15.C57 Optimization Methods Final Project

MIT Sloan School of Management

Fall 2024

Annapurni Subramanian, Malcolm Gaynor, Michelle Zhou

Designing an Optimal Subway Line for Detroit: Station Placement and Network Flow Optimization

1 Problem Statement

The goal is to design an optimal subway line for Detroit, Michigan that connects densely populated residential areas to high-density job locations. This problem is divided into two parts:

- Station Placement: Identify the best locations to establish stations based on population and job density as well as proximity to key locations such as universities, museums, stadiums, and concert venues.
- Network Flow Optimization: Establish a network flow that minimizes travel distances from populated areas to stations and from stations to job locations, ensuring efficient commuter flow.

2 Why Does It Matter?

Detroit Michigan, the largest city in the state of Michigan, and the 26th largest city (by population) in the United States (just below Boston at 25) does not currently have a subway system. [6] [3] This likely contributes to the fact that Detroit is considered one of the worst cities in the US for commuters, and for public transport in general. [1] [2] Therefore, designing a subway line, especially one focused on decreasing total commute time for employees in the metro Detroit area, would have a wide range of benefits, including decreasing traffic, transporting the population more efficiently and sustainably, and saving money, both for the citizens and the city as a whole.

3 Data Collection and Processing

3.1 Station Placement Data

To determine station placement, we used data from the United States Census Bureau, specifically home and work population data for Michigan (LODES- Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics). [4] Each location in the census data is assigned a unique geocode. The geocode is a 15-character string comprised of the state FIPS code (2) + county FIPS code (3) + census block code (6) + census block code (4). That code uniquely identifies that block among all other more than 11 million census blocks covering the U.S. wall to wall. We filtered Detroit's census block data from the entire Michigan dataset using these geocodes, yielding 324 census blocks, which we used as station candidates. For each census block, we extracted centroid coordinates (longitude and latitude) using the TIGER/Line® Shapefiles provided by the Census Bureau. [5] These centroids allowed us to calculate pairwise distances between stations. Pre-calculated distances were then stored to simplify computations in the subsequent optimization process. This data also includes the lattitude and longitude of six key locations in Detroit (the airport, a concert venue, the University of Detroit Mercy, two museums, and the football stadium), as we want to ensure that the subway system services these locations.

3.2 Network Flow Optimization Data

To conduct the network flow optimization, we required paired data detailing the travel patterns of individuals, i.e., the home and work locations of people associated with each station. Since no reliable, high-quality paired dataset was available, we employed a synthetic assignment algorithm. This algorithm randomly assigned individuals from each station's residential population to work locations. Although this synthetic assignment introduces assumptions, it provides a foundational dataset for optimization. If the Detroit Transportation Department acquires more accurate paired travel data in the future, the methodology outlined in this paper can be replicated using real-world data to enhance its precision. This approach ensures reproducibility and allows flexibility for future improvements with better data availability.

4 Formulation

The optimization problem for designing an efficient subway system in Detroit involves two key components: station placement and network flow optimization. These components are mathematically modeled to minimize total travel distances while ensuring an effective connection between residential areas, job locations, and key landmarks in the city. The formulation encapsulates these objectives through a combination of binary decision variables, flow constraints, and distance-based minimization, ensuring that the solution is both practical and computationally feasible.

4.1 Station Placement Formulation

The station placement component focuses on identifying optimal locations for subway stations based on population density, job density, and proximity to significant landmarks like the airport, university, stadiums, and cultural venues. A binary decision variable determines whether a station is built at a given location. The objective is to minimize the total distance traveled by residents to their nearest station while ensuring that all high-density population centers and job hubs are adequately served. Constraints are introduced to cap the number of stations and ensure that stations are placed strategically near key locations, enhancing accessibility and commuter convenience.

The following indices were used for the formulation:

- $i = 1, \ldots, I$: Population density index (home locations of residents)
- $j = 1, \ldots, J$: Job density index (job locations of residents).
- $k = 1, \ldots, K$: Candidate station index (locations evaluated for optimal station placement)
- $m = 1, \ldots, M$: Key location index (universities, concert venues, museums, stadiums)

The following parameters were used for the formulation:

- $d1_{ik}$: Euclidean distance between population center i and station k
- $d2_{jk}$: Euclidean distance between job center j and station k
- $d3_{mk}$: Euclidean distance between key location m and station k
- p_i: Population at index i
- w_j: Number of workers at index j

The following decision variables were used for the formulation:

- $x_k \in \{0, 1\}$: Binary variable, where $x_k = 1$ if station k is built, and 0 otherwise
- $a_{ik} \geq 0$: Amount of people going from population center i to station k
- $b_{jk} \geq 0$: Amount of people going from job center j to station k
- $c_{mk} \in \{0, 1\}$: Binary variable, where $c_{mk} = 1$ if people are going from key location m to station k, and 0 otherwise

The objective function is defined to minimize the total distance traveled by individuals, encompassing the journey from their home to the nearest station and from the station to their workplace. Constraint (1) ensures that the total number of people traveling from a population center i to any station k equals the population p_i at that center. Constraint (2) ensures that the total number of people traveling from any station k to a job center j equals the total number of workers w_j at that job center. This ensures that all workers are assigned to a station. Constraint (3) ensures that each key location m (such as an airport, university, or stadium) is assigned to exactly one station k. This ensures that every key location is connected to the subway network. Constraints (4)–(6) Link the decision variables a_{ik}, b_{jk}, and c_{mk} to the binary station placement variable x_k. These constraints ensure that people or flows can only travel to or from stations that are built. Constraint (7) Limits the total number of stations that can be built to 25, ensuring feasibility within a given budget or infrastructure constraint. Constraint (8) Ensures that the total weighted distance from key locations to their assigned stations does not exceed a threshold 0.02. This guarantees accessibility for key locations. Constraints (9)–(12) Define the non-negativity and binary nature of the decision variables. Below is the formulation:

$$\min \quad \sum_{i=1}^{I} \sum_{k=1}^{K} d1_{ik} \cdot a_{ik} + \sum_{j=1}^{J} \sum_{k=1}^{K} d2_{jk} \cdot b_{jk}$$

Subject to:

$$\sum_{k=1}^{K} a_{ik} = p_i, \quad \forall i \in \{1, \ldots, I\} \tag{1}$$

$$\sum_{k=1}^{K} b_{jk} = w_j, \quad \forall j \in \{1, \ldots, J\} \tag{2}$$

$$\sum_{k=1}^{K} c_{mk} = 1, \quad \forall m \in \{1, \ldots, M\} \tag{3}$$

$$a_{ik} \leq M \cdot x_k, \quad \forall i \in \{1, \ldots, I\}, \forall k \in \{1, \ldots, K\} \tag{4}$$

$$b_{jk} \leq M \cdot x_k, \quad \forall j \in \{1, \ldots, J\}, \forall k \in \{1, \ldots, K\} \tag{5}$$

$$c_{mk} \leq M \cdot x_k, \quad \forall m \in \{1, \ldots, M\}, \forall k \in \{1, \ldots, K\} \tag{6}$$

$$\sum_{k=1}^{K} x_k \leq 25 \tag{7}$$

$$\sum_{m=1}^{M} d3_{mk} \cdot c_{mk} \leq 0.02 \tag{8}$$

$$a_{ik} \geq 0, \quad \forall i \in \{1, \ldots, I\}, \forall k \in \{1, \ldots, K\} \tag{9}$$

$$b_{jk} \geq 0, \quad \forall j \in \{1, \ldots, J\}, \forall k \in \{1, \ldots, K\} \tag{10}$$

$$x_k \in \{0, 1\}, \quad \forall k \in \{1, \ldots, K\} \tag{11}$$

$$c_{mk} \in \{0, 1\}, \quad \forall m \in \{1, \ldots, M\}, \forall k \in \{1, \ldots, K\} \tag{12}$$

4.2 Network Flow Optimization Formulation

Before initiating the network flow optimization, once the stations to be built have been identified, all individuals from different home locations are assigned to the closest station. Similarly, all individuals who work at different locations are assigned to a closest work station that is built. The model assumes that no individual both lives and works at the same station, ensuring that all flows involve commuting between distinct stations.

The objective of the network flow optimization is to minimize the total distance traveled by commuters, weighted by the flow of people between stations. This ensures the system is designed to optimize efficiency, reduce commuting times, and streamline the flow of commuters between residential areas, job centers, and key locations within the network.
The following indices are used in the formulation:

- $i, j \in \{1, \ldots, I\}$: Station indices.
- $z \in \{0, 1\}$: 0 if outward flow, 1 if inward flow.
- $l \in \{1, \ldots, L\}$: Home location indices.

The following parameters are used in the formulation:

- p_l: Population at home location l.
- b_{lu}: Number of people living at home location l and working at station u.

- d_{ij}: Euclidean distance between stations i and j.

The following decision variables are used in the formulation:

- $x_{ij} \in \{0,1\}$: Binary variable, where $x_{ij} = 1$ if a connection exists between station i and station j, and 0 otherwise.
- $f_{ijlz} \geq 0$: Continuous variable representing the flow of people who live at home location l traveling between stations i and j, where $z = 0$ represents outgoing flow and $z = 1$ represents incoming flow.

The objective of the network flow optimization is to minimize the total distance traveled by commuters, weighted by the flow of people between stations. Constraint (13) ensures that connections between stations are bidirectional, meaning $x_{ij} = x_{ji}$. constraint (14) ensures that the total flow leaving each home location l must equal the population at that location. constraint 15) controls the total flow reaching work locations must meet the work demand. constraint (16) ensures that if no connection exists between stations i and j, then the flow must be zero. constraint (17) ensures that each station has exactly two outward connections. Constraint (18) is the flow balance constraint that ensures The total flow into a station equals the total flow out plus the work demand at that station. Below is the formulation:

$$\min \quad \sum_{l=1}^{L}\sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{z=0}^{1} f_{ijlz} \cdot d_{ij}$$

subject to:

$$x_{ij} = x_{ji}, \quad \forall i,j \in \{1,\ldots,I\} \tag{13}$$

$$\sum_{j=1}^{I} f_{ljl1} = p_l, \quad \forall l \in \{1,\ldots,L\} \tag{14}$$

$$\sum_{j=1}^{I} f_{ijl0} \geq b_{lu}, \quad \forall l \in \{1,\ldots,L\}, u \in \{1,\ldots,I\} \tag{15}$$

$$f_{ijlz} \leq M \cdot x_{ij}, \quad \forall i,j \in \{1,\ldots,I\}, l \in \{1,\ldots,L\}, z \in \{0,1\} \tag{16}$$

$$\sum_{i=1, i\neq j}^{I} x_{ij} = 2, \quad \forall j \in \{1,\ldots,I\} \tag{17}$$

$$\sum_{l=1}^{L}\sum_{i=1,i\neq j}^{I} f_{ijl0} = \sum_{l=1}^{L}\sum_{i=1,i\neq j}^{I} f_{jil1} + \sum_{l=1}^{L} b_{lu}, \quad \forall j \in \{1,\ldots,I\} \tag{18}$$

5 Results & Discussions

Station Placement

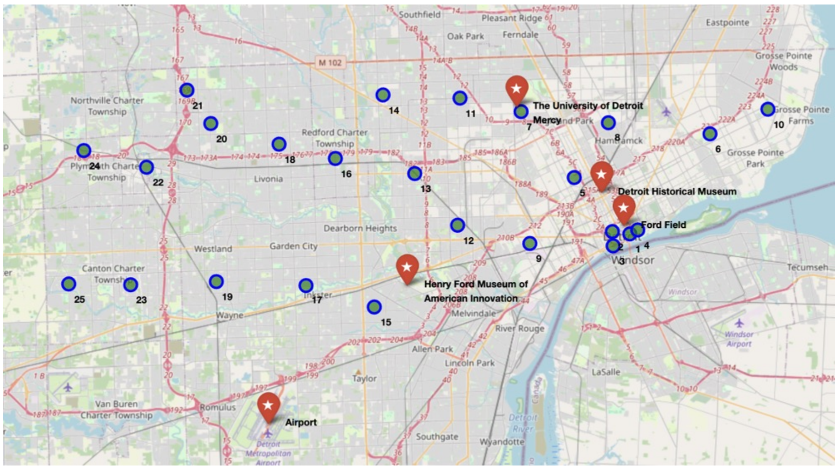

Figure 1: Locations of Selected Stations and Key Locations in Detroit

Among the 324 potential locations, our optimization model selected 25 stations for the Detroit subway system. As shown in Figure 1, the results demonstrate a well-aligned station placement strategy, with stations scattered in high-population areas and strategically located near key areas to maximize convenience for commuters and travelers.

In particular, the selected stations are located close to high-demand locations, ensuring accessibility for residents and visitors alike. For example: station 7 is placed within a block of University of Detroit Mercy, making it convenient for students, faculty, and staff. Two stations 12 and 15 are located near the Detroit Historical Museum, which improves the access of tourists and cultural enthusiasts. Also several stations are concentrated in the downtown area, which features a high density of traffic, businesses, and iconic landmarks such as Ford Field.

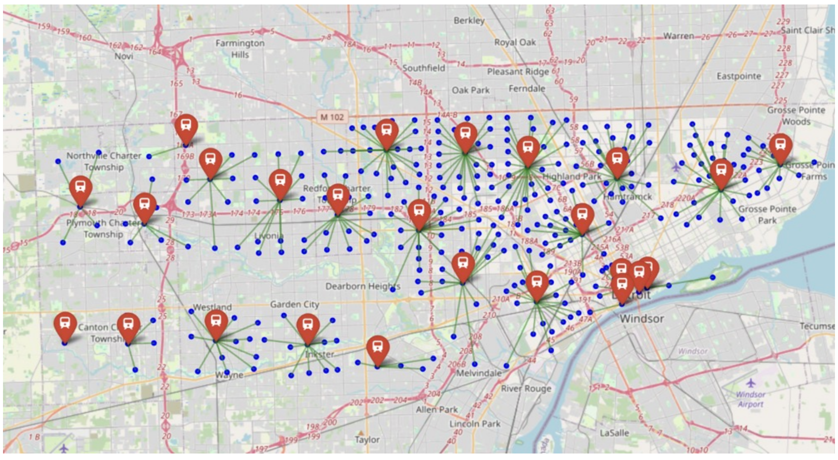

Figure 2: Commuting Paths for Each Census Block

Figure 2 visualizes how people who live or work in each census block will travel to the stations we built, highlighting the effectiveness of our station placement strategy. The selected stations are strategically located to minimize travel distances while ensuring coverage of high-density population areas and key landmarks, which proves the success of our optimization formulation. These stations will serve the city efficiently, minimizing the total distance residents must travel to their nearest station.

Subway Network Connections

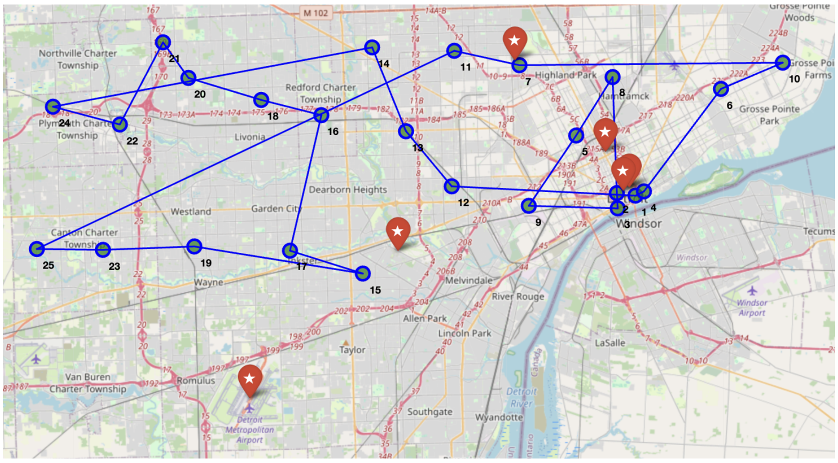

Figure 3: Subway Network Connections for Home-to-Work Travel

Figure 3 illustrates the optimized connections of the subway network designed to ensure that all residents can travel efficiently from their homes to their workplaces with minimal distance. Given that the Gurobi optimization output already ensures all residents can get to their work space with minimized distance and time spent, we can make a few refinements on its result based on our practical considerations:

- Hub Development: Stations 16, 20, 2, and 15 will be designated as hubs due to their significant role as intersection points in the network. This hub-based design facilitates better accessibility and reduces congestion in high-traffic areas, as these stations will have 4 connections, instead of the usual 2.

- Direct Downtown Connection: A direct connection between Stations 15 and 12 is proposed to provide quicker access to the downtown area, ensuring reliable access to the high traffic demand in that region, regardless of workplace locations.

- Airport Express Line: An additional express line can be built between Station 15 and the airport to ensure seamless travel for passengers traveling to and from the airport, even though it lies outside of the metro Detroit area.

- Optimized Flow Adjustment: To improve network flow, we can modify the connection between Station 10 and Station 7 to avoid bypassing Station 8.

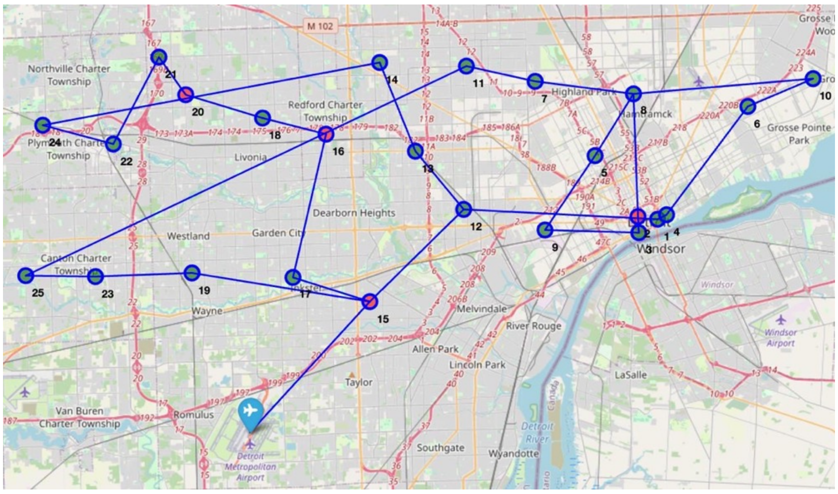

Figure 4: Final result after making "common-sense" changes

These adjustments are informed by both optimization results and common-sense urban planning principles. By integrating practical revisions into the optimization output, like shown in Figure 4, the proposed subway network ensures comprehensive connectivity, reduces travel times, and better aligns with the commuting needs of Detroit's residents and surrounding areas.

6 Impact

If implemented, this subway system would be extremely beneficial for the Detroit citizens. The large amount of commuters, who this implementation is designed to benefit, would have another option outside of driving to commute to work. This would not only decrease travel time and effort for the commuters who choose to use the subway, but also for those who drive (by personal choice or because they live outside of Detroit), as the total traffic on the streets would decrease. This system would help non-working people as well, including students (the implementation includes a stop at University of Detroit Mercy) and tourists (this system brings people directly to the airport, and has stops in close proximity to major tourist destinations). Therefore, this subway system could potentially draw more people to live, work, and visit Detroit, as it would be easier and cheaper to move around the city.

7 Future Scope

First of all, replacing the synthetic data with real data would add legitimacy to the formulation. Also, the common sense additions to the final subway plan could be integrated directly to the formulation. For example, we could allow for multiple hubs (stations with flow to 4 different output stations), and ensure that every station is within some minimum number of connections to downtown Detroit. Finally, we could take into account more than just occupation and residence data, such as integrating the subway system within the existing public transport infrastructure, or more directly include the needs of the non working population to the formulation.

References

[1] Felton, R. How Detroit ended up with the worst public transit. *Detroit Metro Times*. 2014. Available at: `https://www.metrotimes.com/news/how-detroit-ended-up-with-the-worst-public-transit-2143889`. Accessed: Dec 6, 2024.

[2] Keenan, T. Study: Detroit is the Worst Metro Area for Commuters in 2022. *DBusiness*. 2022. Available at: `https://www.dbusiness.com/daily-news/study-detroit-is-the-worst-metro-area-for-commuters-in-2022/`. Accessed: Dec 6, 2024.

[3] Reindl, J. A subway system in Detroit? Here are 6 times the city tried — and failed. *Detroit Free Press*. 2024. Available at: `https://www.freep.com/story/money/business/2023/12/03/detroit-subway-system-people-mover/71592145007/`. Accessed: Dec 6, 2024.

[4] United States Census Bureau. Longitudinal Employer-Household Dynamics. *US Census Bureau*. 2021. Available at: `https://lehd.ces.census.gov/data/`. Accessed: Dec 6, 2024.

[5] United States Census Bureau. TIGER/Line Shapefiles. *US Census Bureau*. 2024. Available at: `https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html`. Accessed: Dec 6, 2024.

[6] Wikipedia. List of United States cities by population. *Wikipedia*. Available at: `https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population`. Accessed: Dec 6, 2024.