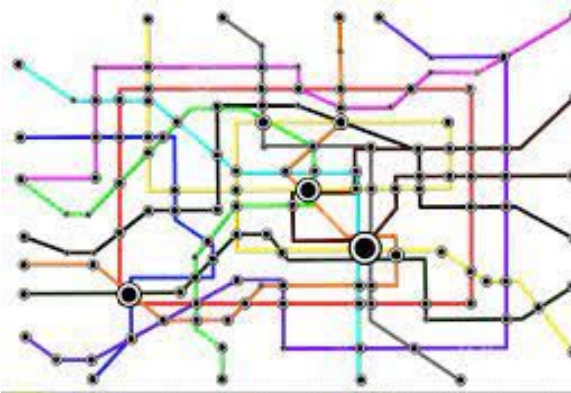


Network Analysis on Modern City Subway

Wheaton Wang, Kyle Yu, Brandon Bai.

Overview

- Subway serves as an essential infrastructure of the public transit system.
- Subway network is composed of stations(Node) and routes(Edge).
- Designing a subway network is complex and costly.




Goals

- By analyzing existing subway networks, we want to...
 - Find correlations between subway station/route distribution and other important factors, such as population distribution, daily population flow, and etc.
 - Decide factors of a successful subway network.
 - Be able to design an efficient subway network.



Steps

- Datasets collection: Station/Route datasets and other datasets (population distribution datasets for Milestone 1).
 - Generate subway network graphs on Gephi and calculate graph properties.
 - Visualize the network.
 - Analyze the relationship between graph properties and other factors.
 - Make conclusions and predictions.
 - Investigate potential improvement of the subway network.
- 

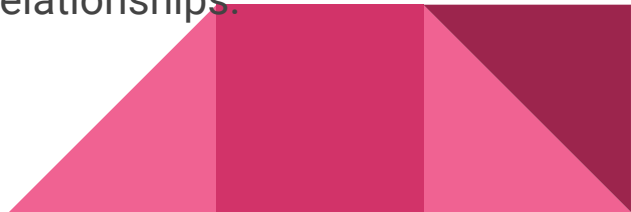
Assumptions

- Assume any route between two stations is undirected.
- Subway network datasets are posted on March 05 2021 on MTA's website. We assume the datasets are up to date.
- For Milestone 1, Population distribution datasets are generated from the 2020 American Community Survey. We assume the datasets are accurate and up to date.



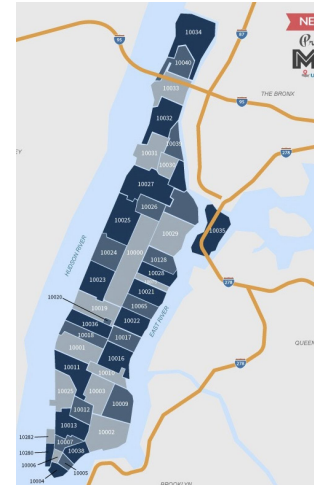
Challenge

- Data collection is difficult.
 - Manually process and create datasets for each city.
- Incorporate population distribution into the subway network graph is hard.
 - Focus on population distribution based on zip codes.
- Hard to draw conclusions/relationships between network properties and other factors.
 - Try more network properties (5 and more) to find the relationships.



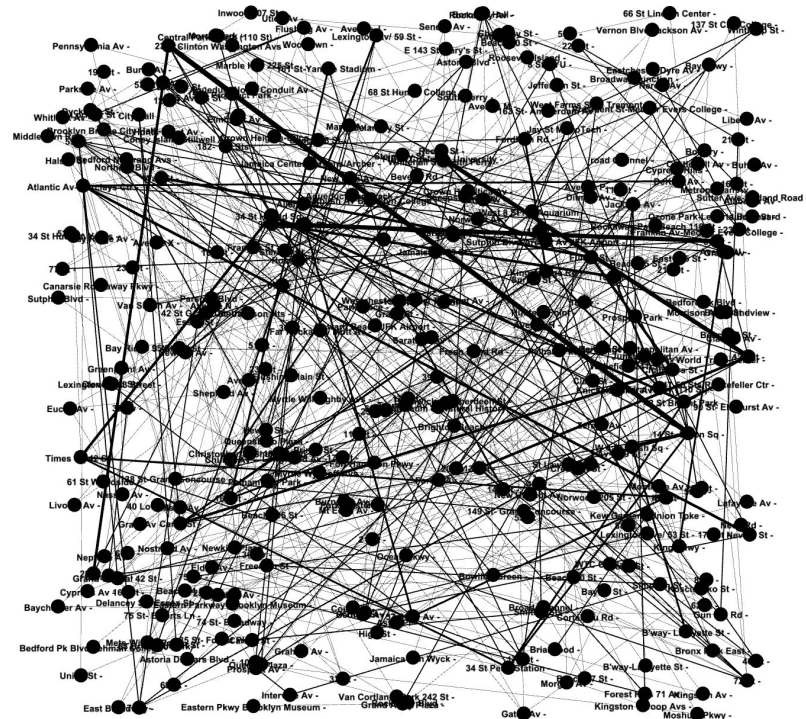
Milestone 1

- For Milestone 1, we focus on the New York City Subway Network.
- Aim to find the relationship between graph properties (such as betweenness centrality) and population distribution.



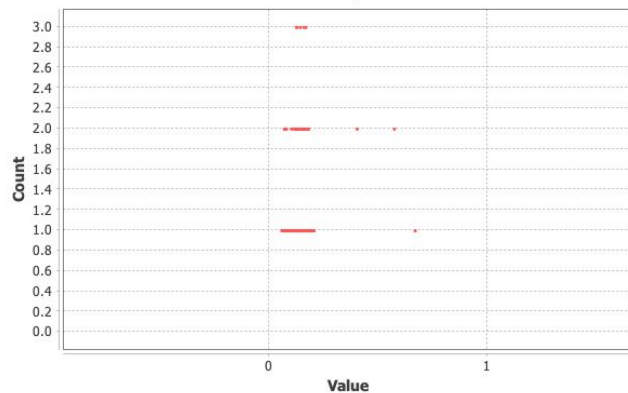
Preliminary Results

- Nodes: 343
- Edges: 487
- Average Degree: 2.84
- Network Diameter: 27
- Graph Density: 0.008
- Highest Degree: 14
- Highest Eccentricity: 27
- Highest Betweenness: 19482.275693

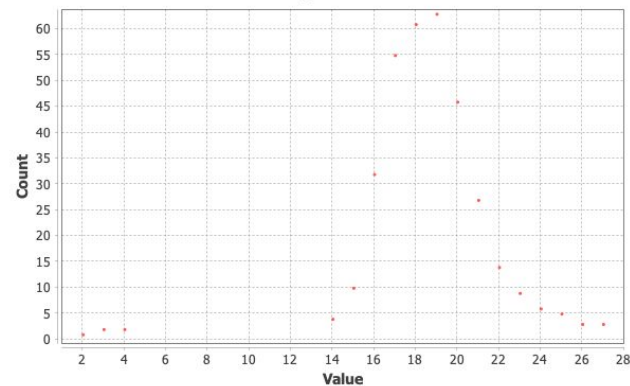


Preliminary Results - Visualize and Analyze

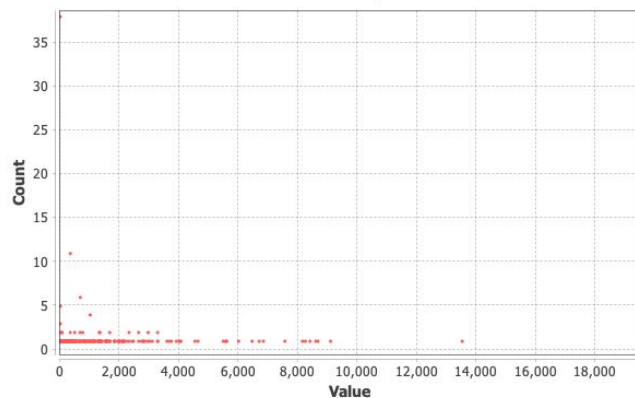
Closeness Centrality Distribution



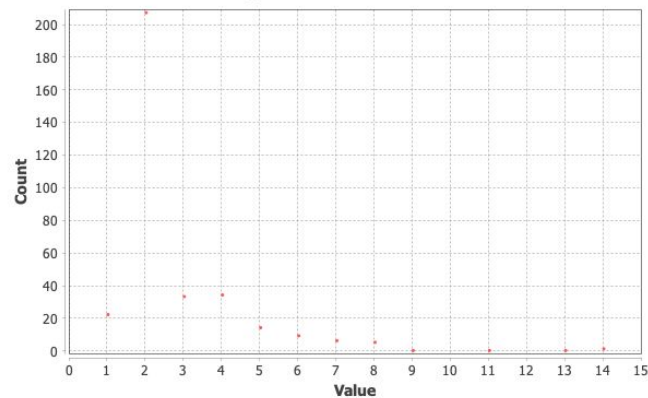
Eccentricity Distribution



Betweenness Centrality Distribution

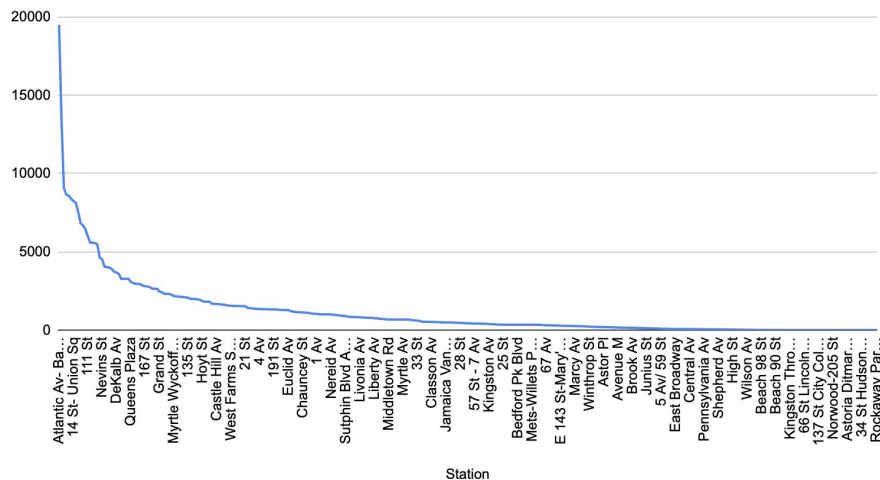


Degree Distribution

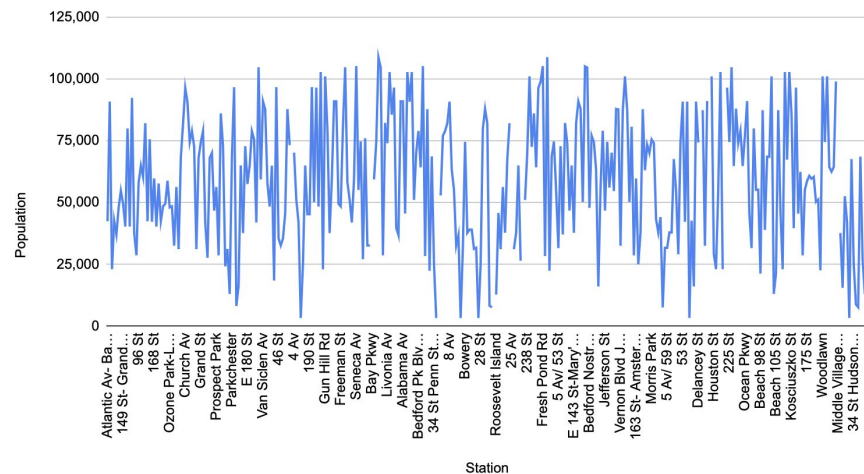


Preliminary Results - Betweenness vs Population

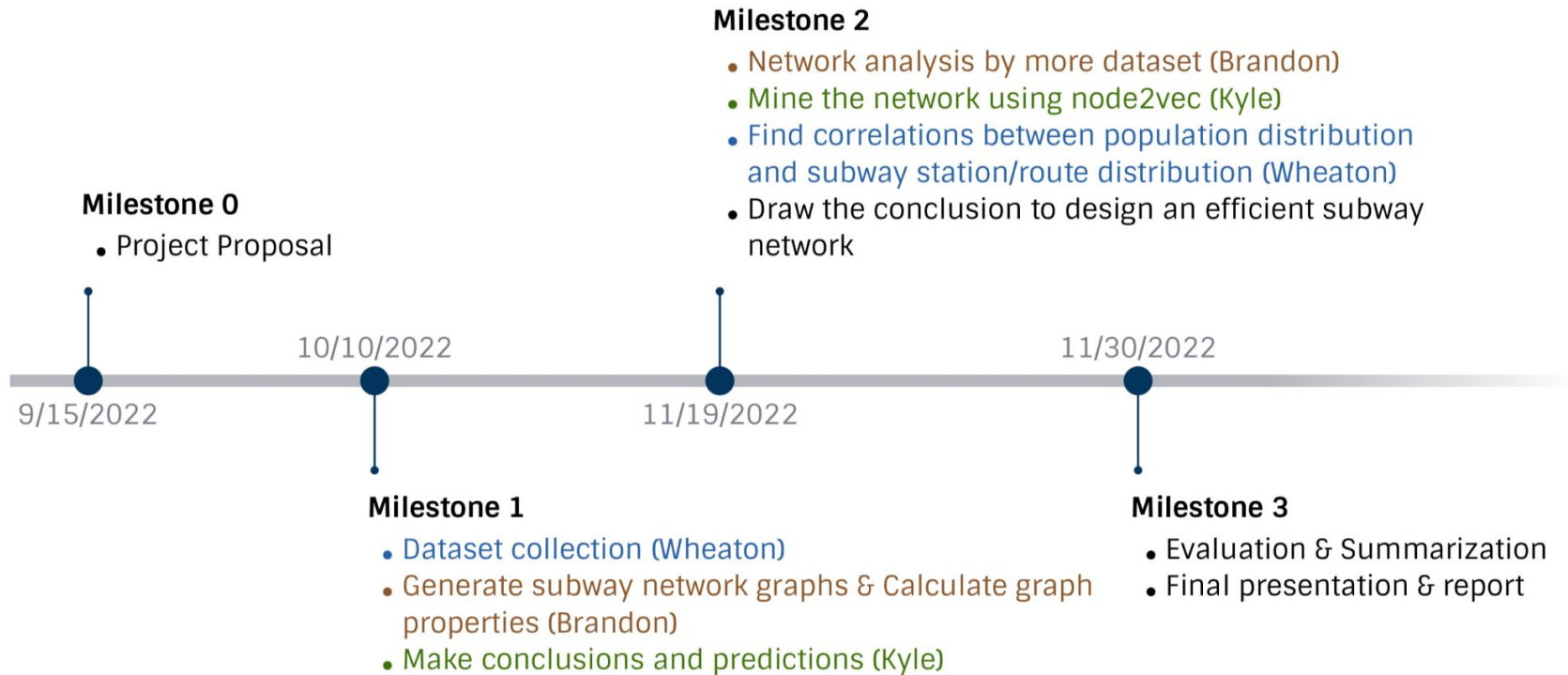
Betweenness vs. Station



Population vs. Station



Timeline



Conclusion

Subway Network

- undirected weighted edge
- Station name, population, region

Visualize & Analyze

- Population vs station traffic
- Location vs station traffic

Milestone 1 Conclusion

- NYC subway stations and population are weakly correlated.



References

[1] <https://new.mta.info/maps/subway-line-maps>

[2] https://www.newyork-demographics.com/zip_codes_by_population



Thank you!

