University of Stuttgart
Germany

Complex Network Systems

Communities

Ilche Georgievski
ilche.georgievski@iaas.uni-stuttgart.de
Room: U38 0.353

2019/2020
Winter

Types

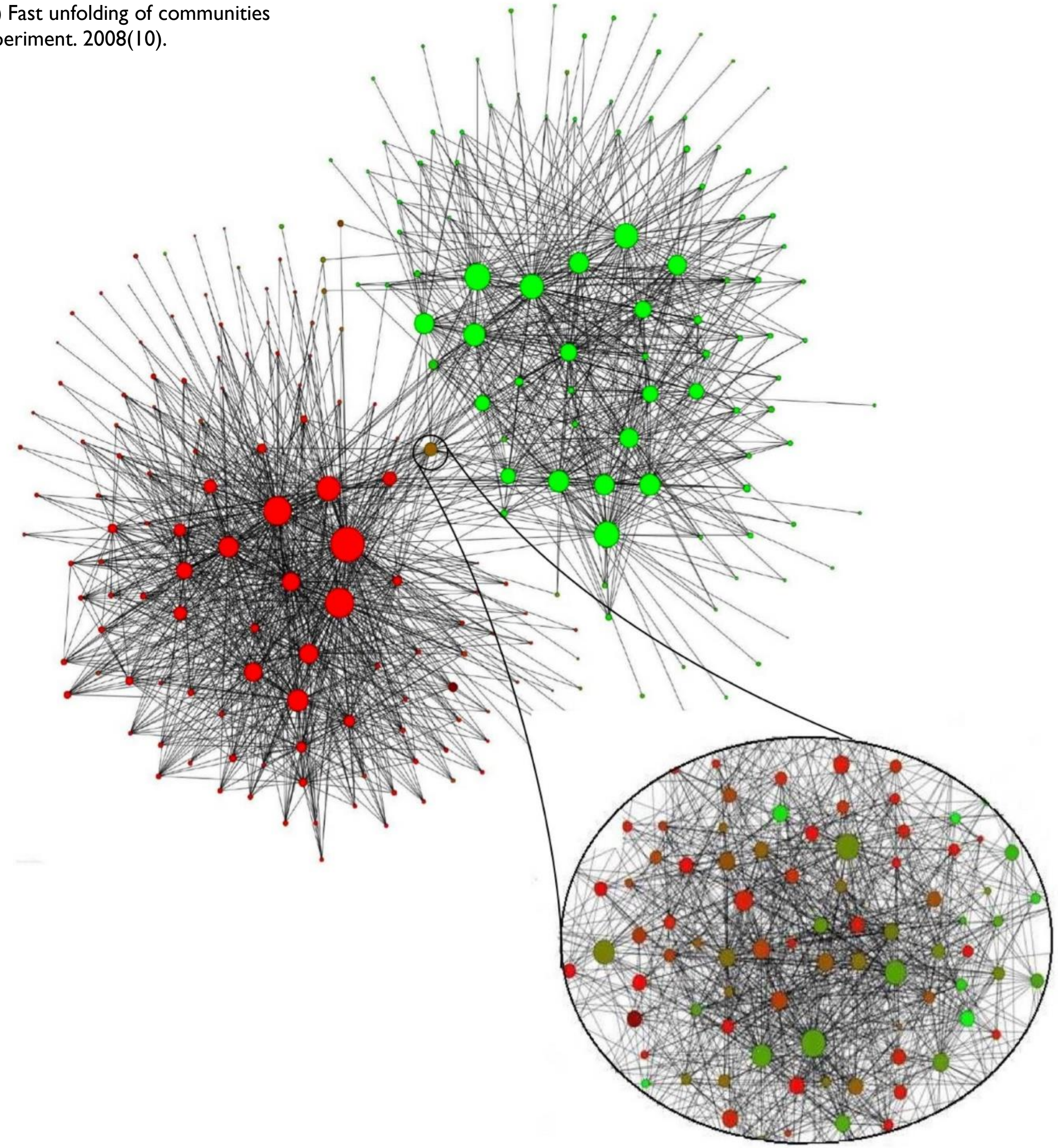
Graph-level metrics

- Size
- Density
- Paths and distances
- Neighbourhoods
- Egocentric network
- Clustering coefficient
- Transitivity
- Cores
- Cliques
- Communities

Node-level metrics

- Closeness centrality
- Betweenness centrality
- Degree centrality
- Eigenvector centrality
- Katz centrality
- PageRank

Blondel, V. D., Guillaume, J., Lambiotte, R., and Lefebvre, E. (2008) Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment. 2008(10).

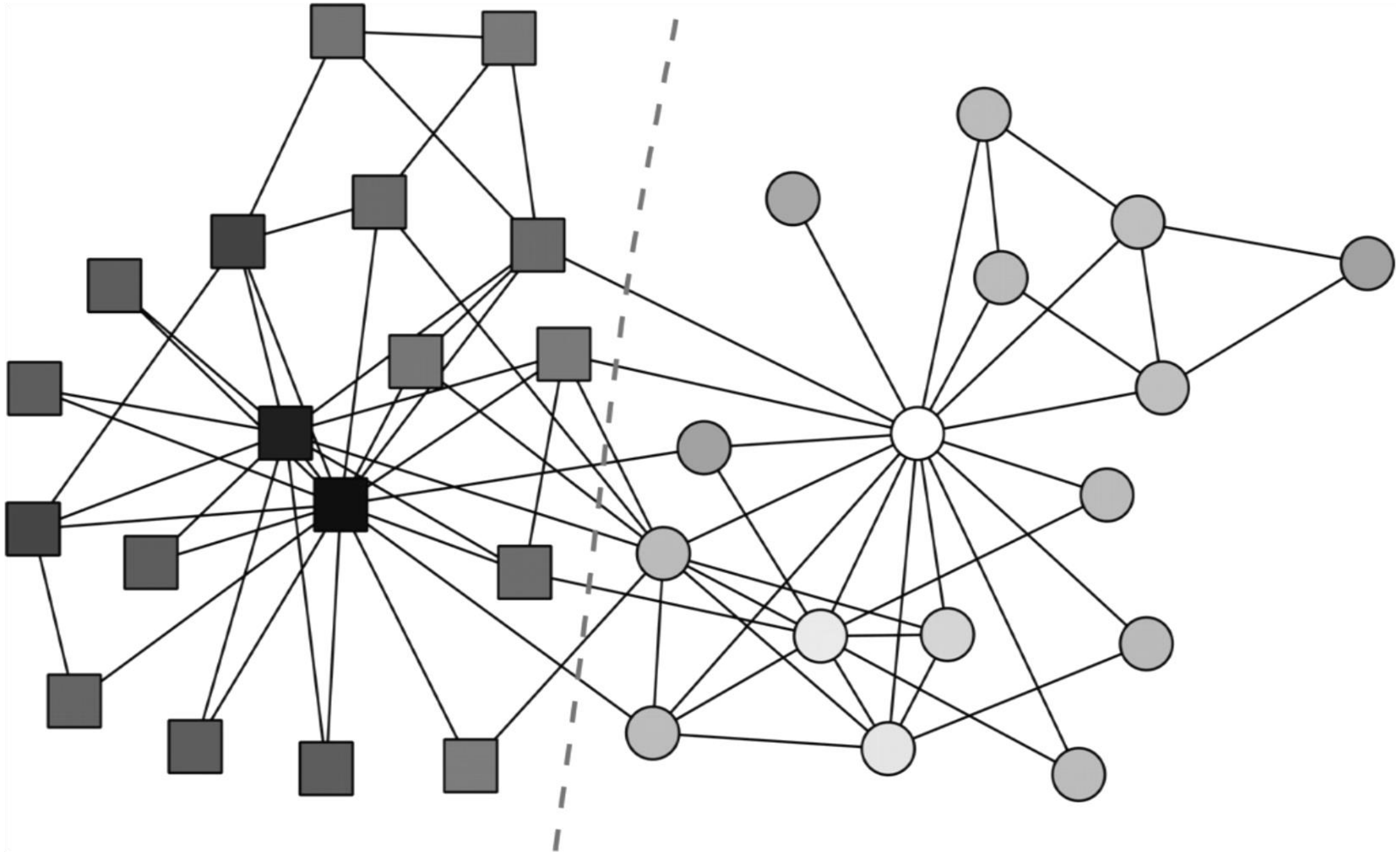


What is a community?

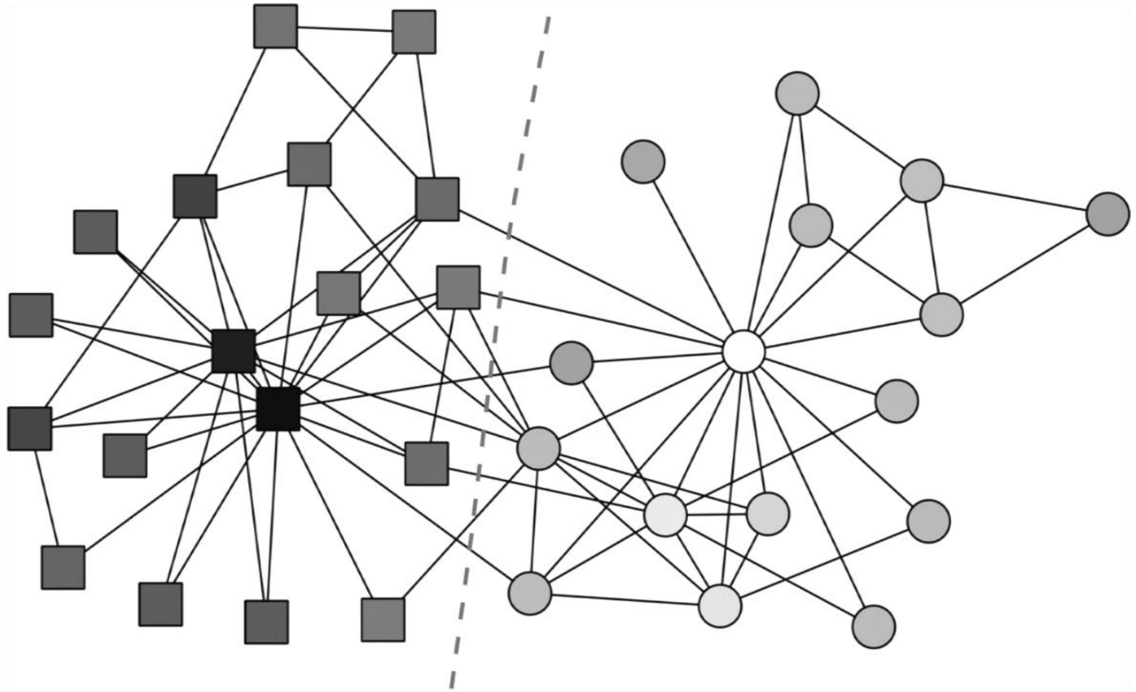
A group of nodes that are densely connected internally and sparsely connected to other groups

Zachary, W. W. (1977) An Information Flow Model for Conflict and Fission in Small Groups, Journal of Anthropological Research, 33(4), pp. 452-473.

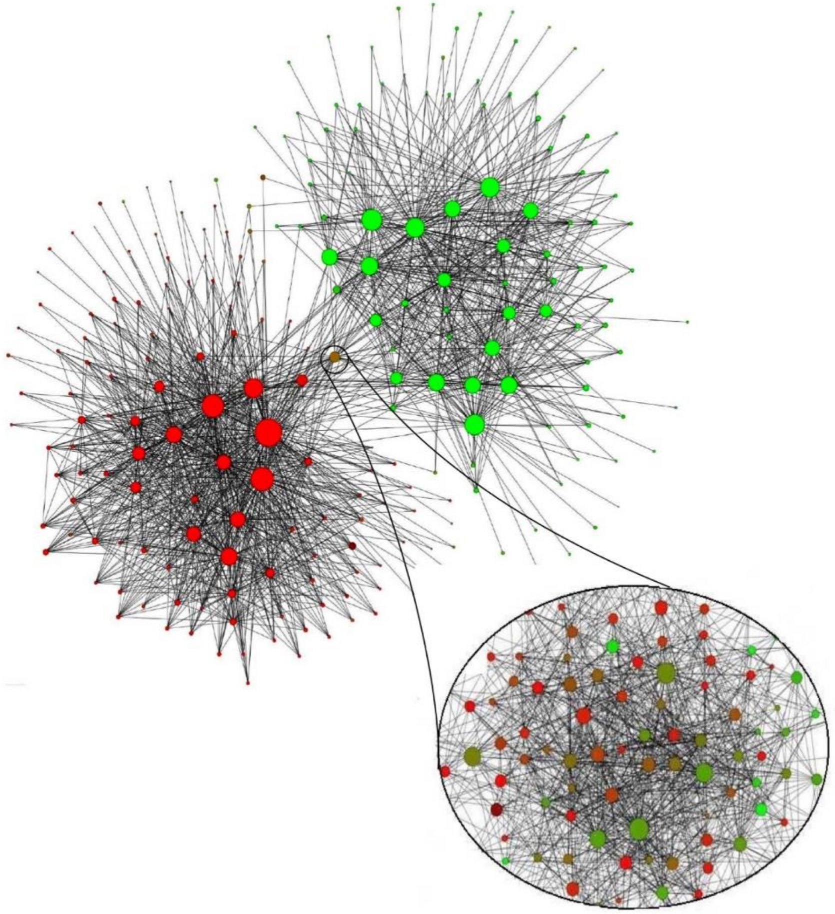
Zachary's Karate Club



Auxiliary information



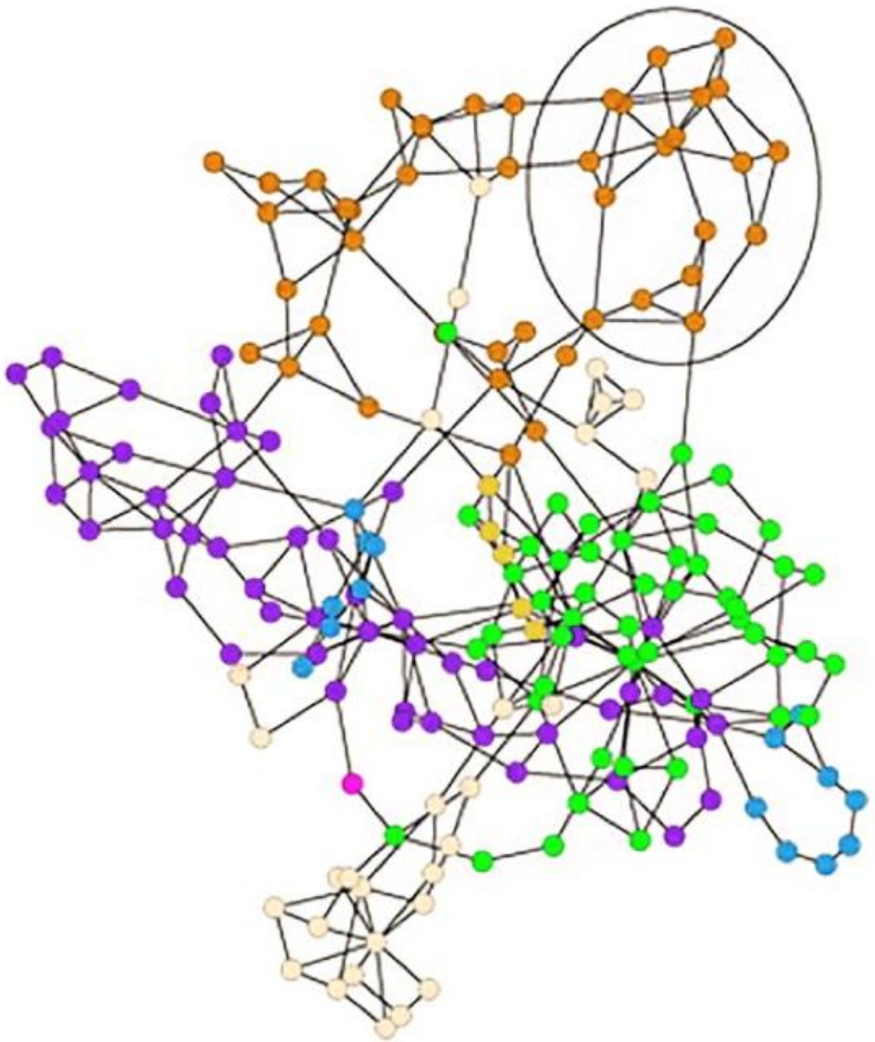
Breakup of the club

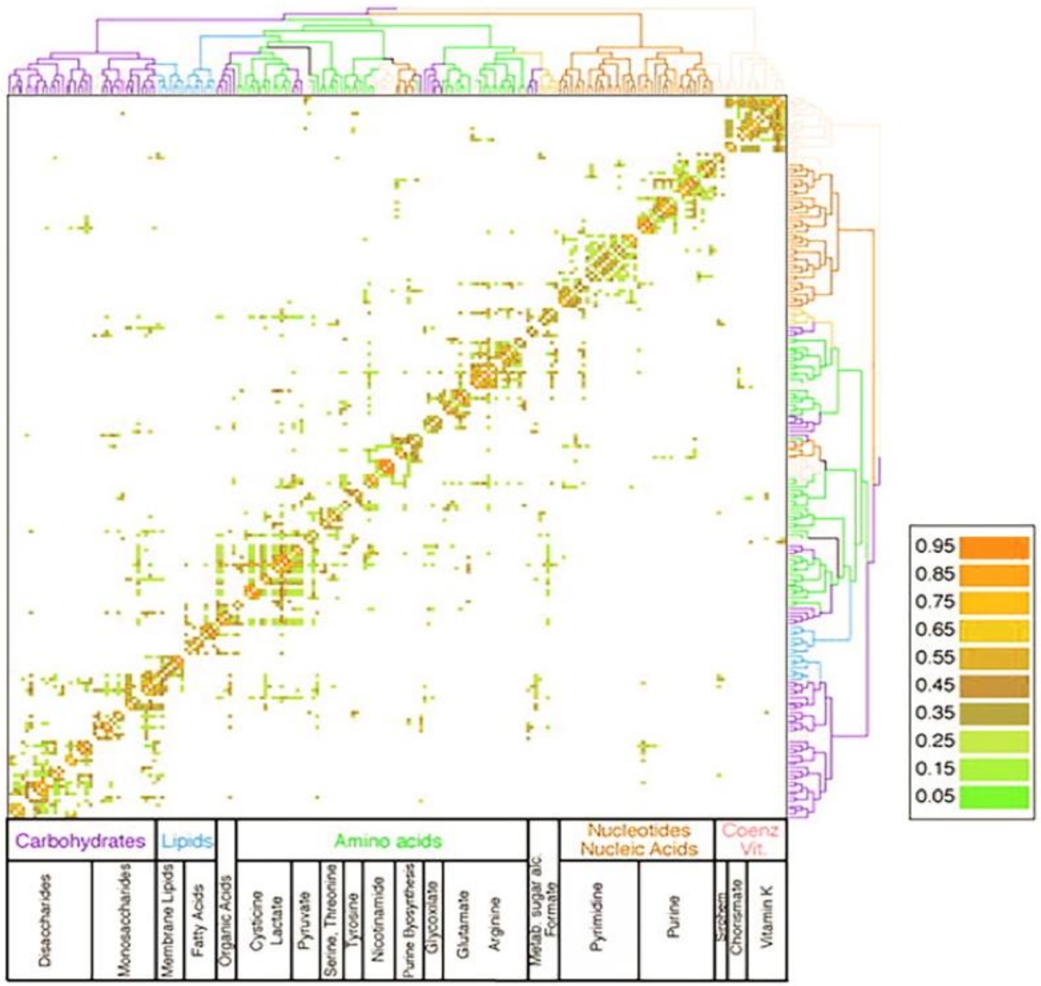


Language spoken

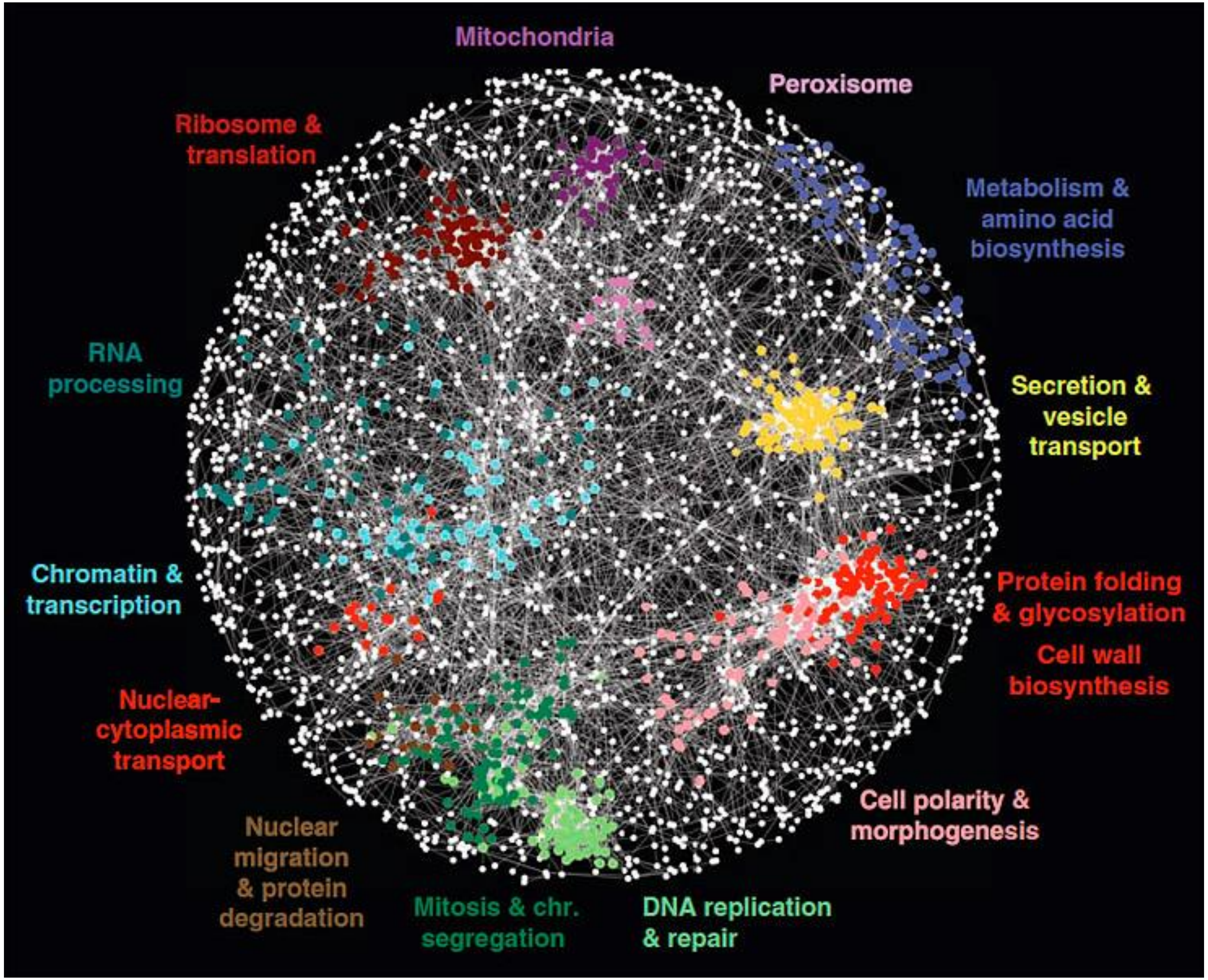
E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási (2002) Hierarchical organization of modularity in metabolic networks. Science, 297:1551-1555.

E. coli metabolism





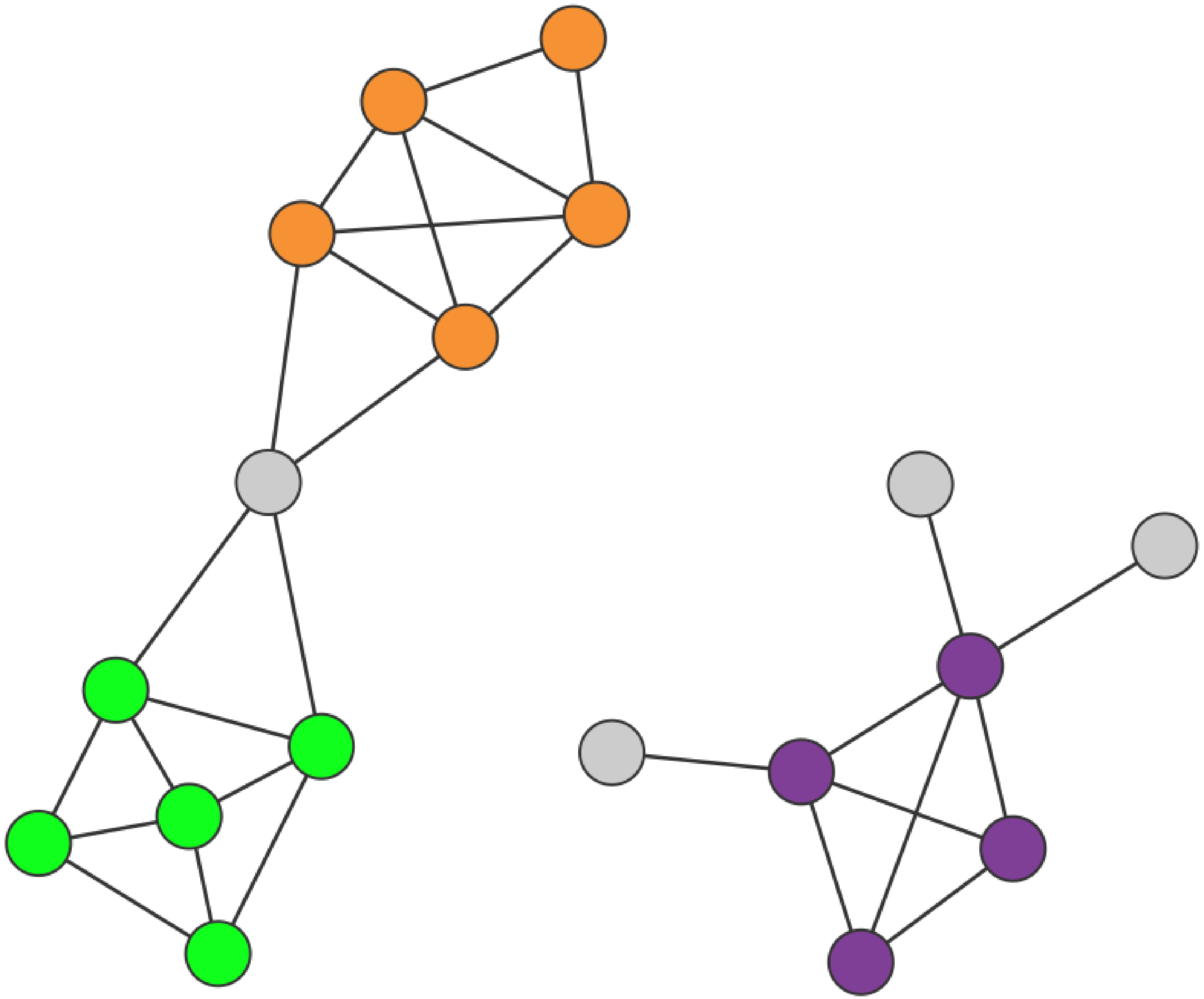
Constanzo, M. et al. (2010) The Genetic Landscape of a Cell, Science, 327(5964): 425-431.



H1: *A network's community structure is uniquely encoded in its wiring diagram*

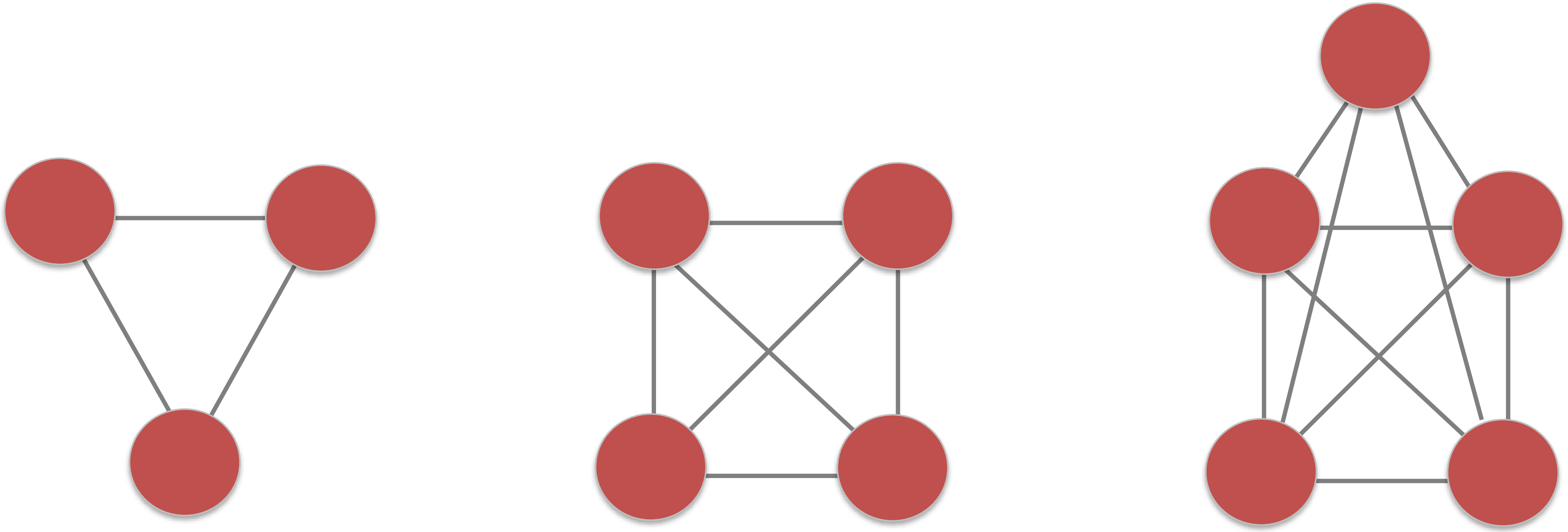
How to define communities?

H2: *A community is locally dense connected subgraph in a network*

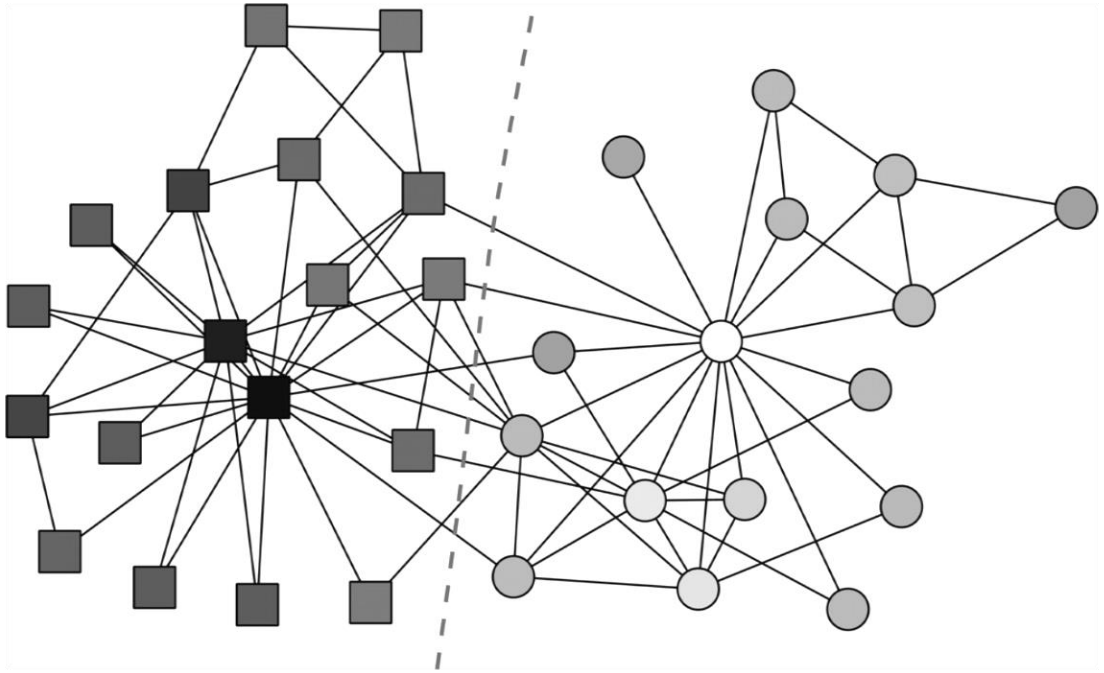


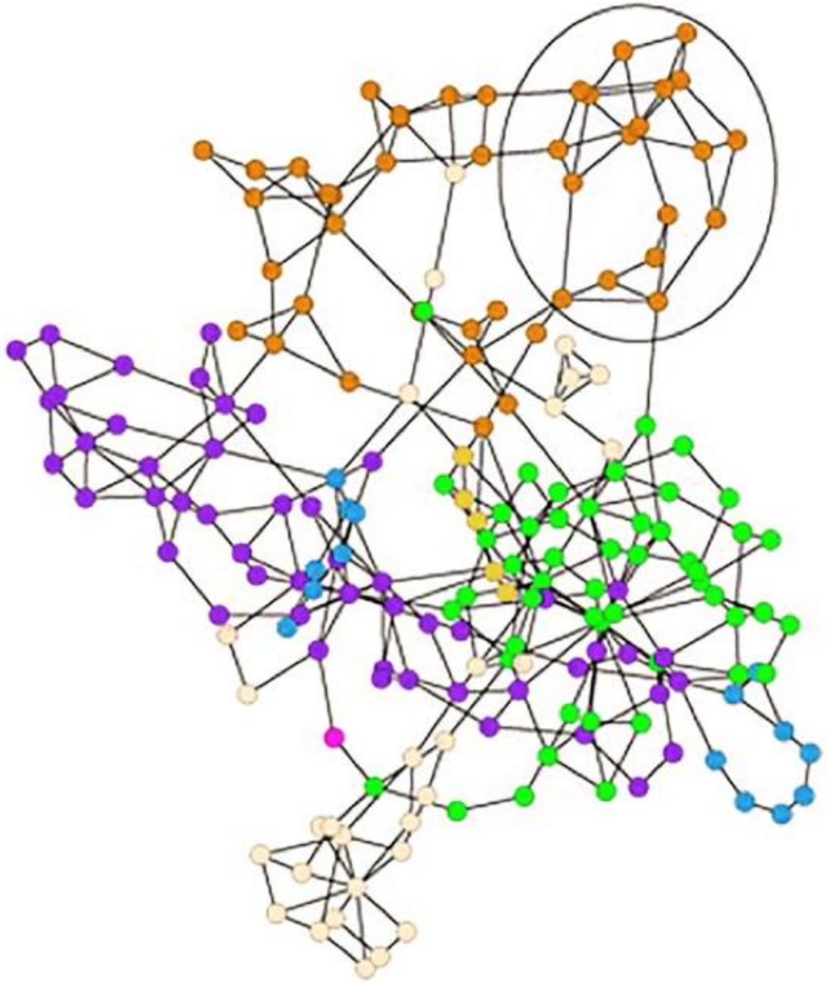
R.D. Luce and A.D. Perry. A method of matrix analysis of group structure. Psychometrika, 14:95–116, 1949.

Cliques as communities

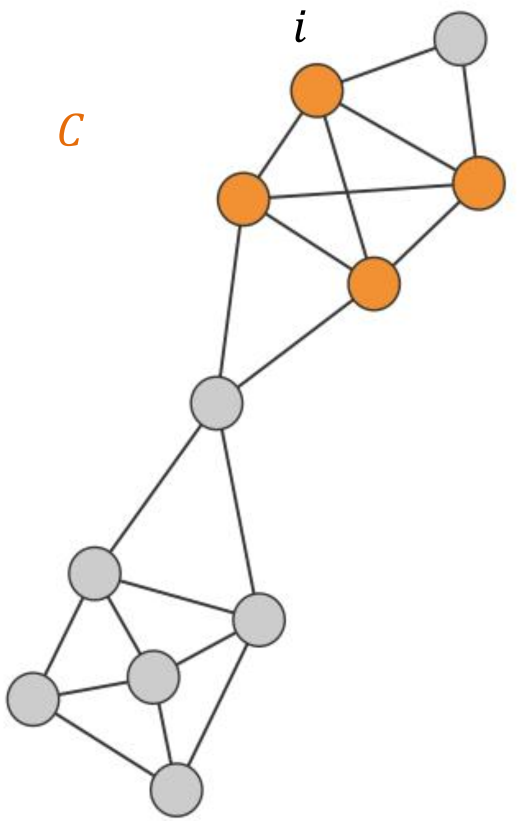


A clique satisfies H2: it is connected subgraph with maximal link density





Internal and external degrees



$k_i^{int} =?$ $\quad k_i^{int} = 3$

$k_i^{ext} =?$ $\quad k_i^{ext} = 1$

What does it mean for i when $k_i^{ext} = 0$ in C?

What does it mean for i when $k_i^{int} = 0$ in C?

Strong community

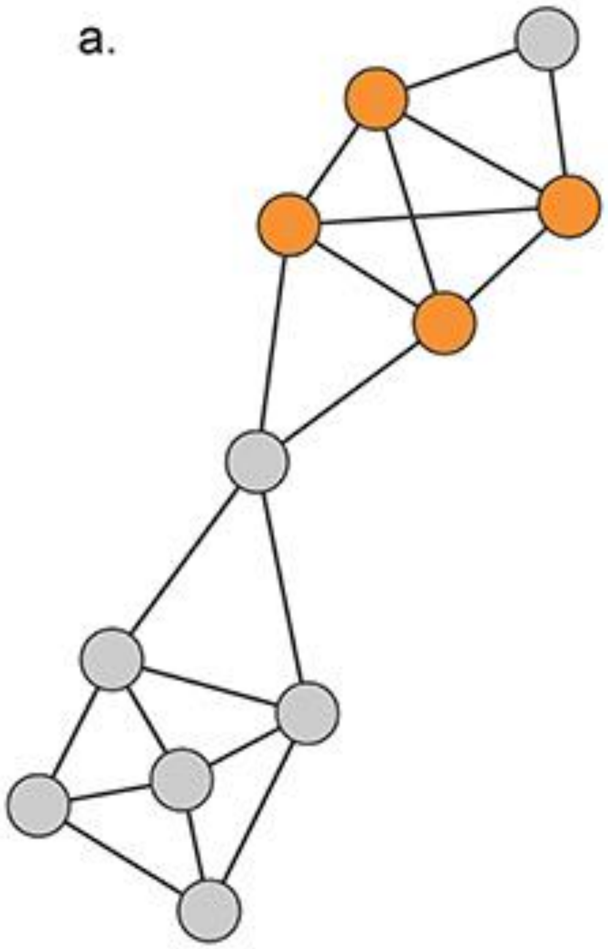
$$k_i^{int}(C) > k_i^{ext}(C)$$

Each node of C has more links within the community than with the rest of the graph

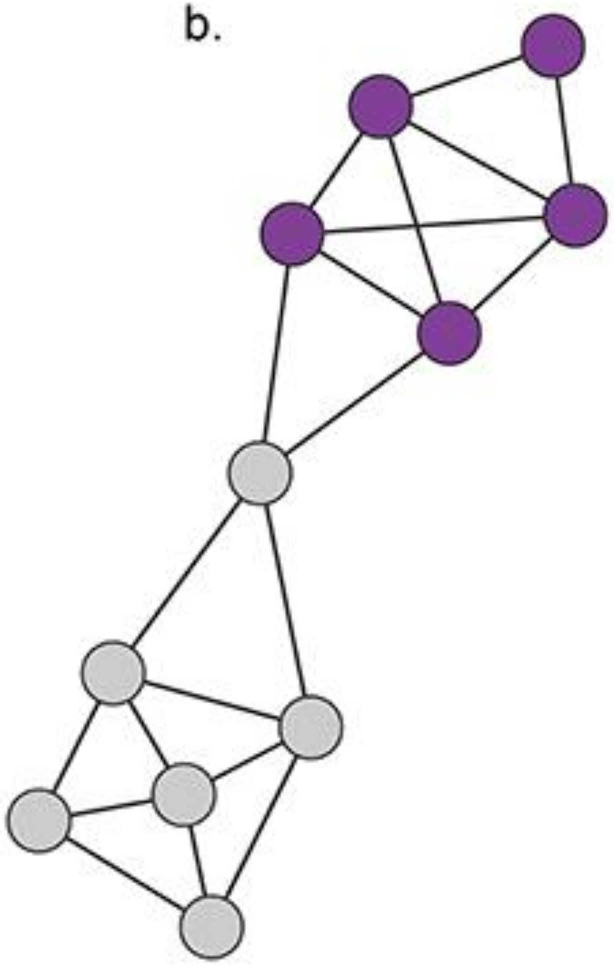
Weak community

$$\sum_{i \in C} k_i^{int}(C) > \sum_{i \in C} k_i^{ext}(C)$$

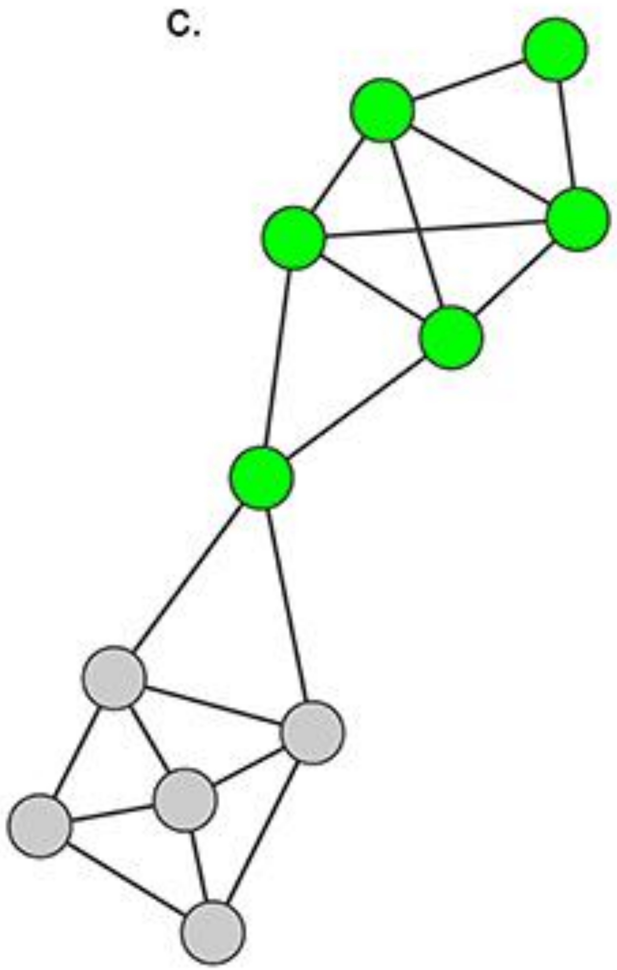
The total internal degree of C exceeds its total external degree



Clique



Strong



Weak

How many ways can we group the nodes of a network into communities?

Graph partitioning

Community detection

Graph bisection problem

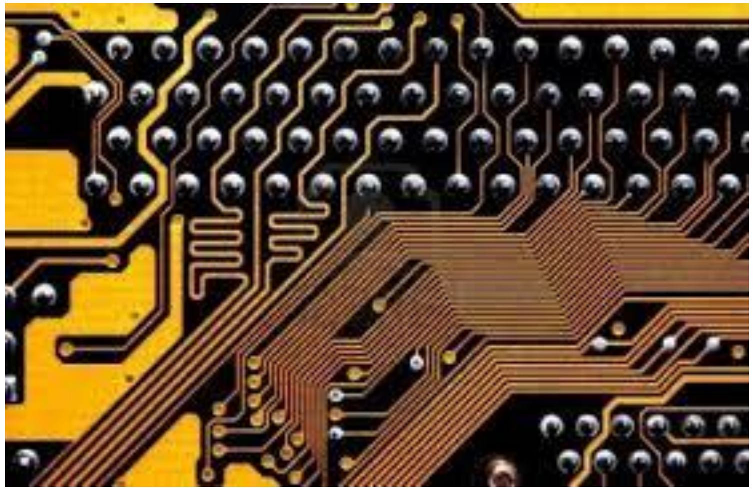
Divide a network into two non-overlapping subgraphs such that the number of links between the nodes in the two subgraphs is minimised

cut size

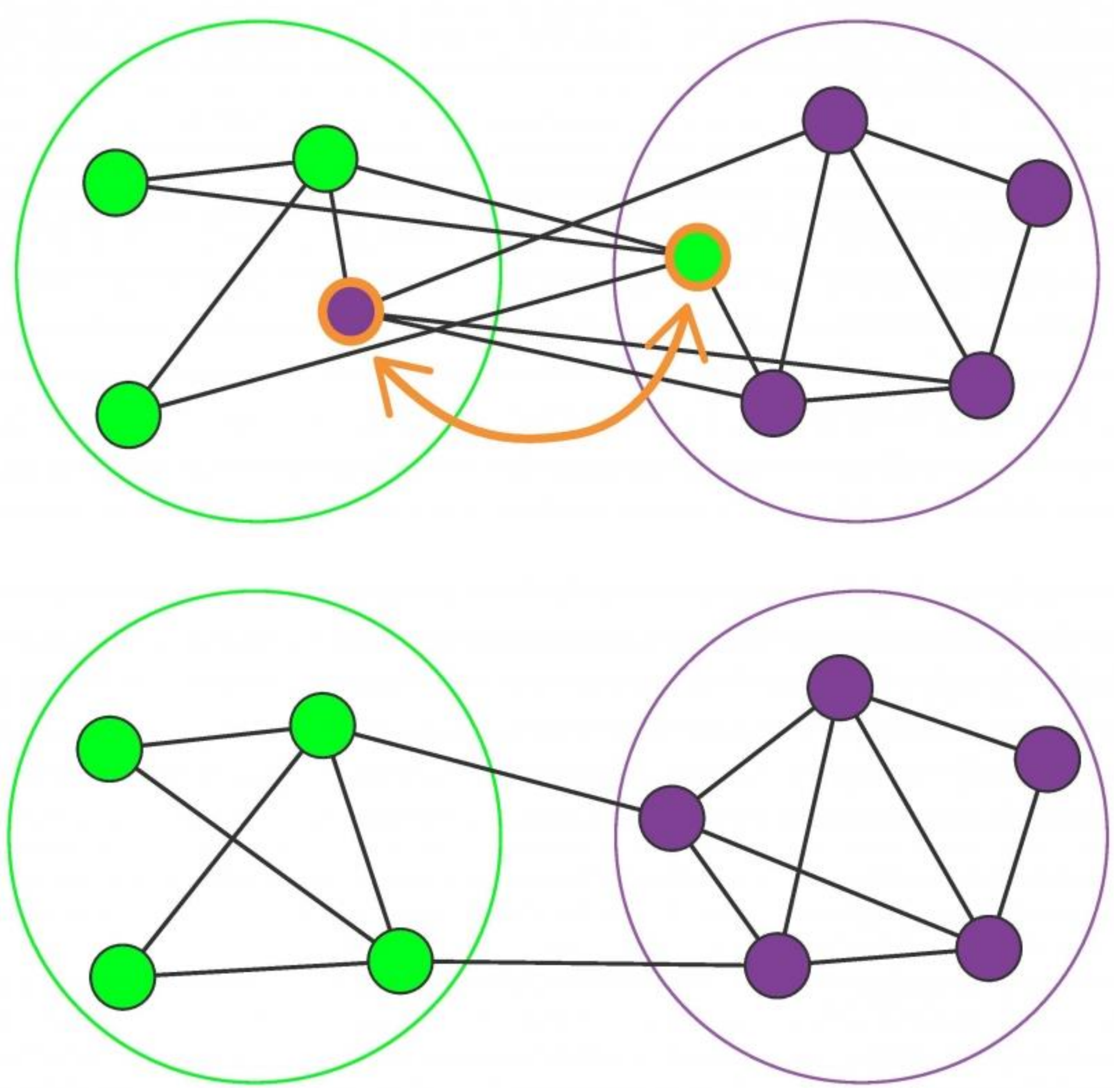
Partitioning problems

- Place 2.5 billion transistors on a chip such that their wires do not intersect

- Partition computation into subtasks assigned to individual processors such that the communication between them is minimised



Kerninghan-Lin algorithm



Computational complexity

N – number of nodes in a network

N_1 – number of nodes in the first subgraph

N_2 – number of nodes in the second subgraph

$$number\ of\ distinct\ partitions = \frac{N!}{N_1!\,N_2!}$$

Assume that $N_1 = N_2 = N/2$

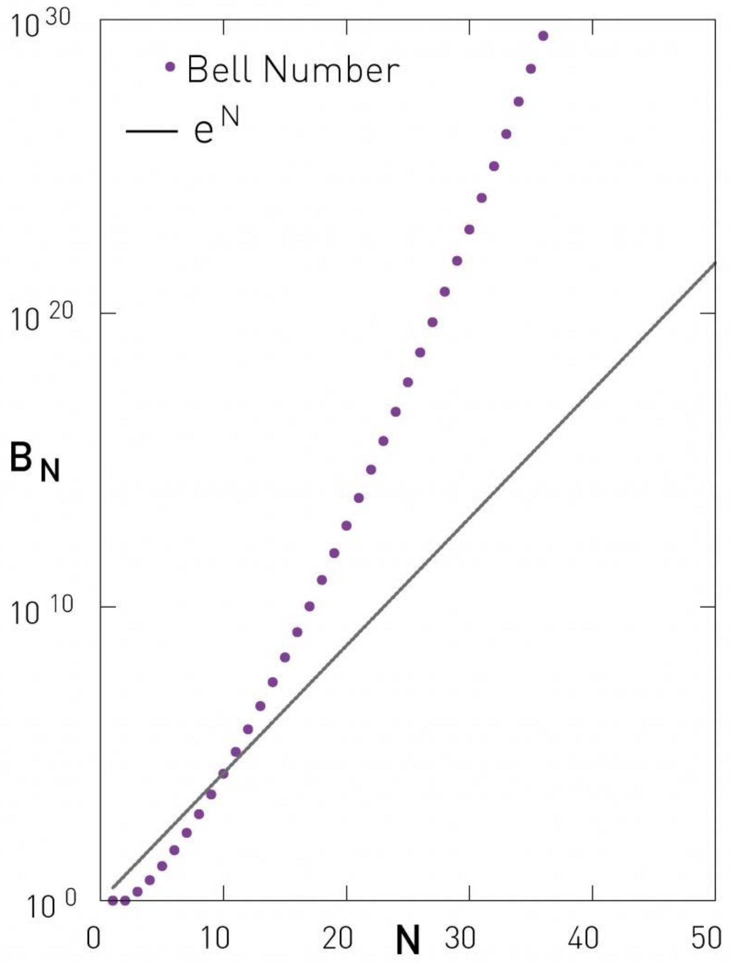
$$\frac{2^{N+1}}{\sqrt{N}} = e^{(N+1)\ln 2 - \frac{1}{2}\ln N}$$

$N = 10$ and $N_1 = N_2 = 5$ => 252 partitions ($10^{-3}s$)
$N = 100$ and $N_1 = N_2 = 50$ => 10^{29} partitions (10^{16} years)

Community detection

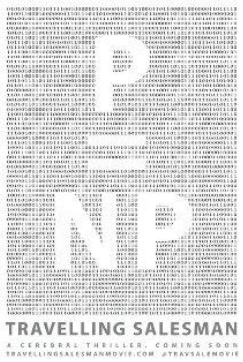
Partition is a division of a network into groups of nodes such that each node belongs to one group only

$$B_N = \frac{1}{e}\sum_{j=0}^{\infty}\frac{j^N}{j!}$$



It is impossible to inspect all partitions of a large network

NP-complete



Hierarchical clustering

Modularity

H3: *Randomly wired networks lack an inherent community structure*

Modularity

$$Q = \frac{1}{2M} \sum_{i,j}^{N} (A_{ij} - P_{ij}) \delta(C_i, C_j)$$

Original network

Expect number of links if the network is randomly linked

Relative to a specific partition

Random network $P_{ij} = \frac{k_i k_j}{2M}$

P_{ij} can take into account

weights

Fortunato, S. (2010) Community detection in graphs. Physics Reports, 486(3-5):75–174.

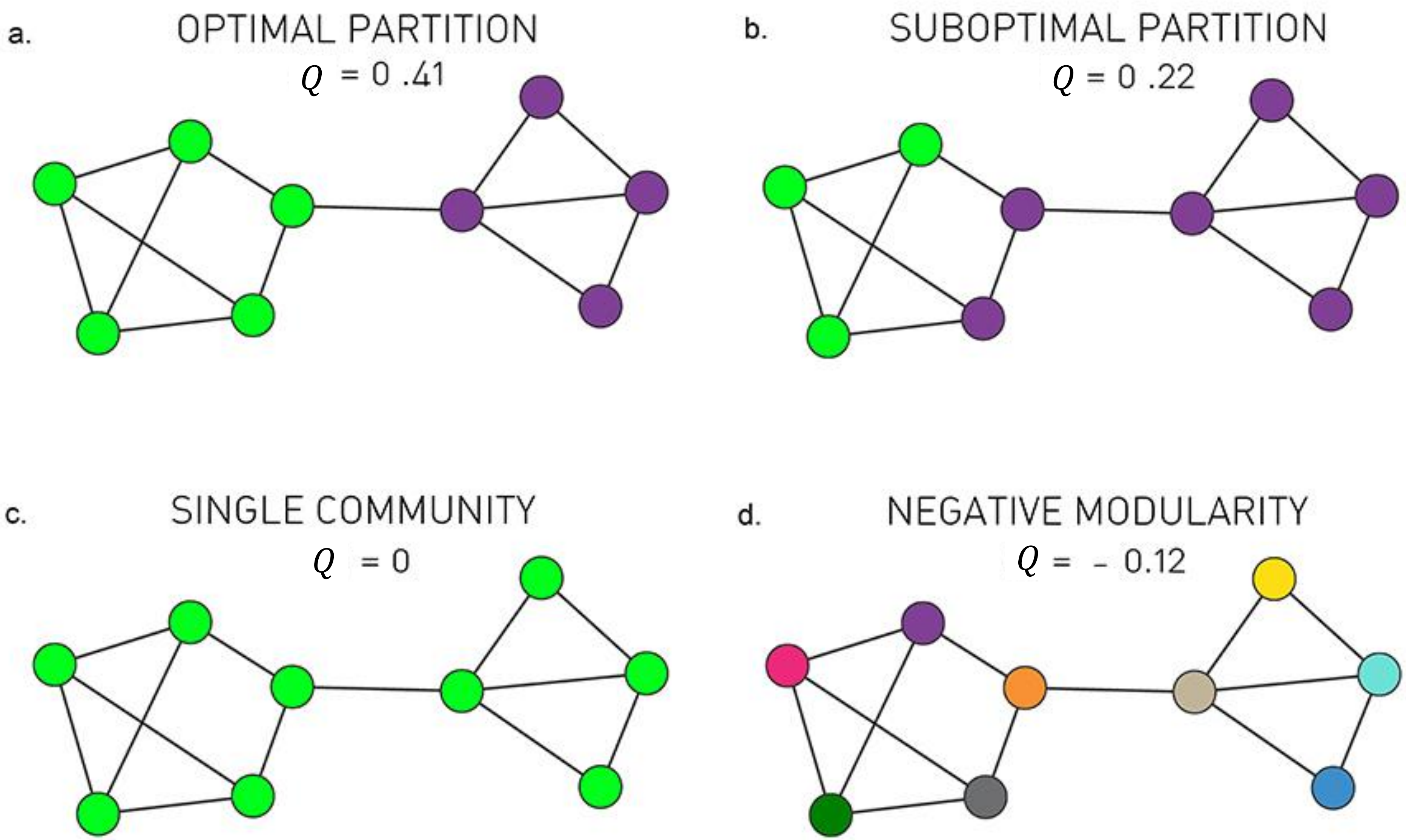
directions

Fortunato, S. (2010) Community detection in graphs. Physics Reports, 486(3-5):75–174.

attributes or space

Expert, P., Evans, T. S., Blonder, V. D., Lambiotte, R. (2011) Uncovering space-independent communities in spatial networks. PNAS, 108(19):7663–7668.

Which partition $\{C_c, c = 1, n_c\}$?

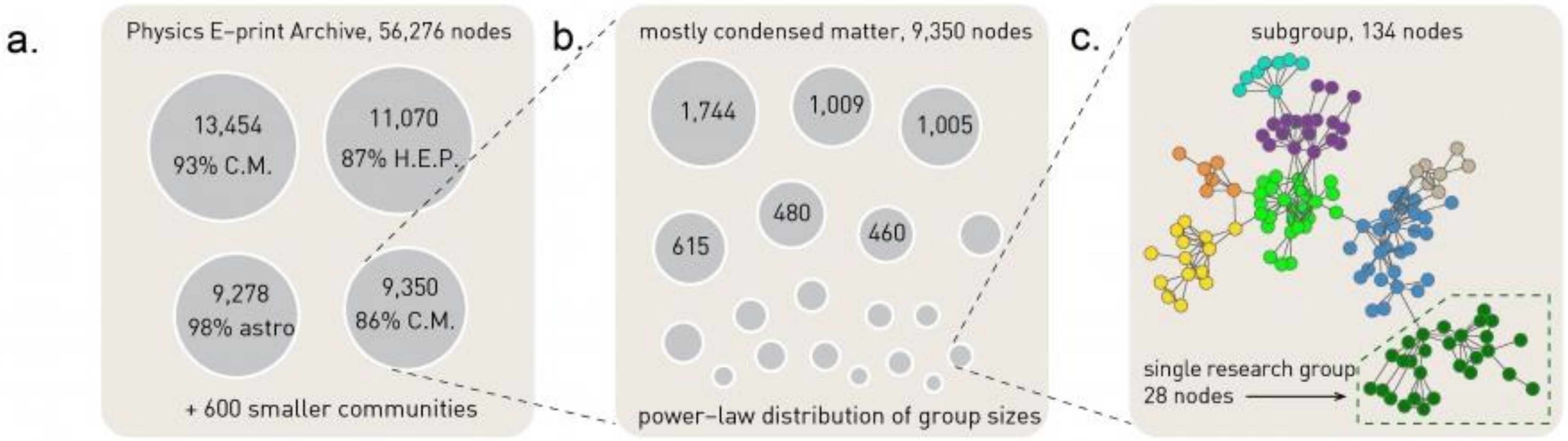


H4: *For a given network, the partition with maximum modularity corresponds to the optimal community structure*

Find partition $\{C_c, c = 1, n_c\}$ maximises Q

Greedy algorithm

1. Assign each node to a community of its own
2. Inspect each pair of communities connected by at least one link and compute ΔQ
3. Identify the community pair for which ΔQ is the largest and merge them. Note that the modularity of a particular partition is calculated from the full topology of the network
4. Repeat step 2 until all nodes are merged into a single community, recording Q for each step
5. Select the partition for which Q is maximal


a.
Physics E-print Archive, 56,276 nodes
13,454
93% C.M.
11,070
87% H.E.P.
9,278
98% astro
9,350
86% C.M.
+ 600 smaller communities
b.
mostly condensed matter, 9,350 nodes
1,744
1,009
1,005
480
615
460
power-law distribution of group sizes
c.
subgroup, 134 nodes
single research group
28 nodes

Computational complexity

- Step 1-2 (calculation of ΔQ for M links): $O(M)$
- Step 3 (matrix update): $O(N)$
- Step 4 (N-1 community mergers): $O((M+N)N)$

- For sparse networks $O(N^2)$ M.E.J. Newman. Fast algorithm for detecting community structure in networks. Physical Review E, 69:066133, 2004.
- Optimised implementation $O(Nlog^2N)$ A. Clauset, M.E.J. Newman, and C. Moore. Finding community structure in very large networks. Physical Review E, 70:066111, 2004.

- Louvain algorithm $O(M)$ V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. J. Stat. Mech., 2008.

Community detection with modularity

- Fraction of edges that fall within the given communities minus the expected fraction if edges were distributed at random, while conserving the nodes degrees
- Measure of relative density in the network: a community has high density relative to other nodes within the community but low density with those outside
- Gives an overall score of how fractious your network is, which can be used to partition the network and return the individual communities

Community detection with modularity

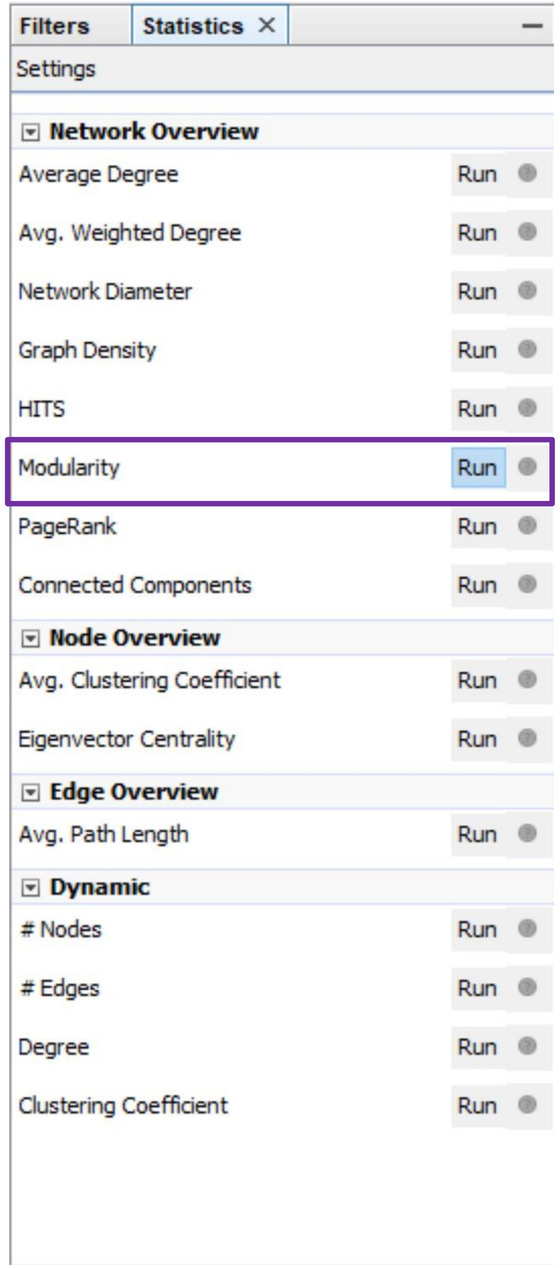
- Modularity of 0.6 and above corresponds to networks that have a clearly visible community structure

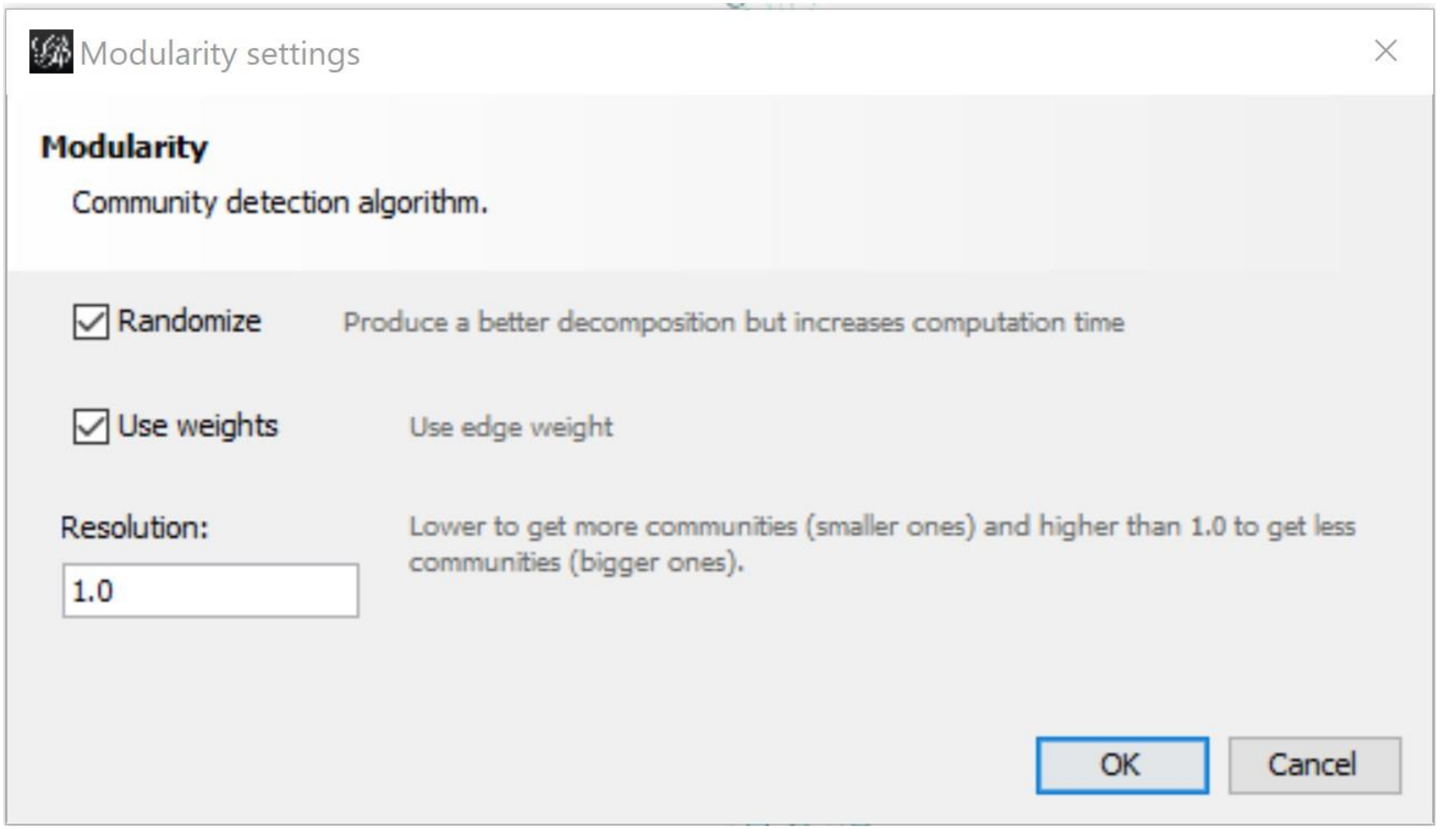
- Very dense networks are often more difficult to split into sensible partitions

Modularity in NetworkX

- Requires installing an additional package called `python-louvain`
  - Documentation: https://perso.crans.org/aynaud/communities/api.html
- Externally visible name of the module is `community`
- It uses the Louvain algorithm that maximises network modularity
  - `community.best_partition(G)`
    - tries to determine the number of communities
    - assigns each node a number (>=0), corresponding to the community it belongs to
    - computes the partition of the graph nodes that maximises the modularity
  - `community.modularity(partition, graph)`
    - compute the modularity of a partition of a graph

Modularity in Gephi



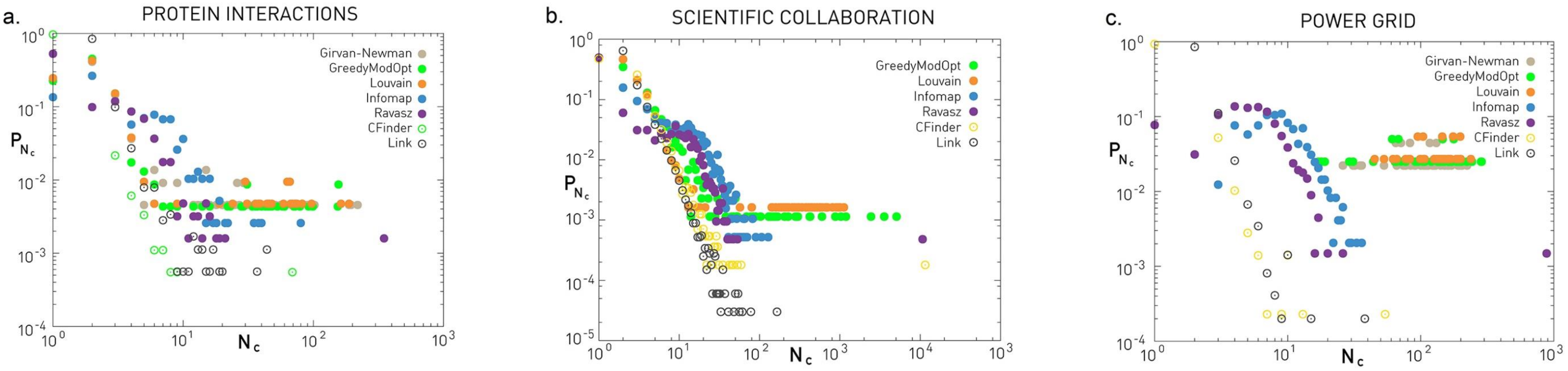


Resolution assigns self-loops to nodes to increase or decrease the aversion of nodes to form communities

Characterising communities

- Community size distribution
- Communities and link weights
- Community evolution

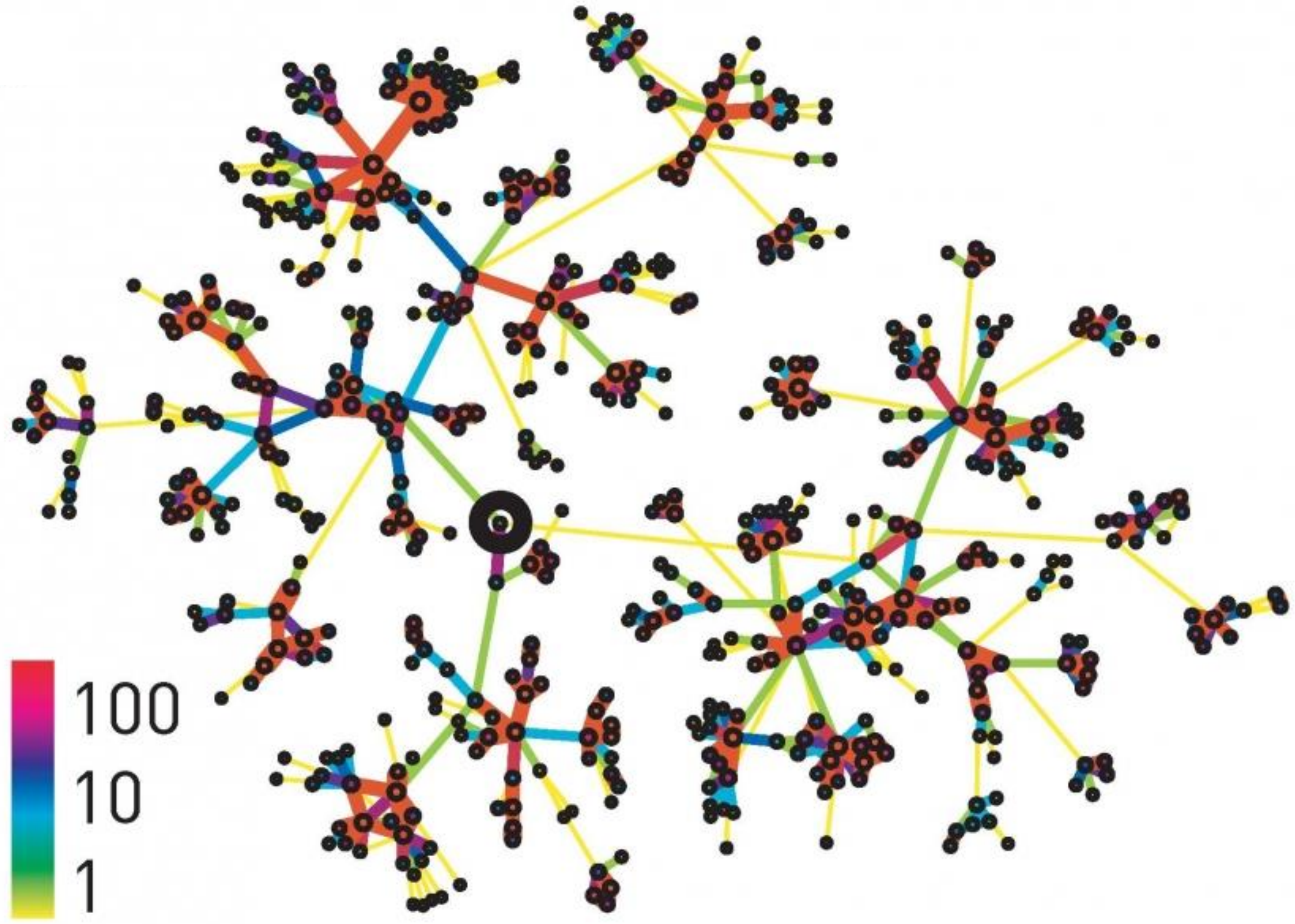
Community size distribution



J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. PNAS, 104:7332, 2007.

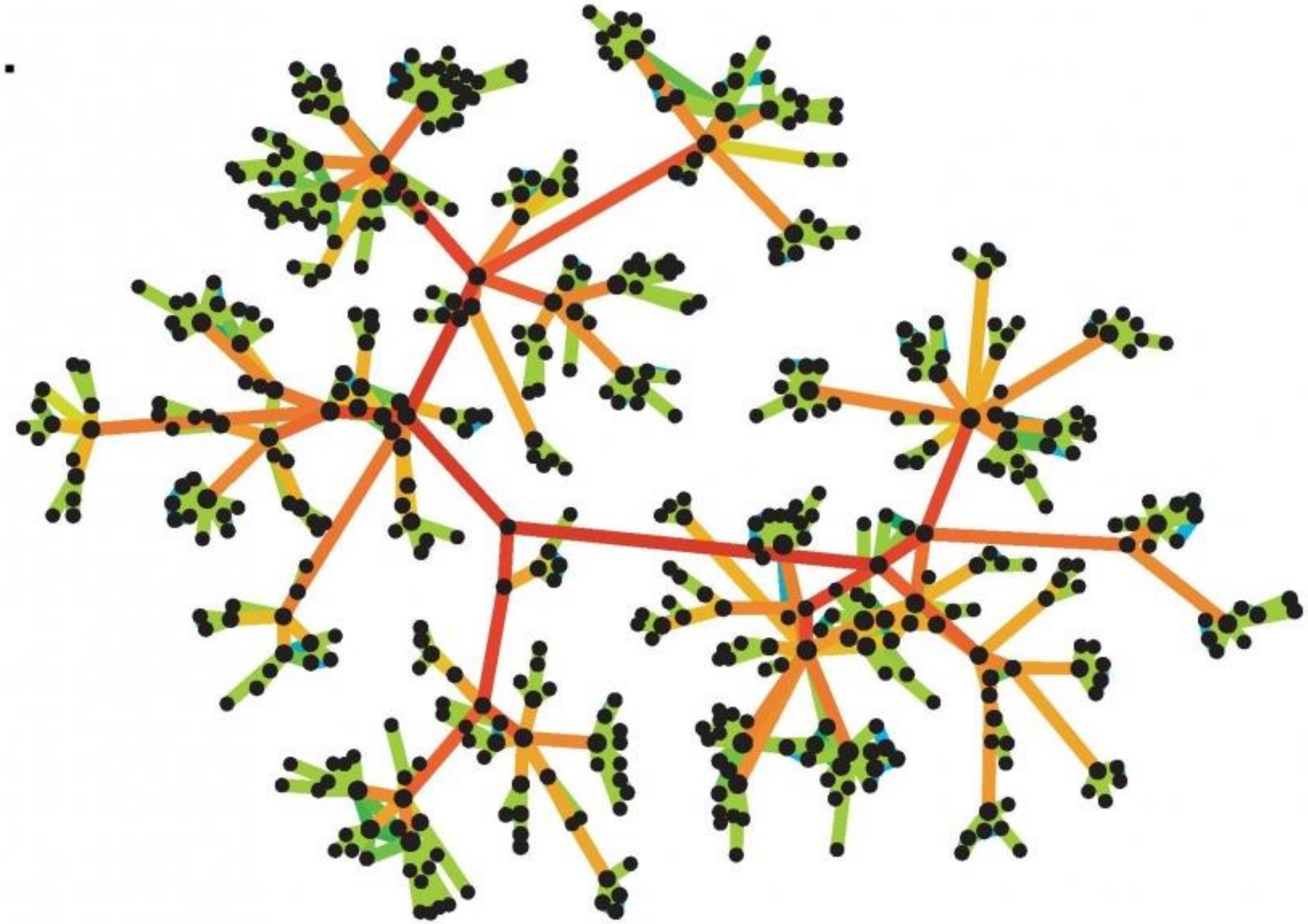
Communities and link weights

Weak tie hypothesis



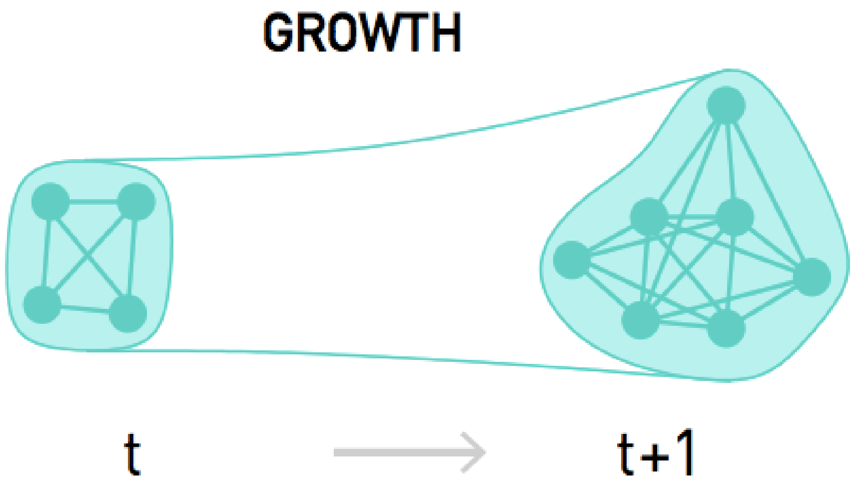
Communities and link weights

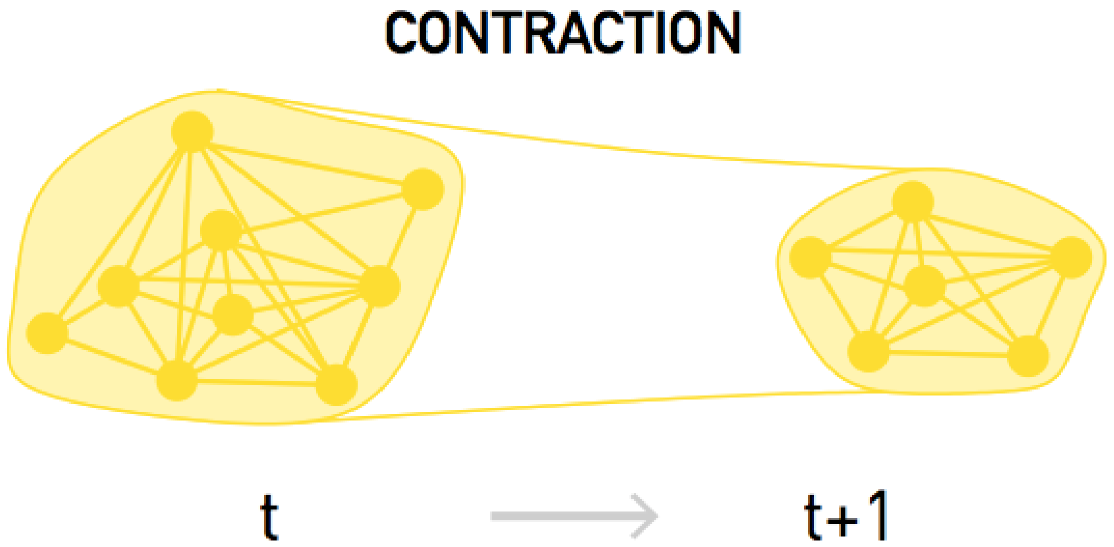
b.



G. Palla, A.-L. Barabási, and T. Vicsek. Quantifying social group evolution. Nature, 446:664-667, 2007.

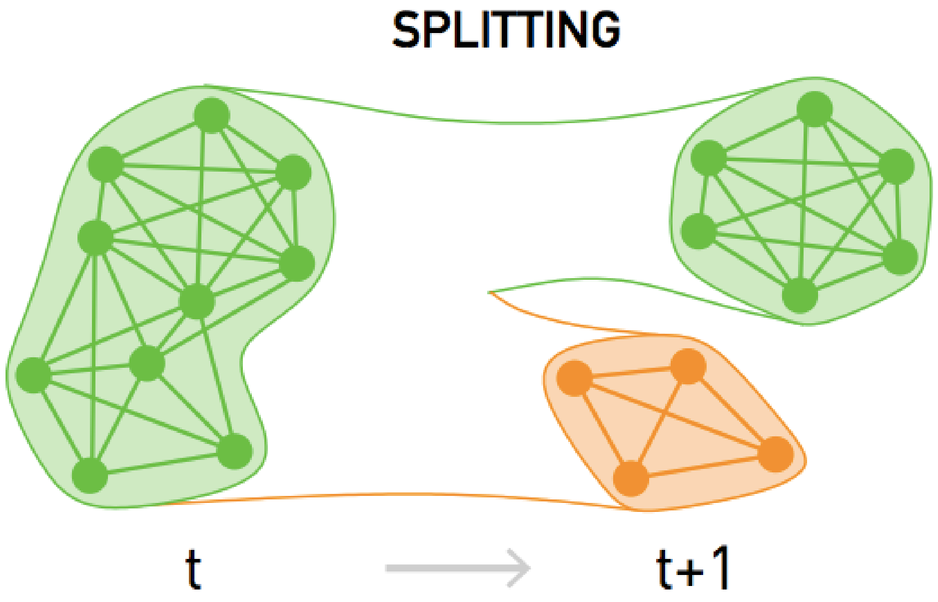
Community evolution

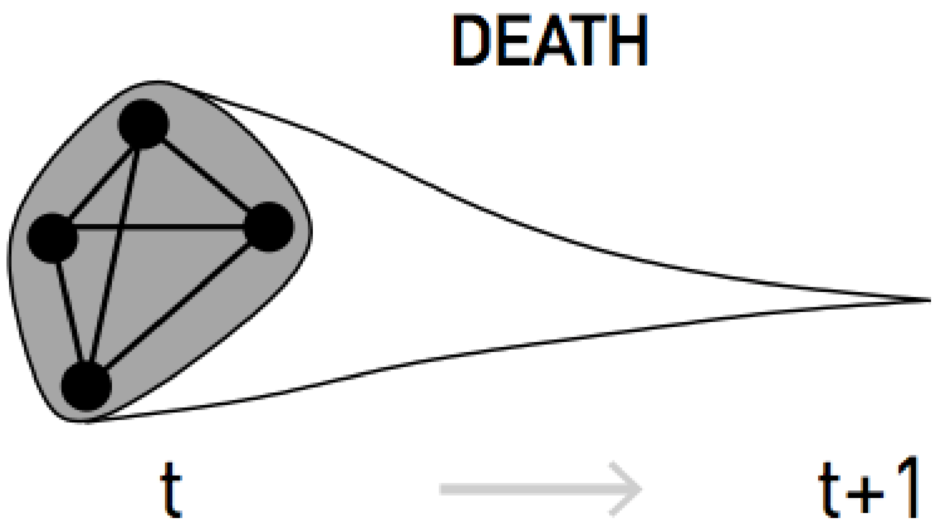




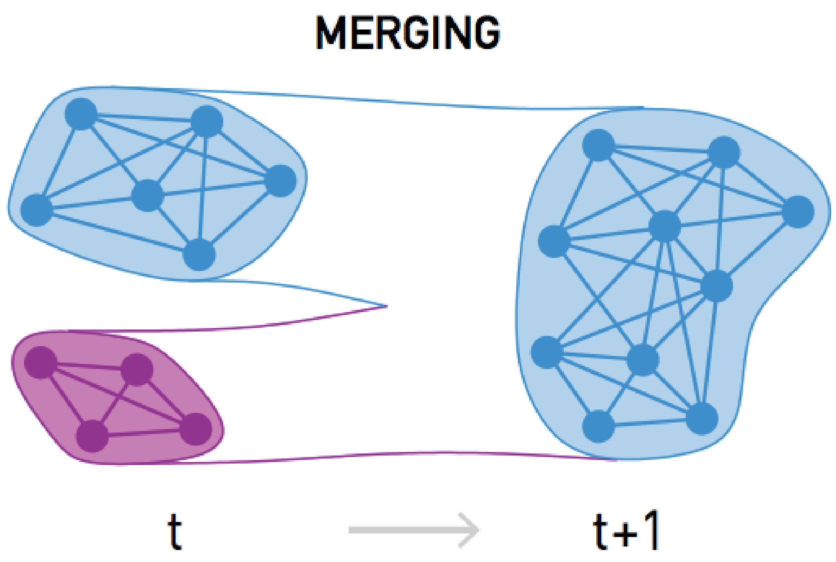
G. Palla, A.-L. Barabási, and T. Vicsek. Quantifying social group evolution. Nature, 446:664-667, 2007.

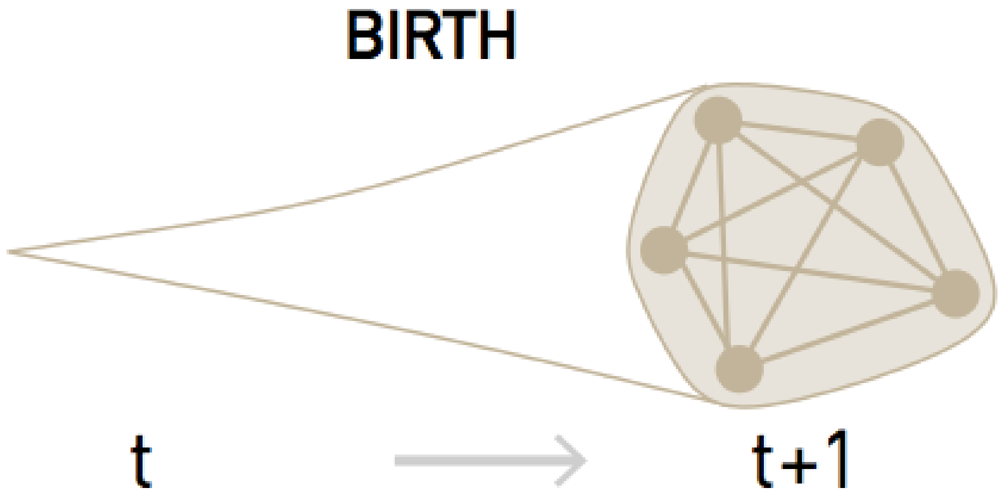
Community evolution



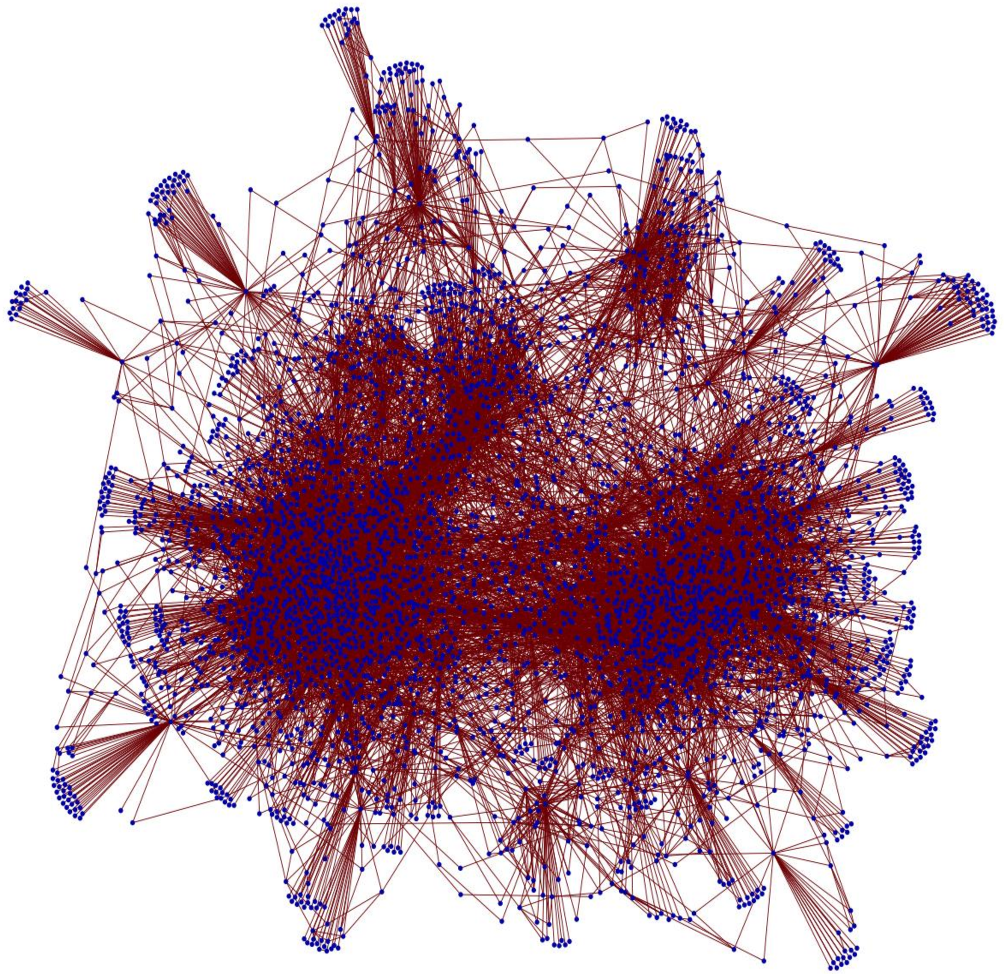


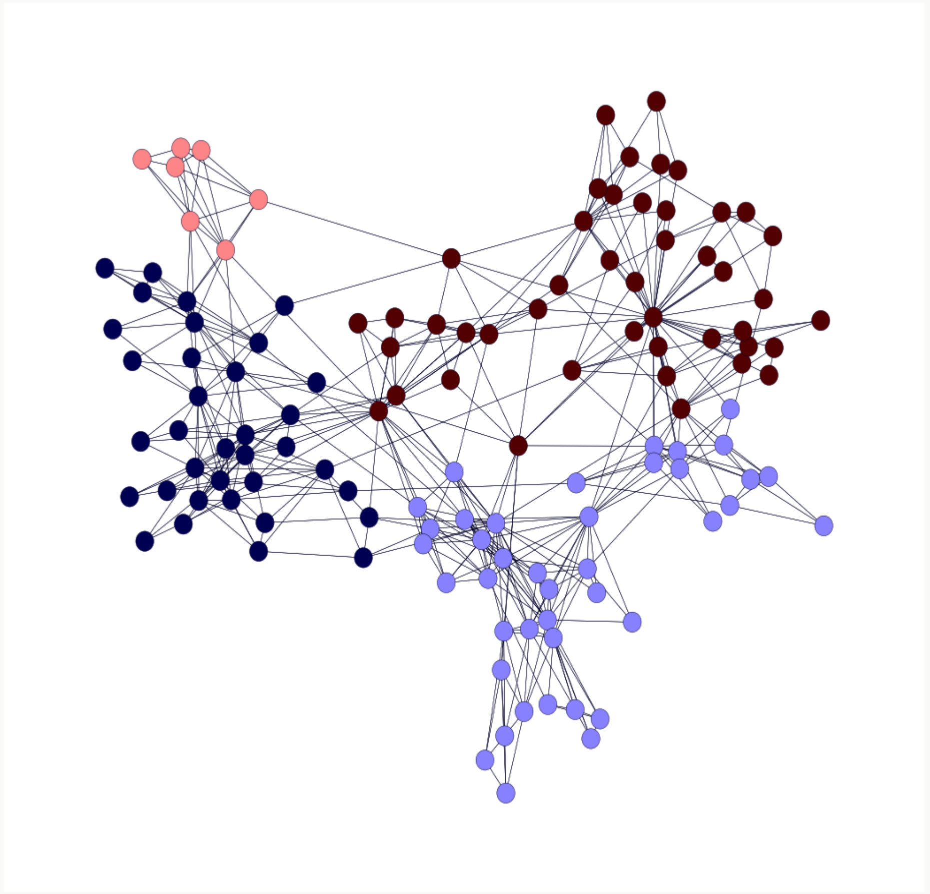
Community evolution

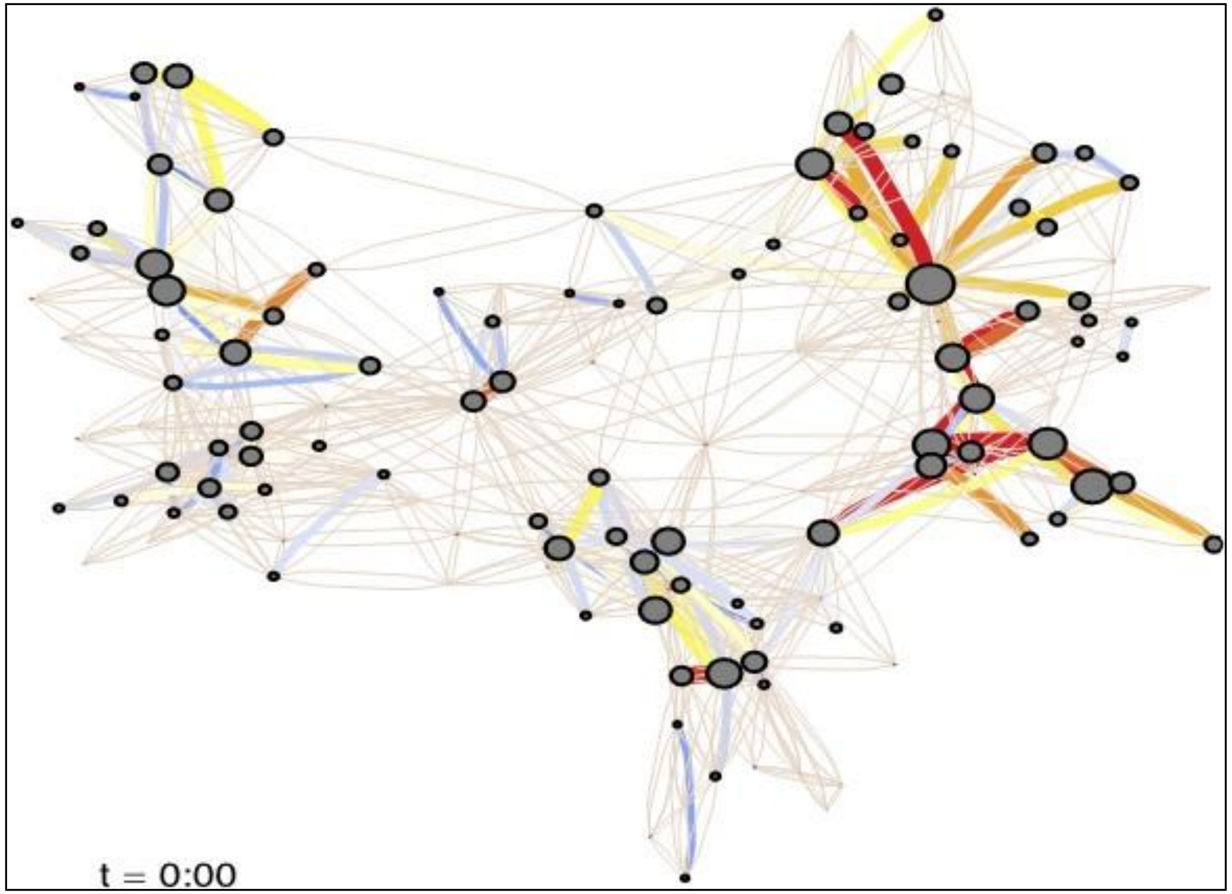


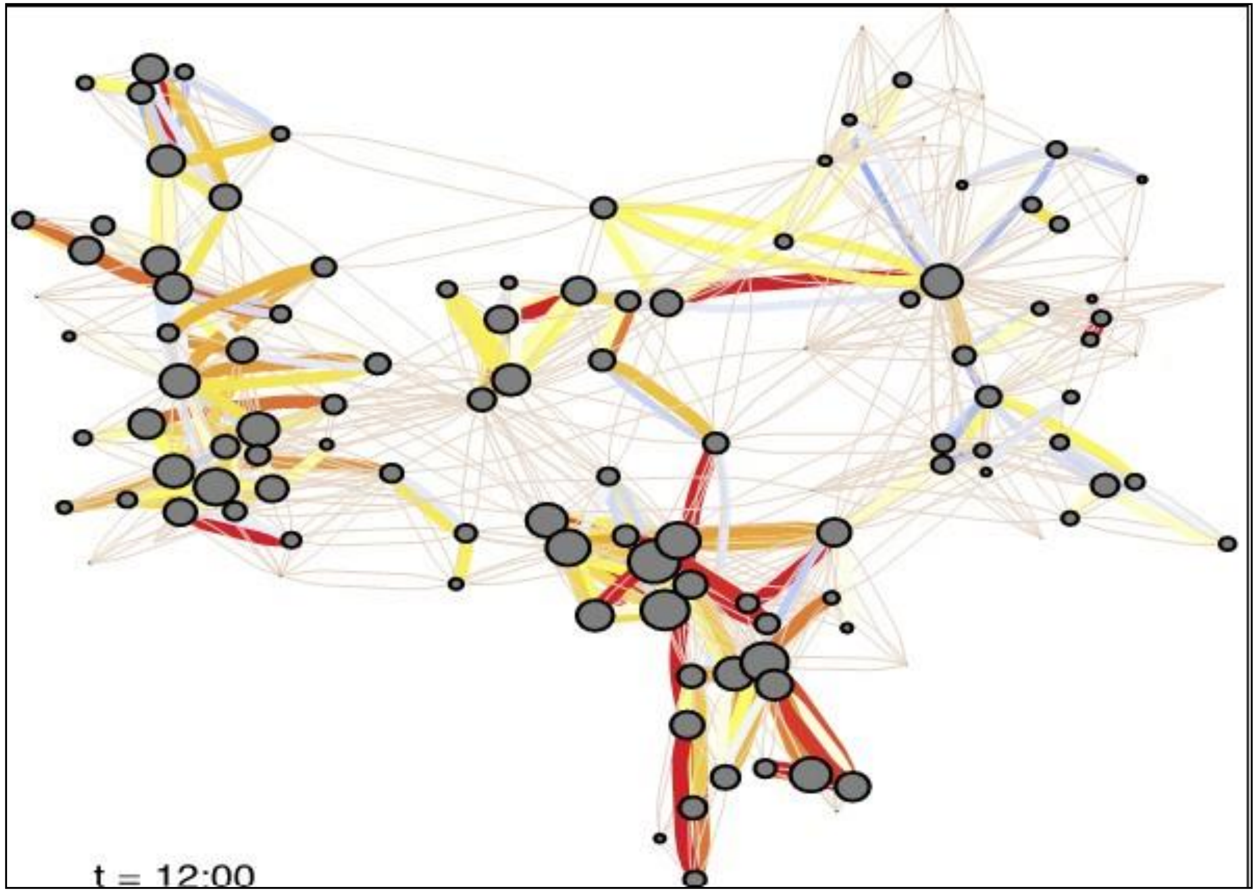


Do communities matter?







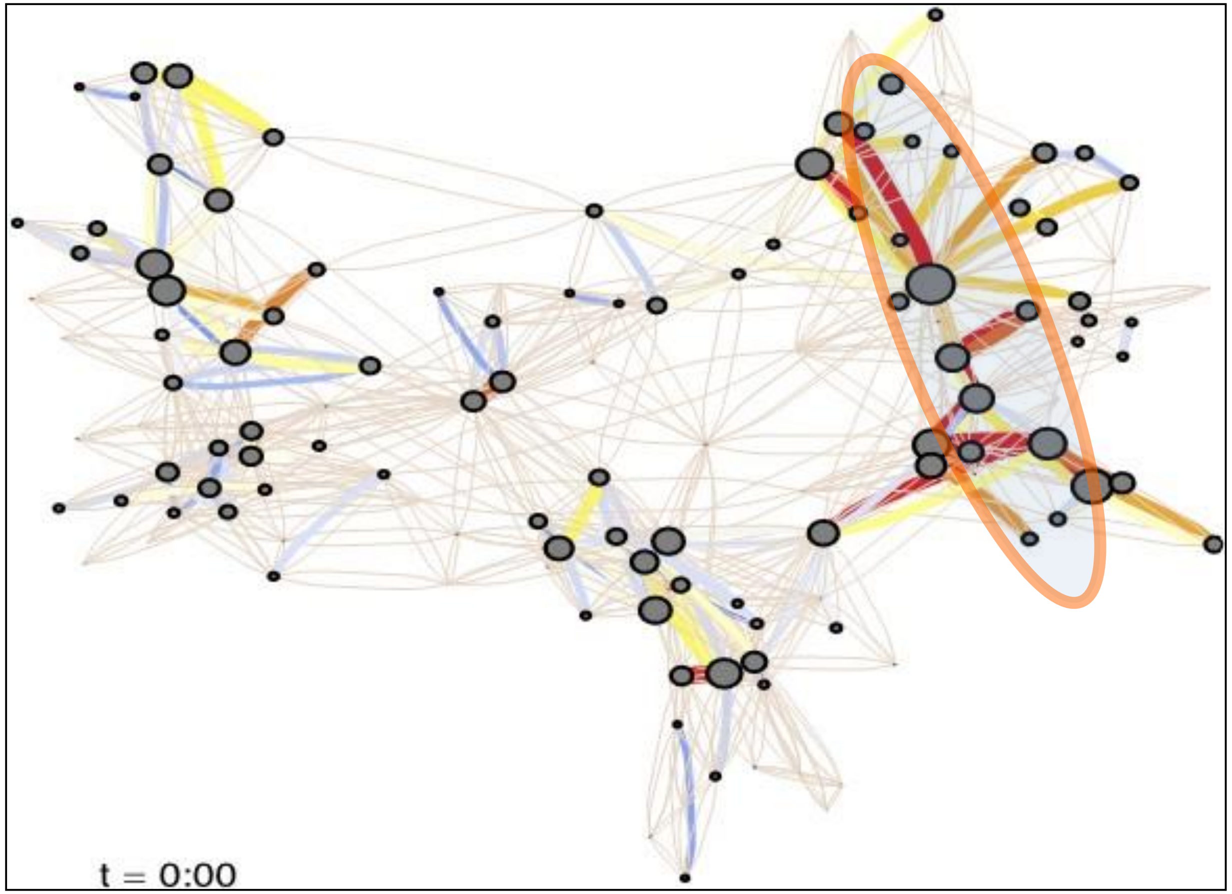


Noon

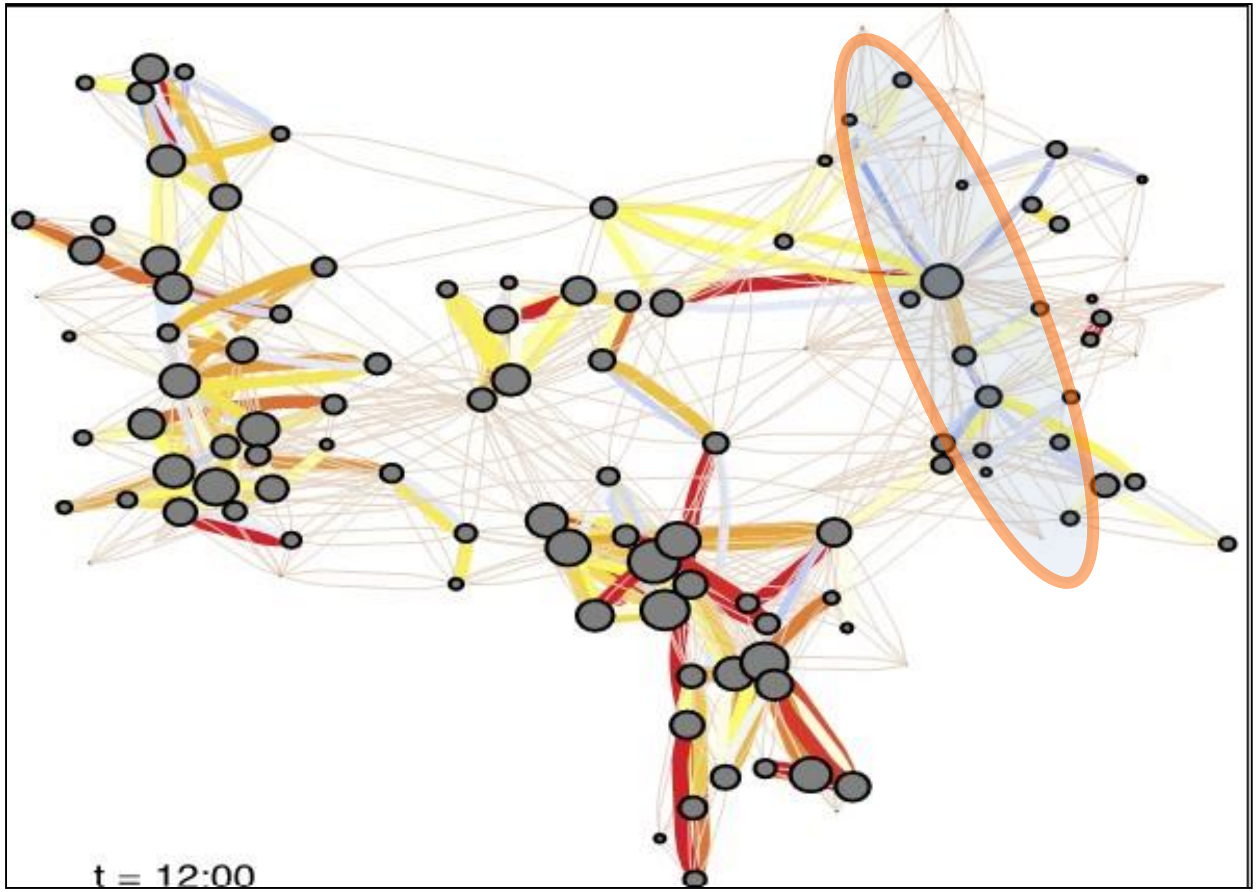
Midnight

Busy at midnight

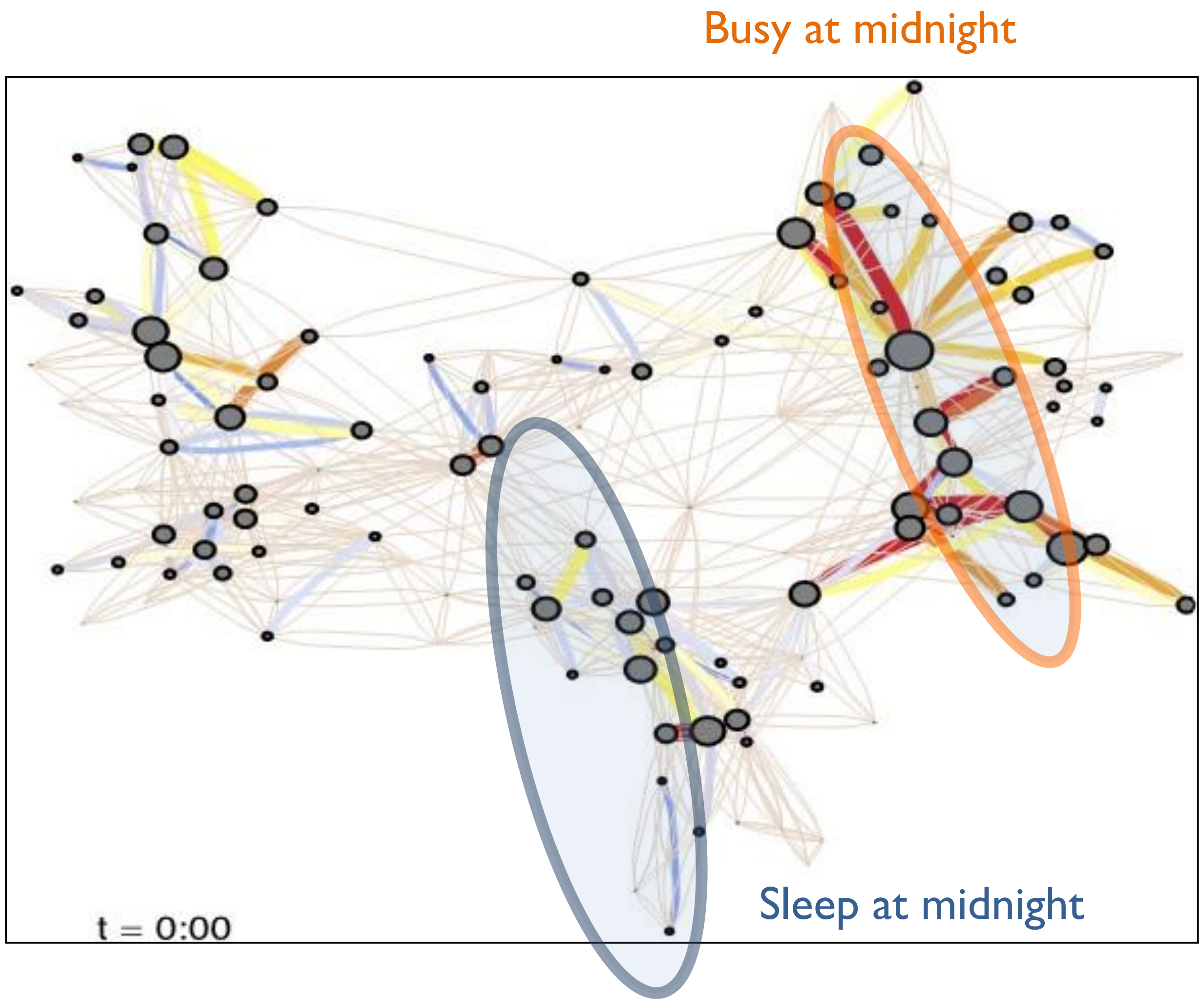
Sleep at noon



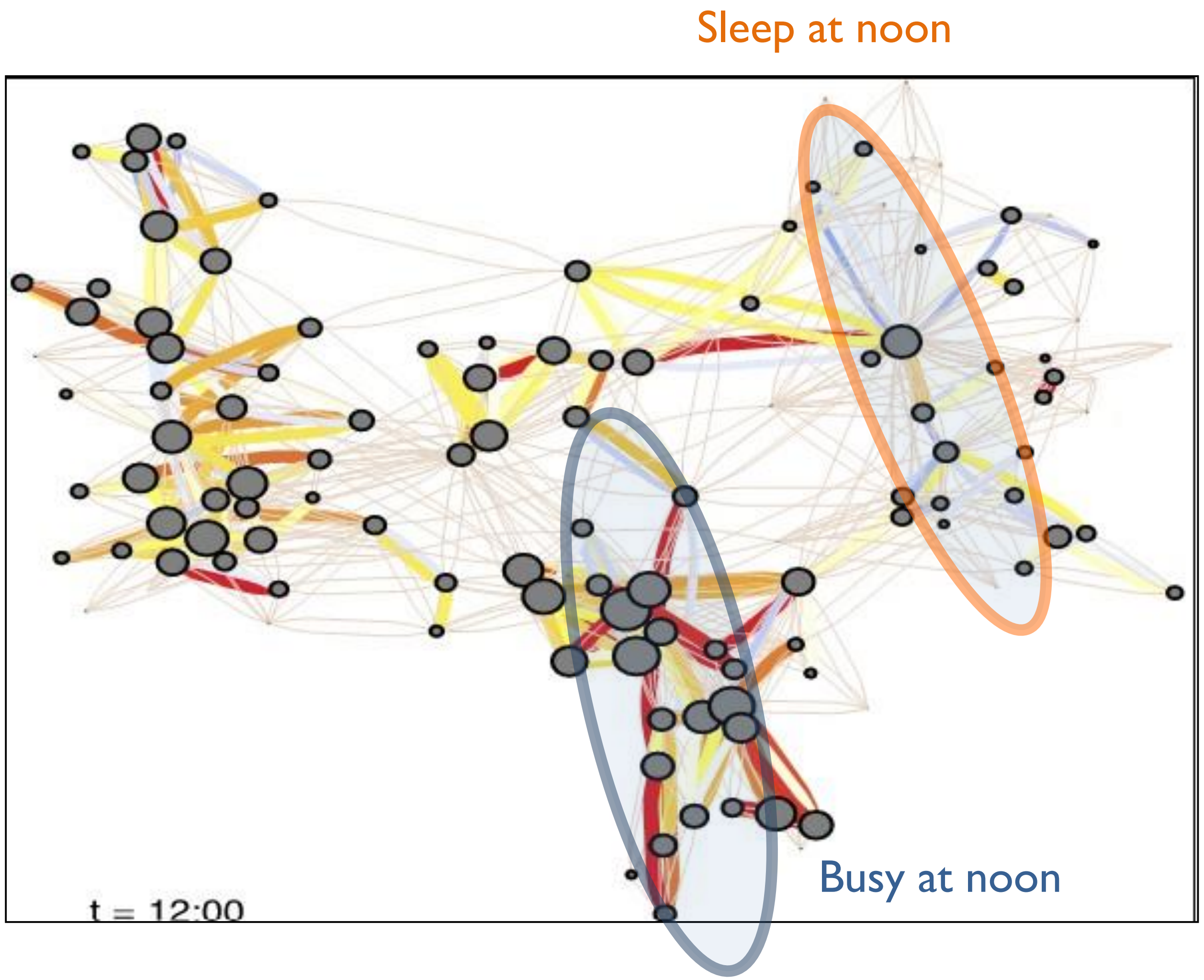
Midnight



Noon



Midnight



Noon

Sources

- Zinoviev, D.. Complex Network Analysis in Python: Recognize, Construct, Visualize, Analyze, Interpret, The Pragmatic Bookshelf, 2018.
- Barabási, A. Network Science, http://networksciencebook.com
- Cornelius, S. P., Towlson, E. K., Barabási, A., Communities Part 1 and Part II, Network Science.