INTRODUCTION TO NETWORK SCIENCE

János Kertész
janos.kertesz@gmail.com

8. WEIGHTED, SIGNED AND DIRECTED NETWORKS

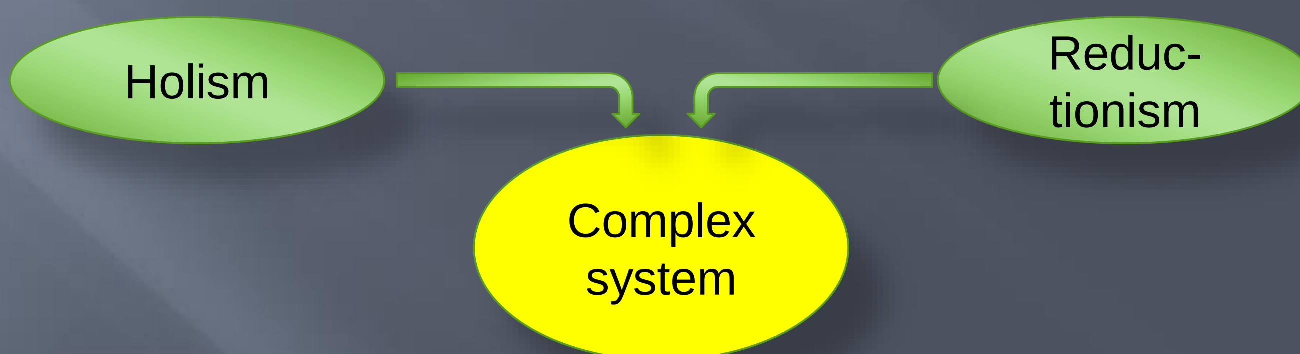
Why weighted networks?

Complex systems: Many interacting units such that the resulting behavior is more than a mere sum (brain, cell, society…)

Much is known about the *interactions* but *complex behavior* often still puzzling

Networks: Scaffold of complexity

Useful to concentrate on the carrying NW structure (nodes and links): Holistic approach with very general statements



Why weighted networks?

Step towards reductionism: Weighted NW-s
Interactions have different intensities:
Let us characterize them by a single real number:
weights on the edges

Weighted NW = fully connected NW with some
$w_{ij} = 0$.

First: No negative weights, $w_{ij} > 0$.

Negative weights: signed networks, e.g., negative sentiments towards a person. See later.

Why weighted networks?

Weights:

- Social relationship (intensity)
- Collaboration networks (joint papers)
- Mobile phone communication data
  (Call duration or frequency)
- Vehicular traffic network (throughput)
- IATA data on air transportation (passingers/year)
- Metabolic networks (chemical flux)
- Correlation based financial data (correlation coeff.)
- Topological role (betweenness)

etc.

Weighted network characteristics

We have to generalize the concepts and notions developed for binary networks.

Adjacency matrix A_{ij} → weight matrix w_{ij}

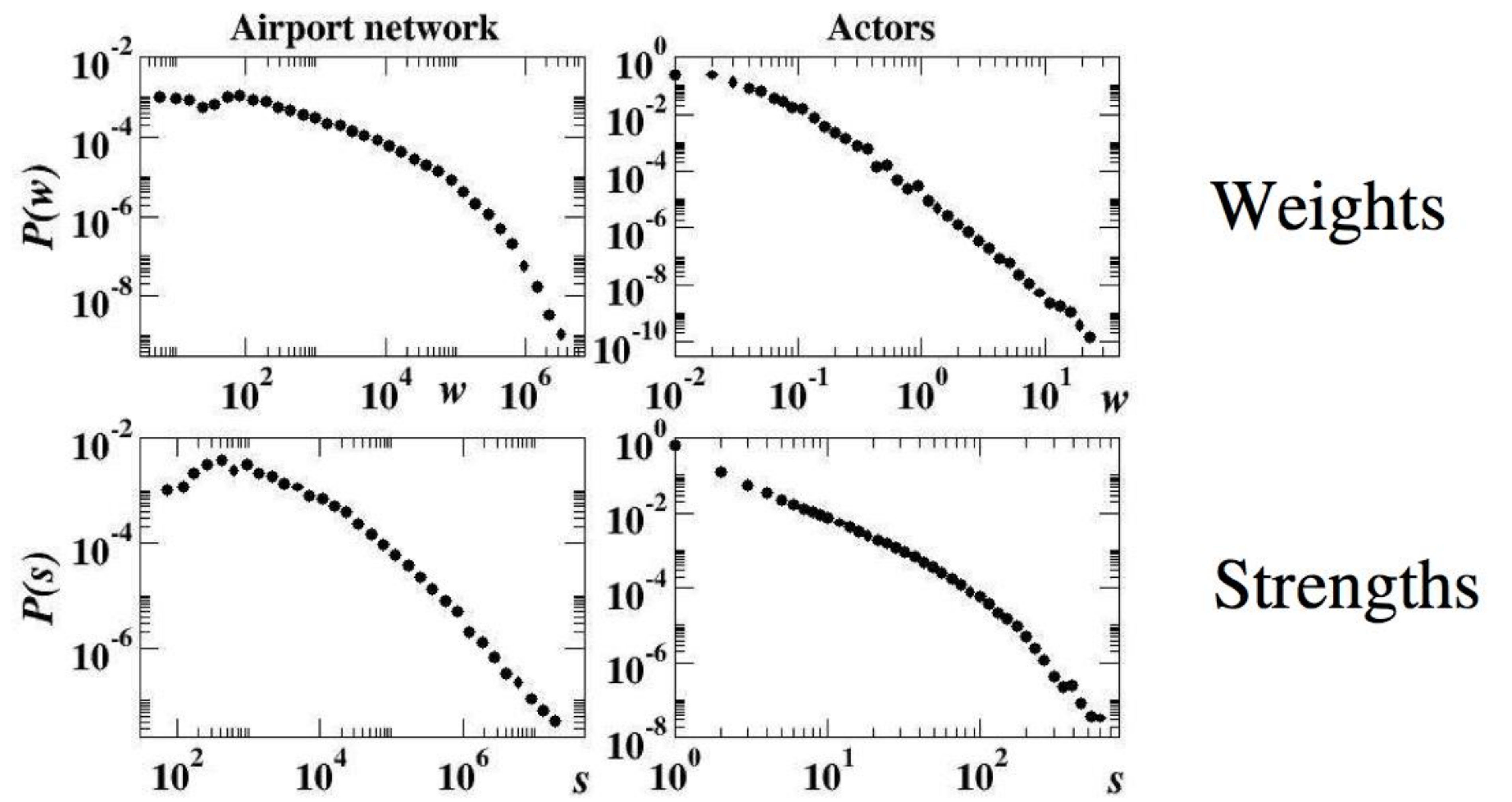
Degree of node i : k_i → strength s_i , e.g., traffic at a node

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

Degree distribution → strength distribution

(Of course, we can consider the degrees in a weighted network too.)

Weighted network characteristics



Broad, fat tailed distributions

Weighted network characteristics

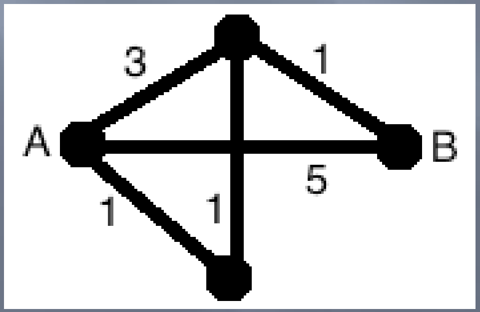
Length of a path $P(i \rightarrow j)$ → weight of path $P(i \rightarrow j)$

$$d_{ij} = \sum_{e_{mn} \in P(i \rightarrow j)} A_{mn}$$

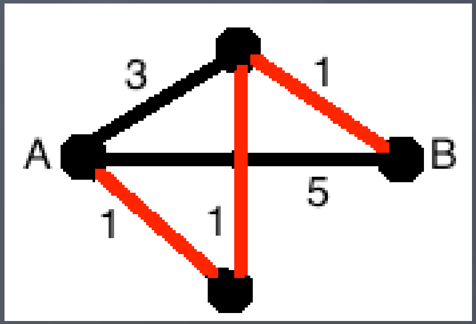
$$t_{ij} = \sum_{e_{mn} \in P(i \rightarrow j)} w_{mn}$$

If weight is considered as passage time, we can ask for the first passage time = min t_{ij} which is the counterpart of the distance in the weighted network.

Shortest path ≠ path with first passage time!



d_{AB}=1



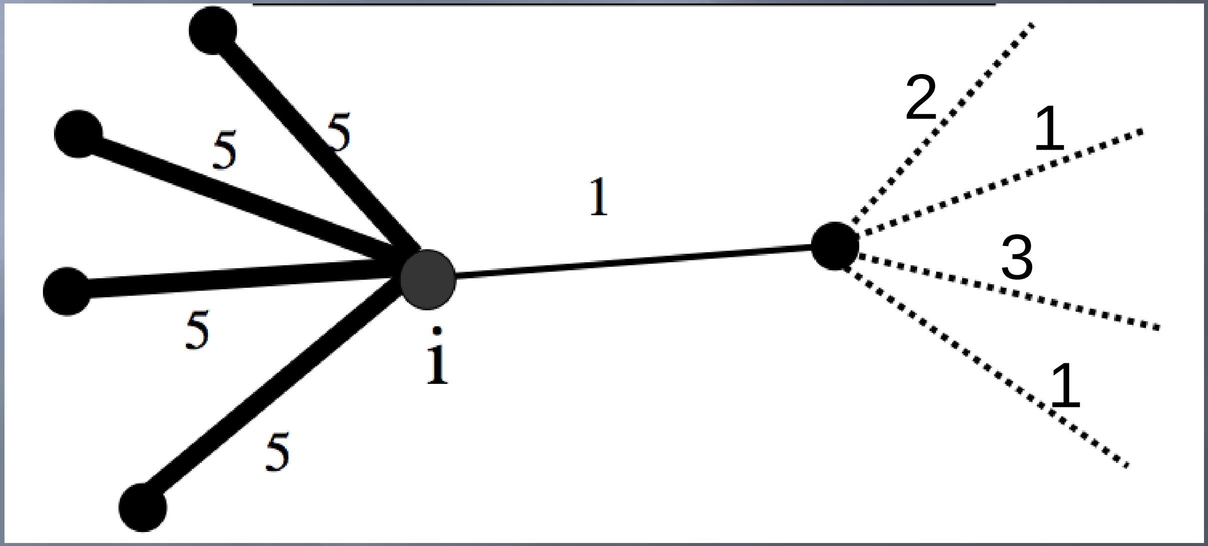
t_{ij}=3

Can be used to calculate betweenness centrality

Weighted network characteristics

Assortativity: average degree of neighbors

$$k_{nn}(i) = \frac{1}{k_i} \sum_{j \in nn(i)} k_j = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j$$



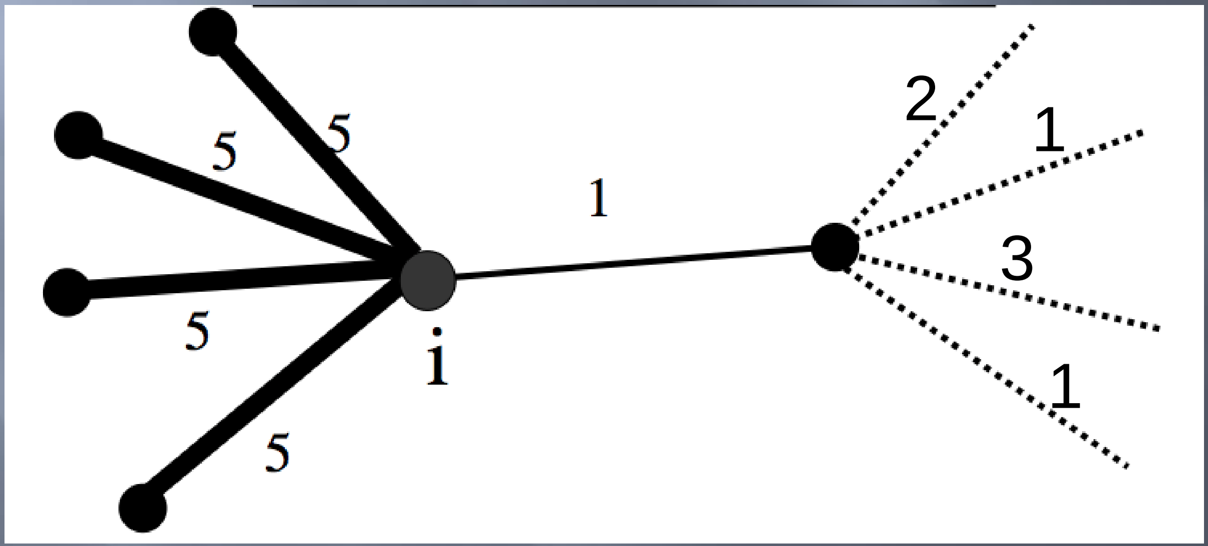
$k_{nn}(i) = 1.8$

Weighted network characteristics

Weighted assortativity:

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

If this is an increasing function of k, high degree nodes tend to be linked with high weight links.



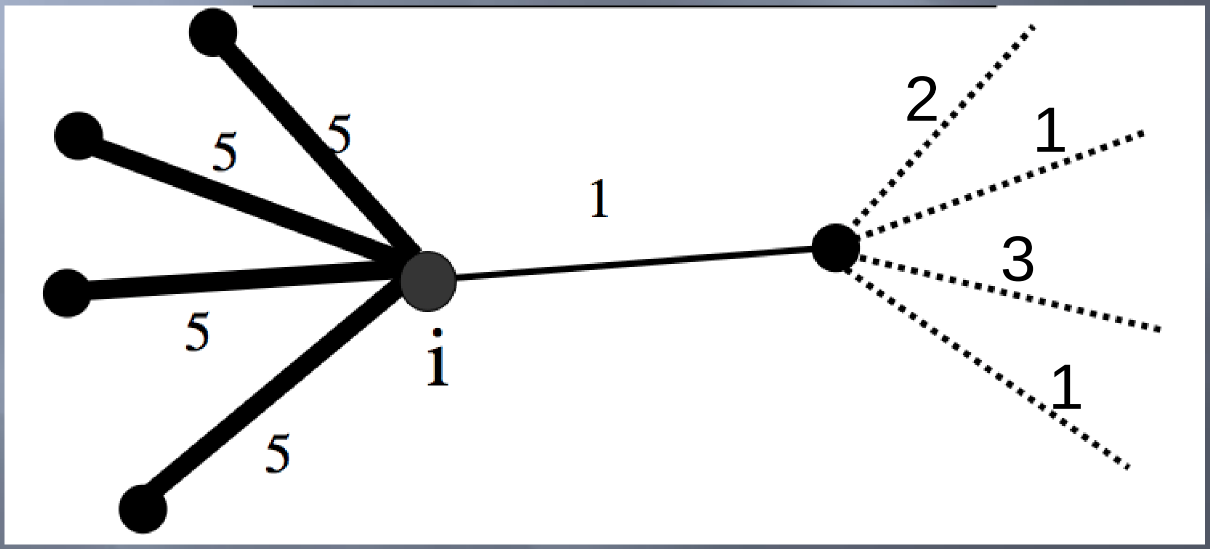
$$k_{nn}^{w}(i) = 1.19$$

Weighted network characteristics

Another possibility:

$$s_{nn}^{w}(i) = \frac{1}{s_i} \sum_{j=1}^{N} w_{ij} s_j$$

If this is an increasing function of *s*, high strength nodes tend to be linked with high weight links.



$$s_{nn}^{w}(i) = 2.29$$

Ambiguity in generalization!

Weighted network characteristics

New concepts and notions needed

Participation ratio

$$Y_2(i) = \sum_{j \in V(i)} \left[\frac{w_{ij}}{s_i} \right]^2 \begin{cases} 1/k_i \text{ if all weights equal} \\ \text{close to 1 if few weights dominate} \end{cases}$$

A. Barrat et al. 2005

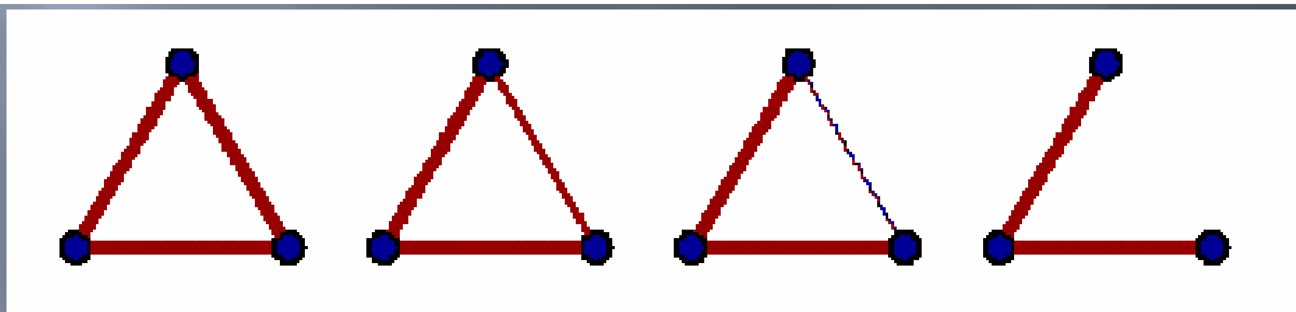
Weighted network characteristics

Subgraph characteristics: Intensity

Define *intensity* $I(g)$ of a particular subgraph g with vertices v_g and links ℓ_g as the *geometric mean* of the weights in g:

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

where $|\ell_g|$ is the number of links in ℓ_g



$I_{\triangle} > I_{\triangle} = 0$

Weighted network characteristics

Subgraph characteristics: Coherence

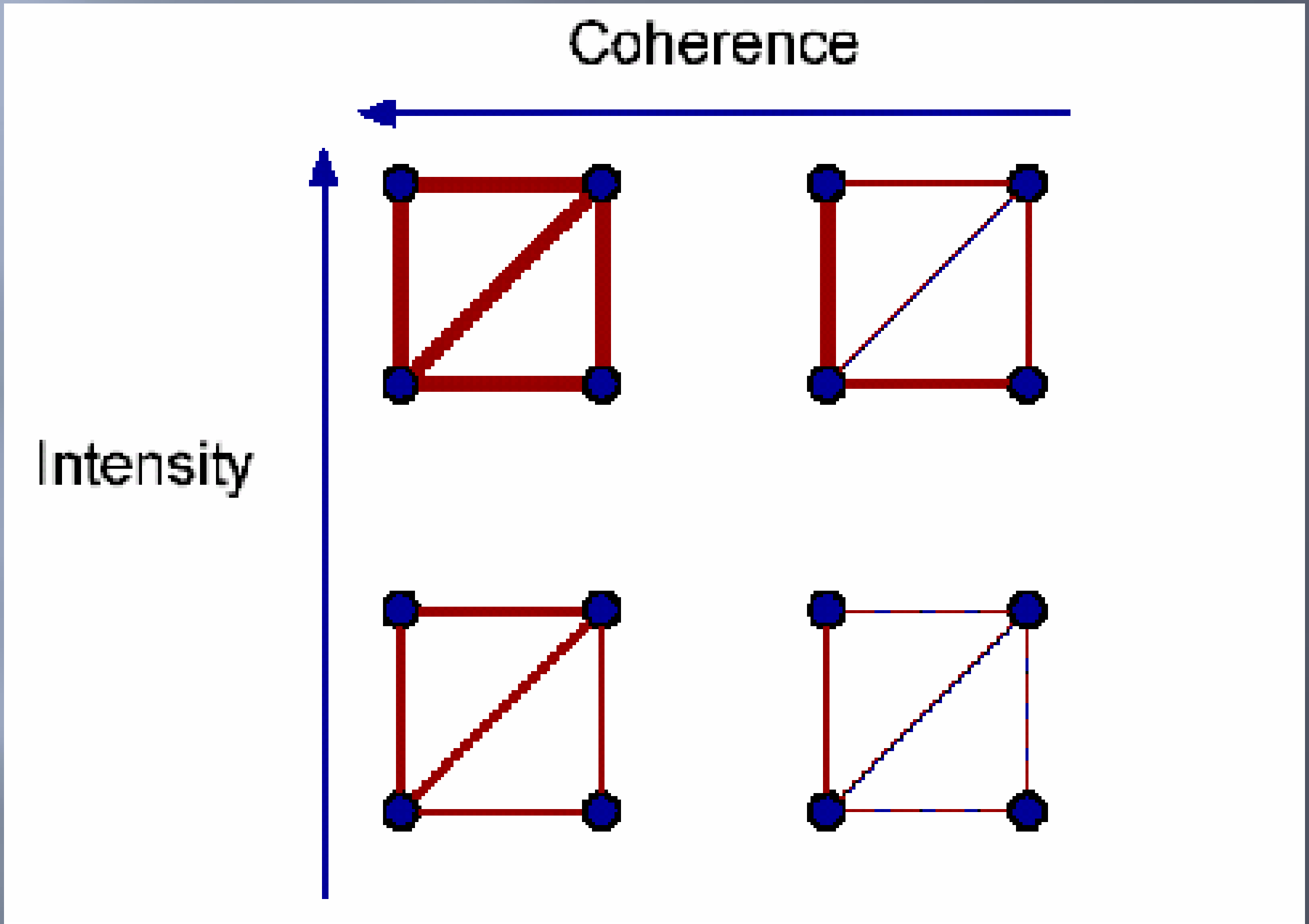
Subgraph intensity $I(g)$ may be low because one of the weights is very small, or because most or all of the weights are small

To distinguish between these two extremes, we introduce *coherence* $Q(g)$ for subgraph g as the ratio of the geometric to the arithmetic mean of the weights:

$$Q(g) = \frac{I(g)}{\frac{1}{|\ell_g|}\sum_{(ij)\in \ell_g} w_{ij}} \qquad (2)$$

Due to inequality btw arithmetic and geometric mean, $Q(g) \leq 1$and equality only holds for perfect homogeneity.

Weighted network characteristics



$0 \leq Q(g) \leq 1$, and the closer it is to 1, the more coherent are the interactions

If the $\boldsymbol{w_{ij}}$-s are normalized with the max $\boldsymbol{w}$ $0 \leq I(g) \leq 1$, too.

Onnela et al. 2007

Weighted network characteristics

Total and average quantities are naturally defined:

Total:

$$I_M = \sum_{g \in M} I(g)$$

$$Q_M = \sum_{g \in M} Q(g)$$

E.g. average intensity of subgraphs at node *i*:

$$\bar{I}_i = \frac{1}{n_i(M)} \sum_{g \in M} I(g)$$

Where $n_i(M)$ is the number of subgraphs of type *M* at *i*

Weighted motifs

Unweighted motif: Set of all topologically equivalent subgraphs in a NW

Motif *z* scores:

$$z_M = (N_M - \langle n_M \rangle)/\sigma_M$$

where N_M is the number of subgraphs in motif M in the empirical network, $\langle n_M \rangle$ and σ_M are its expectation and standard deviation, respectively, in the reference ensemble

Motifs with significantly high score are expected to play important functional role.

Weighted motifs

For weighted networks the number of occurrence is replaced by total motif intensity

Statistical significance is now measured using the *motif intensity score*

$$\tilde{z}_M = (I_M - \langle i_M \rangle)/(\langle i_M^2 \rangle - \langle i_M \rangle^2)^{1/2},$$

where i_M is the total intensity of motif M in one realisation of the reference system

Motifs showing statistically significant deviation from the reference system are called high/low intensity motifs

Weighted motifs

What should be chosen as null model? (strongly influences the result!)
Depends on what we are intested in.

- For unweighted problems: Correlations->
Null model: P(k) fixed, no correlations
- Weighted: Relation between weights and topology
Null model: Fixed topology, randomized weights

For more general null models see Serrano et al. cond-mat/0609029

Weighted motifs

An example:

Cellular metabolism can be represented as a directed network of intracellular molecular interactions

The network consists of nodes (X_i, Y_j), which represent the chemicals and they are linked if connected by a metabolic reaction of the bacterium *Escherichia coli* grown in glucose

The chemical flux f of reactions provides an overall measure of their relative activity

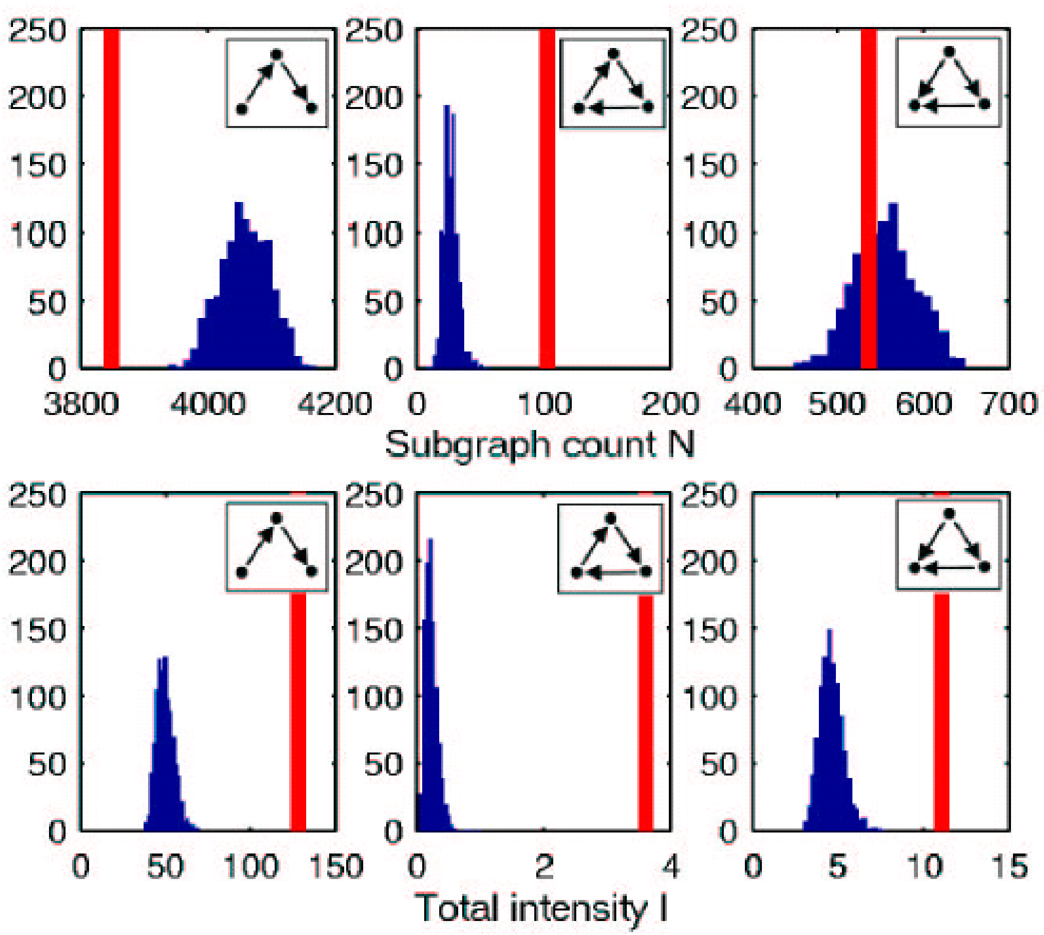
Biochemical reaction: $x_1 X_1 + \cdots x_n X_n \rightarrow y_1 Y_1 + \cdots y_m Y_m$

Define the weights as $w_{ij} = (y_j / x_i) f$, reflecting the rate at which X_i is converted into Y_j

Weighted motifs

TOP: Subgraph counts (unweighted); $z = -5.4, 12.8, -0.5$

BOTTOM: Motif intensity scores (weighted); $\tilde{z} = 14.8, 33.8, 9.0$



Even the sign changes!

Onnela et al. 2006

Weighted clustering coefficient

Definition of unweighted clustering coefficient at node *i:*

$$C_i = \frac{2n_{\text{D}}}{k_i(k_i - 1)}$$

where k_i and t_i are the degree and the number of triangles at that node.

Density of triangles.

Much is known, e.g. often C(k) ~ $1/k$

Weighted clustering coefficient

How to generalize to the weighted case?
- for $w > 0$, $w \to 1$ the weighted $\tilde{C}$ -> unweighted C
- $\tilde{C} \in [0,1]$
- A triangle's contribution is 0 if any of its w_{ij}-s is 0.

Suggestion:

$$\tilde{C}_i = \frac{2}{k_i(k_i-1)} \sum_{j,k} (\tilde{w}_{ij}\tilde{w}_{jk}\tilde{w}_{ki})^{1/3}$$

Advantage: $\tilde{C}$ factorizes:

$$\tilde{C}_i = \overline{I}_i C_i$$

where

I_i is the average intensity of the triangles at *i*

Weights are normalized

$$\tilde{w}_{ij} = \frac{w_{ij}}{\max w}$$

Weighted clustering coefficient

This was not the only, even not the first suggestion:
Barrat et al 2004:

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

Onnela et al 2005:

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

Zhang & Horvath 2005:

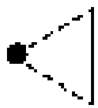
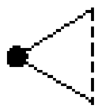
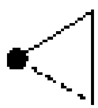
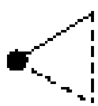
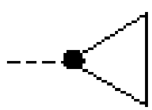
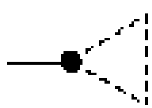
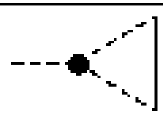
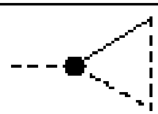
$$\tilde{C}_{i,Z} = \frac{\sum_{j,k} \hat{w}_{ij} \hat{w}_{jk} \hat{w}_{ik}}{\sum_{j \neq k} \hat{w}_{ij} \hat{w}_{ik}}$$

Weighted clustering coefficient

Feature	$\tilde{C}_B$	$\tilde{C}_O$	$\tilde{C}_Z$
1) $\tilde{C} = \mathrm{C}$ when weights become binary	X	X	X
2) $\tilde{C} \in [0, 1]$	X	X	X
3) Uses global $max(w)$ in normalization		X	X
4) Takes into account weights of all edges in triangles		X	
5) Invariant to weight permutation for one triangle		X	
6) Takes into account weights of edges not participating in any triangle	X		X

Saramaki et al. 2007

Weighted clustering coefficient

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

—— : $w = 1$, - - - : $w = \epsilon \ll 1$

All of them have got problems

B: weak triangles full in

O: weights of links not in triangles ignored

Z: inconsistency

Weighted clustering coefficient

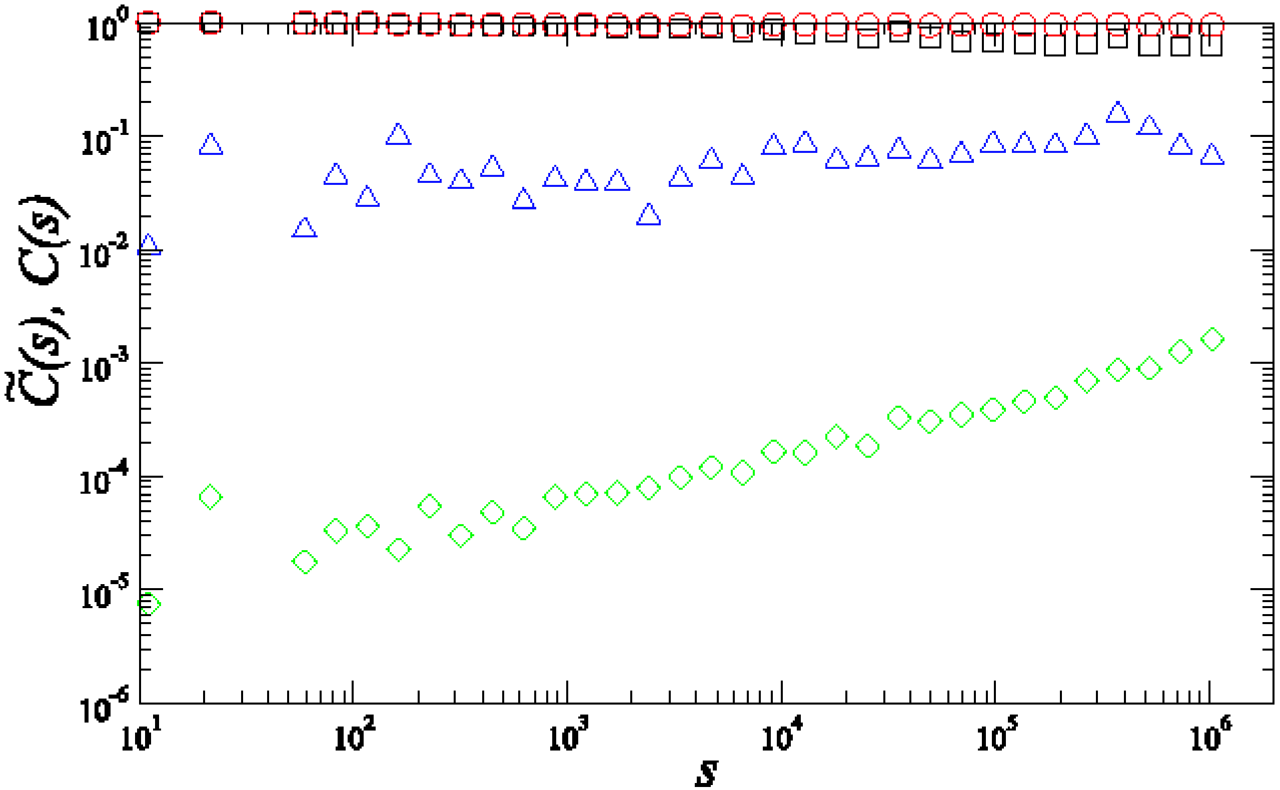
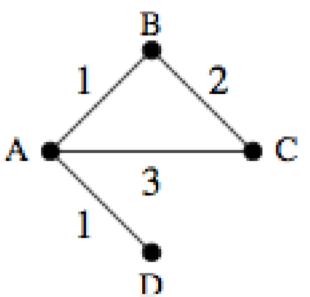


FIG. 2: Clustering coefficients computed for the international trade network (ITN) as function of vertex strength s: Unweighted $C(\square)$ and weighted $\tilde{C}_B$ ($\circ$), $\tilde{C}_O$ ($\diamond$), and $\tilde{C}_Z$ ($\triangle$).

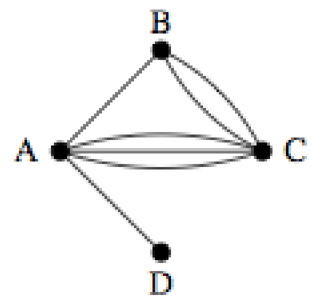
Saramaki et al. 2007

Weighted networks and multigraphs

If the weights are natural numbers, one can consider them as multiple links, i.e., map the weighted graph to a multigraph.



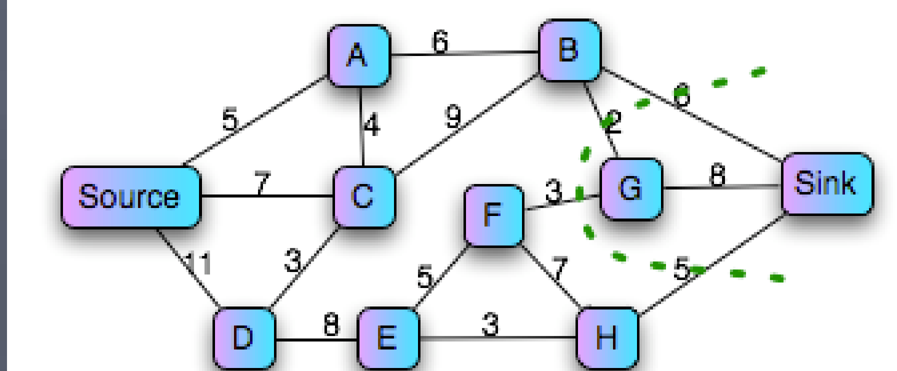
$$\equiv \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 3 & 1 \\ 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \end{array}$$



$$\equiv \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 3 & 1 \\ 1 & 0 & 2 & 0 \\ 3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \end{array}$$

Newman 2004

Consequence: Simple proof of max flow/min cut theorem: The maximum flow between two nodes is given by the weight of minimum edge cut set.



True for real weights too.

Community structure in weighted networks

Some methods are based on weights (hierarchical clustering)

We have to generalize the other methods.

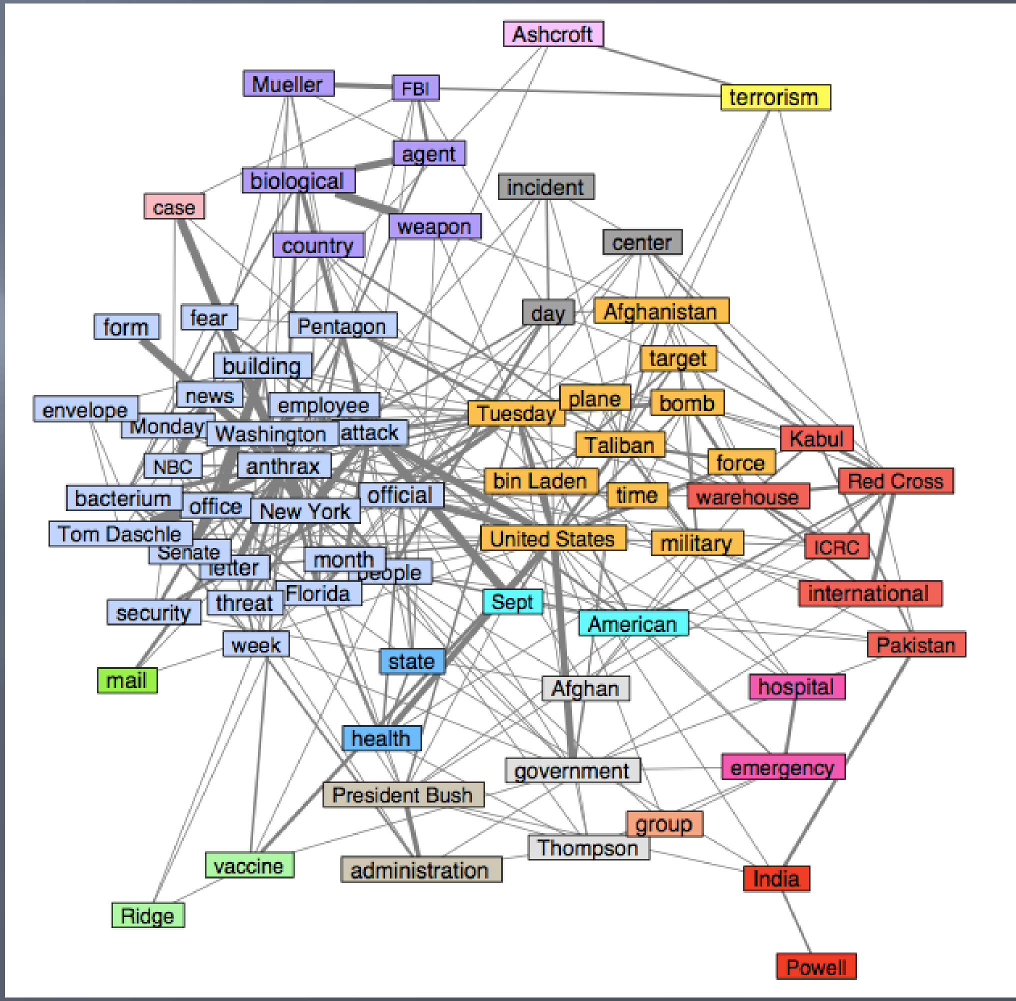
Modularity:

$$Q = \frac{1}{2L}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2L}\right)\delta\left(C(i),C(j)\right)$$

$$Q = \frac{1}{2S}\sum_{i,j}\left(w_{ij} - \frac{s_i s_j}{2S}\right)\delta\left(C(i),C(j)\right)$$ with $$S = \frac{1}{2}\sum_i s_i$$

Community structure in weighted networks

Weights can be considered as similarity measures from which a dendrogram can be constructed. Modularity tells, where is the optimal cut of the dendrogram.



Analysis of Reuthers newswire most frequent words.

Newman 2004

Community structure in weighted networks

Thresholding: A trivial way to map a weighted network into a unweighted one is to ignore the links having weights smaller than a threshold.

Then all unweighted methods can be applied…

Weak community for weighted NW-s: Total in-weight exceeds total out-weight.

Local methods based on this definition can be immediately applied.

Community structure in weighted networks

Weighted clique percolation communities:

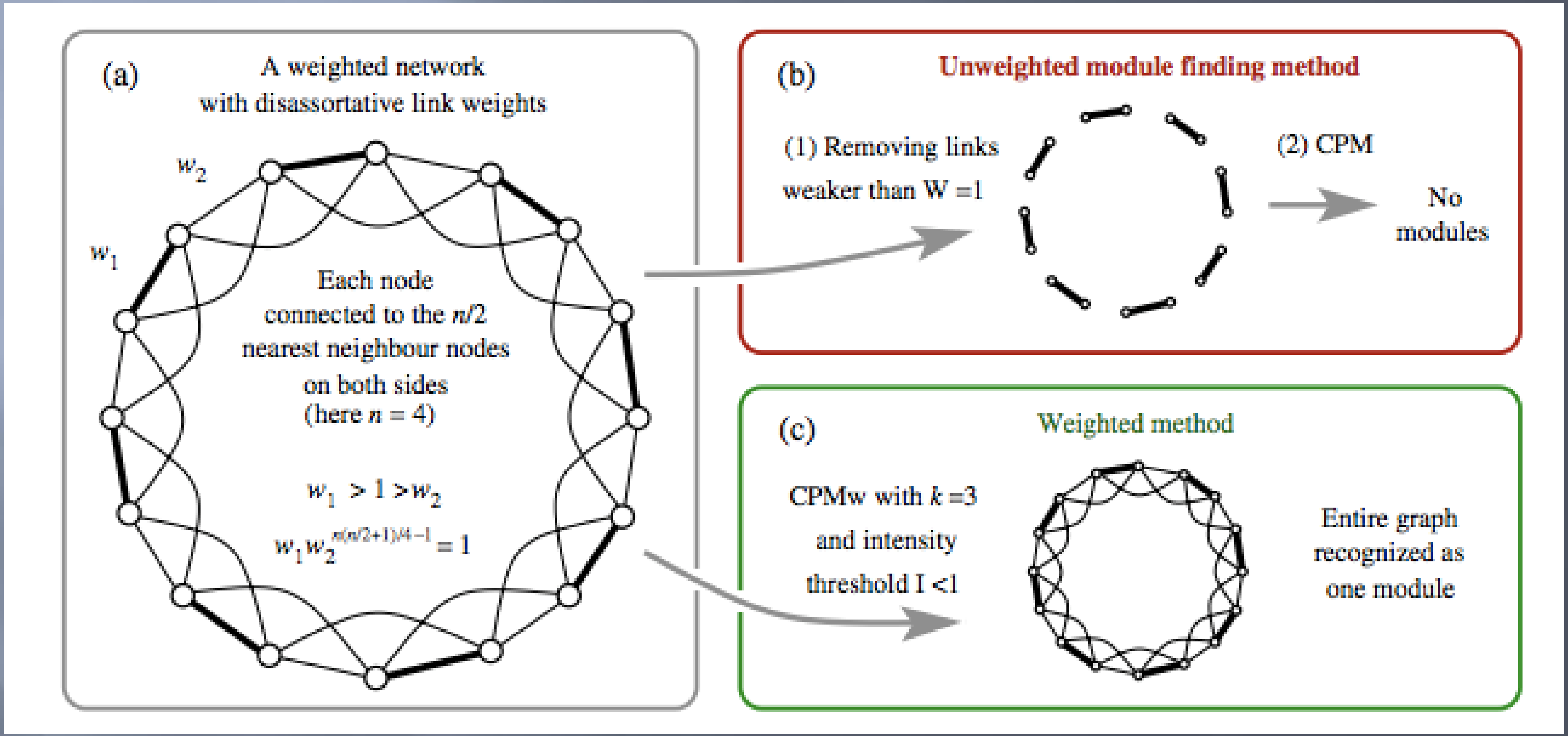
Links in the cliques may have very different weights.
By thresholding the cliques are easily destroyed, esp. for disassortative networks.

Use thresholding for the intensity of the cliques!

$$I(C) = \left(\prod_{i<j} w_{ij} \right)^{2/(k(k-1))}$$

for a *k*-clique

Community structure in weighted networks



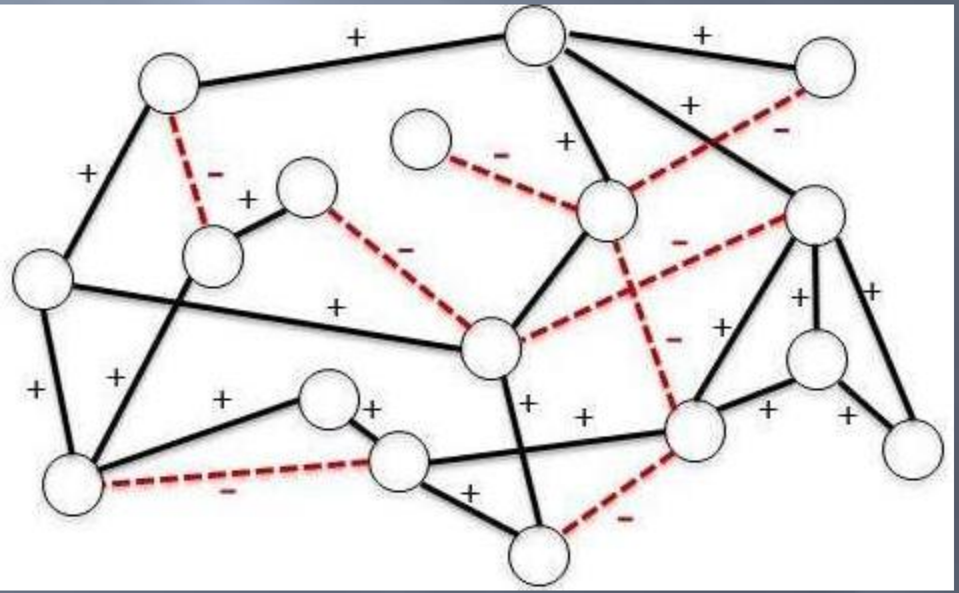
Palla et al. 2007

Signed networks

What if weights can be both positive and negative?

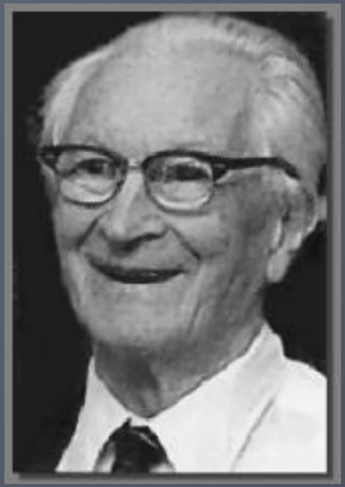
Examples:

- Social networks: love ←→ hate
- Political science: ally ←→ enemy
- Economy: cooperator ←→ competitor
- Biology: stimulator ←→ inhibitor



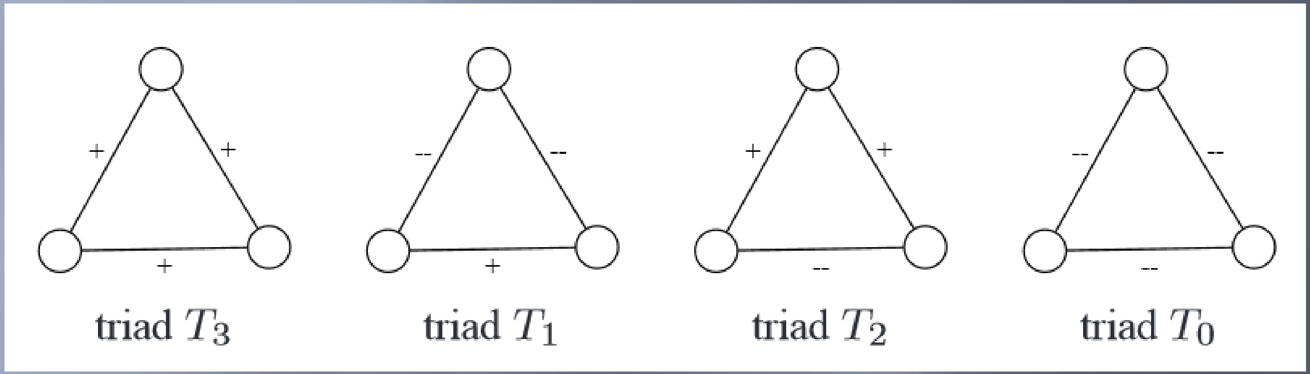
http://users.cecs.anu.edu.au/~u5549252/research_demo.html

Signed networks



F. Heider

Fritz Heider's theory of structural balance:



Balanced: T3 and T1; imbalanced: T2 and T0

Any plaquette with an odd number of negative bonds is "frustrated":

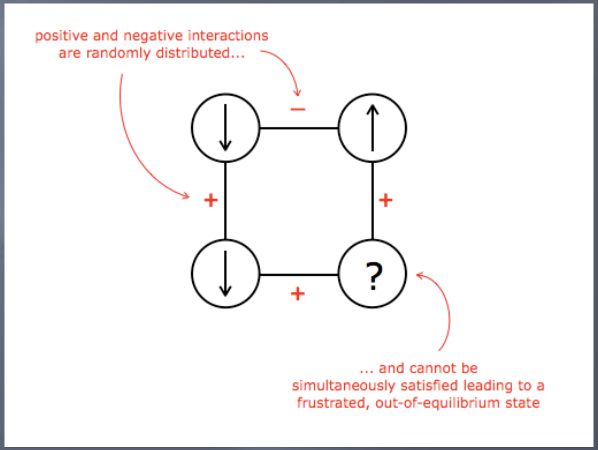
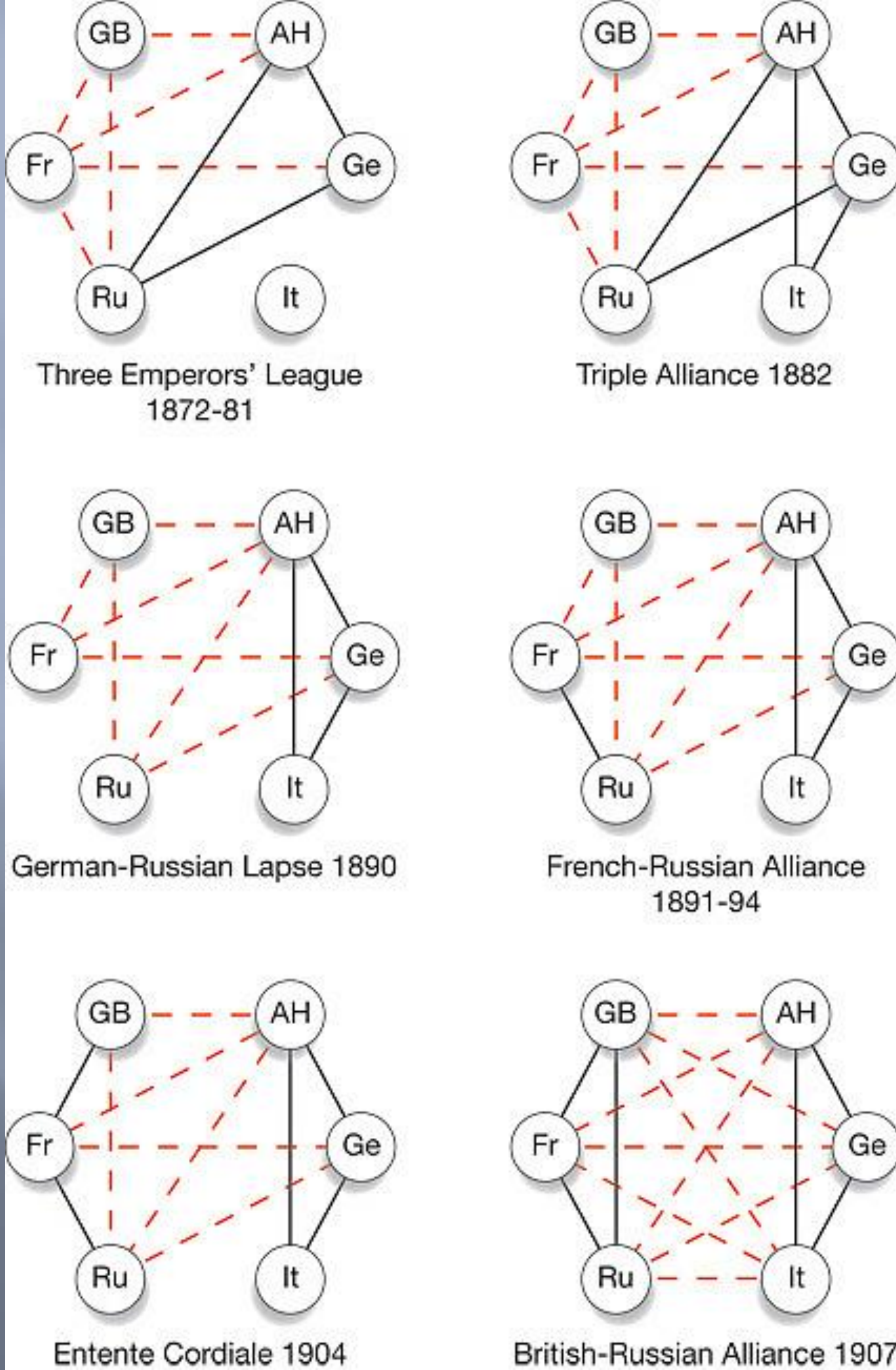


Figure taken from physics: spin glass theory



Steps to form the system of allies before WWI in the light of balance theory

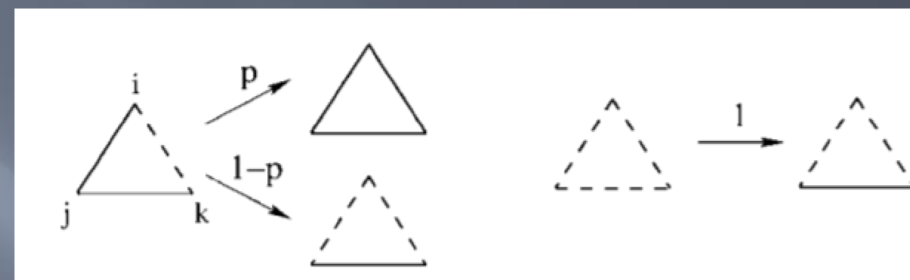
Dyn. models: $\mathcal{H} = -\Sigma_{ijk} s_{ij} s_{jk} s_{ki}$

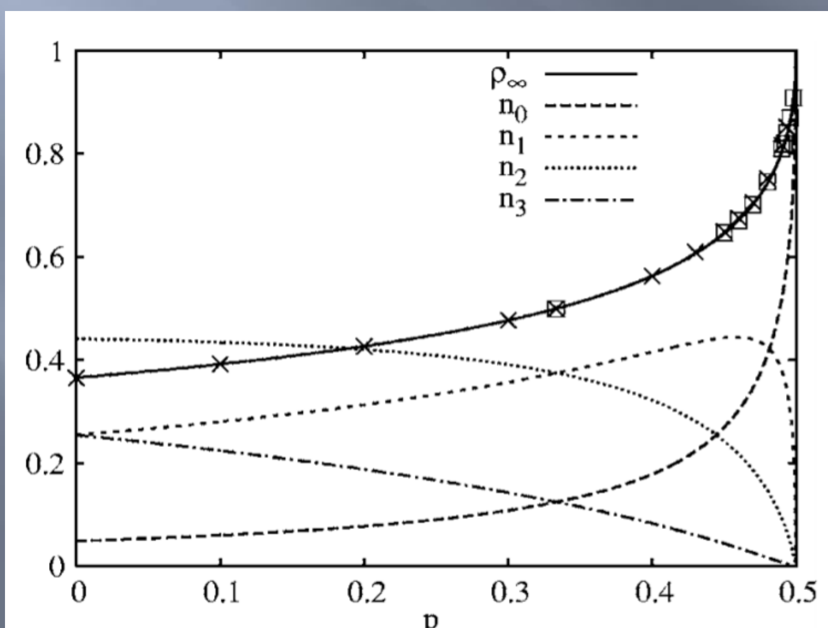
Constrained triad dyn (CTD): flip link sign if advantageous

~ logN steps to balance

Local triad dynamics (LTD)

Fast convergence: log N





ρ_0 density of friendly links
n_i density of Δ-s with
i unfriendliy links
Phase transition at $p = 1/2$

Adam, Karpivsky, Redner 2005

Signed networks

Fritz Heider's theory of structural balance:
Balanced triangles are more prevalent than imbalanced ones. The theory describes mechanism that remove imbalance from the network by rewiring.
James Davis: Weak balance theory: Only T2 is forbidden.
However, signed links in social networks may carry different meaning than love and hate. E.g., they may indicate (subjective) status. A → B is pos if A thinks that B has higher status than A and neg. if lower. This is a signed directed network.

Balance theory and status theory may lead to opposite conclusions:

A→B B→C ⇒ C→A (B.T.)

A→B B→C ⇒ C→A (S.T.)

Leskovec et al. 2010

Signed networks

By counting the triads one can decide, which of the theories is adequate for a dataset.
Three datasets studied:
- Epinions (product review)
- Slashdot (user-submitted and evaluated news stories about science and technology-related topics)
- Wikipedia voting

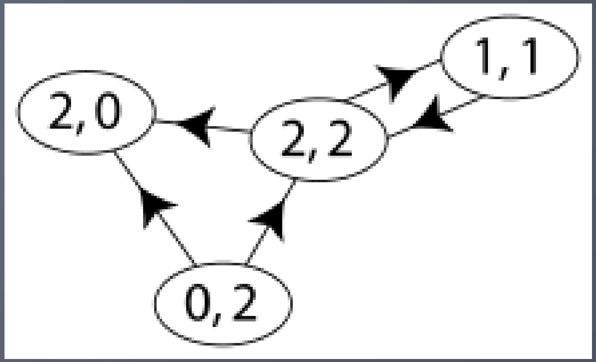
Symbol	Meaning
T_i	Signed triad, also the number of triads of type T_i
Δ	Total number of triads in the network
p	Fraction of positive edges in the network
$p(T_i)$	Fraction of triads T_i, $p(T_i) = T_i/\Delta$
$p_0(T_i)$	A priori prob. of T_i (based on sign distribution)
$E[T_i]$	Expected number of triads T_i, $E[T_i] = p_0(T_i)\Delta$
$s(T_i)$	Surprise, $s(T_i) = (T_i - E[T_i])/\sqrt{\Delta p_0(T_i)(1 - p_0(T_i))}$

Triad T_i		$\|T_i\|$	$p(T_i)$	$p_0(T_i)$	$s(T_i)$
Epinions					
T_3	$+++$	11,640,257	0.870	0.621	1881.1
T_1	$+--$	947,855	0.071	0.055	249.4
T_2	$++-$	698,023	0.052	0.321	-2104.8
T_0	$---$	89,272	0.007	0.003	227.5
Slashdot					
T_3	$+++$	1,266,646	0.840	0.464	926.5
T_1	$+--$	109,303	0.072	0.119	-175.2
T_2	$++-$	115,884	0.077	0.406	-823.5
T_0	$---$	16,272	0.011	0.012	-8.7
Wikipedia					
T_3	$+++$	555,300	0.702	0.489	379.6
T_1	$+--$	163,328	0.207	0.106	289.1
T_2	$++-$	63,425	0.080	0.395	-572.6
T_0	$---$	8,479	0.011	0.010	10.8

More Davis than Heider
but! directed network!

Directed networks

In-degrees and out-degrees

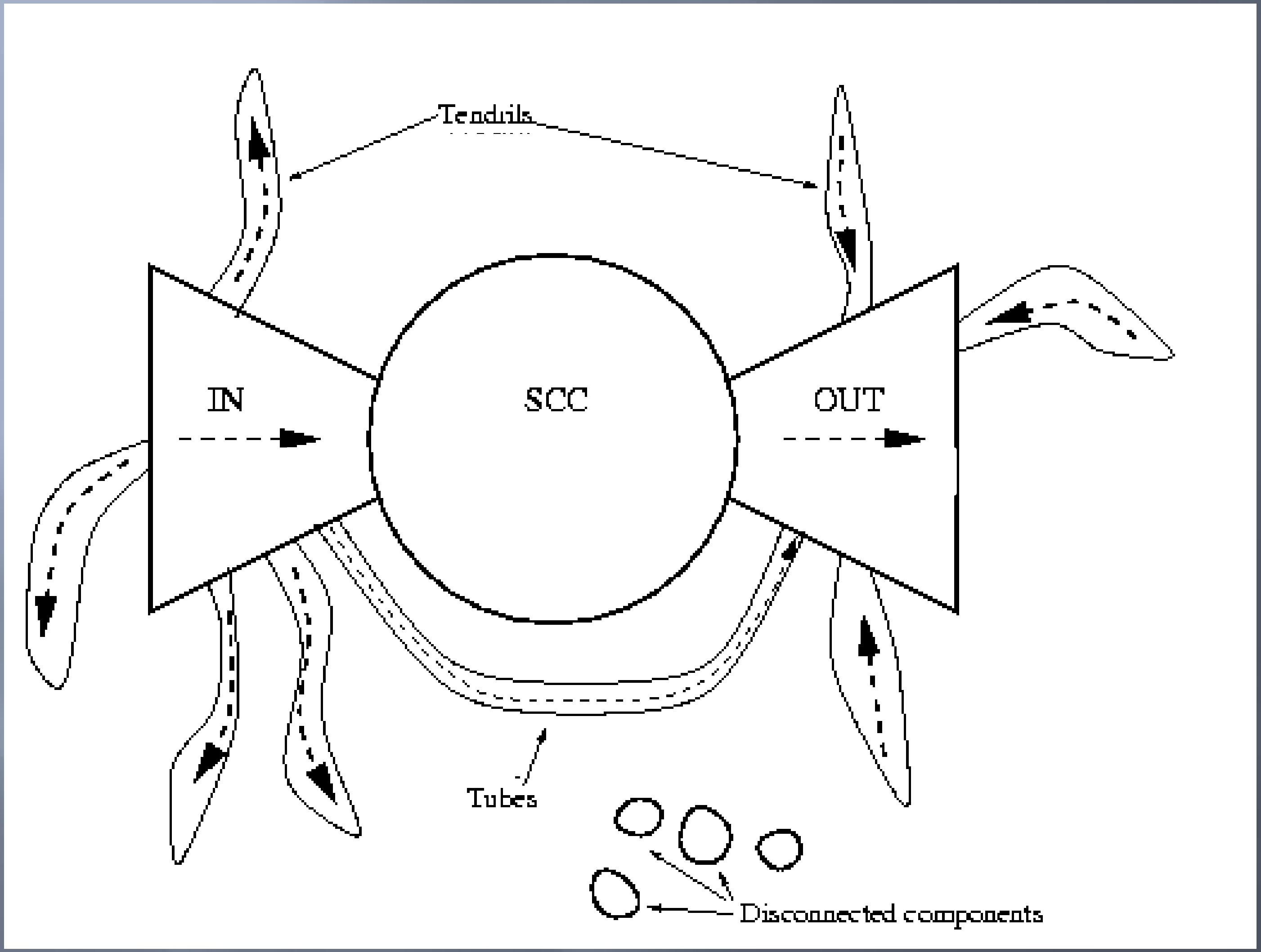


(Not to be confused with those introduced in the context of communities.)

What are the components?

Not trivial, no transitivity

Directed networks



bow-tie

SCC: strongly connected component

Communities in directed networks

It depends! E.g., on WWW people interested in a topic may be counted to a community but they are not reachable for each other via URL links.
If mutual influence is asked for then there must be a path in both directions.

The community definition depends on what we are interested in, and the algorithm has to be adjusted accordingly!

Communities in directed networks

Modularity: First create a directed equivalent of the configuration model (similar to the bipartite case): Given the $\{k_i^{in}\}$ and $\{k_i^{out}\}$ sequences, *out* stubs have to be paired with *in* stubs. Condition: $\sum k_i^{in} = \sum k_i^{out}$. The prob. that a node with k_j^{out} degrees is connected to a node with k_i^{in} degrees is $\frac{k_j^{out} k_i^{in}}{L}$. Thus the directed modularity is:

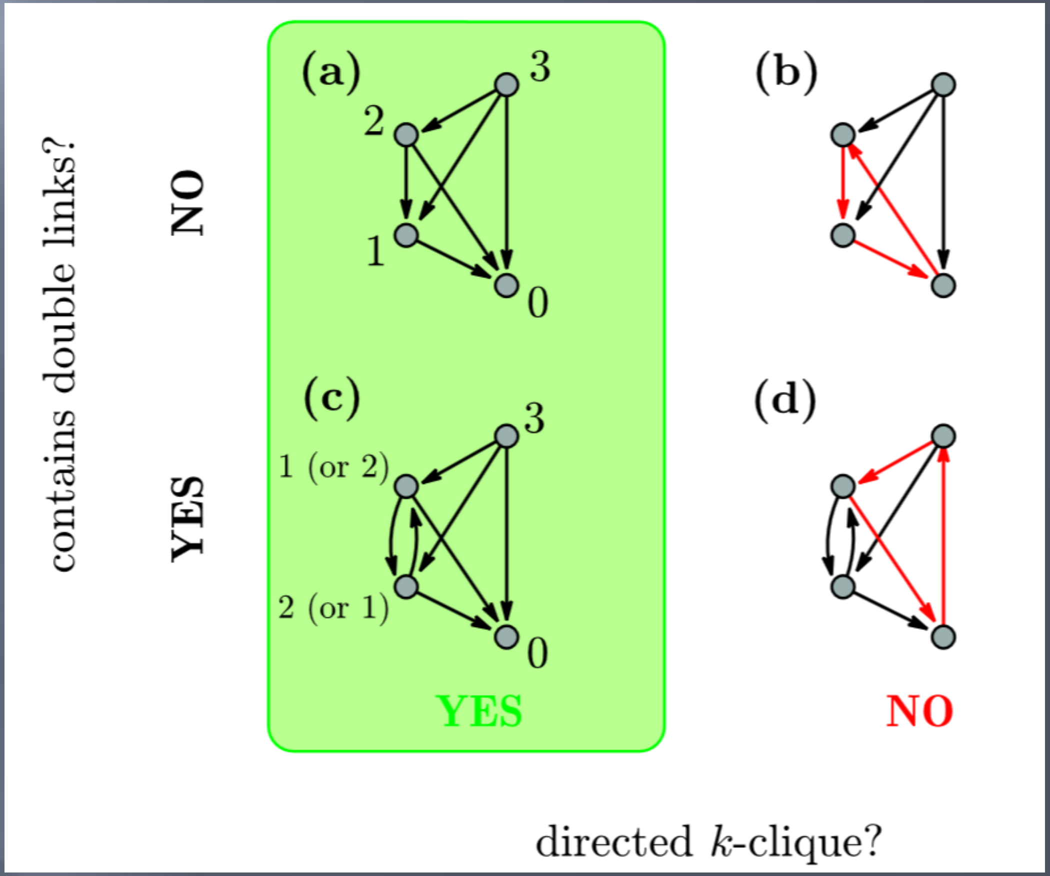
$$Q = \frac{1}{L} \sum_{i,j} \left(A_{ij} - \frac{k_i^{in} k_j^{out}}{L} \right) \delta(C_i, C_j)$$

Communities in directed networks

k-clique percolation method: One has to define the directed cliques:

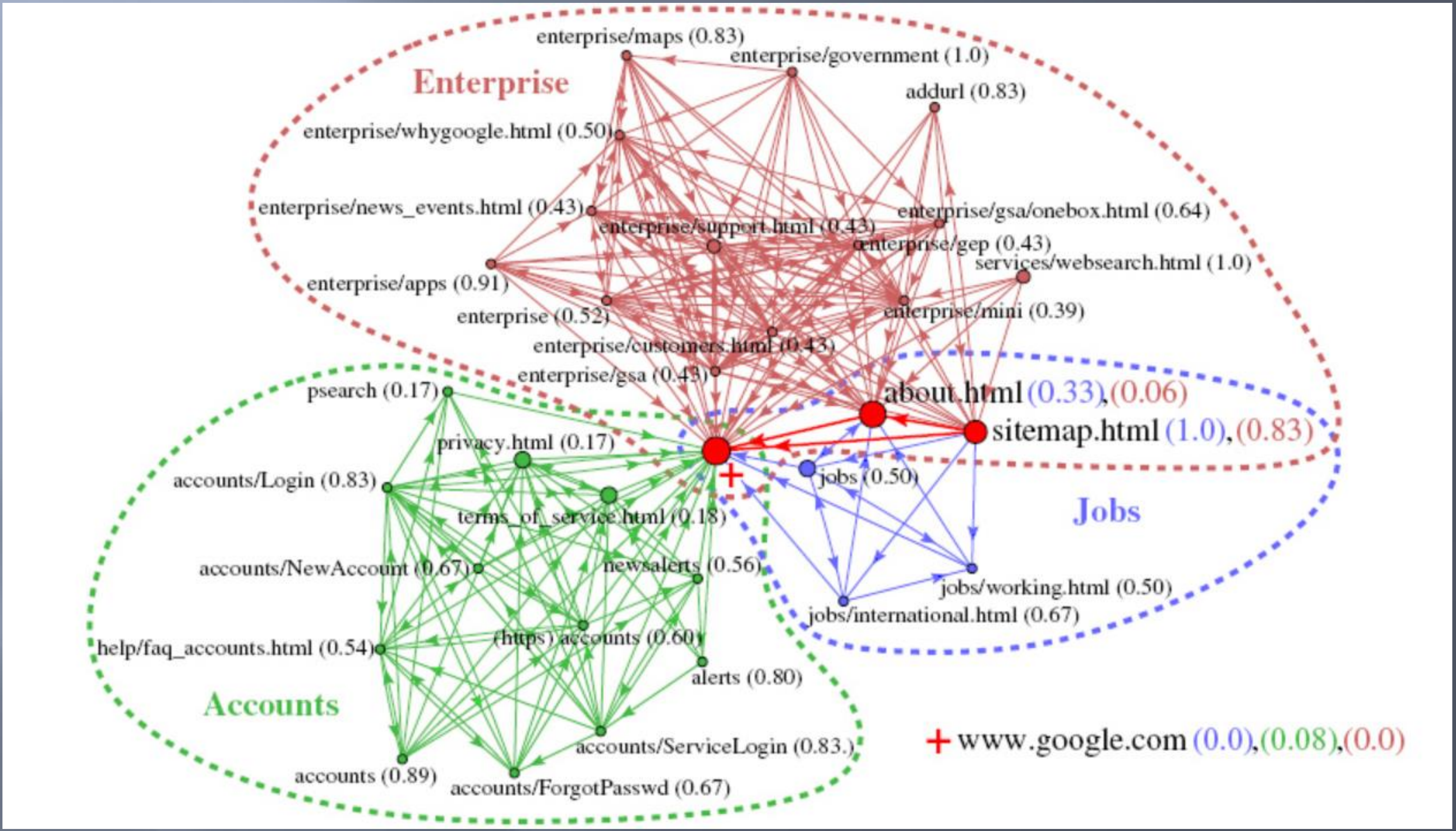
One possibility:
Flow from high rank (larger out degree) nodes to lower ones

For bidirectional links one is ignored



Communities in directed networks

Google pages



Pages within ≤ 3 steps from google.com

Palla et al. 2007

Homework

Analyze the weighted network of the co-appearances of the characters in Les Miserables (downloadable from http://www-personal.umich.edu/~mejn/netdata/)

Calculated the weight distribution and the distribution of the intensities and the coherences of triangles.