

CENG 465
Fall 2016-2017

Assignment #4

(Programming assignment about network analysis)

Due Date: January 22, 2017, 11:55PM
(submit source code via ODTU-CLASS)

Finding the Highest K-Core Subgraph in a Protein-Protein Interaction Network

In this assignment your goal is to find the highest k -core subgraph in a given protein-protein interaction network. The interaction network is given as a set of undirected edges that comprise a graph. There will be no self-edges in this graph. A k -core in a graph is a subgraph in which all the nodes in that subgraph have at least degree k . In other words each node in a k -core has at least k immediate neighbors. Use the $O(n^3)$ algorithm described in class to find the highest k -core subgraph of the given network. Use your implementation to find the highest k -core subgraph of the following human protein-protein interaction network.

http://www.ceng.metu.edu.tr/~tcan/ceng465_fl617/Assignments/HumanPPI_BioGRID.txt

Your program should output the following:

- a) The value of k of the highest k -core in the network
- b) The gene names of the nodes in the highest k -core

Hint: Using an adjacency matrix (or adjacency list) representation for the PPI network will be very efficient.

Notes: The network data is given as a list of undirected edges. Each line of the file represents an undirected edge that connects a pair of tab separated gene ids. There are no self-edges.

Example output:

For the network at:

http://www.ceng.metu.edu.tr/~tcan/ceng465_fl617/Assignments/yeast_PPI.txt

the output may look like the following:

The highest k -core is a 4-core and there are 85 proteins in that 4-core. The proteins are:
YOR095C YML029W YNL334C YNL333W YPL135W YNL219C YLR291C YBL091CA YGR267C
YOR284W YML064C YMR096W YDL164C YMR095C YKL189W YEL042W YGR261C YBR080C
YKL103C YGL153W YPL094C YER087CA YDL239C YOR353C YLR245C YJL025W YGL225W
YPR113W YJL218W YML055W YPL049C YBR072W YNL288W YDR100W YPL088W YGR177C
YBR261C YOL082W YJL058C YNL044W YNL315C YDR256C YJR003C YJL097W YOR380W
YPR105C YMR308C YMR079W YDR520C YFL060C YML042W YDR448W YAR027W YLR423C
YHL018W YAL032C YMR071C YPL070W YDR480W YGR120C YOL034W YCL028W YKR034W
YLR268W YGL053W YFL059W YML110C YBR057C YNL229C YIR038C YHR113W YHR111W
YIR033W YDR479C YER086W YNL263C YHL004W YPR159CA YPL020C YJL117W YNL189W
YJL199C YJL112W YCR007C YDR311W

An example program to read a network as an Adjacency List can be found at:

http://www.ceng.metu.edu.tr/~tcan/ceng465_f1617/Assignments/NetworkReader.java

Deliverables:

Your source code.

Submission:

Submit the deliverables via ODTU-Class before the deadline.

Late Submission Policy:

Penalty: 20 points per day.