THE SUBWAY CHALLENGE

World Record Route Generator

01 TRANSLATION

02 POSTMAN

03 STORAGE

04 LIMITATIONS

Denzel S. Williams

Data Science = Solving Problems = Happiness

The Challenge

Navigate the entire New York City Subway System, stopping at all **472 stations**, in the shortest time possible.

The Challenge

The record for the 469-Station Challenge is **21 Hours, 28 Minutes, 14 Seconds** set by Matthew Ahn in 2016. Nobody holds the record for the 472-Station Challenge.

The goal of this project is to use graph theory to determine a set of paths that could potentially set a world record in the 472-Station Challenge.

THE APPROACH

TRANSLATION

Model the MTA Subway System as a weighted undirected graph.

CHINESE POSTMAN

Modify a preexisting algorithm to suit the needs of the challenge.

STORAGE

Store the different routes in an SQLite database.

01

TRANSLATION

Nodes (Vertices)

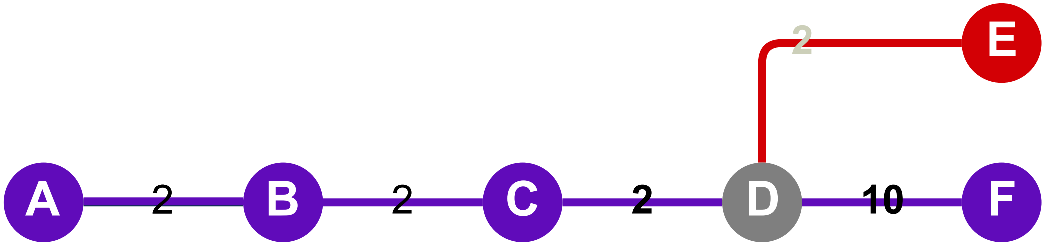
The Nodes in the graph represent the stations of the MTA Subway System

Edges

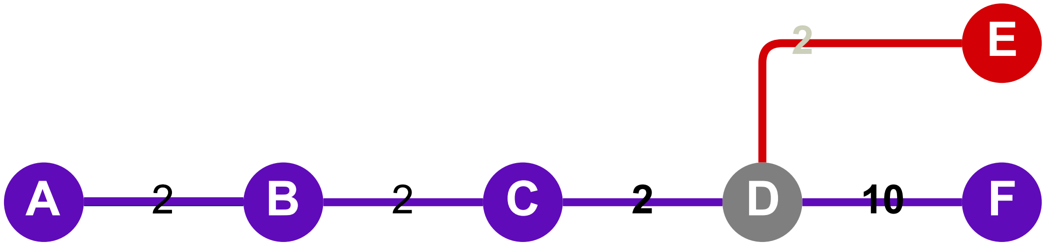
The Edges in the graph represent the train tracks that connect stations to each other. The weight on an edge is the time to travel between stations.

All stations DON'T matter. Some stations do not need to be modeled as nodes.

The network can be “reduced” by removing stations that don’t require a decision.

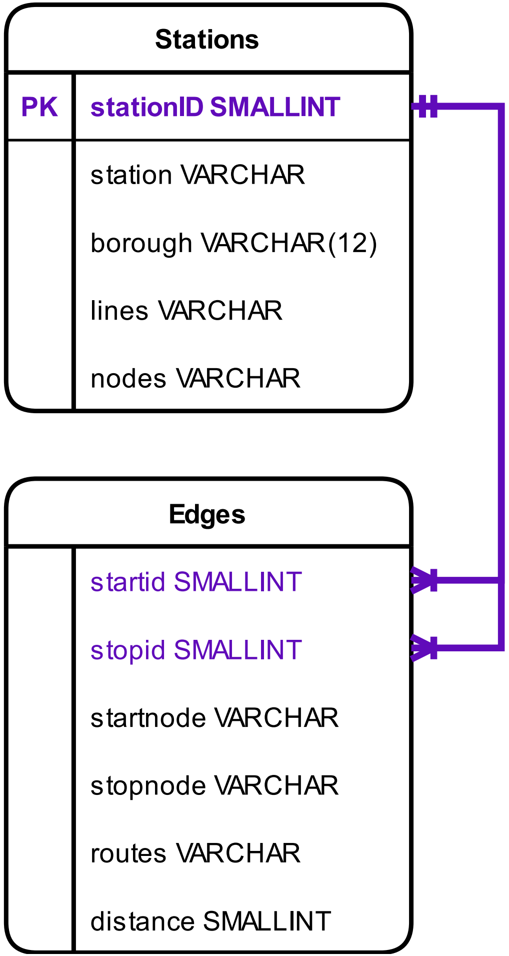


On the way from A to D you have **NO CHOICE** but to stop at stations B and C. Therefore, they are not needed.

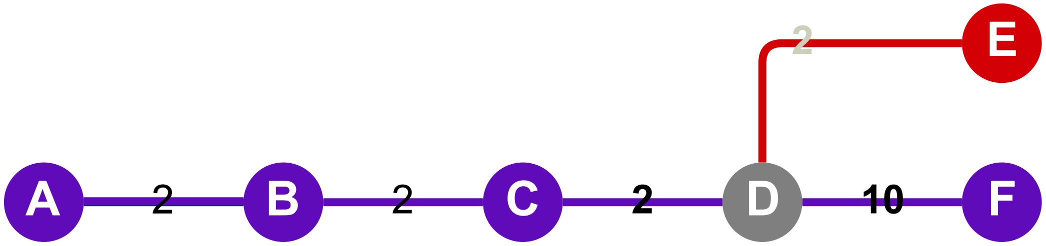


The stations remaining are called "**decision points**". Of the 472 stations only 79 of them are decision points.

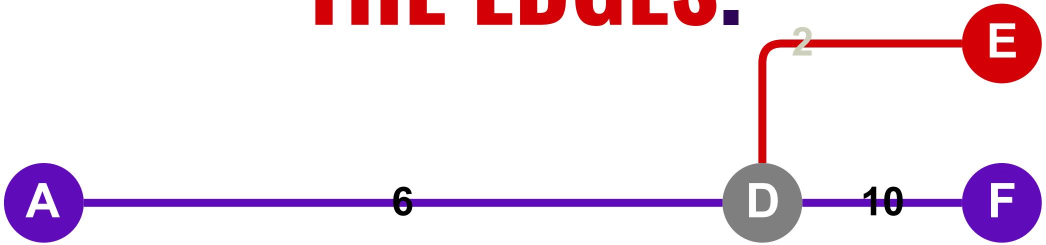
The Reduced Network was translated and stored in two CSV files to be used by the algorithm and in two tables to be read by a human.



Distance in the Edges table is the TIME between stations



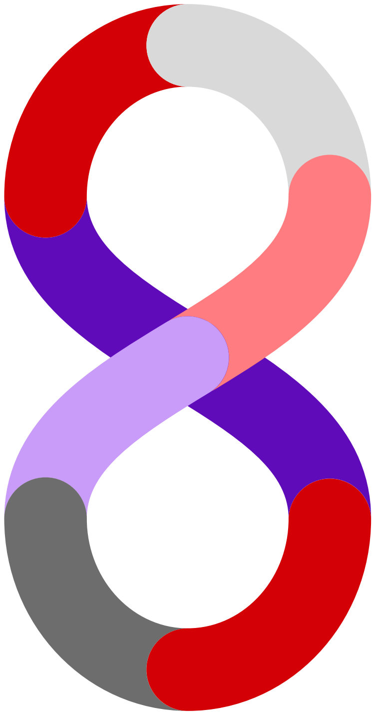
By reducing the network, the focus shifts from nodes to edges. The ONLY way to get to all the stations is by traversing ALL THE EDGES.



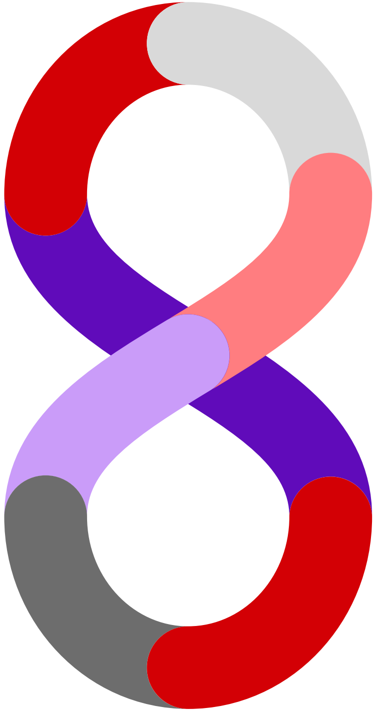
CHINESE POSTMAN

02

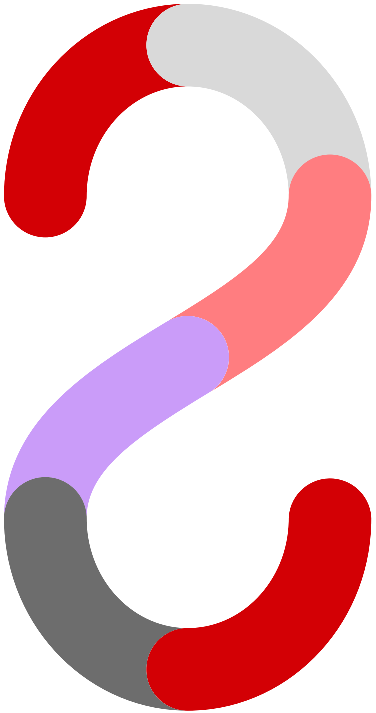
The postman’s job is to deliver all of the town’s mail using the **SHORTEST** route possible. He must pass each street **AT LEAST ONCE** and then **RETURN TO THE ORIGIN.**



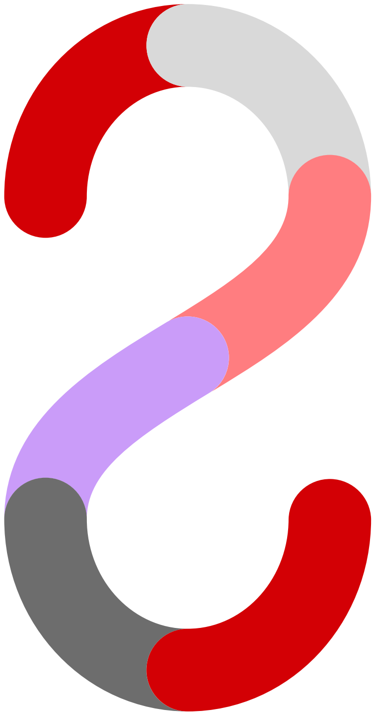
In graph theory, solving the Chinese Postman Problem (CPP) is to find the shortest **CIRCUIT** that visits every edge of a graph.



The Subway Challenge is the same problem, **EXCEPT** we don't need to return to the origin. We only need to "pass each street" **AT LEAST ONCE.**



Our CPP is trying to find the shortest **PATH** that visits every edge of a graph.



Which means we don’t solely care about where to START the challenge, we also need to consider where to END it.

All decision points **DON’T** qualify. Due to a theorem, only certain stations can be used as start/end nodes, specifically **odd-degree** nodes.

CPP DIFFERENCES

Regular CPP

Find a **circuit** that traverses each edge.

Odd Nodes

Turns **ALL** odd-degree nodes even.

Algorithm

Uses the **Euler Circuit** function in NetworkX

Subway CPP

Find a **path** that traverses each edge

Odd Nodes

Turns all **BUT TWO** odd-degree nodes even. *They will be used as the start and end stations.*

Algorithm

Uses the **Euler Path** function in NetworkX

Of the 79 decision points, 58 of them qualify. Meaning, there are 1653 possible routes that would complete the challenge.

03

STORAGE

PROGRESS REPORT

TRANSLATION

The subway was translated into CSV files: one for nodes the other for edges.

LOADING

Both files were uploaded to the database and the edge file was loaded into the program.

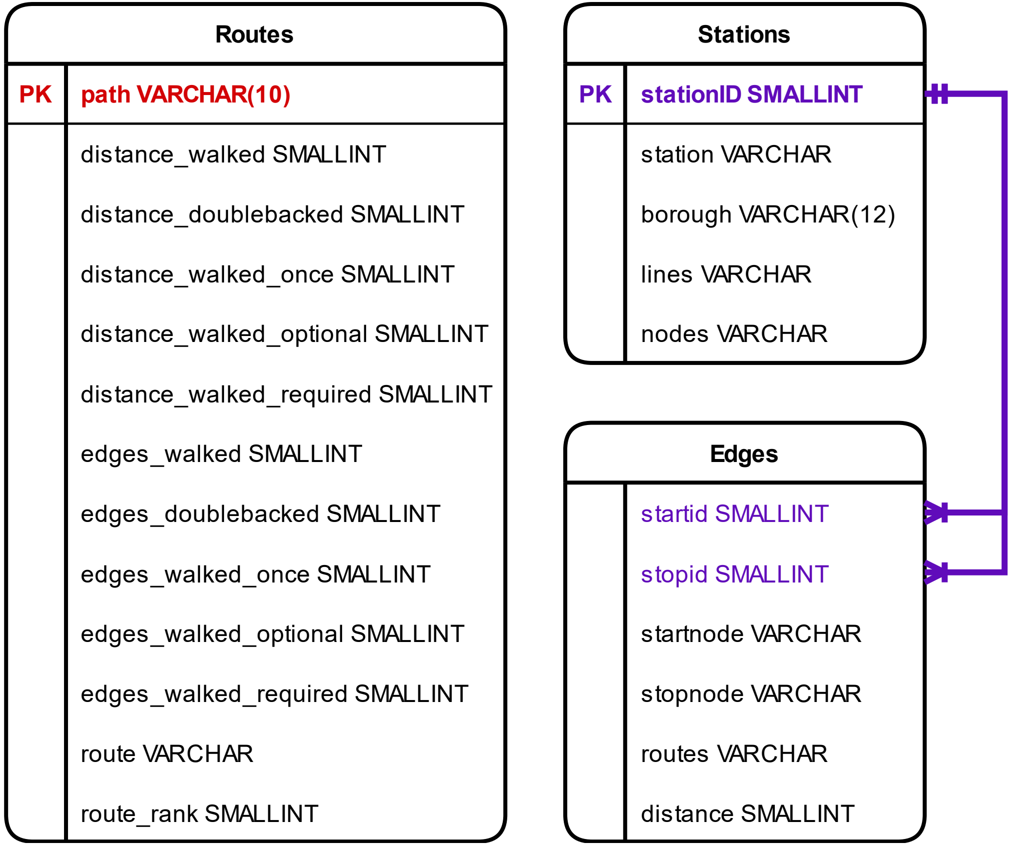
ROUTE FINDING

All 58 odd-degree nodes were paired with each other creating 1653 routes.

STORAGE

The 1653 routes, their metrics, and the steps in the path were integrated into the SQLite database as their own table.

The Full Database



The metric used to rank the routes is the **distance walked**. This metric represents the time it will take to complete the challenge.

Routes	
PK	path VARCHAR(10)
	distance_walked SMALLINT
	distance_doublebacked SMALLINT
	distance_walked_once SMALLINT
	distance_walked_optional SMALLINT
	distance_walked_required SMALLINT
	edges_walked SMALLINT
	edges_doublebacked SMALLINT
	edges_walked_once SMALLINT
	edges_walked_optional SMALLINT
	edges_walked_required SMALLINT
	route VARCHAR
	route_rank SMALLINT

The metric used to rank the routes is the **distance walked**. This metric represents the time it will take to complete the challenge.

Routes	
PK	path VARCHAR(10)
	distance_walked SMALLINT
	distance_doublebacked SMALLINT
	distance_walked_once SMALLINT
	distance_walked_optional SMALLINT
	distance_walked_required SMALLINT
	edges_walked SMALLINT
	edges_doublebacked SMALLINT
	edges_walked_once SMALLINT
	edges_walked_optional SMALLINT
	edges_walked_required SMALLINT
	route VARCHAR
	route_rank SMALLINT

You only spend 14.75 Hours traversing unique edges. The rest of time (6+ Hours) is spent double backing.

The "best" route starts at Wakefield-241 St *(2-Train)* and ends at Woodlawn *(4-Train).* The route takes **20.65 Hours** to complete with 150+ steps to follow.

37th Place

The ranking of the route that used the same start and end pair that Matthew Ahn used.

122 Routes

The number of routes that could potentially beat Ahn's 469-station time in the 472-station challenge.

LIMITATIONS & NEXT STEPS

04

LIMITATIONS

LATE-NIGHT MAP

To simplify the process, the Late-Night Subway Map was used instead of the Full Sevice Map. Therefore, none of the paths can be applied exactly as stated.

TIME VARYING NETWORK

Both the late-night map and the regular map are valid. Decisions are different when different maps are in session.

WAIT TIMES

The algorithm doesn’t understand that train switching is an expensive task that adds to the overall time. Too much train switching can make an optimal route sub-optimal. Additionally, wait times also vary throughout the day.

SUBWAY TRANSFERS

There are useful transfers directly between stations that couldn't be modeled.

ARTIFICAL EDGES

To improve his time, Matthew ran between disconnected stations because it was faster than using the train. These "running edges" can be added to the network.

NEXT STEPS

I
CHALLENGE
YOU TO
BEAT THE
RECORD