CS-E5740
Complex Networks

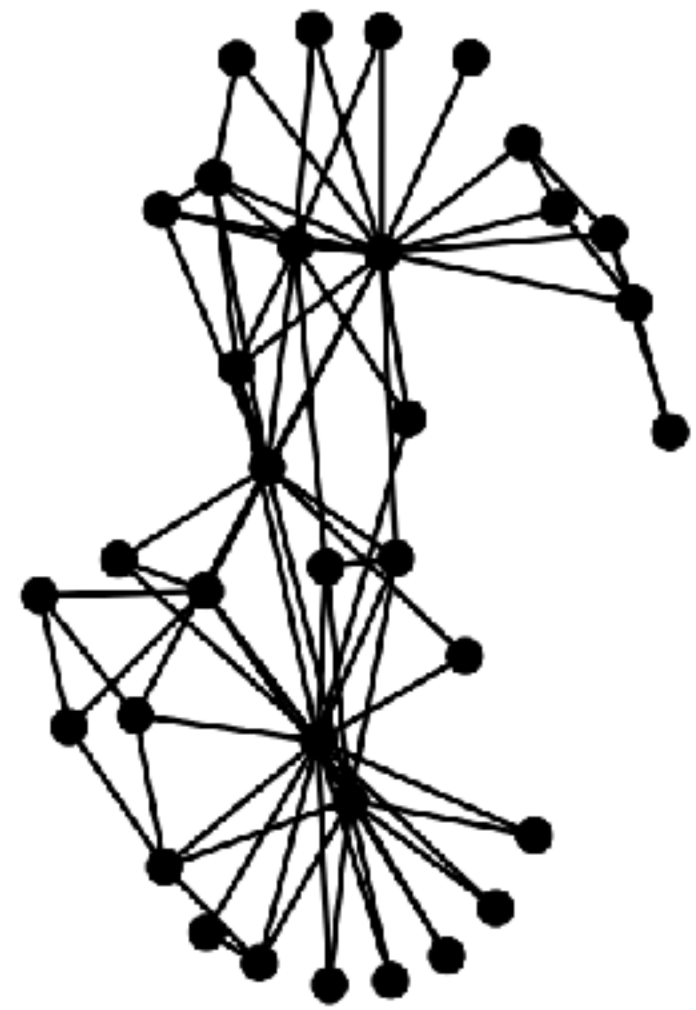
Weighted networks

Course outline

1. Introduction (motivation, definitions, etc.)
2. Static network models: random and small-world networks
3. Growing network models: scale-free networks
4. Percolation, error & attack tolerance of networks, epidemic models
5. Network analysis: key measures and characteristics
6. Social networks & (socio)dynamic models
7. Weighted networks
8. Clustering, sampling, inference
9. Temporal networks & multilayer networks

Are simple graphs enough?

VS



Networks with Weights

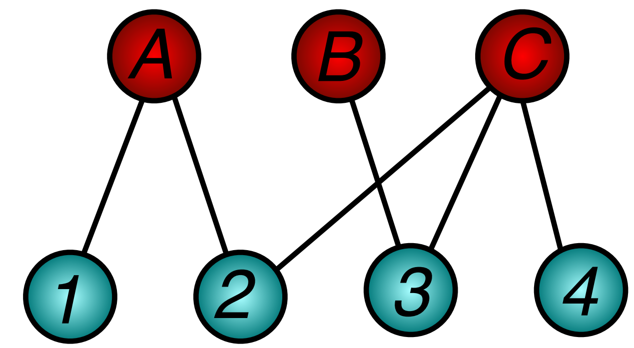
The network view:

- Elements ⇔ vertices
- Interactions ⇔ edges
- Interaction strengths ⇔ edge weights

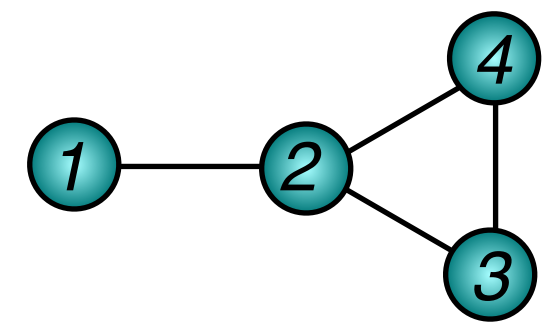
Vertex	Edge	Weight
person	friendship	closeness
neuron	synapse	synaptic strength
WWW	hyperlink	none
company	ownership	% owned
gene	regulation	level of regulation

From bipartite networks to weighted networks

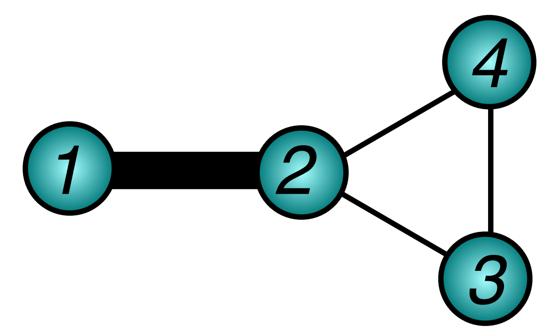
- Bipartite networks can be projected into weighted networks in multiple ways
  (δ_i = 1 iff i connected to p otherwise 0,
  n_p = degree of p)



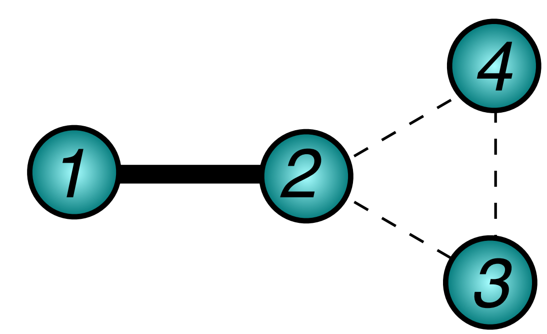
$$w_{ij} = \sum_p \delta_i^p \delta_j^p$$



$$w_{ij} = \sum_p \frac{\delta_i^p \delta_j^p}{n_p - 1}$$



$$w_{ij} = \sum_p \frac{\delta_i^p \delta_j^p}{n_p(n_p - 1)}$$

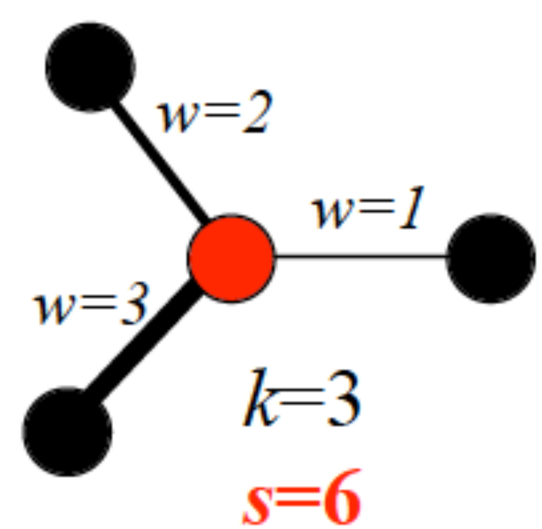


Weighted Networks: Fundamentals

- Let us denote the weight between i and j by w_{ij}
- (Usually) $w_{ij} \geq 0$ and $w_{ij} = 0$ means that there is no edge
- The weights w_{ij} form **a weight matrix W**, analogous to the adjacency matrix
- For undirected networks, $W^T = W$
- If weights $w_{ij} = \{0,1\}$, W=A

- The notion of degree is readily generalized for weighted networks; the **strength** of a vertex is defined as

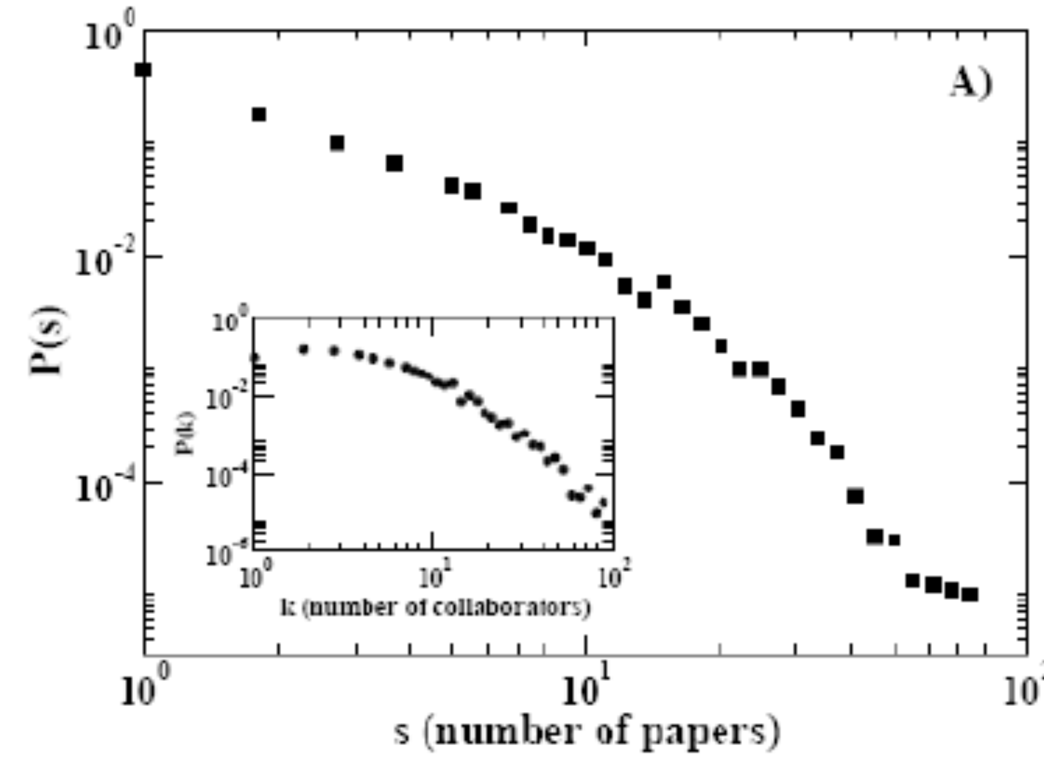
$$s_i = \sum_{j=1}^{N} w_{ij}$$



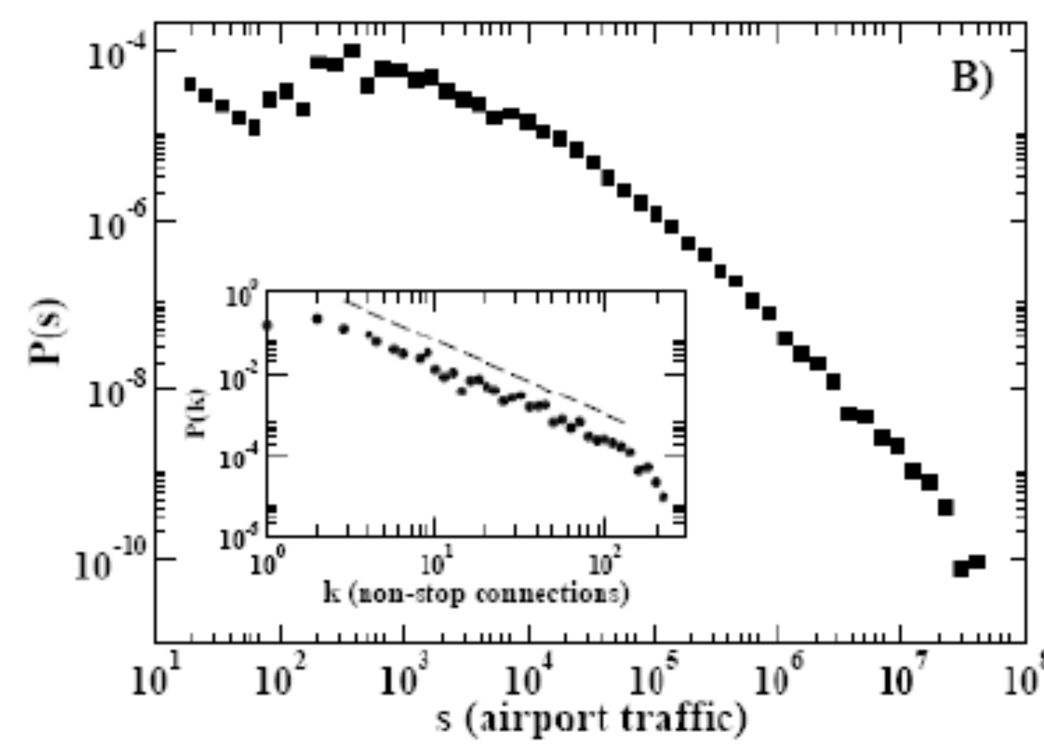
Strength distributions of real-world networks

- Just as for the degree, we can investigate the strength distributions P(s) of networks
- And (unsurprisingly), these tend to be broad and have power-law-like tails
- Evidently this has to do with the fact that degrees and strengths are related

scientific collaborations



world-wide airport network

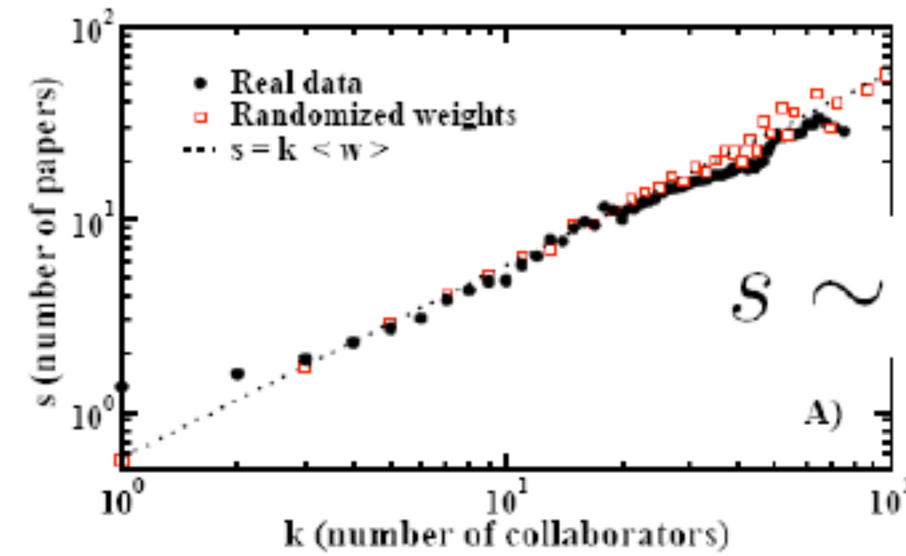


A. *Barrat*, M. *Barthélemy*, R. Pastor-Satorras, and A. *Vespignani* (2004). "The architecture of complex weighted networks". *PNAS* 101 (11): 3747–3752

Strength-Degree Correlations

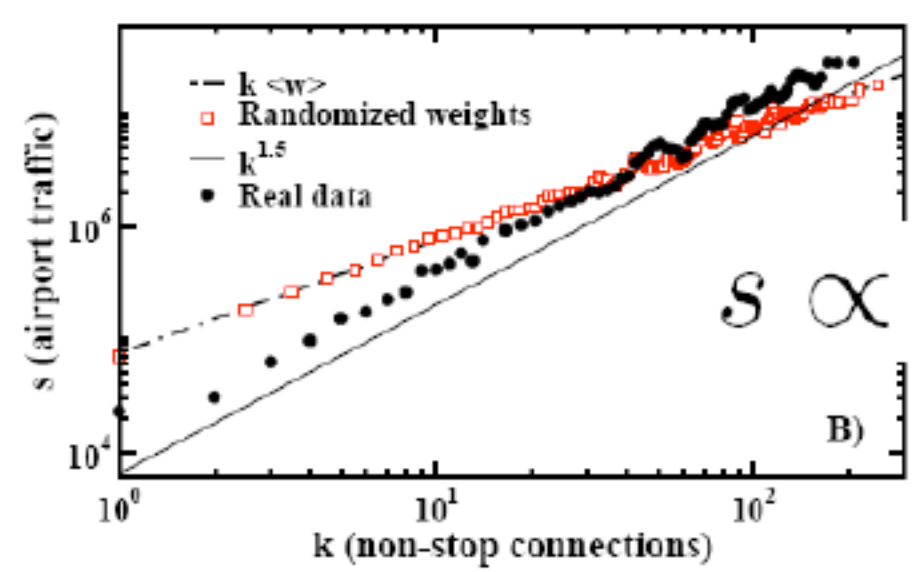
- How does the strength of nodes depend on degree?
- If on average $\langle s(k)\rangle \sim \langle w\rangle k$, i.e. there is a linear dependence, link weights and node degrees do not correlate
- If $\langle s(k)\rangle \sim k^{\beta}$,
  - high-degree nodes have higher weights if $\beta > 1$
  - high degree implies lower weights if $\beta < 1$

scientific collaborations



$s \sim k\,\langle w\rangle$

world-wide airport network



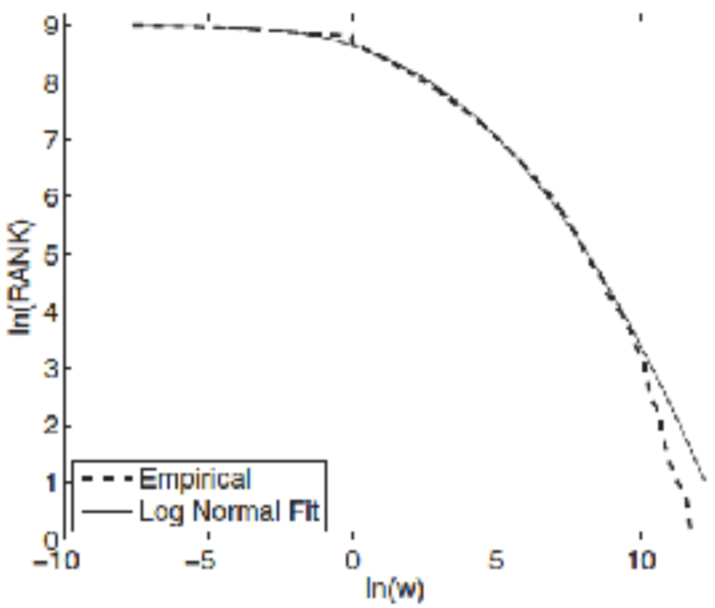
$s \propto k^{1.5}$

A. *Barrat*, M. *Barthélemy*, R. Pastor-Satorras, and A. *Vespignani* (2004). "The architecture of complex weighted networks". *PNAS* 101 (11): 3747–3752

Weight distributions are typically broad, too

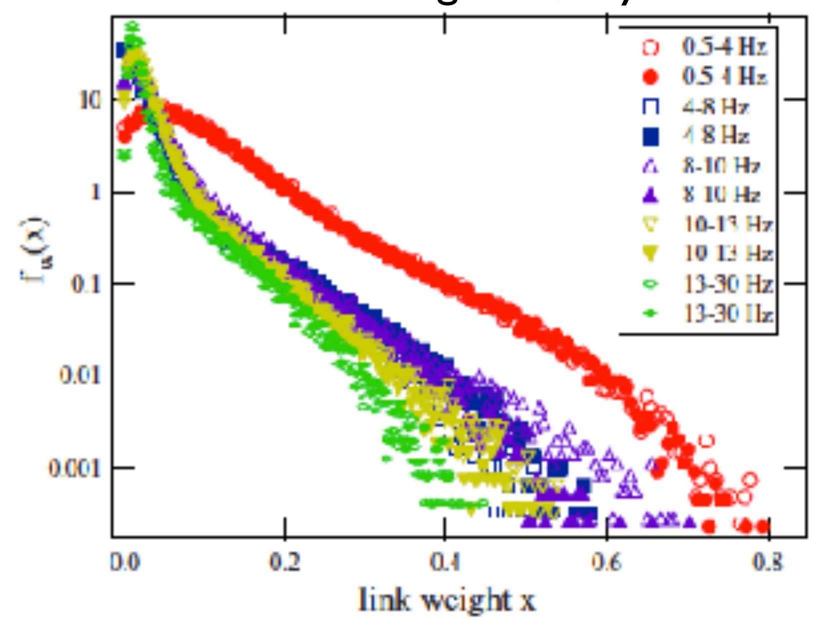
world trade network

Fagiolo et al, Phys. Rev. E 79, 036115 (2009)



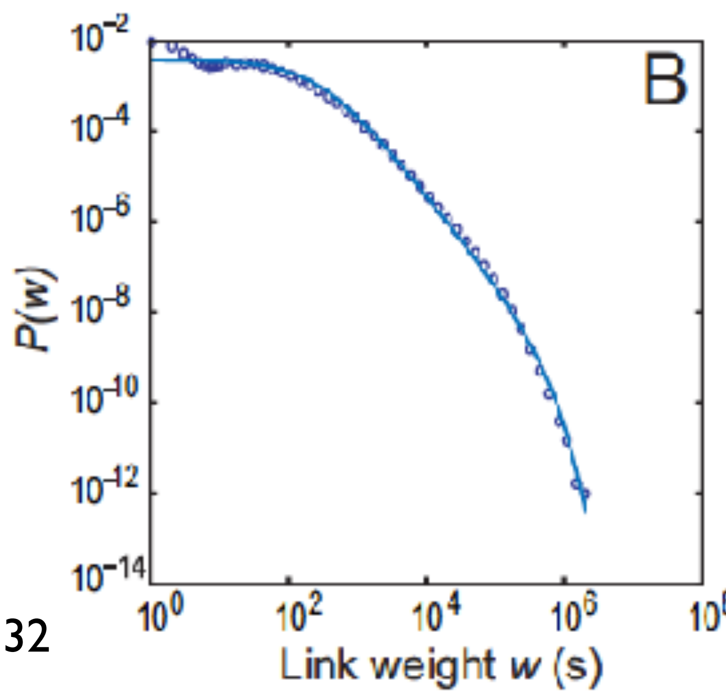
brain functional networks

Wang et al, Phys. Rev. E 82, 021924 (2010)



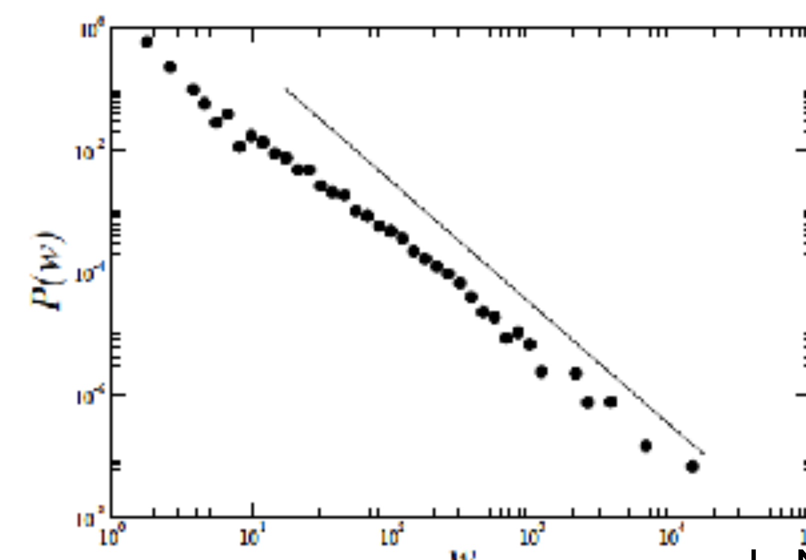
social tie strengths

Onnela et al, PNAS 104, 7332 (2004)



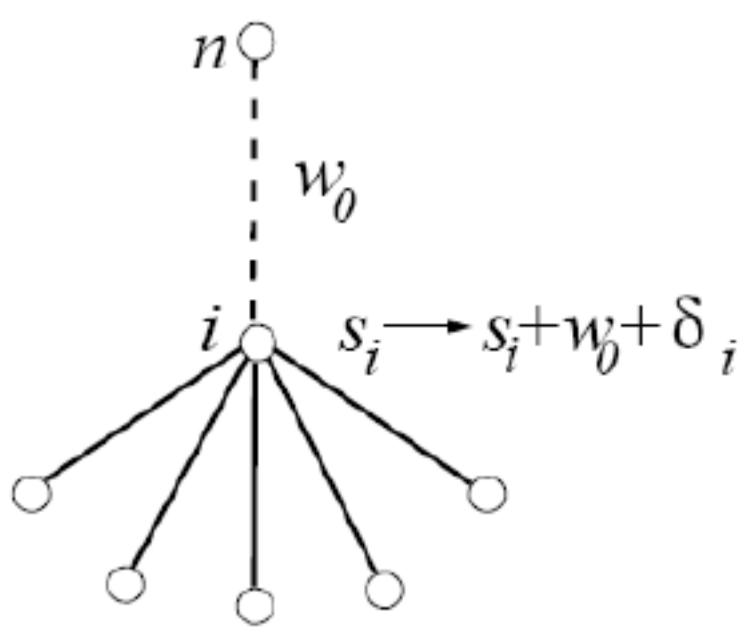
inter-urban traffic

de Montis et al, Environment and Planning: B, 34:905-924 (2007)



The Barrat-Barthélemy-Vespignani Model

- A "generalized" version of the BA model
- Power-law distributions of degrees, strengths and weights



1. Take a small seed network
2. Create a new vertex with m edges of weight w_0
3. Connect the m edges to existing vertices with a probability proportional to their strength s:

$$\pi_i = \frac{s_i}{\sum_j s_j}$$

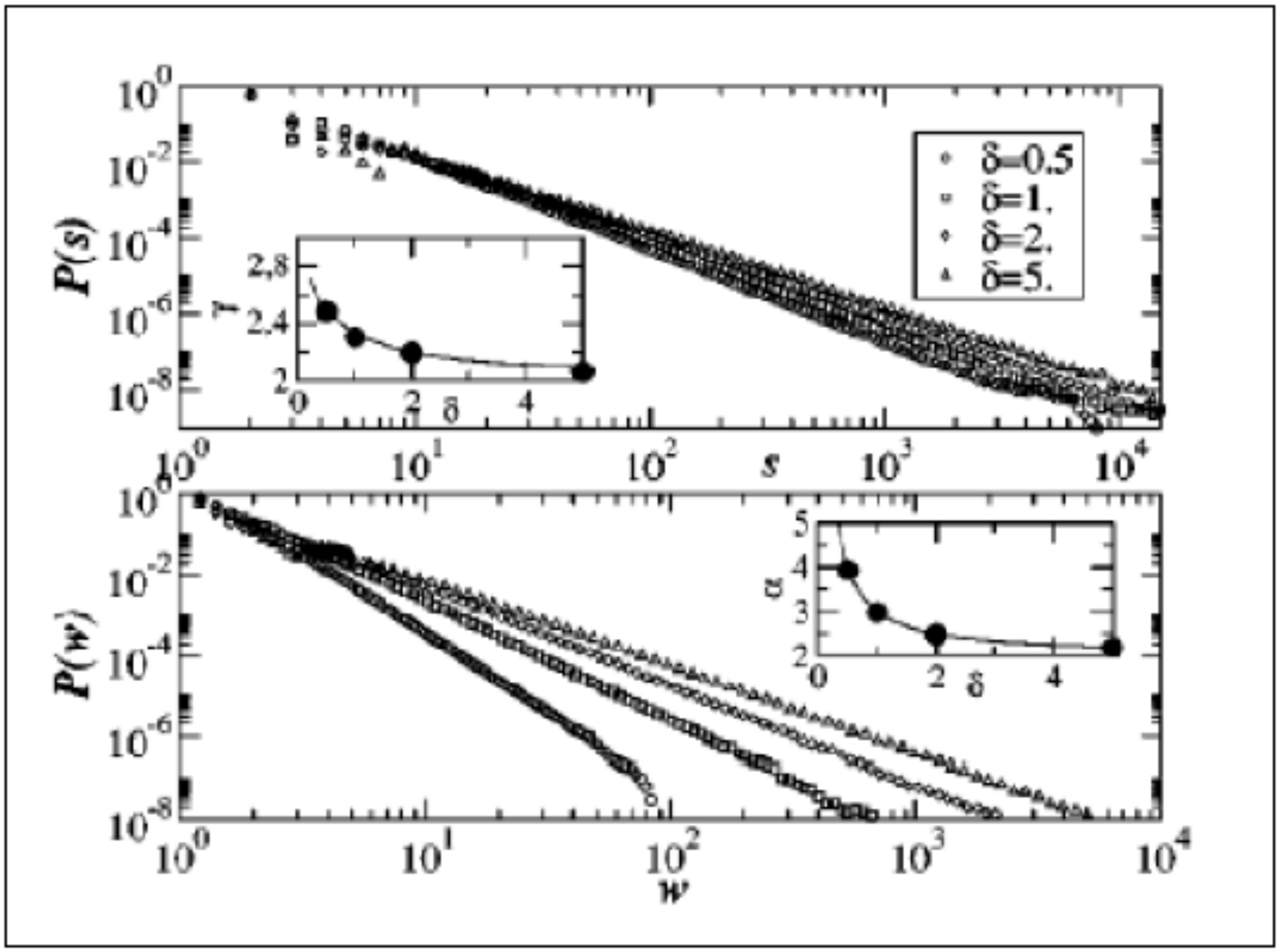
4. Update weights of edges of the selected vertices according to

$$\Delta w_{ij} = \delta_i \frac{w_{ij}}{s_i}$$

5. Repeat 2-4.

Barrat, Barthelemy, Vespignani, Phys Rev Lett **92**, 228701 (2004)

The Barrat-Barthélemy-Vespignani Model



- The exponents of strength, degree, and weight power laws depend on the weight addition parameter δ
- The strength and degree power-law exponents $\gamma \in [2,3]$
- Recall that for natural networks with $P(k) \propto k^{-\gamma}$, $\gamma \in [2,3]$, whereas

  (almost) all unweighted models yield $\gamma \geq 3$
- ...so one possible explanation for real-world exponents is that weights are involved!

How to generalize measures for weighted networks?

- For some quantities simply weighting the contributions of edges by $w_{ij}/(\sum_j w_{ij})$ works fine
- For example, one can consider the weighted average nearest-neighbour degree

$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_j$$

- Evidently, all weighted quantities should be equal to their unweighted counterparts if $w_{ij} = \{0,1\}$
- Often, the actual meaning and definition of weights plays a role
- E.g. should "weighted path lengths" be defined as $l=\sum w_{ij}$ or $l=\prod w_{ij}$?
- Some other quantities can be defined in numerous ways (which might indicate that the wrong question is being asked...)

Case Example: The Weighted Clustering Coefficient

Barrat & Barthélemy & Vespignani:

$$\widetilde{C}_{i,B} = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{jk} a_{ik}$$

Onnela, Saramäki, Kaski, Kertész:

$$\widetilde{C}_{i,O} = \frac{1}{k_i(k_i - 1)} \sum_{j,k} (\hat{w}_{ij} \hat{w}_{ik} \hat{w}_{jk})^{1/3}$$

Zhang et al:

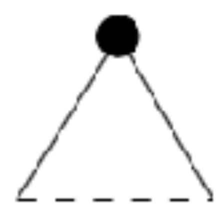
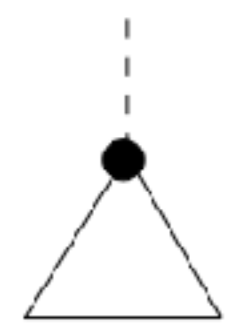
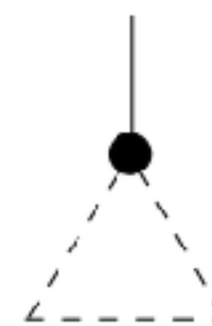
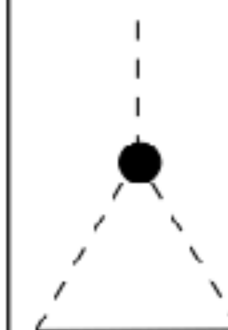
$$\widetilde{C}_{i,Z} = \frac{\sum_{j,k} \hat{w}_{ij} \hat{w}_{jk} \hat{w}_{ik}}{\left(\sum_k \hat{w}_{ik}\right)^2 - \sum_k \hat{w}_{ik}^2}$$

Holme et al:

$$\widetilde{C}_{i,H} = \frac{\sum_{j,k} w_{ij} w_{jk} w_{ki}}{\max(w) \sum_{j,k} w_{ij} w_{ki}} = \frac{\mathbf{W}_{ii}^3}{(\mathbf{W} \mathbf{W}_{\max} \mathbf{W})_{ii}}$$

...same idea, different formulas, different behaviour...

Case Example: The Weighted Clustering Coefficient

							
$\widetilde{C}_B$	1	1	1	~1/2	~0	1/3	~1/2
$\widetilde{C}_O$	~0	~0	~0	1/3	~0	~0	~0
$\widetilde{C}_Z$	1	~0	1	~1	~0	1/3	~0

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

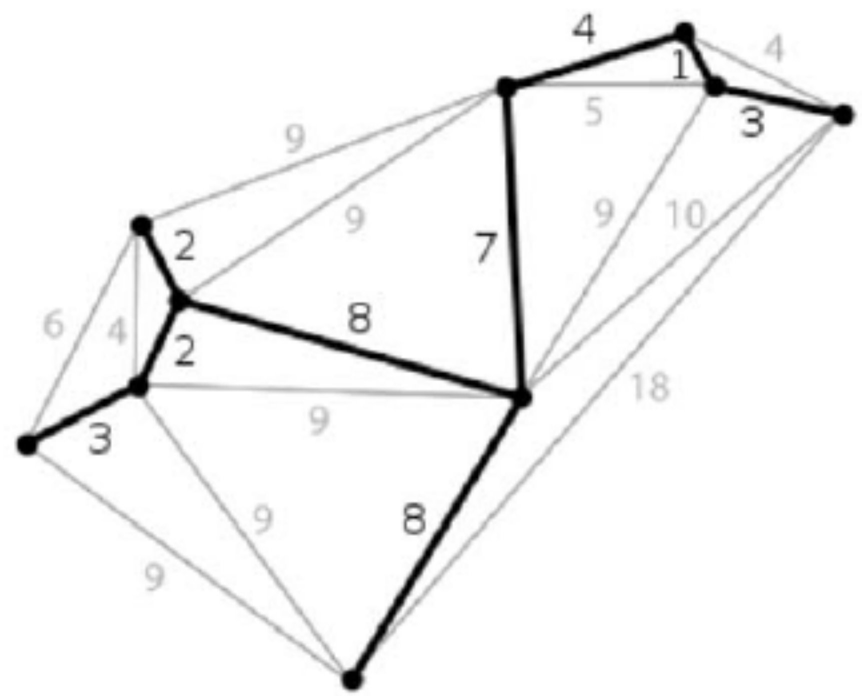
- There are different scales to this problem: the global, macroscopic scale of overall connectivity and the mesoscopic scale of clusters and modules

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

- <u>Maximal (minimal) spanning tree:</u> a tree built using a subset of the original links, so that
  1) the tree contains all the nodes of the original network
  2) the sum of link weights is maximized (minimized)



network has N nodes
⇒ spanning tree has N-1 links

Computing MST's

- These two greedy algorithms both give correct result.
- Presented here for minimal spanning tree.

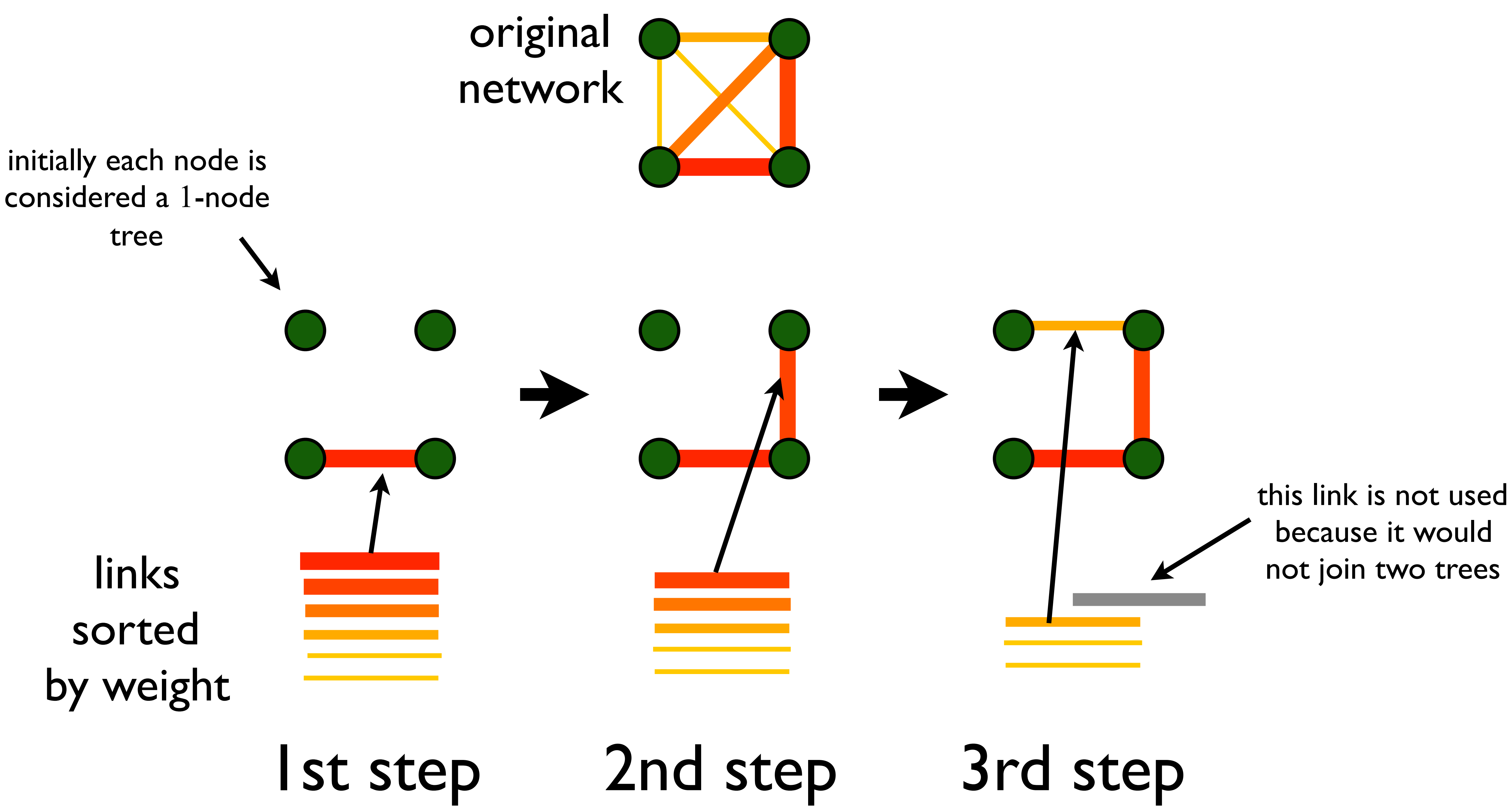
Kruskal's algorithm

- Let L be a list of edges sorted by weight.
- Start with an empty forest F of n nodes.
- Repeat until F is a tree:
  - Take an edge with the smallest weight from L. If it joins two trees in F, add it, otherwise discard it.

Prim's algorithm

- Pick any node v_0 to start. Initialize $T = (V, E)$ as $V = \{v_0\}$, $E = \emptyset$
- Repeat until all nodes are are included:
  - Select an edge (v_i, v_j) with minimal weight such that exactly one of the nodes is in V. Add to T.

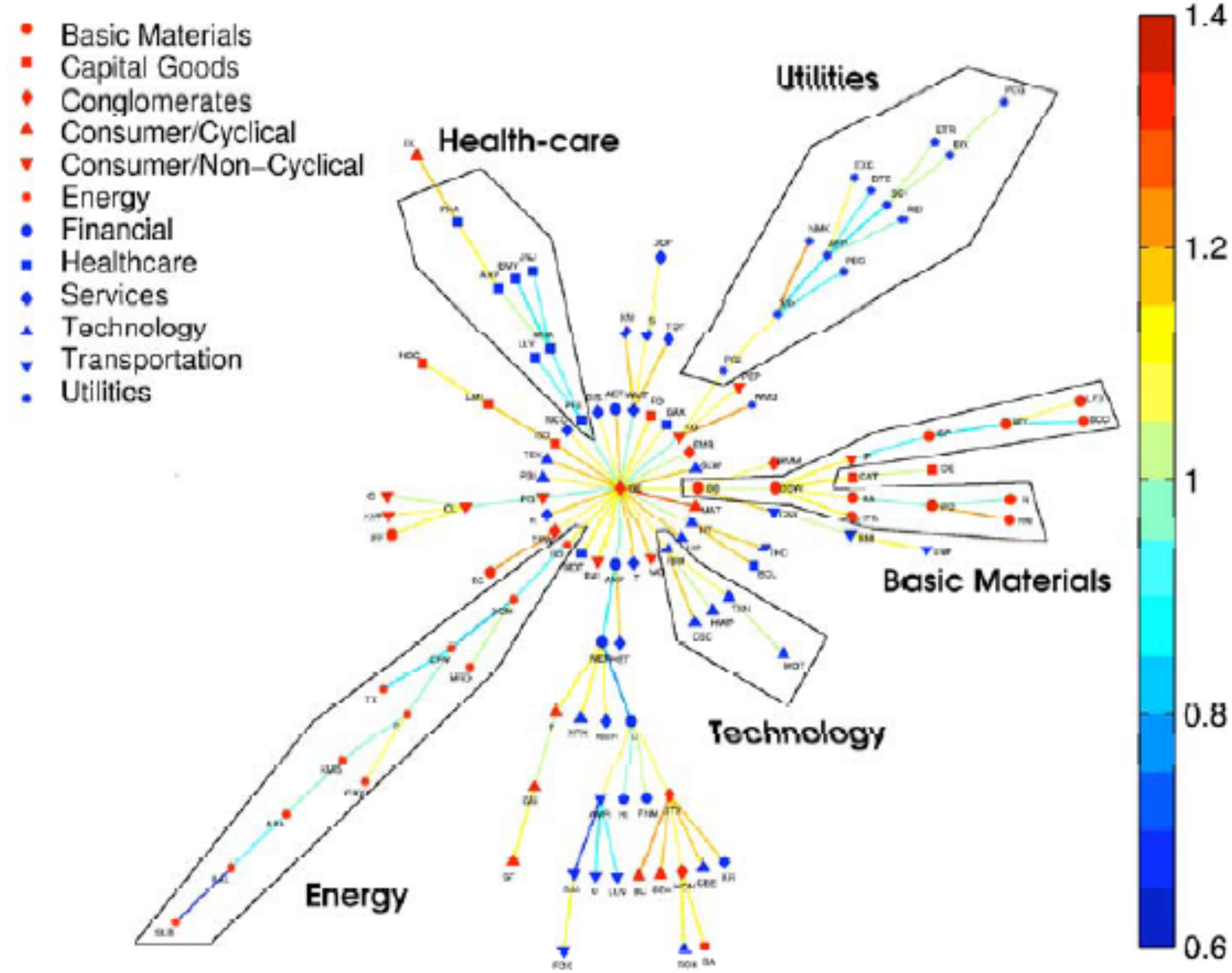
Maximal Spanning Tree: Kruskal's algorithm



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

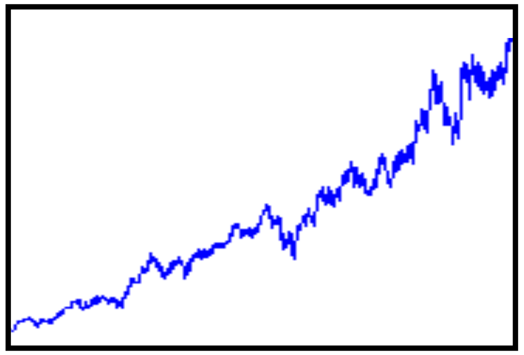


Onnela et al. Phys. Rev. E 68, 056110 (2003)

Weighted networks from time series
Example: stock networks

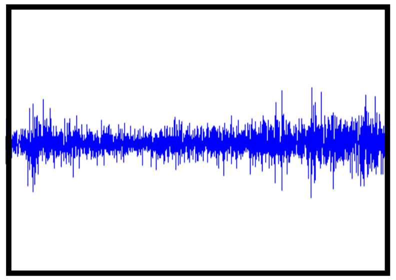
1. Starting point: time series of prices for N stocks

$$P_i(t)$$

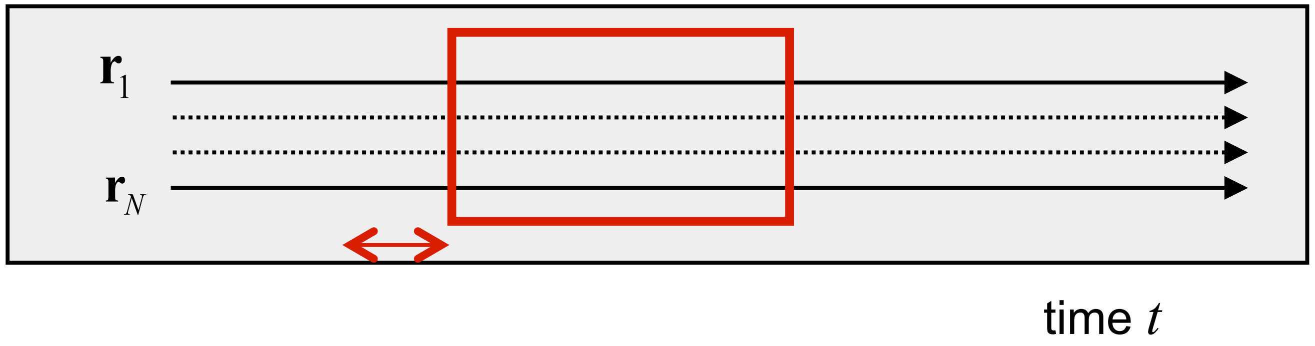


2. Calculate log-returns (usually done for stock prices, not feasible for all time series)

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



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



Weighted networks from time series Example: stock networks

4. Calculate correlation coefficients between time series within your windows

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

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

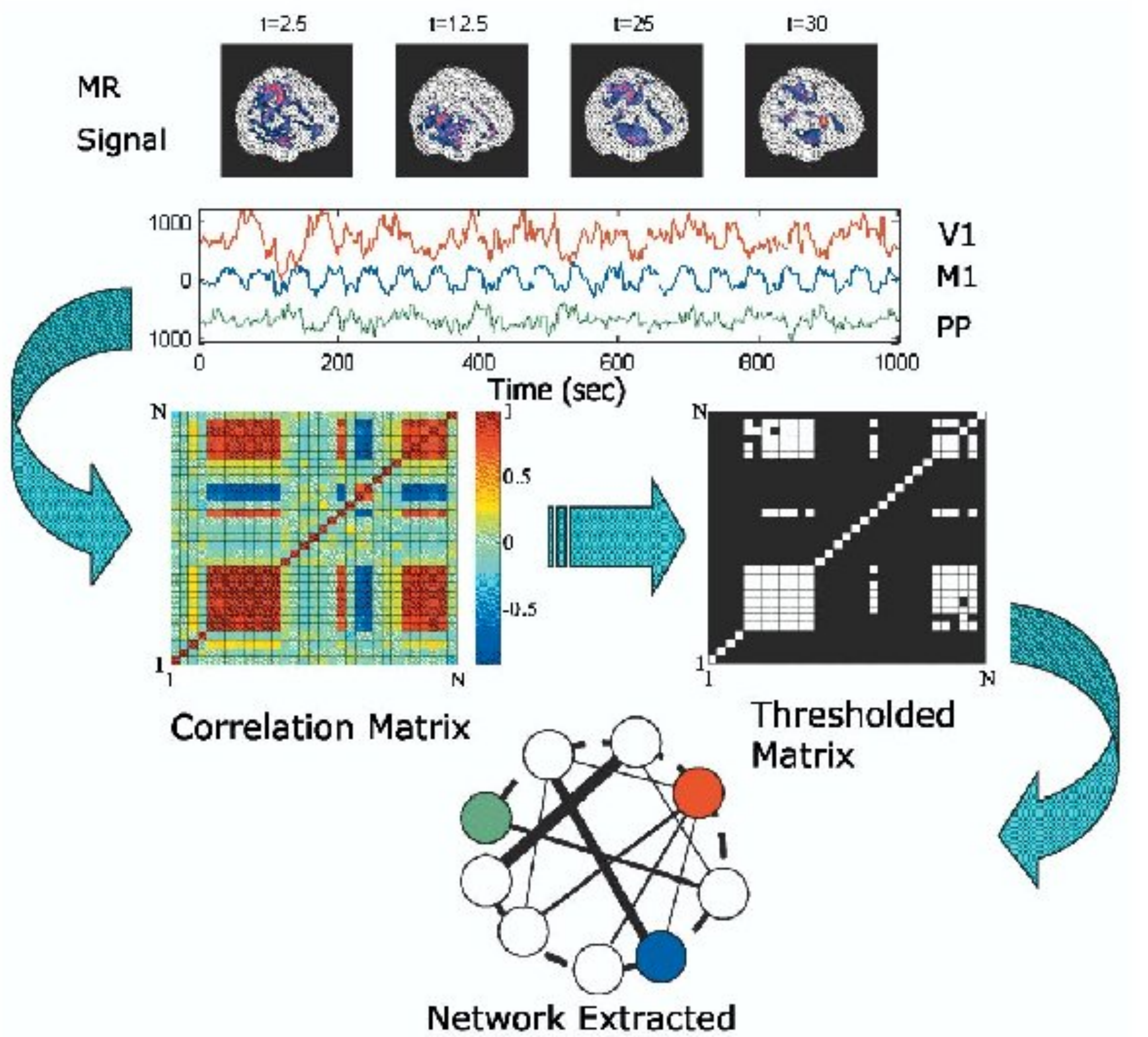
5. If negative coefficients exist, take absolute values

6. This correlation matrix is your weight matrix W_{ij} for a window

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

Networks from time series
Example: Brain Functional Networks

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



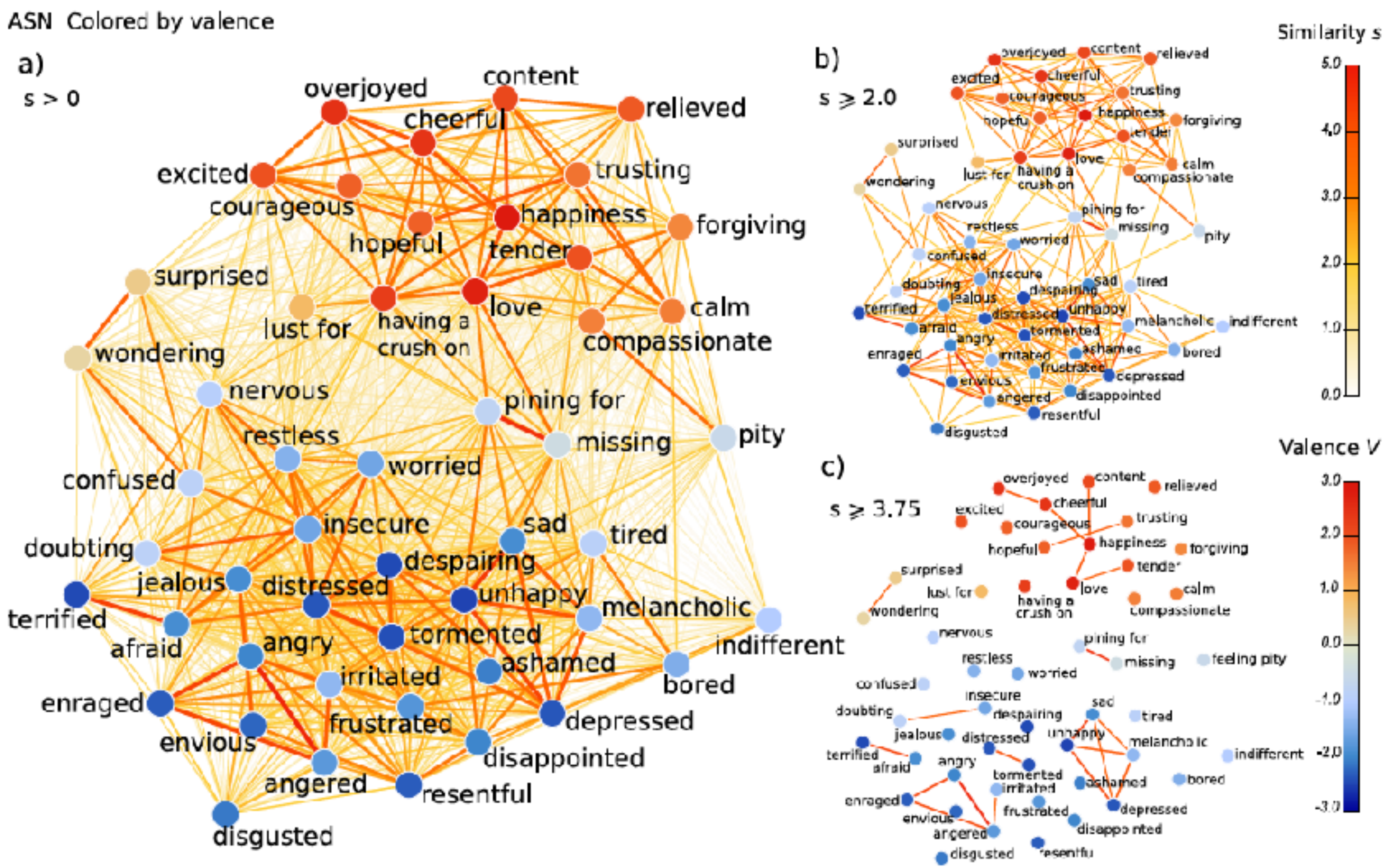
Matrices to Networks: Thresholding

- 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



[Networks of emotion concepts](), R. Toivonen, M. Kivelä, J. Saramäki, M. Viinikainen, M. Vanhatalo, and M. Sams, *PLoS One* **7(1)**, e28883 (2012)

Advanced methods for pruning dense weight matrices

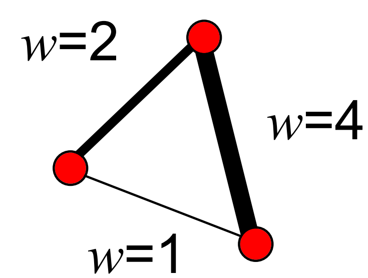
- Statistical significance of links:
  - Tumminello et al, Statistically Validated Networks in Bipartite Complex Systems, PLoS One **6**: e17994 (2011)
  - M. Ángeles Serrano et al, Extracting the multiscale backbone of complex weighted networks, PNAS **106**, 6483 (2009)
- Graphs embedded on surfaces (instead of spanning trees):
  - Tumminello et al, A tool for filtering information in complex systems, PNAS **102**, 10421 (2005)
- Dynamics determining link weights:
  - Lambiotte et al, Flow graphs: interweaving dynamics and structure, Phys Rev E 84, 017102 (2011)

The local role of weights

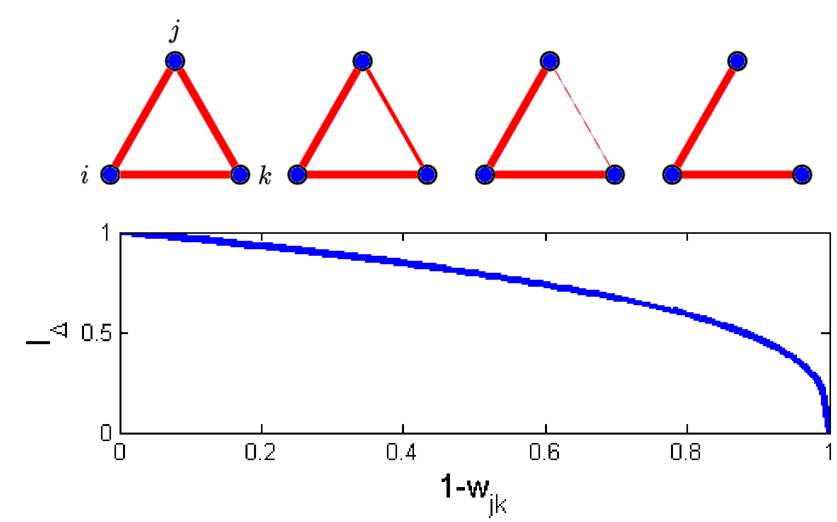
Weighted Subgraphs: Intensity

- 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

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



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



- Measures the "weight" of a subgraph as geometric average of weights of its edges

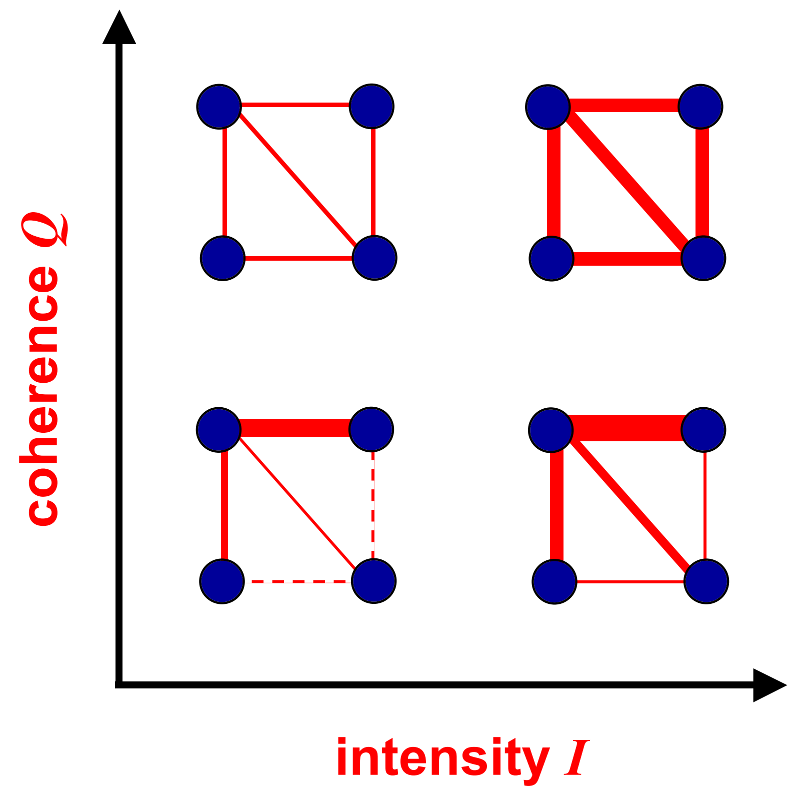
- Becomes low if any of the weights is low

[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

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

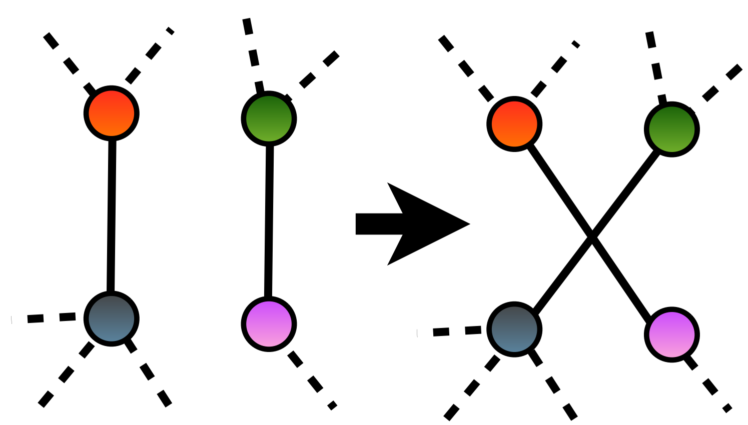


- Measures how equal ("coherent") the weights are
- If all weights $w_{ij} = w$, $Q(g)=1$

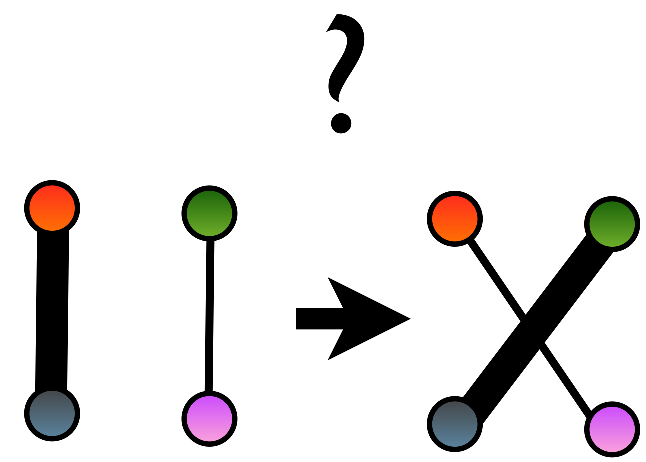
[Onnela, Saramäki et al, Phys Rev E **71**, 065103 (2005)]

Weighted networks & reference models

- Unweighted networks: the configuration model randomizes networks preserving degrees



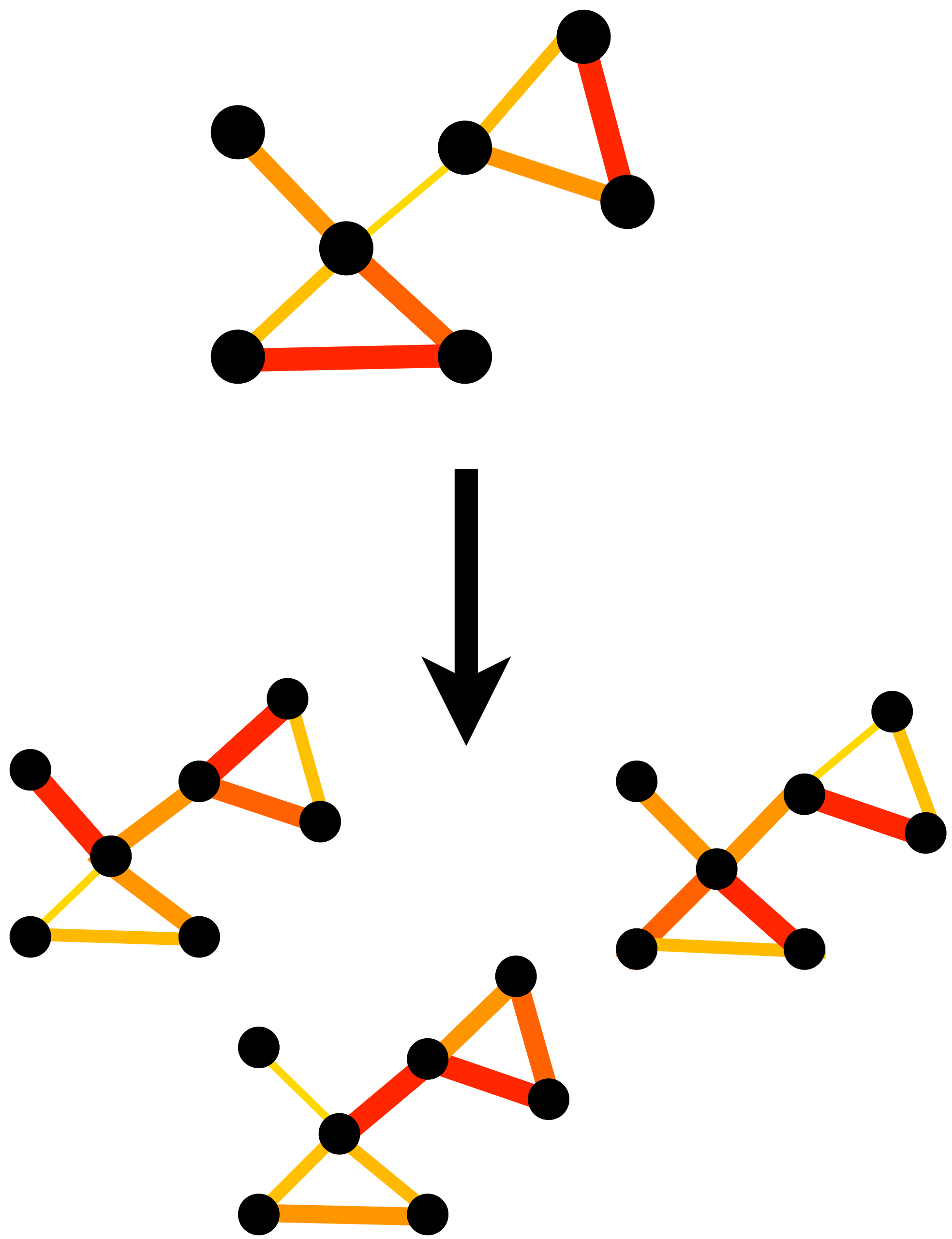
- Weighted networks: what to preserve? Degrees, strength, weights?



- One cannot rewire links and preserve node strengths.

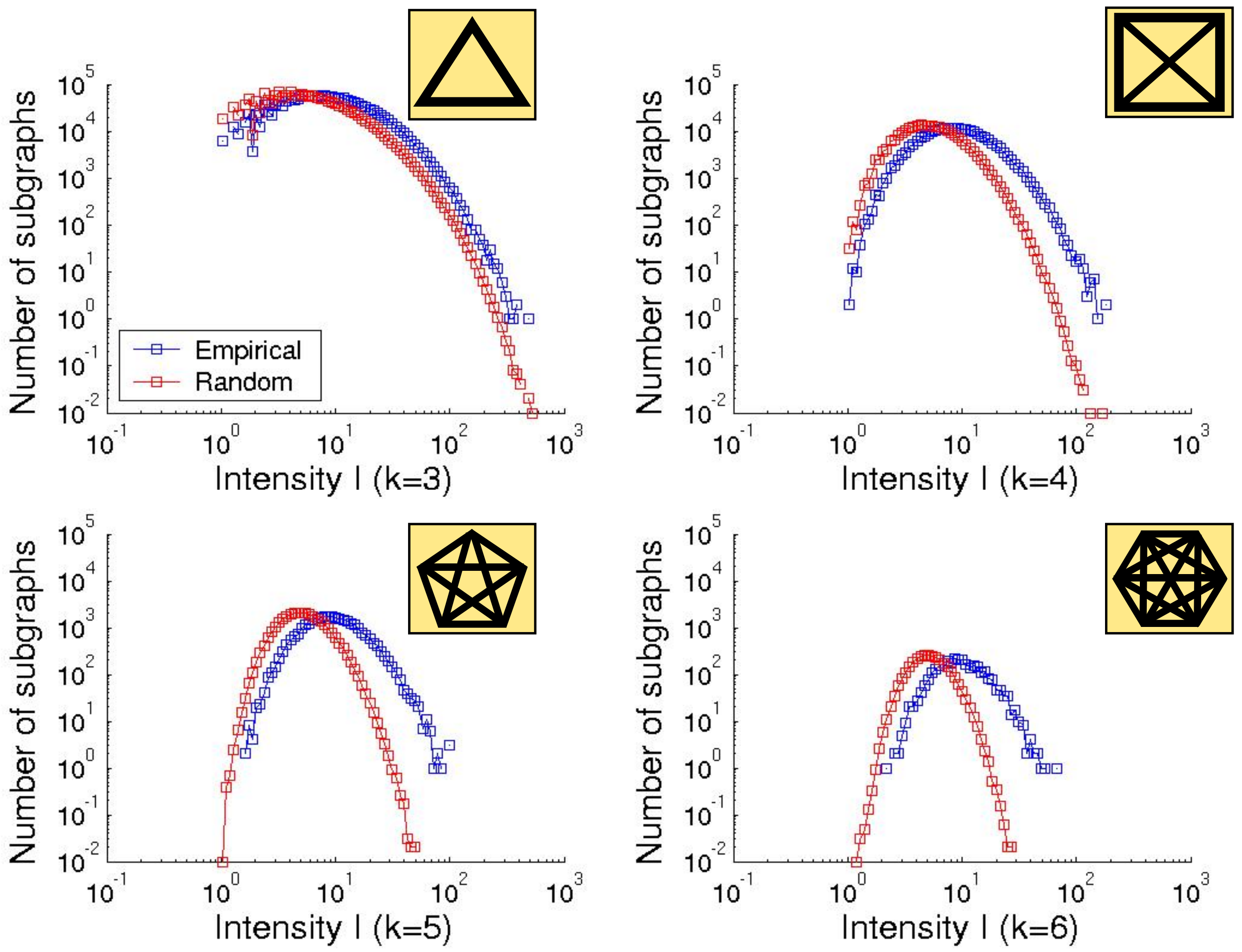
Weighted networks: weight shuffling

- Randomly displace all link weights in the network
- Retains degrees and topology
- Destroys the strength distribution
- Retains the overall weight distribution
- Destroys weight-topology correlations!



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)

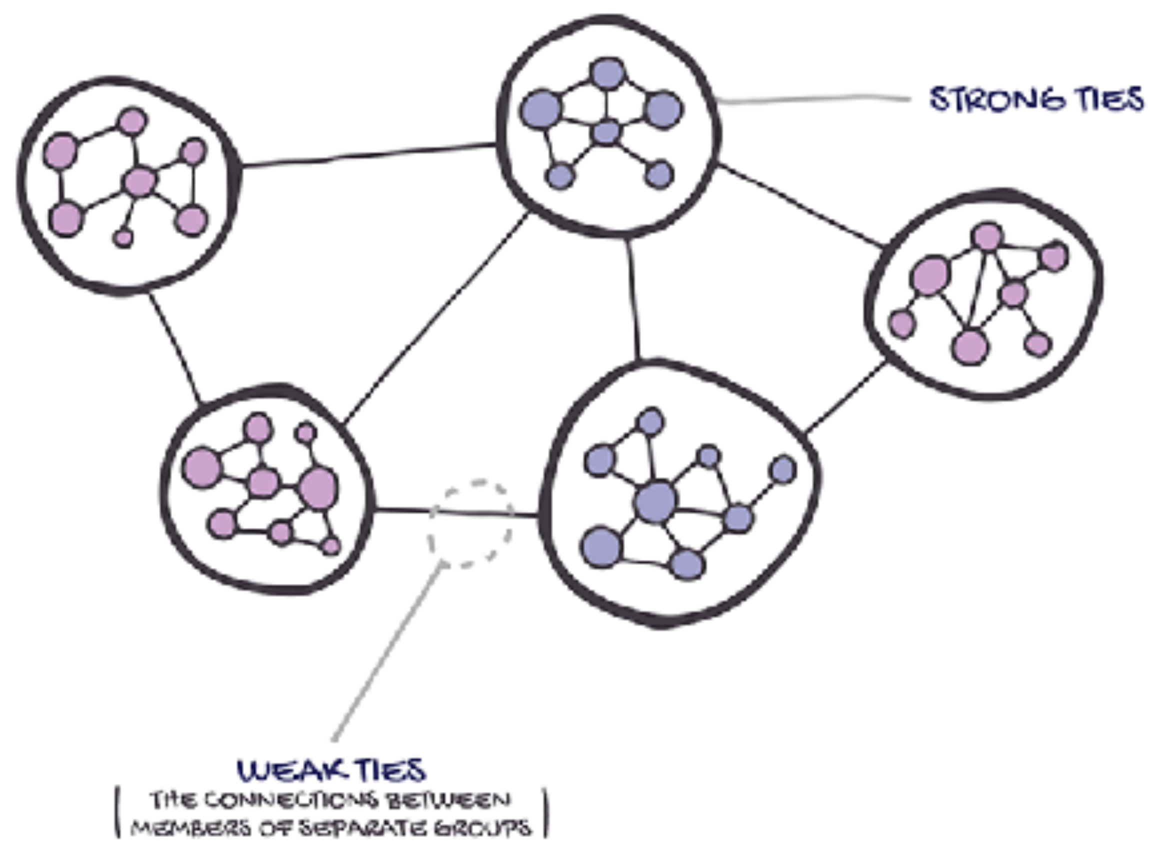
Global role of weights

Weak/Strong Ties: Granovetter's Hypothesis

- **Hypothesis**:
  "The overlap between the circles of friendship of two persons depends directly on the strength of their tie"

- **Strong links are inside social groups** (overlap = triangles!)

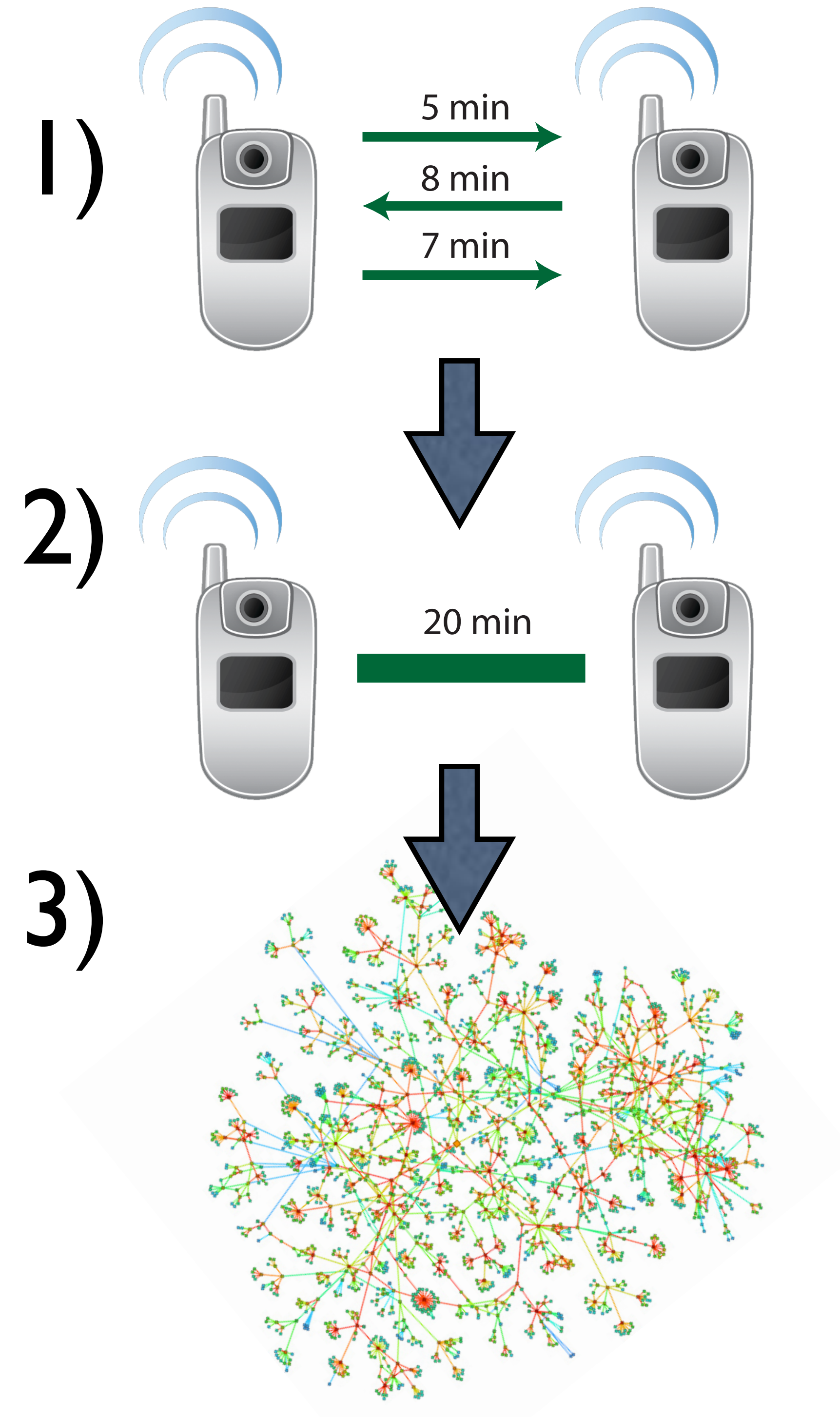
- **Weak links connect social groups** (like Burt's bridges).



[figure from: http://blog.headresourcing.com/networking-and-the-strength-of-weak-ties/]

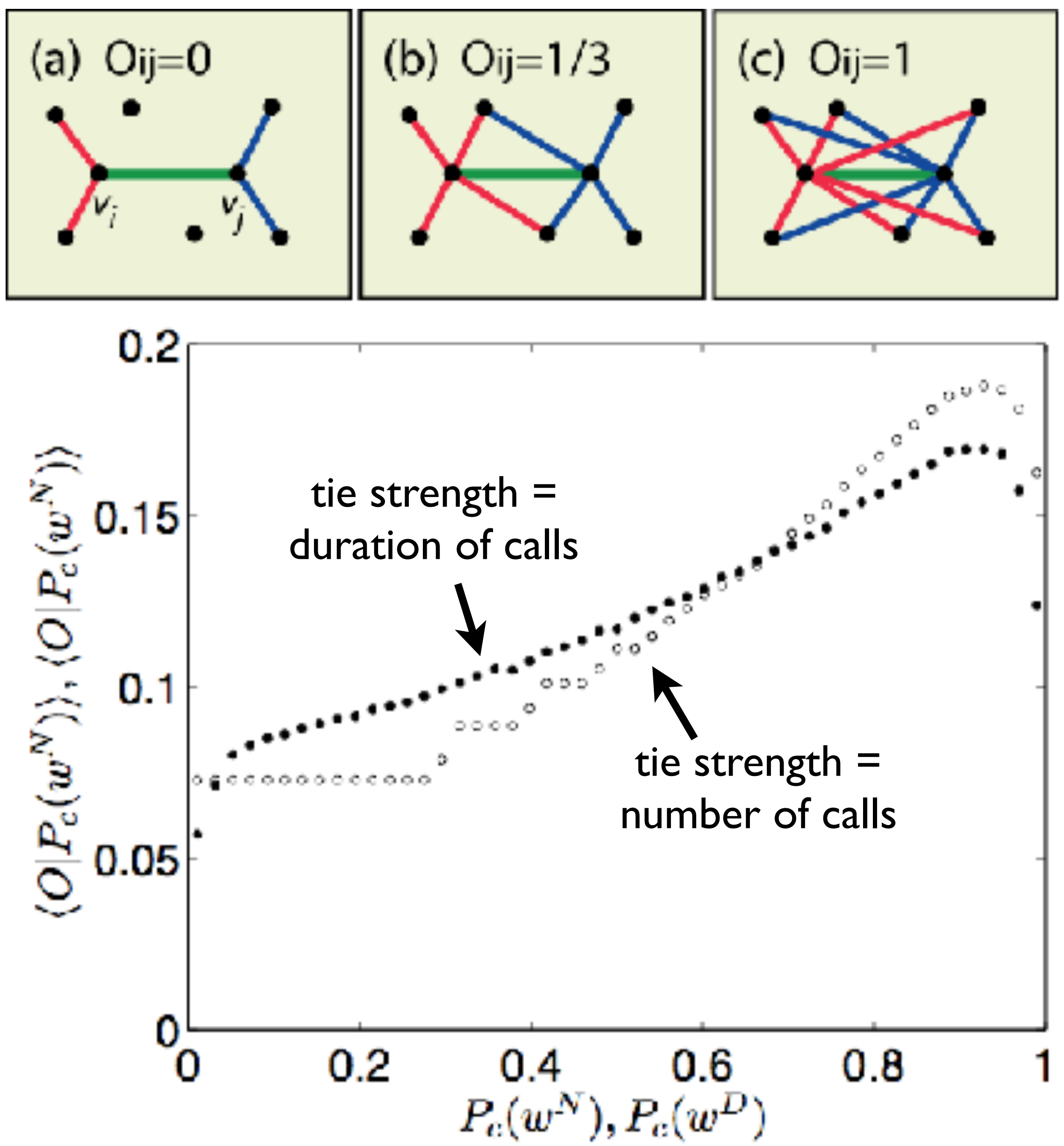
The Granovetter hypothesis in mobile communication networks

- source data: call records of an European mobile operator
- ~7 million subscribers
- data for 18 weeks
- network construction: link persons (nodes) if they have called each other
- the strength of the link: total call minutes between A and B



Onnela, Saramäki, *et al.*,
Proc. Natl. Acad. Sci. (USA) **104**, 7332 (2007),
New Journal of Physics **9**, 179 (2007)

Verifying the Granovetter hypothesis



Onnela, Saramäki, *et al.*,
Proc. Natl. Acad. Sci. (USA) **104**, 7332 (2007),
New Journal of Physics **9**, 179 (2007)

- Use the overlap O_{ij}

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

- Calculate average overlap as function of link weight
- There is an increasing tendency, i.e. the hypothesis holds

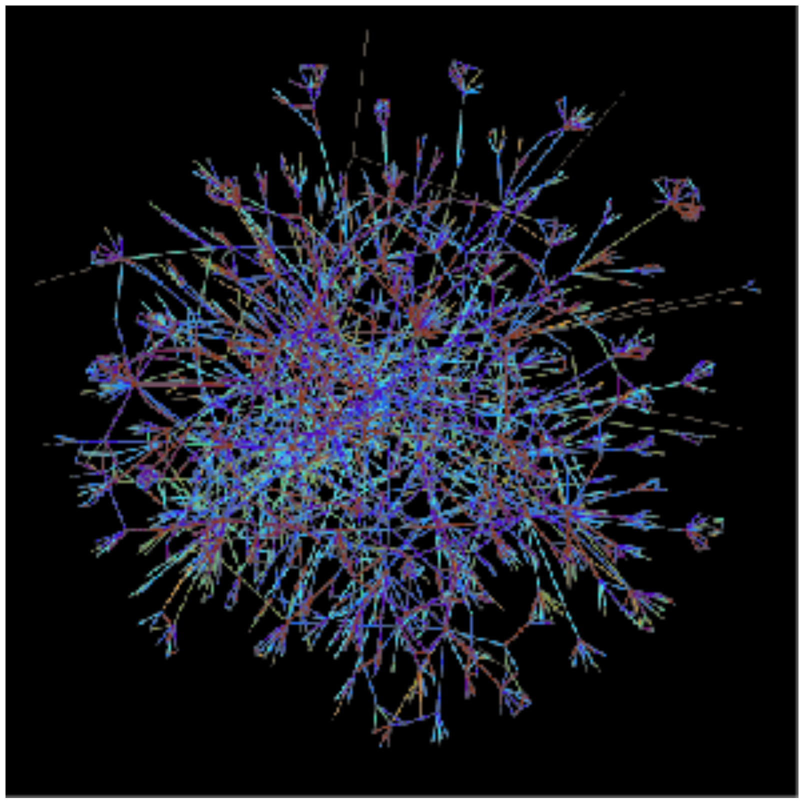
Percolation & weight-topology correlations

- Remove links by order of increasing/decreasing weight
- If one order of removal breaks down the network faster than the other, those links are more important for overall connectivity

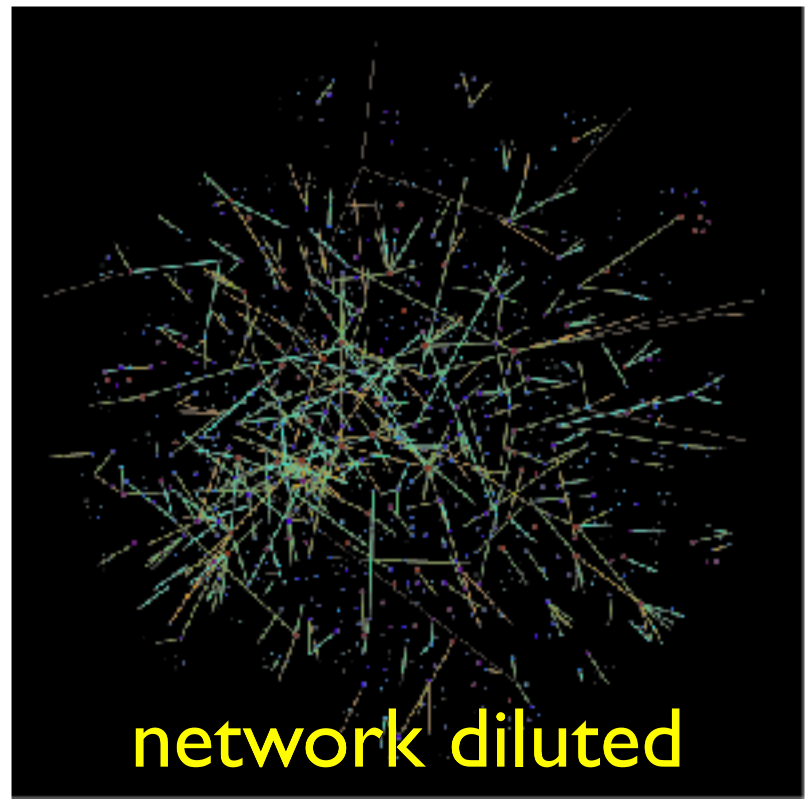
- Example: social network from phone calls
- Nodes = people
- Weights = call minutes between people
- Millions of nodes

Role of weak links in connectivity

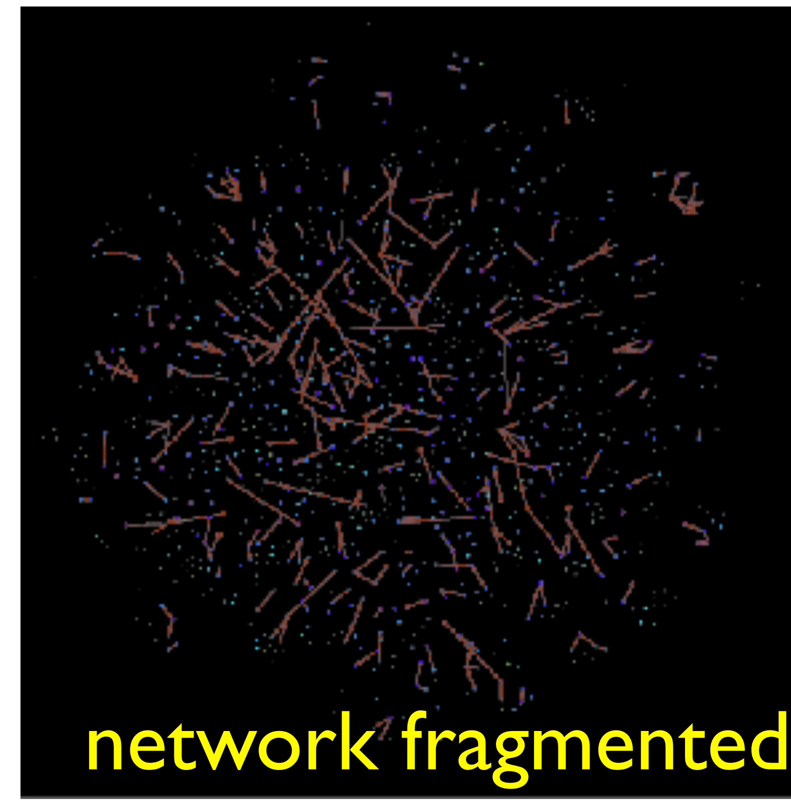
a small subnetwork



80% strongest links removed

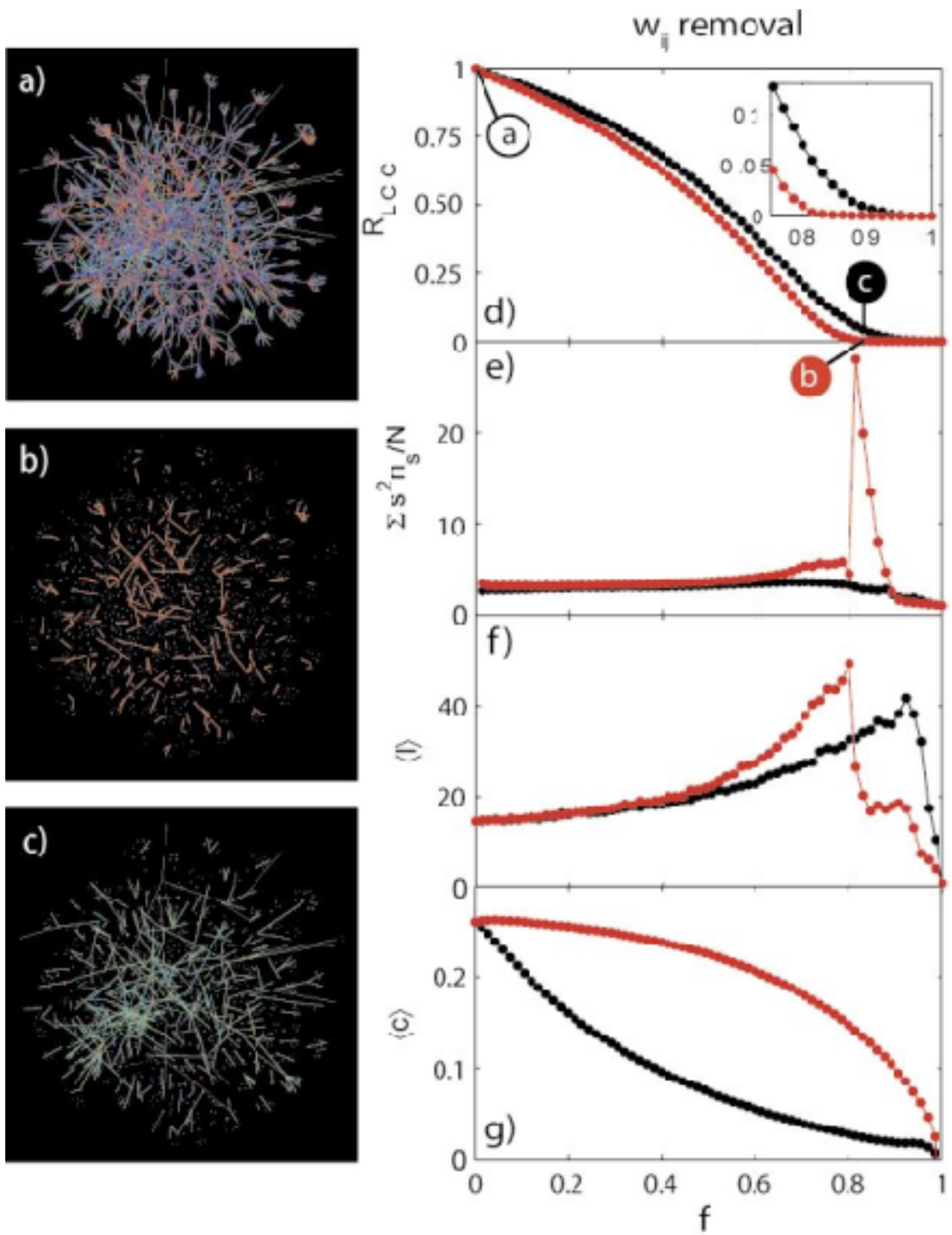


80% weakest links removed



Onnela, Saramäki, *et al.*,
Proc. Natl. Acad. Sci. (USA) **104**, 7332 (2007),
New Journal of Physics **9**, 179 (2007)

Percolation Analysis: Empirical Example



fraction of removed links

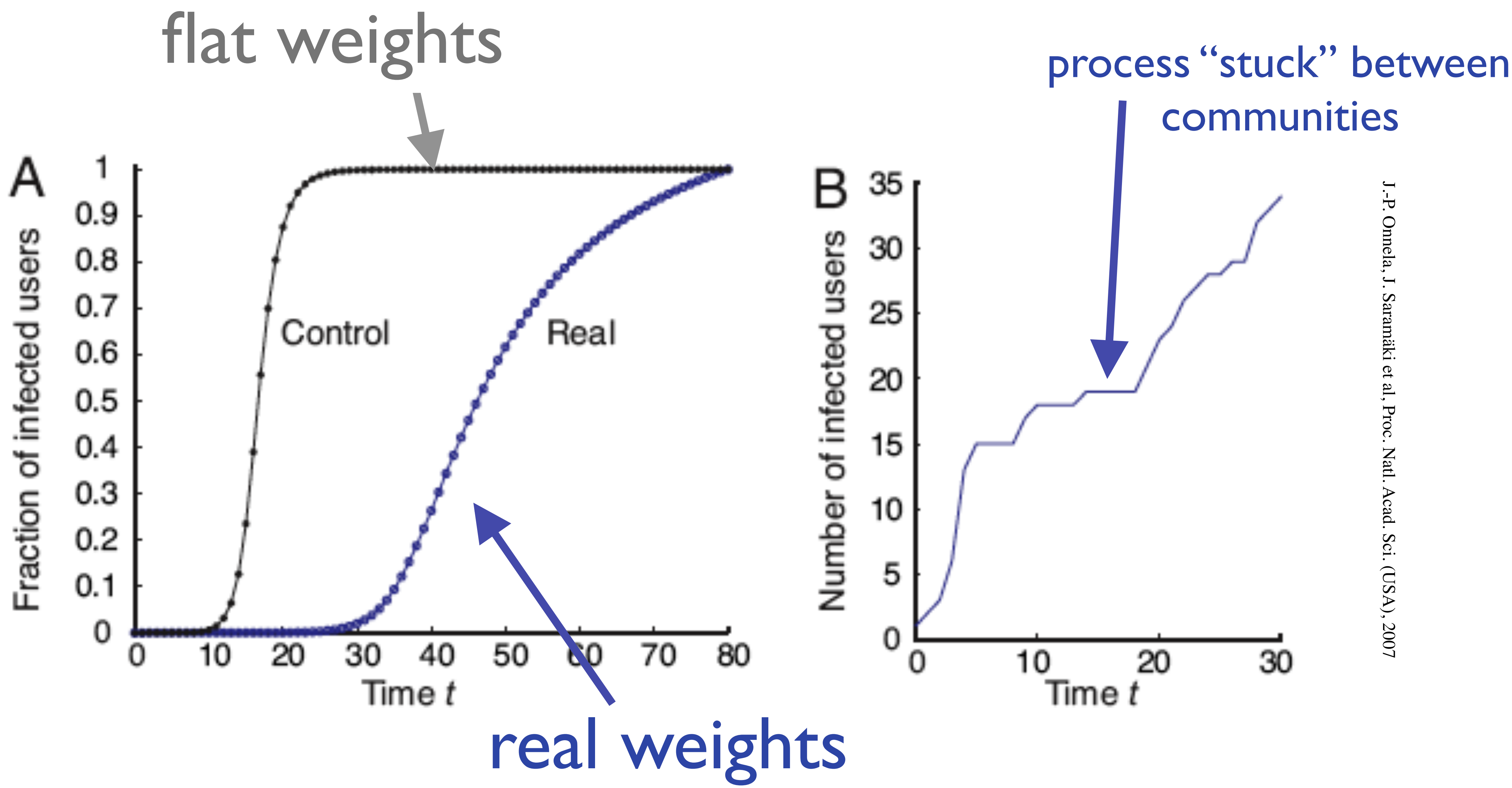
red: weak links removed first

black: strong links removed first

- **Giant component size R_{LCC}**
  - Def: % of nodes in the largest connected component
  - Weak link removal: network collapses when f~0.8
- **Susceptibility S**
  - Divergence indicates collapse of network
- **Average shortest path $<l>$ in L_{CC}**
  - Diverges at percolation transition
- **Clustering coefficient $<C>$**
  - Def: fraction of interconnected neighbours, averaged over network
  - Decreases faster on strong link removal

Weight-topology correlations & spreading dynamics

- simulated SI spreading on the aforementioned network
- infection probability per unit time proportional to weight



J.-P. Onnela, J. Saramäki et al, Proc. Natl. Acad. Sci. (USA), 2007