NATIONAL UNIVERSITY OF COMPUTER & EMERGING SCIENCES
FAST - PESHAWAR CAMPUS

Subject: AL 2002 - Artificial Intelligence Lab
Instructor: Muhammad Saood Sarwar

Lab Task: Data Manipulation and Graphs

Question 1

You are provided with a dataset named `indian-cities-dataset.csv`, which consists of the following columns:

- **source**: The starting city of a connection.
- **target**: The destination city of a connection.
- **distance**: The distance between the source and target cities.

Your task is to analyze this dataset using graph by performing the following operations:

1 Tasks

1. **Graph Representation**:
   - Construct a graph using NetworkX where cities act as nodes and distances act as weighted edges.
   - Visualize the graph.
2. **Graph Analysis**:
   - Compute and display the shortest path between two user-defined cities.
   - Find the city with the highest number of connections.
   - Calculate the average shortest path length between all connected cities.
3. **Statistical Calculations**:
   - Find the longest and shortest direct connection in the dataset.
   - Calculate the average, median, and standard deviation of all distances.
4. **Advanced Operations**:
   - Identify if the graph is connected.
   - Find all possible paths between two given cities.
   - Convert the graph into an adjacency matrix and display it.

Question 2

You are required to analyze the distances between cities in the dataset using NumPy and Matplotlib. Perform the following tasks:

1. **Statistical Analysis of Distances:**
   - Load the distances from the dataset into a NumPy array.
   - Compute and print the mean, median, mode, and standard deviation of distances.
   - Identify the top 5 largest distances and their corresponding city pairs.
2. **Data Visualization:**
   - Create a histogram to show the distribution of distances.
   - Generate a scatter plot where the x-axis represents different city connections and the y-axis represents the distance.
   - Use a line plot to visualize the cumulative sum of distances.
3. **Data Transformation:**
   - Normalize the distance values using Min-Max scaling.
   - Apply log transformation to handle skewness in distance data.

Question 3

Write a Python program to plot two or more lines and set the line markers. The code snippet gives the output shown in the figure 1.

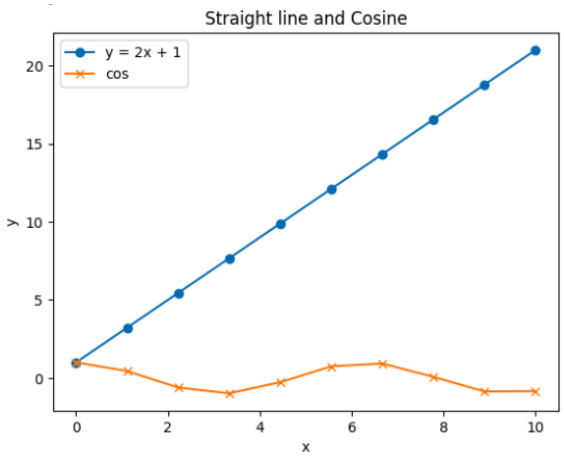

Figure 1: Chart

2 Instructions

- Use Python with the following libraries: `pandas, networkx, matplotlib, numpy`.
- Ensure your code is modular and well-commented.
- Submit your Jupyter Notebook along with pdf.