

CENG 465
Spring 2020-2021

Assignment #4

(Programming assignment about network analysis)

Due Date: July 13, 2021, 11:59PM
(submit source code and your output 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 are no self-edges in the input interaction networks on which your program will be tested. A k -core in a graph is a subgraph in which all the nodes **in that induced subgraph** have at least degree k . In other words, each node in a k -core has at least k immediate neighbors. Given an input network, use the $O(n^3)$ algorithm described in class, to find the number of proteins for each value of k such that k is the highest k -core they belong to. Report your results for the following test network.

http://www.ceng.metu.edu.tr/~tcan/ceng465_s2021/Schedule/PathwayCommons.txt

Your program should output the number of proteins at each k -core starting with $k = 1$ and going until the number of proteins for larger values of k becomes 0 (zero).

Hint: If you use the Hashtable/Map/Dictionary based adjacency list representation for the PPI network we discussed in class, your program should report the result and finish in a few seconds.

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. You may use any language of your choice for implementing the assignment. You are also free to use available libraries.

Example outputs:

For the network at:

http://www.ceng.metu.edu.tr/~tcan/ceng465_s2021/Schedule/yeast.txt

the output should look like the following:

```
For k = 1 there are 1222 proteins.  
For k = 2 there are 442 proteins.  
For k = 3 there are 217 proteins.  
For k = 4 there are 85 proteins.
```

For the network at:

http://www.ceng.metu.edu.tr/~tcan/ceng465_s2021/Schedule/BioGRID.txt

the output should look like the following:

```
For k = 1 there are 1745 proteins.
For k = 2 there are 1215 proteins.
For k = 3 there are 916 proteins.
For k = 4 there are 743 proteins.
For k = 5 there are 641 proteins.
For k = 6 there are 591 proteins.
For k = 7 there are 539 proteins.
For k = 8 there are 562 proteins.
For k = 9 there are 452 proteins.
For k = 10 there are 473 proteins.
For k = 11 there are 439 proteins.
For k = 12 there are 350 proteins.
For k = 13 there are 389 proteins.
For k = 14 there are 352 proteins.
For k = 15 there are 311 proteins.
For k = 16 there are 286 proteins.
For k = 17 there are 249 proteins.
For k = 18 there are 274 proteins.
For k = 19 there are 241 proteins.
For k = 20 there are 213 proteins.
For k = 21 there are 202 proteins.
For k = 22 there are 242 proteins.
For k = 23 there are 235 proteins.
For k = 24 there are 190 proteins.
For k = 25 there are 174 proteins.
For k = 26 there are 170 proteins.
For k = 27 there are 192 proteins.
For k = 28 there are 191 proteins.
For k = 29 there are 176 proteins.
For k = 30 there are 217 proteins.
For k = 31 there are 228 proteins.
For k = 32 there are 242 proteins.
For k = 33 there are 215 proteins.
For k = 34 there are 165 proteins.
For k = 35 there are 128 proteins.
For k = 36 there are 119 proteins.
For k = 37 there are 101 proteins.
For k = 38 there are 121 proteins.
For k = 39 there are 81 proteins.
For k = 40 there are 70 proteins.
For k = 41 there are 74 proteins.
For k = 42 there are 68 proteins.
For k = 43 there are 98 proteins.
For k = 44 there are 30 proteins.
For k = 45 there are 65 proteins.
For k = 46 there are 23 proteins.
For k = 47 there are 18 proteins.
For k = 48 there are 105 proteins.
```

For $k = 49$ there are 41 proteins.
For $k = 50$ there are 44 proteins.
For $k = 51$ there are 4 proteins.
For $k = 52$ there are 112 proteins.
For $k = 53$ there are 101 proteins.
For $k = 54$ there are 4 proteins.
For $k = 55$ there are 1 proteins.
For $k = 57$ there are 2 proteins.
For $k = 58$ there are 7 proteins.
For $k = 60$ there are 3 proteins.
For $k = 61$ there are 1 proteins.
For $k = 62$ there are 2 proteins.
For $k = 63$ there are 74 proteins.

Deliverables:

A zip file containing your source code and your output for the test network as a text file.

Submission:

Submit the deliverables via ODTU-Class before the deadline.

Late Submission Policy:

Penalty: 15 points per day.