A static and dynamic analysis of the Anaheim Road Transportation Network

Rajarshi Basak

December 2018

1 Introduction

In large cities or urban conglomerations, where movement (of people or goods) by foot is not a viable option, vehicular movement is the way forward. With the advancement of technology, vehicles are now easier to afford and maintain, and are the lifelines for people who commute and transportation of heavy goods alike.

A road transportation network consists of junctions represented by vertices and roads or streets connecting the junctions represented by edges between the vertices. Essentially it is a spatial network, since the vertices and edges associated with it are spatial elements which are linked to geometric objects (which are the vertices of the network).

2 Motivation

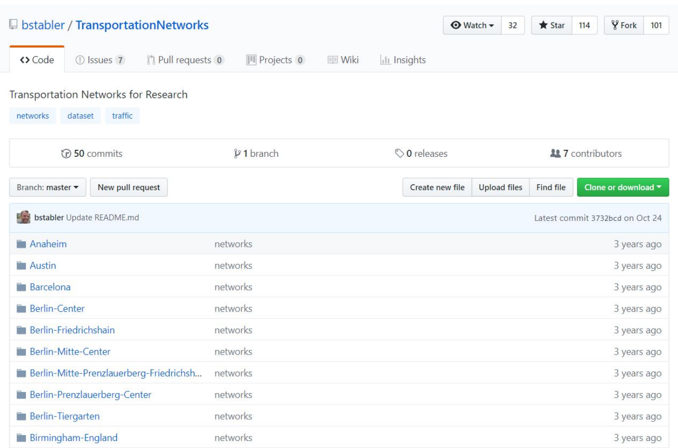
The motivation of this project is to employ some of the interesting tools and metrics of network science to analyze an important road transportation network, and employ these analyses in an attempt to improve the network itself. The importance of nodes and edges (using centrality measures), and the structure of the network (by virtue of quantities like clustering coefficient and diameter) are important and interesting to note. Enhancements in the network include, but are not limited to

- Identifying the importance of junctions and roads in the network in context of traffic flow
- Pinpointing bottlenecks in the network in context of heavy traffic flow

3 The dataset

After some research on the *Index of Complex Networks* website (at https://icon.colorado.edu/), the data set for this project was sourced from Ben Stabler's gitub page (at https://github.com/bstabler/TransportationNetworks). Ben Stabler is developer of transportation systems modeling software in Oregon, and his repository in Github titled *Transportation Networks* has a large number of data sets for road transportation networks around the world, covering cities and urban areas in the USA, Canada, Germany, Spain, United Kingdom, etc.

(a) Ben Stabler, Github



(b) Transportation Networks, Github

Figure 1: Screenshots of Ben Stabler's page on Github showing his profile picture, and his *Transportation Networks* repository also on Github

3.1 The Anaheim Network

From the list of available road networks at Ben Stabler's Github repository on *Transportation Networks*, the Anaheim road network was chosen to conduct analyses. The network data was provided by *Jeff Ban* and *Ray Jayakrishnan*.

Listed in the Anaheim folder are three .tntp files, *Anaheim-flow.tntp*, *Anaheim-net.tntp* and *Anaheim-trips.tntp*. The first two files have edge data for the network - each line on both of these files have an edge represented by the start and end nodes, and attributes such as 'volume', 'cost' on the *Anaheim-flow.tntp* file, while others such as 'power', 'speed', 'free flow time' in the *Anaheim-net.tntp* file.

The plot of the Anaheim network shown below is from 1992, and it consists of 416 nodes (junctions), 914 edges (roads) and 38 zones. A spectral plot if the network (using a built-in function from networkx) was constructed to generate the plot below, which places nodes of comparable eigenvector centralities together.

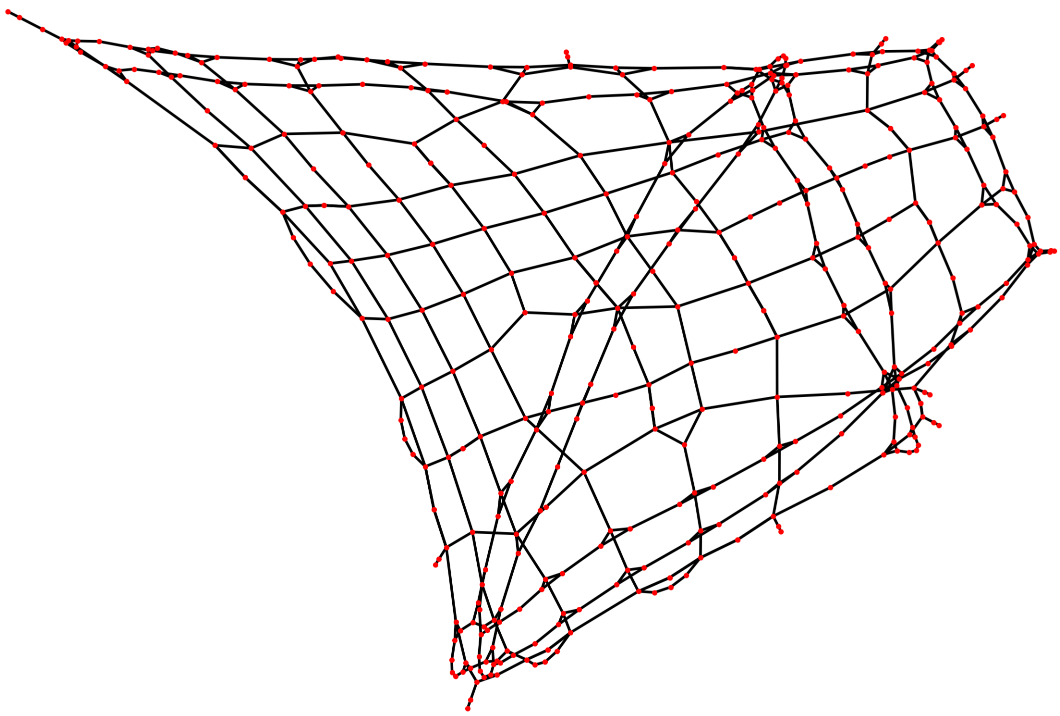


Figure 2: Schematic of the Anaheim Network from 1992 with nodes of comparable eigenvector centralities placed close to each other

4 Static Analysis

The first set of analyses involve the network as it is presented to us. In other words, here we explore certain parameters in the network to give us deeper insights on its structure and function.

4.1 Measures of Clustering, Degree and Diameter

The table below lists the values of some of the measures for clustering, degree and diameter for the network.

Attribute	Value
Average Degree	3.048
Average Neighbor Degree	3.36
Number of triangles	162
Clustering coefficient	0.108
Number of squares	114
Diameter	26
Degree Assortativity	-0.049

While average degree of the network is 3.048, the average neighbor degree is 3.36, which shows that the network (although not a social network) exhibits the friendship paradox!

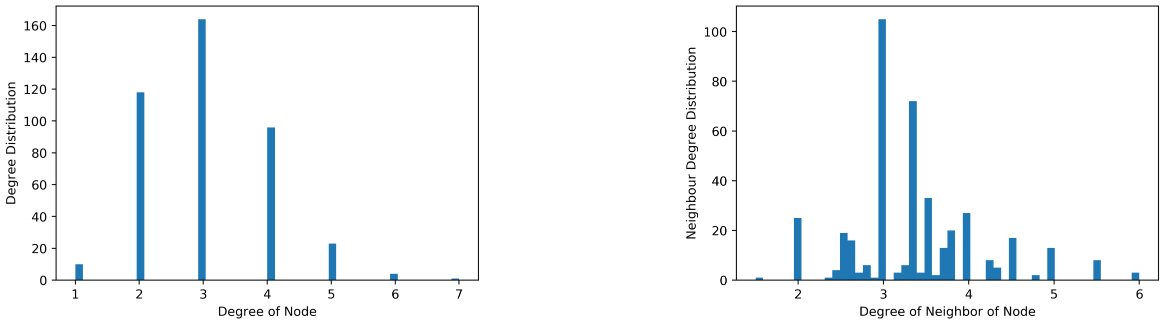
There are 162 triangles in the network, and these can be attributed to highway entries and exits, which are mostly in the shape of triangles, or simply triangular loops of roads. These also contribute to the clustering coefficient, which stands at a modest 0.108.

What is more interesting to note is the number of squares, which has a count slightly below the number of triangles at 114. The mesh/grid like structure of the network accounts for this.

The diameter of the network is 26, which does not scale logarithmically with the number of nodes in the network. The property of logarithmic scaling for the diameter is typically for social networks, and hence it is something we do not expect to observe here.

Finally, there is no assortative mixing by degree in the network, as indicated by the value of the degree assortativity, which actually has a negative value but is very close to 0.

The following plots show the degree distribution and the average degree distribution of the nodes.



(a) Degree distribution

(b) Neighbor degree distribution

Figure 3: Plots showing the degree and neighbour degree distribution for the Anaheim network

4.2 Centralities of Nodes

It is interesting first to identify the most important nodes in the graph. Centrality measure is the metric to investigate, and the different types of centrality, when evaluated for both the unweighted and the weighted graphs, throw light on the key infrastructure nodes in the transportation network.

The **degree centrality** of a node is simply the degree of node, or the number of edges a node has in the graph.

The **betweenness centrality** B_C of a node i in a graph is given by

$$B_C(i) = \sum_{s \neq v \neq t \epsilon V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

The **closeness centrality** of a node is the average length of the shortest path between the node and all other nodes in the graph. For a node i it is given by the relation

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

where $d(j,i)$ is the distance between the vertices x and y. The definition of harmonic centrality reverses the sum and reciprocal operations from the definition of closeness centrality. It is given for a node i by

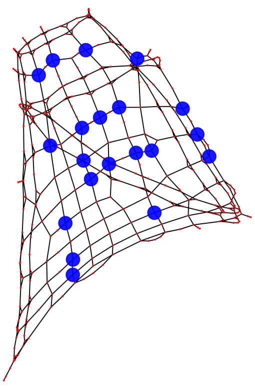
$$H_C(i) = \sum_{j \neq i} \frac{1}{d(j,i)}$$

Eigenvector centrality assigns relative scores to all vertices in a network given the fact that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low scoring nodes.

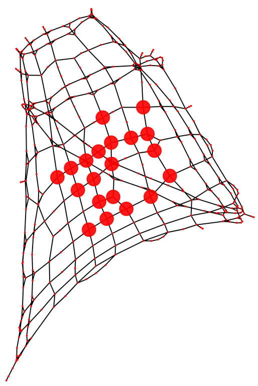
The table below shows the top 20 nodes for each of the four different centrality measures described above for the un-weighted network.

Degree	Harmonic	Betweenness	Eigenvector
330	319	358	317
303	321	321	329
337	317	305	328
299	330	299	330
317	356	333	343
266	320	266	342
267	318	277	355
269	329	315	316
273	333	357	354
302	355	384	371
304	316	356	372
308	344	319	356
341	357	390	299
329	358	385	370
332	343	386	341
333	328	355	315
361	303	388	327
369	315	327	387
385	305	317	388
373	354	375	273

There seem to be some nodes which appear in the top 20 list for almost all the centrality measures, such as 317, 330, 355, 333, and 329.

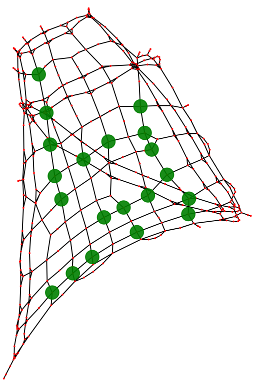


(a) Degree

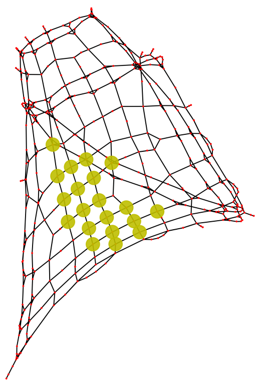


(b) Harmonic

Figure 4: Spectral plots of the Anaheim network showing the top 20 nodes according to degree centrality (left) and the top 20 nodes according to harmonic centrality (right) in the unweighted network



(a) Betweenness



(b) Eigenvector

Figure 5: Spectral plots of the Anaheim network showing the top 20 nodes according to betweenness centrality (left) and the top 20 nodes according to eigenvector centrality (right) in the unweighted network

The first plot shows the nodes with the highest degree centrality. Since most of the network has a grid like structure, most of the nodes have a degree of 4, with a few of degree 5, a handful having 6 and one node with degree 7. The node with degree 7, the nodes with 6, and few nodes with degree 5 have been shown. The top 20 nodes according to harmonic centrality are clustered towards the center of the network. Since the harmonic centrality is a modification of closeness centrality, where nodes are ranked in inverse order for farness (i.e. the more central a node is, the closer it is to all other nodes), the placement of the nodes in the plot makes sense. A similar distribution is observed when nodes are ranked according to eigenvector centrality - though a slightly different set of nodes. Finally the plot on the bottom right shows the top 20 nodes with the highest betweenness, and it is clear that while all of these nodes are not visibly in the center of the network (some are), none of them are at the visible periphery - they are all situated in rings that are somewhat in the center as they should be, since they act as a bridge for the shortest path between two other nodes.

For the weighted network, the following table shows the top 20 nodes for the different centrality measures.

Degree	Harmonic	Betweenness	Eigenvector
303	20	358	317
337	62	321	329
330	397	305	328
266	413	299	330
267	14	333	343
269	23	266	342
273	414	277	355
302	21	315	316
304	15	357	354
308	8	384	371
341	2	356	372
329	412	319	356
332	22	390	299
333	5	385	370
361	19	386	341
369	380	355	315
385	257	388	327
373	416	327	387
389	415	317	388
378	254	375	273

While the betweenness and eigenvector centralities remain unchanged from the un-weighted network, the harmonic centralities change drastically to the extent that there is no overlap between these and any of the other centralities for the weighted network.

The plot showing the top 20 nodes with the highest harmonic centralities (below) shows that most of the nodes are on or near the top and bottom horizontal highways.

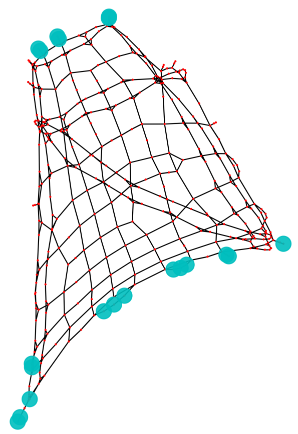
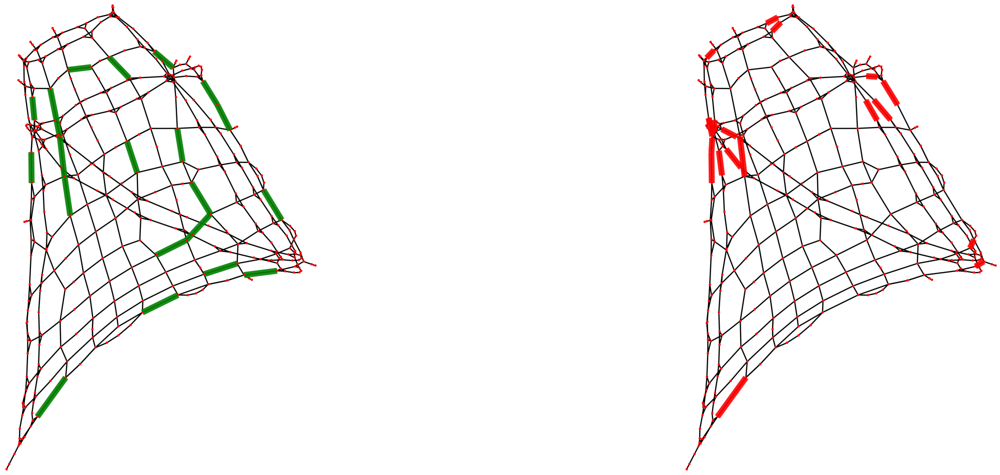


Figure 6: Spectral plot of the Anaheim network showing the top 20 nodes according to harmonic centrality in the weighted network

4.3 Betweenness centrality of edges

The betweenness centrality of an edge is similar to that for a vertex in a network. Instead of checking for the existence of a vertex in the shortest path, we instead check for the existence of the edge under question.

To evaluate the betweenness centrality of the edges in the network, the 'edge-current-flow-betweenness-centrality' module from networkx was used. This method of edge betweenness centrality calculation draws an analogy from electric circuits. In an electric circuit, there are junctions connected by wires or electrical components, and a finite current flows along these wires, which have a finite resistance. If we were to visualize the electric circuit as a network, the junctions can be thought of as vertices, and the wires as edges. Since current flow is inversely proportional to resistance, the greater the resistance of an 'edge', the smaller the current that will flow through that 'edge'. Similarly, in a transportation network, the greater the 'cost' of an edge (or a road connecting two junctions), the 'lesser' the traffic that should flow through that 'edge', or the lesser its contribution to the shortest path between any two node pairs.



(a) Un-weighted

(b) Weighted

Figure 7: The top 20 edges with the highest betweenness (according to the current flow model) in the un-weighted (edges in green, left) and the weighted (edges in red, right) network

The schematic above shows the top 20 edges with the highest betweenness in both the un-weighted and the weighted network. The edges in green on the left are for the un-weighted network, while those in red on the right are for the weighted network.

It is unclear as to why the edges (marked in green) actually have the highest betweenness in the un-weighted network. The only inference that can be made is that most of these edges lie not on the periphery, but more to the center of the network. Hence they should have higher betweennesses than other edges that lie on the boundaries of the network.

However, for the weighted network, there are some interesting obserevations, apart from the fact the edges with the highest betweenness centralities are different from those in the un-weighted network. The network has several highways in it, which can be identified from the triangle shaped exits located periodically on the junctions of the highway:

- Two running horizontally from one corner to another on the top and bottom
- One running horizontally somewhat in the top center of network, parallel to the previous two
- One running vertically from top left corner to bottom left corner
- One running vertically from top right corner to bottom right corner with a short horizontal path in between
- One running diagonally from bottom right corner to mid-left center of the network

Close observation reveals that these edges (with the highest centralities) are located near the entry and exit points of the above-mentioned highways. This implies that highways play a crucial role in traffic flow when we consider the weights, as a large chunk of the traffic flows along the highways.

4.4 Shortest Paths between junctions

Since the network has a geometric structure, an useful property to investigate is the shortest path between node pairs.

The shortest path between a start and end node is, as the name suggests, the minimum number of hops (edges) required to be traversed to reach the end node from the start node. In an unweighted network, the shortest path will always be the minimum number of edges (and the minimum number of nodes) one needs to traverse to reach the end node from the start, while in a weighted network (where the edges have weights), the shortest path between the start and end node is that path which incurs the least cost of traversing due to the weights of the edges.

Shown below are the histograms for the length of the shortest paths in the unweighted and the weighted network. For the weighted network, the *Anaheim-flow.txt* file was used, and the 'cost' attribute was the weight of the edge that was taken into account.

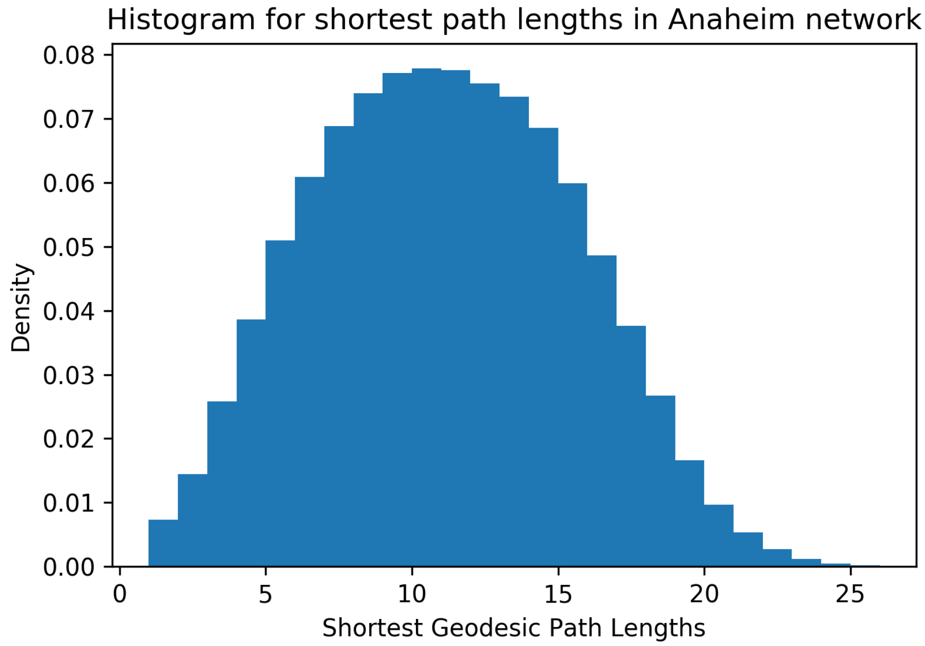


Figure 8: A histogram of shortest path lengths in the unweighted Anaheim network

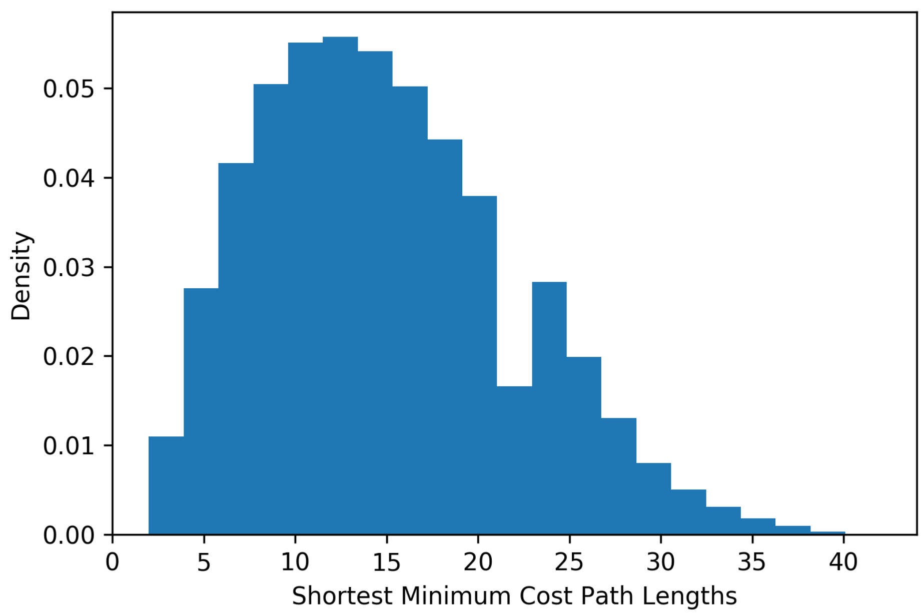


Figure 9: A Histogram of shortest path length in the weighted Anaheim network, with 'costs' as the edge weights

Since there are n nodes in the graph, there are $n^2 - n$ pairs of nodes (not including pairs of nodes with the same node), and hence $n^2 - n$ shortest path to be calculated. This number turns out to be 172640 for $n = 416$.

For the unweighted network, the range of shortest path lengths vary from 1 to 26, with the highest fraction being of length = 11, and the fraction being close to 0.08. The histogram also shows that the path lengths are almost evenly distributed on both sides of this central value which accounts for the largest number of path lengths.

Things change slightly when we switch to the weighted network. The shortest path lengths can be as large as 42 this time, but the highest fraction is is still for path length = 11 - only the fraction goes down to account for the other values of the path lengths. The distribution is also more uneven around the central value now, with more path lengths having larger values than 11 compared to those having values lesser than 11.

This suggests that the shortest paths between two node pairs is not always the same (path, or of the same length) when we consider the unweighted and the weighted graph. When we take into account the 'weights' of the edges, the shortest paths are sometimes larger than what they are when we do not consider the weights.

4.5 Equality of shortest paths

A natural question to ask is how many of these shortest paths are identical for both the weighted and the unweighted network. The metric to explore is the fraction of shortest paths that are equal for pairs of nodes that are separated by a certain distance, or number of hops.

The algorithm for this analysis is as follows:

```
Load both the graphs (unweighted and weighted)
Iterate through all pairs of nodes in the unweighted graph
    Find shortest path using Dijkstra's algorithm
Iterate through all pairs of nodes in the weighted graph
    Find shortest path using Dijkstra's algorithm
Iterate through the shortest paths between node pairs for unweighted and weighted graph
    Count paths that are equal for a given node pair
    Keep count of the separation between the node

Plot the fraction of node pairs for which the shortest paths are equal
as a function of the separation (or geodesic path length) between nodes
```

A large fraction (possibly the largest) of node pairs are expected to have identical shortest paths when the separation between them (geodesic distance) is only 1, and the fraction should decrease when the distance between the nodes in the network grow. What we see, however, is in slight contrast to our expectations.

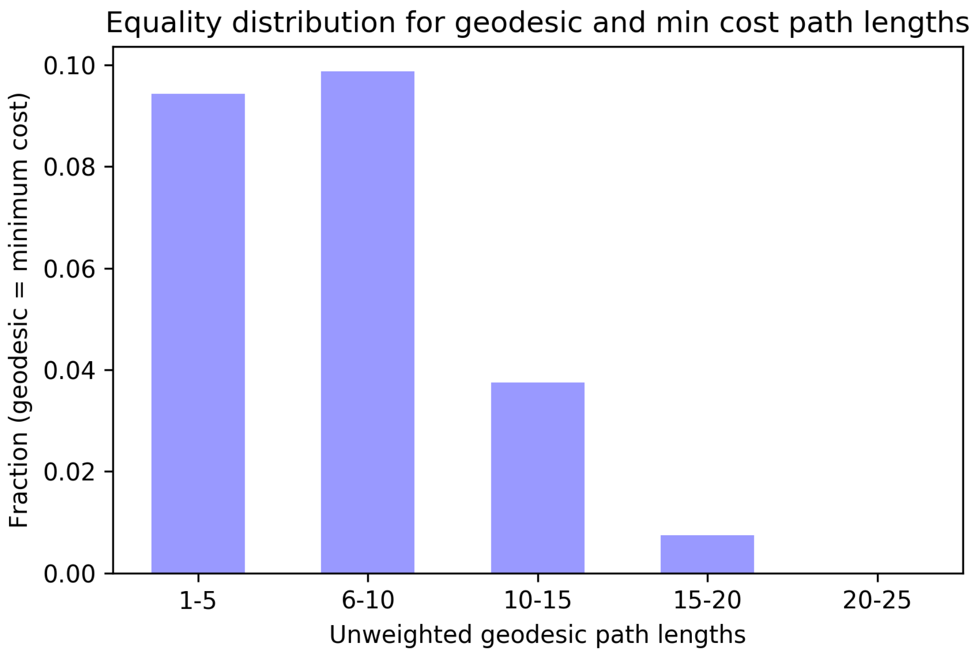


Figure 10: Plot showing the fraction of shortest paths that are identical for node pairs as a function of the separation (or geodesic distance between) the node pairs for bins of length 5

Categorizing the separation of nodes into bins of 1-5, 6-10, and so on, the observed trend is that the maximum fraction is in the 6-10 bin. This implies that the largest fraction (of identical shortest paths) is not for nodes that are separated by a single edge.

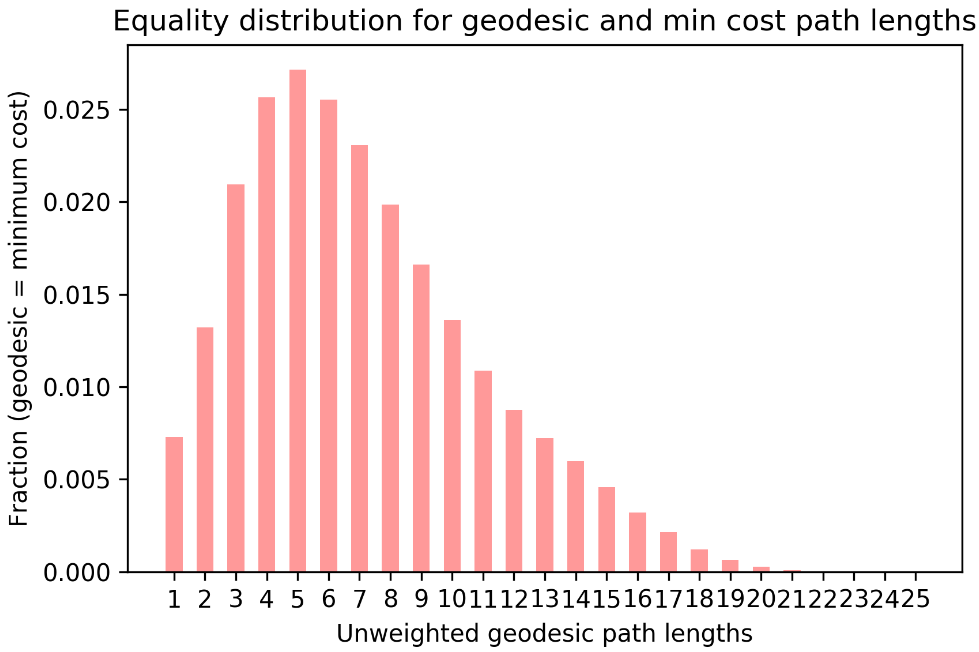
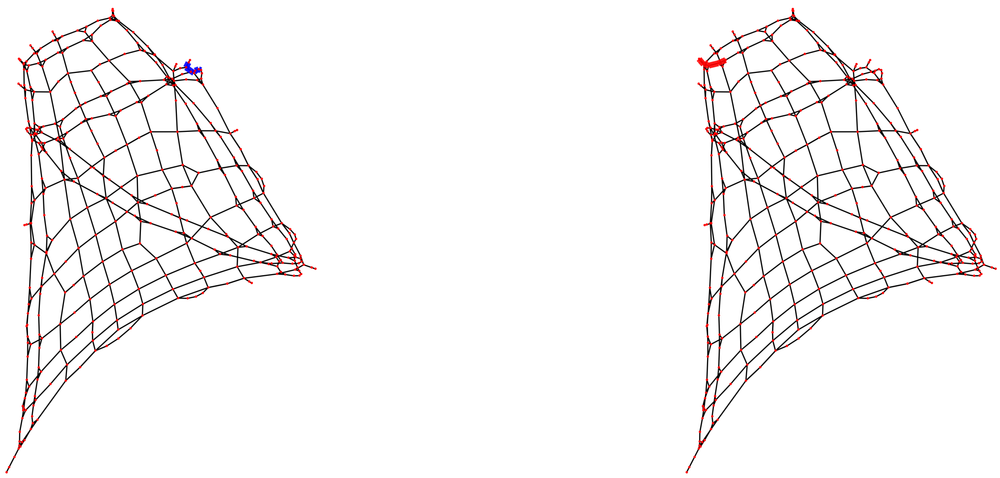


Figure 11: Plot showing the fraction of shortest paths that are identical for node pairs as a function of the separation (or geodesic distance between) the node pairs for individual lengths

Individual length categorization reveals that the fraction equal are for those node pairs that are separated by (or have geodesic distances of) 5 units. The rise in the fraction equal from geodesic path length = 1 to geodesic path length = 5 is steep, and the consequent fall is gradual.

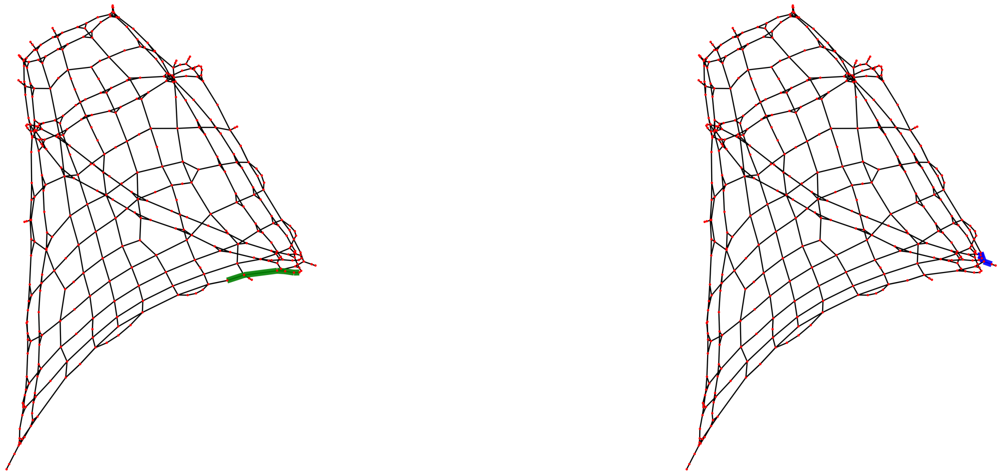
Since path length = 5 has the highest fraction of node pairs which have identical shortest paths, it would be interesting to observe some of these paths in the network.

It is evident from the figures below that these shortest paths of length = 5 appear at the periphery of the network, which is where the entry and exit points of the highways are located. Since the highways have a large traffic flowing through them, it is expected that the shortest paths for node pairs near the entry and exit of highways would not change whether we take the weight of an edge into account or not.



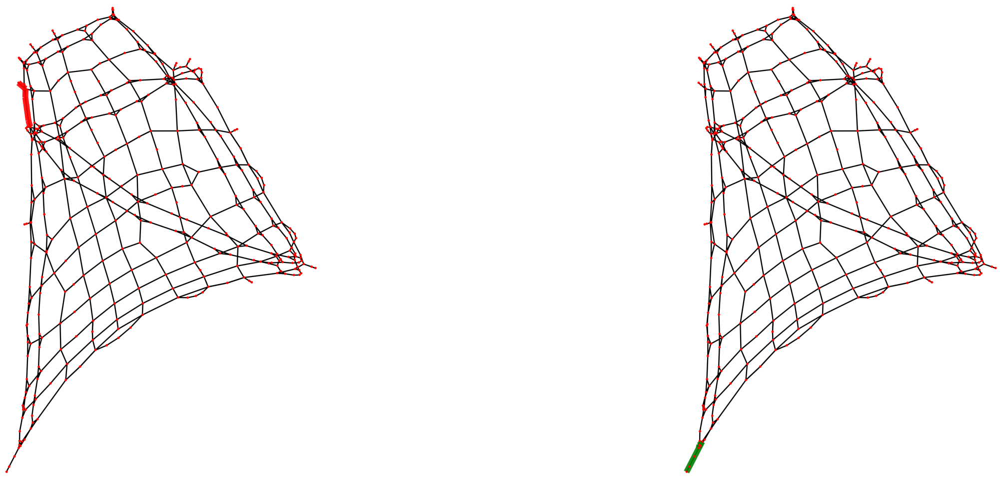
(a) (b)

Figure 12: A few of the Shortest paths of length = 5 shown in the actual network



(a) (b)

Figure 13: A few more of the Shortest paths of length = 5 shown in the actual network



(a) (b)

Figure 14: Even more of the Shortest paths of length = 5 shown in the actual network

5 Dynamic Analysis

In this section, the network was perturbed from its original state, and analyses were made thereafter to compare the network before and after the perturbation.

There are various possible ways of perturbing the network - including (but not restricted to) removing nodes, removing edges, removing both nodes and edges, and altering the values of the attributes of nodes and/or edges. We could also add nodes and/or edges to the network.

Since our data-set consists of attributes associated explicitly with the edges (and not the vertices), it made sense to remove the edges and not the vertices.

5.1 Importance of Edges for geodesic based rerouting

The dynamical importance of an edge (or of the two nodes that the edge connects) is the drop in overall transport speed when we block all traffic along that edge (or between the nodes connected by that edge). This is equivalent to forcing everyone to detour around these two nodes. We leverage the betweenness centrality of edges to explore this.

A few questions that spring to mind are as follows:

- Is betweenness a good predictor of the increase in overall cost from deleting that edge? How can we tell?
- How can we estimate using betweenness centralities for nodes and/or edges, the increase in cost when an edge is removed?
- Can we modify the mathematical estimates for cases when we use the weighted ('cost') network instead of the un-weighted ('adjacency') network with no weights?

Here we try to answer some of these questions.

Let the betweenness centrality of an edge $(i \rightarrow j) = C_B(ij)$, that is

$$C_B(ij) = \sum_{s,t \epsilon V} \frac{\sigma(s,t|i \rightarrow j)}{\sigma(s,t)} \tag{1}$$

where $\sigma(s,t)$ represents the shortest path between nodes s and t in the network, and $\sigma(s,t|i \rightarrow j)$ represents the shortest path between the nodes s and t that include the edge (i $\rightarrow j$).

Then the number of traffic paths that would have to be rerouted if that edge were to be removed is given by

$$\sum_{s,t \epsilon V} \sigma(s,t|i \rightarrow j) \tag{2}$$

$$=> C_B(ij) \sum_{s,t \epsilon V} \sigma(s,t) \tag{3}$$

Hence higher the value of C_{Bij}, the greater the number of paths that are required to be rerouted.

When we remove an edge the overall cost in the network can either stay the same (if all the shortest paths in the network are independent of that edge, or the shortest path lengths do not increase since rerouting does not involve increasing the number of edges - it only replaces the removed edge with another edge, when the network has multi-edges) or it can increase. This is because by definition, the shortest path between two nodes are such that there cannot be any shorter path. This suggests that the edge betweenness centrality should be a 'good' predictor.

Now we define the initial overall cost as the ratio of the sum of all shortest path lengths in the original network to the number of unique node pairs in the network. Mathematically, it is given by

$$\frac{\sum_{s \neq t \epsilon V} d(s,t)}{n(n-1)} \tag{4}$$

The final overall cost is defined as the ratio of the sum of all shortest path lengths in the network after removing the edge to the number of unique node pairs in the network, which is denoted by

$$\frac{\sum_{s \neq t \epsilon V} d'(s,t)}{n(n-1)} \tag{5}$$

Hence the change in overall cost is given by

$$\frac{1}{n(n-1)} \Big[\sum_{s \neq t \epsilon V} d(s,t) - \sum_{s \neq t \epsilon V} d'(s,t) \Big] \tag{6}$$

$$=> \frac{1}{n(n-1)} \sum_{s \neq t \epsilon V} [d(s,t) - d'(s,t)] \tag{7}$$

Since the shortest paths between nodes that do not involve the edge $i \rightarrow j$ do not change when the edge $i \rightarrow j$ is removed, the above equation can be written as

$$=> \frac{1}{n(n-1)} \sum_{s \neq t \epsilon V} [d(s,t|i \rightarrow j) - d'(s,t|i \rightarrow j)] \tag{8}$$

Here $d(s,t|i \rightarrow j)$ and $d'(s,t|i \rightarrow j)$ represent the shortest path lengths between nodes s and t before and after removal of the edge $i \rightarrow j$ respectively.

The quantity $\sum_{s \neq t \epsilon V} [d(s,t|i \rightarrow j) - d'(s,t|i \rightarrow j)]$ is the change in length of all shortest paths in the network between what they were initially and what they are after removal of the edge.

An estimate of the this quantity is given by the value of the betweenness centrality of the edge. The change in overall shortest path length of the network can be plotted as a function of the betweenness centrality of the removed edge for a few edges, and curve can be extrapolated to find the change in overall cost of the shortest path length given the betweenness centrality of the the edge under question can be read off from the x-axis.

Another metric that can be used to estimate the quantity under question is the clustering coefficient of the network. Since the clustering coefficient is the ratio of the number of closed paths of length 2 to the number of paths of length 2, which essentially reveals the density of triangles in the network, and is a value between 0 and 1 for the network. If the clustering coefficient of the network is 1, then there are as many triangles as there paths of length 2, and hence removing an edge will only increase the overall change in cost by an amount proportional to the number of shortest paths that contains the removed edge times 1. This is because for each of those shortest paths, there is a two node detour for the edge removed (since there has to be a triangle for every node pair). If the clustering coefficient is a small number, then a large detour is required when an edge is removed - Hence the greater the clustering coefficient, the lesser the change in overall shortest path length in the network when an edge is removed.

Finally, we can also look at the diameter of the network. The more 'small world' the network, the more the existence of random connections in the network, and hence the higher the chance that the removal of an edge will not affect the overall shortest path length in the network by a substantial amount.

To investigate the effect of removing edges from the network, the following protocol was adopted.

Algorithm

```
Find sum of all shortest paths in the network

For each edge in the network
    Remove it from the network
    For all pairs of nodes in the network
        Fetch the shortest path between the node pair
    Find sum of all shortest paths after edge removal
    Find difference in sum of all shortest paths before and after removal
    Fetch the betweenness centrality of the removed edge
    Store betweenness centrality and 'difference in sum of all shor-
    test paths before and after removal' for plotting later for edge
    Add removed edge to the network

Plot the 'difference in sum of all shortest paths before and after rem-
oval of edge' as a function of the betweenness centrality of the edge
```

Since we were calculating the shortest paths for all pairs of nodes after removing each edge from the network, the algorithm initially consumed more time than we could let it afford, and minor modification was made later to cut down on run-time. If the removed edge was not there in the shortest path, then the shortest path need not be recalculated after removal of the edge. However, if it was there, then removing it will change the shortest path and hence the shortest path length, and hence it needs to be recalculated.

Both the un-weighted and the weighted network was analyzed using the above-mentioned protocol.

We expected to see a strong correlation between the change in overall shortest path length due to the removal of an edge with the betweenness centrality of that edge. In other words, the higher the betweenness centrality of the removed edge, the more it should affect the transportation speeds (and hence cost) in the network - that is the higher the change in overall shortest path lengths in the network.

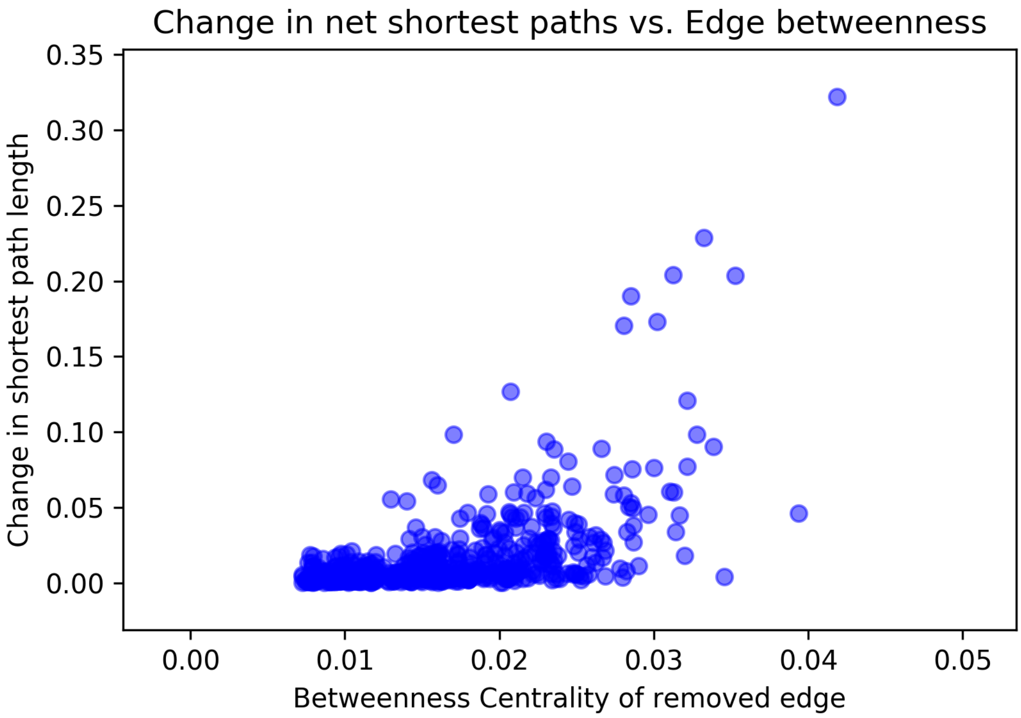


Figure 15: Plot showing the Change in overall shortest paths (per node pair) in the network as a function of the Betweenness centrality of the edge removed in the un-weighted network

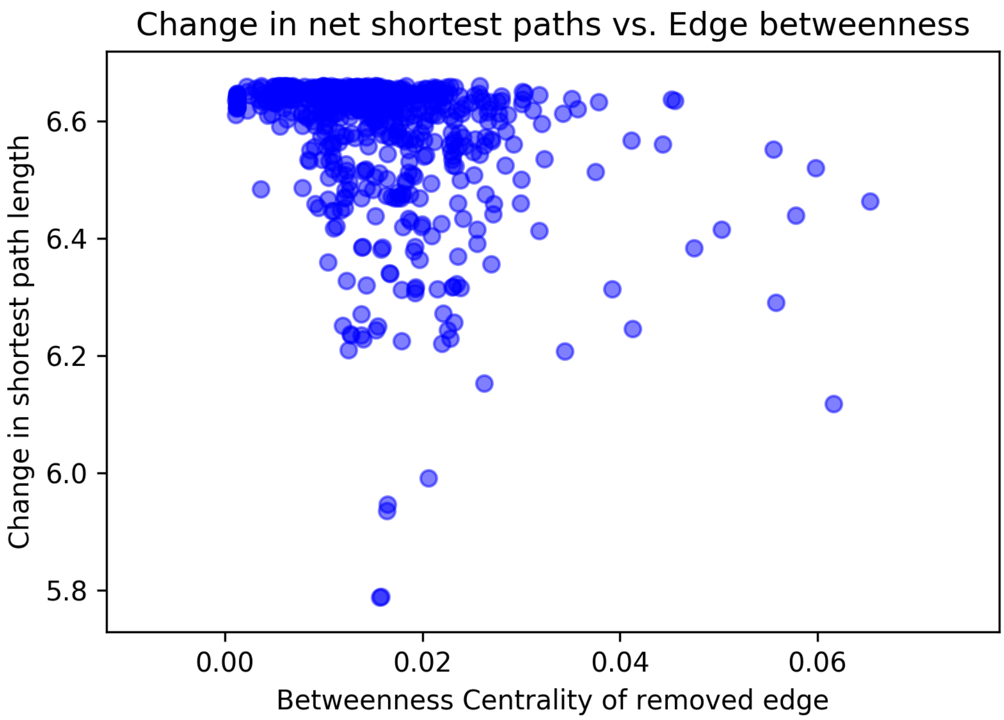


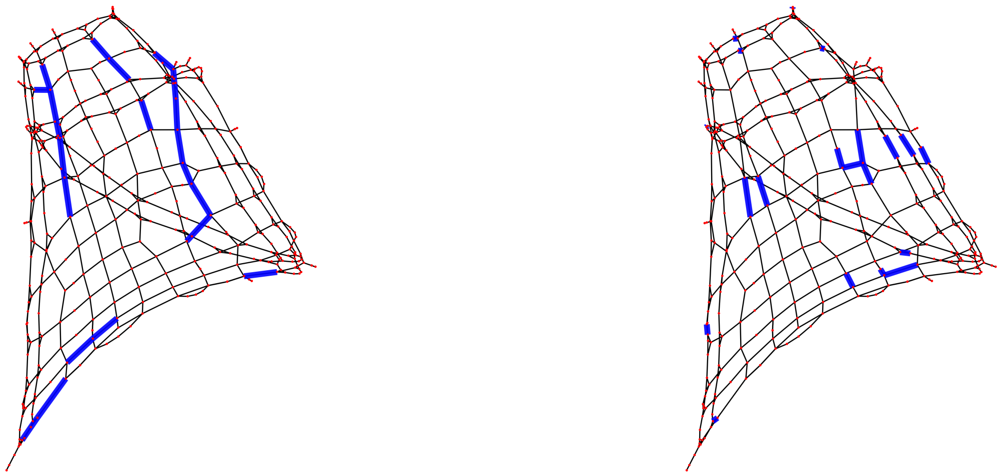
Figure 16: Plot showing the Change in overall shortest paths (per node pair) in the network as a function of the Betweenness centrality of the edge removed in the weighted network

It was observed (from the first plot above) that in the network without weights, there is a (weak) positive correlation between change in overall shortest path length with the betweenness centrality of the removed edge. The overall cost in shortest path per node pair in the original network was 10.63. Most edges did not contribute drastically to the change in overall shortest path, even with moderately high values of betweenness, as is observed from the large chunk located towards the bottom left of the graph. There are a few points that are located slightly higher in the graph, and slightly more to the right, implying that they have a modest contribution to the change in overall shortest path.

Finally, only a few edges contribute linearly - and these are represented by the points placed on the straight line running at a 45 degree angle in the plot.

The correlation between the two quantities of interest in more interesting when we consider the weights, though. There is a positive correlation, like before, shown by the linear set of points in the graph. However, there are a large number of points clustered toward the top right and center, indicating that most edges (with low or moderate betweenness) contribute to a large extent to the change in overall shortest path length.

5.2 Correlation of dynamical importance with betweenness



(a) Un-weighted

(b) Weighted

Figure 17: The top 20 edges with the highest change in net shortest path length per node pair in the un-weighted (left) and the weighted (right) network

It is evident from the figure above that there is a clear difference in the edges that contribute to highest change in overall shortest path length of the network on their removal between the un-weighted and the weighted network. In the case for the weighted network, the edges seem to be grouped together, suggesting zones for high traffic flow. Some of these zones are on highways, while others are not.

On comparison with Figure 6, there is some overlap between edges for highest betweenness centrality and highest change in net shortest path lengths when the edge is removed for the un-weighted network, and no such overlap in the weighted network. However, for the weighted network, the edges for both the highest betweenness centrality and highest change in net shortest path when the edge is removed seem to be (a) located near each other, and (b)located in certain zones in the network. The latter indicates higher traffic flow in these 'zones' in the network.

6 Results and Conclusions

From the static and dynamic analyses, several conclusions can be drawn about the structure and function of the network.

- It has a grid like structure representing blocks separated by roads, and highways on the borders (including one running diagonally).
- The degree distribution displays a slight amount of 'friendship paradox', usually seen in social networks.
- Most nodes have a degree of 4, with a few having degree 5 or 6.
- The un-weighted network has some vertices which show up in top 20 rankings for different centrality measures. The weighted network does not have any such vertices.
- The maximum shortest path length is 26 (hops) in the un-weighted network, and 42 (hops) in the weighted network.
- Investigation of equality of shortest paths between un-weighted and the weighted network reveals an interesting distribution.

- There exists a positive correlation between the change in overall shortest paths and the betweenness centrality of the edge removed, for both the un-weighted and weighted networks. In the un-weighted network, however, most edges do not contribute drastically to the change the overall shortest path, while in the weighted network it is the other way around.

- Apart from betweenness centrality, measures of clustering and diameter can be employed to estimate the change in overall cost on the removal of an edge from the network.

7 References

[1] *Models and algorithms for road network design: a review and some new developments*; Hai Yang Michael G. H. Bell; Pages 257-278; https://doi.org/10.1080/01441649808717016

[2] *A review of urban transportation network design problems*; Reza Zanjirani Farahani, Elnaz Miandoabchi, W. Y. Szeto, Hannaneh Rashidi; https://doi.org/10.1016/j.ejor.2013.01.001

[3] *The Structure of Interurban Traffic: A Weighted Network Analysis*; Andrea De Montis, Marc Barthélemy, Alessandro Chessa, Alessandro Vespignani; https://doi.org/10.1068/b32128

[4] *New applications of the betweenness centrality concept to reliability-driven routing*; Juan Segovia, Eusebi Calle, Pere Vila; http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.212.471rep=rep1type=pdf

8 Notes

- *It is to be noted that one of the reasons for using the Anaheim road network from 1992 instead of any other network is its relative size compared to the other networks. Most other networks had thousands of vertices and tens of thousands of edges, while the number of nodes and edges in the Anaheim network is small enough to run most algorithms in a short period of time, yet large enough to reveal some of the trends that we expect.*

- The code used for the algorithmical analysis and the plots can be found at the authors github repository 'Anaheim-Road-NAM', the link for which is (https://github.com/basakrajarshi/Anaheim-Road-NAM)

9 Acknowledgements

I would like to thank Professor Daniel Larremore for helpful conversations and ideas.

Finally this project would not have been possible without the support of my family (mother, father and brother) and the countless hours of discussions that I had with my father.