Break Through the Network Mist

Team:
Brieg L'Hostis
Qiyun Wu
Xiaoxuan Shang
Zhenning Li

Content

1. Introduction

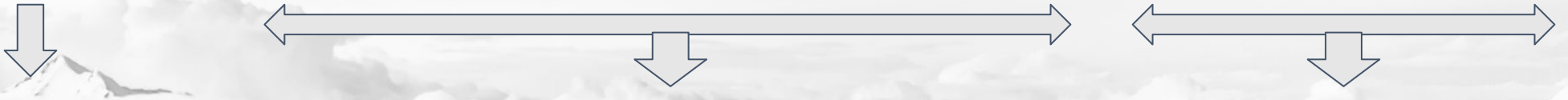
Motivation

Although bike sharing is become more and more popular around the world, making profit is still difficult for those companies because of some inappropriate bike distributions and riders' deliberate saving behaviours like switching bikes under some limits. Citi Bike, the largest bike-sharing system in New York was analysed instead to give some insights about sharing bike systems in metropolises. After studying networking behaviours between different stations in different time, companies can make optimization of bike supply and distribution and adapt pricing according to time and routes.

Data

We used data of August 2021, December 2021 and January 2022 published on Citi Bike website to analysed networks considering seasonality effect.

	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	distance	diff_time
0	2022-01-06 18:58:23	2022-01-06 19:02:31	Madison St & 1 St	HB402	6 St & Grand St	HB302	40.738790	-74.039300	40.744398	-74.034501	743.404166	4.133333
1	2022-01-05 13:14:34	2022-01-05 13:23:33	7 St & Monroe St	HB304	11 St & Washington St	HB502	40.746413	-74.037977	40.749985	-74.027150	995.098561	8.983333
2	2022-01-13 16:07:54	2022-01-13 16:10:20	7 St & Monroe St	HB304	Mama Johnson Field - 4 St & Jackson St	HB404	40.746413	-74.037977	40.743140	-74.040041	403.440901	2.433333
3	2022-01-06 15:45:41	2022-01-06 15:52:04	JC Medical Center	JC011	Lafayette Park	JC078	40.716540	-74.049638	40.713464	-74.062859	1165.934400	6.383333
4	2022-01-16 15:36:30	2022-01-16 15:53:24	Morris Canal	JC072	Washington St	JC098	40.712419	-74.038526	40.724294	-74.035483	1345.564908	16.900000



start and end times

start and end station names, ids

start and end station coordinates

Graph and Feature Definition

- For each **node N:**
  - **id**: the code name of the station represented by the node;
  - **traffic**: the number of trips which started or ended at the node's station.

- Each **edge E = ($N1$, $N2$)** will have the following features:
  - **distance**: the distance between the two stations at the extremities of the edge;
  - **duration**: the average duration in minutes of a trip between the two stations at the extremities of the edge;
  - **edge_traffic**: the number of trips between the two stations at the extremities of the edge recorded in the data.

2. Related Works

Related Works

- *Data-driven Bicycle Network Analysis Based on Traditional Counting Methods and GPS Traces from Smartphone*

- **Highlights**: They analyzed transport demand of cyclists to quantify the road network weaknesses of a city aiming to identify infrastructure improvements

- **Similarity**: Both aim at city efficiency using cyclists routes data, and both used similar features (betweenness centrality etc.) to measure the extent to which a bridging role (a route) in a network

- **Difference**:
    - We considers more about the distance and nodes clustering instead of different route choices for cyclists
    - They used GPS data which gives more nodes information

Related Works (Cont'd)

- *Green travel mobility of dockless bike-sharing based on trip data in big cities: A spatial network analysis*

- **Highlights**: They constructed several networks according to different time periods in a weekday, and used Betweenness, PageRank, Network efficiency etc. to build up node features for the regression model to detect the travel mobility of dockless bike sharing.

- **Similarity**: Both used similar ways to build up features and models

- **Difference**:
  1. We considers more on the changes of biking system over time (16 months) instead of only focusing on one week
  2. Our starting point is to find out the distribution changes over time to maximize company profit, they aim to forecast the demand in a day and in a week.

3. Methodology

Methodology

1

Data loading & preprocessing

- Aggregate data needed (for a specific month or a whole year)
- Time difference & distance

Methodology

1

Data loading & preprocessing

- Aggregate data needed (for a specific month or a whole year)
- Compute the node and edge features

2

Preliminary network analysis

- Compute simple network statistics
- Visualize the bike network on the map of Jersey City

Representation of the network on the map of Jersey City for the data of January 2022

Methodology

1

Data loading & preprocessing

- Aggregate data needed (for a specific month or a whole year)
- Compute the node and edge features

2

Preliminary network analysis

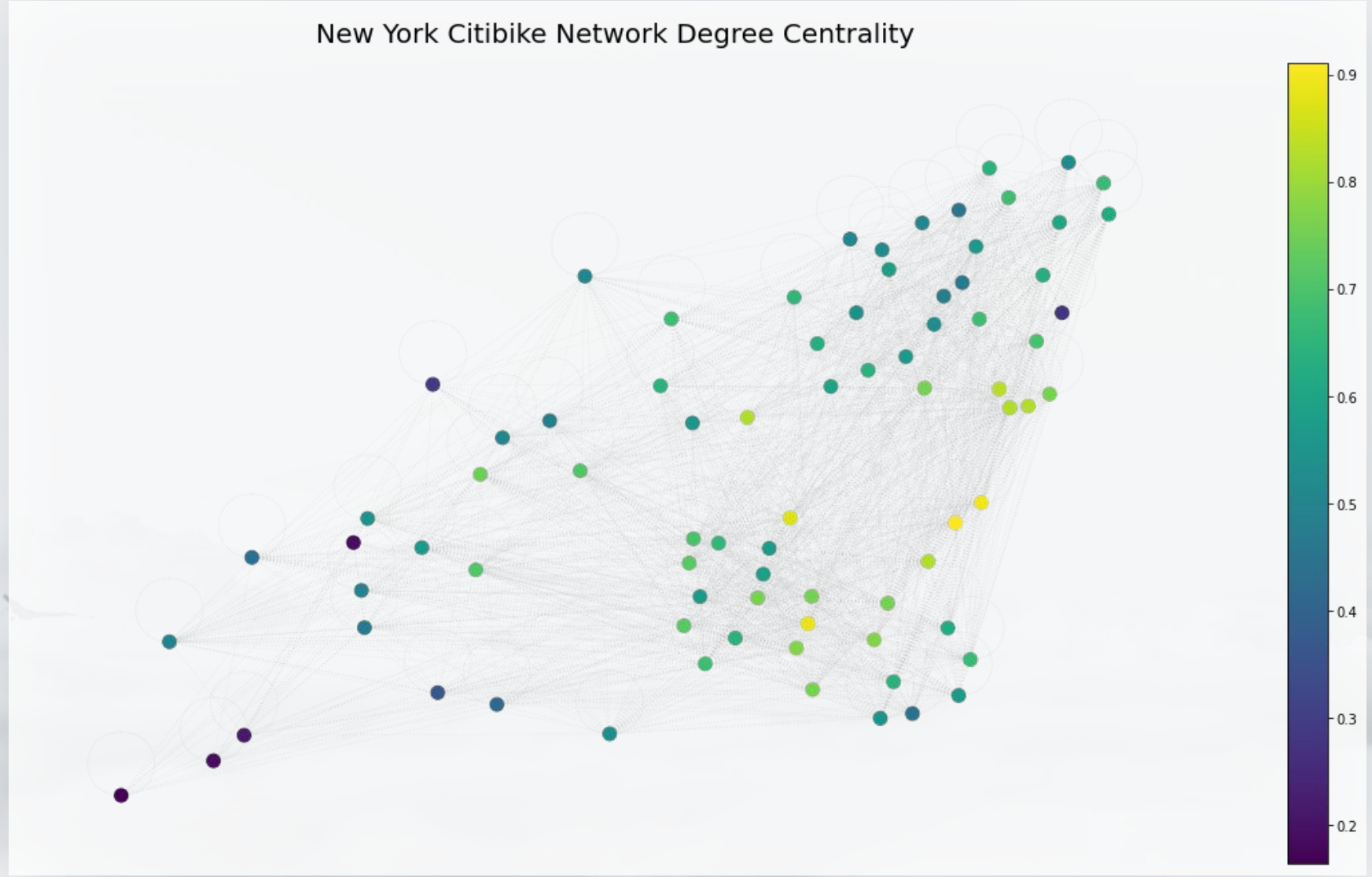
- Compute simple network statistics
- Visualize the bike network on the map of Jersey City

3

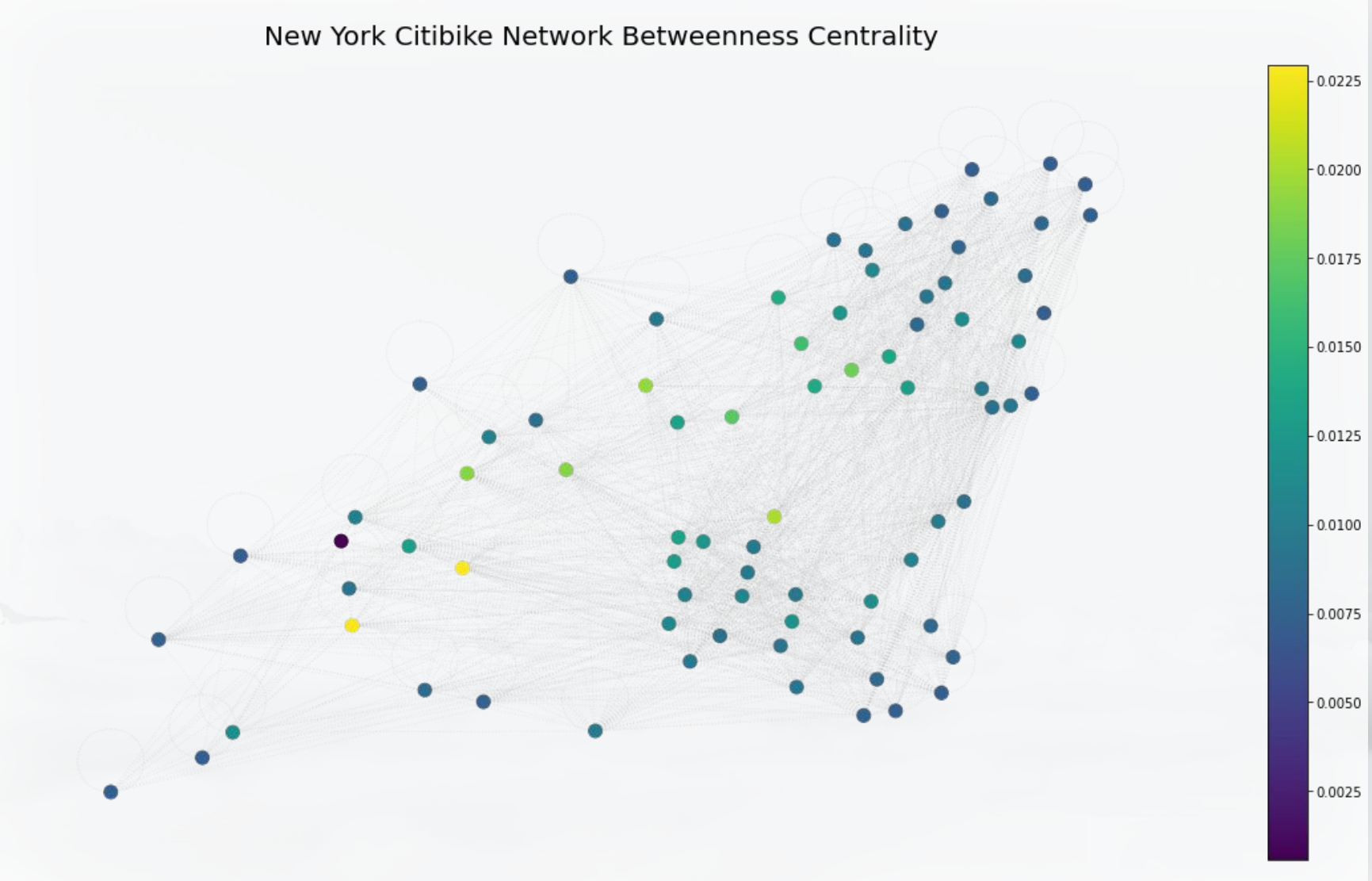
Network centrality measures

- Apply different measures to describe the importance of the nodes
- Display them on the graph to compare the methods

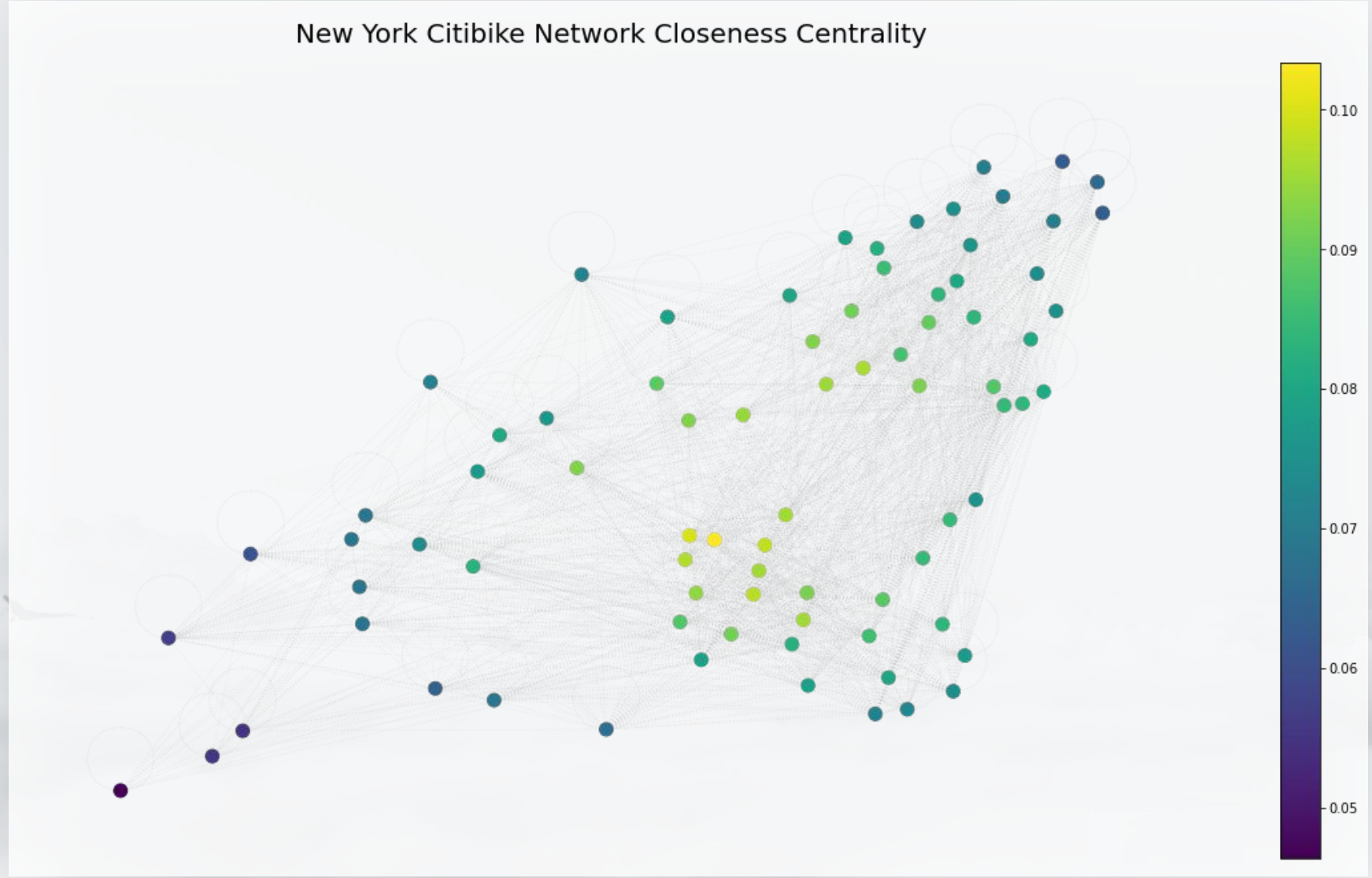
Degree Centrality



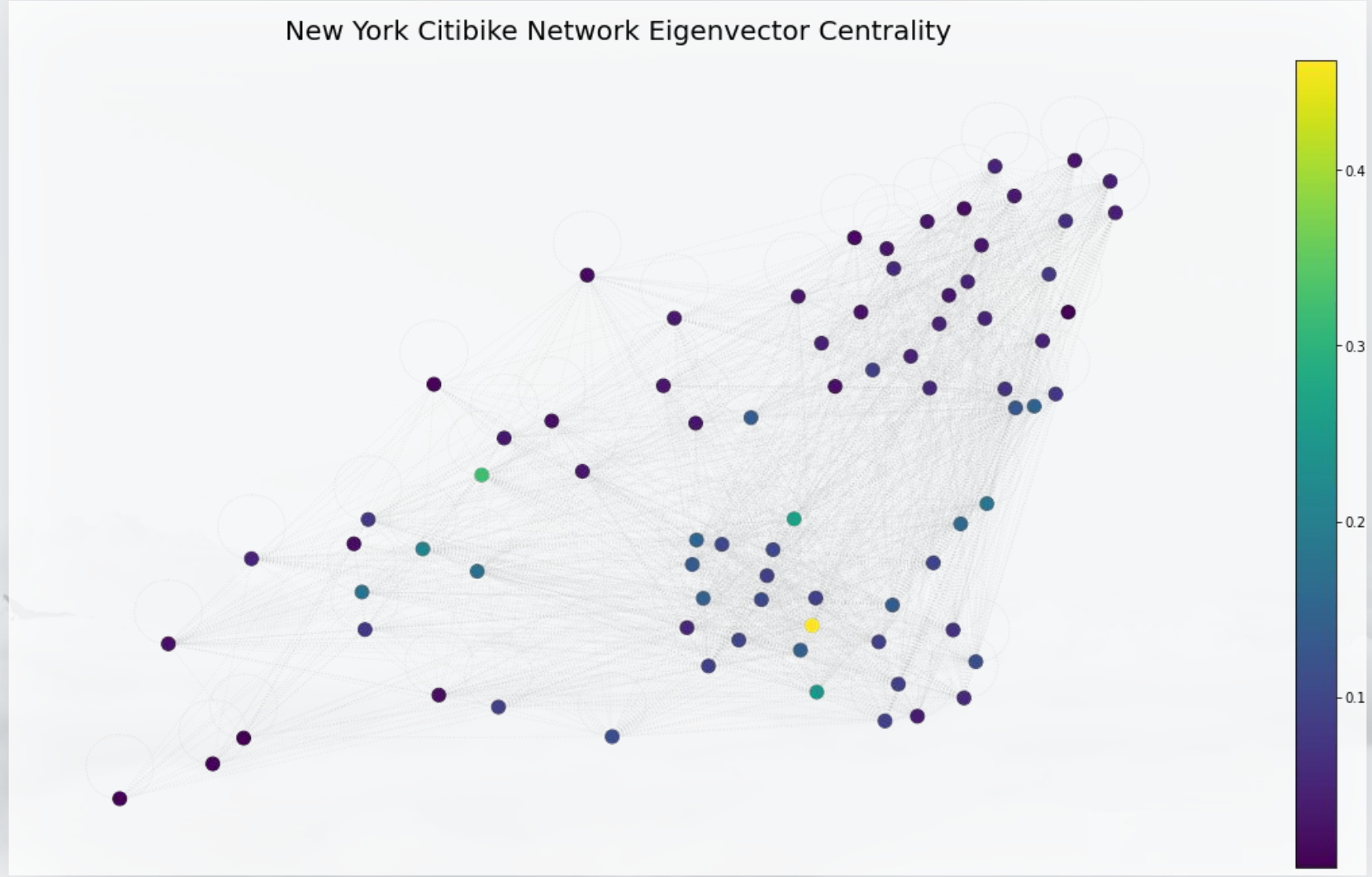
Betweenness Centrality



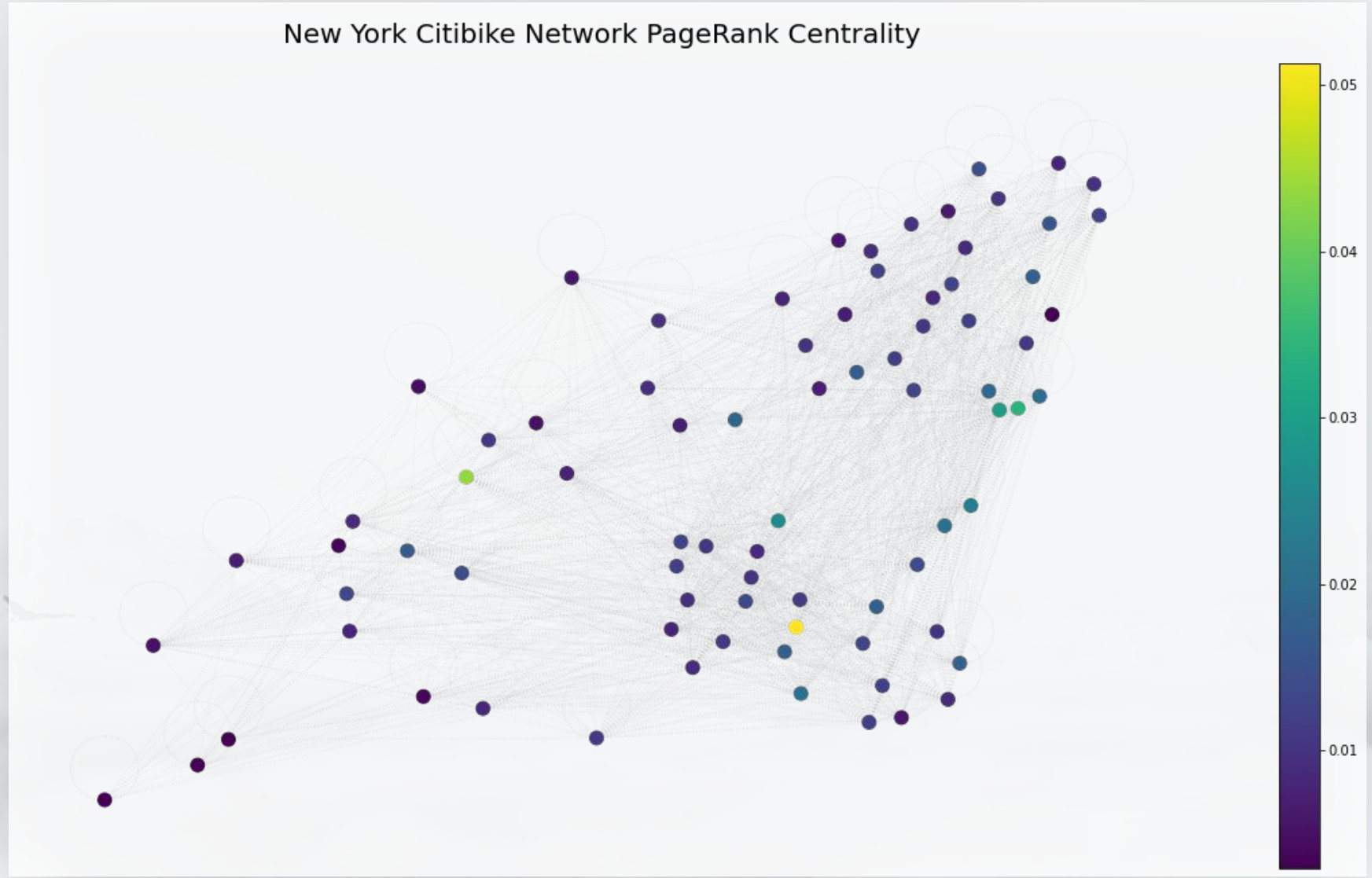
Closeness Centrality



Eigenvector Centrality



PageRank


New York Citibike Network PageRank Centrality
0.05
0.04
0.03
0.02
0.01

Methodology

1

Data loading & preprocessing

- Aggregate data needed (for a specific month or a whole year)
- Compute the node and edge features

2

Preliminary network analysis

- Compute simple network statistics
- Visualize the bike network on the map of Jersey City

3

Network centrality measures

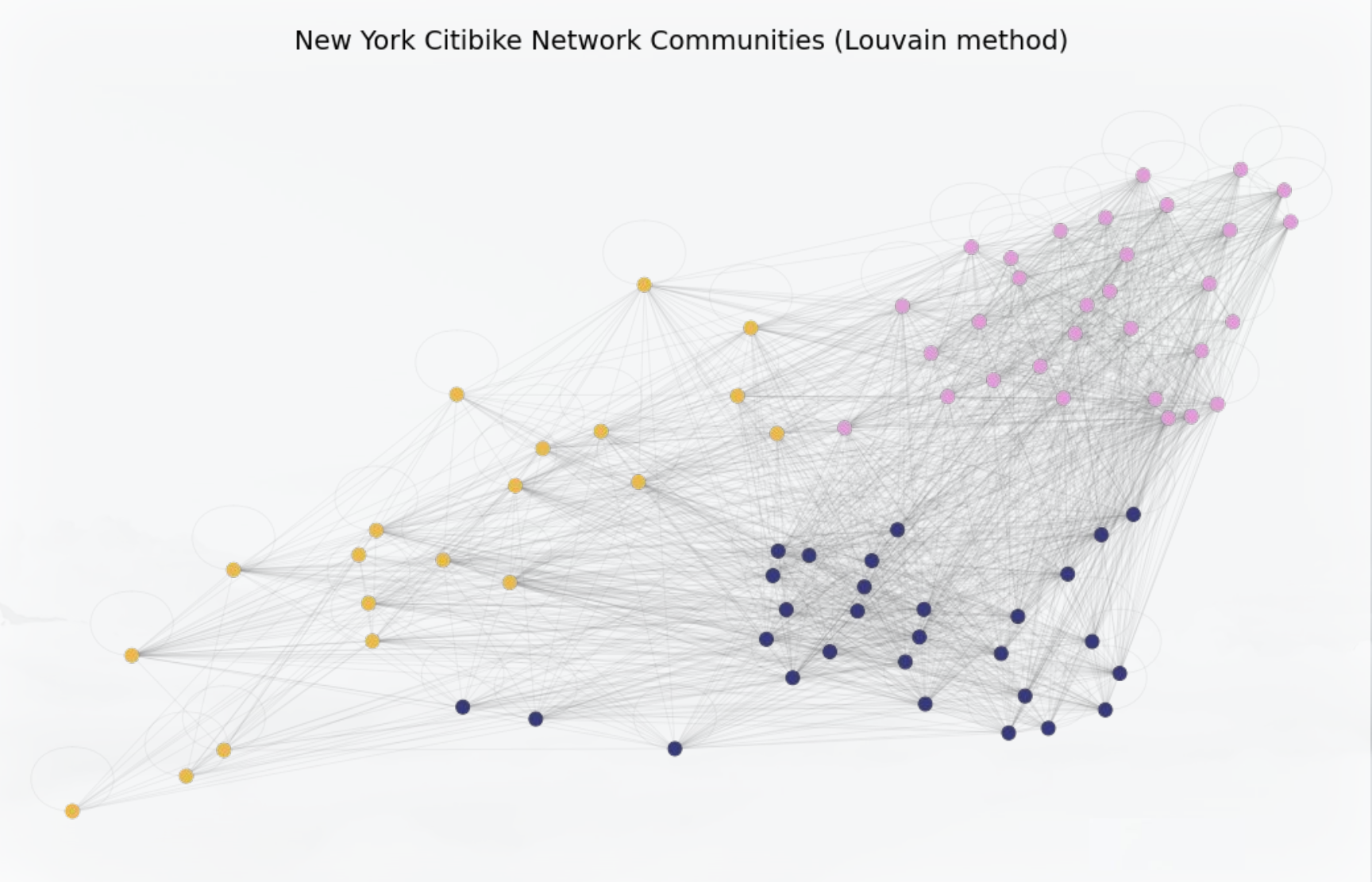
- Apply different measures to describe the importance of the nodes
- Display them on the graph to compare the methods

4

Community detection

- Implement different algorithms to find communities in the nodes
- Combo method; Louvain method; Fluid communities method

Community Detection - Louvain Method



Methodology

1

Data loading & preprocessing

- Aggregate data needed (for a specific month or a whole year)
- Compute the node and edge features

2

Preliminary network analysis

- Compute simple network statistics
- Visualize the bike network on the map of Jersey City

3

Network centrality measures

- Apply different measures to describe the importance of the nodes
- Display them on the graph to compare the methods

4

Community detection

- Implement different algorithms to find communities in the nodes
- Combo method; Louvain method; Fluid communities method

5

Time analysis

- Launch an interactive dashboard to display yearly, monthly and daily data for comparisons

Thank you!