SNA 2018 - HW1

Einav Ram - Ziv Levi - Arnon Kahani

203546445 – 304974645- 305509556

Data structure and clustering coefficient algorithm logic

Data structure

1. To represent the input network, we choose a combination of a HashMap and ArrayList. Each vertex is an entry (key) in the HashMap and the adjacent vertices (neighbors) are kept in an ArrayList as the value of the vertex key.
2. The decision for this data structure was due to several considerations:
    a. Finding the adjacent vertices quickly
    b. The amount of the neighbors is given by the size of the ArrayList
    c. Counting the number of "mutual-friends" required only checking if they exist in both lists.
    d. Relativity to adjacency matrix (AM) this data structure is small in memory, we considered large network such as Facebook were the AM will be very sparse and will require large amount storage.

Logic

1. The full algorithm can be seen in the implementation but the high level idea of the algorithm is to iterate over all nodes i.e. entries in the HashMap and for each "neighbor" in the ArrayList (value of the entry) go over the node neighbors and count how many "mutual-friends" they have.
2. Pseudo code for section 2.a
    a. graph = HashMap<Integer,ArrayList<Integer>>
    b. for node in graph:
        i. counter ← 0
        ii. k = node.value.length
        iii. for neighbor_outer in node.value:
            1. neighbor_of_neighbors ← graph.get(neighbor)
            2. for neighbor in node.value:
                a. If neighbor_of_neighbors.contains(neighbor):
                    i. counter ← counter + 1
        iv. coefficients ← $\frac{counter}{k*(k-1)}$

Networks

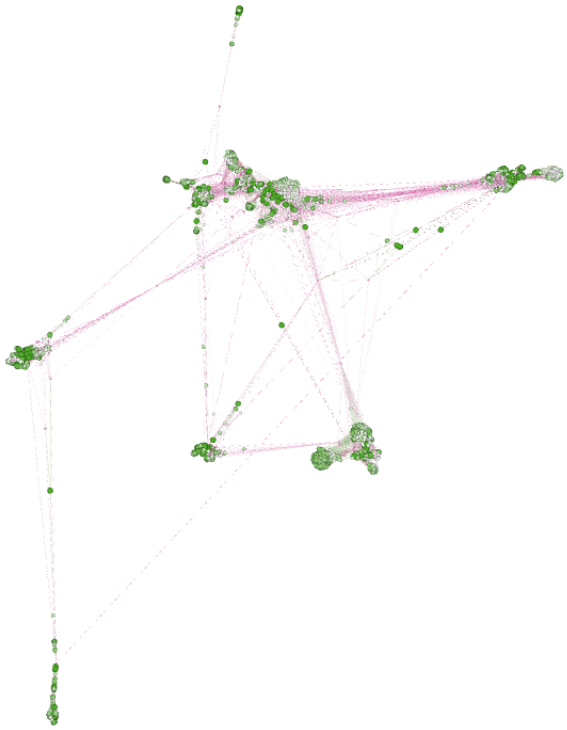
The network images were crated in Neo4J.
ACC - The average clustering coefficient
Colors = Green(High CC) → Purple(Low CC)
Size = Large (High CC) → Small (Low CC)

Egonets-Facebook

URL: https://snap.stanford.edu/data/egonets-Facebook.html

Description: Dataset from Facebook collected from survey participants, the data contains friend list and features per node i.e. Facebook account that are encoded to anonymized
the data and the participants account is also encoded as a number for the same reason.
Nodes: 4039
Edges: 88234
ACC: 0.6055467186200871

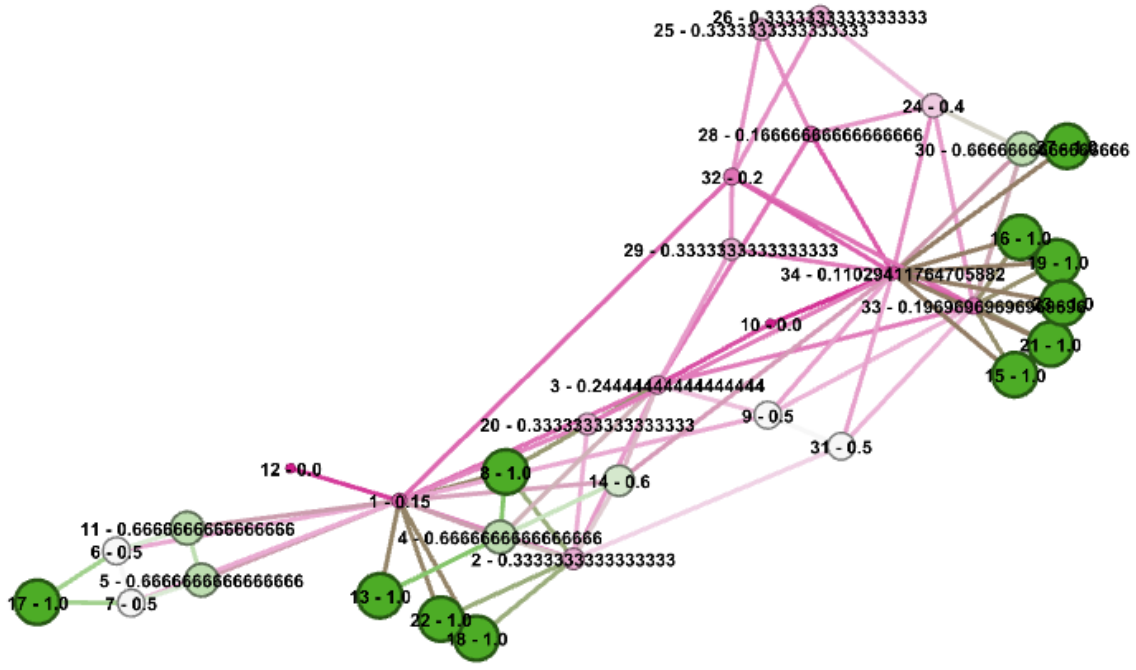
Node ID	Clustering coefficient
71	1
70	1
63	1
52	1
47	1
46	1
44	1
42	1
35	1
33	1



Karate

Nodes: 34
Edges: 78
ACC: 0.5706384782076823

Node ID	Clustering coefficient
27	1
23	1
22	1
21	1
19	1
18	1
17	1
16	1
15	1
13	1



Les Mesrables

URL: https://github.com/gephi/gephi/wiki/Datasets
Description: Dataset from coappearance weighted network of characters in the novel Les Miserables
the data and the participants account is also encoded as a number for the same reason.
Nodes: 77
Edges: 254
ACC: 0.5731367499320134

Node ID	Clustering coefficient
33	1
30	1
22	1
21	1
20	1
19	1
18	1
17	1
12	1
3	1