# Road Network Analysis in Pennsylvania

Introduction

The field of road network analysis is crucial in city management. Not only it shows the effectiveness of transportation, by analyzing the trend of road design, it could tell us how the city is laid out, the population distribution, as well as the geographical characteristic of the city. But today, by analyzing the road network in Pennsylvania, we are going to answer the question of “how is the road network of Pennsylvania structed in terms of connectivity, and what does this reveal about the state’s transportation infrastructure?” By analyzing the joint and roads, we also want to gain insight of the transportation structure to understand the overall efficiency of Pennsylvania’s road design and characteristic, to better understand the city.

The data set we are using is a Pennsylvania road network dataset that uses nodes to represent intersections and endpoints and uses undirected edges to represent roads that are connecting the intersections and endpoints. The datasets have 1088092 nodes and 1541898 edges in total. At first glance, there are 500,000 more edges than roads meaning averagely every node is connecting to 1.42 edges leading them to other nodes. Average of 1.42 is obviously not enough for a working transportation system, but certainly this is not the case, some nodes would have more than 6 edges connected while some would have only one. And using rust, we are now going to investigate in the data that are more meaningful and comprehensive than the mean.

Design and functions

The first function I used is to count the connected Components. It uses ‘connected components’ function form petgraph, it calculates the number of connected components, a subgraph in which any two nodes are connected by roads. The total number of connected components shows the segmentation of the road network. A big number means a very fragmentized network with many isolated nodes and roads, while a small number means a very centralized network.

Result: 206. A small number of 206 means a centralized network with 206 clusters of roads. A total of a million nodes is concluded in 206 of groups show a mostly interconnected structure with high efficiency.

The function finds the nodes more than four connection simply gives you the total number nodes that connect to more than four edges. We want this data simply to show how many multi-ways nodes exist, and it shows some aspect of the effectiveness of network. \

Result: 809035

A large number shows most of the nodes are connected, and not only that, they have a large number of neighbors. Shows high connectivity.

The function distance criteria find the total number of nodes that is 1 distance or 2 distance away from the initial node. and at the same time, the number of these nodes are at least 5000 greater than then initial node. We assume that the number is assign to each nodes base on their geographical location. For example, at the beginning of the dataset we can see 0 is connected to 1, and 2 is connected to 3. We also assume the number is assigned continuously on a straight line. For example, 0 is connected to 1 and 1 is connected to 2, but 0 is not connected to 2. So we make an assumption that a road is effective and can be leading to many ways if it connects directly to the nodes that have a great difference in their assigned number. If it returns great number, it shows many roads fulfills the criteria showing a effective network.

Result: 1connection of 150218 and 2 connections of 597174. The number is not bad at all. Meaning a great number of nodes could be reach to not only the next nodes linearly, but also to a handful difference place in different direction. The efficiency is shown.

The shortest path analysis shows the nodes that a starting node could get to under 3 edges. We need to enter an initial node number, and the returns will give you all the nodes that it can get to under 3 edges.

The example nodes 456 shows a reasonable connectivity under a distance of 3. Not only it can be connected to the neighbor like 455 and 460, it can also be connected to 11321, that makes the whole network fluent and interconnected.

The degree distribution function is a more comprehensive function than the ‘find the nodes more than four connections’, for different degree of connections, it gives you the total number of nodes that have the degree. For example, degree of 3 means the nodes that have 3 edges connecting to it. It’s a straightforward data that shows the patterns and overall structure of the network.

The result shows a great diversity of distribution. We can see 80 nodes of degree 14 and surprisingly great number of nodes of degree 5 and 8. Which shows great connectivity and efficiency.

Outputs:

Total number of nodes: 1088092

Total number of nodes with more than four connections: 809035

Number of nodes meeting the distance-1 connection criteria: 150218

Number of nodes meeting the distance-2 connection criteria: 597174

Number of connected components: 206

Degree Distribution:

Degree 2: 188317 nodes

Degree 14: 80 nodes

Degree 12: 1237 nodes

Degree 18: 4 nodes

Degree 6: 532686 nodes

Degree 4: 90740 nodes

Degree 10: 7759 nodes

Degree 16: 13 nodes

Degree 8: 267256 nodes

Distance from node 456 to node 455 is 1

Distance from node 456 to node 463 is 3

Distance from node 456 to node 195 is 2

Distance from node 456 to node 460 is 1

Distance from node 456 to node 465 is 2

Distance from node 456 to node 459 is 1

Distance from node 456 to node 462 is 3

Distance from node 456 to node 457 is 2

Distance from node 456 to node 456 is 0

Distance from node 456 to node 464 is 3

Distance from node 456 to node 468 is 3

Distance from node 456 to node 458 is 2

Distance from node 456 to node 11321 is 3

Distance from node 456 to node 461 is 3

Distance from node 456 to node 188 is 3

Conclusion

The analysis reveals that the road net work of Pennsylvania is highly interconnected with a significant number of intersections and hubs. The presence of many nodes connected to distant nodes suggests and extensive and integrated network. Although it’s just an investigation, we can use the data for further design such as city design and population analysis.