ROAD IMPORTANCE IN AND AROUND BOSTON

CARMEN CASTAÑOS, MARGO DAWES

INTRODUCTION AND NETWORK STORY

The Boston Area Road Network is a collection of 9643 nodes with 21905 directed connections among them. Each node represents an intersection found at a specific latitude and longitude and each link represents a road connecting two given intersections. Two-way streets are represented by two directed links while one-way streets have only one link. The edges have attributes for capacity (vehicles per hour), speed limit (miles per hour), and travel time (minutes). These properties can further give an approximation of the length of each street segment by converting the travel time to hours and multiplying it by the speed limit.

When we assemble this network and display it using visualization software (Gephi), we can see the major roads in and surrounding the city of Boston through the lens of the intersections that connect them. In doing so, we hope to intuit road types by the depiction of differently sized intersections and to learn about the most important roads in the metropolitan area.

There are a number of perspectives from which to consider the importance of certain routes, but particularly relevant in this part of the country is snow clearance on prioritized arteries. Major roads are cleared before minor roads, but how is "major" classified? Are these the highest capacity routes or the routes included in the greatest number of commutes or something else entirely? Can we see these routes just by looking at the degree of the intersections? We will attempt to answer some of these questions in this report.

RELEVANT LITERATURE

The Network Analysis of Urban Streets: A Primal Approach explores the notion of some streets being more important than others due to their centrality. Motivated by a dearth of investigation into centrality in urban studies, much of the paper is devoted to discussions of various indices of centrality—closeness, betweenness, straightness, and information.

The first measure of centrality, closeness, is based on the degree of each node and suggests that the most important intersections are simply the ones connected to the most other intersections in the network. However, this measure fails to address the geographical limitations of urban road networks. As the degree of any given intersection is constrained by physical closeness to other intersections, all intersections are likely to have degrees within an order of magnitude of one another and very few are likely to be connected to other intersections a significant geographical distance away. We expect this measure to give artificial importance to some streets in dense urban areas (where walking or taking the train might even be possible alternatives to driving after a snowstorm) while discounting low-degree intersections that connect many suburban individuals to work closer to the city. We also expect it to affect average shortest path length calculations especially in comparison to common network models.

The second measure, betweenness, indicates the number of shortest paths that pass through a certain node. Nodes with high betweenness centralities are thereby critical in the functioning of a number of routes and should likely be prioritized highly during snow removal. This calculation can be skewed by the use of geodesic path lengths versus non-geodesic path lengths (network shortest paths), but regardless of approach many nodes should consistently demonstrate high betweenness.

The third measure, straightness, considers the efficiency of a route (mostly with regard to distance rather than time). By drawing a "virtual straight route" between two nodes, one can then compare the deviation of the actual connection to determine the route's efficiency.

The last measure, information centrality, contemplates the importance of a node's contribution to the network. By considering information propagation with an active node and then again with the node deactivated, one can measure the network's ability to respond to a loss. Specifically, this measure is defined as the relative drop in efficiency caused by the removal of a specific node (or edge). In our case, this measure would manifest as a prediction of the consequences of electing not to remove snow from a particular intersection or road segment.

For all the measures outlined above, it is worth noting that calculating the centrality of an edge rather than a node could give more intuitive information without requiring much in the way of programmatic edits. Plowing a road that has been identified as important, for example, is more obvious (and critical) than plowing an important intersection.

The paper focuses on four cities: Ahmedabad, Venice, Richmond, and Walnut Creek. In the first two, which are significantly less orderly than Richmond (whose grid resembles the road network of New York City), betweenness and information centrality appear to indicate the roads that should be cleared first. For Boston streets (somewhere in between), betweenness alone should be sufficient for identifying "important" streets as more ideal grids exhibit negligible agreement between information and betweenness centrality.

METHODS AND RESULTS

Some basic properties of the Boston Area Road Network are given below.

Average degree: $< k > = 4.53$

Average in- and out-degree: $< k_{in} > = < k_{out} > = 2.26$

Average clustering coefficient: $< C > = 0.149$

Average shortest path: $< l > = 48.69$

THE PROPERTIES

To find the average degree, we used the formula $\frac{2L}{N}$, which considers the degree of a node as the number of adjacent nodes (rather than considering the in-degree or out-degree specifically).

Clustering coefficient algorithms are not defined for directed graphs, so we had to cast our graph onto an undirected but otherwise equivalent graph to find the average clustering. One possible limitation of this function is its directive to seek triangles rather than squares. For any given node, it calculates the number of neighbors that share a common link, but with a road network it might be more appropriate to calculate the number of neighbors separated by only one other node as well.

Finally, the normal average_shortest_path_length function does not work on unconnected graphs like the Boston Area Road Network. To use this function, we found the largest connected component of the graph (i.e. a subgraph that represents enough of the network to allow for the calculation of path length between nodes). This value is higher than we have come to expect for other networks due to the geographical organization of roads. As direct connections between two nodes a great geodesic distance from one another are extremely unlikely, the path from an intersection on one side of Boston to the other necessarily needs to pass through dozens of other nodes. This is discussed further in the next section.

FINDING AN APPROPRIATE MODEL

The best model to describe the Boston Area Road Network is small-world network (Watts-Strogatz model). If our road network were perfect, we could say each node was connected to four other nodes (k=4) with no discrepancies. In a small-world network, nodes organized in a ring or grid are connected to their immediate neighbors with some uniform k. The nodes are then redirected to some other random node in the network with probability p. This mimics the tendency for some intersections in a real network to be connected to other intersections further away. Our conceptual justification is corroborated by the properties we calculated from a Watts-Strogatz model of our network and their similarities to the ones given above. See Table 1 below for a comparison of the small-world network with the Boston road network. For the sake of comparison, we also included properties calculated from an Erdős-Rényi model (random graph) to show the lack of applicability.

Network	Average Degree, <k>	Average Clustering, <C>	Average Shortest Path Length, <l>
Boston Roads	4.53	0.149	48.67
Watts-Strogatz	4.53	0.156	8.24
Erdős-Rényi	4.58	0.0003	6.02

Table 1: Summary table of network properties and modeled equivalents

The average clustering from the random graph is what makes it inapplicable as a model for the Boston Area Road Network as it is several orders of magnitude smaller. Note that the <k> value given in the Watts-Strogatz model is an input parameter rather than a calculated property.

The average shortest path length property is vastly different from those calculated from the models due to the reasons outlined in the previous section. A high shortest path is inherent to the structure of the network, while the small-world model we are using tends to produce low shortest paths. Consider two intersections separated by a great distance: in either of our models, there is some probability that those two intersections are connected directly. In real life, this probability is almost nonexistent. As such, a small-world network is the best of the two, but not ideal for modeling the Boston Area Road Network. An ideal model would be a grid, rather than a ring network. For the scope of this paper, however, a small-world model will be sufficient.

DEGREE DISTRIBUTION

Next, we compare the degree distribution of our road network to the distribution of the small-world network (see Fig. 1). The most notable difference between the two is the presence of two peaks in the road network distribution compared to only one in the model. We expect to see the shape exhibited in the model as it approximates a normal distribution, but the peaks in the real network are fascinating as they indicate a high number of intersections with degrees k=3 and k=6.

We can therefore assume a significant quantity of T-intersections and intersections at the convergence of three streets. The preponderance of these intersection types among all those in the Boston area (as opposed to intersections at the convergence of two streets as we expected) is unexpected, however, and in future analyses we would like to consider possible inaccuracies in our understanding of the intersection degrees.

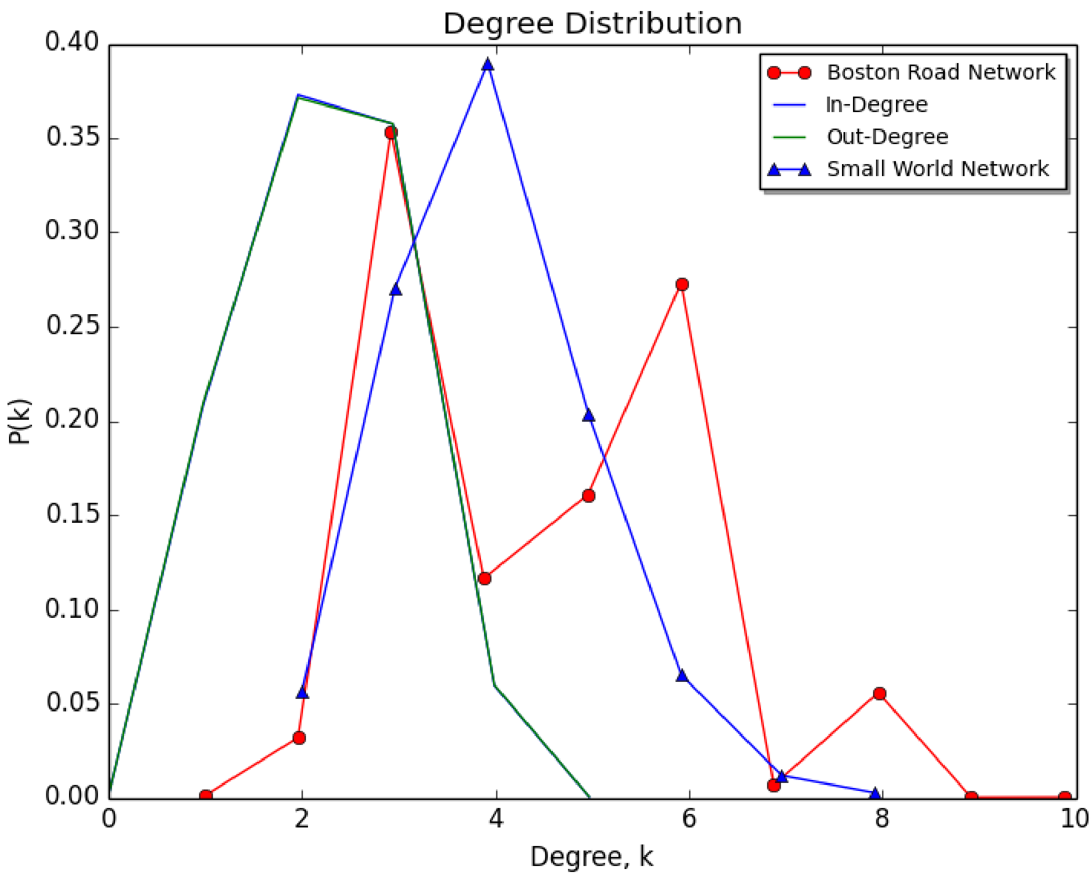

Figure 1: Degree distribution of Boston Area Road Network compared to the small-world model, which approximates a normal distribution. In- and out-degree are roughly equivalent.

BETWEENNESS CENTRALITY

Having determined that betweenness is the best measure of centrality (i.e. importance) for a road network like Boston's, we calculated the betweenness of each edge and node using the following equation:

$$C_i^b = \frac{1}{(N-1)(N-2)} \sum_{j,k \in N} \frac{n_{jk}(i)}{n_{jk}}$$

N is the number of nodes in the network, $n_{jk}$is the total number of links between node *j* and node *k,* and $n_{jk}(i)$is the number of links between node *j* and *k* that contain node *i*. The same applies for edges of the network.

The betweenness of each node and edge is stored as an attribute. The figures below (Figs. 2 and 3) show the network by node degree and by node betweenness, respectively. Figure 4 demonstrates the relationship between degree and betweenness.

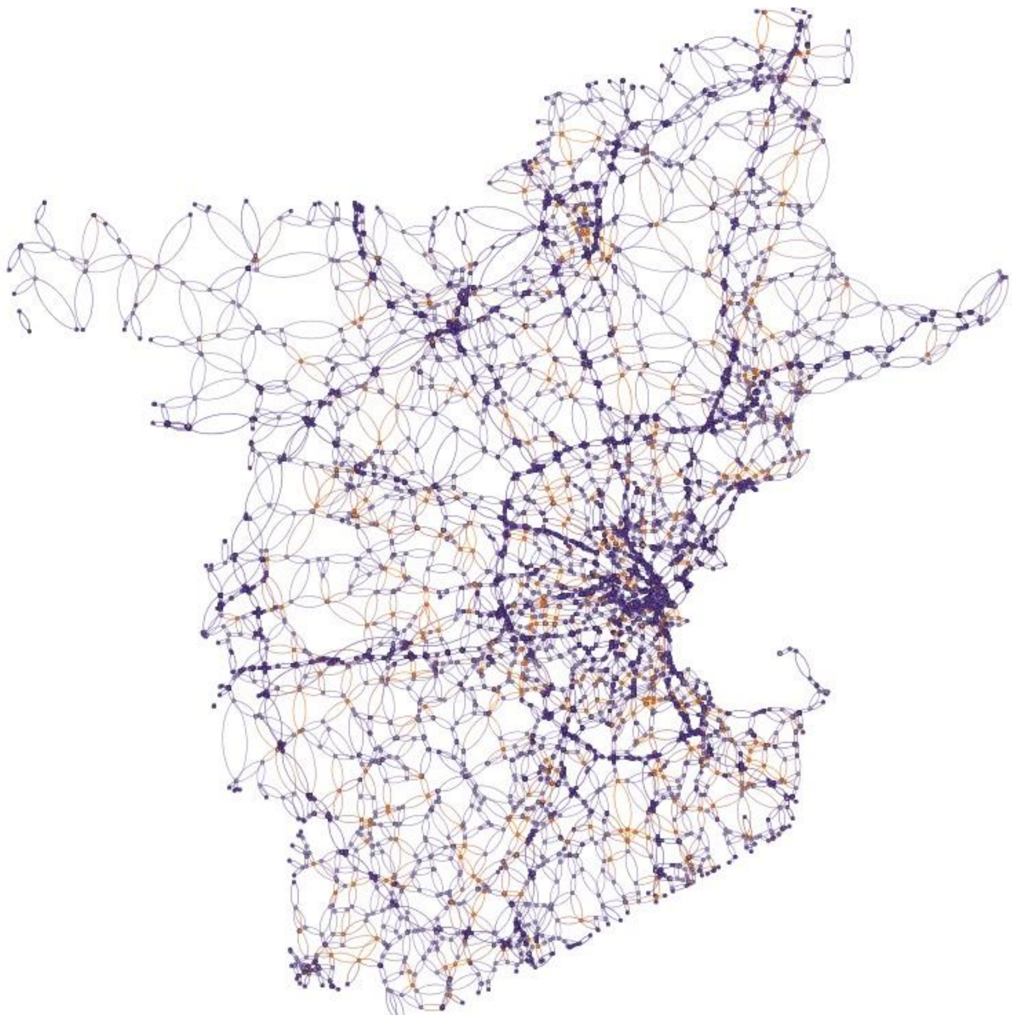

Figure 2: Node degree map of Boston Area Road Network. Purple nodes represent low-degree intersections while orange nodes represent high-degree intersections.

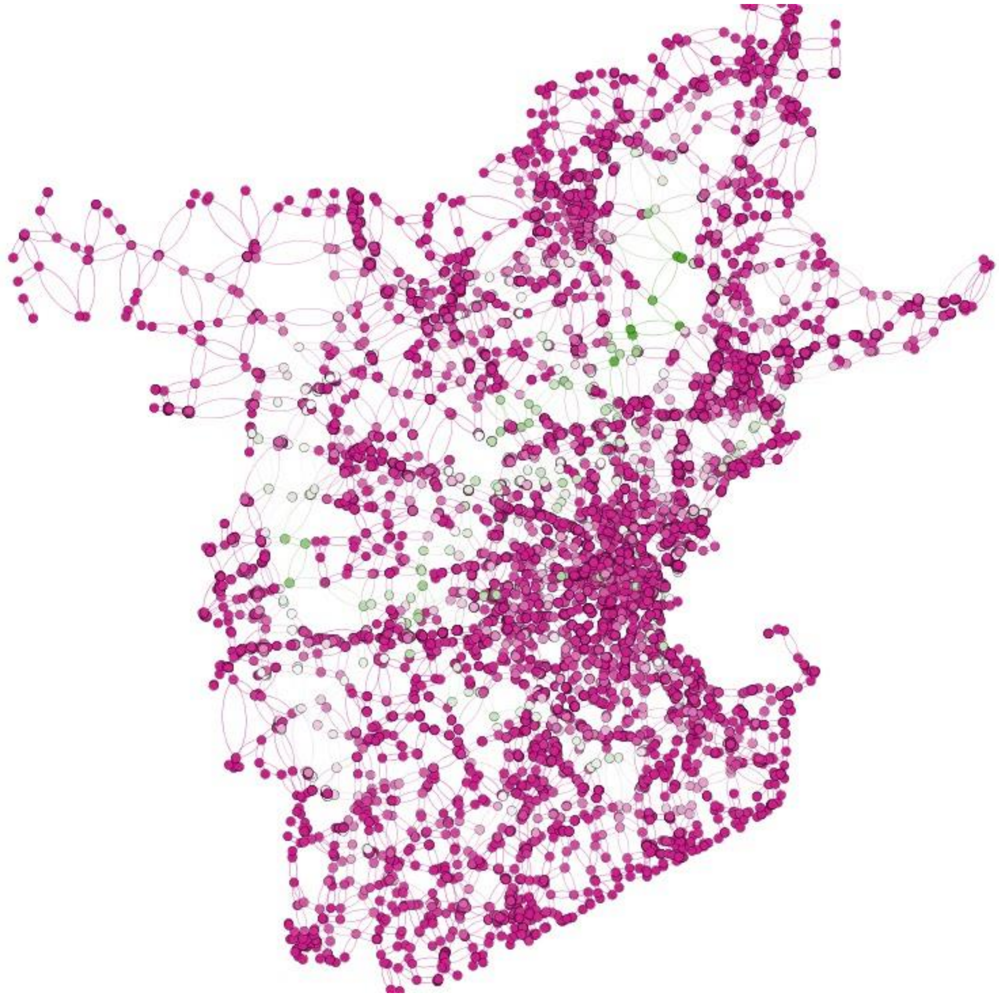

Figure 3: Node betweenness map of Boston Area Road Network. Pink nodes represent low-betweenness intersections while green nodes represent high-betweenness intersections.

FUTURE WORK AND CONCLUSIONS

From the node degree map, we can see a dense concentration of nodes that clearly delineate Boston. Despite the expectation that downtown Boston would have a higher concentration of orange (high-degree) nodes, most of the intersections are purple (low-degree). However, there are more orange nodes in Boston than outside of Boston, and part of the issue could be attributed to range issues in Gephi (i.e. the software's interpretation of high-degree appears to extend up to nodes with degree k=12, meaning even the significant number of nodes with degree k=6 may be interpreted as low-degree).

We can further see the ring roads surrounding Boston, I-95 immediately outside of the city and I-495 about 25 miles out, as well as the arterial spokes connecting them. These are collections of low-density nodes, but arguably constitute some of the most "major" roads in the metropolitan area. The correlation is logical as interstates have access ramps, not massive intersections, but other low-density nodes clearly do not correspond to important streets, which is why we use the measure of betweenness to help narrow our focus.

Comparing the two maps, we can see a significant correlation between node degree and betweenness. The primary difference is in the interstates (most notably I-95, I-495, and possibly Rt. 1 northbound), which are low-degree as discussed above but high-betweenness. Ideally, the most "important" roads would have both high degree and high betweenness, but as we can assume the interstates are in fact highly-used despite being low-degree, the high-betweenness roads should almost certainly be given highest priority in snow removal.

This network story focused on preparation for inclement weather, but we think this analysis could be used in the preliminary stages of scenarios on either end of the spectrum: emergency preparation and long-term maintenance programs, for example. For events even less certain and predictable than snowstorms—e.g. hurricanes, human-made threats, etc.—an understanding of road segment centrality can help emergency preparedness teams identify points of congestion as well as useful but underutilized alternative routes. For events more certain than weather—e.g. road surface deterioration, potholes from snowplows, etc.—a map of the most "important" streets in the Boston metropolitan area can help federal, state, and local road maintenance teams prioritize cyclical rehabilitation. We expect some correlation between the roads identified as most important for the purposes of snow clearance and those identified as priorities for maintenance. There may be less correlation with suitable emergency evacuation routes, but we think centrality is still integral to such analysis.