

Joseph Allen
Final Project Writeup

The `'create_graph'` function is designed to read data from a CSV file named "Placement_Data_Full_Class.csv" and create a graph using the `'petgraph'` library. The necessary imports are included at the beginning of the code snippet: `'petgraph::Graph'`: This is used to create a graph data structure, `'csv::Reader'`: this is used to read CSV data, and `'serde::Deserialize'`: This is used to deserialize the CSV data into a struct. A custom struct `'Record'` is defined to represent each row of the CSV data. The struct has fields corresponding to the columns in the CSV file. With `'Reader::from_path("Placement_Data_Full_Class.csv").unwrap()'`. The CSV data is deserialized using a for loop. For each row in the CSV data: the row is deserialized into a `'Record'` struct using `'result.unwrap()'`. A new node is added to the graph using `'graph.add_node(record.sl_no)'`. The `'sl_no'` field of the `'Record'` struct is used as the identifier for the node as it corresponds to the user id.

Next, is the clustering module which reads a CSV file containing student data, extracts the `etest_p` and `status` columns, and performs K-means clustering with 2 clusters (Placed and Not Placed). The cluster labels are then printed. Then it creates a graph where each student is connected to their degree type, and computes the degree centrality for each node (student). The degree centrality is then printed for each node. This means how employment is clustered based on the students' score on their employment tests.

The `average_distance` module calculates the average distance between nodes in a given graph using DFS. The `compute_average_distance` takes a reference to a graph as an argument and returns a `f64` value representing the average distance between the nodes. `Total_distnsnce` and `count` are mutable variables that keep track of the total distance and number of edges traveled. For each visited node the distance between the current node and visited node is calculated using `graph.find_Edge(node, nx).unwrap()`.

The distance module reads the CSV file containing student data, and creates a graph in which every student is connected to their degree type and computes the degree centrality for each node. It also calculates the pairwise Euclidean distances between the data points, and then calculates the average distance using the extracted `etest_p` and `mba_p` columns which is the employability and mba grade of the student.