Complex Networks IV

Plan of the lectures

I. Introduction

II. Networks: basic definitions

III. Models

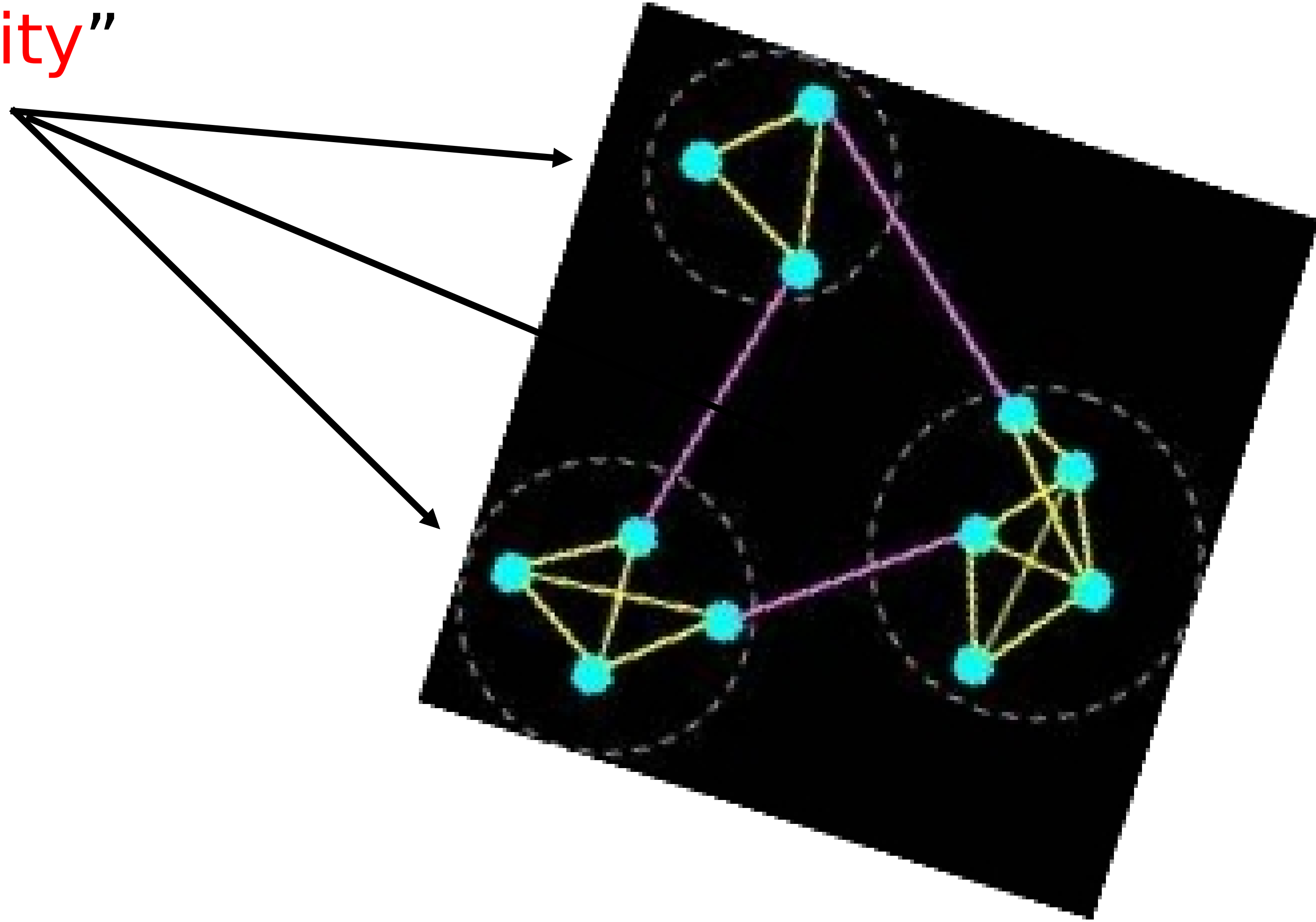
IV. Community Detection

Communities

Complex Networks typically show a high degree of modularity. These modules are called "communities"

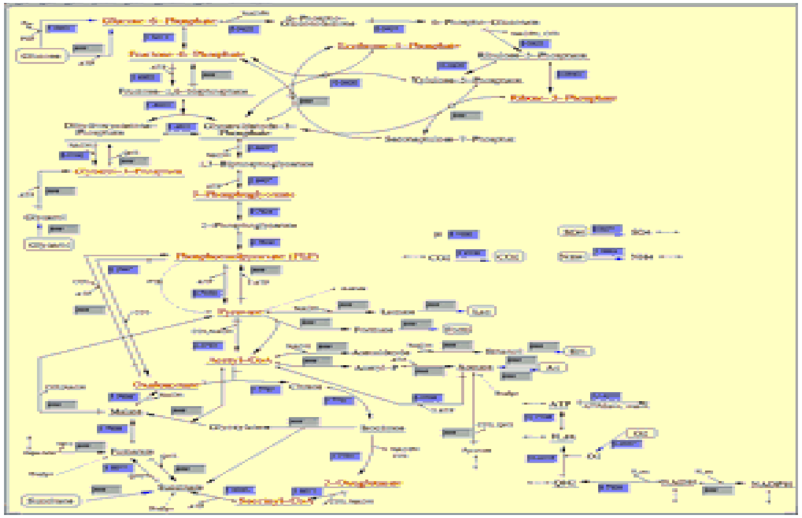
The goal of this lecture is to give a rigorous definition of modularity and a set of recipes to perform community detection.

“Community”

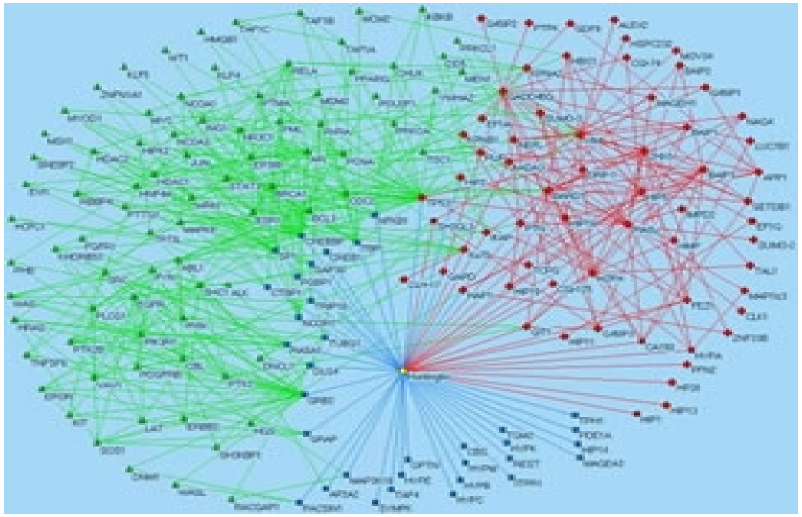


More link “within” a community than outside

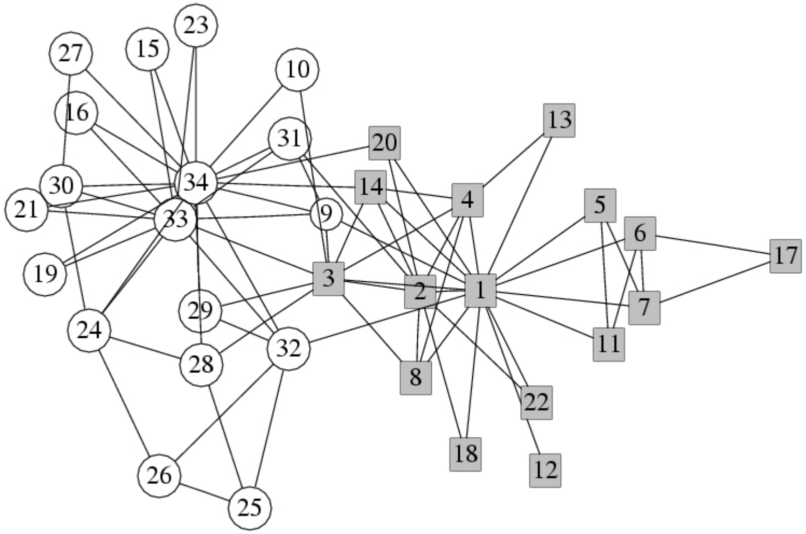
Metabolic



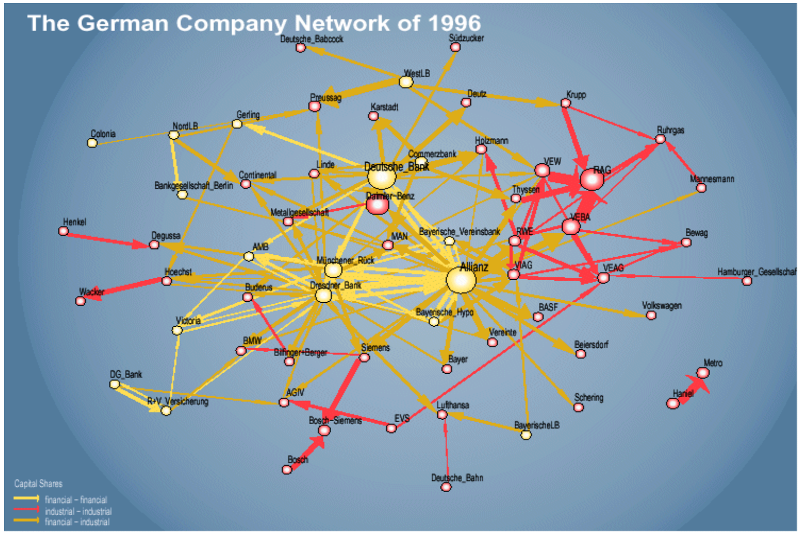
Protein-protein



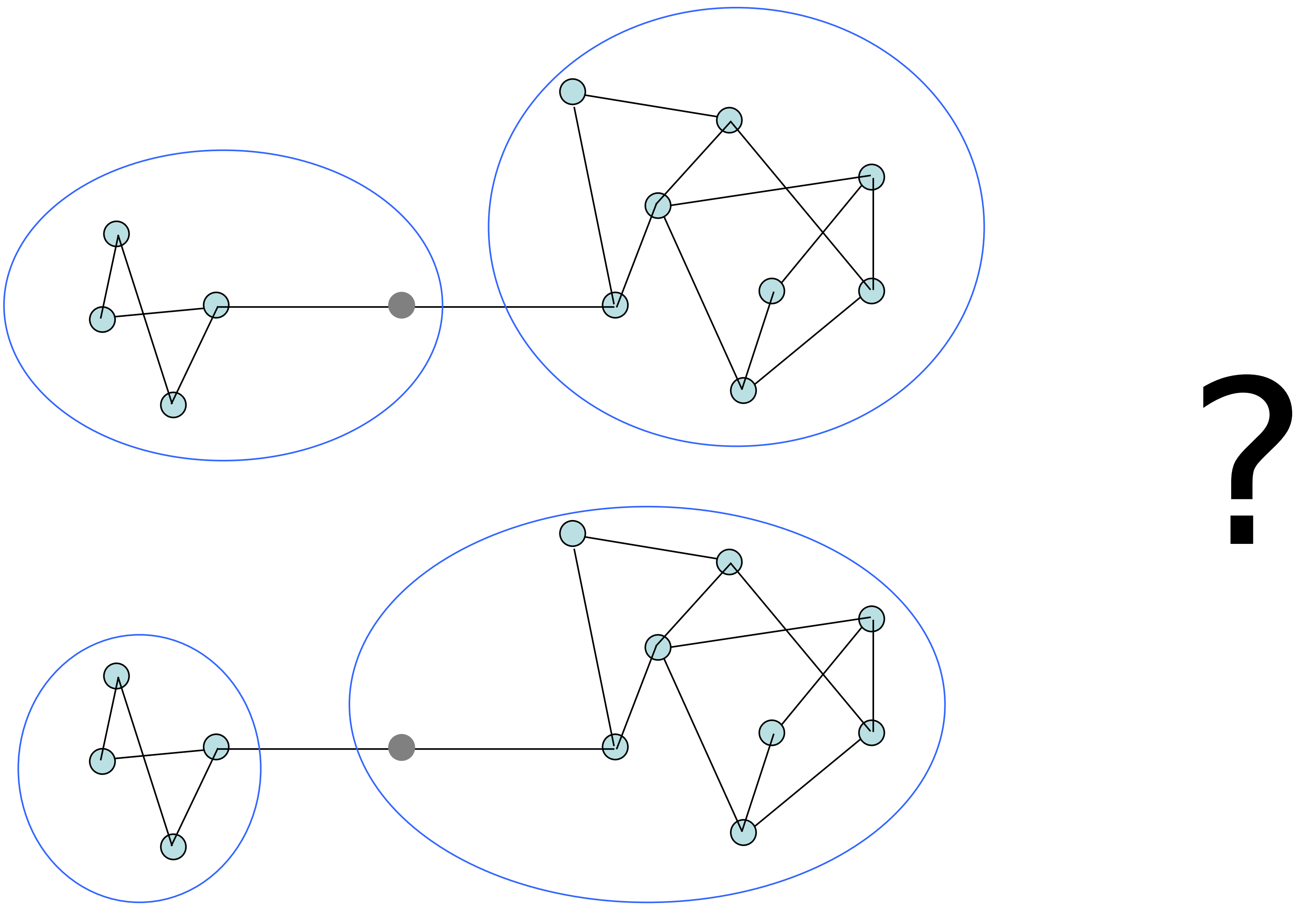
Social



Economical



How can we compare different partitions?



A few important remarks:

1) There are several different ways to perform community detection, depending on the quantity one wants to optimize

2) The problem is NP complete, there is no exact solution, only heuristic approximations

3) Finding communities is more and more difficult as the network becomes larger. One has to face a trade-off between efficiency and quality of the community organization.

Modularity

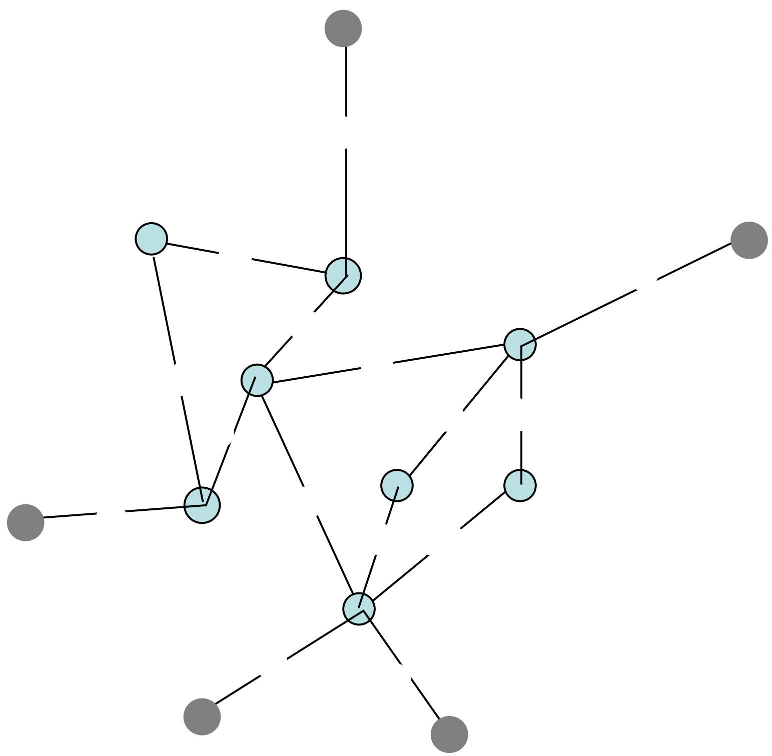
$$Q = \frac{1}{L}\sum_{i=1}^{n}\left(l_i - \frac{d_i^2}{4L} \right)$$

n= num. di communities according to a given partition

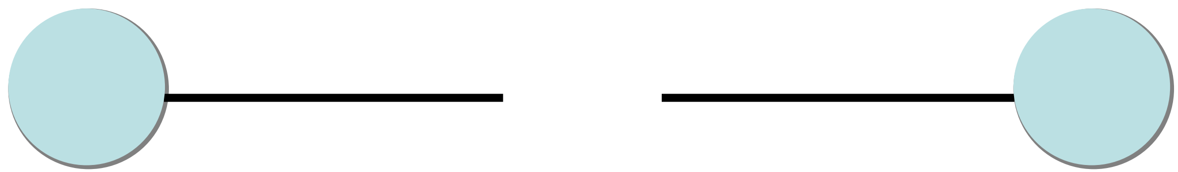
L= total number of links, d_i = number of nodes in module i

l_i = # of links in module i

$\frac{d_i^2}{4L}$ = expected number of links in module i given the number of nodes.



$$\frac{d_i}{2L} =$$ Probability that a link chosen at random ends in module i.



$\frac{d_i}{2L}$ $\frac{d_i}{2L}$

d

$$\frac{d_i}{2L}\cdot\frac{d_i}{2L}=\frac{d_i^2}{4L^2}=$$

Probability that the link is inside module i

$$\frac{d_i^2}{4L^2}\cdot L=\frac{d_i^2}{4L}=$$

Expected number of links in module i

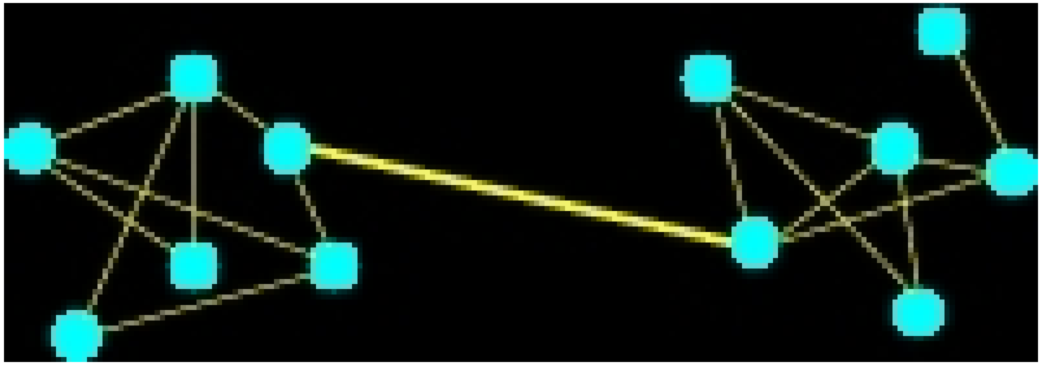
Girvan-Newman algorithm

M. Girvan & M.E.J Newman,
PNAS 99, 7821-7826 (2002)

Method: cut the link which connect different communities so as to isolate them.

How can we identify these links?
Betweenness!!

Link-betweenness: number of paths which must go through a given link



Iterative Algorithm:

1. Evaluate the link betweenness for each link
2. Eliminate the link with highest betweenness
3. Evaluate the link betweenness of all remaining links
4. GO TO 2

This algorithm gives a hierarchy of partitions,
which is the right one?

The best choice is the one with
highest modularity Q

M.E.J. Newman & M. Girvan, Phys. Rev. E
69, 026113 (2004)

General tools for network analysis

igraph
is open source and can be programmed in R, Python, Mathematica and C++   (also contains seven different community detection algorithms)
https://igraph.org/

Stanford Network Analysis Project (SNAP)
written in C++ with a Python interface
http://snap.stanford.edu/

NETWORKX
https://networkx.org/documentation/stable/reference/introduction.html

SCIKIT-NETWORK
https://scikit-network.readthedocs.io/en/latest/

Repositories

Networkrepository
Thousands of networks on different contexts (Bio-socio-economics…)
https://networkrepository.com/index.php

Networks
hundreds of benchmark networks
https://networks.skewed.de/about

Community detection tools

OSLOM
http://www.oslom.org/

INFOMAP
https://www.mapequation.org/infomap/

LOUVAIN
https://en.wikipedia.org/wiki/Louvain_method
https://scikit-network.readthedocs.io/en/latest/tutorials/clustering/louvain.html

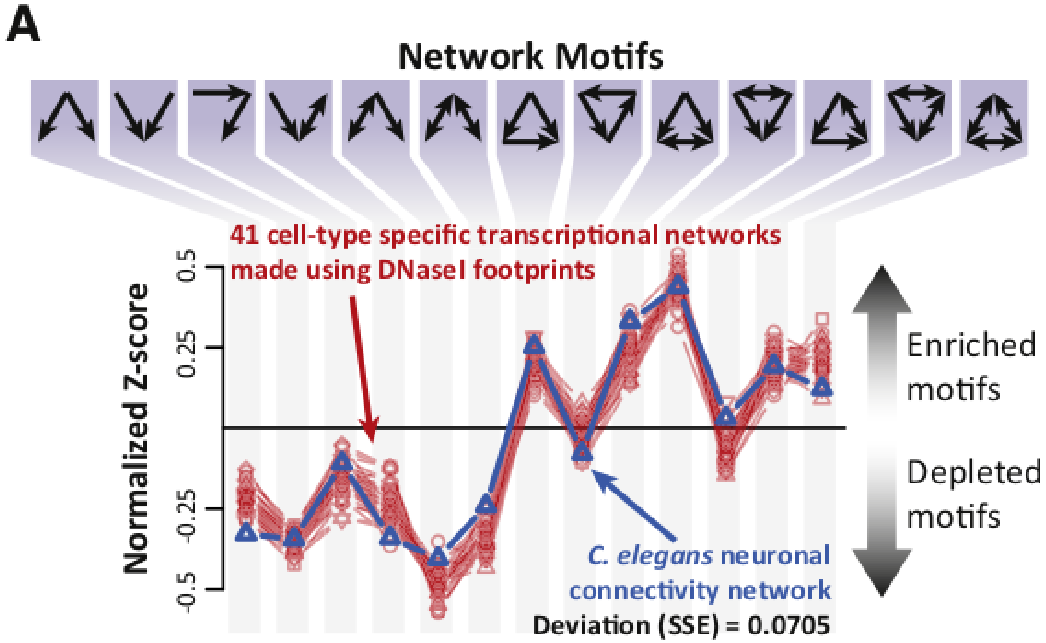
Motif detection tools

MFINDER

https://www.weizmann.ac.il/mcb/UriAlon/download/network-motif-software

Allows to estimate z-score with a reshuffling null model

$$z = \frac{x - \mu}{\sigma}$$



Useful References

1) M. Girvan; M. E. J. Newman (2002). "Community structure in social and biological networks". Proc. Natl. Acad. Sci. USA. 99 (12): 7821–7826. arXiv:cond-mat/0112110.

2) S. Fortunato (2010). "Community detection in graphs". Phys. Rep. 486 (3–5): 75–174. arXiv:0906.0612.

3) Peixoto, Tiago (2014) "Hierarchical Block Structures and High-Resolution Model Selection in Large Networks". Physical Review X. 4 (1): 011047. arXiv:1310.4377

4) Martin Rosvall; Carl T. Bergstrom (2007). "An information-theoretic framework for resolving community structure in complex networks". PNAS 104 (18): 7327–7331. arXiv:physics/0612035.

5) Milo et al. (2002) "Network Motifs: Simple Building Blocks of Complex Networks" Science 298, 824.