INTRODUCTION TO NETWORK SCIENCE

János Kertész
janos.kertesz@gmail.com

7. MOTIFS AND MODULES

Structure and function

How do complex systems function?

Complex networks should reflect the function of the systems

How is the topology related to the function?

Important task: Identifying units which are topologically closely related – they are expected to have functional role.

Microscopic
Mesoscopic scales
Macroscopic

Structure and function

Microscopic
Mesoscopic structures
Macroscopic

Microscopic: node properties + interaction (dyad)
Very systems specific

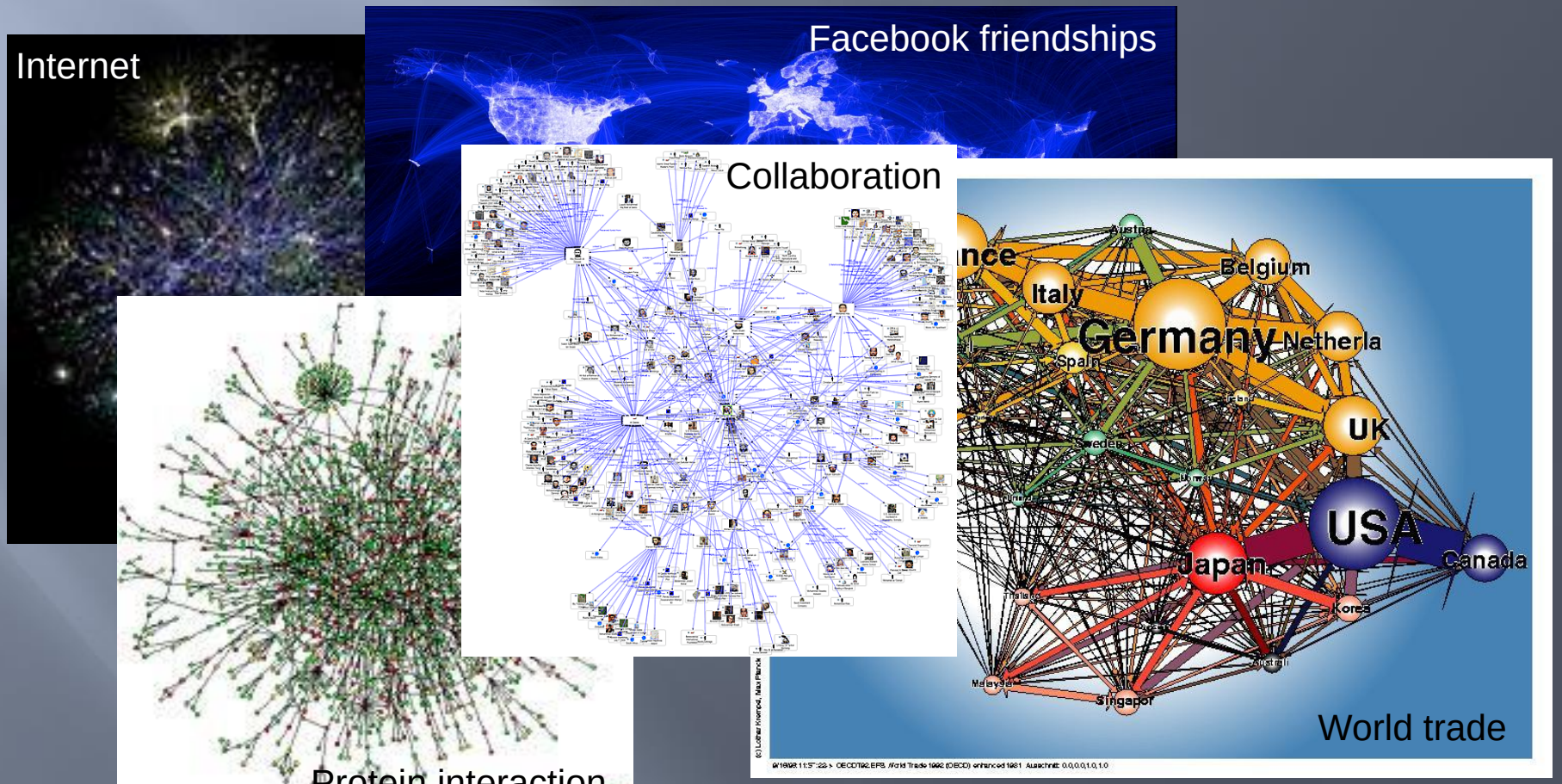
Macroscopic: the network as a whole. Global characterization, qualitative universality, robustness etc.

Mesoscopic: Structures on intermediate scales.

Mesoscale structures: System specific + universal features

Structure and function

So far we have mainly focused on overall structures



Structure and function

Mesoscopic structures: subgraphs of the original, usually large graph.

Subgraph of $G=\{V,E\}$: $G'=\{V',E'\}$ with $V'\subseteq V; E'\subset E$

such that $i, j \in V'$ for $\forall e_{ij} \in E'$

For mesoscopic structures we assume that *N'* << *N*

Two approaches:

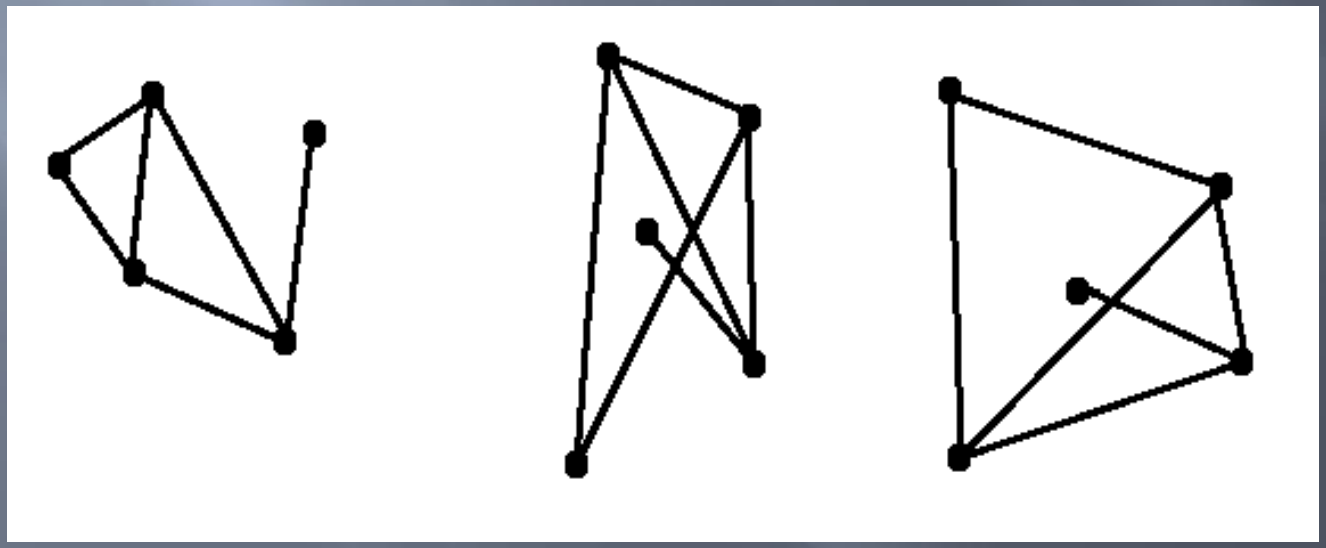
i) Define a type of subgraph, identify topologically equivalent occurrences and check how significant this class of subgraphs is: motifs

ii) Consider the particular subgraphs after identifying them: egocentric networks, communities (often identification is the challenge)

Network motifs

Idea: If a type of subgraph (e.g., a triangle) occurs significantly often in a large network, we can expect that such subgraphs have an important role in the functioning.

Two graphs are topologically equivalent or isomorphic if there is an appropriate numbering of nodes such that the adjacency matrices become identical.



Necessary:
N and degrees the same
But not sufficient

Network motifs

Motif: set topologically equivalent subgraphs in a network
Cardinality of a set: number of elements

If the cardinality of a motif is significantly high, we expect that the motif has an important role in the function of the complex system the network is mapped out from.

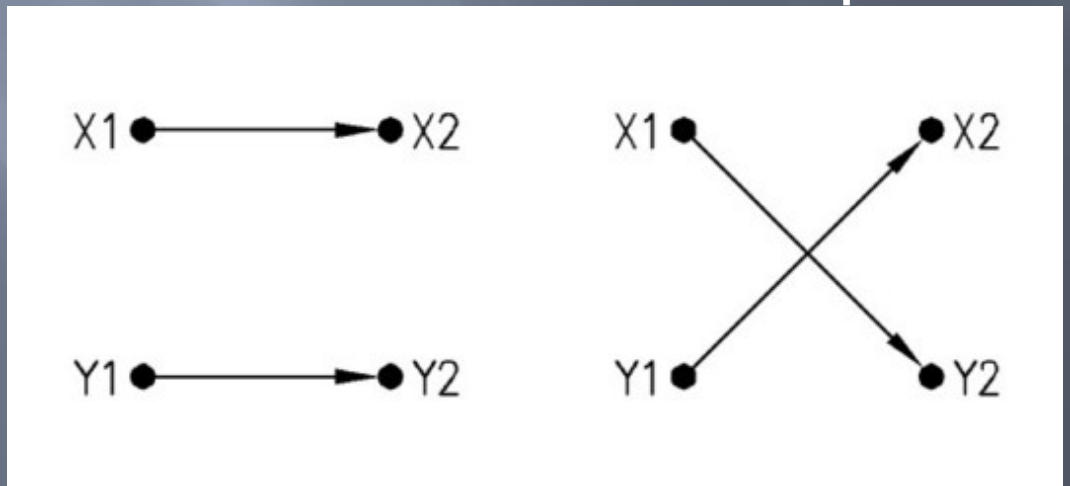
What is “significantly high”?
We have to compare to a reference system.

Significantly high means that we have a null hypothesis that a given cardinality of a motif stems from a reference system. If we can exclude this hypothesis then there is an additional origin for the effect – possibly the function of the system.

Network motifs

Usually we take a random network as a reference system assuming that correlations are caused by the function. Simplest: ER (*N, L* fixed). This is too simple and far from the global, universal observations (broad degree distribution). Thus the common reference system is the configuration model.

The configuration model can be considered as a result of a randomization process: link switching



Randomizing and preserving degrees

Network motifs

We need a measure for the importance of a motif.

The null model is an ensemble, with means and a variances. We compare the empirical cardinality $N_m(emp)$ to the mean cardinality of the ensemble $N_m(rnd)$, and judge about the significance of the deviations by comparing them to the standard deviation σ_m

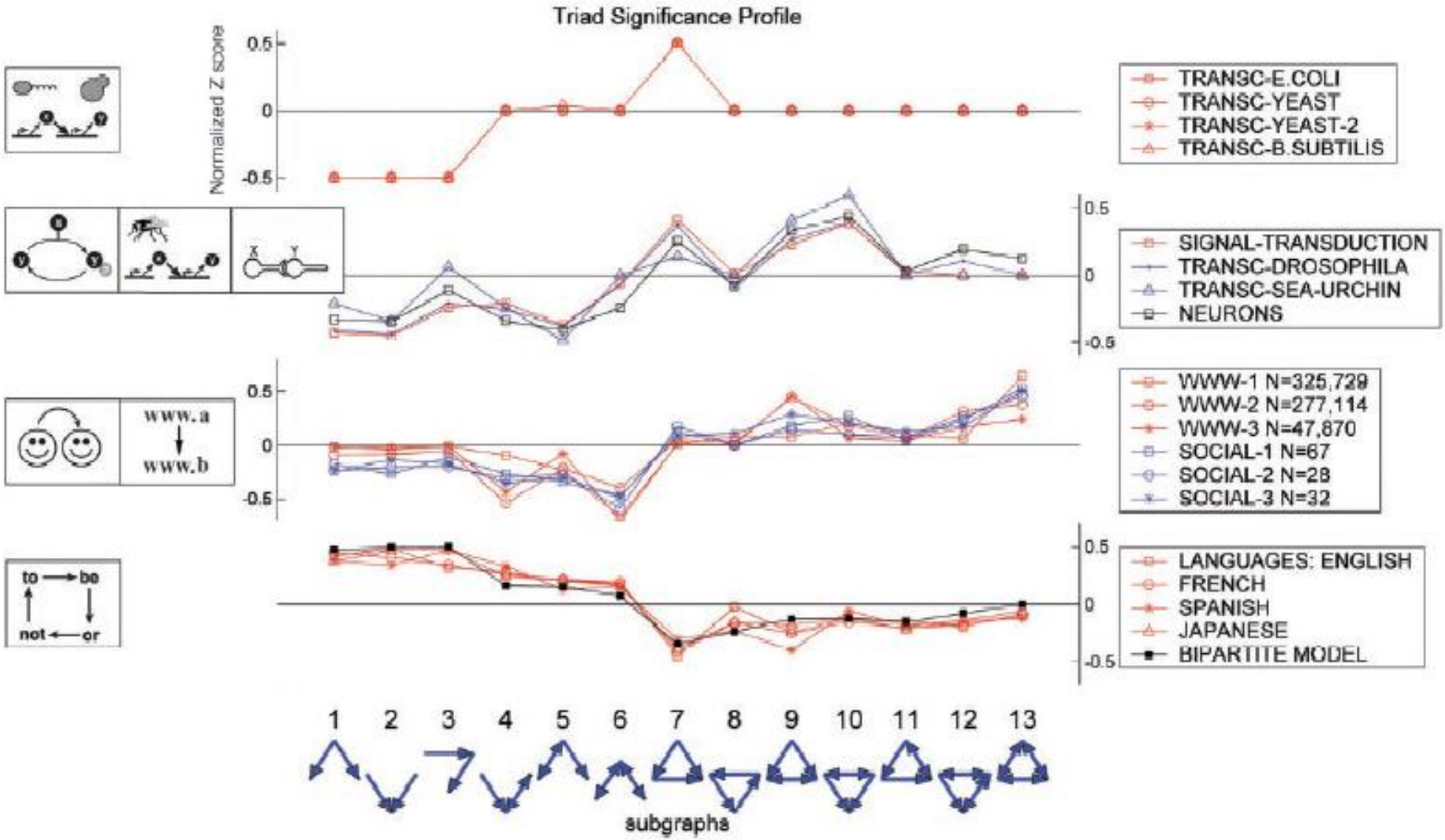
Measure: z-score

$$z_m = \frac{N_m(emp) - N_m(rnd)}{\sigma_m}$$

Uri Alon

R. Milo et al. Science, 298:824-827 (2002)

Network motifs

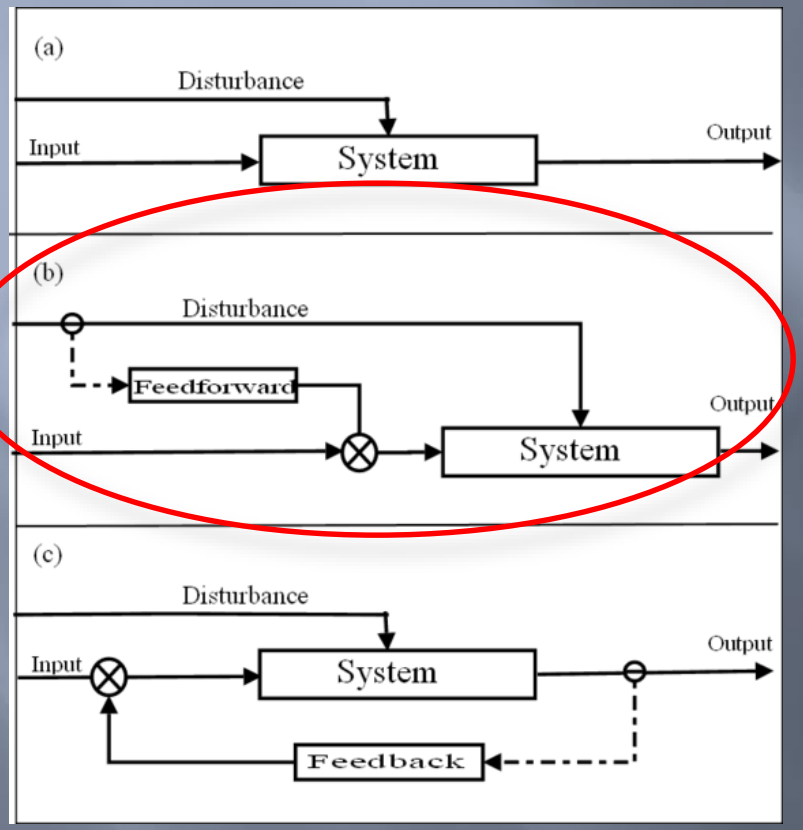


Different types of networks seem to have different typical motifs

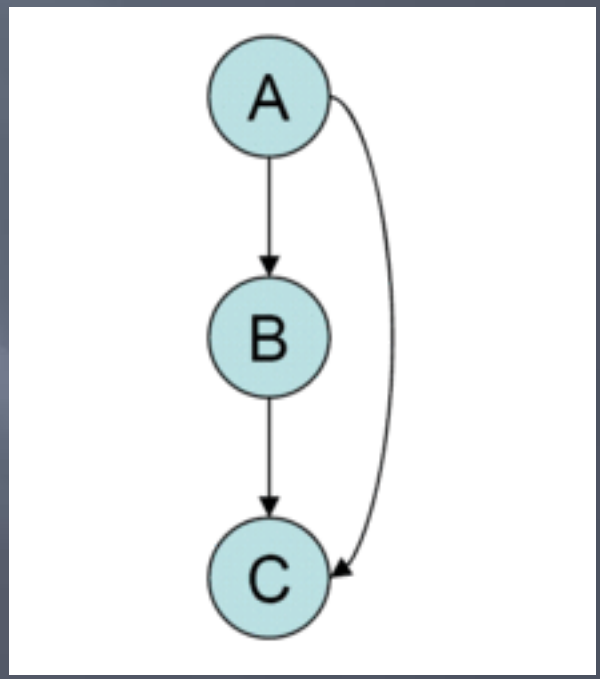
R. Milo et al. Science, 298:824-827 (2002)

Network motifs

E.g., in regulatory networks (gen transcription) some motifs are typical.



Basic regulatory schemes



Typical feed-forward subgraph

Network motifs

Another measure of the significance of the motif frequency: p-value statistics

Take the empirical network with *N*(*emp*)

Generate the ensemble by link switching; make *M* measurements or samples (perform enough switches between two measurements).

$$p = \frac{1}{M}\sum_{i=1}^{M}\theta\left(N_i - N(emp)\right); \qquad \text{with } \theta(x) = \begin{cases} 1 \text{ if } & x > 0 \\ 0 \text{ otherwise} & \end{cases}$$

Here N_i is the cardinality of the motif in sample *i*

Network motifs

Important: How to choose the reference system?

E.g., in a biological system one may conclude that significant overrepresentation of a motif reflects evolutional advantage. However, genetic networks are embedded in space and neighboring nodes interact easier than far laying ones – but this aspect is entirely ignored if the reference system is the configuration model.

There is generally special care needed when choosing a null model!

Yael Artzy-Randrup, et al. Science (2004).

Network motifs

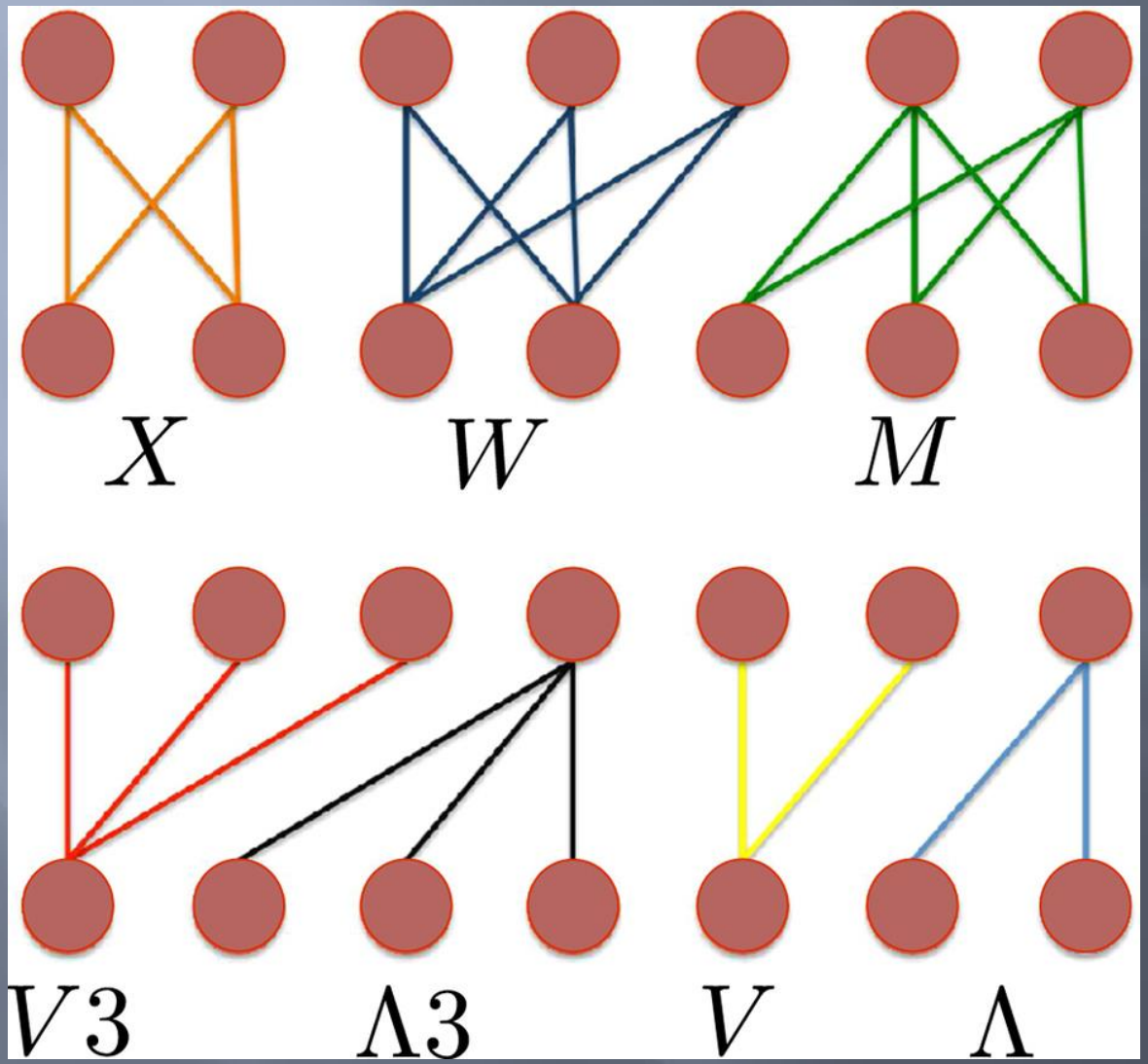
Technically, subgraphs have to be identified and there are many programs around doing this.

Usually directed networks are considered and *N~3-5*

This motif approach has been extensively applied in biology and not so much in social sciences.

The observation that social networks have high clustering is about a specific motif.

Motifs in bipartite networks



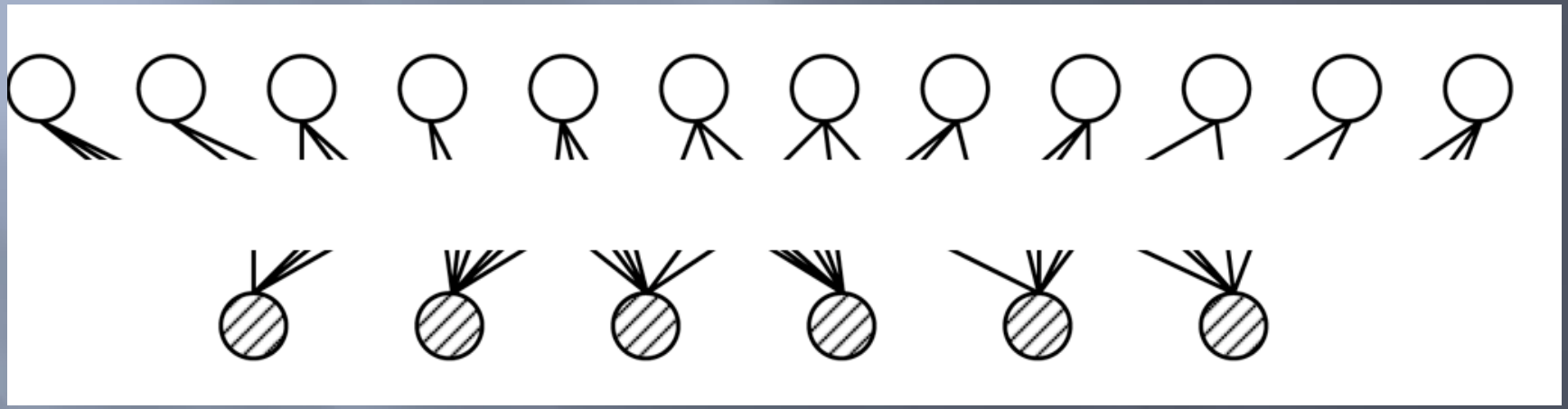
The simplest undirected motifs

Saracco et al. (2015)

Motifs in bipartite networks

What is the proper null model?

BiConfiguration model as obtained by matching always stubs from *A* with ones from *B.*



$$\sum_{i \in A} k_i = \sum_{j \in B} k_j$$

Motifs in bipartite networks

For advanced students:

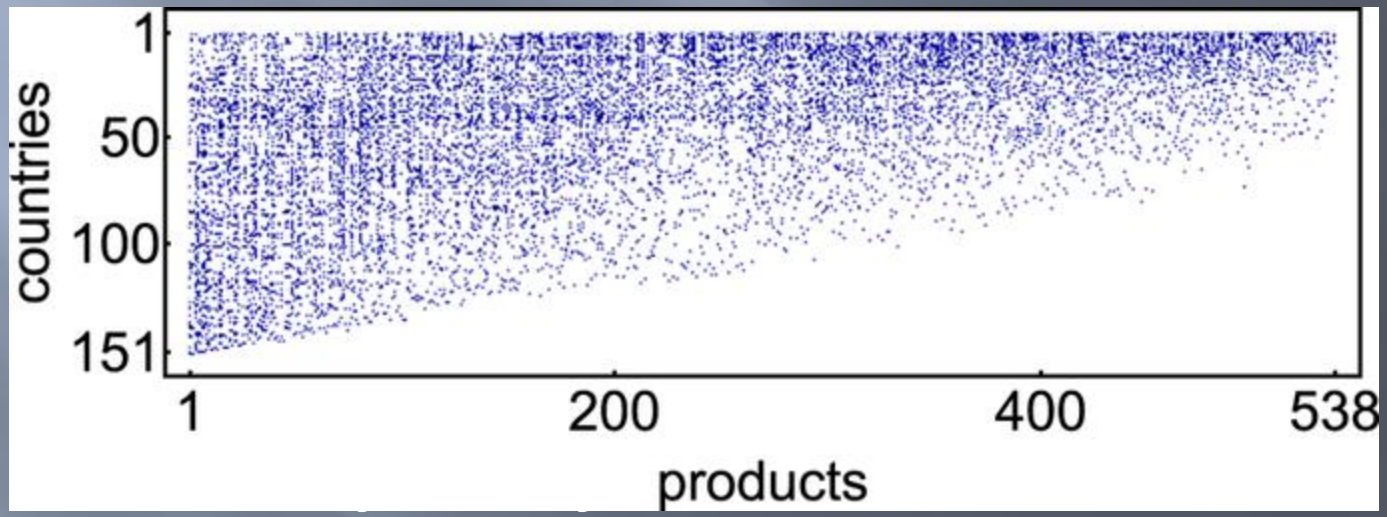
The Shannon entropy *S* of an ensemble of networks: $S = -\sum_M P(M)\ln P(M)$, where M is a biadjacency matrix

The most random configurations define an ensemble, for which $S = \max$, where restrictions or constraints $\boldsymbol{C}(M)$ should be considered. E.g., constraints are the degree sequences in the sets *A* and *B.* We arrive at distribution containing as many parameters as constraints (Lagrange multipliers). Their values are calculated by maximum likelihood method. For details see the orig. paper by Saracco et al. (2015)

Motifs in bipartite networks

Example: World Trade Network
Set *A*: Countries
Set *B*: Products

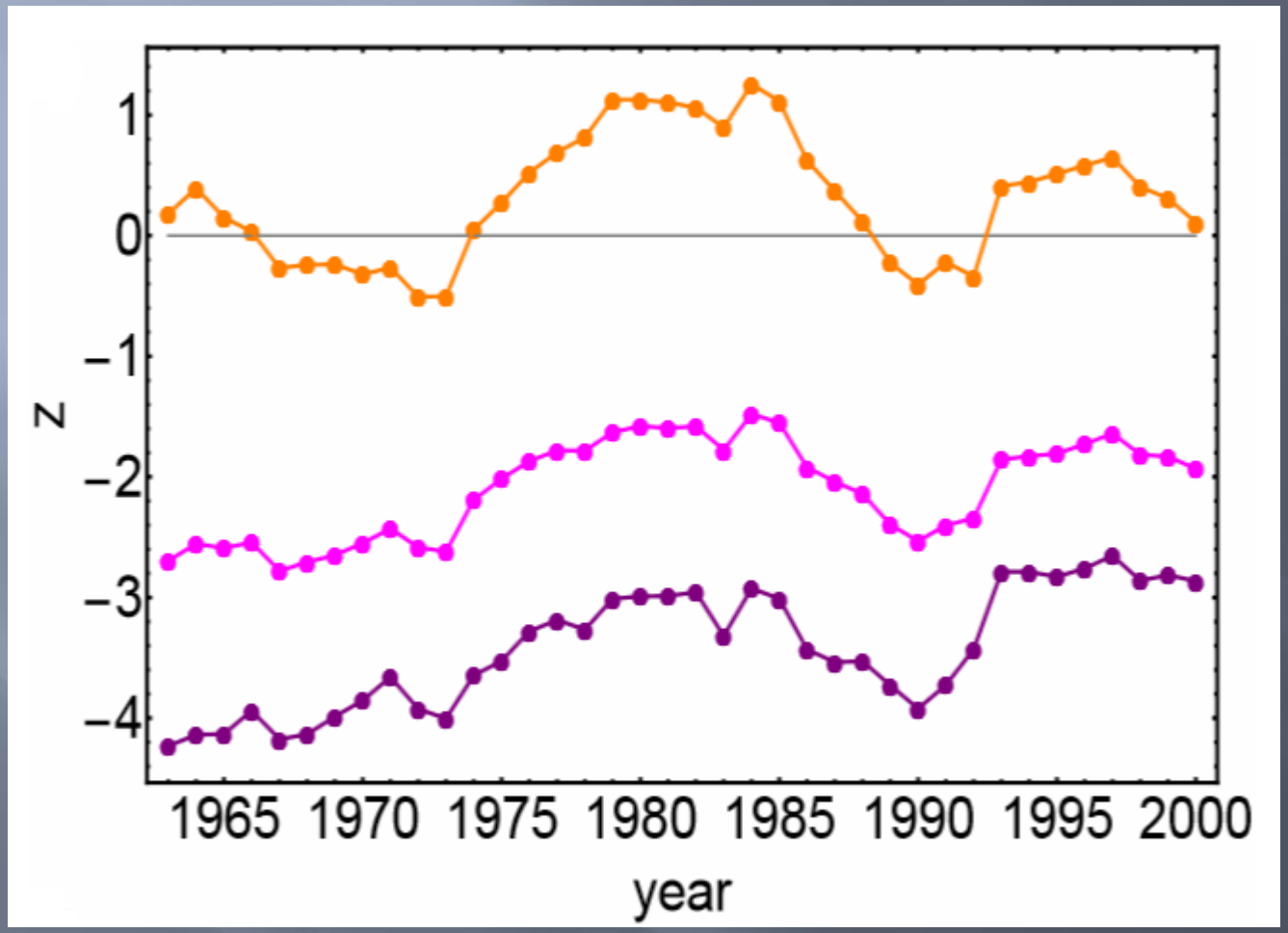
Biadjacency matrix:



Countries ordered according to the diversity of export

Motifs in bipartite networks

Having the null model *z*-scores can be calculated.



Egocentric networks

Egocentric networks

Part of a collab. network

Definition:

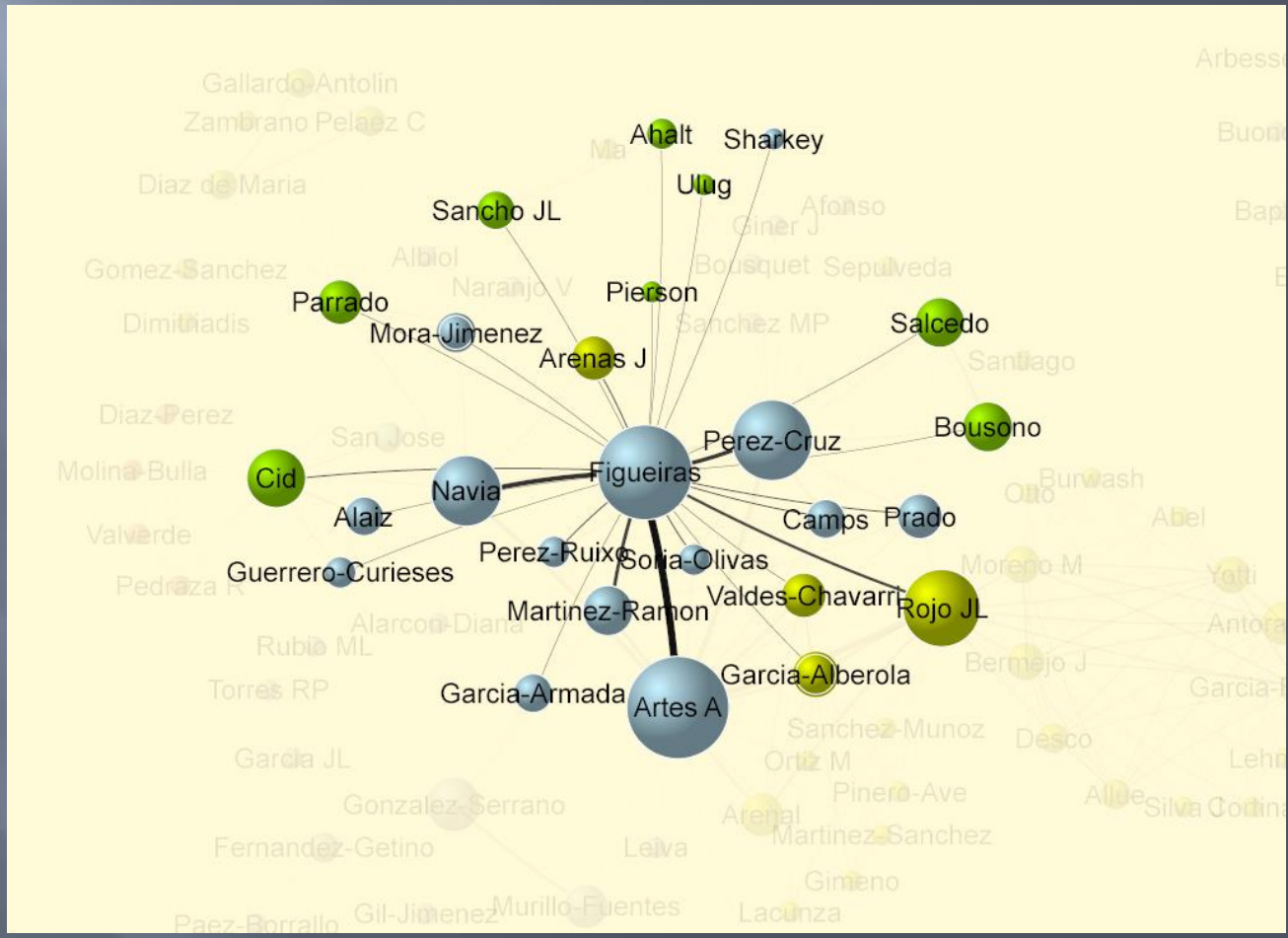
$$G_i^{ego} = \{V', E'\}$$

$$i \in V'$$

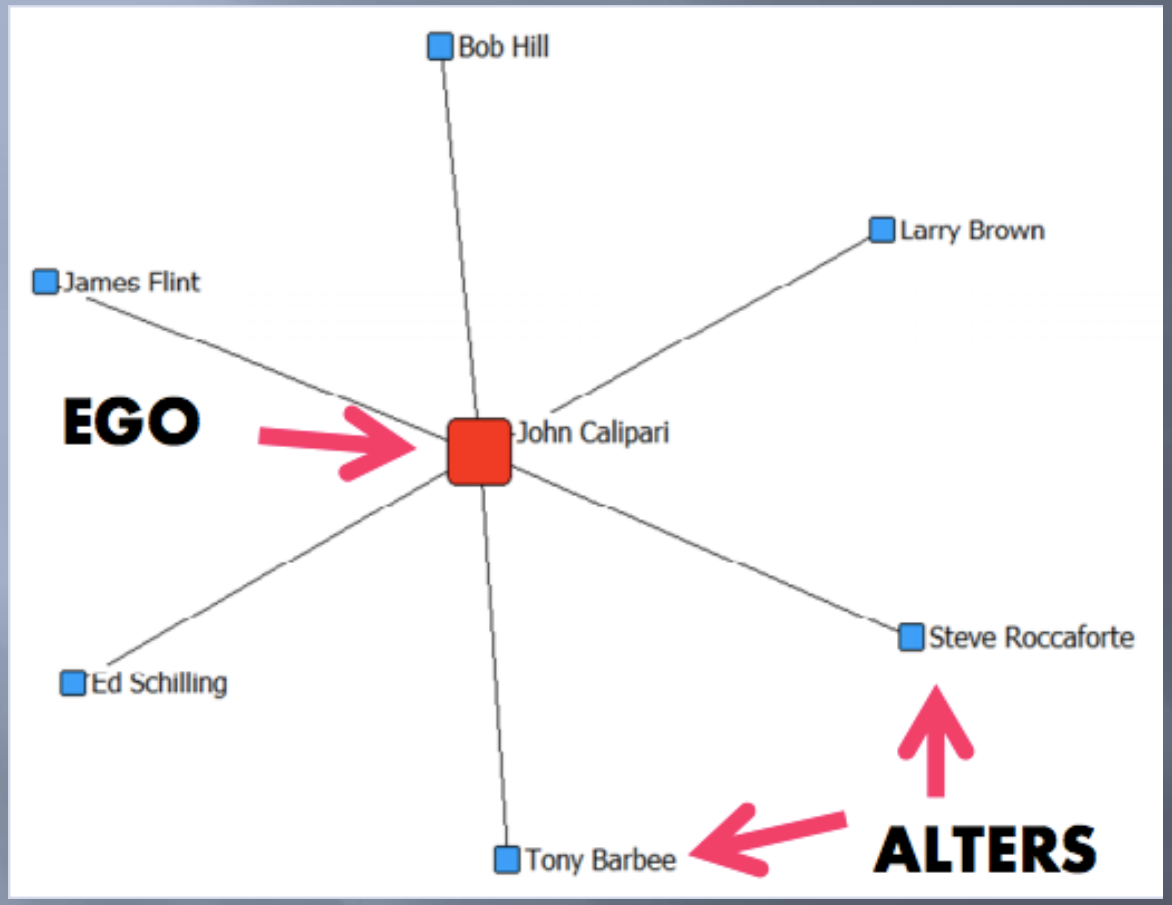
$$\forall j \in V' \text{ iff } \exists e_{ij} \in E$$

$$e_{ij} \in E'$$

Ego + all neightbors and links to them



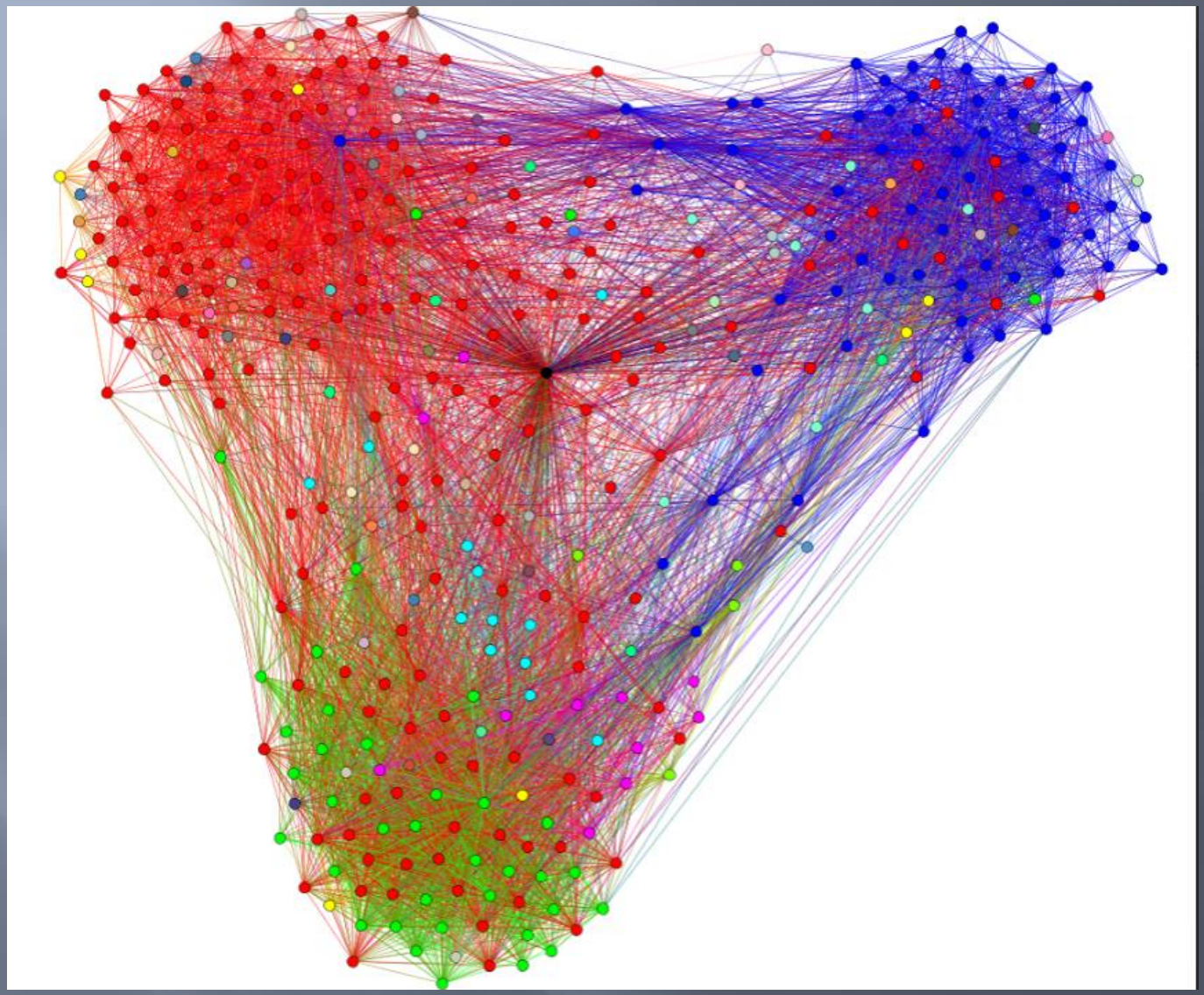
Egocentric networks



There is a focal node: EGO
This is the node we are interested in

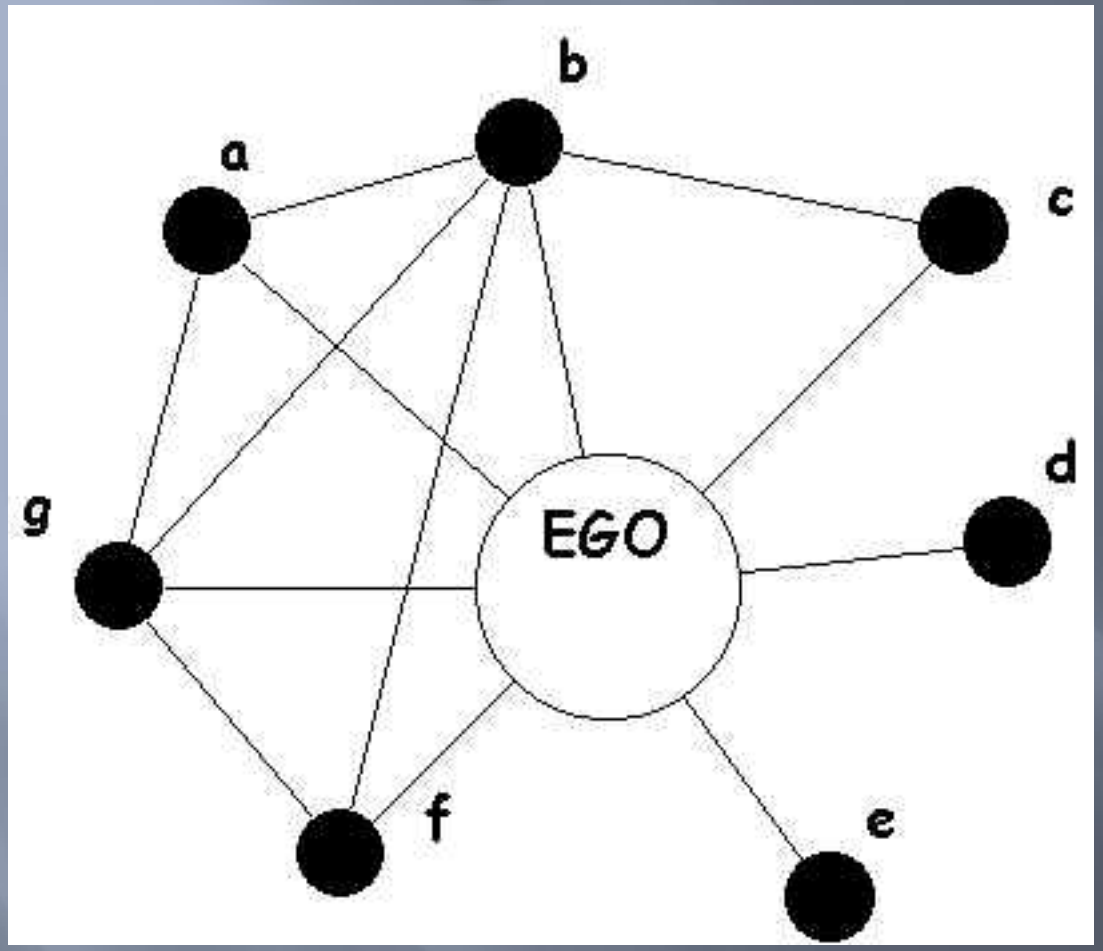
Neighbors (friends, business partners, linked pages etc.) ALTERS

Egocentric networks



Ego NW from an OSN

Egocentric networks



We may include links between alters (but not more)

$$G_i^{\text{ego}} = \{V', E'\}$$

$$i \hat{\text{I}} \; V'$$

$$\text{"} \, j \hat{\text{I}} \; V' \;\; \text{if} \;\; \$e_{ij} \hat{\text{I}} \; E$$

$$e_{jk} \hat{\text{I}} \; E' \;\; \text{if} \;\; \text{both} \, j, k \hat{\text{I}} \; V'$$

Enables to consider C_i clustering of ego

Each alter is an ego in his/her egocentric network

Egocentric networks

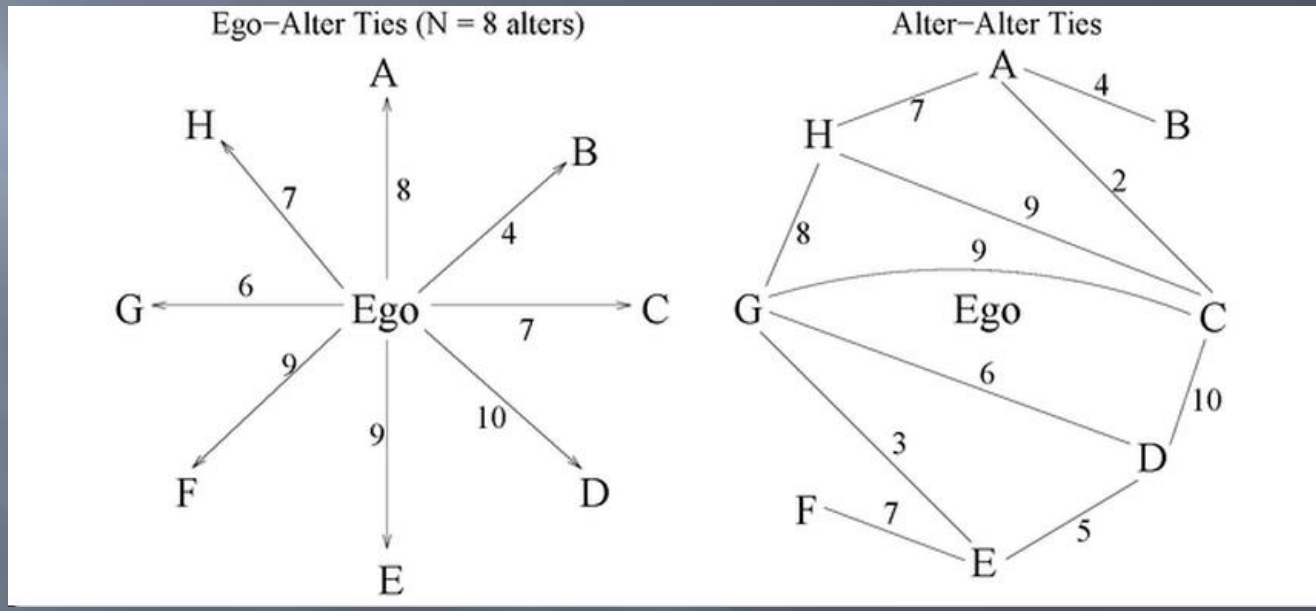
If topological properties are combined with attributes of the ego, unique "portrait" can be obtained.

- Ascribed characteristics
  - Sex
  - Age
  - Race
  - Place of birth
  - Family ties
  - Genetic attributes
- Chosen characteristics
  - Income
  - Occupation
  - Hobbies
  - Religion
  - Location of home
  - Amount of travel,
- Social outcomes: Personality, acculturation, well-being, social capital, social support
- Health outcomes: Smoking, depression, fertility, obesity

Egocentric networks

Egocentric networks are widely used in social sciences, often based on surveys

Survey about US households contained questions: name 4 persons whom you spent most time during last month and 4 whom you discussed important issues. Grade your relationships with them and also those between them. The study pointed out correlations between social embeddedness and health status of egos.



A. James O'Malley et al. PLoS ONE, 2012

Modular structure

We had 3 “universal” properties of networks:

a) High clustering
b) Small average distance
c) Broad degree distribution

There is one more!

Empirical networks are **modular**

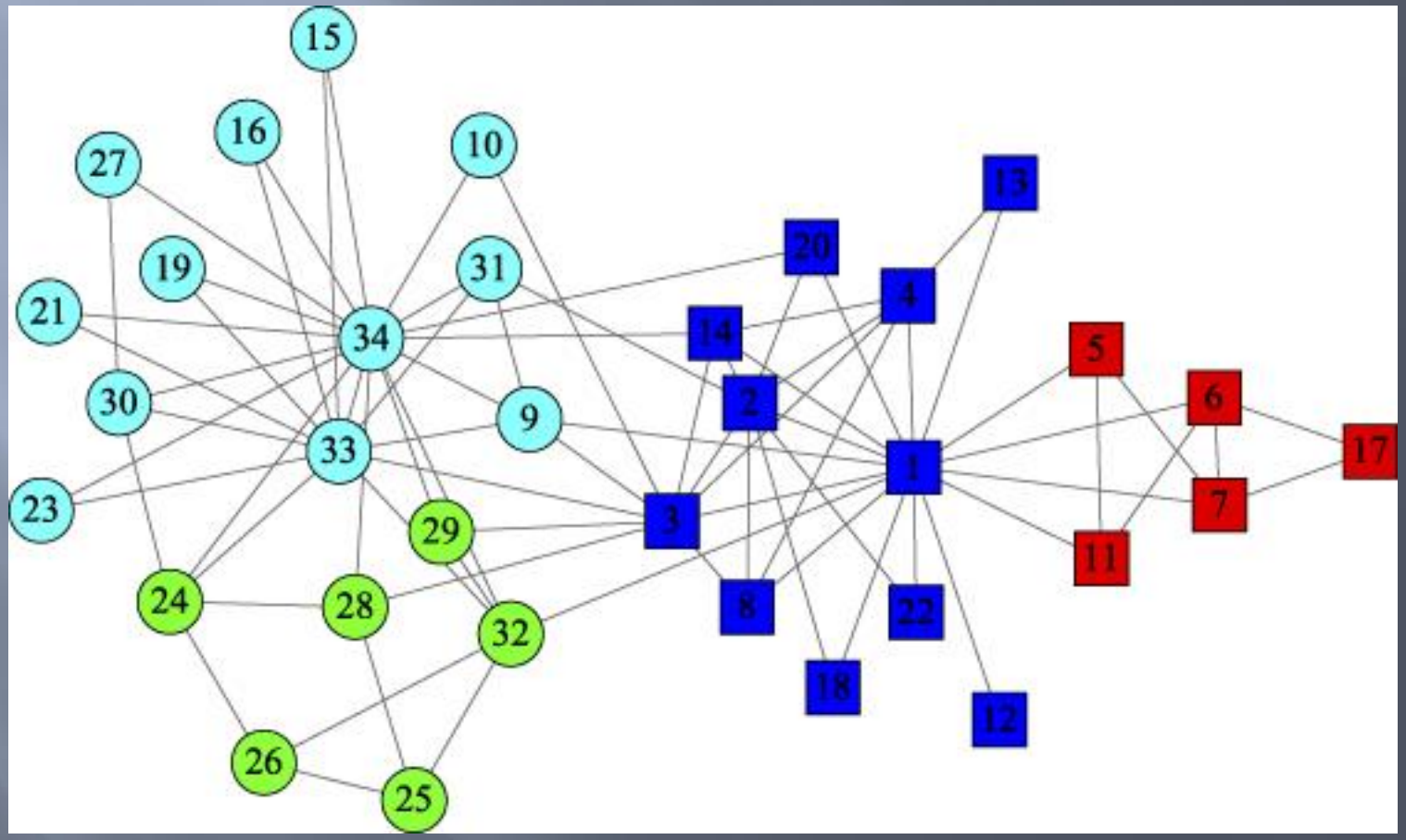
Modular structure

Real networks are very inhomogenous not only on the microscopic scale (degree): There are densely and loosely wired parts.

Dense parts can be assumed to have a correlated role in the functioning of the network: Such part is a module or community.
Communities exist in social networks (families, teams, friendship circles etc.), or functional modules in metabolic networks, thematic groups in www, industrial branches in economy etc.

Modular structure

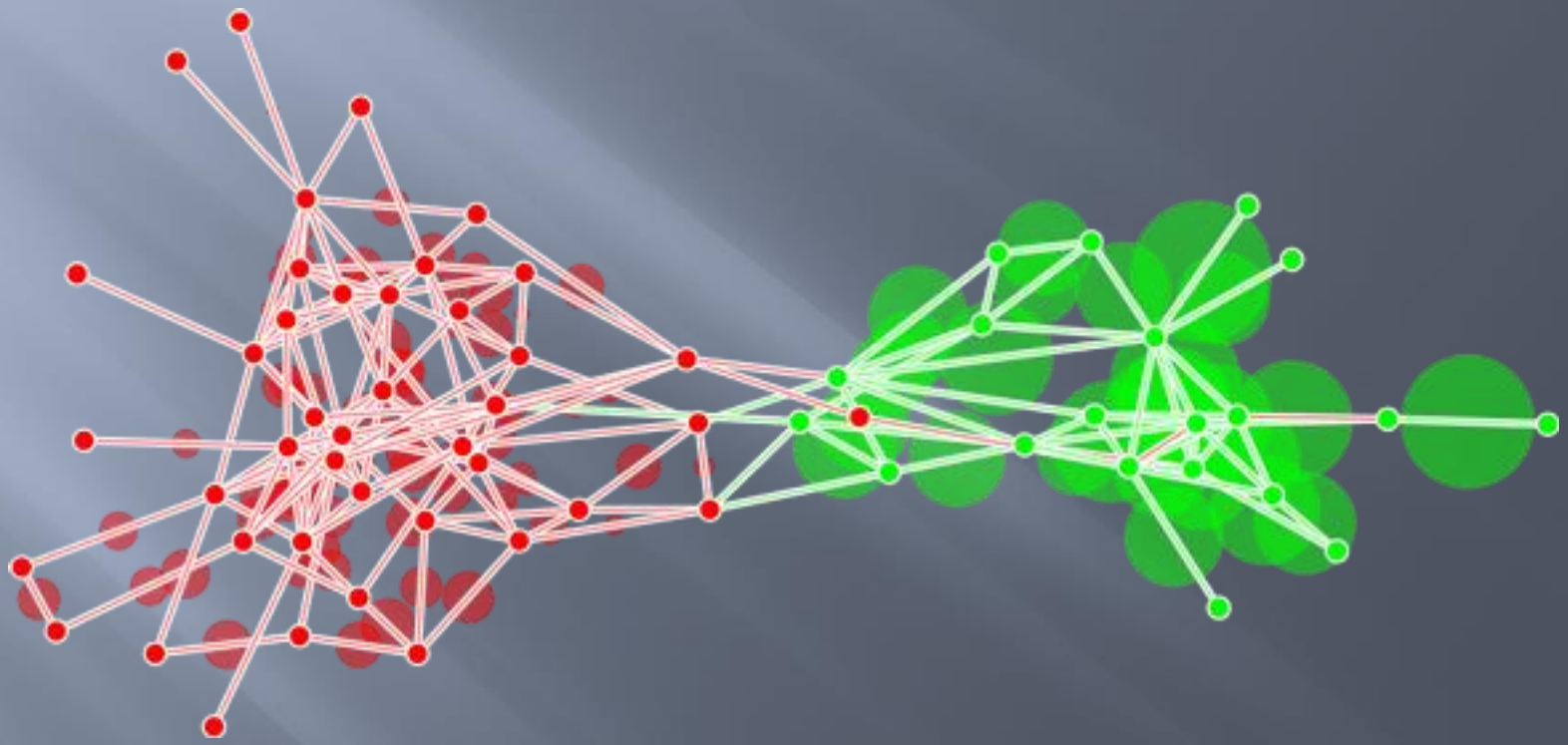
Zachary karate club



Zachary W W, 1977 J. Anthropol. Res. 33 452 (1977)

Modular structure

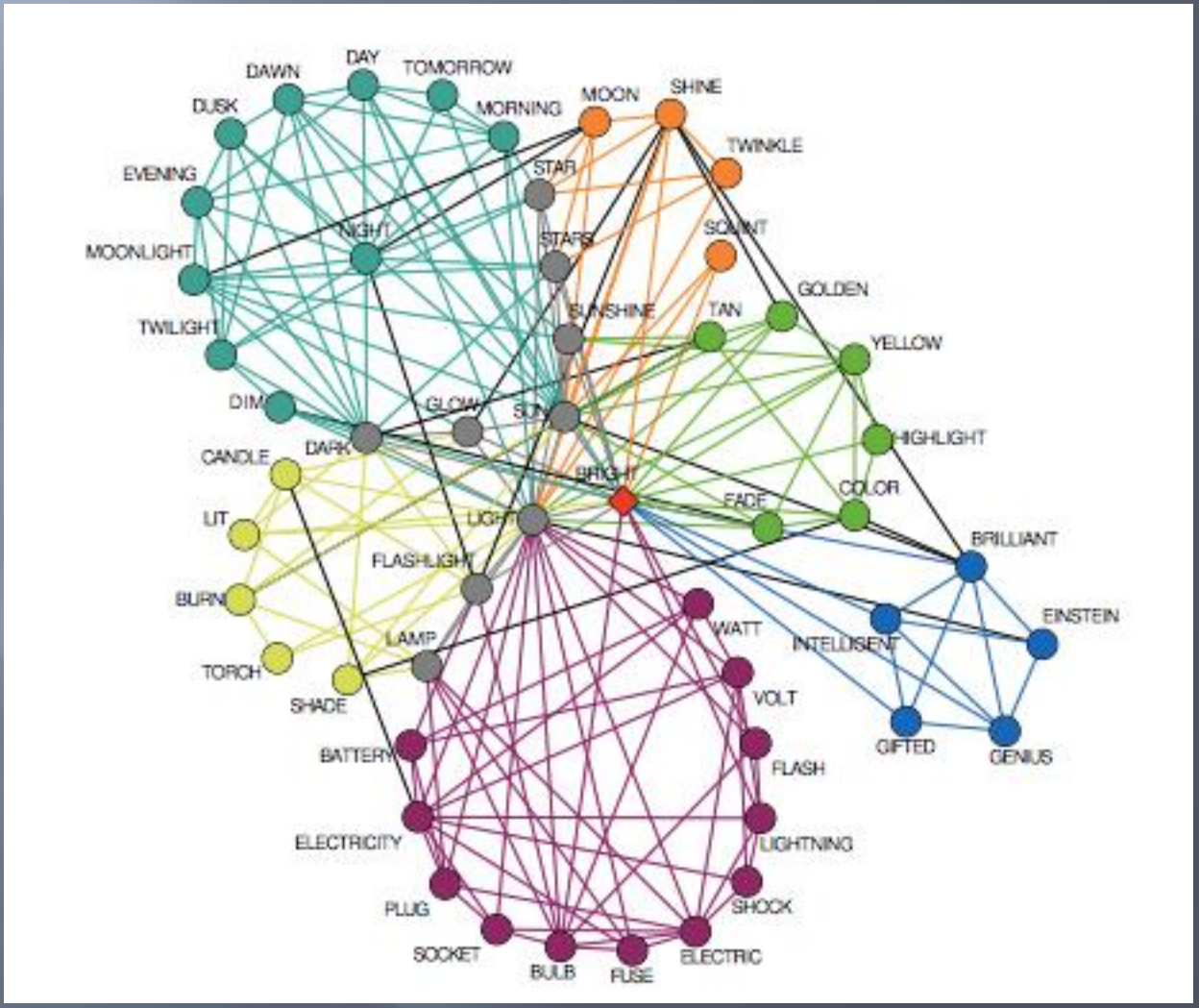
Dolphin network



D. Lusseau, et al. Behavioral Ecology and Sociobiology 54, 396-405, 2003)

Modular structure

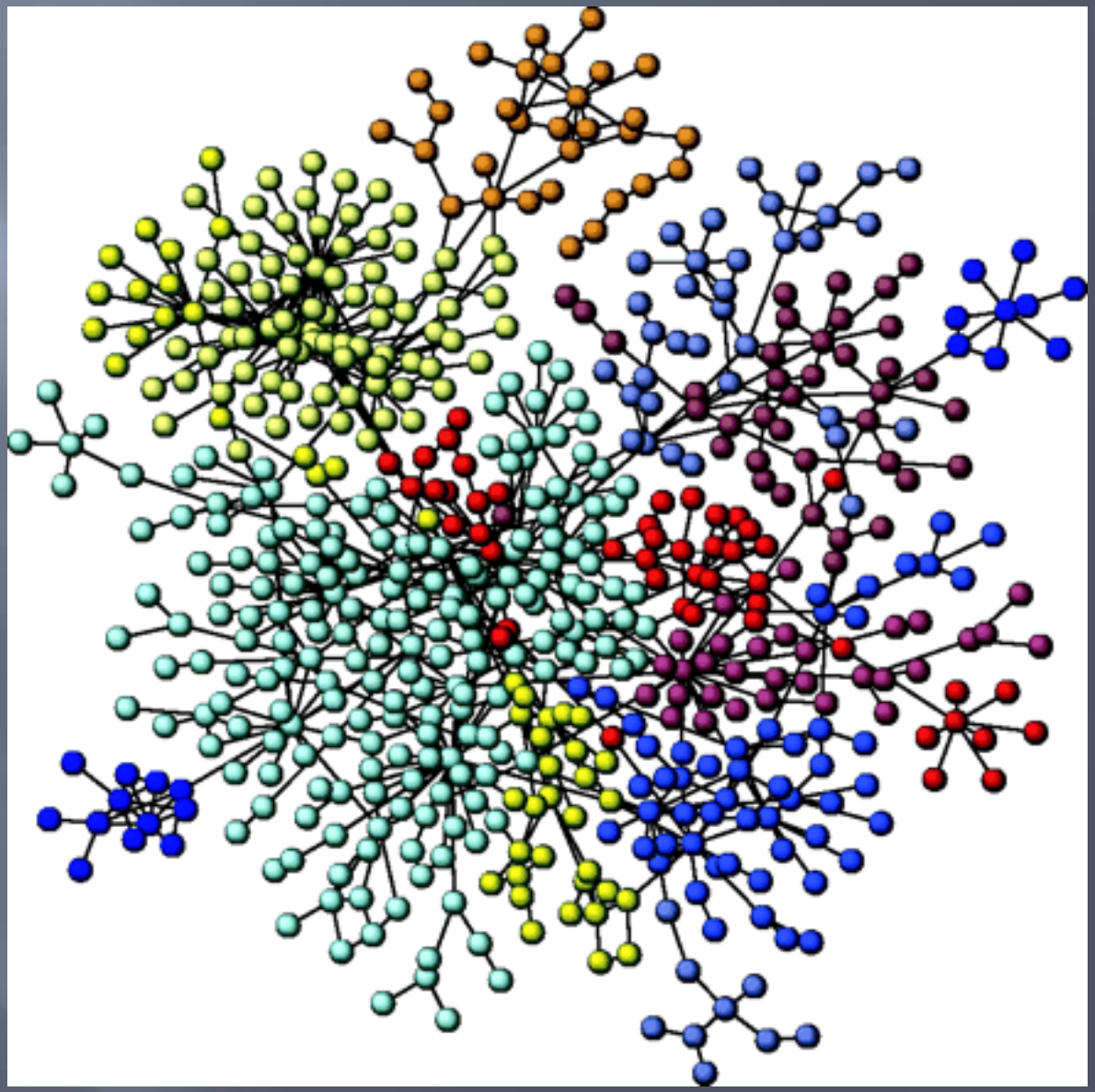
Word association



Tibély (2011)

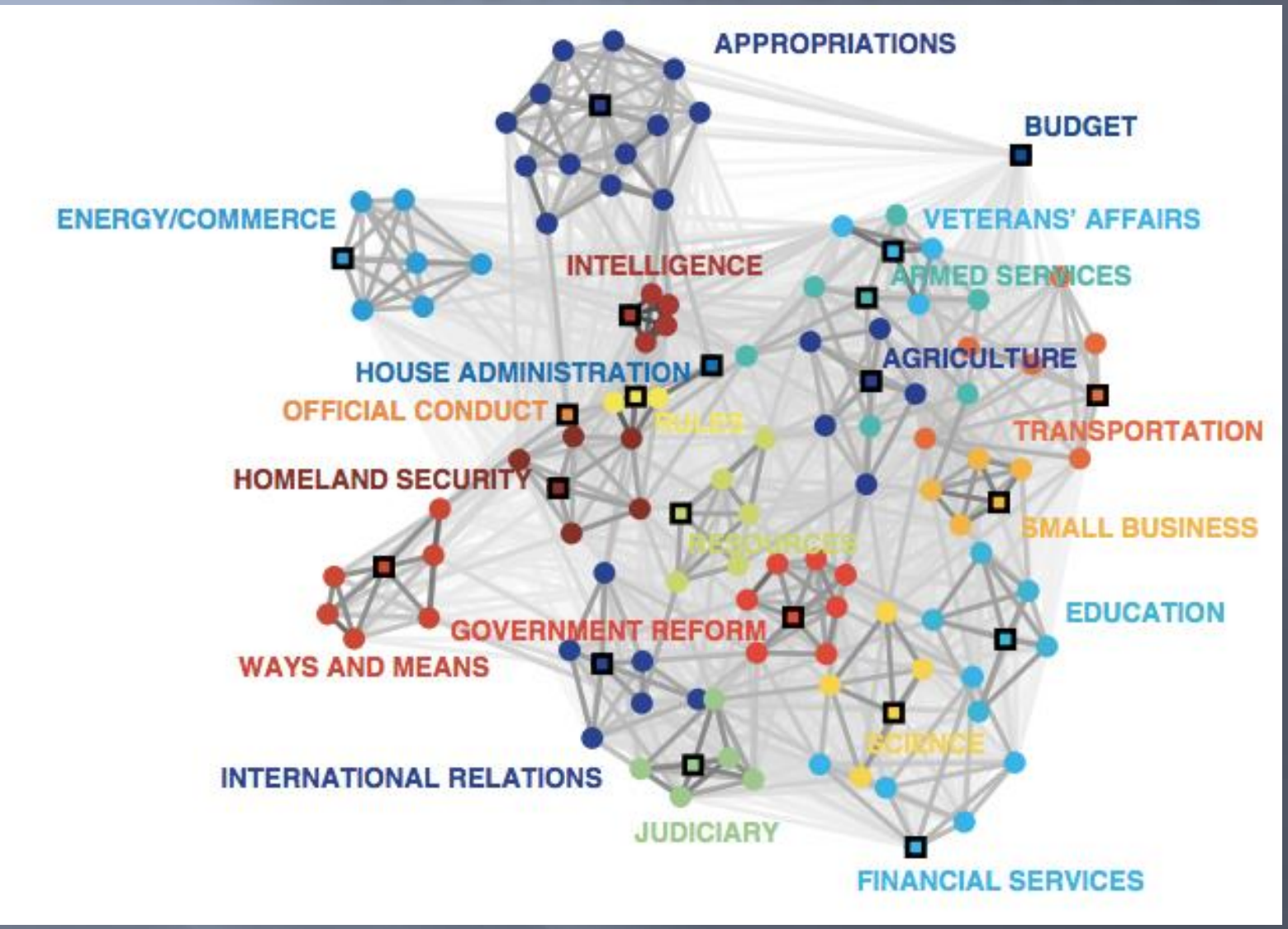
Modular structure

Protein-protein interaction NW



R. V Solé and S. Valverde 2008

Modular structure



Committees in the 108th US House of Representatives
Link: Common membership

Mason A. Porter Physica A (2007)

Modular structure

Communities (modules): families, friendship circles, departments, sport clubs etc.

Intuitive „definition”: There are more links inside the community than going out of it.

Local definitions:
Strong community: Each node has more neighbors inside than outside
Weak community: Total degree within the community is larger than the total degree out of it.

Global definition: The community structure found is optimal in a global sense

F. Radicchi et al., PNAS 101, 2658 (2004)

Modular structure

Major challenge: Knowing only the topology of the network (the adjacency matrix) how to identify the communities? The task of community detection

Unsolved problem, in the focus of present research

Hundreds of competing methods

- Global
- Local

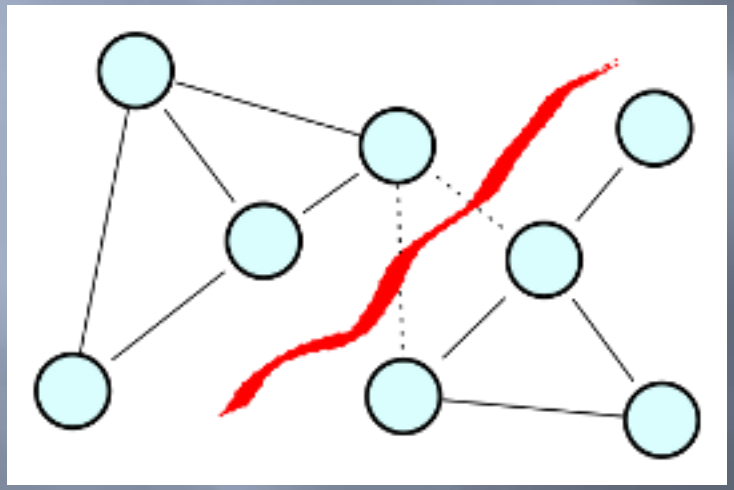
Main problems:

- Hierarchies of communities,
- Resolution
- Overlapping communities

Modular structure

Global methods

Graph partition: cutting links such that the network gets separated into disjoint pieces.



Related to interesting math problems, like

k-partition: How to separate a graph into *k* parts with minimal links cut

Uniform graph partition: $k = 2$ such that the two parts have the same size

NP complete problems, only approximate, heuristic algorithmic solutions exist.

Modular structure

Global methods

If the communities are disjunct and every node belongs to a community then finding them is equivalent to a partition.

Every partition can be considered as a community structure, some are better than others…

How to decide if a partition leads to a good community structure?

Use a global “goodness” criterion

Modular structure

Global methods

Modularity (Newman and Girvan, 2004)

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

A is the adjacency matrix, k_i is degree of node *i*, *L* is the total number of edges in the network, the Kronecker delta indicates that both nodes *i* and *j* have to be in the same module; the summation runs over all pairs of nodes.

Modular structure

Global methods

Modularity

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

The adjacency matrix is compared to the probability of having a link between nodes *i* and *j* in the configuration model. This is the reference system.

If there is a link the contribution is positive. The smaller the probability would be in the configuration model, the higher the contribution – and the opposite if the link is missing.

Modular structure

Global methods

Modularity

$$Q = \frac{1}{2L} \sum_{i \neq j} \left(A_{ij} - \frac{k_i k_j}{2L} \right) \delta(C_i, C_j) \rightarrow Q = \sum_{s=1}^{n} \left(\frac{\ell_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right)$$

where the sum now runs over the modules. Here l_s is the number of edges within the module, while d_s is sum of degrees of vertices within the modules. The first term is the link density within the module, the second one is the expected link density in the configurational model. Q>0 means we have found modules. Q = max: we have found the best structure!

Modular structure

Global methods

Modularity (Newman and Girvan, 2004)

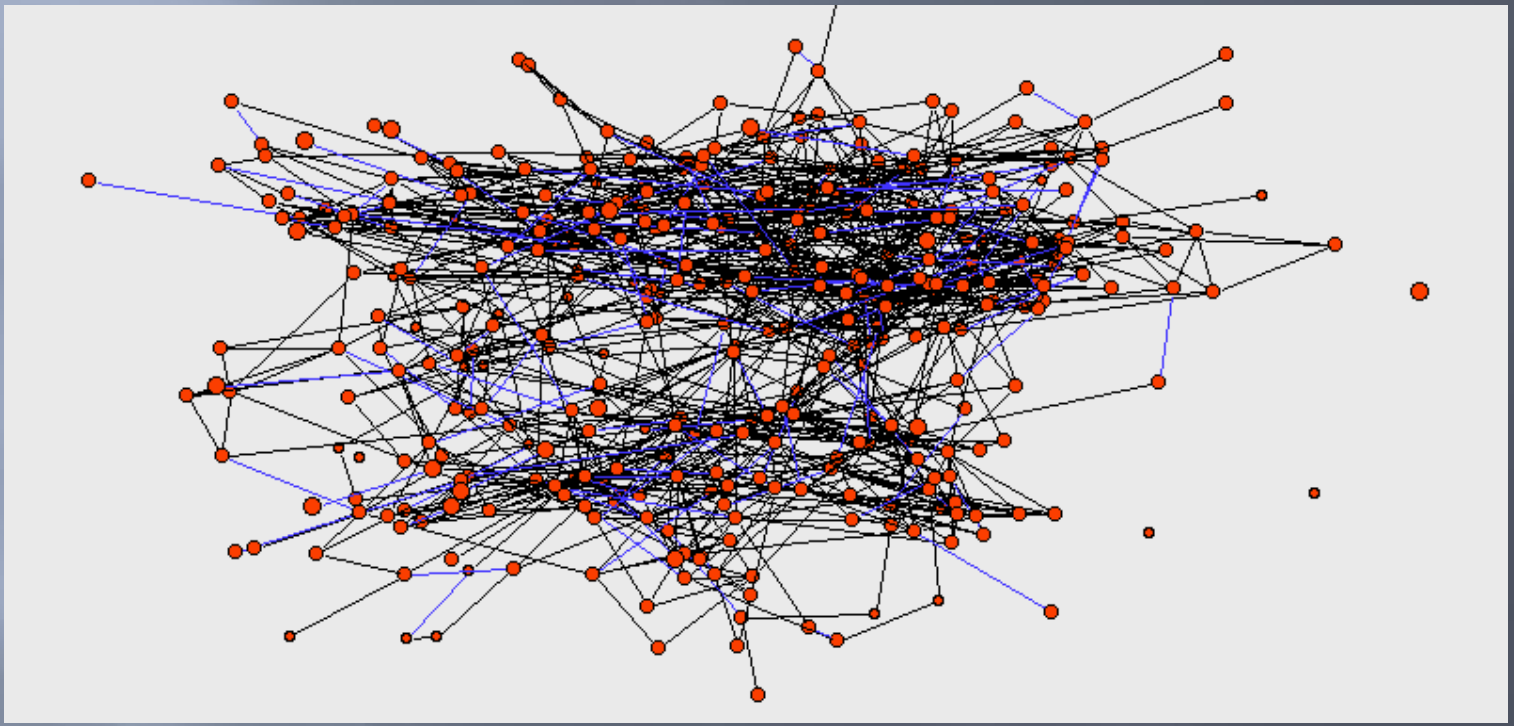
$$Q = \sum_{s=1}^{n} \left(\frac{\ell_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right)$$

How to find a proper partition?
Which links should be cut?

High betweenness links connect densely wired parts.
1st NG algorithm: Order links according to their betweenness and start cutting the highest. Go through and select the partition with the highest *Q*!

Modular structure

Global methods



Friendship and marriage ties in a town. Friendship ties are in black, marriage ties in blue.

Very tedious task!

Modular structure

Global methods

Modularity (Newman and Girvan, 2004)

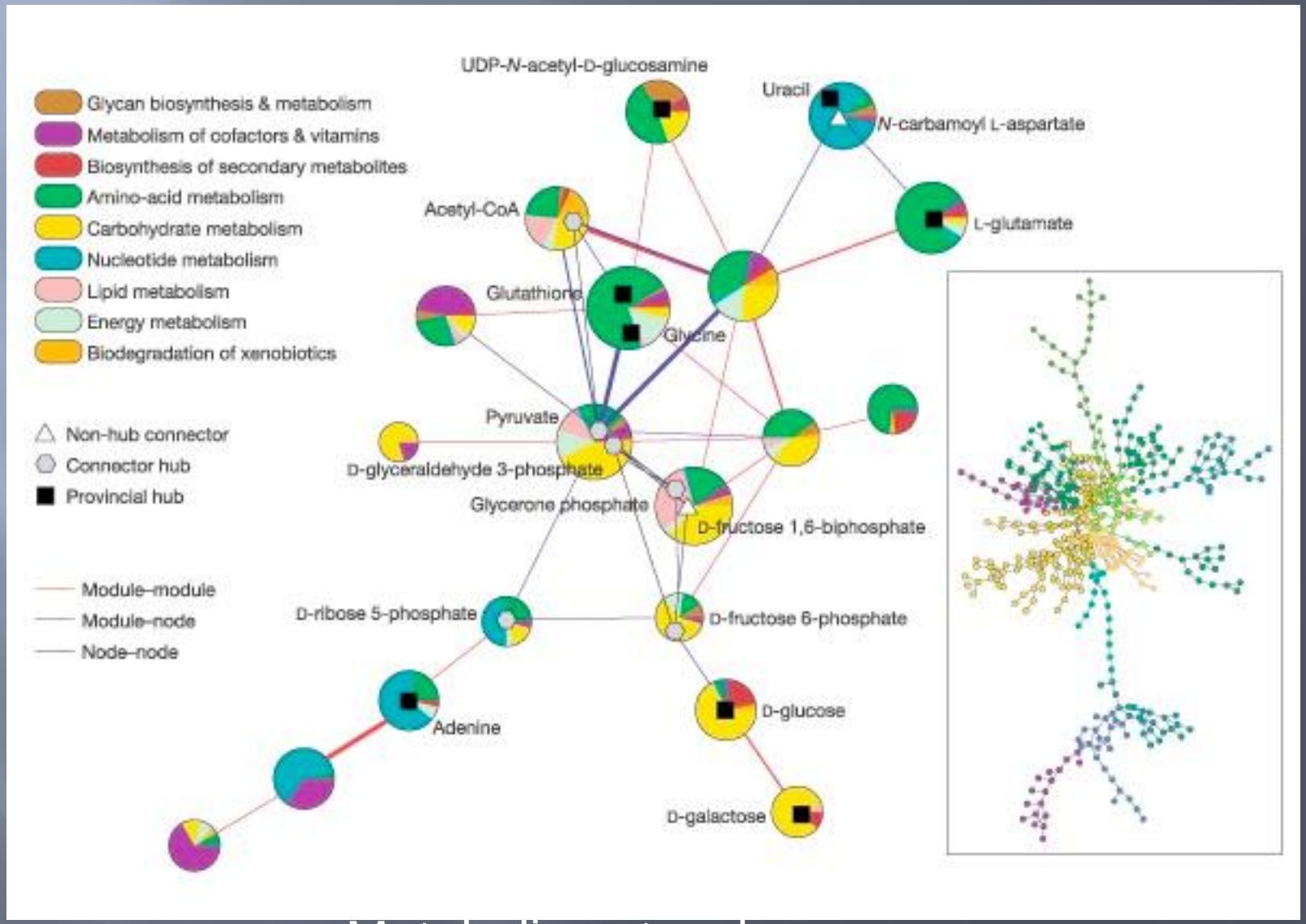
Turn it upside down: Find the best partition by optimizing *Q*!

NP-complete problem.

Approximate, heuristic solution: Greedy algorithm: Start from *N* clusters, each node is a cluster and calculate *Q*. Group nodes such adding an edge results in maximal increase of *Q*. Stop if no way to increase *Q*. Polynomial running time. Leuvain: nlogn

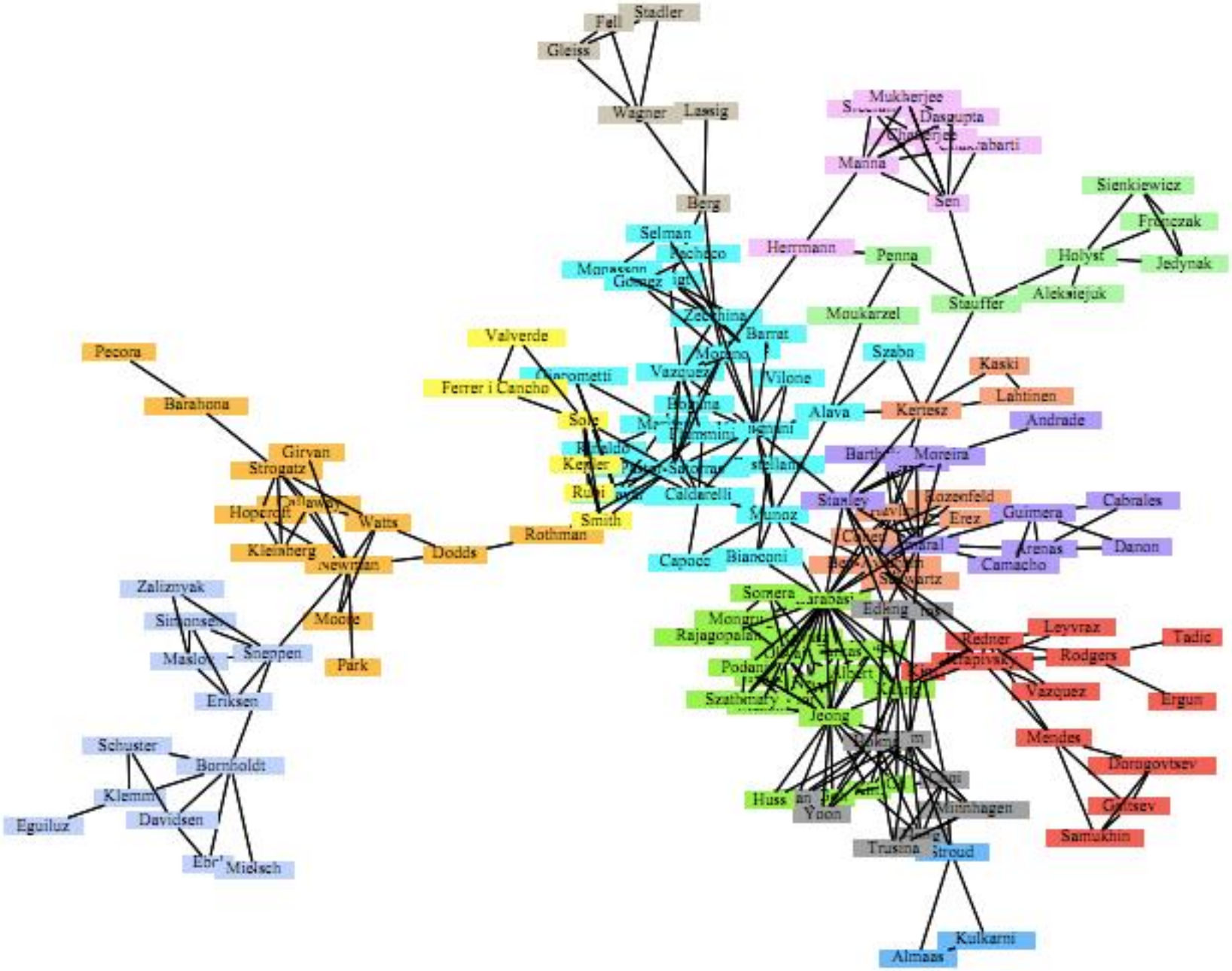
What kind of error do we make?

Modular structure



Metabolic network

Guimera & Amaral 2004



Collaboration network

Newman 2003

Modular structure

Resolution limit of the modularity method

Small, plausible communities cannot be found in a large network

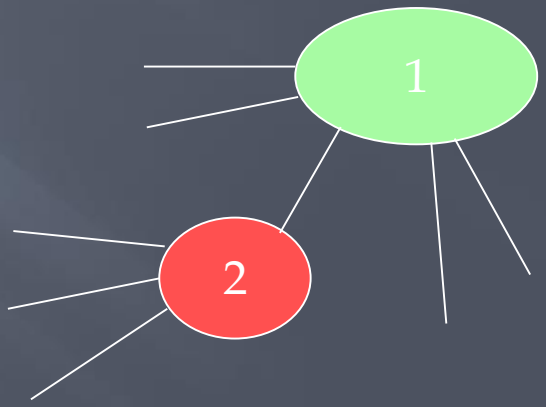
$$Q = \sum_{s=1}^{n} \left(\frac{\ell_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right)$$

where l_s is # links inside module s
d_s is the total degree in C_s
n is # modules

When is it worth considering two connected communities as a single one?

$$\frac{l_1 + l_2 + l_{12}}{L} - \frac{(d_1 + d_2)^2}{(2L)^2} > \frac{l_1}{L} + \frac{l_2}{L} - \frac{(d_1)^2 + (d_2)^2}{(2L)^2}$$

$$\Delta Q = \frac{l_{12}}{L} - \frac{2 d_1 d_2}{(2L)^2} > 0 \qquad 2L > d_1 d_2$$

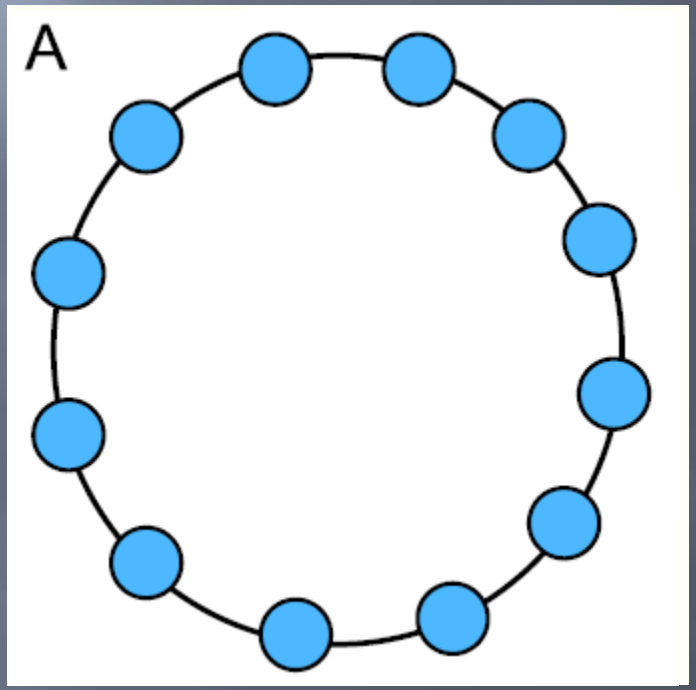


Even if the small communities are cliques and a single link connects them….

Frotunato-Barthelemy 2008

Modular structure

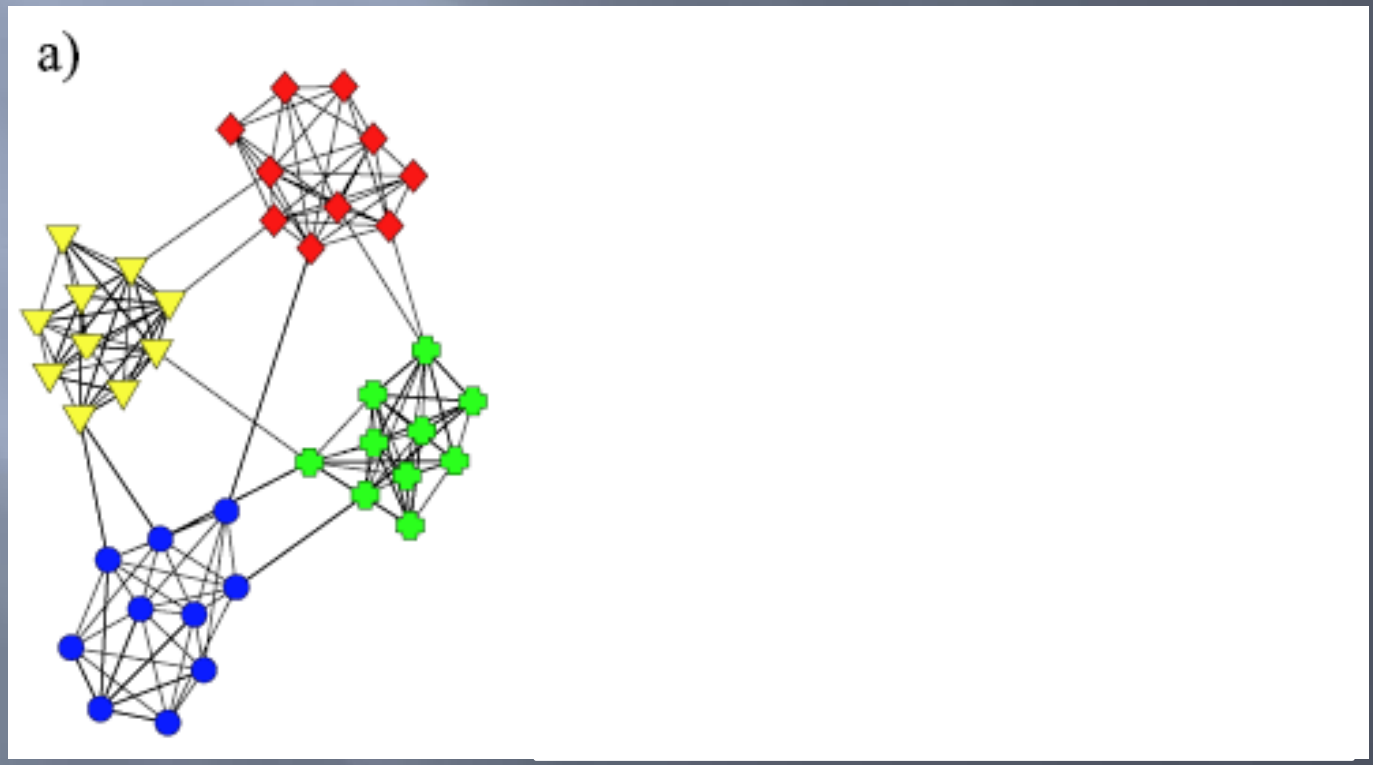
d_s is characteristic of the module size. Assuming equal size moduls we conclude that $d_s < \sqrt{L}$ size modules cannot be seen by the *Q*-based method. This is a resolution limit.



Circles symbolize a m-cliques. Just by changing the length of the chain (changing *L*) it may be worth considering pairs of cliques as single communities. Unphysical: limitation of the modularity appr.

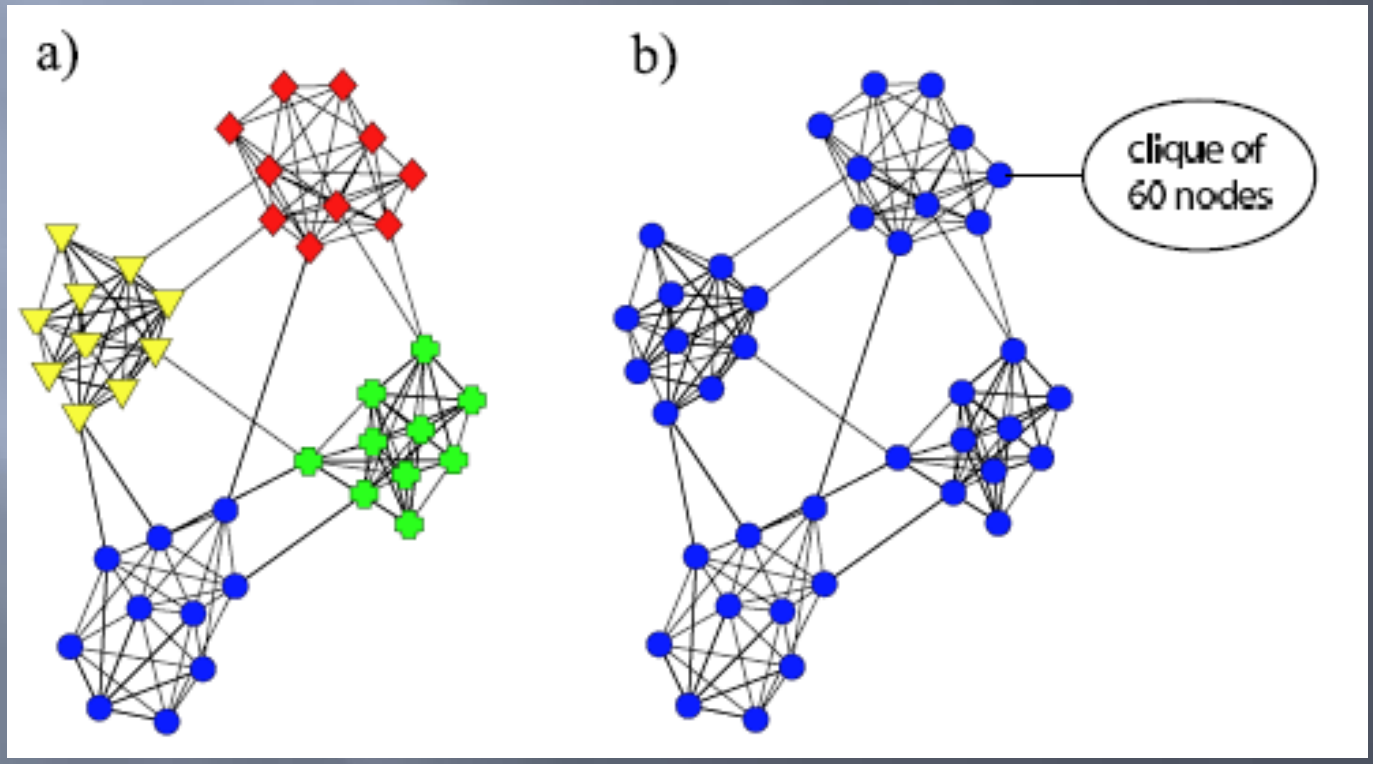
Modular structure

Illustration of the resolution problem:



Modular structure

Illustration of the resolution problem:



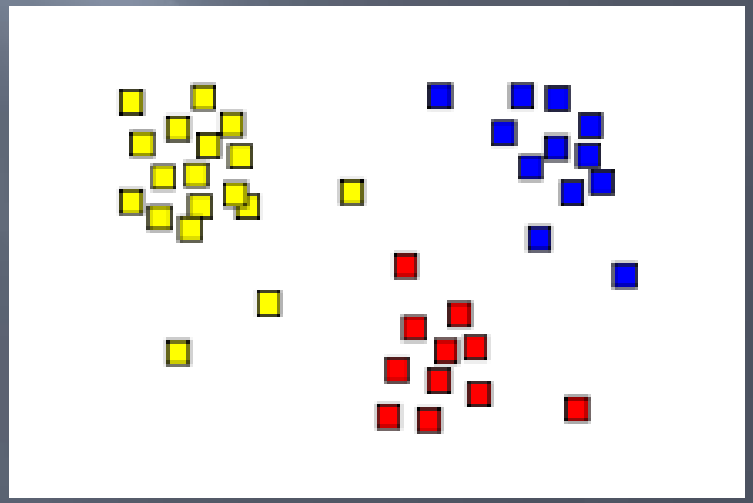
Action at distance!

Modular structure

Further global methods: clustering algorithms

Related to general problems of classification in computer science, artificial intelligence, pattern recognition, data mining, etc.

Recognizing groups of objects, their patterns etc.



Wikimedia

Modular structure

Start with a similarity measure

This can be the Euclidean distance between the points or betweenness centrality of the links connecting nodes etc.

Many algorithms. For some we need to know the number of clusters (modules, communities), which is usually not the case.

- Family of hierarchical clustering algorithms
- Centroid based clustering (fixed number of clusters)
- Density based clustering etc.

Modular structure

Agglomerative hierarchical clustering

Start from the individual nodes (bottom up), i.e., *N* clusters

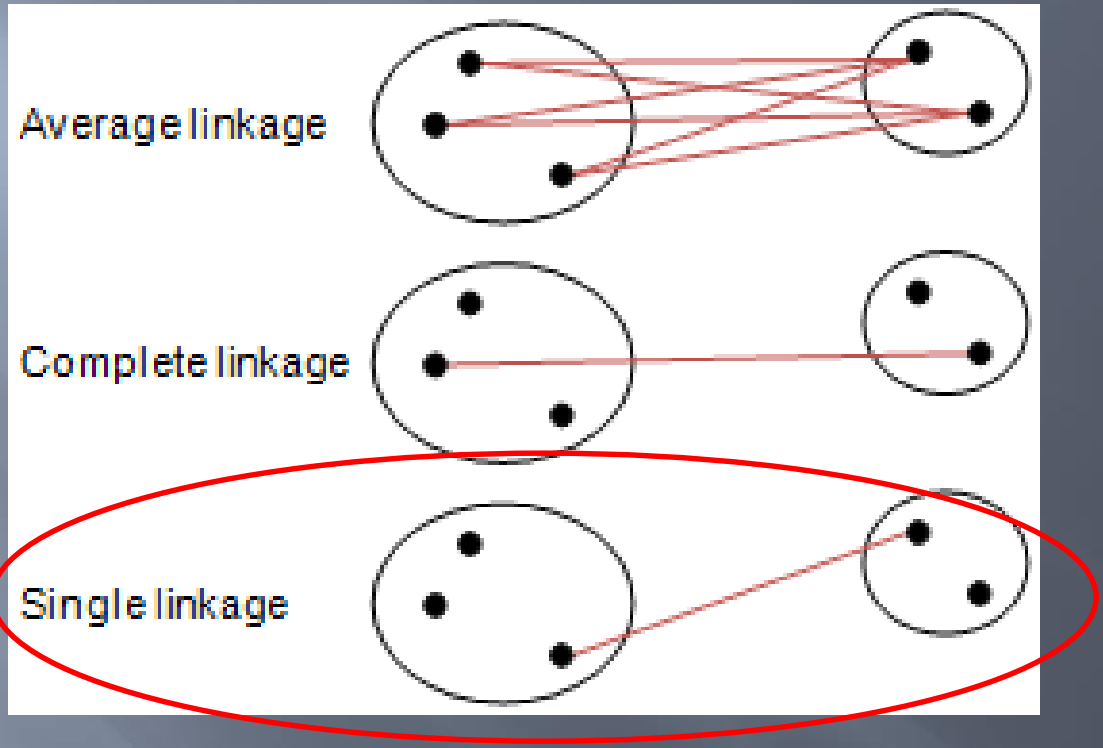
Link two closest clusters. (Distance=1/similarity)

Node distance already defined. How to define distance between clusters? Several ways:

Names	Formula
Maximum or complete linkage clustering	$\max \{ d(a,b) : a \in A, b \in B \}.$
Minimum or single-linkage clustering	$\min \{ d(a,b) : a \in A, b \in B \}.$
Mean or average linkage clustering, or UPGMA	$\frac{1}{\|A\|\|B\|} \sum_{a \in A} \sum_{b \in B} d(a,b).$
Minimum energy clustering	$\frac{2}{nm} \sum_{i,j=1}^{n,m} \|a_i - b_j\|_2 - \frac{1}{n^2} \sum_{i,j=1}^{n} \|a_i - a_j\|_2 - \frac{1}{m^2} \sum_{i,j=1}^{m} \|b_i - b_j\|_2$

Wikpedia

Modular structure



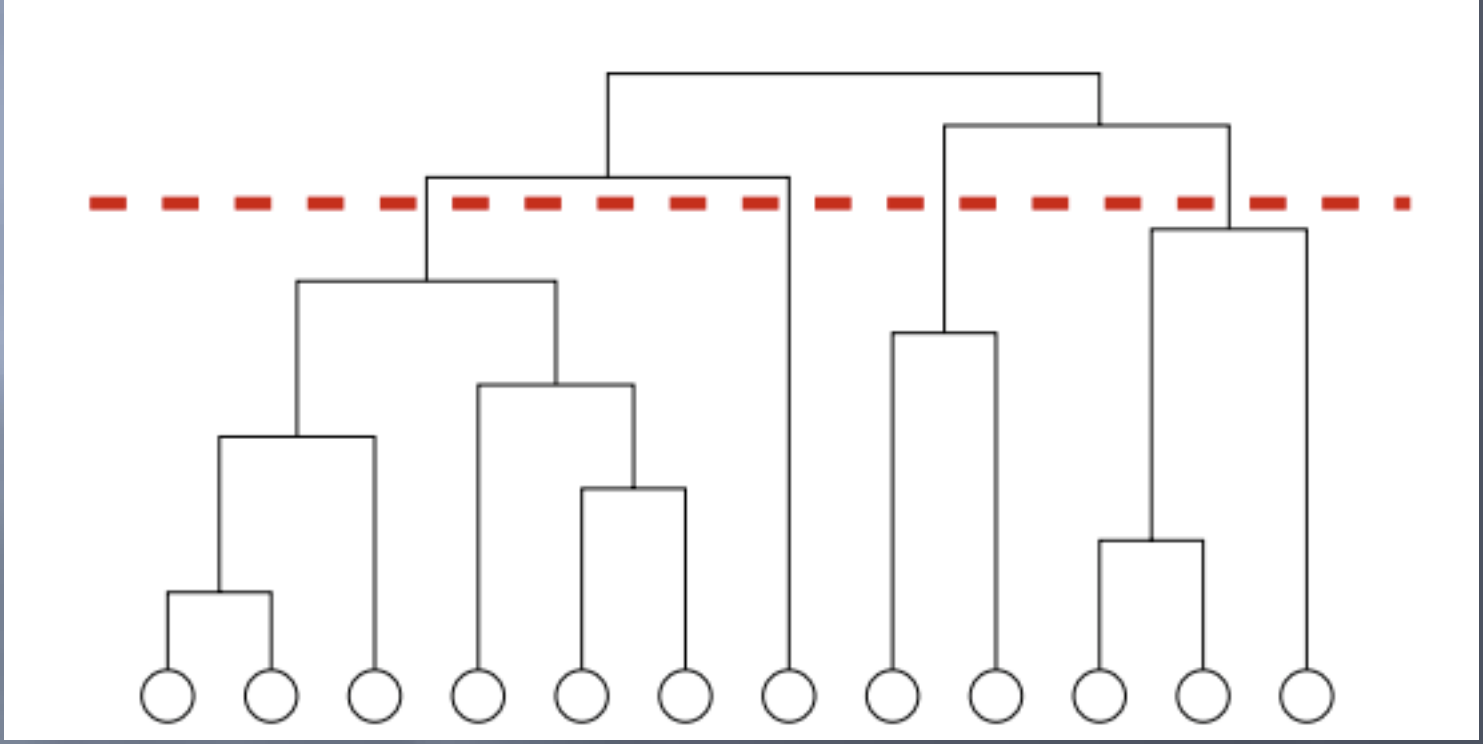
Mostly used: Single linkage

The NG method based on betweenness is such

Modular structure

In the original single linkage method there is no *Q*

The result is a dendrogram:



Depending on the cutting level one gets modules

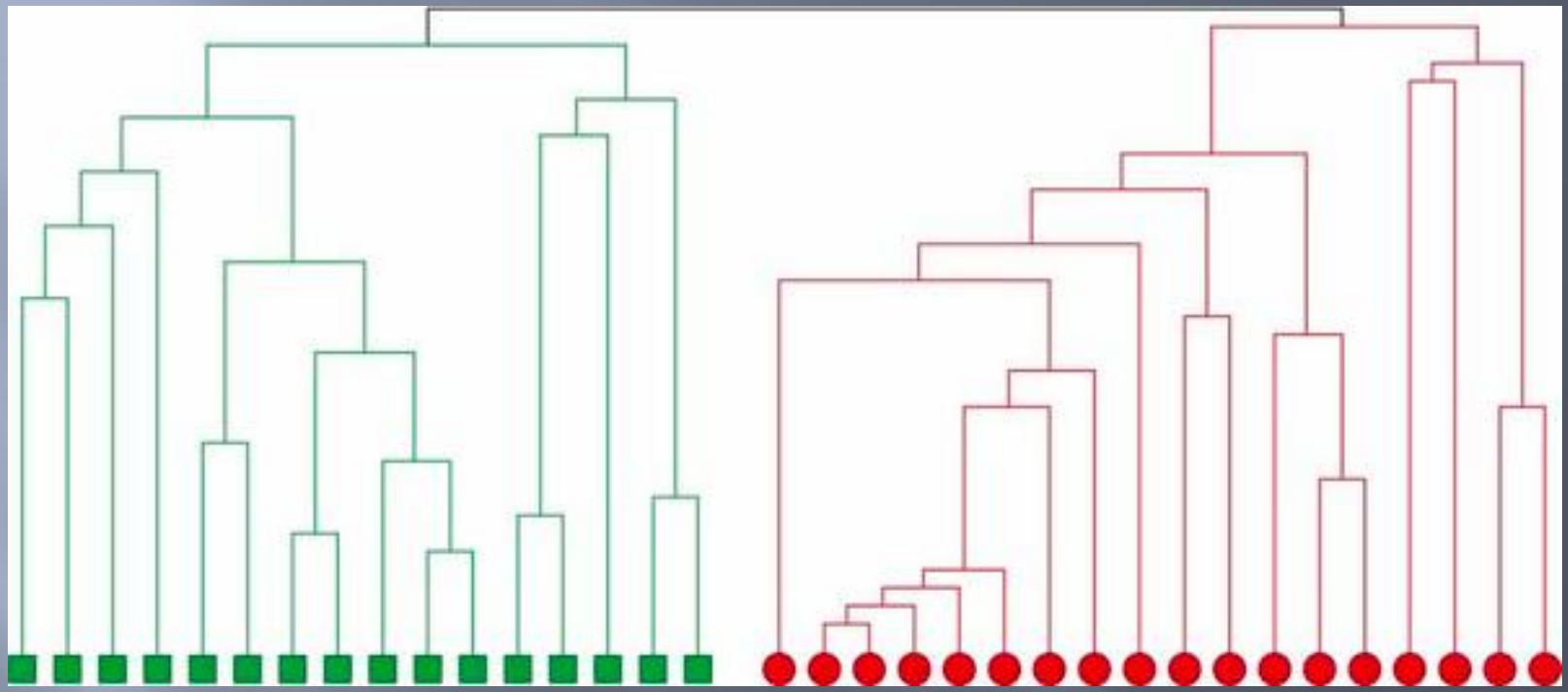
Modular structure

Single linkage

Advantages:
Relatively simple
Fast algorithms (if similarity given)
Number of modules controlled
Hierarchical relationship automatically

Disadvantages:
No *a priori* cutting level
Meaning of clusters unclear
Hierarchy can be artificial
Important links may be missed

Modular structure



Dendrogram of the Zachary karate club
Similarity: Intimacy as declared by the participants

The two major clusters correspond to the splitting

Newman 2012

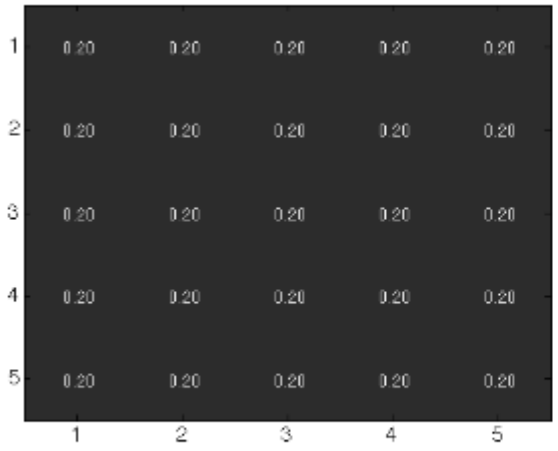
Modular structure

Block models
(Holland et al. Social Networks, 1983)

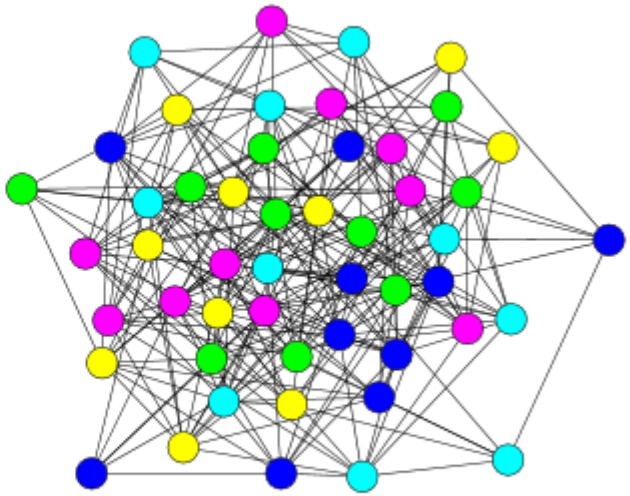
Block models are **generative probabilistic models** which can produce networks with give properties (similarly to configuration networks). Here modular properties.

Task: Construct a networks with *N* nodes ordered forming a **fixed number of** *S* modules. There is an *S*X*S* matrix with elements $M_{s,z}$ showing the probability of having a link between modules *s* and *z*. Nodes within a module are stochastically equivalent.

Block models

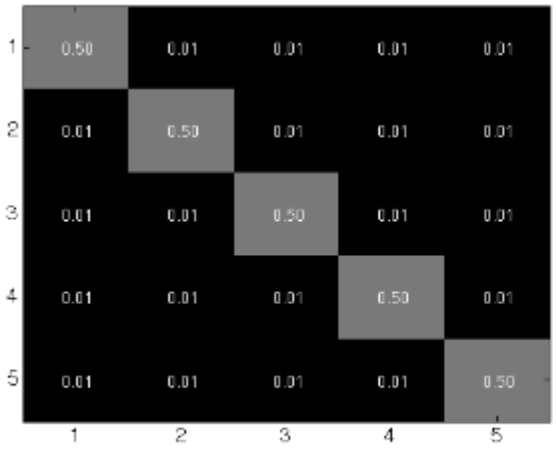


stochastic block matrix

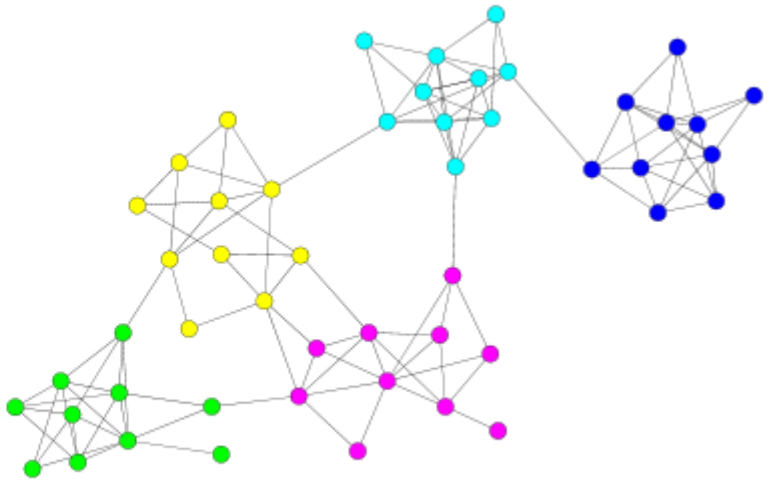


random graph

ER

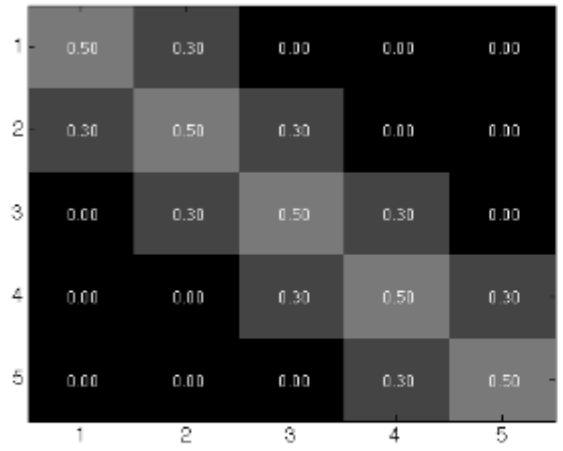


stochastic block matrix

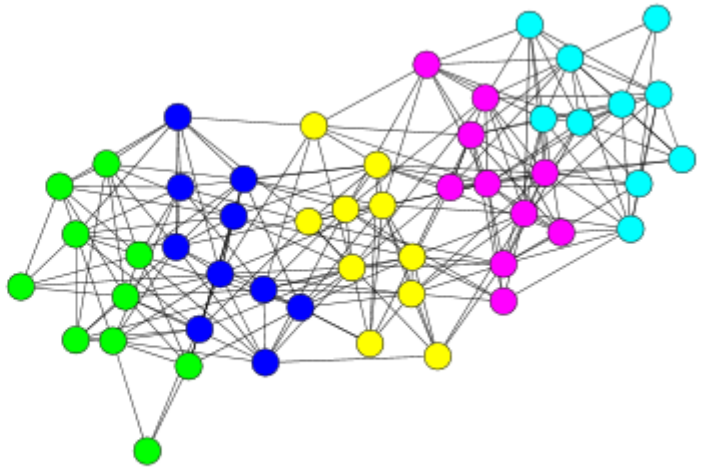


communities

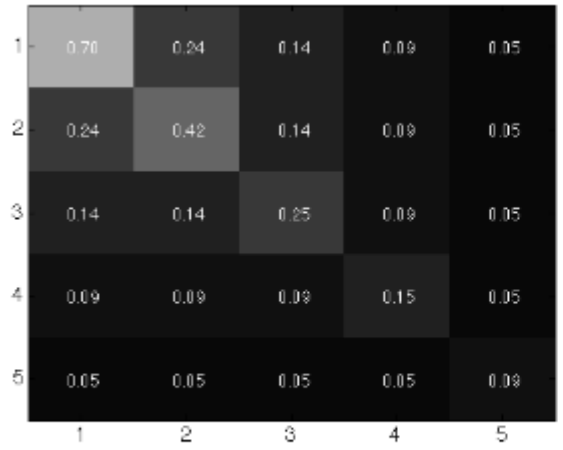
Block models



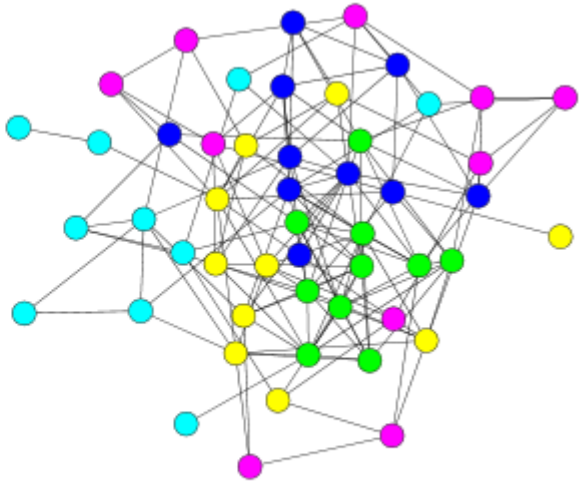
stochastic block matrix



ordered communities



stochastic block matrix



core-periphery structure

Flexible method

Modular structure

Block models define an ensemble. One nw is a sample.

For finding communities we have to revert the method.
Q: What would be the block model, for which the empirical network could be a sample.
We ask for *S* and *M* (statistical inference).

Maximum likelihood: We choose *S* and *M* such that probability of generating the empirical nw will be max.

The probability (likelihood) that the empirical model is created this way is:

$$\mathcal{L}(\mathrm{G}|S,M) = \prod_{i<j} M_{s_i,s_j}^{A_{ij}} \left(1 - M_{s_i,s_j}\right)^{1-A_{ij}}$$

Where s_i is the community, which node *i* belongs to.

Modular structure

We are interested in the maximum of this probability (maximum likelihood). Simplification using stoch. equiv.:

Let the number of nodes in module s be N_s . Then the number of possible links between modules s and z is

$$N_{s,z} = \begin{cases} N_s N_z & \text{if } s \neq z \\ N_s(N_s - 1)/2 & \text{if } s = z \end{cases}$$

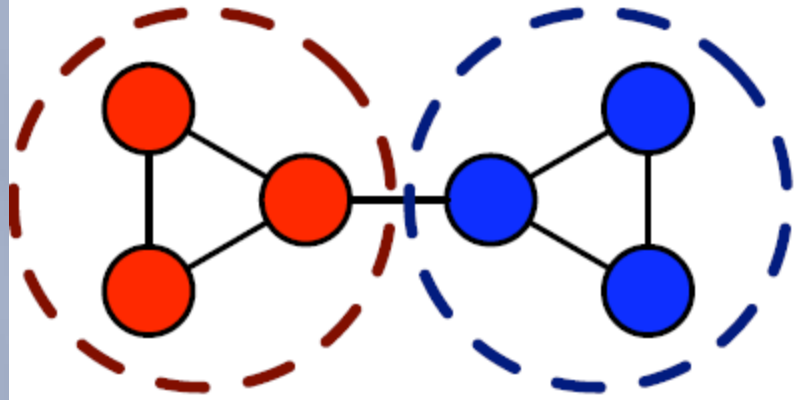
If the number of empirical links between s and z is $E_{i,j}$

$$\mathcal{L}(G|S,M) = \prod_{s,z} \left(\frac{E_{s,z}}{N_{s,z}} \right)^{E_{s,z}} \left(1 - \frac{E_{s,z}}{N_{s,z}} \right)^{N_{s,z} - E_{s,z}}$$

Depends on the partition

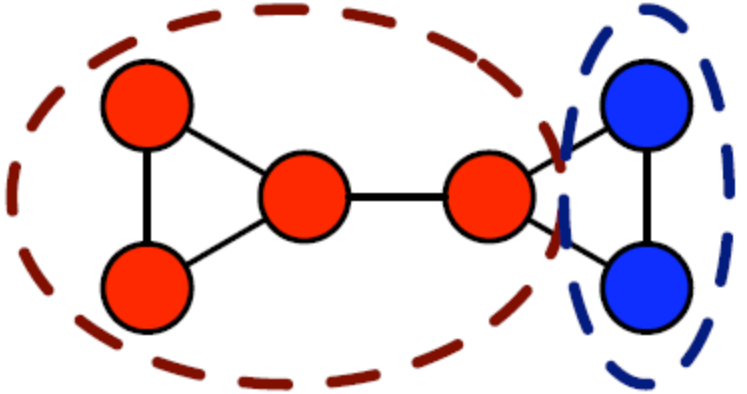
This is another optimization method, where the variables are the N_s –s (S is fixed.)

An example:



$$\mathcal{L}_{\text{good}} = 0.043304\ldots$$
$$\ln \mathcal{L}_{\text{good}} = -3.1395\ldots$$

M_{good}	red	blue
red	3/3	1/9
blue	1/9	3/3



$$\mathcal{L}_{\text{bad}} = 0.000244\ldots$$
$$\ln \mathcal{L}_{\text{bad}} = -8.3178\ldots$$

M_{bad}	red	blue
red	4/6	2/8
blue	2/8	1/1

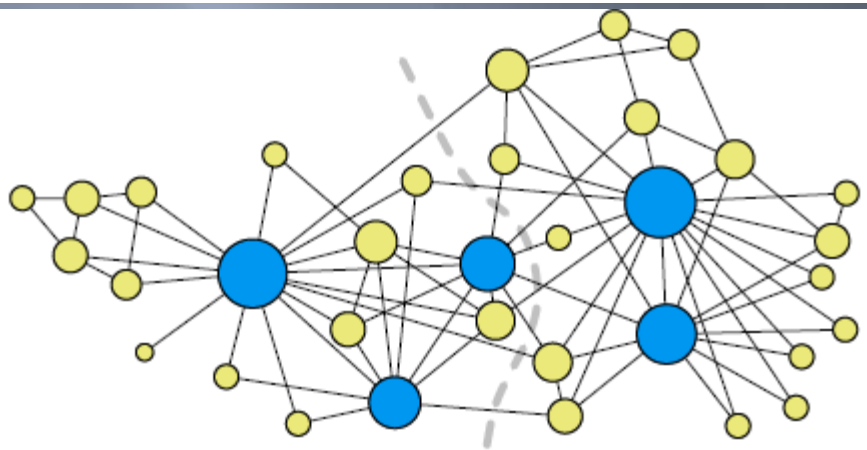
By increasing *S* we have more parameters and the fit becomes easier → for *S* = *N* we have *M*=*A*, with perfect fit. Some knowledge or principle is needed.

A. Clauset

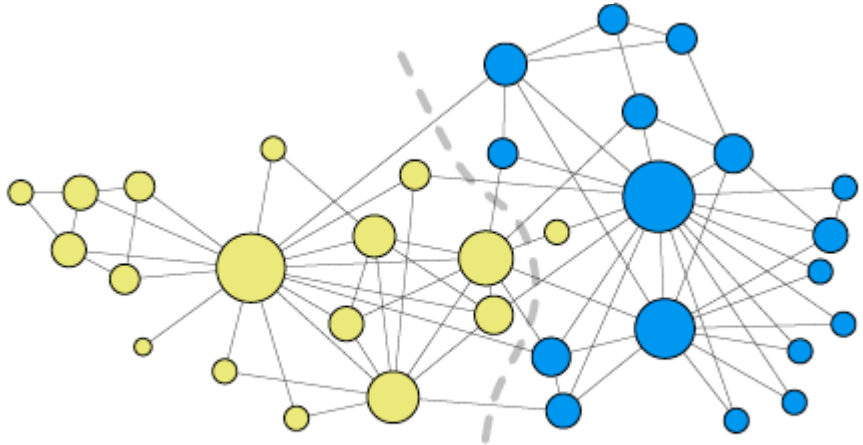
For broad degree distributions the assumption of stochastic equivalence fails. The block model will try to group high degree nodes and low degree nodes together, irrespective of the modular structure. *Degree corrected stochastic block model* introduces new variables for the degree and makes the optimization by taking them into account (κ_s degree in bl. s).

$$\ln \mathcal{L}(\mathrm{G}(\mathrm{k}_i)|S,M) = \sum_{s,z} \frac{E_{s,z}}{2L} \ln \frac{E_{s,z}/2L}{(\kappa_s/2L)(\kappa_z/2L)}$$

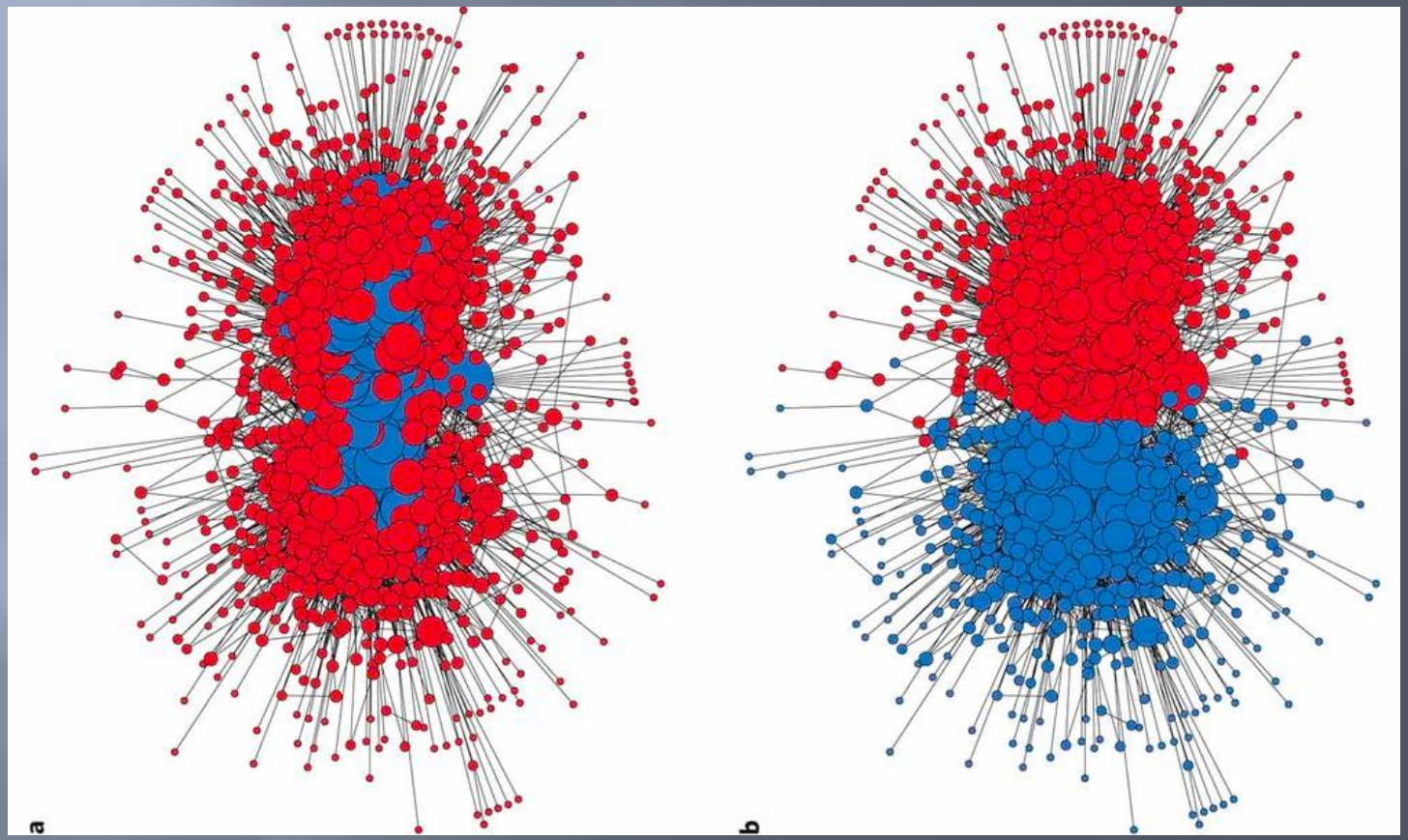
Karrer, Newman
Phys Rev. E, 2010



karate club, with SBM $S = 2$



with degree correction



Political blogs in the US a) simple stochastic block model b) degree corrected block model leads to the expected liberal-conservative splitting

Adamic, L. A. & Glance, N. 2005

Modular structure

Local methods

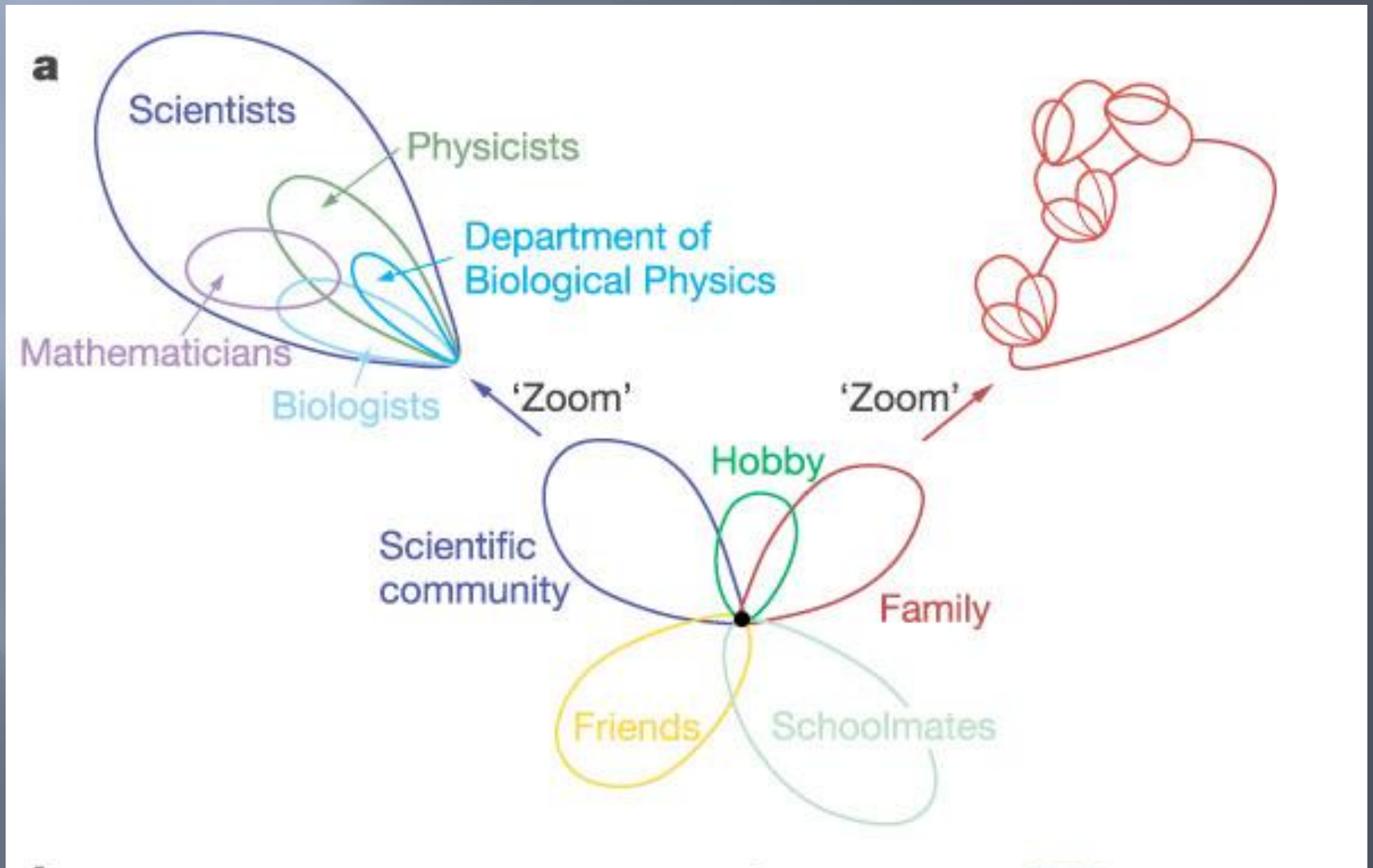
We may be interested in the question: given a node, which community does it belong to?

We can start from the definition:

Strong community: Each node has more neighbors inside than outside
Weak community: Total degree within the community is larger than the total degree out of it.

Modular structure

Nodes can belong to more than one modules!
Communities overlap

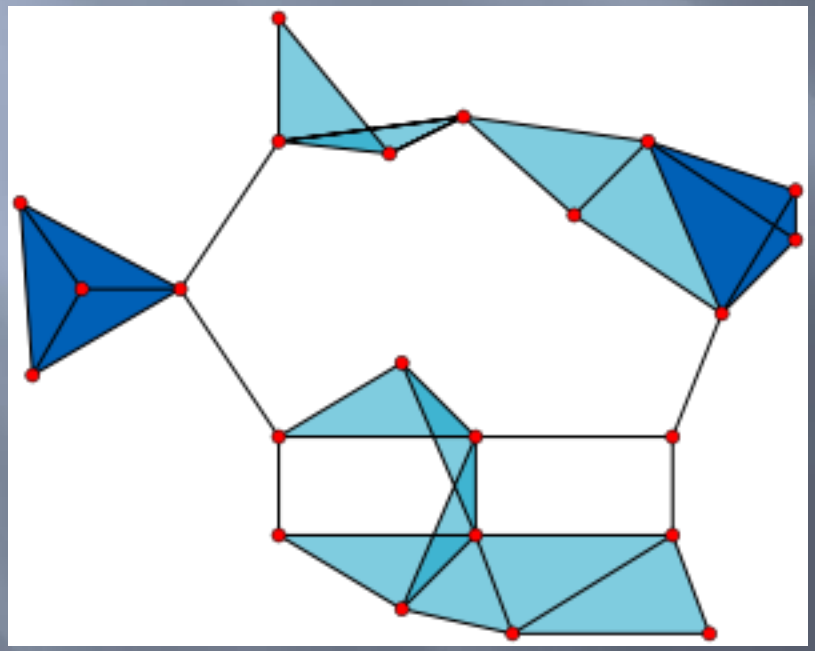


Palla et al 2005

Modular structure

k-clique percolation

k-clique: fully connected subgraph



3-cliques: 19
4-cliques: 2

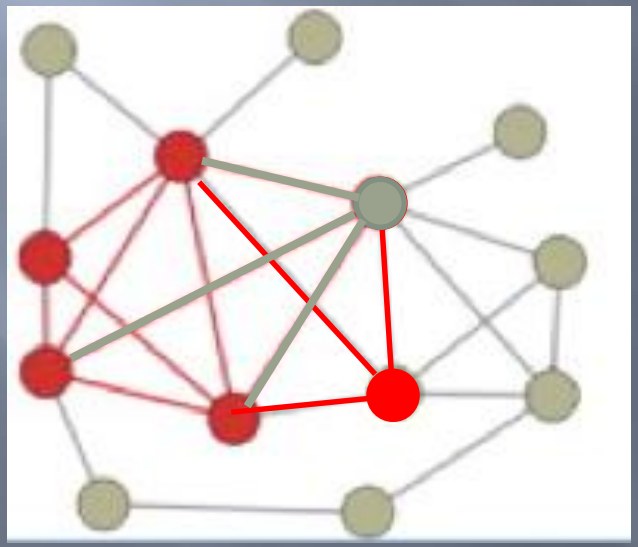
All subset of *k'*<*k* nodes in a *k*-clique is part a *k'*-clique

Modular structure

A clique is naturally considered to be a community (has been used as such in sociology)

In a *k*-clique, we have *k* different (*k*–1)-cliques.

A *k*-clique can be connected to another *k*-clique through a common (*k*–1)-clique:

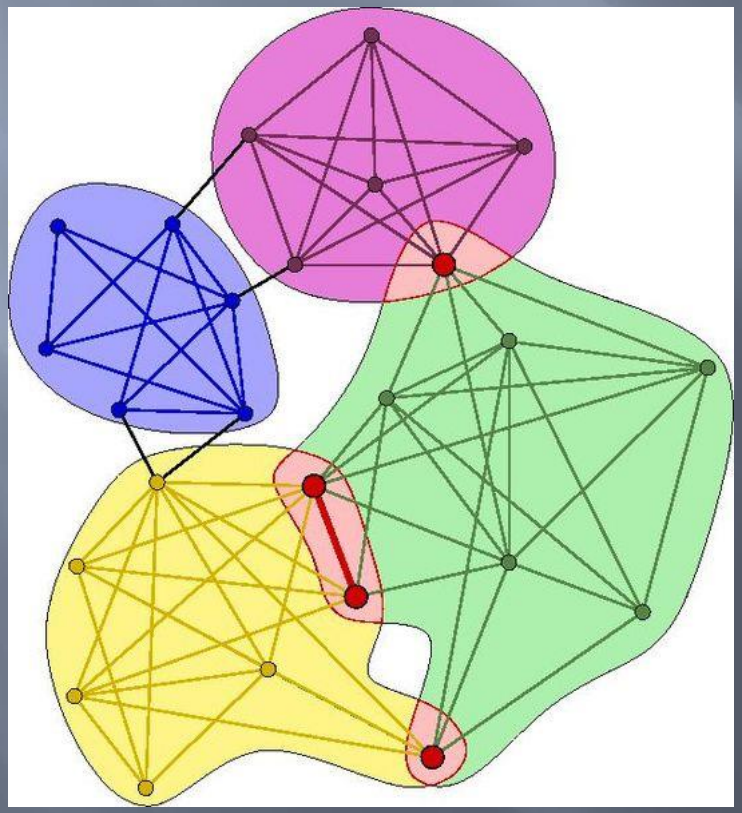


percolation unit: *k*-clique
connection: shared (*k*–1) clique
community: *k*-clique-cluster

“Rolling” the triangle: 4-clique cluster

Modular structure

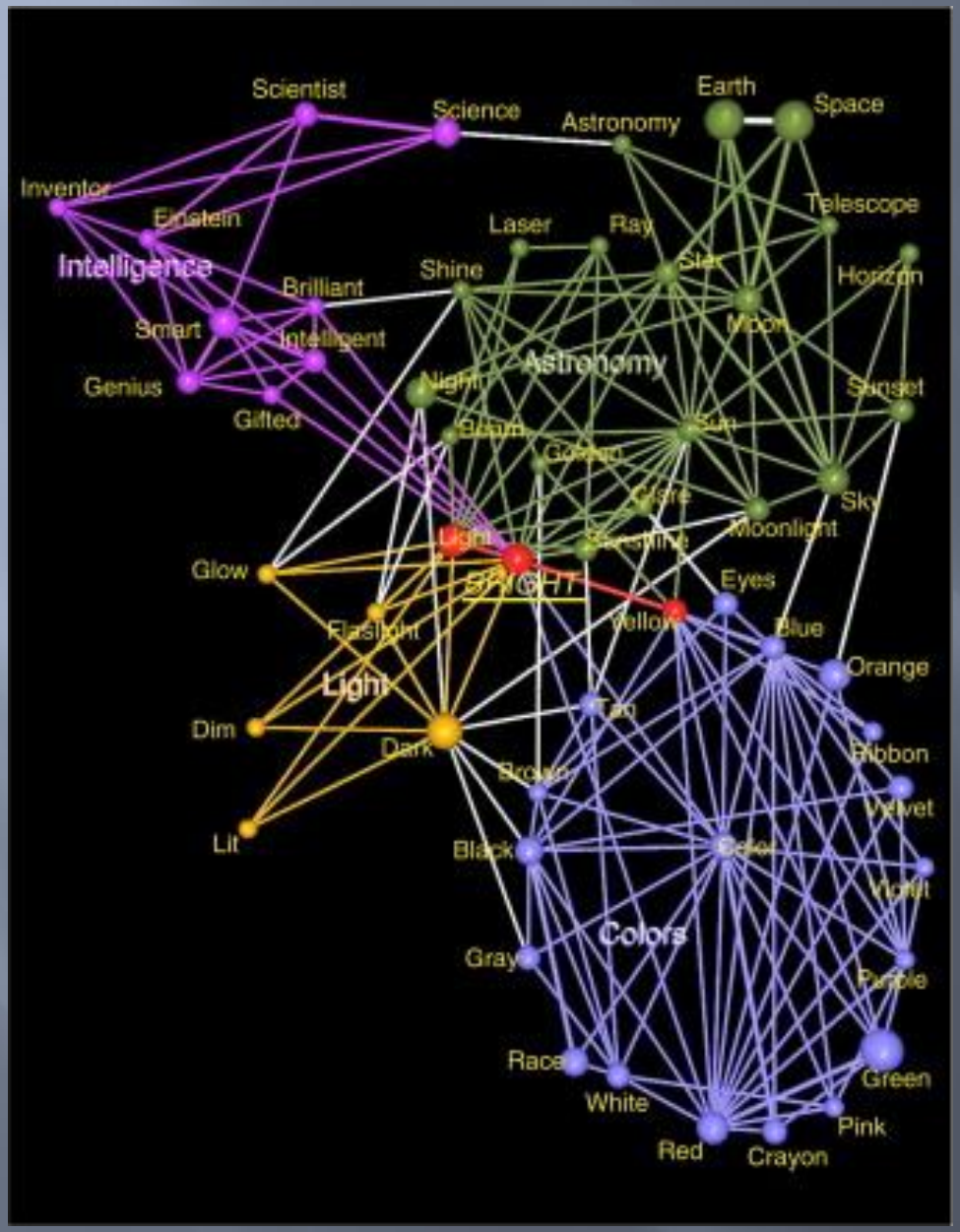
A *k*-clique community is defined as the union of nodes, which is constructed by starting from a *k*-clique and joining nodes such that they are connected to the a *k*-clique of the already existing community by at least *k*–1 links.



Construction of
4-clique communities

There are nodes, which belong to more than one community:
overlap
k is a parameter

Modular structure

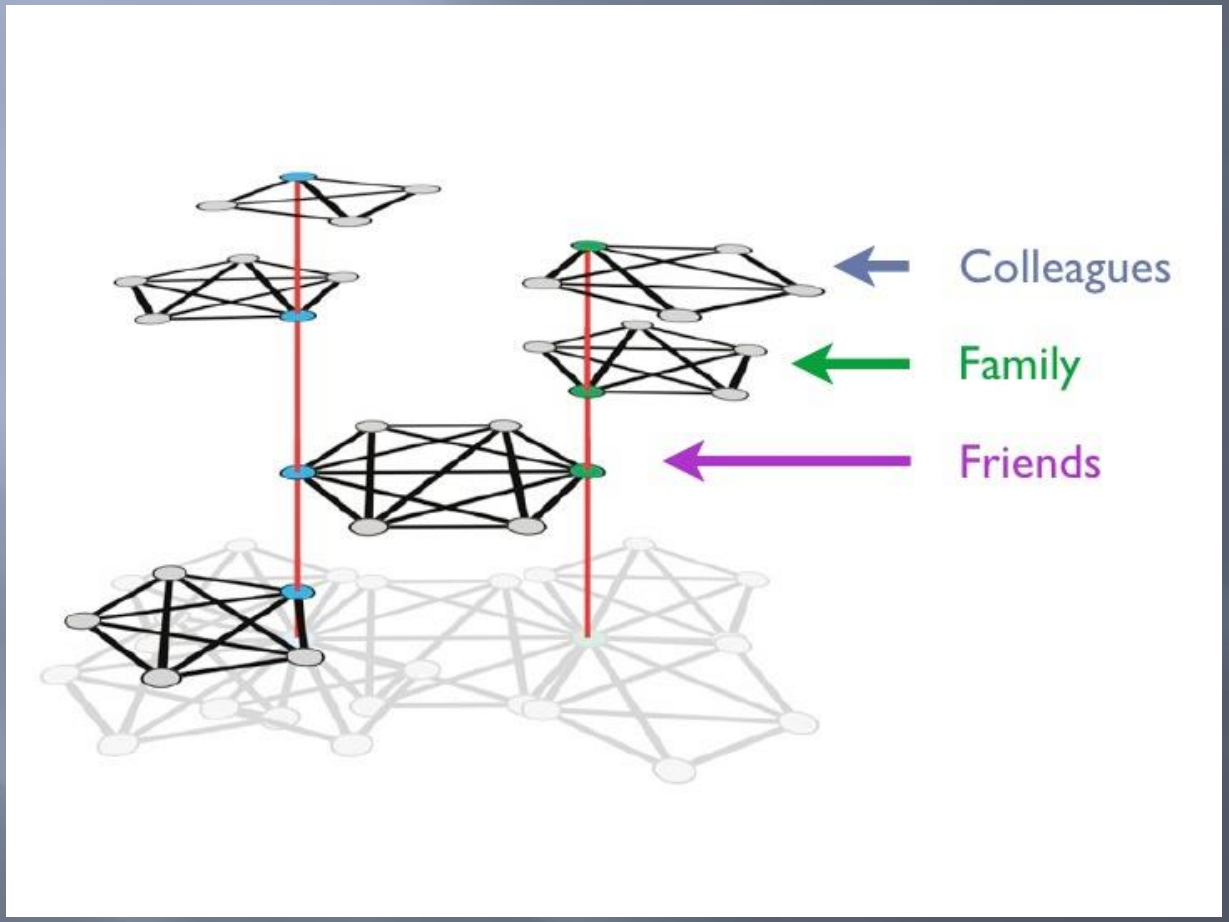


Word association communities as detected by the clique percolation method

Palla et al 2005

Modular structure

Origin of overlap: Multiplex networks



Nodes have multiple roles: overlap

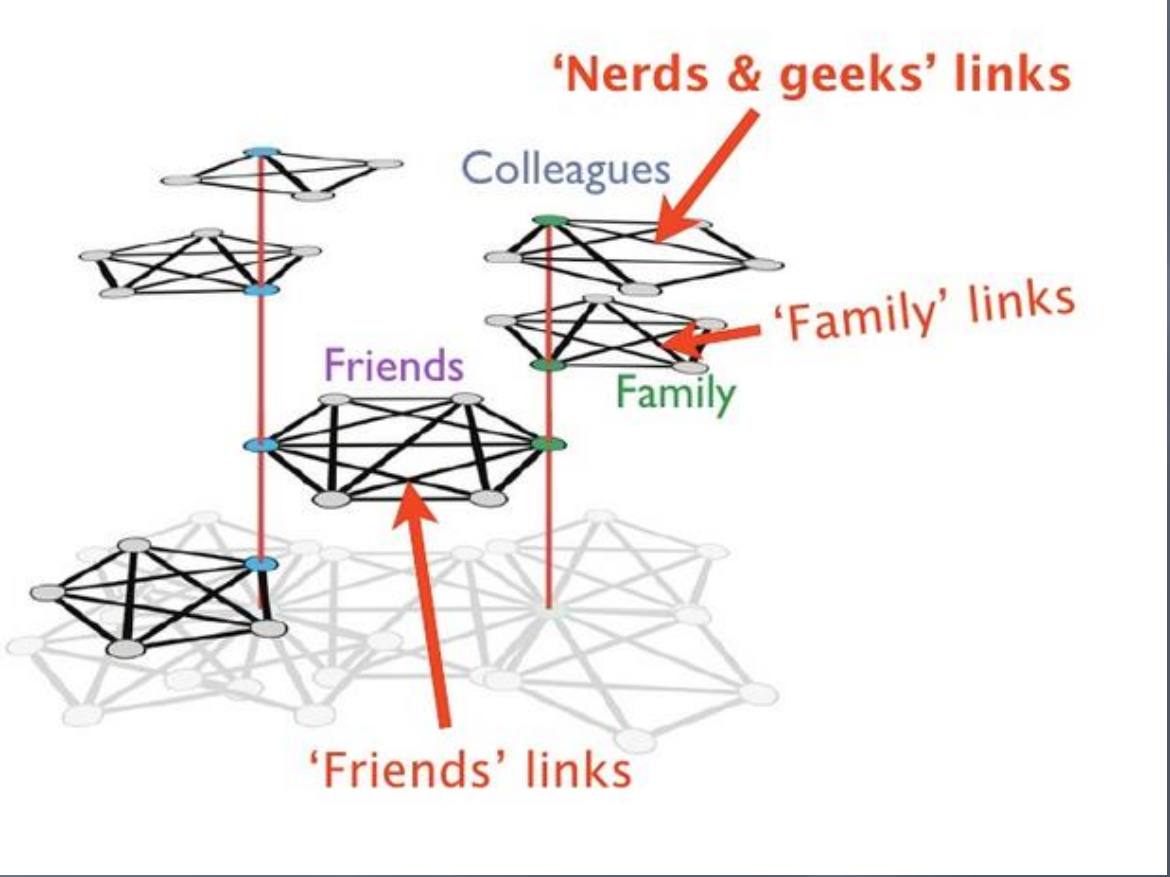
Links are specific to the relationship

Ahn et al. 2010

Modular structure

Community: Group of densely interconnected nodes

Community: Group of similar neighboring links: “link community”



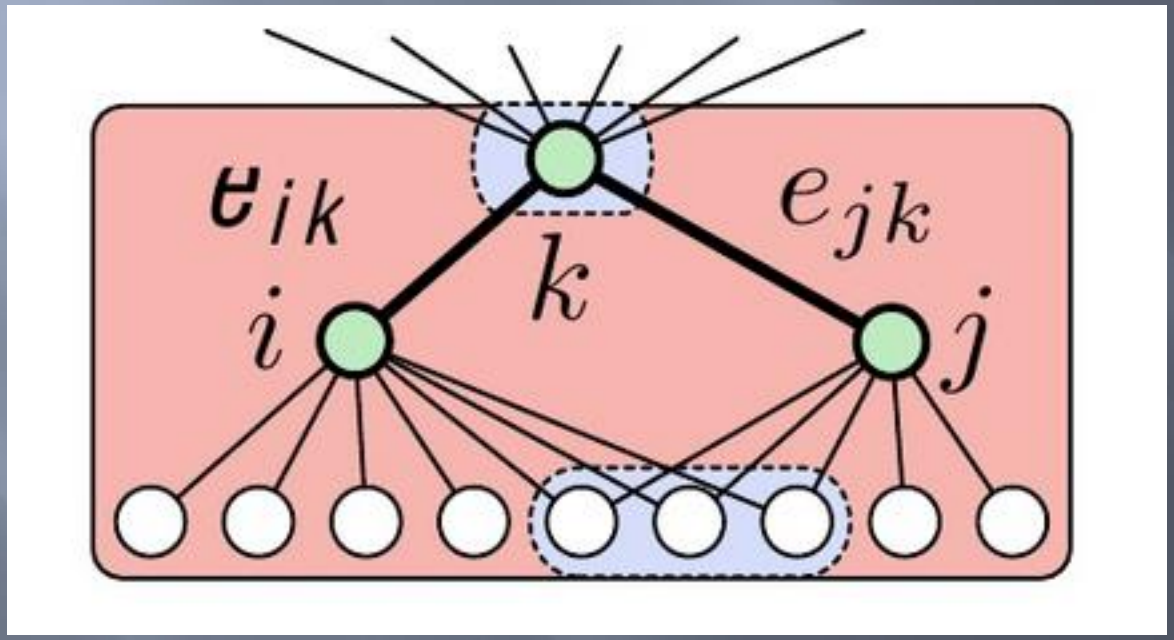
Ahn et al. 2010

Modular structure

Similarity between links e_{ik} e_{jk} :

$$S(e_{ik}, e_{jk}) = \frac{\left|n_+(i) \bigcap n_+(j)\right|}{\left|n_+(i) \bigcup n_+(j)\right|}$$

Where $n_+(i)$ is the set of node i and its neighbors



$$S(e_{ik}, e_{jk}) = \frac{4}{12}$$

Ahn et al. 2010

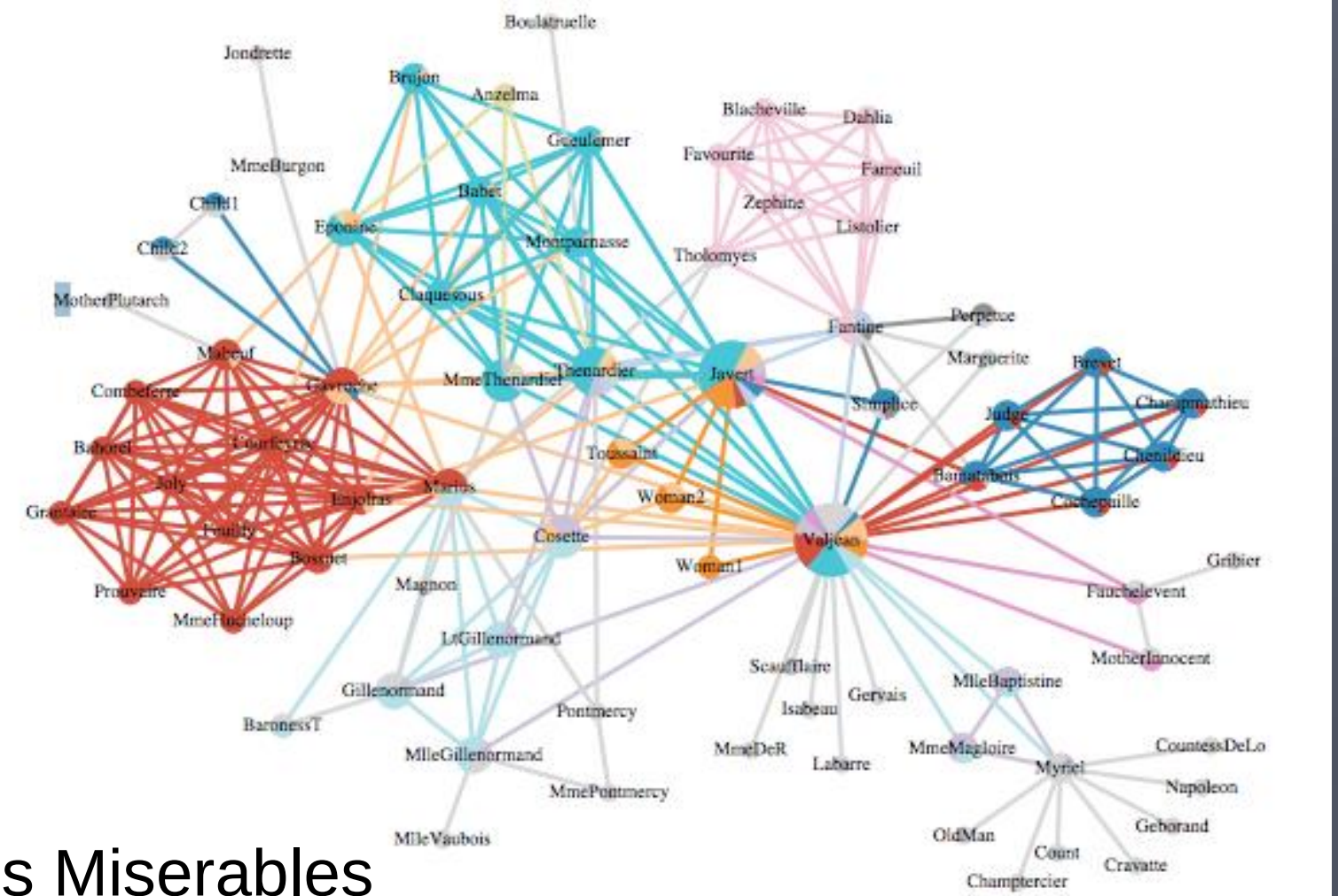
Modular structure

This is a similarity measure for the links. Use hierarchical clustering!

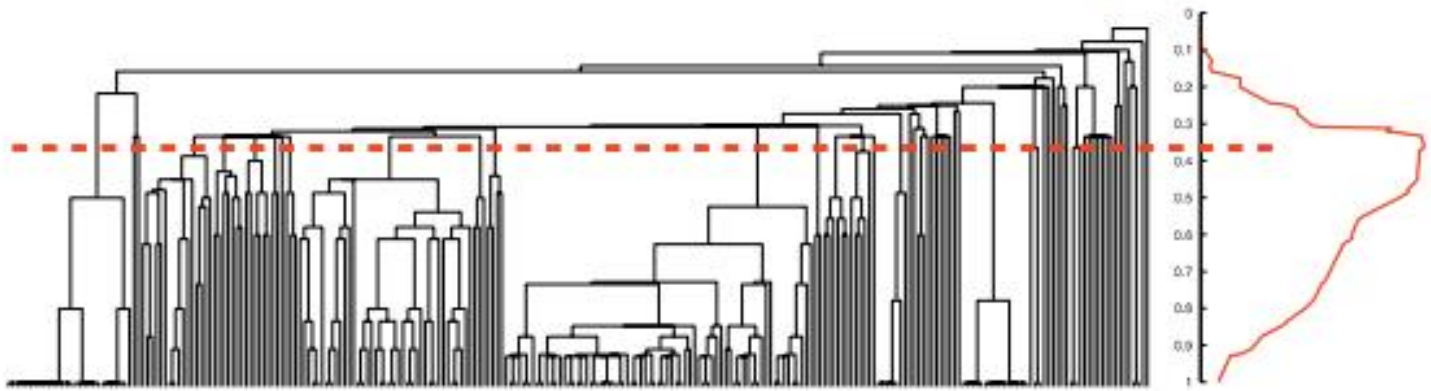
Where to cut the dendrogram?
Maximal “partition density”, i.e., average link density in communities.

This method seems to be able to uncover both hierarchies and overlaps

Ahn et al. 2010



Les Miserables



Ahn et al. 2010

Modular structure

How to decide if a method is good or not?

Testing community detection methods

Comparing particions (computer sci.)

- Normalized mutual information
- Jaccard index

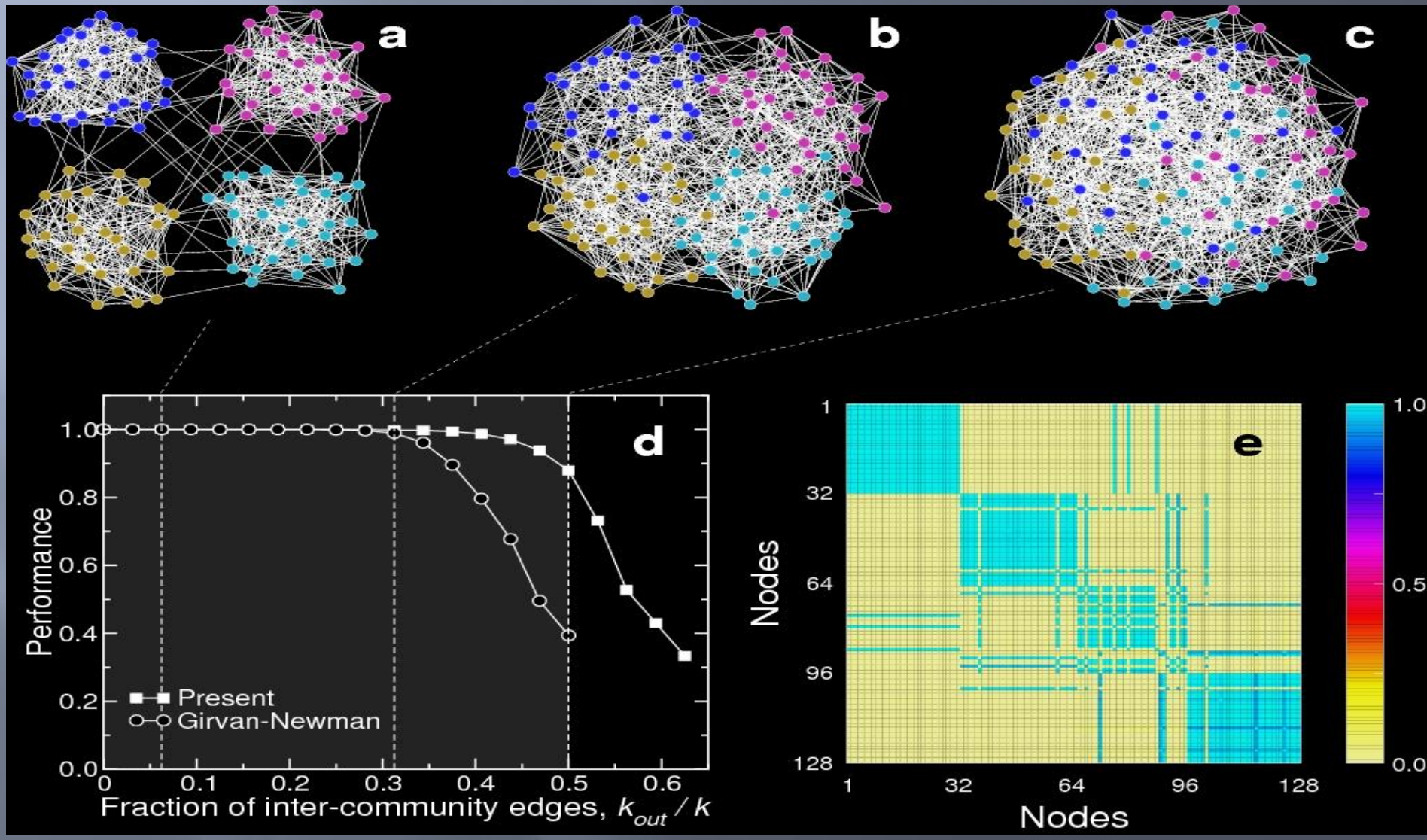
$$I_J(A,B) = \frac{n_{11}}{n_{11} + n_{01} + n_{10}}$$

n_{11}: # pairs in same cluster in A,B
n_{01}: # pairs in same cluster in A
n_{10}: # pairs in same cluster in B

Banchmarks: Control the modular charakter and test how the method works

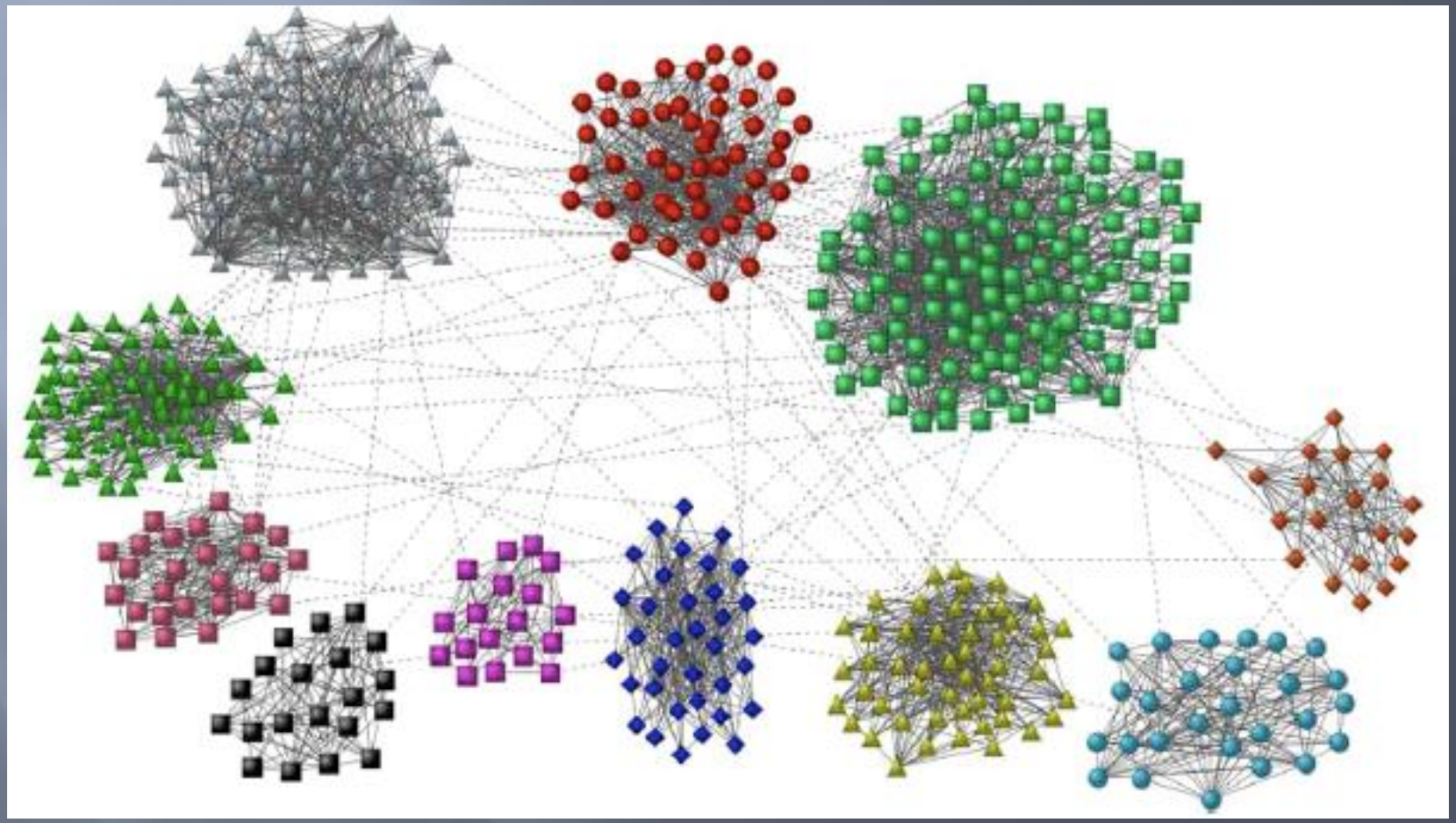
Modular structure

Girvan Newman 2004



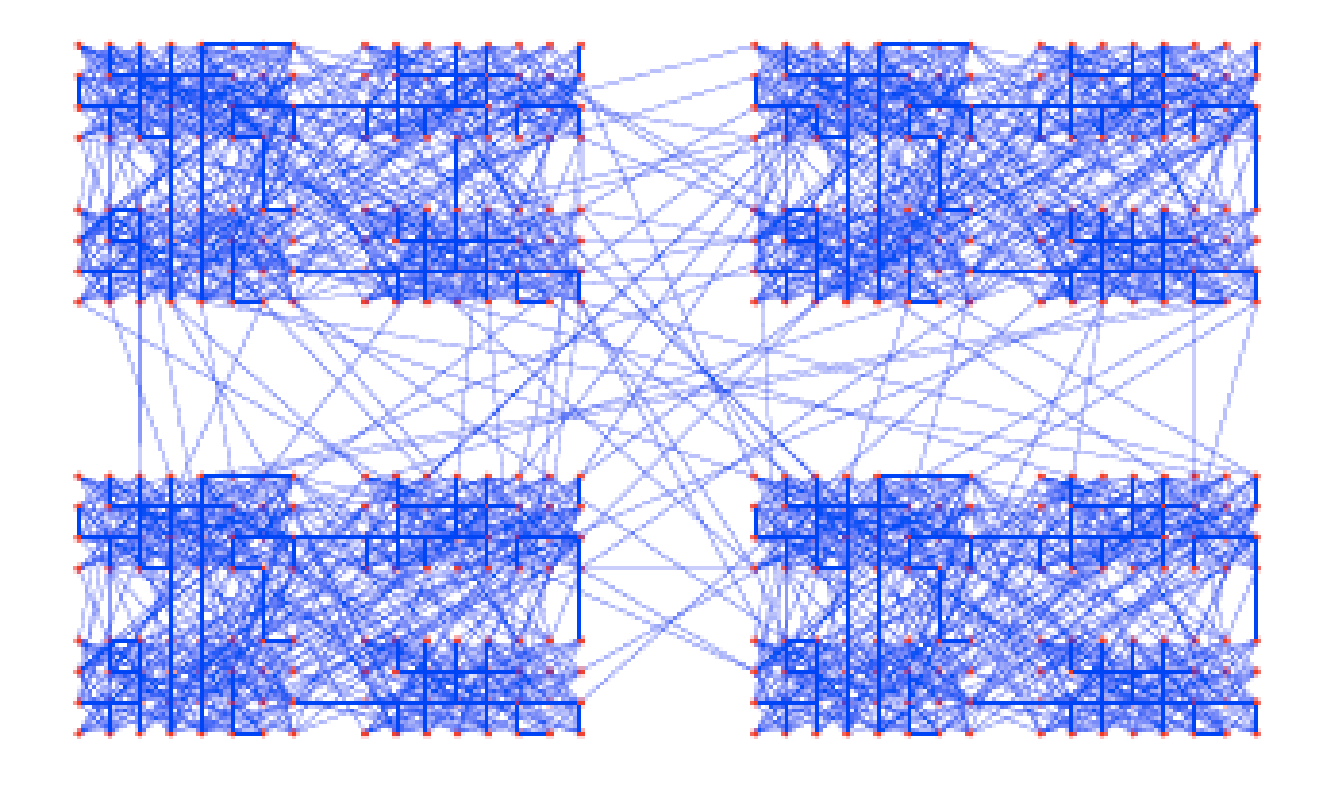
From Fortunato (2009)

Distribution of module sizes (resolution test)



A. Lancichinetti, S. Fortunato, F. Radicchi

Hierarchical structure



Final remarks

Topology, function (and weights) are closely related

Challenge: Identifying functional units from topology

„Ill posed problem”

Criteria, hierarchies, overlaps

Many methods: Which one is good, better, best

Important viewpoint: Computational efficiency

NO BEST METHOD! ADJUST METHOD TO THE PROBLEM!
Review: S. Fortunato, Phys. Rep. 2010
Update: S. Fortunato – M. Hirc: Physics Reports 2016

Summary

Mesoscale structures are expected to reflect the function of the complex system

Motifs can be ordered into classes according to the system categories

Egocentric networks are crucial for characterization of nodes

Community (strongly wired parts of the networks): challenge to detect

Partitions: modularity (efficient heuristics, problems with resolution/hierarchy)

Agglomerative clustering (similarity measure needed, where tom cut?)

Block modeling: Generative approach – maximum likelihood parameter determination. Fixed number of moduls (but T. Peixoto!)

Problem of overlapping communities: clique percolation and link communities

Homework

Analyze the email dataset from Alex Arenas' site:
http://deim.urv.cat/~alexandre.arenas/data/welcome.htm

Compare the community structures as they result from the modularity optimization algorithm (take, e.g., the Louvain algorithm,
https://pypi.python.org/pypi/louvain)
and the link clustering algorithm by Ahn et al.:
http://barabasilab.neu.edu/projects/linkcommunities/