An Exploration of the Geometry of Road Networks Through Embeddings in the Euclidian Plane

Miles Aronow, Computer Science, Wesleyan University

Faculty Advisor: Saray Shai, Computer Science

Introduction

- Half of the global population already lives in cities, and by 2050 two-thirds of the world's people are expected to live in urban areas.
- The speed and scale of urbanization brings tremendous challenges in developing sustainable cities.
- In this project, we seek to understand the road networks of cities through a geometric lens. Specifically, we explore the dimensionality of urban road networks by embedding them in the Euclidean plane.
- Road networks are embedded in the Euclidian Plane through Multidimensional Scaling (MDS)
- "The general aim of multidimensional scaling is to find a configuration of points in a space, usually Euclidean, where each point represents one of the objects or individuals, and the distances between pairs of points in the configuration match as well as possible the original dissimilarities between the pairs of objects or individuals."(Cox)
- We embed the nodes of the City networks with dissimilarities between nodes as the shortest path through the road network
- Embeddings are evaluated by two metrics, Stress and Distortion
- $\text{Stress}_D(x_1, x_2, \ldots, x_N) = \left(\frac{\sum_{i,j} \left(d_{ij} - \|x_i - x_j\|\right)^2}{\sum_{i,j} d_{ij}^2} \right)^{1/2}$
- Distortion=1-R^2
- Stress and distortion are both on a scale from 0-1 with lower values indicating better embeddings
- Applications of this research can be used to analyze traffic patterns, as certain geometries can lead to certain congestion patterns.

Research Questions

- Can city road networks be embedded effectively in the two-dimensional Euclidean plane?
- How do small connected portions of larger city networks embed in comparison to random samples of nodes from the larger city network?
- What patterns can be observed in embeddings as the size of the network grows?

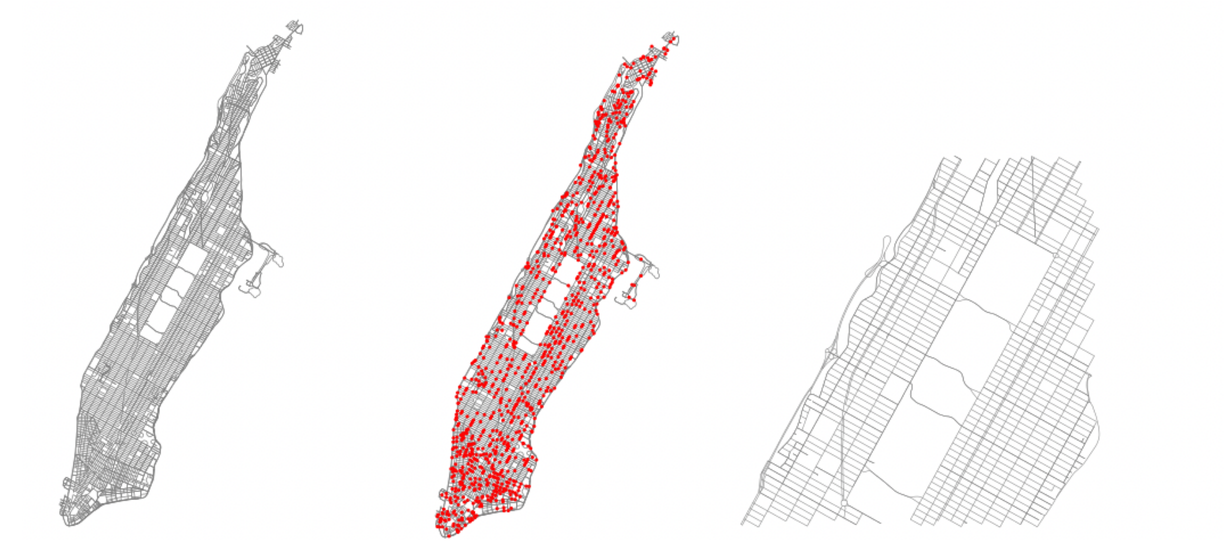
Further Exploration

- Growing networks by taking an increasingly larger random sample
  - An obvious continuation of the experiment is to grow the network through increasingly larger random samples of nodes as opposed to the method of expanding the network from the center.
  - For cities such as Buenos Aires or Boston where the experiment was able to capture a large portion of nodes, it would be hypothesized this method would converge to similar results
  - For cities such as London, in which the experiment only captured a small portion of nodes, perhaps this would show a better picture of the network as a whole
  - Furthermore, the pattern as the sample size grows could be analyzed for potentially interesting results
- Using travel times for the dissimilarity between nodes
  - In the experiment, the dissimilarity between nodes was given by the length of the shortest path through the network
  - This is a very naïve way of measuring dissimilarity; two nodes a mile apart through downtown traffic and two nodes a mile apart on a highway should not have the same dissimilarity
  - The experiment could be repeated with dissimilarity equal to the time it takes to get from one node to another

Results

Embedding Performance for Random Node Samples, and Graph Subsets

Fig 1 (a). The Manhattan Road Network (b). A Random Sample of 1000 Nodes. (c). A Subgraph of Approximately 1000 Nodes



The MDS algorithm is computationally expensive, and city road networks are large. Manhattan has only around four thousand nodes, but cities such as London have over a hundred thousand. Thus it is helpful to use a subset of the data to create the dissimilarity matrix inputted to the MDS algorithm. In this experiment two such methods of sub setting the data were examined. In the first method, a random sample of 1000 nodes was taken. The shortest paths through the in the entire graph were found for the sampled nodes to create the distance matrix. In the second method, one random node was selected from the graph. A subgraph was created of all the nodes within a bounding box of a set distance, such that the graph had on average 1000 nodes. The dissimilarity matrix was created for all the nodes in the subgraph using the shortest paths within the subgraph. Each method has it's pros and cons. Computing the matrix is faster for the subgraph method, since shortest paths do not have to be found through the entire graph. On the other hand, the sample is more representative of the whole network.

Figure 2(a). Embedding Results for Random Sample (b). Embedding Results for a Subset by Radius

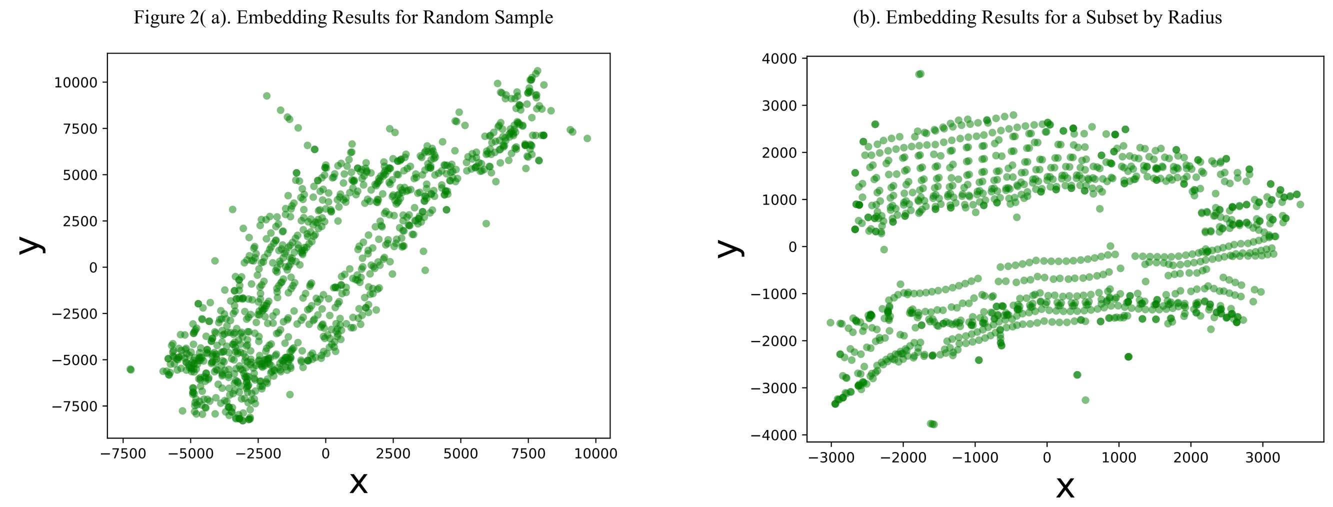


Figure 2.(a) Shows the results of an embedding of the random sample shown in Figure 1(b), and Figure 2(b). shows the results of the subgraph shown in Figure 1(c). In both embeddings, the general shape of the network, and features such as Central Park can be observed. One can observe that orientation is meaningless.

Figure 3 (a (b).

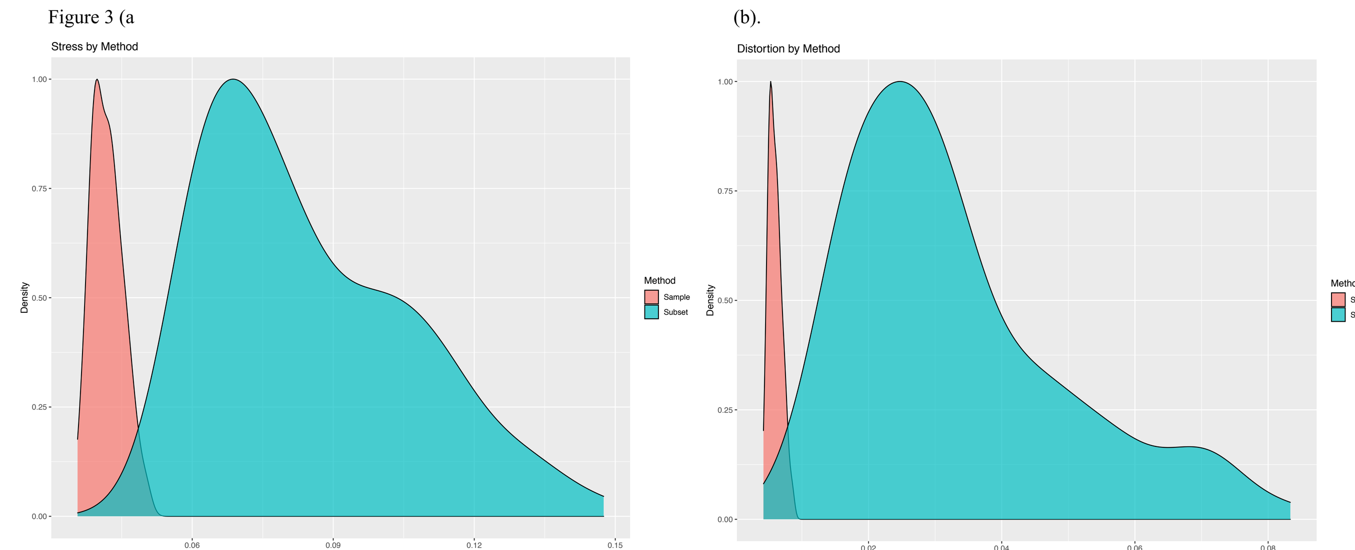
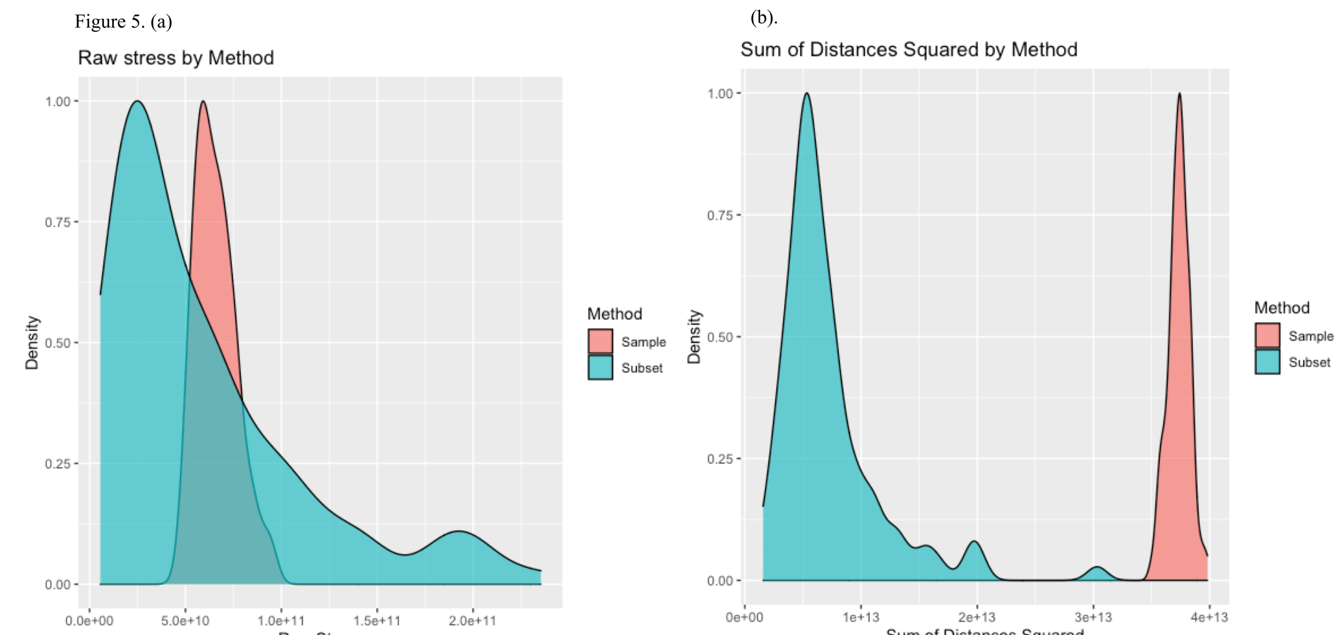


Figure 4. Embedding Results

Method	Mean Distortion	Distortion SD	Mean Stress	Stress SD
Sample	.0419	.000980	. 00597	.00330
Subset	.0846	.0164	. 0329	.0219

As observed in the graph, stress and distortion behave very similarly for both methods. The sample method embedded more effectively with a lower mean stress and distortion values as well as a much lower variance. ANOVA showed a difference a significant lower mean for both stress and distortion for the random sample method, with both p values less than 2*10^-16.

Figure 5. (a) (b).



Recall the formula for stress: $\text{Stress}_D(x_1, x_2, \ldots, x_N) = \left(\frac{\sum_{i,j} \left(d_{ij} - \|x_i - x_j\|\right)^2}{\sum_{i,j} d_{ij}^2} \right)^{1/2}$

The denominator is constant for each individual embedding. Thus, the algorithm only needs to minimize the numerator. This value is sometimes called the raw stress. The average value for raw stress is less for the subset method. However, the sum of distances squared is far greater for the random sample as it includes nodes from all around the city, not just nodes constrained to a small radius. This helps explain the differences in performance. The subset method is unable to have low enough raw stress to compensate for a smaller sum of distances squared.

Embedding Performance of Various Cities, as Networks Grow

Figure 6.

Shown above are the portions of the eight cities studied in the experiment. Starting from the center of the city, the networks were grown in intervals of 100 meters until they surpassed 5000 nodes. An embedding was computed five times for each network as there is a randomness component to MDS. Due to computational limits the most nodes we could analyze was 5000 nodes. For some cities, such as Buenos Aires, this was enough to capture a significant portion of the city. For other cities, such as London, it was only a small portion. This is demonstrated to the right in figure. The whole city is shown, with green being the network studied, and red, the rest of the city.

Figure 7. Portion of Network Studied for Buenos Aires and London

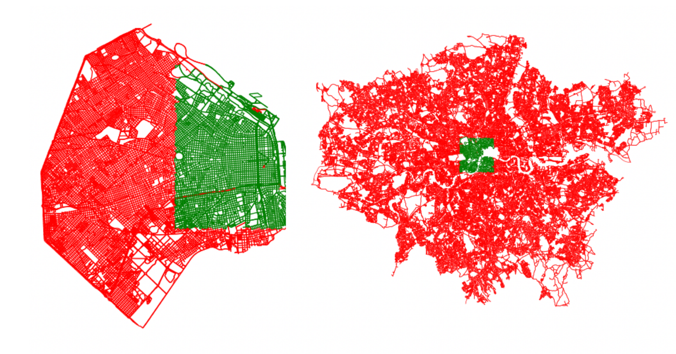
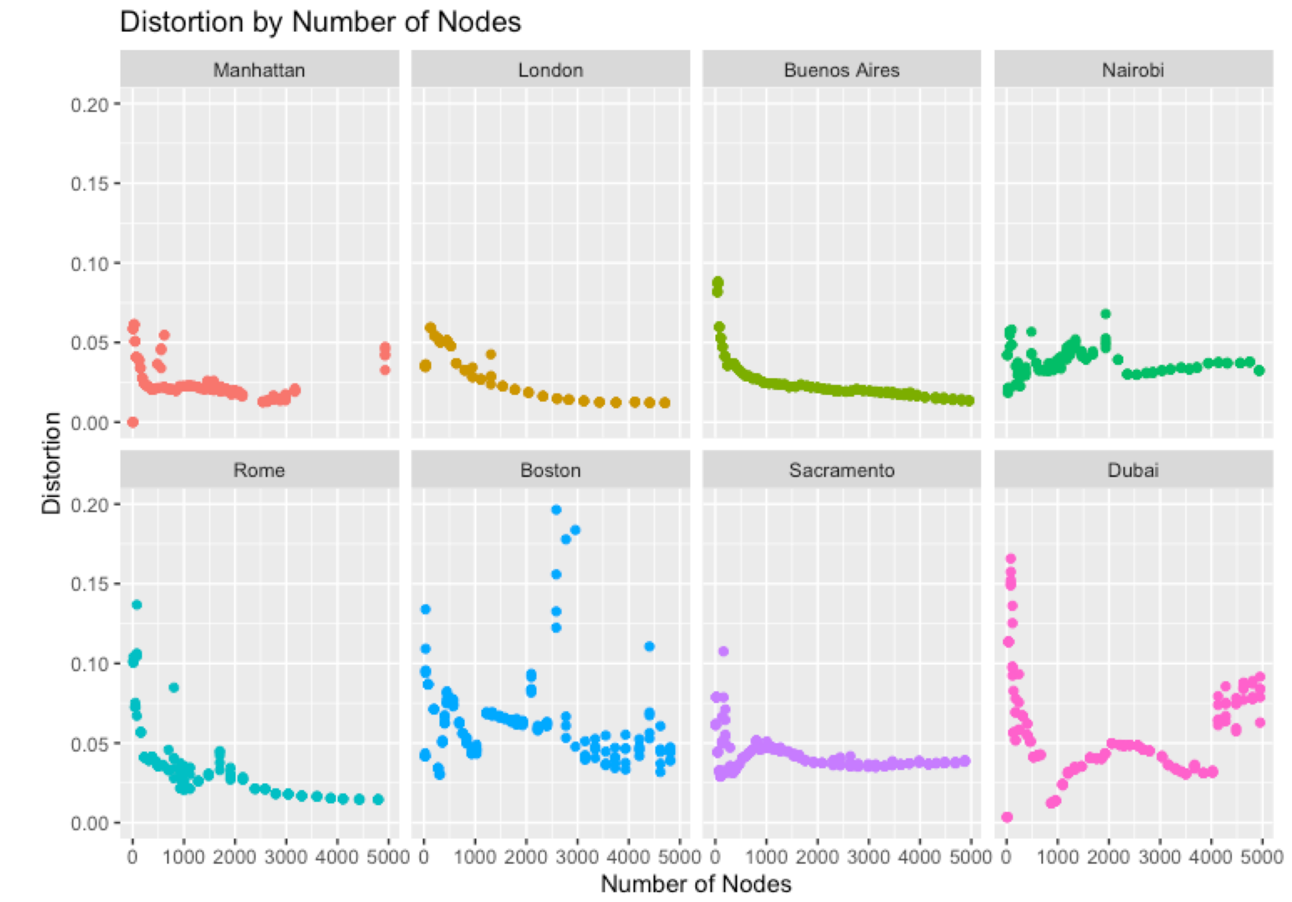


Figure 8 (a).

(b).



The shape of the stress and distortion plots were very similar. This is not surprising as they are both metrics of how effective an embedding is. Both stress and distortion showed a trend of lower values as the number of nodes increased. This shows that embeddings are getting more effective as the network grows. For London and Buenos Aires the pattern is clear. For Manhattan, Sacramento, and Dubai, a general negative trend is observed, but with more variation. For Manhattan this is not surprising given it's structure. As seen in Figure 9, the city graph expanded to include parts of New Jersey and Brooklyn across the river, connected only by one bridge. Boston and Nairobi showed the least clear patterns. Perhaps for Boston it was the harbor that accounted for this. For Nairobi, perhaps the lack of density was to blame.

References

Cox M., Cox T. (2008) Multidimensional Scaling. In: Handbook of Data Visualization. Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg

Batty, Michael, (2008/02/08) The Size, Scale, and Shape of Cities. Science