About Problem

Introduction:

- **Community detection is vital in network analysis, aiming to reveal underlying structures by identifying clusters of densely interconnected nodes.**
- **It is applicable across various domains like social networks, biological networks, transportation networks, and information networks.**

Domain and Use Case:

- In social networks such as Facebook and Twitter, helps in identifying user groups with similar interests
- In biological networks,it aids in understanding protein-protein interactions and gene regulatory networks, offering insights into biological processes and diseases.
- In transportation networks, it assists in identifying closely linked locations or traffic patterns for optimizing routes and resource allocation.

Importance and Relevance:

- Community detection is crucial for understanding organizational principles and dynamics within complex systems.
- It provides insights into information flow, behavior propagation, and resource distribution in networks, benefiting various applications like social network analysis, targeted marketing, disease spread modeling, and network optimization.

Challenges

- Networks are often massive, sparse, and dynamic, making it challenging to identify meaningful communities efficiently.
- Overlapping communities and noise in data add complexity and increase the risk of identifying spurious or misleading communities.

Data-Driven Solutions

- Leveraging large datasets within networks.
- Machine learning techniques can adapt to the complexities of real-world networks, learning from data to detect communities in an automated and scalable manner.
- Integration of additional information such as node attributes or temporal dynamics enhances the accuracy and relevance of community detection outcomes.

Current Scenario

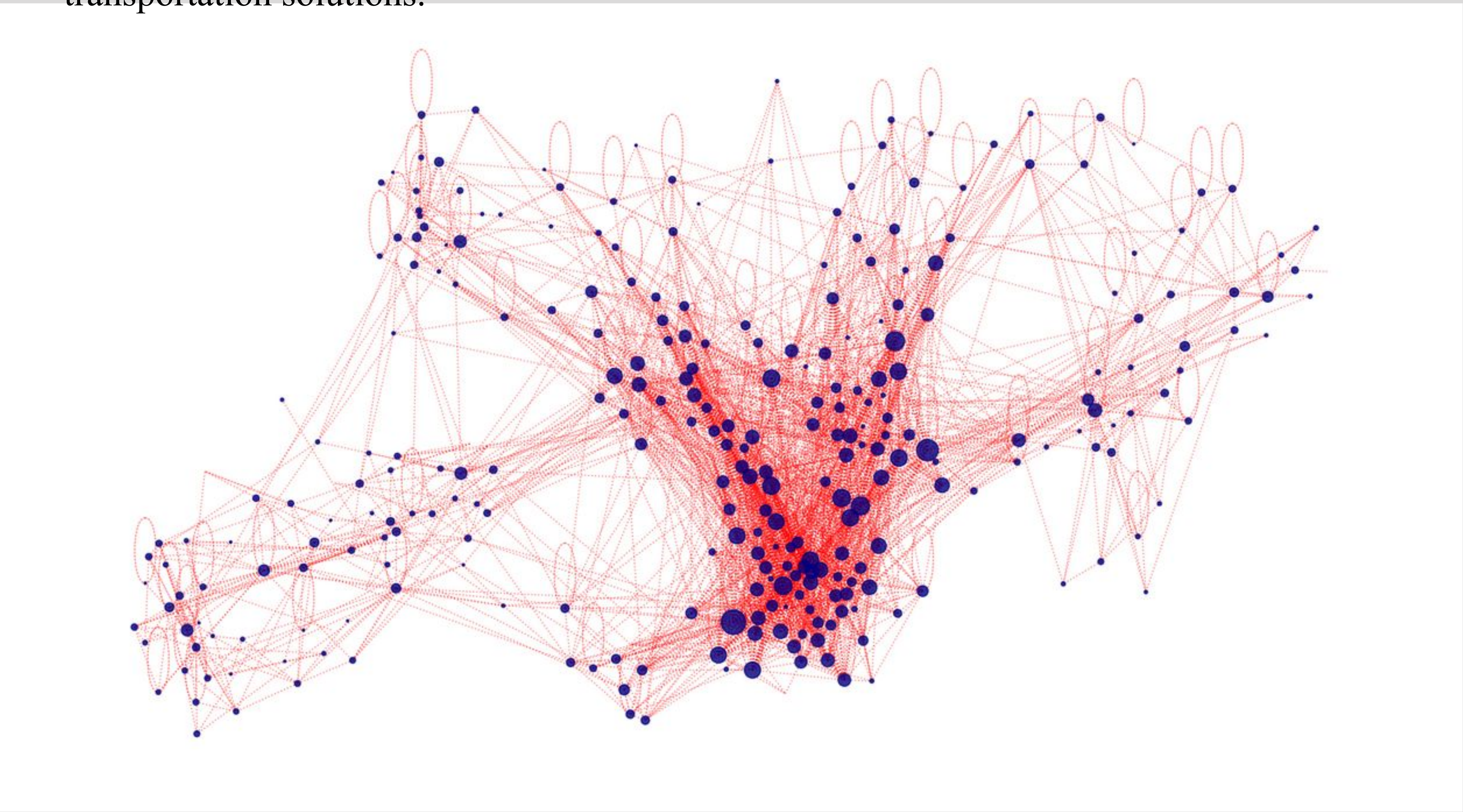
- Various algorithms exist for community detection, such as modularity optimization.
- Recent advancements include deep learning-based approaches like graph neural networks (GNNs) for community detection.
- Community detection algorithms have been applied across diverse domains, offering valuable insights into network structures and dynamics.

Limitations of Existing Approaches

- Many algorithms struggle with efficiently representing and processing sparse network data.
- Some algorithms face challenges in scalability when dealing with large-scale networks, resulting in increased computational costs.
- Existing approaches often lack accuracy in detecting overlapping communities, affecting the quality of results.
- Algorithms may be sensitive to noise and outliers in data, leading to less robust community detection outcomes and also lack of adaptability to dynamic network structures, necessitating frequent re-computation of community assignments.

Idea

- The project aims to analyze the Helsinki City Bike network data comprehensively. It seeks to understand bike usage patterns, optimize station locations, and improve overall network efficiency through data science techniques. The goal is to contribute to better urban mobility planning and sustainable transportation solutions.



Improvemet Suggestions

1. **Data Structure Enhancement:**
   - Implement Graph Data Structure: Utilize a graph data structure to represent the bike network, with stations as nodes and connections between stations as edges. This structured representation enables efficient traversal and manipulation of the network for algorithmic analysis.

2. **Algorithmic Refinement:**
   - Optimize Girvan Newman Algorithm:
     - Enhance Edge Removal Strategy: Modify the edge removal strategy in the Girvan Newman Algorithm to prioritize edges based on additional criteria such as edge weight or node degree. This refinement can lead to more meaningful community detection within the bike network.
     - Parallelize Computation: Implement parallel computing techniques to accelerate the computation of betweenness centrality and edge removal steps in the Girvan Newman Algorithm, particularly for large-scale bike networks.
   - Fine-Tune Louvain Algorithm:
     - Refine Modularity Calculation: Fine-tune the modularity calculation in the Louvain Algorithm by incorporating edge weights or considering higher-order network structures to capture more nuanced community patterns.
     - Incorporate Multi-Level Optimization: Implement multi-level optimization strategies within the Louvain Algorithm to improve its scalability and efficiency, especially for networks with millions of nodes and edges.

Applied Algorithms

Github Link- https://github.com/DSA-IITJ-2024/ideathon-code-submission-Avio3570

Used algorithms: 1. Grivan Newman Algorithm (Hierarchical Divisive Method)

2. Louvain Algorithm (Modularity Optimization)

1. Grivan Newman Algorithm (Hierarchical Divisive Method):

- How it works:
  - The Girvan Newman algorithm aims to detect communities or clusters within a network by iteratively removing edges with the highest betweenness centrality.
  - Betweenness centrality measures how often a node lies on the shortest path between other nodes in the network.
  - By iteratively removing edges with high betweenness centrality, the algorithm breaks the network into communities or clusters.
  - This process continues until the network is divided into the desired number of communities.

Betweenness Centrality $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$

σ_{st} is the total number of shortest paths from node s to node t.

$\sigma_{st}(v)$ is the number of those paths that pass through v.

Applied Algorithms

2. Louvain Algorithm (Modularity Optimization):

- How it works:
  - The Louvain algorithm is a modularity-based method for community detection in networks.
  - It optimizes a quality function known as modularity, which measures the density of connections within communities compared to connections between communities.
  - The algorithm iteratively optimizes modularity by moving nodes between communities to maximize the overall modularity score.
  - It continues this process until no further improvement in modularity is possible.

Modularity Calculation:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

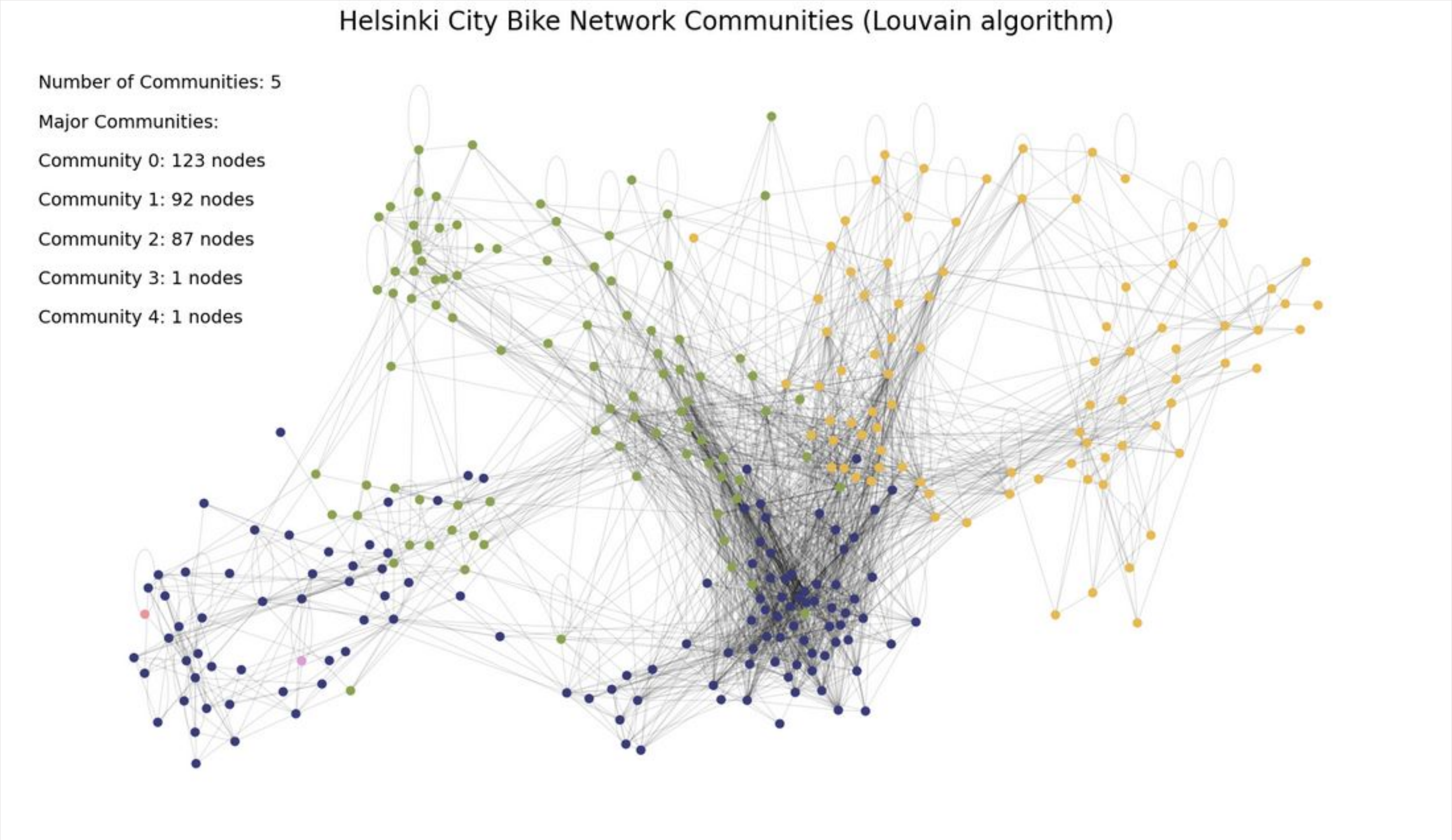
A_{ij} is the adjacency matrix element that represents the connection between nodes i and j.

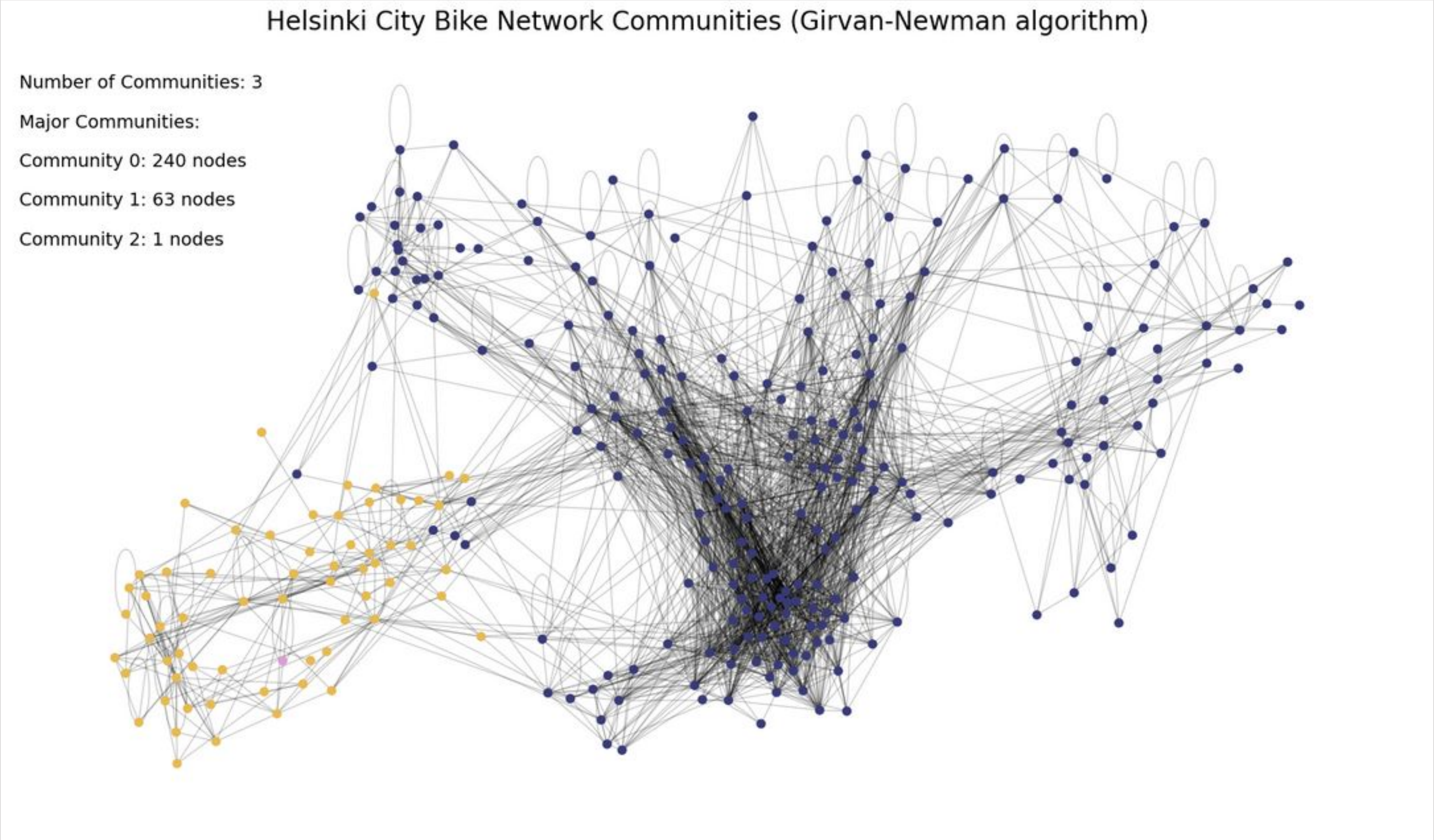
k_i and k_j are the degrees of nodes i and j, respectively.

m is the total number of edges in the network.

$\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community, and 0 otherwise.

Results:





Time complexity:

1. Grivan Newman Algorithm: O(N^2)
2. Louvain Algorithm: O(N)

Results:

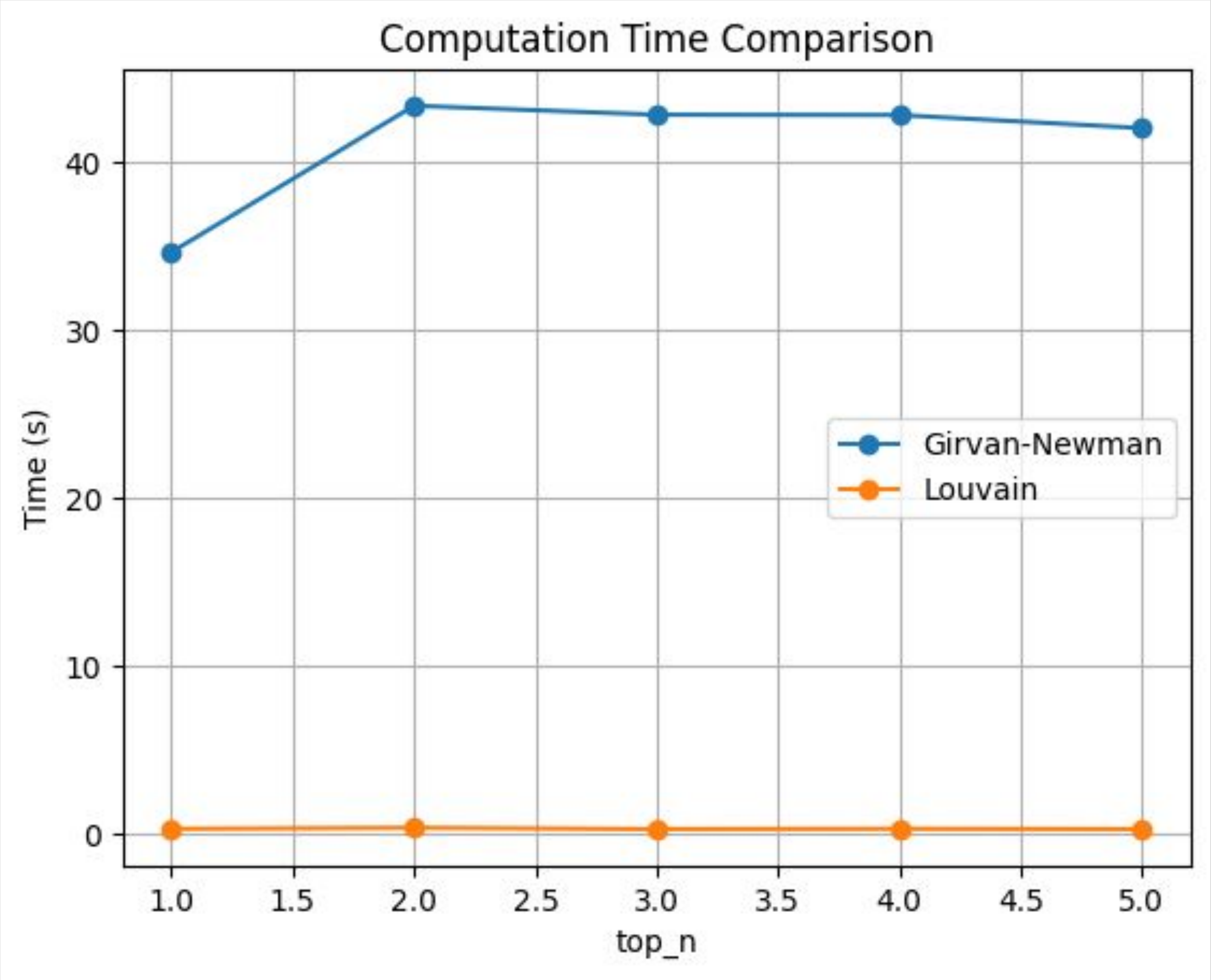
1. GRIVAN NEWMAN ALGORITHM:

```
Time for computing the communities with top_n=1: 22.80 s
Time for computing the communities with top_n=2: 23.69 s
Time for computing the communities with top_n=3: 22.51 s
Time for computing the communities with top_n=4: 22.47 s
Time for computing the communities with top_n=5: 22.43 s

Total time for computing the communities: 113.90 s
```

2. LOUVAIN ALGORITHM:

```
Time for computing the communities with top_n=1: 0.27 s
Time for computing the communities with top_n=2: 0.31 s
Time for computing the communities with top_n=3: 0.26 s
Time for computing the communities with top_n=4: 0.24 s
Time for computing the communities with top_n=5: 0.17 s

Total time for computing the communities: 1.25 s
```

Comparison:



Conclusion:

- **Key Findings/Learnings:**
  - Gained insights into usage patterns and dynamics of the Helsinki City Bike network.
  - Successfully identified distinct communities within the bike network using algorithms.
  - Highlighted optimization opportunities for route planning, station placement, and resource allocation.
  - Uncovered factors influencing bike usage, aiding in the design of targeted interventions.
- **Scope of Future Extension:**
  - Implement real-time data integration and analysis for dynamic operational adjustments.
  - Develop predictive models for demand forecasting to enhance resource allocation.
  - Integrate bike network analysis with urban planning efforts for informed decisions.
  - Continuously monitor user feedback and behavior to inform user-centric solutions.
- **Innovation Summary:**
  - Our innovation lies in leveraging data science techniques to analyze the Helsinki City Bike network comprehensively. By applying advanced algorithms for community detection and optimization, we uncovered actionable insights to improve the efficiency, accessibility, and sustainability of urban bike transportation. Our approach not only enhances the current bike network but also sets the stage for future advancements in urban mobility planning and sustainable transportation solutions.