Project Results Update

I expanded the original dataset to include census blocks in DC as well as the DC Metro Area counties in Maryland and Virginia. In MD, the counties are PG and Montgomery, while in Virginia they consist of Arlington County, Alexandria City, Fairfax County, Fairfax City, and Falls Church City. (I am currently leaving out Virginia's Loudoun County, Prince William County, Manassas City, and Manassas Park City for reasons I will get into further down.)

Figure 1. Map of census blocks in the metropolitan DC area, as defined above.

I have improved the network creation algorithm. As a reminder, in the last update, the network was constructed by creating a Gabriel graph of the identified station locations. This resulted in a graph with excessive connections compared to a realistic metro map, so my first effort was to limit these connections. My improvement in this regard is using a minimum spanning tree (MST) on the Gabriel graph. This results in a much sparser network without the high number of redundancies in the Gabriel graph.

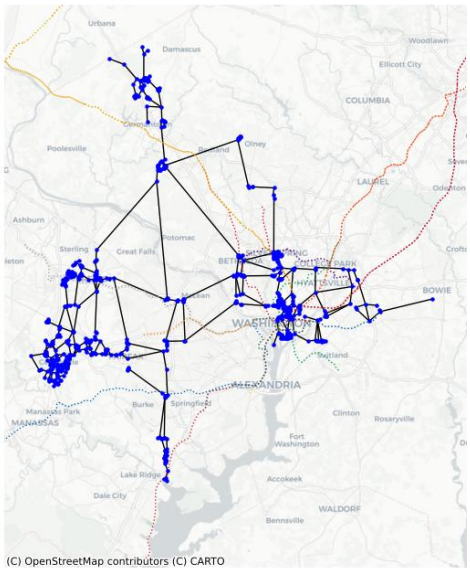

Figure 2. Metro network on the expanded dataset, with connections limited to minimum spanning tree.

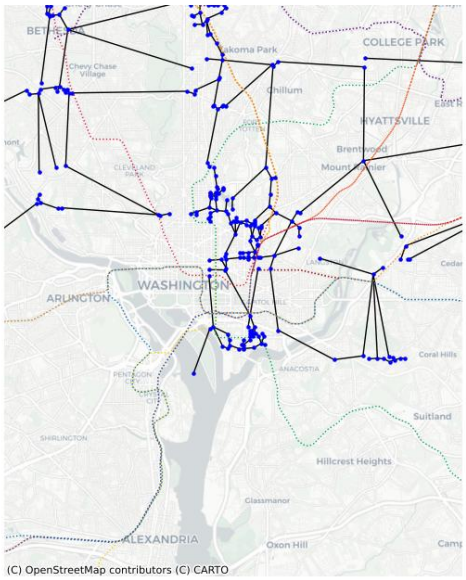

Figure 3. Same data as above, zoomed in on DC urban core.

My next improvement addresses the high number of clusters still visible in the minimum spanning tree. I use the Louvain community detection algorithm to detect clusters and then merge them, creating a much more efficient network.

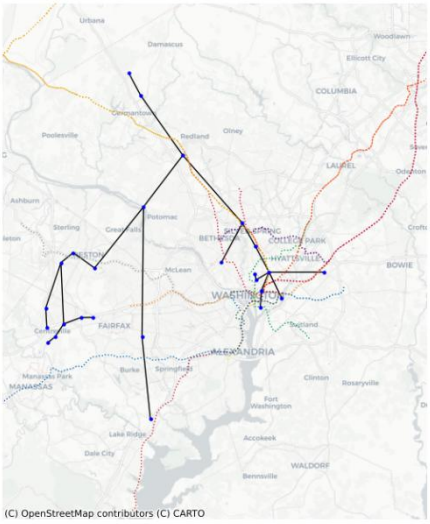

Figure 4. Minimum spanning tree network with Louvain community merging applied.

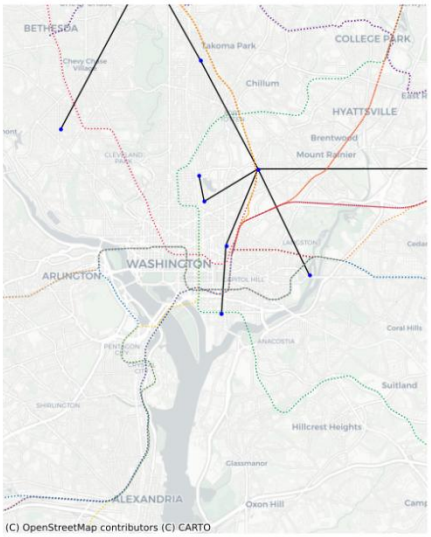

Figure 5. Same data as above, zoomed in on DC urban core.

As I iron out the evaluation metric more, I have created a visual, qualitative evaluation criterion here by overlaying my network creation over an urban map and adding the shapes of the current metro network.

So far, the network does connect some areas that I would consider high density together, and even provides a link I have long considered a critical absence in the WMATA network: a connection between Montgomery County, MD and Fairfax County, VA.

Although in my initial project planning documents I stated that I didn't want to place too much importance on recreating existing transit networks, I envision my transit network being roughly a superset of the existing network. However, there are many sections of the real-world network that my algorithm has not successfully recreated. Furthermore, the network still fails to connect the underserved areas I mentioned in the project introduction: Georgetown and Anacostia.

I'm also thinking that the census block data is a bit too granular, at least given the algorithm I'm currently using. The algorithm mostly does a good job of identifying areas in need of public transit, but I can think of two main problems. First, there are large clusters of stations in dense areas, indicating that the algorithm does a poor job of spacing stations out, which is what the partitioning was intended to do. I am considering remedying this by using a point elimination algorithm to merge points by closest edge distances.

The second is there are a lot of points on outlier census blocks, which are in some cases not indicative of transit need. For example, there are some very small blocks that have a very high calculated population density, even if they only have 50 people in them. I am thinking of remedying this by considering looking at census tract and block group data as well.