

Winter term 2015/16

Bioinformatics II

Assignment Sheet 5

If you have questions concerning the exercises, please write to our mailing list:
vl-bioinf@lists.iai.uni-bonn.de.

Please submit your results for this and all previous assignment sheets by December 1st, before class (9.30 a.m.). Submit a single .zip or .tar.gz archive that should contain all your code, as well as screenshots and answers to questions, clearly sorted by assignment. Send that archive by e-mail to mdkhatami@gmail.com.

Exercise 1 (Graph Visualization, 25 Points)

In the final part of the first project, you will learn how to use the graph visualization package Graphviz via its Python bindings.

- Create another reduction of our original dataset. Start with the reduction from sheet 4, which contained the four classes t-CS-sal, c-CS-sal, t-CS-mem, and c-CS-mem, but kept all proteins. Now, compute the average expression levels for each mouse by averaging all samples that have been taken from the same mouse. (5P)
- Install the Graphviz library and its Python interface. Generate and visualize some simple graph. You can find the software and its documentation at <https://pypi.python.org/pypi/graphviz>. (5P)
- Create a graph from the reduced data and visualize it. Represent each mouse as a node in the graph. Insert an edge between two mice whenever the Pearson correlation between their average expression levels exceeds the threshold $\rho > 0.98$. (4P)
- Modify the visual attributes of the graph nodes so that their color and shape reflect the classes the mice belong to. Pick the visual encoding so that there is a clear relationship to the colors and markers you used in the scatter plots in previous assignments. Modify the visual attributes of edges to reflect the magnitude of the correlation. (4P)
- Produce an alternative visualization in which nodes that correspond to the same class are clustered together as a subgraph. (3P)
- Answer the following questions and indicate which visualization made it easier to answer each one:
 - At the selected threshold, which nodes are disconnected from the rest of the graph? Do they relate to the outliers you observed in the PCA analysis on sheet 4? (2P)
 - Within which class are the nodes connected particularly strongly? Is there a class for which you see hardly any connections between members of that class? (2P)

Good Luck!