Prof. Jörg Stelling
D-BSSE
ETH Center UNG H14
Universitätstrasse 19
CH-8092 Zürich
https://www.csb.ethz.ch

Computational Systems Biology
Autumn 2025

Assignment 1

(Issue: 26-Sept-2025)

1 Reading the text source file

The text file `allyeastJan14_2002-nr.txt` (available on the course homepage) contains protein-protein interaction information. Specifically, each line contains an (undirected) edge of the graph, represented by two vertex names.

a) Use MATLAB to convert it to the standard graph format (v, e), where v contains the vertex names, and e is an upper triangular matrix representing the undirected edges, that is, $e(i, j) = 1 \iff$ edge from $v(i)$ to $v(j), i < j$.

b) Store the resulting variables (v, e) into a `.mat` file, using the `save` command.

2 Graph characteristics

a) Compute the vertex degrees and store them in a vector $vdeg$, where $vdeg(i)$ is the degree of vertex $v(i)$.

b) Plot the histogram of the vertex degrees in a log-log scale. A graph is called scale-free if the vertex degree distribution is following the power law. How can you see this in the log-log plot?

c) Map the vertices to k-cores by implementing the algorithm found in the lecture slides. The outcome should be a vector $vcore(1..n)$ that assigns a k-core at each one of the n vertices of the graph.

d) What is the number of k-cores that you have come up with and what is the minimum and the maximum k value observed?

Useful MATLAB functions (use `help <keyword>` or `doc <keyword>`): `find, sort, sum, unique, hist, loglog`

Submission:

Solutions in MATLAB and Python will be provided. Please address questions to

alix.moawad@bsse.ethz.ch