**Project Proposal**

Title: "Global Education Network Analysis: Graph Clustering and Representative Selection"

Objective:

The objective of the project is to analyze the UN global education statistics dataset from 2020 using graph clustering and partitioning techniques to identify clusters of similar countries based on education-related metrics and to find the best representatives for each cluster. Additionally, the project will evaluate how satisfied you are with the selection of representatives, given your previous beliefs about the network.

### **Project Outline:**

1. Data Preprocessing:
   * Load the CSV data from the UN global education statistics dataset.
   * Clean the data by handling missing values and filtering out irrelevant columns.
   * Normalize the data to ensure that different metrics are on a comparable scale.
2. Graph Construction:
   * Construct a graph where each node represents a country or region.
   * Create edges between nodes based on similarity in education metrics (e.g., enrollment rates, literacy rates, expenditure, student-teacher ratios).
   * Calculate edge weights based on the similarity of the metrics using a similarity metric such as cosine similarity.
3. Graph Clustering and Partitioning:
   * Apply a community detection algorithm such as the Louvain method or modularity-based clustering to the graph to identify clusters of similar countries.
   * Partition the graph into clusters and evaluate the modularity to measure the quality of the clustering.
4. Identify Best Representatives:
   * Within each cluster, identify the best representatives using centrality measures (e.g., degree centrality, closeness centrality).
   * Select the top k representatives from the clusters based on their centrality scores.
5. Evaluate Satisfaction:
   * Compare the selected representatives with your prior beliefs about the network (e.g., which countries you expected to be key players in education).
   * Evaluate how satisfied you are with the selection of representatives based on your expectations and the actual results.
6. Data Visualization:
   * Visualize the graph and the clusters using graph visualization tools.
   * Create visualizations of the selected representatives and their positions within the graph.
7. Analysis and Insights:
   * Analyze the clusters and representatives to identify trends, patterns, and outliers in global education statistics.
   * Provide insights into global education policies and how different countries compare.
8. Report and Documentation:
   * Document your findings, methodology, and insights in a report.
   * Provide recommendations for future research and potential policy implications.

### **Potential Challenges:**

* Data Cleaning: Handling missing values and inconsistencies in the dataset.
* Choosing the Right Similarity Metric: Selecting an appropriate metric for measuring similarity between countries.
* Clustering Algorithm Selection: Choosing the best clustering algorithm and parameters for the given data.
* Interpretation of Results: Making sense of the clusters and representatives in a meaningful way.

### **Conclusion:**

This project aims to provide a better understanding of global education statistics through graph clustering and partitioning. By finding meaningful clusters and representatives, you can gain insights into educational trends and potential areas for policy improvement.

**Project Written Submission**

The project leverages a data set containing global education statistics from the United Nations (UN) for the year 2020. The main goal is to analyze the data using graph clustering and partitioning methods to gain insights into relationships between different countries or regions based on their education statistics. The project involves creating a network where each node represents a country or region, and the connections (edges) between the nodes are weighted based on similarity in education metrics. The aim is to identify clusters within the network and evaluate how well the chosen clusters represent different aspects of the global education landscape.

The code provided achieves this through the following key steps:

1. Data Loading and Preprocessing:
   * The program begins by loading data from a CSV file containing the UN global education statistics for 2020.
   * It processes the data to handle missing values and prepares it for graph construction.
2. Graph Construction:
   * Once the data is loaded, the code constructs an undirected graph (petgraph::Graph) where each node represents a country or region, and the edges between nodes are weighted based on some metric (e.g., similarity in education statistics).
   * A mapping between node indices and country names is created to facilitate the interpretation of clusters later.
3. Graph Clustering and Partitioning:
   * The program uses the Louvain method for community detection to cluster the graph nodes.
   * It produces clusters of nodes that can represent groups of countries or regions with similar education statistics.
4. k-Means Clustering:
   * In addition to graph clustering, the program uses the smartcore library to perform k-means clustering on the graph's adjacency matrix.
   * This additional clustering method provides another layer of analysis on the data.
5. Counting Nodes and Edges:
   * The program includes a function to count the number of nodes and edges in the graph, providing insights into the size and complexity of the network.
6. Testing:
   * Tests are written to verify the functionality of each key function in the program.
   * These tests validate data loading, graph construction, clustering, and other functions to ensure the code behaves as expected.

The project provides a comprehensive analysis of the global education data set by applying advanced clustering and partitioning techniques. These methods allow for the identification of patterns and trends within the data, offering valuable insights into the similarities and differences in education across various countries and regions. The code serves as a robust starting point for further analysis and customization depending on the project's goals and specific requirements.