COMP 314: Algorithms and Complexity
Lab work 5: Using Graph Libraries in Python

1 Purpose

To be familiar with Python Graph Libraries.

2 Tasks

Students are required to accomplish the following tasks using a Python graph library such as Networkx[1], igraph[2], pygraphviz[3], python-graph[4] etc.

1. Download a network data from Network Repository[5]. Choose a small network, e.g. a graph with about 50 nodes and less than 1000 edges. There are several small neworks in Animal Social Networks[6] section. Import this graph using Networkx or other library. Draw this graph, e.g. using matplotlib. Then, compute the following network properties:

   (a) **Number of nodes and edges**

   (b) **Average degree**

   (c) **Density**: Ratio of the number of edges to the number of possible edges in a network, given by:

   $$D = \begin{cases} \dfrac{2|E|}{|V|(|V|-1)}, & \text{if } |V| > 1 \\ 0, & \text{otherwise} \end{cases}$$

   where E is the set of edges, and V is the set of vertices.

   (d) **Diameter**: Length of the shortest path between the most distanced nodes.

[1] `https://networkx.org/`
[2] `https://igraph.org/`
[3] `https://pygraphviz.github.io/`
[4] `http://code.google.com/p/python-graph/`
[5] `http://networkrepository.com/`
[6] `http://networkrepository.com/asn.php`

(e) **Clustering coefficient**: Average of the clustering coefficients of all the nodes. The clustering coefficient of the i^{th} node is

$$C_i = \begin{cases} \dfrac{2e_i}{k_i(k_i-1)}, & \text{if } k_i > 1 \\ 0, & \text{otherwise} \end{cases}$$

where k_i the number of neighbours of the i^{th} node, and e_i is the number of connections between these neighbours.

Note: these network properties must have been already implemented in many of the graph libraries. You are required to implement them yourself using the given formulas. You may use the implementation provided by the library only for checking if your implementation is correct.

2. Download 5 other networks (with at least 5000 nodes) from Network Repository[7] or Stanford Large Network Dataset Collection[8].

   (a) Compute the network properties listed in Task 1 for these networks.

   (b) Plot the degree distribution of all of these networks. The degree distribution $P(k)$ of a network is the fraction of nodes in the network with degree k.

   Answer the following questions:

   (a) From the network properties, what can you say about the networks you have selected?

   (b) Did you find any pattern in the degree distributions of the networks? In any case, can you come to any conclusion about the networks from their degree distribution?

[7] `http://networkrepository.com/`
[8] `https://snap.stanford.edu/data/`