Outline Team 6
Can Network Analysis Pedal Hannover Towards Better Bike Sharing?

Akshita Sharma, Aleksanteri Kylmäaho, Emil Schallwig,
Jannis Wiederhöft, Yannick Oefler

University of Mannheim, 68131 Mannheim, Germany

Keywords: Directed Weighted Graphs, Degree Centrality, Betweenness Centrality, Closeness Centrality, Rank Prestige, K-Core Analysis, Louvain Methode, Girvan-Newman Method.

1 What is the problem you are solving?

Public transportation is becoming increasingly complex as more and more people concentrate in cities, making it progressively difficult to travel from point A to point B. At the same time, cities are confronted with the challenge of becoming more climate-friendly. Bike rental systems are a rapidly growing solution to this problem. Users can rent a bike at designated stations for a rental fee and return it to any station after their ride, all with just a few clicks on their phone. This provides a convenient, environmentally friendly, and flexible way to travel. To ensure these systems operate smoothly, careful, data-driven planning is essential. In our project, we focus on the bike rental stations in Hannover, called *Bike Citizens*. Specifically, we aim to address the following challenges:

Q1: Where is the demand for bicycles highest? By identifying the most frequently used stations, we can pinpoint potential shortages. In such cases, the company should consider manually reallocating bikes to address the demand.

Q2: Where should the bike storage be located? It is in the company's best interest to position storage facilities centrally within the network, allowing for quick and efficient bike redistribution whenever demand arises.

Q3: How should the company divide the bike network into groups of stations to be serviced by different maintenance teams? This question focuses on structuring the network so that maintenance tasks can be managed efficiently and cost-effectively, with specific teams assigned to designated groups of stations.

Q4: Where should the company place bike repair kits? Strategically positioning do-it-yourself repair kits at key locations can not only reduce maintenance efforts but also enhance user comfort and satisfaction.

Q5: Where should the city build new bike lanes? The decision of the population to switch from car or tram to bike depends on various factors, including the time it takes to travel from A to B. Therefore, we aim to identify popular routes that currently have low average speeds. We suggest enhancing bike lanes along these routes.

2 What data will you use?

We will use data from the Geospatial Data Management of the Hannover Region. The GML files provide information on intensity, and average speeds of bike rides tracked by the Bike Citizens app, in aggregated form based on the street network graph for the Hannover region for the respective year (see title). Each quarter is represented as a separate layer, with a total of four GML files for each quarter from 2020 to 2024.

The GML files contain multiple `ogr:featureMember` entries. Each `featureMember` describes an individual bike ride. Each `gml:LineString` describes a route with a `gml:posList`, which provides the coordinate pairs (latitude and longitude) for the course of the bike ride. Each record has a unique `ID` that identifies the bike or station. The `YEAR` and `MONTH` attributes indicate the year and month of the ride. `TRACKS_FWD` and `TRACKS_BAC` refer to the number of rides in the forward and backward directions along the route. `SPEED_REL` is an attribute that provides the relative speed of the bike ride, representing the efficiency of the trip.

3 How to solve the problem?

Our approach involves representing the quarterly Bike Citizens WFS information as weighted directed graphs. This method allows us to effectively apply network analysis techniques to understand the detailed bicycle usage patterns revealed by the trip counts, segment coordinates, and speed data within the GML files.

Specifically, we'll construct these graphs where Nodes represent the unique start- and end-points derived from the segment coordinates `gml:posList`. Each road segment will be represented by two directed edges connecting its endpoint nodes, reflecting the two possible directions of travel. The weight of each edge directly corresponds to the usage frequency, using the `TRACKS_FWD` and `TRACKS_BAC` counts provided. Additionally, the `SPEED_REL` value will be attached as an edge attribute, which is crucial for addressing *Q5* regarding potential bike lane improvements.

3.1 What preprocessing steps will be required?

To prepare this data for meaningful analysis, we'll follow several preprocessing steps. First, we'll load the GML data for each of the four quarters of 2024 using a graph processing library like `NetworkX`.

Next, we need to decide how to handle the time dimension. We can either create separate graphs for each quarter (Option A) to study seasonal variations or combine all data into a single, annual graph (Option B) for a clearer view of overall network structure and yearly demand. While Option A offers seasonal insights, we plan to initially focus on Option B. Creating this aggregated annual graph involves identifying all unique nodes (junctions) and directed edges present across all four quarters. We will then sum the `TRACKS_FWD`/`TRACKS_BAC` counts for each edge over the year to get the total annual weight, reflecting overall usage. The `SPEED_REL` attribute will be aggregated by calculating a usage-weighted average for each edge. This approach implicitly captures dynamic changes, such as potential station relocations, through the resulting shifts in traffic flow observed at the affected junctions over the year. Integrating historical data from 2020-2023 remains a possibility for extension.

Although not a major focus, a certain amount of data cleansing is still necessary. For instance, we ensure that slightly differing geographical coordinates are treated as a single node by applying a suitable tolerance threshold. Additionally, we identify and review nodes that are only connected to road segments with zero reported trips in both forward and backward directions. These may point to parts of the network where no bicycle activity was observed, potentially indicating underused infrastructure.

3.2 Which algorithms you plan to use?

To address the question *"Q1: Where is the demand for bicycles highest?"*, we propose analyzing the in-degree and out-degree of each node. We hypothesize that if a node's out-degree significantly exceeds its in-degree, it indicates a shortage of bikes at that location. In such cases, the company should consider manually reallocating bikes to meet the demand.

In addition, we plan to compute the prestige rank of each node beforehand. The idea behind using prestige is to identify key locations such as train stations or universities, where ensuring bike availability is particularly critical. These high-prestige stations are typically destinations that are frequently reached from other prestigious stations. At such hubs, users are more dependent on finding available bikes—and more likely to be frustrated when none are available—so it is especially important that their in-degree matches the out-degree approximately.

To answer the question *"Q2: Where should the bike storage be placed?"*, we plan to use closeness centrality as a metric. The goal is to place storage facilities in central positions within the network, enabling the company to redistribute bikes quickly and efficiently whenever demand arises.

To address the question *"Q3: How should the company divide the bike network into groups of stations to be serviced by different maintenance teams?"*, we will apply k-core analysis. Stations that belong to high k-cores are more densely connected within the network, making them ideal candidates for bundled maintenance alongside nearby stations. We also consider incorporating more advanced community detection methods such as Louvain or Girvan-Newman to further refine the grouping of stations.

Additionally, we aim to address the question, *"Q4: Where should the company place bike repair kits?"* To answer this, we plan to measure betweenness centrality.

Stations with high betweenness centrality serve as critical hubs in the network, facilitating connections between different parts of the system. These locations are ideal candidates for placing repair kits, as they are likely to be visited by many users.

Finally, we aim to address the question *"Q5: Where should the city build new bike lanes?"*. To answer this, we will analyze the average relative speed (`SPEED_REL`) in conjunction with the most heavily used edges in the network. Our hypothesis is that routes with high traffic but low average speeds may require infrastructure improvements—such as enhanced or extended bike lanes—to better meet demand and improve overall efficiency. We reserve the option to apply more advanced methods and metrics to address all questions, should our further research uncover appropriate techniques.

4 How will you evaluate, measure success?

We recognize that there is no established benchmark in our context, which makes our evaluation challenging. Due to the absence of a clear success metric, we will compare our results as accurately and critically as possible against real-world conditions. Specifically, we will examine the city's infrastructure to assess the feasibility and practicality of our proposals. However, we will also conduct further research to explore the possibility of identifying additional evaluation methods.