Project Results Update

New Data

I have significantly expanded the dataset that the algorithm considers. It now incorporates points of high transit potential across multiple categories, including employment, entertainment, healthcare, government, tourism, and military. I sourced this comprehensive dataset by accessing the REST APIs from Open Data DC and Maryland state data resources. I am currently working on adding the equivalent data for Virginia.

Network Algorithm Improvements

- Adjustable Louvain Community Agglomeration

I introduced a "resolution" parameter to the Louvain community detection algorithm. This parameter allows control over the granularity of community merging. By increasing the resolution from 1 (in the previous update) to 10, I achieved a finer partitioning, resulting in a slightly greater number of stations. This adjustment provides a more detailed and realistic representation of transit hubs.

Figure 3. Louvain Community clustering with resolution 1, old value.

Figure 4. Updated Louvain Community clustering with resolution 5.

- Removal of Minimum Spanning Tree pruning approach

I stopped using the minimum spanning tree (MST) algorithm from the previous update. While MSTs ensure network connectivity, they produced networks that were too sparse and irregular.

- New graph pruning approach

To replace the MST pruning strategy, I implemented a basic greedy algorithm to reduce node degree in a way that I designed to reflect some basic real-world transit design considerations.

First, the algorithm removes edges where the angle between adjacent connections is 30 degrees or less. This reflects the real-world tendency for transit route interchanges to favor perpendicular or near-perpendicular angles. For every pair of edges from a given vertex, if the angle between them is found to be 30 degrees or less, the longer of the two edges is removed.

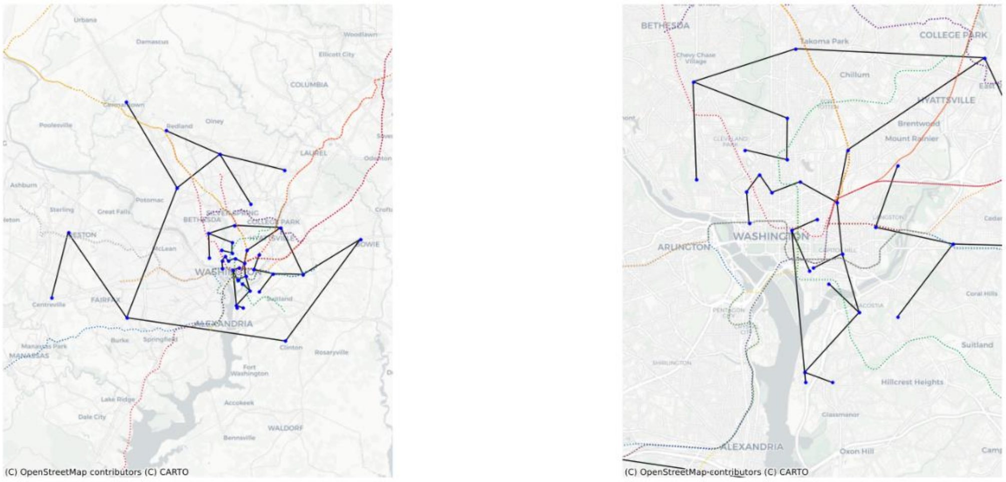

Figure 1. Old version of transit networks pruned using a minimum spanning tree.

Figure 2. New transit networks with improved pruning methods.

Second, for any nodes still exceeding a degree of 4, the algorithm prunes the longest remaining edges. This constraint reflects the typical structure of transit stations, which generally have at most four directions of travel, even if multiple lines interline within each direction (e.g., the intersection of the blue, orange, and silver lines with the green and yellow lines at L'Enfant Plaza is a good example).

In both of these steps, I choose to remove the longer of the conflicting edges because I want the network to tend towards closer connections.

- Transit Line Generation via Random Walks

I initiated work on the most exciting development so far: generating realistic transit lines. I achieve this using a sequence of random walks with some constraints. Each walk prioritizes maintaining the straightest path possible while avoiding revisiting vertices or edges. Subsequent walks may intersect previously traversed vertices but cannot reuse edges. Walks continue until they reach a user-specified maximum distance (based on edge weights representing spatial distances).

So, the overall pipeline is as follows: first, I load the points of transit interest into a graph structure. I then construct a Gabriel graph on the points with resolution of 5. I then prune the resulting Gabriel graph, and perform 12 random walks on the graph to create 12 transit lines.

Figure 5. First output of random walk transit networks. (Note that I am still working to resolve the color mismatch between plots.)

Discussion:

The updated algorithm produces more realistic and coherent transit networks, particularly across DC and Maryland. These networks now capture important structural features such as ring routes and suburb-to-suburb connections, reflecting real-life transit dynamics more comprehensively.

This week's progress is an exciting step forward, but I continue to look towards determining how to compare or score the networks produced.

The new data that I have incorporated seems to create a network that connects important areas. The litmus test of Anacostia and Georgetown now being connected is finally passed! The network now even connects to College Park, and with lines from College Park in several directions!

One issue is that there is a large discrepancy between the shortest lines and the longest lines. I will work to make sure the line length is largely more consistent.