Assignment for Project
Darkweb Analysis using Graph Neural Networks

1 Assignment A

Task: For a given graph, develop a clustering algorithm that can divide nodes into different groups. These groups are formed such that nodes within a group are highly similar, whereas nodes between groups are dissimilar.

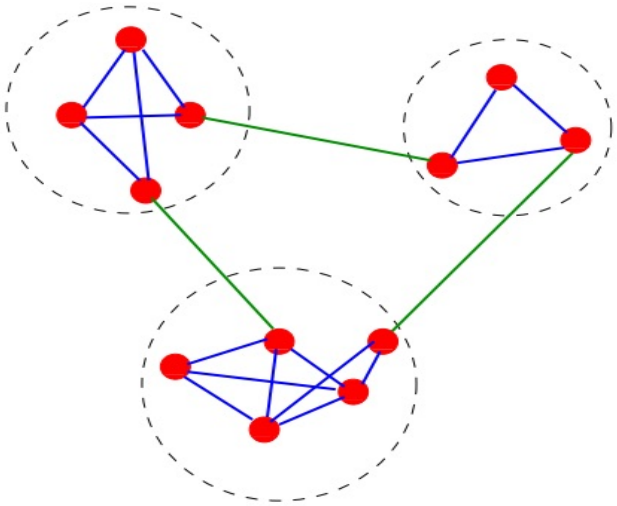

Figure 1: A simple graph with three clusters (dashed circles)

In the given example (shown in Fig.1), the clusters are formed based on the densities of links among nodes. Nodes within the clusters are densely connected, whereas nodes between clusters are sparsely connected.

Students are advised to explore at least 2-3 ways to clusters the nodes in a given network and implement their algorithms. It is allowed to use any traditional algorithm as well as machine learning based algorithms from the state-of-the-art literature to solve the given problem.

Please use the "Facebook Large Page-Page Network" dataset to test your algorithm. You can download dataset from
`https://snap.stanford.edu/data/facebook-large-page-page-network.html`

Deliverable:

- Source code.
- A maximum of 2-3 page report that discuss the implemented method and results.
- The third page of the report can have references of research papers that you studied for the project.

2 Assignment B

Please note that this assignment is designed only for those students who do not possess programming skills at this moment, but will gradually learn over the course of time. This is an exciting opportunity for the students to learn and explore a new research domain. We highly encouraged them to use their theoretical knowledge to the solve the given problem

Task: node2vec is a neural network based learning algorithm. It learns continuous and low-dimensional feature vector for each node in a given network. The feature vector of a node is learned by maximizing the likelihood of network neighborhoods of that node. The algorithm is explained in the article attached with email.

The article discuss the following three main concepts:

- classical search strategies (BFS, DFS) and their shortcomings
- flexible neighborhood strategy using random walk
- skip-grap neural network architecture to train feature learning model

We want students to do in-depth study the node2vec algorithm and understand the underlying mathematical concepts. Using your graph theory and mathematical skills, prepare explanation of the each step of the algorithm in a manner that it is easy to understand and implementable. In addition, we would highly appreciate if you could come up with some improvements or new suggestions in the algorithm.

The pdf of article is attached with email. The code of node2vec is available at `http://snap.stanford.edu/node2vec/`

Deliverable:

- Presentation slides. One can include handwritten notes also to explain maths involved in the algorithm.
- A maximum of 2-3 page report that discuss and summarize the benefits of using node2vec over other algorithms.

We highly recommend submitting the work even if it is not fully complete or at the initial stage. Please be noted that the assignment is not the only selection criteria.

Do not hesitate to clear the doubts related to the assignment; You can reach us at `vinti.agarwal@pilani.bits-pilani.ac.in`. -

We wish you the best!!