Time Series Data Clustering of Minnesota Bike Sharing System and Operation Strategy

Yufeng Zhang

zhan4879@umn.edu

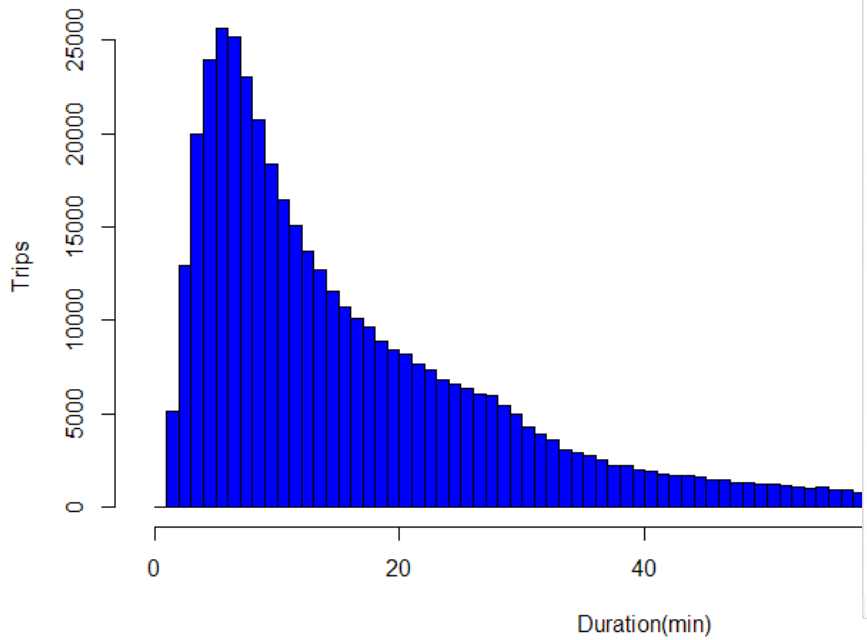
Department of Civil, Environmental, and Geo-Engineering

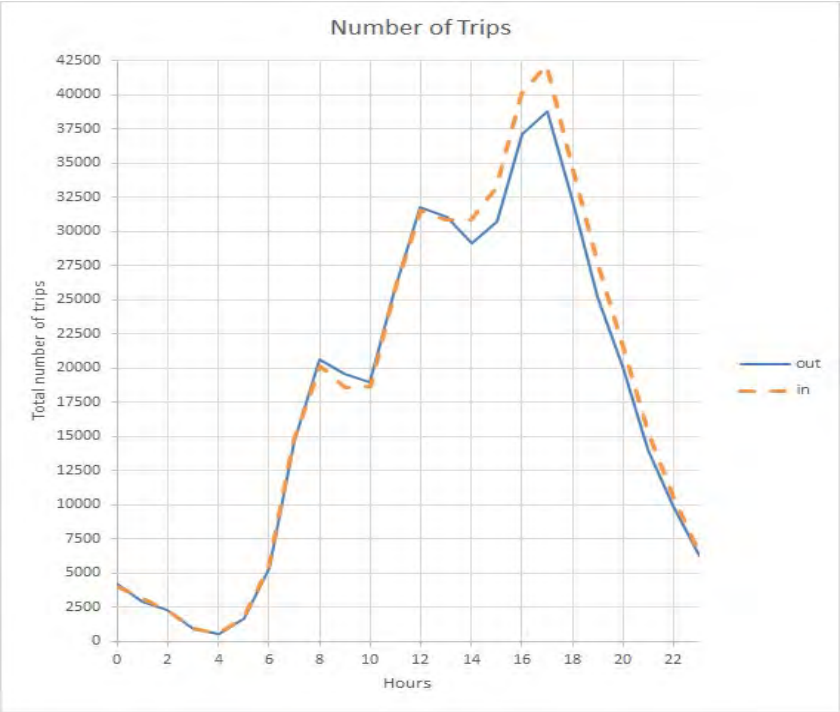
Outline

1. Introduction
2. Bike Station Clustering
3. Rebalancing Vehicle Routing Strategy
4. Conclusions

Introduction

- **Bike Sharing System in MN: Nice Ride**
- **202 (190) stations**
- **430,210 trips in 2016**





Goal and Methods

- **Goals**

1) To identify usage patterns in BSS

2) To find a routing strategy for rebalancing vehicle

- **Methods**

1) Bike station clustering using Poisson Mixture Model

2) Vehicle routing optimization problem for finding the routing strategy

Station Clustering

- **Input: time-series data**

X_{sdh}^{out}: departure count for station $s \in \{1, \cdots, S\}$ during day $d \in \{1, \cdots D\}$ and at hour $h \in \{1, \cdots, 24\}$

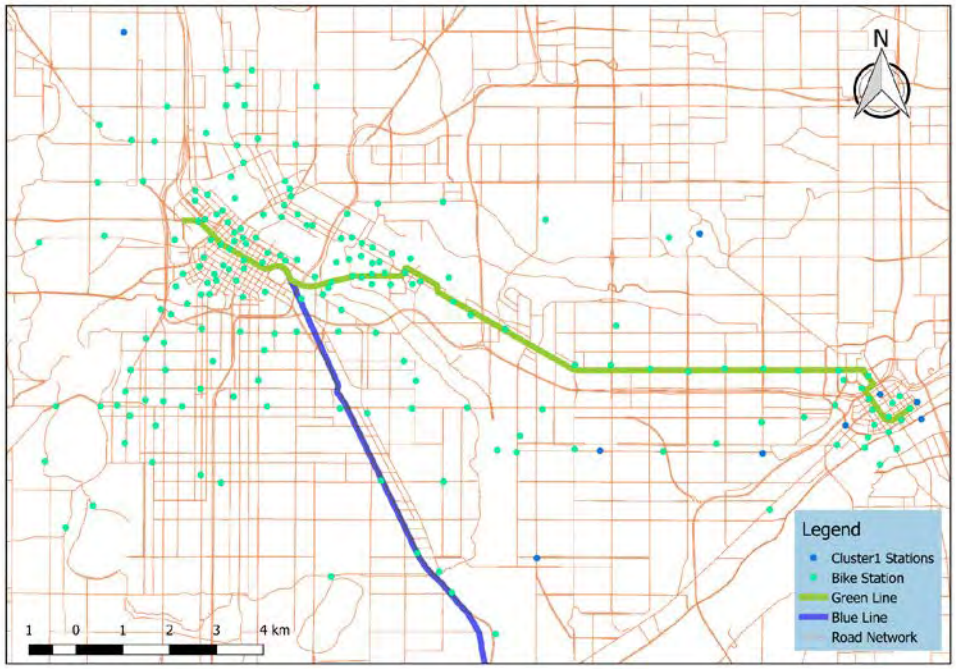
X_{sdh}^{in}: arrival count for station $s \in \{1, \cdots, S\}$ during day $d \in \{1, \cdots D\}$ and at hour $h \in \{1, \cdots, 24\}$

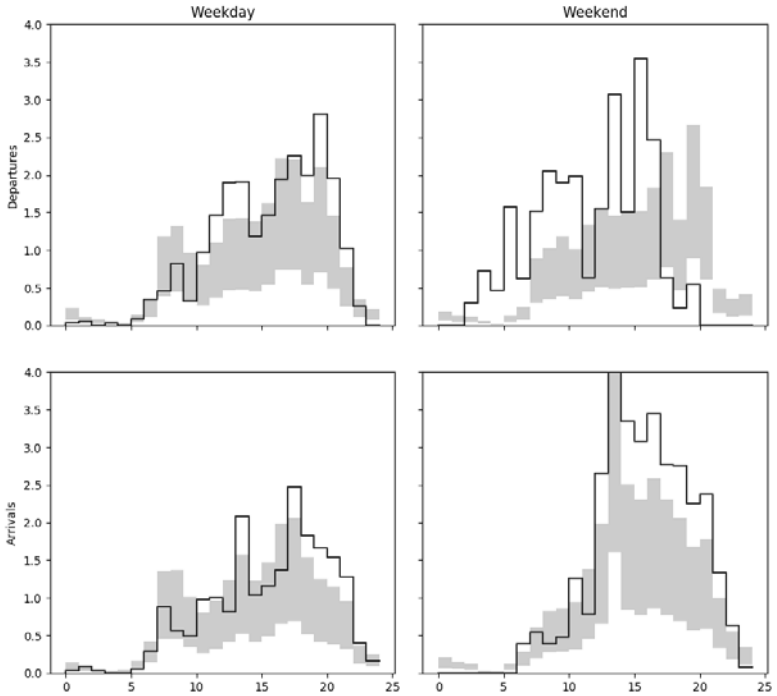
A vector $\boldsymbol{X}_{sd} = \left(\mathrm{X}_{sd1}^{in}, \cdots, \mathrm{X}_{sd24}^{in}, \mathrm{X}_{sd1}^{out}, \ldots, \mathrm{X}_{sd24}^{out}\right)$ describes the arrival and departure activity of station s during day d

- **Assumption:** $X_{sdt} \mid \{Z_{sk} = 1, W_{dl} = 1\} \sim Poisson(\alpha_s \lambda_{klt})$
- **Output: Pick up and drop off rates on both weekend and weekday, λ_{klt}**
- **EM algorithm**

Clustering Results (1)

- **Weekend Destination**

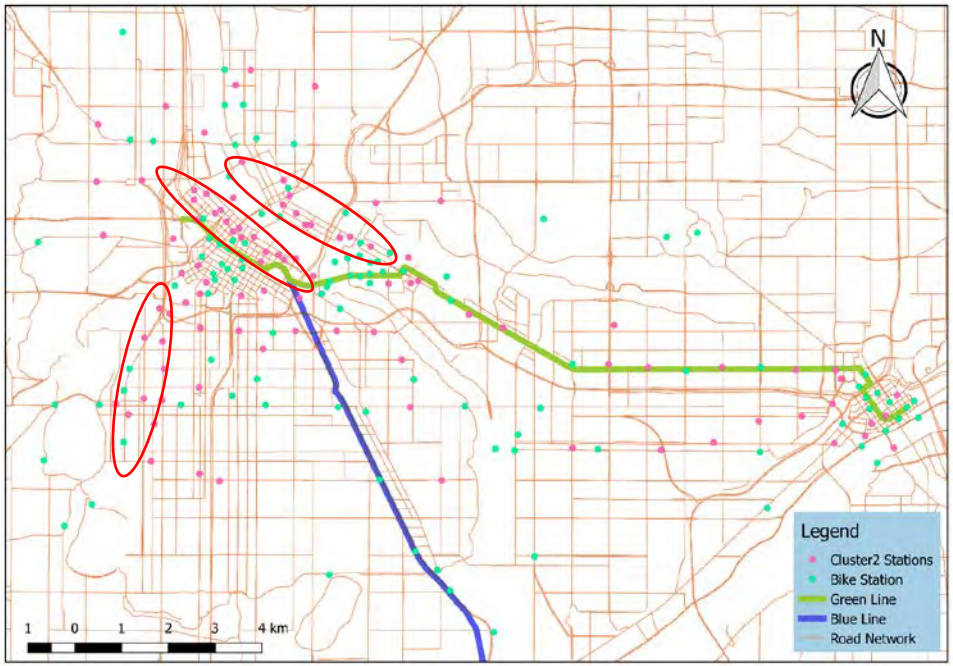


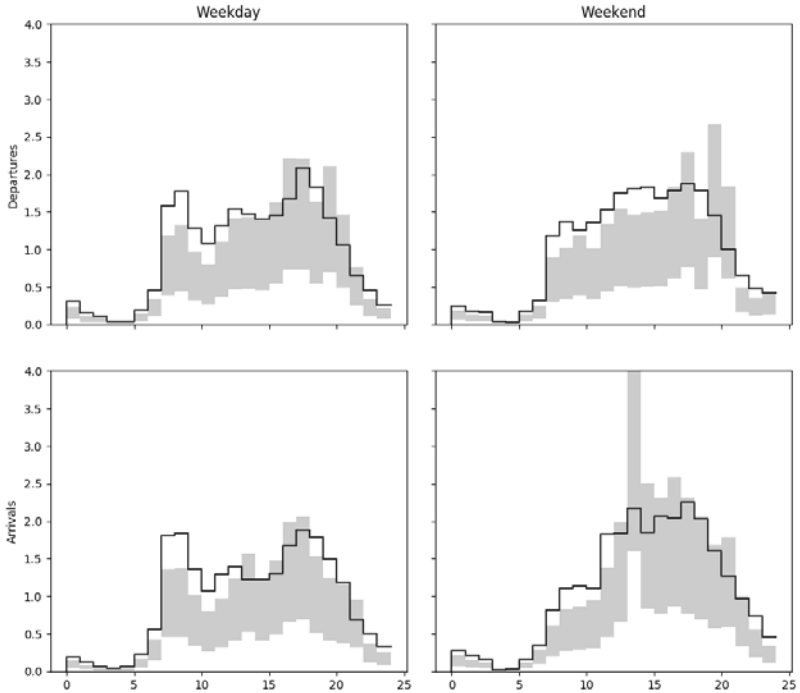


10 stations

Clustering Results (2)

- **Bimodal**

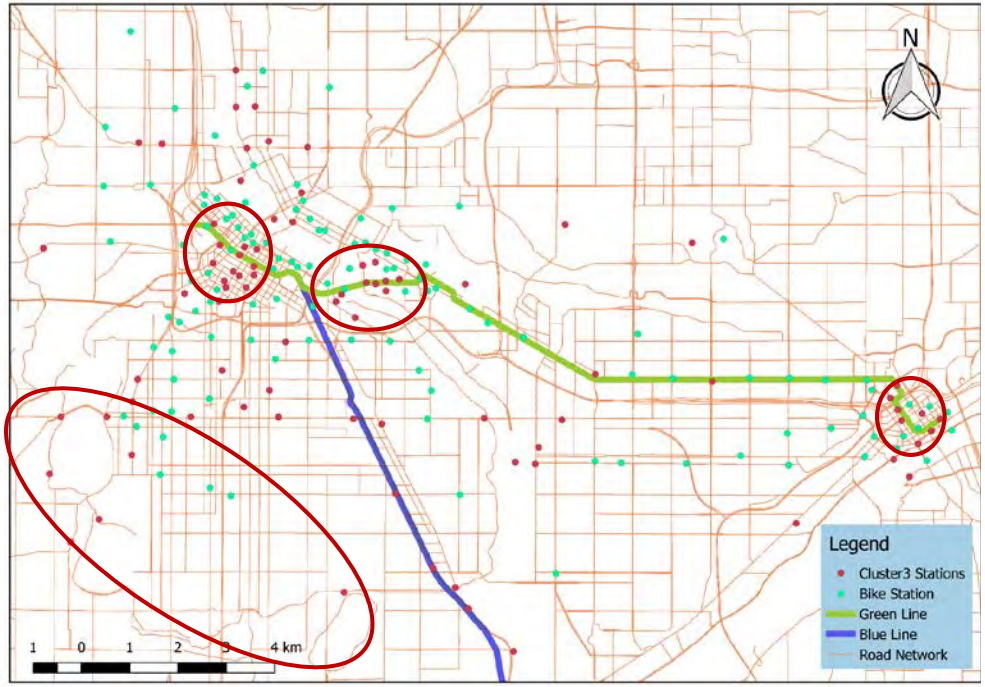


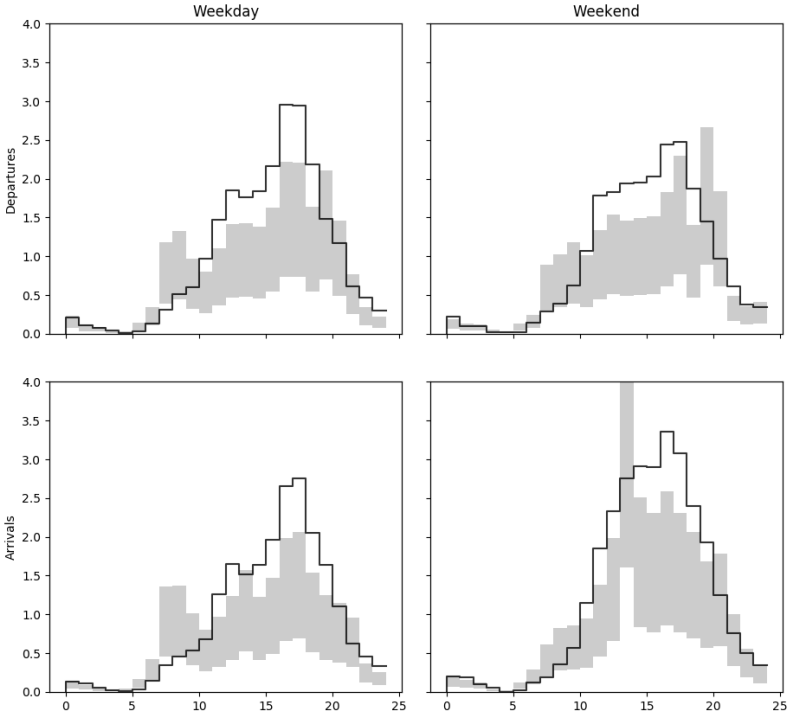


97 stations

Clustering Results (3)

- **Single Peak**





83 stations

Rebalancing vehicle routing strategy

- **Mathematical formulation**

$$\text{Min} \quad \sum_v \sum_{ij} c_{ij} x_{ij}^v - \alpha \sum_s DS_s \tag{1}$$

$$\text{s.t.} \quad \sum_j x_{ij}^v - \sum_j x_{ij}^v = \begin{cases} -1 & i = R \\ 0 & i \in \mathcal{S} \setminus \{R, S\} \\ 1 & i = S \end{cases}, \qquad \forall v \in \mathcal{V} \tag{2}$$

$$y_{ij}^v \leq CV^v x_{ij}^v \qquad \forall (ij) \in \mathcal{A}, \forall v \in \mathcal{V} \tag{3}$$

$$L_i^v \leq \min\{\sum_{j \in \mathcal{S}} y_{ji}^v, \quad \bar{n}\} \qquad \forall i \in \mathcal{S}, \forall v \in \mathcal{V} \tag{4}$$

$$\sum_{v \in \mathcal{S}} L_i^v \leq CS_i - I_i - D_i \qquad \forall i \in \mathcal{S} \tag{5}$$

$$U_i^v \leq \min\{CV^v - \sum_{j \in \mathcal{S}} y_{ij}^v, \quad \bar{n}\} \qquad \forall i \in \mathcal{S}, \forall v \in \mathcal{V} \tag{6}$$

$$\sum_{v \in \mathcal{V}} U_i^v \leq I_i + D_i \qquad \forall i \in \mathcal{S} \tag{7}$$

$$L_i^v - U_i^v = \sum_{j \in \mathcal{S}} y_{ji}^v - \sum_{j \in \mathcal{S}} y_{ij}^v \qquad \forall i \in \mathcal{S}, \forall v \in \mathcal{V} \tag{8}$$

$$d_{ij}^v \leq T x_{ij}^v \qquad \forall (i,j) \in \mathcal{A}, \forall v \in \mathcal{V} \tag{9}$$

$$\sum_{j \in \mathcal{S}} d_{ij}^v - \sum_{j \in \mathcal{S}} d_{ji}^v = \sum_{j \in \mathcal{S}} x_{ij}^v t_{ij} + (L_i^v + U_i^v) h \qquad \forall i \in \mathcal{S}, \forall v \in \mathcal{V} \tag{10}$$

$$DS_s \leq \min\{0, I_s + D_s\} \qquad \forall s \in \mathcal{S} \tag{11}$$

$$x_{ij}^v \in \{0, 1\} \qquad \forall (ij) \in \mathcal{A}, \forall v \in \mathcal{V} \tag{12}$$

$$y_{ij}^v, I_i, L_i^v, U_i^v \in \mathbb{Z}_+ \qquad \forall (ij) \in \mathcal{A}, \forall i \in \mathcal{S}, \forall v \in \mathcal{V} \tag{13}$$

$$d_{ij}^v \in \mathbb{R}_+ \qquad \forall (ij) \in \mathcal{A}, \forall v \in \mathcal{V} \tag{14}$$

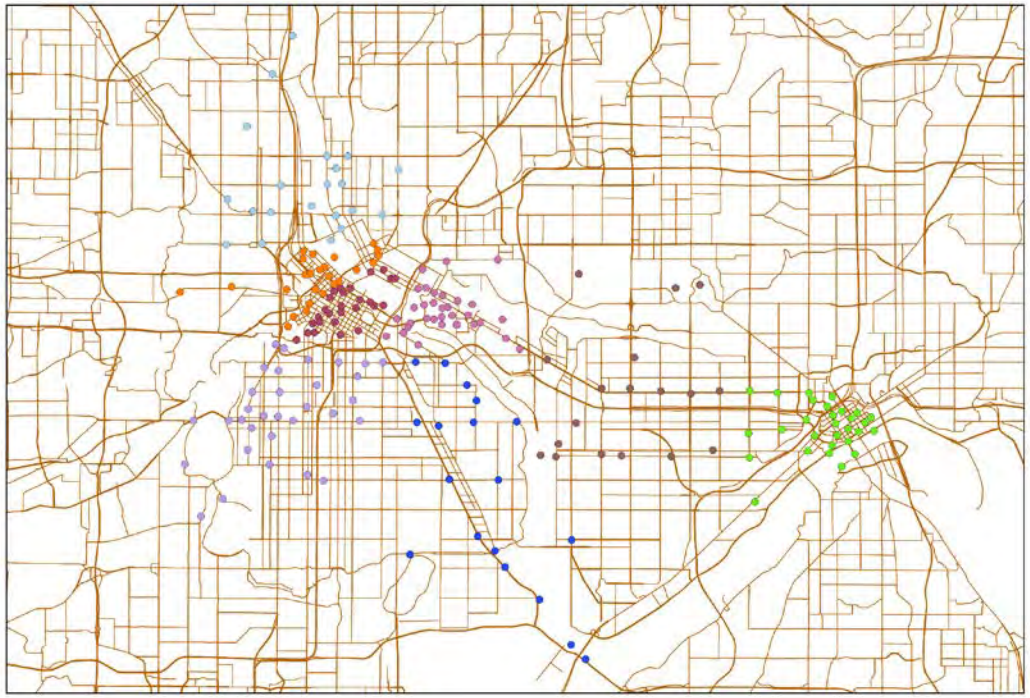
Rebalancing vehicle routing strategy

- Rebalancing vehicle travelling among stations to pick up and drop off bikes to meet the demand
- Decide the route that minimizes travel cost but satisfy demands
- Problem size:

12,000 nodes

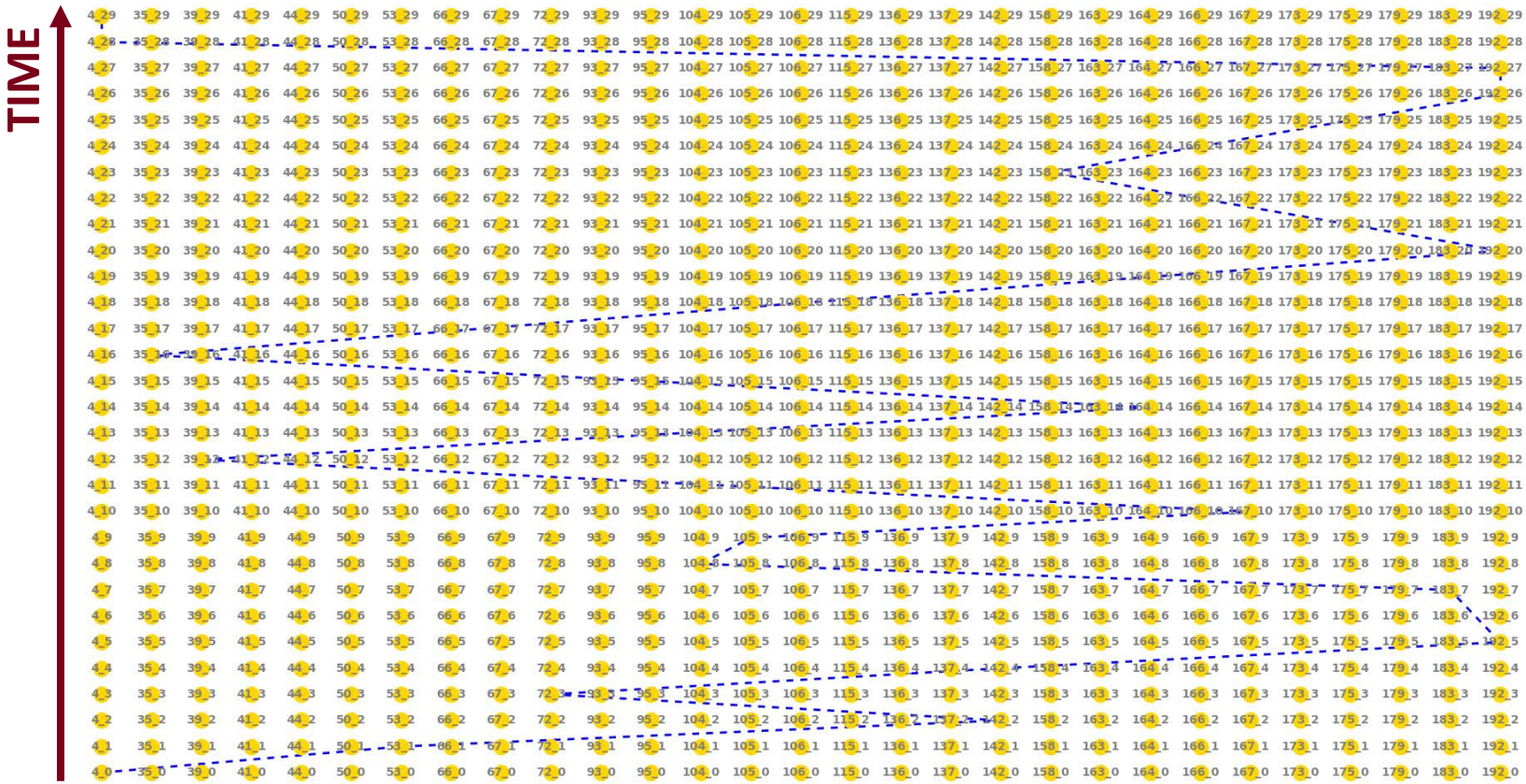
420,000 links

1 hour planning horizon



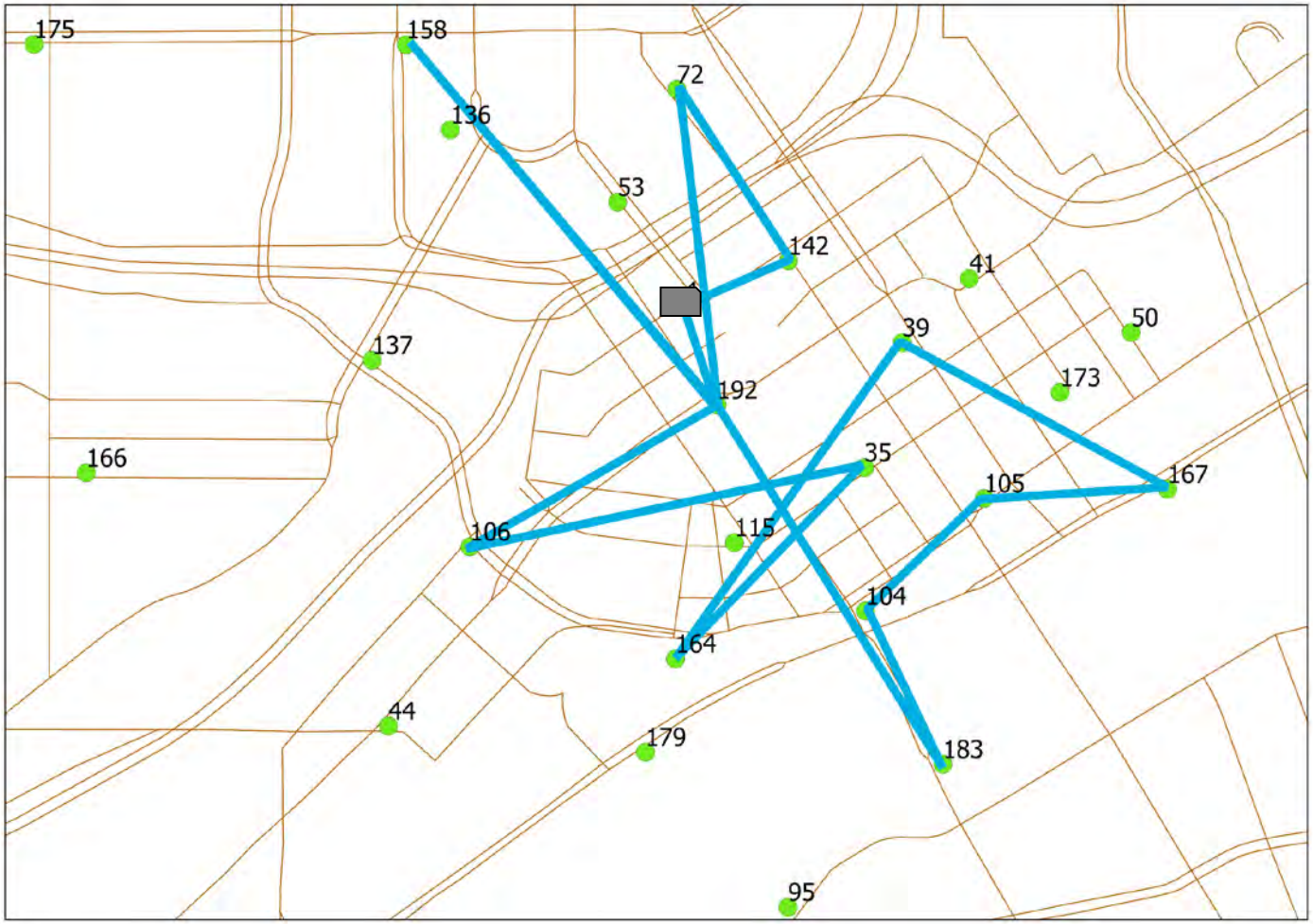
Routing Results (1)

- **Routing- Time Expanded Network**



Routing Results (2)

- **Routing- Physical Network**



Station ID	Demand	Stock
66	1	6
179	-6	6
95	1	6
192	3	6
41	-3	6
50	-5	6
173	-2	6
106	-6	6
166	-3	6
136	4	6
137	-5	6
53	5	2
104	-4	8
44	-1	6
93	0	6
105	-7	7
158	-4	6
39	5	2
164	3	10
183	-7	7
175	-5	6
163	-3	6
142	1	0
72	-4	4
167	5	10
67	-6	6
4	0	20
115	3	6
35	-5	6

Conclusion

- Bike stations are clustered into 3 categories based on its usage profile
- Using estimated demand from clustering, the vehicle routing problem gives optimal rebalancing routes that meet the demands and at the same time lowers the travel cost.

Thanks!