Bioinformatics III

Prof. Dr. Volkhard Helms
D. Gaidar, Maryam Nazarieh, D. Nguyen, T. Will
Saarland University
Chair for Computational Biology
Winter Semester 2015/2016

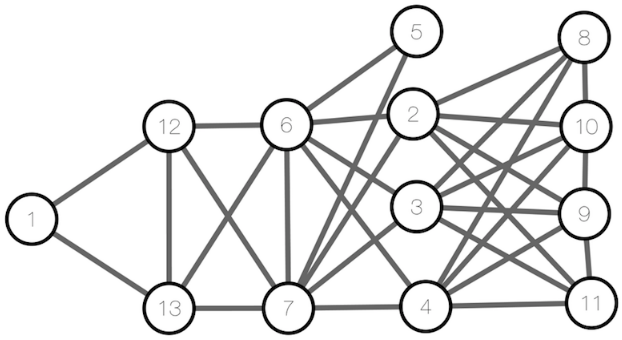
Exercise Sheet 4

Due: November 20, 2015 13:15

Submit your solutions on paper, hand-written or printed at the *beginning* of the lecture or in building E2 1, Room 3.09. Alternatively you may send an email with a single PDF attachment. If possible, please include source code listings. Additionally hand in all source code via mail to maryam.nazarieh@bioinformatik.uni-saarland.de.

Exercise 4.1: Graph Modular Decomposition (50pts)

A module of a graph $G = (V, E)$ is a set $X \in V$ of vertices where all vertices in X have the same neighbors in $V \backslash X$.



- Every single vertex of V satisfies the definition of a module is a trivial module. List the trivial modules in the graph given above.
- A module X composed of a subset of vertices of G overlaps another module Y composed of a subset of vertices of G if $X \cap Y \neq \{\}$, $X \backslash Y \neq \{\}$ and $Y \backslash X \neq \{\}$. Overlapping modules are called weak modules. List two weak modules in the given graph.
- The modules that do not overlap any other module are called strong modules. Hint: this graph contains one largest strong module of size 4. List all nodes which make it.
- A module is called series if all included nodes are direct neighbors of each other. List all nodes which make series modules in the given graph and specify the series modules.
- A module is called parallel if all included nodes are non-neighbors. List all nodes which make a parallel module and specify the parallel module.
- A module is called prime if it does not fulfill the conditions either of a series or of a parallel module. List all nodes which make a prime module and specify the prime module.
- Draw a modular decomposition tree of this graph formed by prime, series and parallel modules.

Exercise 4.2: Network Evolution (50pts)

Evolving networks are networks that change as a function of time, either by adding or removing nodes or links over time.

(a) Write a function to read the data from supplementary files as an undirected graph.

(b) Write a function which calculates the number of cliques of size 3, 4 and 5.

(c) For each network provided in the supplementary files, randomly insert or delete edges as function of time (one edge per time, $t = 100$, so that the total of edges remains about constant).

(d) Plot the number of cliques before and after each edge modifications as function of time.

(e) **Calculate P_value**
Start from the original network and randomly shuffle the edges as mentioned below for 100 times:

- for $2 * L$ steps, two edges $e_1 = (v_1, v_2)$ and $e_2 = (v_3, v_4)$ are randomly chosen from the network and rewired such that the start and end nodes are swapped, i.e. $e_3 = (v_1, v_4)$ and $e_4 = (v_3, v_2)$.
- Determine using ($P\_value < 0.05$) whether clique motif(of size 3, 4 and 5) are significantly enriched in the original network. (The P_value is calculated as ratio of the number of random times that a certain motif type is acquired more than or equal to its number in the real network.)

Have fun!