**Project Proposal: Network Analysis of Canadian Wildfires: Investigating Causes**

Atharva Jagtap (39283783), Ishika Agarwal (21368931), Mariya Putwa (28826089), Prateek Balani (68220110)

**Problem Statement:**

Wildfire propagation is a complex and dynamic process influenced by various environmental factors such as wind direction, topography, and vegetation. Traditional models often struggle to capture these interdependencies, limiting the effectiveness of emergency response strategies.

Our project proposes using network analysis to model wildfire propagation by treating geographic regions as nodes and critical factors influencing fire spread (e.g., causes, weather) as edges.

**Research Questions:**

1. What are the most common causes of wildfires in different provinces, and how do these causes affect fire spread and clustering?
2. Which year had the largest fires, and what were their characteristics?
3. Are larger wildfires more likely to have multiple causes, and do they play a more significant role in connecting different fire clusters?
4. Which regions experience the most fires, and what are the primary causes by region?
5. How do provincial policies or geographical features influence the clustering of wildfires and the dominant causes of fire events in Canada?

**Data Collection:**

We will utilize publicly available data from the Canadian government [Wildfire Dataset](https://open.canada.ca/data/en/dataset/a221e7a0-4f46-4be7-9c5a-e29de9a3447e), which includes:

* Locations of wildfires.
* Real-time weather data when the fire was active.
* Fire size, start and end times.

This dataset is crucial for analyzing wildfire behavior and optimizing emergency responses to reduce response times and mitigate wildfire damage.

**Network Definition:**

* **Nodes**: Fire events, defined by Fire Number, Longitude, and Latitude. Each node's weight is based on fire size.
* **Edges**: Causes of fires, such as lightning, human activity, etc.
* **Clusters**: Wildfire clusters will be identified based on provinces and geographical proximity.

**Metrics:**

We will use the following network metrics to analyze the data:

* **Degree Centrality**: Identify fire events with the most connections (i.e., influential events) to understand their role in fire spread.
* **Betweenness Centrality**: Determine which fire events act as bridges or critical points in the spread or response network.
* **Clustering Coefficient**: Measure the tendency of wildfire events to form localized clusters.
* **Network Density**: Evaluate the overall interconnectedness, assessing how frequently wildfires occur in close proximity.

**Analysis Plan:**

1. **Data Preprocessing**: Clean and filter the dataset to focus on relevant attributes like fire location, size, cause, response details, etc.
2. **Network Construction**: Build a network of wildfire incidents based on geographical proximity and factors influencing fire spread (e.g., causes, weather).
3. **Network Metrics Calculation**: Apply metrics such as Degree Centrality, Betweenness Centrality, Clustering Coefficient, and Network Density.
4. **Visualization**: Generate visualizations that depict the relationships between wildfire events and highlight important nodes.
5. **Results Interpretation**: Use network metrics and visualizations to answer research questions and extract key insights about wildfire causes, spread, and clustering.

**Work Breakdown Structure:**

All team members will work equally on every aspect of the project, from preprocessing to interpretation. We will collaborate on the project using GitHub for version control, code sharing, and collaborative work on data analysis, network construction, and visualizations. Each member will regularly contribute to the repository.