

Biological Computation

חישוב ביולוגי

Exercise #2

הגשה ביחידים או זוגות בלבד עד לתאריך 1.6.23

Note: You can discuss the work with other students however you should write all code by yourself / with your exercise partner. Please document your code and also prepare a short (no more than 2 pages) explanation on how the code works and what were the main design and implementation decisions you made. If any code that a student did not write is used, explain where it is taken from and why it's needed.

***In addition to submitting to Moodle prepare a GitHub repository with a readme on how to run the code.**

- 1) a) Write a program (in your favorite programming language) that gets as input a positive integer n and generates all connected sub-graphs of size n .

The output should be a textual file of the following form:

```
n=2
count=2
#1
1 2
#2
1 2
2 1
```

The first two lines output n and the total number (count) of different sub-graphs of size n . Then the sub-graphs themselves are given each starting with a line labelled $\#k$ for motif number followed by all edges, each line $i\ j$ means an edge from source i to target j .

- b) Output the result of your program for $n = 1$ to 4.

c) What is the maximal number n for which your program can complete successfully within no more than 1 hour of computing time?

d) What is the maximal number n for which your program can complete successfully within 2,4,8 hours of computing time?

2) Write a program that gets as input positive integer n and a graph of the format:

1 2

2 3

1 4

The graph in the example contains 4 vertices 1,2,3,4 and directed edges (1,2) (2,3) (1,4). The program should output all sub-graphs of size n and count how many instances appear of each motif. The format of the output of the identified sub-graphs should be like in question 1, where in the line after # k should appear the count of number of instances, count= m if the motif appears m times. Output count=0 if a motif does not appear in the graph.