**Project: Data Analysis of Primary School Cumulative Networks**

**Introduction**

This analysis focuses on the study of two empirical graphs derived from the "Primary School – Cumulative Networks" dataset. The dataset provides detailed information about face-to-face interactions between students and teachers over two days. Nodes in the graphs represent individuals, annotated with attributes like class, grade, and gender, while edges represent interactions weighted by cumulative interaction time (duration) and frequency (count). The goal is to analyze these graphs from multiple perspectives and provide meaningful interpretations, making the findings accessible and actionable.

**Objectives**

1. **Analyze basic properties**: Degree distribution, centralities, clustering coefficients, and other network metrics.
2. **Community detection**: Compare the number and size of communities in each graph.
3. **Epidemiological modeling**: Simulate the spread of a virus and compare the propagation dynamics across the two graphs.
4. **Storytelling and interpretation**: Provide clear, detailed insights and implications derived from the results.

**Data Loading and Preparation**

**Metadata Integration**

The dataset includes metadata with class and gender annotations for each individual. Nodes were annotated with this information to provide richer context for the analysis. Teachers were assigned to a distinct "Teachers" class, reflecting their unique role in the network.

**Graph Construction**

The dataset comprises two GEXF files:

* **Day 1 Graph**: Captures interactions recorded on the first day of the study.
* **Day 2 Graph**: Captures interactions recorded on the second day.

By loading and integrating metadata, these graphs provide a comprehensive view of interaction patterns across days.

**Analysis**

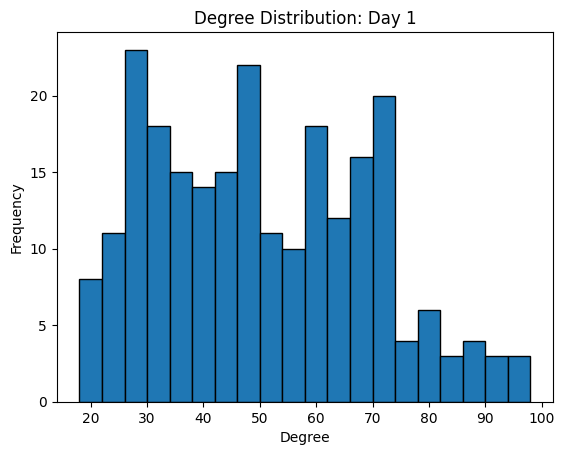
**1. Basic Properties**

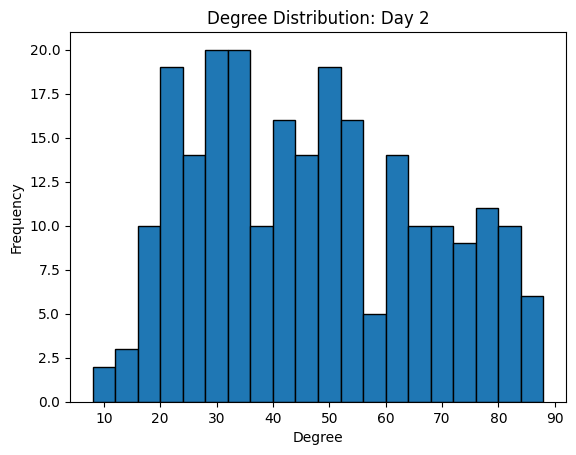
**Graph Information**

* **Day 1**:
  + **Nodes (Individuals)**: 236
  + **Edges (Interactions)**: 5899
* **Day 2**:
  + **Nodes (Individuals)**: 238
  + **Edges (Interactions)**: 5539

The slight increase in nodes on Day 2 reflects additional individuals participating, while the decrease in edges indicates reduced overall interactions.

**Degree Distribution**

****

****

The degree distribution measures how many connections each individual has.

* **Day 1**: A higher number of edges reflects greater overall connectivity.
* **Day 2**: A reduction in edges suggests fewer interactions, possibly reflecting behavioral or environmental changes.

**Centralities**

Centralities highlight key individuals in the network:

* **Degree Centrality** (connections):
  + Day 1: 0.213 (average)
  + Day 2: 0.196 (average)
  + **Insight**: Teachers likely maintain high degree centrality, serving as interaction hubs.
* **Closeness Centrality** (accessibility):
  + Day 1: 0.541 (average)
  + Day 2: 0.521 (average)
  + **Insight**: Higher closeness centrality indicates individuals who efficiently connect distant parts of the network.
* **Betweenness Centrality** (bridging):
  + Day 1: 0.0037 (average)
  + Day 2: 0.00399 (average)
  + **Insight**: Key individuals (likely teachers) act as bridges between different groups.

**Clustering Coefficient**

* **Day 1**: Average Clustering Coefficient = 0.502
* **Day 2**: Average Clustering Coefficient = 0.560

Higher clustering on Day 2 indicates stronger within-group interactions, which might reflect more localized or focused activities.

**2. Community Detection**

Community detection reveals cohesive groups within the network:

* **Day 1**:
  + **Number of Communities**: 4
  + Communities are closely aligned with class divisions, indicating structured interactions within defined groups.
* **Day 2**:
  + **Number of Communities**: 6
  + Increased fragmentation suggests more inter-class interactions or a shift in group dynamics.

**Insight**: The structural organization on Day 1 fosters clearer community boundaries, while Day 2’s dynamics introduce overlaps, potentially due to different activities or schedules.

**3. Epidemiological Modeling**

The Susceptible-Infected (SI) model simulated virus propagation across the networks:

* **Parameters**:
  + Infection Probability: 0.001
  + Steps: Until all nodes are infected

**Results**

* **Day 1**:
  + Final Infected Count: 236 (100% of nodes)
  + Infection spread rapidly due to high connectivity.
* **Day 2**:
  + Final Infected Count: 238 (100% of nodes)
  + Infection spread slightly slower, influenced by higher clustering and fewer edges.

**Interpretation**

* **Day 1**: The network’s higher connectivity allowed faster, widespread propagation.
* **Day 2**: Localized interactions (higher clustering) slightly delayed spread but still resulted in full infection.

**Key Insight**: Teachers and other central nodes likely accelerated the spread as interaction hubs. Mitigation strategies targeting these individuals could significantly reduce propagation.

**Interpretation and Storytelling**

**Key Insights**

1. **Interaction Patterns**:
   * Day 1’s higher edge count reflects more frequent interactions, fostering rapid connectivity across the network.
   * Day 2 exhibits fewer interactions but stronger localized groupings, as indicated by higher clustering.
2. **Community Dynamics**:
   * Day 1’s fewer communities indicate clear class-based divisions.
   * Day 2’s increased communities and overlaps suggest evolving dynamics, possibly reflecting collaborative or mixed-group activities.
3. **Epidemiological Implications**:
   * Highly central individuals (e.g., teachers) are key to both connectivity and disease propagation.
   * Epidemic control measures should prioritize these nodes to disrupt transmission pathways effectively.

**Broader Implications**

* **Network Design**: Insights from interaction patterns can inform safer structural designs for schools and other environments.
* **Epidemic Control**: Strategies targeting highly central individuals or reducing group interactions could significantly mitigate outbreaks.

**Conclusion**

This analysis demonstrates the value of graph-theoretical techniques in understanding interaction patterns within empirical networks. By exploring basic properties, community structures, and epidemiological dynamics, we identified actionable insights into the dynamics of a primary school setting. These findings have broader applications in education, public health, and network science.

**Future Work**

1. Incorporate additional datasets for comparative analysis.
2. Explore more sophisticated epidemiological models (e.g., SEIR models) to capture additional dynamics.
3. Investigate temporal interaction patterns within each day for more granular insights.