Generating Synthetic Road Networks from Various Reduced Dimension Representations

Elaine Zhou, Alec McGlaughlin, Deger Turan

December 9, 2015

1 Introduction

Road networks are a unique and iconic representation of a city that embeds topology, geography, and history. Early Sixties "lollipop" low density suburbs of Irvine or Walnut Creek look much different than self-organized patterns of Los Angeles or San Francisco, or highly planned grid structures of Cairo or Venice (Cardillo).

Recent analysis of road networks include analyzing real-world implications of network properties, and many different attempts of creating a uniform representation of all these urban differences, from turning roads into planar graphs, different representations of road networks. As urbanization continues and our world becomes increasingly connected, it is especially important to develop a metric that allows comparison of between cities of various forms and sizes. Developing new road networks can benefit from learning from previous established patterns and failures.

However, a generalizable representation and quantitative analysis of these networks still has much to expand. As we deepen our understanding of how road networks are structured, we can begin to automate reasonable road networks for entire neighborhoods or cities for a given geography using intersections of interest (important buildings, banks, schools, airports, landmarks, etc) or require certain constraints (direct edges between certain places). This may be particularly useful for cities that have not fully developed their road networks.

Here, we hope to answer two main questions in our research: (1) what network characteristics might distinguish cities from each another in a reduced dimension representation and (2) how well can we use these reduced forms to generate synthetic road networks. Understanding spatial representations and characterizing network parameters of our road networks may eventually help planners generate new road networks and optimize for future developments.

We begin by performing dimensionality reduction on complex road network data. From these k-dimension matrices, we can construct synthetic road networks - we utilize a distribution weighted sampling approach and create pseudo-random graph model. This is based on latent space distances with the same edge density as the original representation. We compare these synthetic networks to existing ones. In order to recommend optimal infrastructure, we compare graph diameter, size of the strongly connected component size, clustering coefficients, and degree distributions across these synthetic networks.

2 Review of Prior Work

We begin our discussion of related with with an overview of network analysis on road networks, and then dive into the applicability of using latent space models to represent road networks.

2.1 Road Network Analyses

Cities and their road networks are always in motion, and exhibit a huge range of diversity in shape (circular, linear, fractal) or street systems (regular, organic, tree-like). This makes storing, obtaining, and understanding road network data for any given particular city is an incredible hassle. The famous resolution of the "Bridges of Konigsberg" problem by Euler proposed that roads be represented as edges and intersections as nodes, and this is a representation that has stuck and maintained for most of road network analysis

(Blanchard). However, Jiang and Kalapala et al. use a dual-representation of the road network, using the intersections as edges, which has revealed to be advantageous for network flow analysis.

Cardillo et. al discovered that by using minimum spanning trees and greedy triangulations, certain structural properties of planar graphs of urban areas reveal certain local and global characteristics of different world cities. Cities of the same class e.g., grid-iron or medieval, exhibit roughly similar properties. This forms the basis of our hypothesis that we can not only model road networks with a few parameters, but generate new ones of a similar class.

Complex networks based study has been used to describe cities. Both Buhl ('06) and Porta ('06) showed that cities behave neither like a classical scale-free network nor a small-world essentially because of its spatial embedding, and how critical the spatial representation stays intact. It is clear that cities need a different model or framework of condensed representation.

2.2 Latent Space Models

Perhaps one of the strongest reasons for using a latent space model is how critically important the spatial representation of a road network stays intact. As an example of a latent space model, multidimensional scaling (MDS) aims to preserve the idea of distance as much as possible. Typically, MDS requires the input of a proximity matrix, such that each entry D_{ij} denotes the distance between a pair of nodes i and j in the network.

Hoff et. al introduced the Latent Space Model, in which nodes are associated with locations in a p-dimensional space, and edges are more likely if the nodes are close together in latent space. The authors denote the matrix of relationships between actors as Y, with entries $y_{i,j}$ denoting denoting the value of the relation between the actors in the latent space model (whether or not the nodes representing the actors are connected in the model network). The authors model these relations as being conditionally independent and thus the presence or absence of a tie (edge) between two nodes in the network, given the unobserved positions of these nodes in some Euclidean space is given by,

$$P(Y|Z, X, \theta) = \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \theta)$$

where X and $x_{i,j}$ are characters which may indicate pairing tendency. In our model, we do not consider any such additional characteristics, and thus we will ignore the $x_{i,j}$ terms in accordance with the authors' direction. The authors parameterize $P(y_{i,j}|z_i, z_j, \theta)$ using the logistic regression model, and thus,

$$P(y_{i,j}|z_i, z_j, \theta) = \frac{1}{1 + \exp(-\alpha d_{i,j})}$$

where $d_{i,j}$ is the distance between actors i and j in the latent Euclidean space, i.e. $d_{i,j} = ||z_i - z_j||_2$. For notational simplicity, we will let $\lambda = -\alpha$. The parameters and latent distances should be set as those which maximize the log-likelihood of the above function. One issue, however, is that the log-likelihood of the above function is generally not concave which makes identification of a global maximum difficult using typical gradient ascent methods. As a result, the authors recommend first estimating the latent distances using a method such as multi-dimensional scaling, then fine tuning the parameters from these distance estimates to maximize log-likelihood. In summary, there are two important steps in the implementation: (1) determine the distance between nodes in the latent space using a dimensionality reduction technique, and (2) fine tuning the model parameters, namely λ. In performing the dimensionality reduction, the goal is to map the nodes into a low-dimensional Euclidean space such that node proximity is preserved in the new space. Such dimensionality reduction can be done in multiple ways - Laplacian Eigenmaps, Isomaps, Local Linear Embedding, etc. The recommended approach is multi-dimensional scaling, or MDS, and which attempts to map a high dimensional matrix indicating pairwise distances or similarities into a lower dimensional space while preserving pairwise distances. Another potential method for finding positions and distances in the latent space is through the use of the Isomap algorithm, which is essentially a generalized version of MDS which deals more effectively with non-linear manifolds than traditional MDS.

3 Approach

3.1 Data Collection

All of our data comes from Utah's "Real Datasets for Spatial Databases: Road Networks and Points of Interest". In our discussion, we focus on the cities of Oldenburg and San Joaquin. By using these two cities as case studies, we show the validity of our methods, as future work is able to extend beyond larger cities and larger data sets.

3.2 Comparing Cities

Since multidimensional scaling requires a proximity matrix, we first generate an NxN distance matrix from the edge list. For nodes i and j in our proximity matrix D_{ij}, each entry represented the euclidean distance between the two nodes. Then using MDS, we can embed the information from the distance matrix into a set of coordinates for each node in a k-dimensional space. We will then use the arrangement of these k-dimensional coordinates in order to infer a connectivity profile characteristic of the network we are trying to model. That is, we utilize the distances between the k-dimensional coordinate points of two nodes to infer a probability that there is a direct connection between these nodes.

We hypothesize that our embedding of the shortest path distances between all pairs of nodes will provide a nodal distribution in k-dimensions which captures associations between nodes as distances, i.e. nodes which are more likely associated will be close to each other in the embedded space.

3.3 Generating Synthetic Road Networks

We leverage this hypothesis to create synthetic model networks in which the closer two nodes are in our embedded space, the more likely they are to be connected in the model network.

Generating synthetic datasets is a fairly common approach when collecting real datasets is tedious. This can happen in social networks - problems that might arise in real dataset collection include avoiding biases, privacy limitations, or distribution issues (Prat-Perez). Similarly, synthetic network generation has also been applied to traffic workload modeling. In these circumstances, synthetic data generators have become an extremely useful tool to test applications or scenarios that simulate realistic data.

We seek to generate model networks which share global characteristics with the original road network representation upon which we base our model. This approach is consistent with the protocol presented by Raftery in his discussions of latent space modeling, where we are defining the distance between two nodes or actors as the euclidean distance between the node coordinates in our embedded space. We discuss the exact procedure below.

4 Methods

4.1 Latent Space Model

Similar to Hoff et al. in their analysis of social networks, we hypothesize that important structural properties of the road networks of cities, i.e. a characteristic profile of the road networks for a given type of city, can be determined by the position of each intersection (node) in an unobserved, "latent", space. Unlike Hoff et al., we can compute the physical distances of the shortest path between any two intersections in a city. Thus, instead of using a maximum likelihood or Bayesian network framework to estimate the positions of nodes within a latent space, we can utilize multi-dimensional scaling and Isomap to reduce the data into an unobserved space in which its position within the space characterizes a given node's relation to other nodes (intersections in the road network representation). Once we perform multidimensional scaling or Isomap, we define $D_{ij}^{(L)}$ to be the euclidean distance from node i to node j in the embedded, k-dimensional, space. We define the probability that any two nodes are connected in our model to be

$$P_{ij} = \frac{1}{1 + exp(\lambda * D_{ij}^{(L)}}$$

where λ is a modeling parameter which tunes the impact of $D_{ij}^{(L)}$ on the probability of nodes i and j being connected in our model. This edge probability can be simply interpreted to imply that intersections which are further apart in the embedded space are expontentially less likely to be directly connected by a road segment than edges which are proximal in the embedded space.

In order to enforce similar density to the original model, and thereby maximize the likelihood of our model, we iteratively tune the λ parameter, starting with a wide range of values and performing a pseudo-binary search until we converged to a value of λ for each network which produced edge density similar to the original. In this way, we are creating a pseudo-random graph model based on latent space distances with the same edge density as the original representation. These networks can be validated against real roads and geographic markers, both of which are critical to evaluating the viability of new roads.

4.2 Defining Distance in the Input Space

Perhaps one of the strongest reasons for using multidimensional scaling or Isomap is that they aim to preserve the notion of distance in the input space as much as possible. Typically, both require the input of a proximity matrix, such that each entry D_{ij} denotes the distance, or similarity, between a pair of nodes i and j in the network. In social networks, for which much latent space modeling research has been done, papers use . Even for papers which use maximum likelihood, however, many start optimization from distance estimates using MDS.

Unlike in the case of social networks, the distances of road networks have physical implications. The benefit of road networks is we can use the basic geometry of the roads and intersection to define the distance between two intersections. Specifically, we define the distance between two intersections (nodes) to be the shortest distance that can be traveled using the road network to reach one intersection from the other. As a result, we are not using true Euclidean distance in the input space, nor pure Manhattan distance but rather a distance measure which is more consistent with the ”reachability” of one intersection from another, i.e. how far one would actually have to travel.

In our D matrix, each D_{ij} represents the shortest distance from intersection i to intersection j as found using Dijkstra's Algorithm. Although there are faster and more efficient ways to compute the D matrix, this is still a standard that many use when computing shortest path algorithms on highways and road networks (Goldberg). These shortest path distances populate our D_{ij} matrix of distances in the high-dimensional input space.

4.3 Multi-dimensional Scaling

As mentioned previously, multi-dimensional scaling takes a distance matrix D as an input and attempts to map the network nodes into a low-dimensional space while preserving pairwise distances. The distance matrix is then converted into a kernel matrix K, which is a centered and scaled version of the distance matrix, as follows:

$$K = -\frac{1}{2} HDH$$

where for a network with n nodes, H is defined by $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. The kernel matrix is positive semi-definite and thus can be written as $K = X^\top X$. Further, we take the singular value decomposition of $X^\top X$ and find that it can be decomposed as:

$$X^\top X = V \Sigma V^\top$$

We then plot the singular values, Σ, to find the ”knee” in the plot at which the additional signal contained in a $d+1$-dimensional low-rank approximation becomes small. We then define $\hat{\Sigma}$ to be the diagonal matrix containing the top d singular values, and $\hat{V}$ to be the corresponding singular vectors. Finally, we can recover the coordinates of our embedding, Y, which can be written as:

$$Y = \hat{\Sigma}^{1/2} V^\top$$

We define the latent distance between two nodes in our network, $D_{ij}^{(L)}$, as the Euclidean distance between the projection of the nodes (given by the coordinates above) in the embedded space.

4.4 Isomap

One limitation of MDS is that it performs linear dimensionality reduction, whereas the Isomap algorithm is a generalized version of MDS that performs nonlinear dimensionality reduction. As a result, Isomap is useful for dimensionality reduction of data which lives in a nonlinear manifold. There are three main steps in the Isomap algorithm: (1) generate a graph where adjacency is defined to be the nearest neighbors of each node in the high-dimensional input space, (2) compute the pairwise distance between all nodes in the nearest neighbors graph, and (3)perform MDS on the resulting distance matrix.

For our application of the algorithm, we first calculated the shortest path distance between all pairs of intersections in the road network. We then constructed our nearest neighbors graph by defining an edge between each node in the graph and its 4-nearest neighbors. Edge weights were defined to be the associated shortest path distances from the original road network. We chose 4 neighbors because in a city, one would expect to commonly find a grid pattern in which each intersection is directly connected to 4 other intersections. We then calculated the shortest path between all pairs of nodes in our weighted nearest neighbors graph and used the resulting distance matrix as the input distance matrix for the MDS algorithm described above.

4.5 Fixing the Latent Space Local Bias

One issue we encountered in the above described methods is that the model graphs which were produced displayed extremely high local connectivity, forming highly connected cliques which were not connected together. As a result, these networks had small strongly connected components, a large number of triads, and low clustering coefficients. There were also nodes with high degree, which is not found in real life road networks. These results were likely caused by the heavy consideration given by the latent space model to the shortest path distance between two nodes in the original road network. The two methods of finding low-dimensional embeddings, MDS and Isomap, both took only the shortest path length between all pairs of intersections as inputs. As a result, it is possible that two intersections on opposite corners of a city block may be more likely to be connected in the model network than two intersections directly connected by a road but which were a greater distance apart than the Manhattan distance between the first two intersections. This is clearly an issue that needs to be dealt with as it can lead to unrealistic connections between intersections, leading to very high node degrees for intersections which are located close together geographically. There are two major problems that needed to be addressed with the original methods, (1) the lack of global connectivity, i.e. the lack of roads between clusters, and (2) the unrealistically high degree of certain intersections.

There are two approaches that we tried to remedy these problems, the first was to increase the value of λ which would decrease the density of the graph, then to add edges randomly to the graph with probability 1/10000 even if they would not be included by the latent space model, thus the probability two nodes are connected becomes $P_{ij} = (1+exp(\lambda * D_{ij}^{(L)})^{-1}$. As before, we tuned the λ parameter to achieve the same edge density while adding the random edges as in the original road network. The reason that we added random edges is that it would increase the chances of forming edges between the highly connected communities and thereby increase global connectivity. In essence, these random edges are meant to reflect expressways or highways within a city that allow an individual to drive commute quickly between distant points in a city. In cities, these highways would be built to connect high-value or high-traffic areas, however, our data does not inform decisions about node values and as a result we resorted to random edge insertion.

The second approach we used was to incorporate the number of roads that an individual would have to take to transit from one intersection to another, i.e. the number of hops between intersections in an unweighted representation of a road network, into our measure of distance in the input space. When commuting in a city, the relative distance or "reachability" of one intersection from another is dependent on both the physical distance that one must travel between intersections and the number of road segments that constitute the path. The more road segments an individual must travel, the greater the number of traffic lights encountered, the more turns required, etc. In this new approach we define B_{ij} to be the number of hops needed to reach intersection i from intersection j and W_{ij} to be the physical length of the shortest path between the two intersections in the original road network. We then define $D_{ij} = W_{ij}B_{ij}$, where D is the distance matrix input to MDS. This newly scaled distance matrix is designed to prevent links between intersections on opposite corners of city blocks and thus increase our model's fidelity to the expected grid pattern of cities.

5 Results and Findings

5.1 Comparison of Strategies

As mentioned in our Methods, we first implemented the latent space model based on embedded coordinates derived from multi-dimensional scaling and Isomap on the physical shortest path lengths between intersections in the original network. In order to evaluate our synthetic model road networks, we compared the size of their largest SCC, the network diameter, number of triads in the network, the clustering coefficient, and the degree distribution. All of the results except for the degree distribution are given in Figure 1.

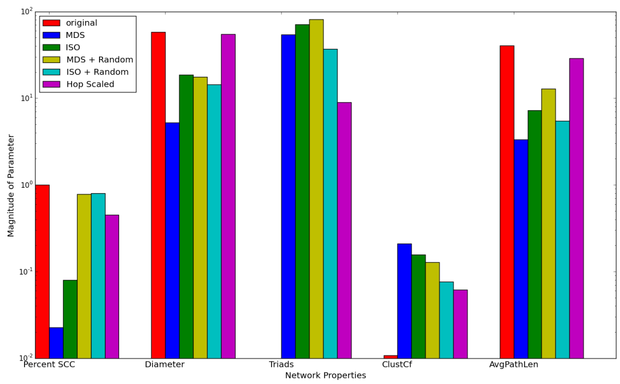

Figure 1: Real and Synthetic Oldenburg Network Properties

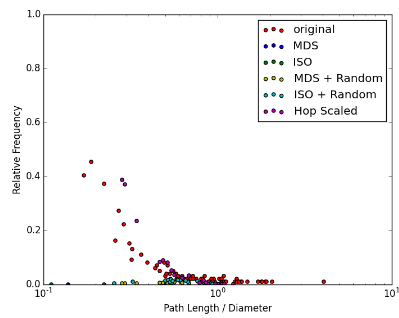

Figure 2: Path Length Distributions

Connectivity. A major issue with the results of our first two implemented methods is that the resulting model networks are plagued by extremely high local connectivity. These networks tend to form highly connected cliques with very sparse connections between cliques, in fact many of the cliques are not connected at all. This local connectivity bias manifested itself in the results in dark blue and green bars in Figure 1. It is clearly evident from this plot that these networks had small strongly connected components, a large number of triads, and low clustering coefficients compared to the real road network. There were also nodes with very high degree as seen in the degree distributions given in Figures 4 and 5, which is not consistent with real life road networks. These results can be attributed to the latent space model's sole reliance on the distances between nodes in the latent, or embedded space. Thus, we need to reconsider how we construct the embedded space or our exclusive use of the latent space model. We explored two different options as described in our Methods, adding random edges to increase global connectivity, and scaling by number of hops to decrease local connectivity. Both of these approaches drastically increased the size of the largest connected component, resulting in model networks closer in connectivity to the original road network.

Diameter and Path Length. Out of all five methods we implemented, the hop scaled adjustment to MDS proved to have the highest fidelity to the original network in both diameter (Figure 1) and overall path length distribution as shown in Figure 2. The hop scaled MDS model also had average path length closer to the original network than any of the other models. All of the other model network creation methods yielded drastically fewer pairs of nodes with short path lengths, and also had slightly lower approximate diameters. These decreases in diameter are due to the fact that the MDS and Isomap networks are not fully connected, and the networks with random edges are expected to have lower diameter (as you would expect in a random graph). These results suggest that the hop scaled MDS method tends to perform best in terms of replicating path lengths from the original network.

Triads and Clustering Coefficient Due to the strong local bias of our latent space model implementation, all of our model networks have significantly higher numbers of triads and higher clustering coefficient than the original network. We were able to reduce the number of triads and the clustering coefficient with our hop scaled MDS model as we initially predicted, however, there is further investigation to be done in order to reduce the number of triads present in the model networks. There are two possible routes in which we can attack the problem in future work: implementing constraints on the degree of nodes in the model or constraints to prevent the formation of triangles, or transform the embedded space or model parameters to encourage the latent space model to resist the formation of triads, as we attempted in our hop scaled MDS method.

Degree Distributions. In the original network, there are no intersections which have degree greater than 5. Further, previous work has noted that an intersection often does not have more than 4 incident roads (Courtat). Many of our model networks, however, have intersections with a large number of incident roads which is unrealistic. We attempted to manipulate the node positions in the embedded space in our hop scaled model in order to generate a more realistic degree distribution and we were largely successful. While the degree distribution of hop scaled model is not exactly the same as the original network, it constitutes a significant improvement from the other models.

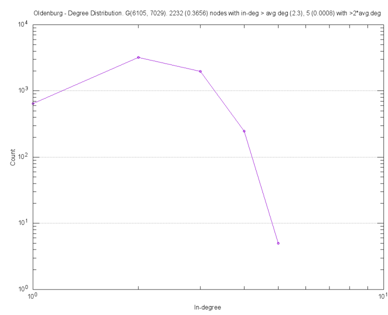

Figure 3: Oldenburg, avg-deg: 2.3

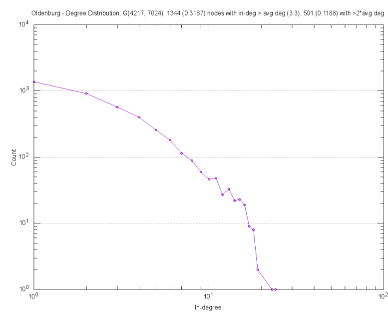

Figure 4: MDS, avg-deg: 3.3, $\lambda = 2.6$

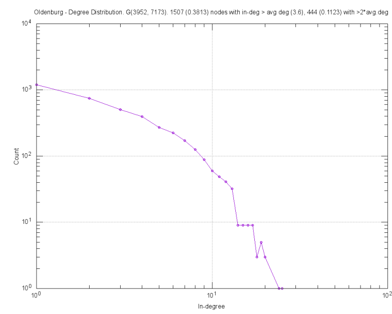

Figure 5: Isomap, avg-deg: 3.6, $\lambda = 7.6$

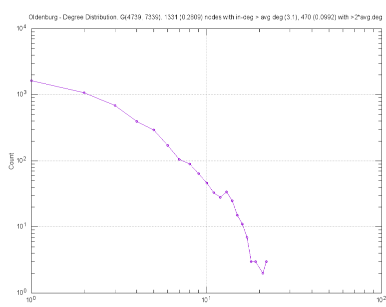

Figure 6: MDS & R, avg-deg: 3.1, $\lambda = 4.0$

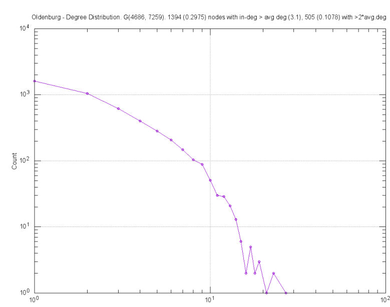

Figure 7: Isomap & R, avg-deg: 3.1, $\lambda = 10.0$

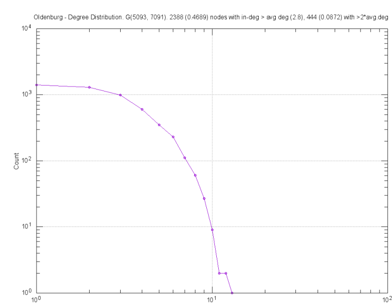

Figure 8: MDS & HS, avg-deg: 2.8, $\lambda = 0.102$

5.2 Visualizing Synthetic Graphs

Because our data set of road networks is inherently geospatial, it is critical to evaluate the success of various methods not only relying on the graph properties, but also taking into account the resulting graph communities formed by the road structure. We must validate with the full network and evaluate – do these networks make sense in the physical world, do they have similarities to real-world road phenomena, such as highways, or small streets in residential neighborhoods? We graphed the synthetic road networks to highlight the roads with the highest edge betweenness centrality scores, since they altogether outline the most critical backbone of the city road map. The intuition is that the main roads (i.e. highways) and roads that must be travelled for many points around the city have high betweenness centrality, since many smaller roads branch off from these points. The thickness of edges in the graphs correspond to the betweenness scores on a logarithmic scale. In almost all models, including the original city, there are a handful of nodes that have betweenness scores that are orders of magnitude higher than most of the nodes. Instead of only highlighting the most important ones, the betweenness centrality on log scale also highlights more regional, yet important roads.

On the other hand, the color of edges, on a spectrum of pink to blue to gray as highest to lowest, reflect betweenness centrality scores on a linear scale, outlining the drastic drop in score as only a handful of nodes are colored. All of the model visualizations are in the appendix, and the two models that yield the best results are below, along with the original model for Oldenburg.

Figure 9: Real Oldenburg Network Visualization

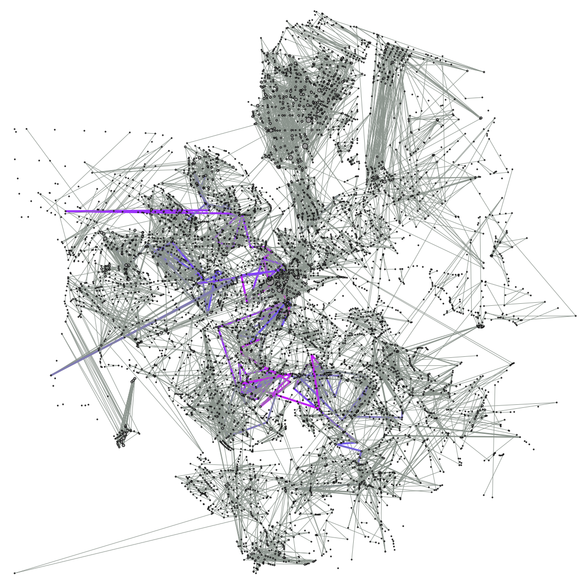

Figure 10: Hop-Scaled Model Oldenburg Visualization

The original Oldenburg model has a clearly defined road in pink that most nodes travel through to connect with other nodes. The central loop, with the west end more critical to the city, branches into smaller and smaller neighborhoods with roads that naturally have smaller betweenness centrality, as they are only used for the local nodes. The presence of several thick edges that are not pink or blue shows that while several side branches also have relatively high betweenness, the main backbone in pink is much higher for the connectivity it adds to the city.

The MDS model has the edges with highest betweenness all fall in the vicinity of the same neighborhood, which consists of the nodes in the original city that have the edges with highest betweenness centrality scores. The model, in a sense, considered those nodes as a "hub" to connect the rest of the connected component through. The fact that the largest weakly connected component is 1.556 percent of the original graph shows that the hub it generates is also not extremely significant in connecting the whole city. Some other connected components also contain fragments of the "pink" segment, of the roads with highest betweenness, thus there is no long chain of an important road in this model.

We also generated an MDS model with a high lambda score, which effectively decreased the diameter from 17 to 12, and average road length from 3.35 to 2.03. As in the southwestern cluster, we see that there are several nodes across two different neighborhoods that have independent edges, while a realistic city would contain a more important road connecting the two. Seeing several parallel edges across communities is a strong indicator that the road generation model is not efficient.

The ISO model does not have roads with highest betweenness following the old nodes, but uses the northern neighborhood of the city as a true hub that connects the rest of the city. With slightly larger WCC, the neighborhood has both the colored edges, and most of the thick edges. This means that the rest of the network has edges that travel into this cluster, if the nodes are connected to others, or use one of the rare roads that connect points across the city.

Both the MDS and ISO models that also incorporate random edge generation yield more hairball-like graphs. While the neighborhood connections are far too complicated to make visual insights, one important quality we can see from these graphs is that the number of thick roads is much higher, which means the betweenness centrality scores have overall flattened, with the spike at the highest end of the spectrum also encompassing a slightly higher number of roads. While the graph property metrics seem closer to the real city, by analyzing the graphs, we see that the intuition of randomly adding in "highways" into the model is not realistic - the randomness should be aimed at connecting specific communities that already have high connectivity in.

The problems in the random edge addition visualization are addressed in the hop-scaled MDS model. We observe several well-connected communities, and two well-dispersed highways that jump between most of the neighborhoods. While it is unreasonable to expect a grid pattern as in a realistic city, the model actually closely mimics the neighborhoods of Oldenburg, by "hairballs" at each neighborhood. While the pink highway does not go circular in connecting all of them in the most efficient way as in the real city, it touches several of the northwestern neighborhoods. Based on this visualization, if an MDS hop-scaled model is combined with an intelligent highway, we create a mapping of a city that both has big weakly connected components, and has the structure of a main road that branches off into smaller structures for neighborhoods.

6 Future Work

6.1 Embedding Information into Latent Space.

As a proof of concept, we created a latent space model of road networks based purely on their physical locations. As seen in our visualizations and graph property comparisons, simply using this as a metric had room for improvement, even when using different graph generation techniques. As shown by Boguna and Raftery, different information can be embedded into the latent space. Perhaps the most relevant information for road networks are road hierarchies (highways, avenues, etc) and natural-forming communities or neighborhoods.

Communities & Hierarchies. Road networks belong to a family of networks that in which there is often overlap such that nodes simultaneous belong to several groups. As Ahn, Bagrow, and Lehmann showed, this has the consequence that a global hierarchy of nodes cannot capture the relationship between overlapping nodes. They proposed using link communities to uncover hierarchically organized community structures while maintaining pervasive overlap. However, Lancichinetti, Fortunato, and Kertesz presented an algirthm that used the local optimization of a fitness function to reveal community structure in the histogram peaks.

Future work could generate several road networks in smaller neighborhoods as we've shown. Since the MDS-hop scale method most successfully generates local communities that are strongly connected, adding the random edge "highways" can remedy the overall connectivity across the several weakly connected components.

6.2 Hyperbolic Mapping

Boguna et. al mapped the Internet to a hyperbolic space, and the resulting map exhibited scaling properties that are theoretically close to the best possible, meaning strongest heterogeneity and clustering, and remarkably robust (resistent to disturbances and damages to the network structure). The hyperbolic space makes sense for the Internet, because the space needed to hold the huge amount of smaller networks is much greater than that needed for the relatively fewer big networks.

Road networks seem to work similarly – we have much fewer "big" networks such as highways, and many more "small" networks such as local roads. As a stretch goal, we hope to also apply hyperbolic mapping to road networks. Aside from the suggestion of highly robust and efficient road networks most similar to that of existing cities, hyperbolic mapping methods yield insight into other complex networks. Instead of trying to split nodes into discrete community sets, the hyperbolic map would naturally yield a continuous measure of similarity between nodes based on hyperbolic distances.

6.3 Generating Synthetic Graphs Based on Physical Points of Interest

Because our models only rely on geospatial orientation of road intersections, future work can incorporate using a reduced dimensionality matrix to generate reasonable road networks for any given geography. In this process, we outlined how distances between road intersections (nodes) led us to generate road networks similar to that of a given city. We imagine that this represents a new location with a similar geography or physical constraints when creating new road networks. Since points of interest (important buildings, banks, schools, airports, landmarks, etc) are often located at busy road intersections, a city planning committee could use the same process to generate road networks for entire neighborhoods or towns but using critical points of interest connections as a starting point. This may be particularly useful for cities that have not fully developed their road networks.

7 Conclusion

We set out to explore the utility of implementing latent space models in conjunction with multi-dimensional scaling and the Isomap algorithm to extract the characteristics of a city's road networks, embed them in a latent space, and then use the latent space embedding to inform the creation of model networks which resemble the original network. Our results indicate that the use of latent space models with the proposed dimensionality reduction techniques is a useful way to advance understanding the characteristics of complex road networks and, in particular, how to construct models of such networks. We were able to reduce the structure of complex road networks into a low dimensional embedding based on distance and proximity properties in the original road network.

Our original approach focused solely on physical shortest path distance between pairs of intersections, however, the models resulting from this simplistic approach did not capture the global network properties of the original road network. These simplistic methods exhibited a strong bias towards local clustering and triad formation. Road networks are unique in that there's a physical constraint for in and out-degree (an intersection rarely joins more than four roads). Thus, more creative approaches are required to create more networks that better reflect real-world constraints. We attempted to correct the local connectivity bias by either adding random edges to increase global connectivity or scaling physical shortest path distance by the number of road segments on the path. The second approach, hop scaling, proved to be the most effective of our model network creation methods.

Ultimately, We hope that future work can expand on what we've done so far in understanding spatial representations and characterizing network parameters of our road networks to optimize for future urban development. There are several promising frontiers which can be expanded upon to increase understanding of road networks and to improve our ability to generate model or synthetic networks. The methods and results we present in this study are a strong proof of concept and take-off point for future study.

Alec: Researching, Coding, and Testing Algorithms; Methods and Results Writeup
Elaine: Analysis of Model Network Properties and Creation of Results Figures; Background, Results, and Future Work for Writeup, Poster
Deger: Prior Work, Creating Visualizations and Graph Analysis and Comparison Section of Writeup

---

8 References

[1] Mokbel, M., et. al "MNTG: An Extensible Web-based Traffic Generator"

[2] Prat-Perez, A., Dominguez-Sal, D. "How community-like is the structure of synthetically generated graphs?"

[3] Boguna, M., Papadopoulos, F., Kriouskov, D. "Sustaining the Internet with hyperbolic mapping"

[4] "Elementary processes governing the evolution of road networks" http://www.nature.com/articles/srep00296

[5] "Grid Unlocked: How Street Networks Evolve as Cities Grow" http://www.scientificamerican.com/article/grid-unlocked-how-street/

[6] Hoff, P., Raftery, A., Handcock, M. "Latent Space Approaches to Social Network Analysis". Journal of the American Statistical Association. Vol. 97, Iss. 460, 2002 http://www.stat.cmu.edu/ brian/905-2009/all-papers/hoff-raftery-handcock-2002-jasa.pdf

[7] Tang, L. Liu, H. "Community Detection and Mining in Social Media."

[8] Cardillo, A., Scellato, S., Latora, V., Porta, S. "Structural properties of planar graphs of urban street patterns." http://www.maths.qmul.ac.uk/ latora/cardillo_urban_066107_pre06.pdf

[9] Raftery, A., Niu, X., Hoff, P., Yeung, K. "Fast Inference for the Latent Space Network Model Using a Case-Control Approximate Likelihood." 2012.

[10] Sarkar, P., Moore, A. "Dynamic Social Network Analysis using Latent Space Models"

[11] Porta, S., Crucitti, P., and Latora, V., Environment and Planning B: planning and design 33, 705 (2006).

[12] Buhl, J., Gautrais, J., Reeves, N., Sol, R., Valverde, S., Kuntz, P., and Theraulaz, G., Eur. Phys. J. B 49, 513 (2006).

[13] Blanchard, P. and Volchenkov, D., Mathematical analysis of urban spatial networks (Springer Complexity, 2009).

[14] A. Penn, Environ. Behav. 35, 30, 2003.

[15] Vamsi Kalapala, Vishal Sanwalani, Aaron Clauset, and Cristopher Moore. 2006. Scale invariance in road networks. Physical Review E, 73(2):026130.

[16] Bin Jiang, 2007. A Topological Pattern of Urban Street Networks. 1 Physica A 384, 647–655.

[17] Goldberg, A. "Highway Dimension and Provably Efficient Shortest Path Algorithms"

[18] Ahn, Y., Bagrow, J., Lehman, S. "Link communities reveal multiscale complexity in networks"

[19] Lancichinetti, A., Fortunato, S. , and Kertesz, J. "Detecting the overlapping and hierarchical community structure in complex networks"