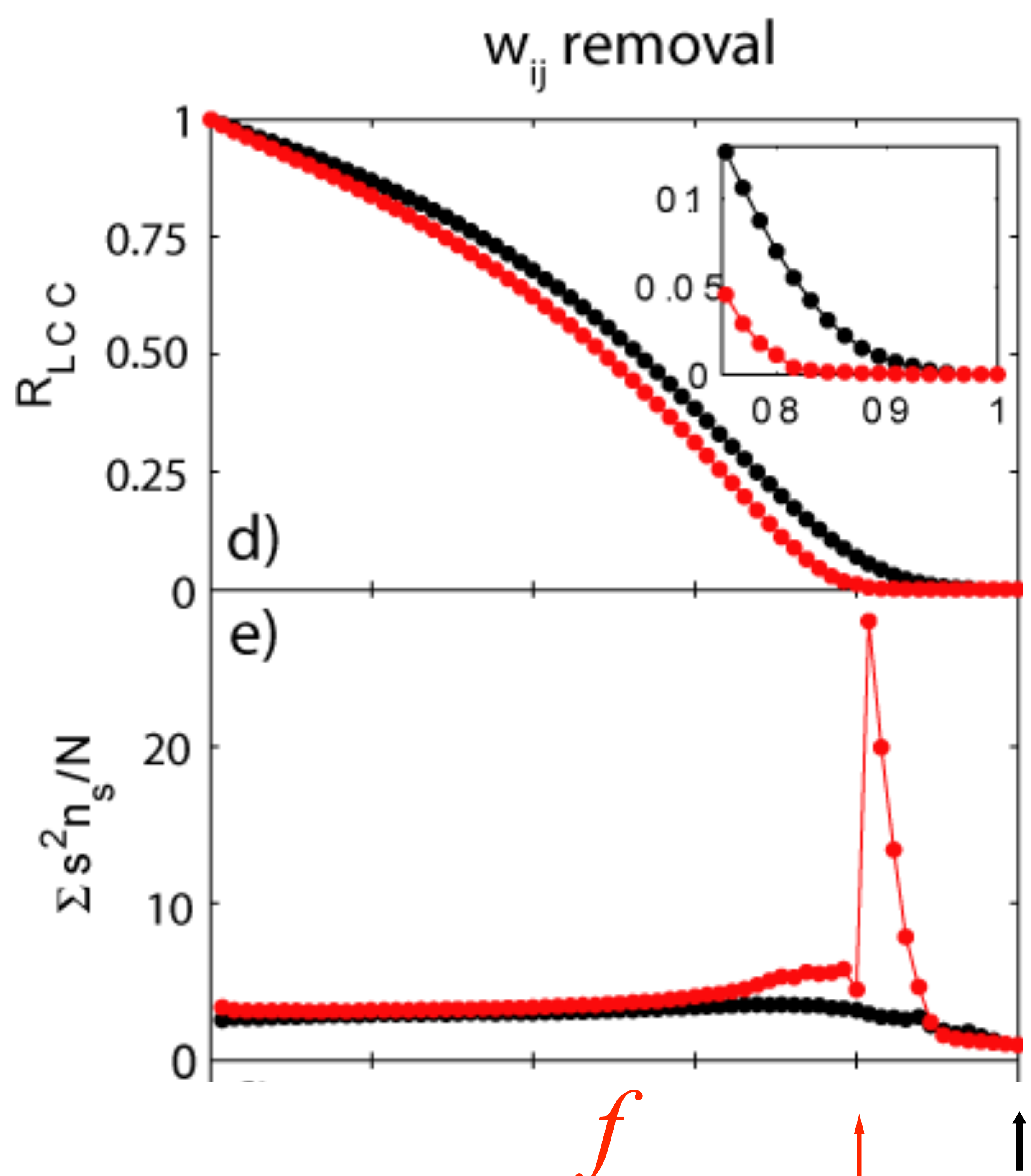
Role of weak ties on the network level

Increasing weight thresholded network (f=0.8)

⇒ 80% of the weakest links removed, strongest 20% remain



Role of weak ties on the network level



$f_c^w(\infty) = 0.80 \pm 0.04$

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Proc. Natl. Acad. Sci. USA* **104**, 7332 (2007)

Weak links first:

- Network fragments at around f=0.8

Strong links first

- No evidence of fragmentation

R_{LCC} = fraction of nodes in largest connected component

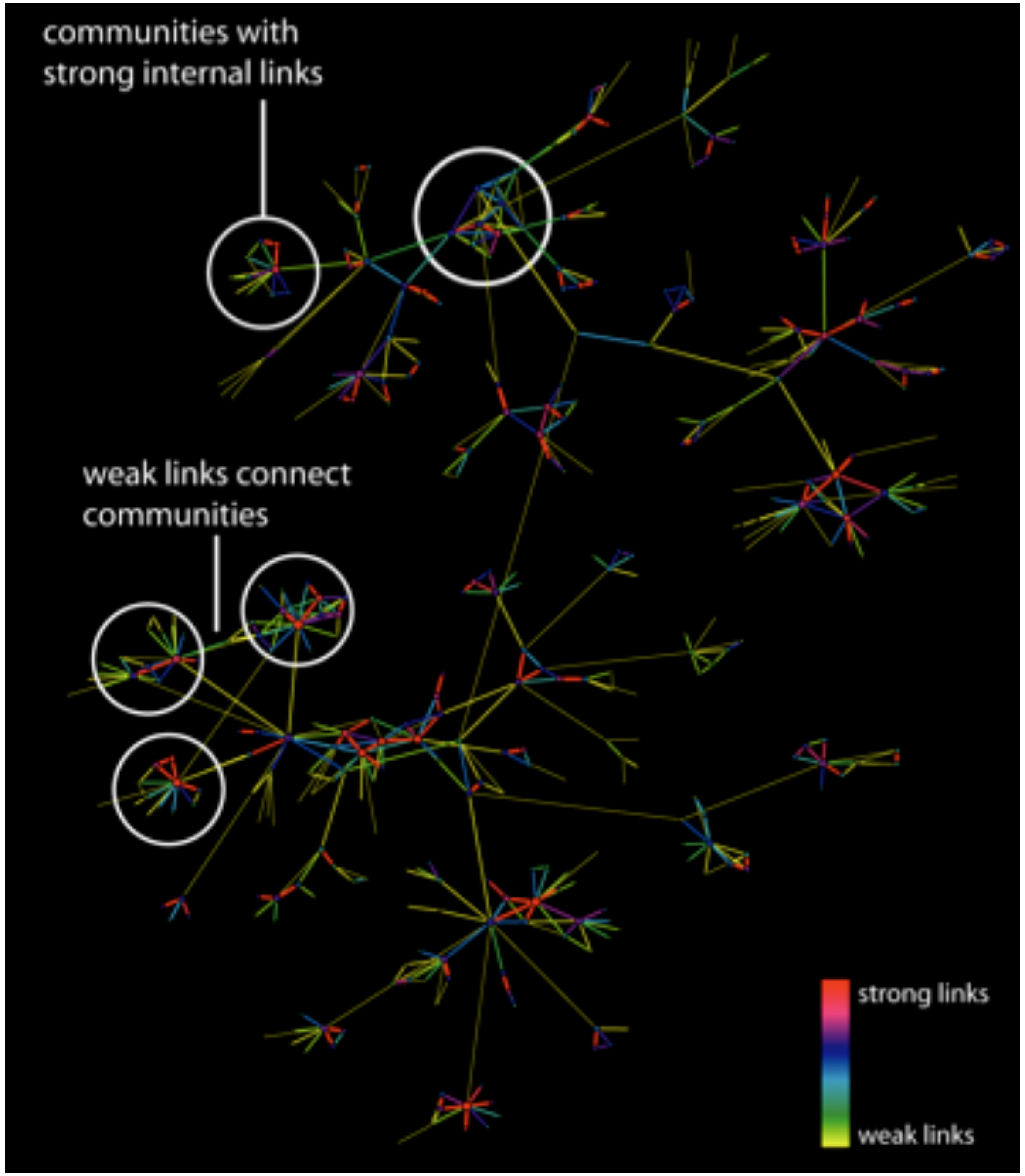
S = susceptibility

- Def: average cluster size (excl. LCC)

$$S = \sum_{s<s_{\max}} n_s s^2 \Big/ \sum_{s<s_{\max}} n_s s$$

Mobile network: summary of observations

- Strong links associated with dense network neighbourhoods (triangles, cliques, etc)
- Weak links connect dense neighbourhoods
- I.e. social groups with strong ties are connected via bridges of weak links
- Weak links crucially important for connectivity of the whole network!



A weighted model based on observations

Kumpula, Onnela, Saramäki et al, Phys Rev Lett **99**, 228701 (2007)

Tie formation mechanisms known in social sciences

Cyclic closure:

- Getting to know people through own friends, their friends, etc
- Decreases exponentially with network distance*, hence one can only consider triangles (becoming friend of a friend's friend)

Focal closure:

- Connections which appear random regarding the network

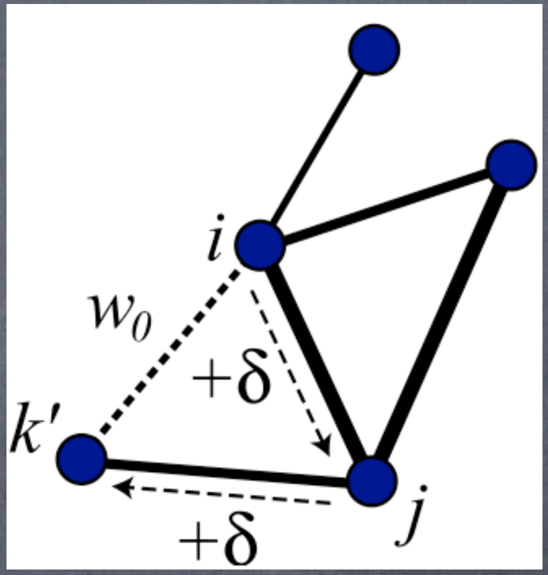
Model

- Use these mechanisms, add tie reinforcement mechanism
- Network of fixed size N, initially random connections

* M. Kossinets et al., "Empirical Analysis of an Evolving Social Network", Science **311**, 88 (2006)

Microscopic rules

- Fixed number of nodes, 3 mechanisms for link creation & deletion
- Rule 1/3: Local attachment + weight reinforcement
  - Pick a random node i
  - Pick another (k) by weighted 2-step random walk



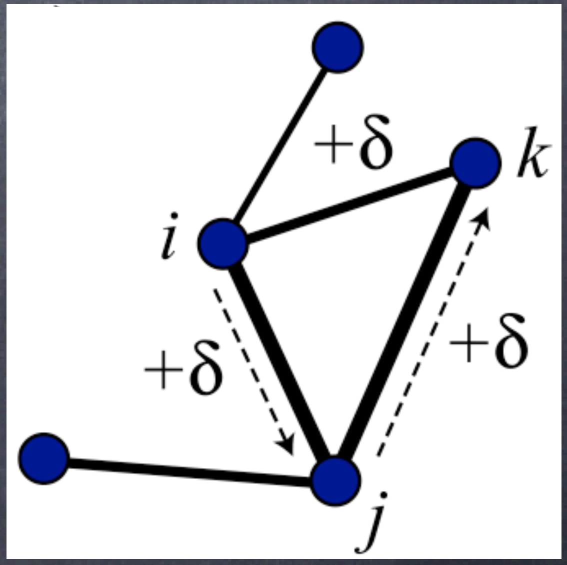
$$P(i \to j) = w_{ij}/s_i$$
$$P(j \to k) = w_{jk}/(s_j - w_{ij})$$
$$w_{ij} \to w_{ij} + \delta$$
$$w_{jk} \to w_{jk} + \delta$$

  - If no triangle (i,j,k) => **Form triangle**

$$P(i,j,k) = p_\Delta$$
$$w_{ik} = w_0 = 1$$

  - If triangle (i,j,k) exists => **Reinforce triangle**

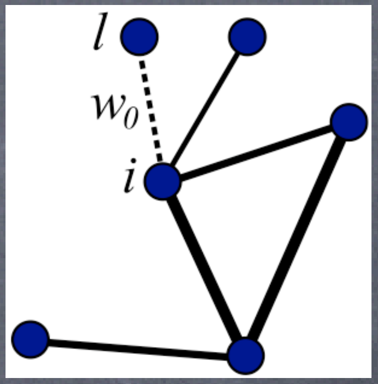
$$w_{ik} \to w_{ik} + \delta$$



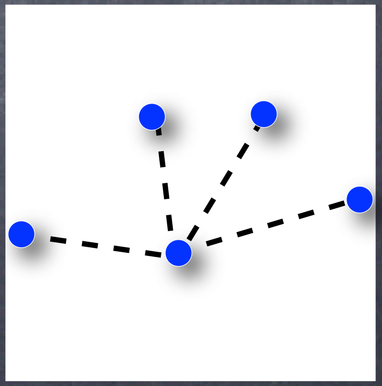
Microscopic rules

- Rule 2/3: Global (random) attachment
  - Pick a random node, connect to a random node with probability p_r (or if its degree=0)

$$k_i = 0 \implies P(i,j) = 1; w_{ij} = w_o = 1$$
$$k_i > 0 \implies P(i,j) = p_r; w_{ij} = w_o$$



- Rule 3/3: Node deletion
  - Pick a random node; delete it with probability p_d
  - Adjacent links are removed
  - Node is returned to the network



$$k_i > 0 \implies P(k_i = 0) = p_d$$

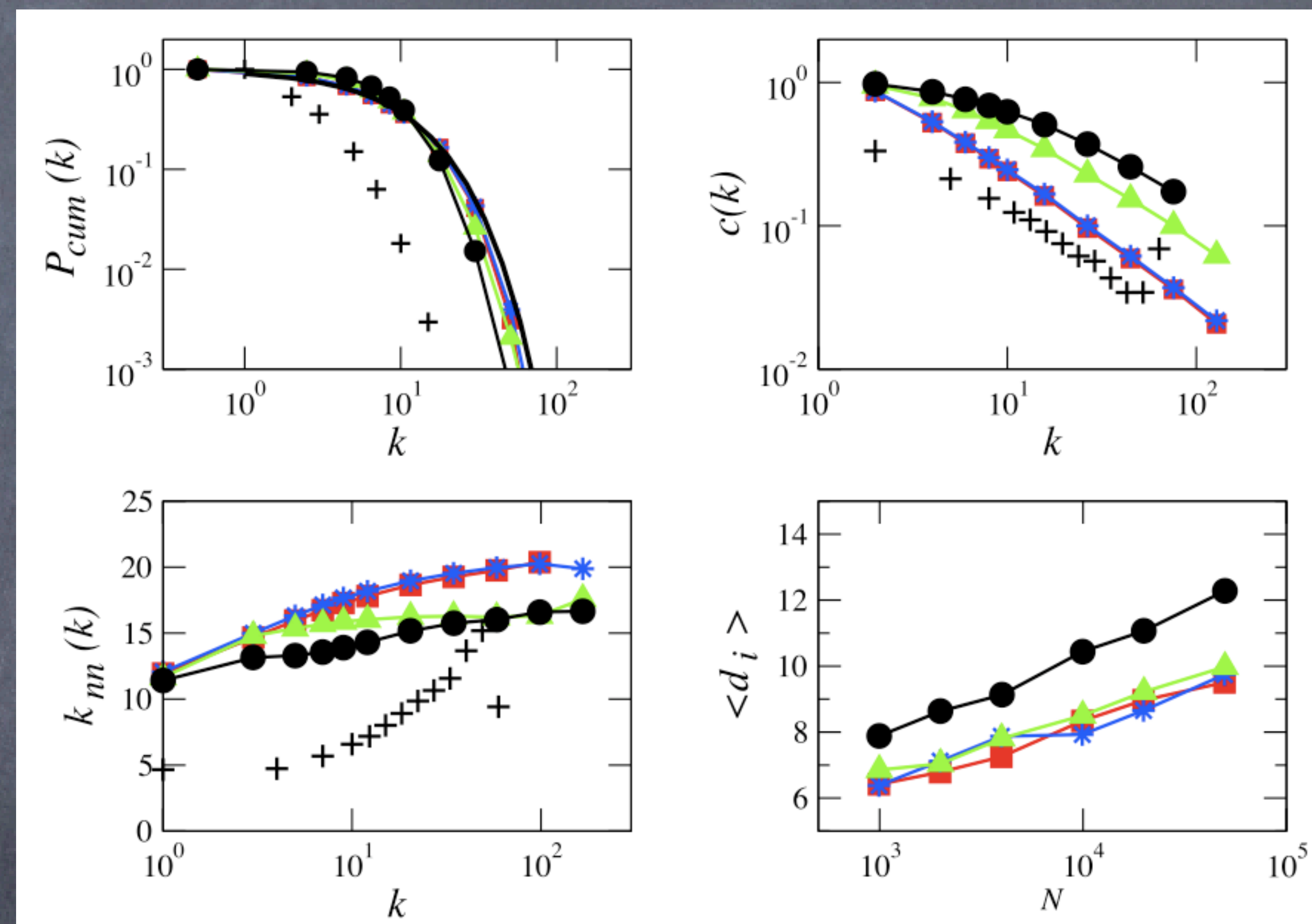
Basic characteristics

(a) Broad degree distribution

(b) High clustering

(c) Assortative

(d) Small world



$\delta = 1$

$\delta = 0.5$

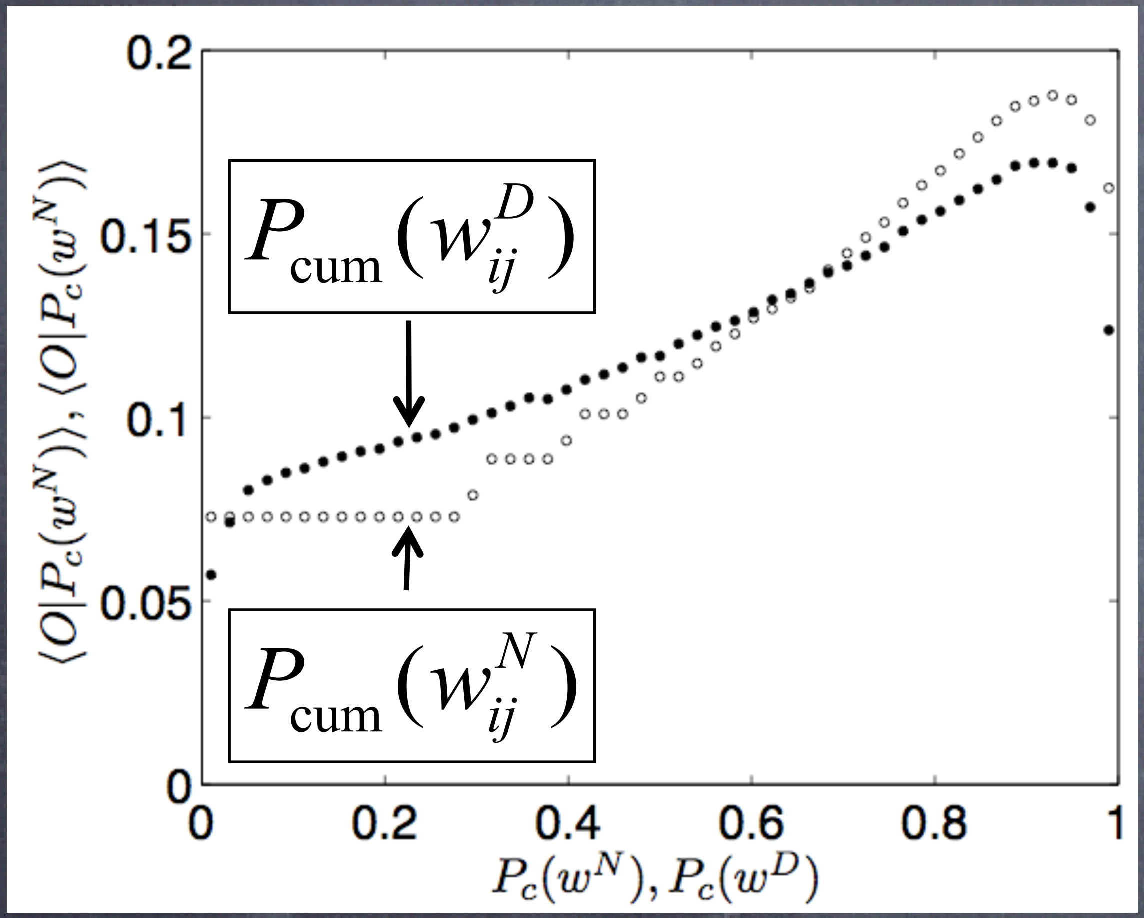
$\delta = 10^{-3}$

$\delta = 0$

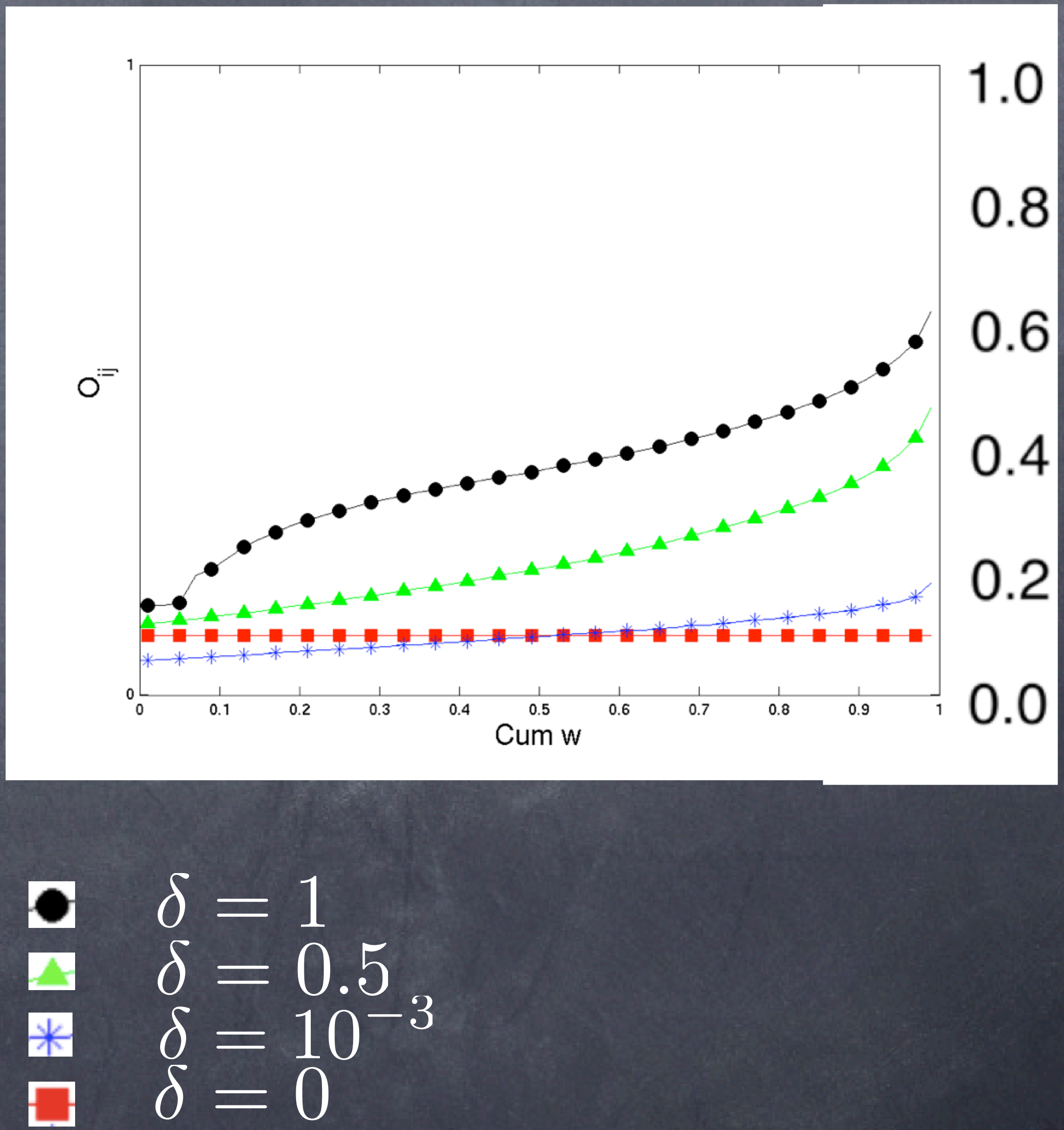
Values of δ are 0 ($\square$), 1×10^{-3} ($*$), 1×10^{-2} ($\triangleright$), 0.1 ($\triangle$), 0.5 ($\triangledown$), and 1 ($\circ$).

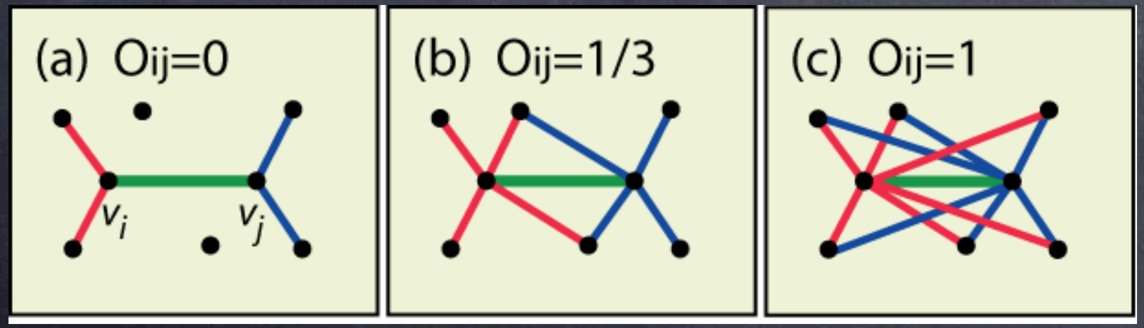
Local structure (overlap)

Empirical



Model

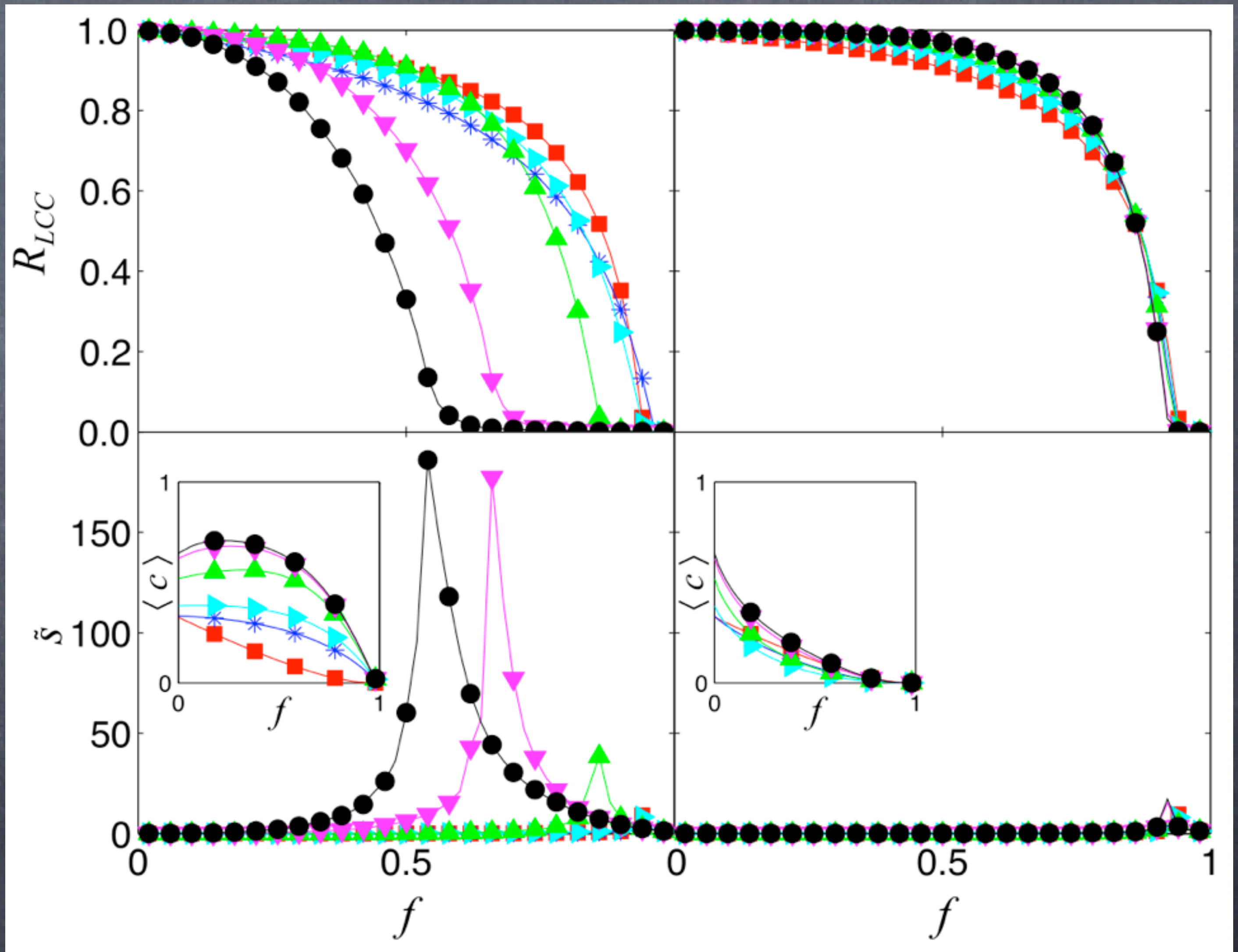




Global structure (percolation)

- Small $\delta < 0.1$ ($\delta = 10^{-3}$, $\delta = 0$)
  - Network disintegrates at the same point for weak and strong link removal
  - Incompatible with WTH
- Large $\delta > 0.1$ ($\delta = 1$, $\delta = 0.5$)
  - Network disintegrates at different points
  - Compatible with WTH

Weak go first | Strong go first



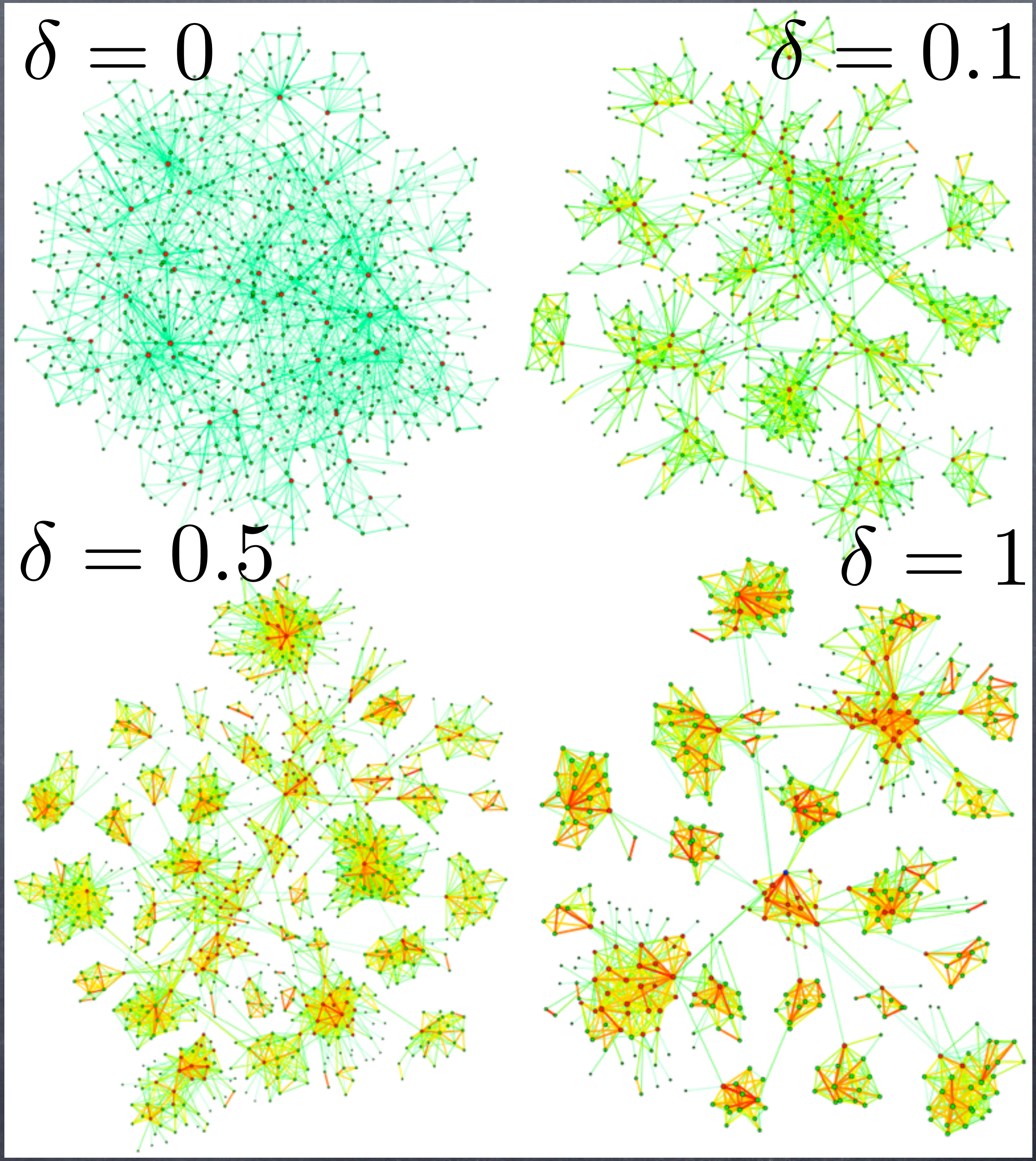
alizations of $N = 5 \times 10^4$ networks. Values of δ are 0 (□), 1×10^{-3} ($*$), 1×10^{-2} (▷), 0.1 (△), 0.5 (▽), and 1 (○).

Communities by inspection

- Increasing δ traps walks in communities, further enhancing trapping effect

  => Clear communities

- Triangles accumulate weight and act as nuclei for communities



Sociodynamic Models

- Mimick social processes taking place on networks
- Usually the outcome of dynamics is heavily affected by network structure
- Edge weights should affect interactions - however, only a few studies of soc. dyn. models on weighted networks exist.
- Examples:
  - SI, SIR (spreading processes) in the context of information/rumours
  - Threshold model (D.J. Watts, PNAS 99, 5766-577, 2002)
  - Opinion formation models: Voter, Majority Rule, Sznajd, language competition models, etc
  - See C. Castellano, S. Fortunato, V. Loreto: Statistical physics of social dynamics, Rev. Mod. Phys. 81, No. 2. (2009), pp. 591-646.