Deciphering Traffic Dynamics: Insights from the Los Angeles Highway Network

Michael Lui & David Davini

***Abstract*—Traffic is an essential phenomenon in urban systems that can be represented as an evolving dynamical system on a fixed road network. Understanding and predicting traffic flow is a high-impact task for improving the quality of life in overcrowded urban areas and systems. In this paper, we specifically analyze traffic flow on the Los Angeles highway network. In particular, we characterize the structure of the LA road network, the nature of LA traffic flow, and the relationship between network structure and traffic flow. We also attempt a machine learning approach to discover parameters for a dynamical system on the road network that predicts LA traffic flow.**

Source code available at: https://github.com/daviddavini/math-168-final-project

I. Introduction

TRAFFIC is an infamous yet omnipresent phenomenon in daily urban life. In America, cars are the main form of travel: 87.5% of all adult Americans drive, and the average American spends nearly 50 minutes driving each day [1]. Since the American transportation system is solely reliant on well-functioning road networks to operate properly, road network congestion is a serious national concern. Traffic harms the environment and damages the economy, causing an estimated 3.3 billion gallons of wasted fuel annually and $166 billion in worker productivity [2].

No city faces worse traffic than Los Angeles. In LA, residents spend an average of 119 hours every year in gridlock traffic, the highest of any U.S. city [2]. Los Angeles is also consistently ranked highly as one of the worst cities for traffic congestion [3] [4] [5]. This motivates us to study the theory and properties of traffic flow in the Los Angeles road network in hopes of identifying possible solutions to the LA traffic crisis. Any solutions to optimizing traffic flow and mitigating congestion would reduce the fuel waste that comes with traffic congestion and decrease average travel time, both of which increase the quality of life for LA residents. A key component of mitigating traffic is having accurate methods of traffic prediction.

II. Background

Traffic prediction is a well-studied topic in the scientific literature. Traditionally, traffic research has focused on *dynamical modeling*, which attempts to describe the mechanics of traffic (eg. using differential equations) and then simulate those mechanics. For instance, in a landmark paper for the field, *Dynamical model of traffic congestion and numerical simulation*, researchers Bando *et al.* modeled traffic congestion in terms of the equations of motion of individual vehicles on the road [6]. They then used this model to investigate the evolution of traffic over time, observing that simulated traffic naturally clusters into low-velocity congestion "bubbles" that travel opposite the road's direction, similar to real-world traffic. Dynamical model are illuminating in that they characterize the underlying causes and behavior of traffic. However, dynamical models are limited in their predictive accuracy by the simplifying assumptions made while modeling.

With the recent rise in popularity of machine learning methods, *data-driven methods* of traffic prediction have become more popular. In particular, researchers have started investigating neural-network models of traffic. In *Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting*, researchers Yu *et al.* propose a novel model architecture that allows for time-series prediction of network data [7]. The new architecture achieves state of the art traffic prediction results compared to fully-connected neural networks and recurrent neural networks, while also being significantly faster to train. This suggests that sparse nature of traffic graphs make them especially suited to prediction via graph neural networks, which can take advantage of the road network structure without much computational overhead. However, data-driven methods are opaque, providing no insight into the underlying mechanics of traffic flow. In this paper we seek to balance these two concerns, employing both mathematical analysis and machine learning methods for traffic prediction.

III. Dataset Description

Our dataset comes from the Caltrans Performance Measurement System which is a system that has more than 39,000 traffic sensor stations that collect real-time data across California traffic systems [8]. From the data Caltrans provides, the research paper *Predicting Los Angeles Traffic with Graph Neural Networks* introduces two subsets of the dataset: PeMSD7(M), a medium-sized dataset comprised of 228 traffic sensors and PeMSD7(L), a larger dataset comprised of 1026 traffic sensors. Both datasets contain data about each sensor's speed at a given timestamp, where each timestamp last 5 minutes, in the time interval of May 1st to June 30th, 2012. The researchers publish the geographical coordinates for the 228 sensors of PeMSD7(M) on the GitHub, however they failed to publish the coordinates for the PeMSD7(L) dataset. As a result, any geographic plots we make correspond to the PeMSD7(M) dataset.

Both datasets also included adjacency matrices defining the edge connections between the stations. The edges of the network were created based on the following criteria: for any two sensor nodes, an edge connects them if and only if the geographic distance between them is less than some threshold distance and the weight of the edge is inversely proportional to the distance. The diagram in Figure 1 visually explains this process.

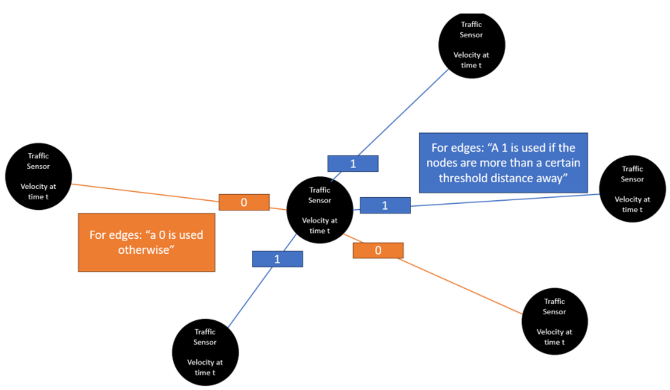


Fig. 1: Diagram explaining the criteria used to construct the network edges.

IV. Methods

During our computational experiments we extensively used several Python libraries: NetworkX, PyTorch, NumPy, Matplotlib, and pandas. Our methods can be roughly separated into four categories: Graph Structure Analysis, Traffic Data Analysis, Correlations between Graph Structure and Traffic Flow, and Machine Learning Models.

A. Graph Structure Analysis

First, we sought to understand the structure of the underlying road network for our dataset by creating visualizations and computing various network statistics.

To start, we plotted the traffic sensor stations geographically on top of a map of the LA region, using the coordinate data included in the PeMSD7(M) dataset. We then plotted the edges between sensor stations with each edge's transparency set to its weight value to attain a complete visualization of our network graph (Figure 2). We also graphed the adjacency matrix for our dataset (Figure 3), and used NetworkX's draw feature to get representations of the graph topology for both PeMSD7(M) and PeMSD7(L), shown in Figure 4

Next, we computed the degree distribution of the road network for the PeMSD7(L) dataset and plotted the distribution's histogram, as well as computed the cumulative degree distribution and plotted it in log-log scale (Figure 5).

Finally, we computed various centrality measurements for the nodes of the road network, in particular degree centrality (Figure 6b), eigenvector centrality (Figure 6a), Katz centrality and PageRank centrality (Figure 6c). We geographically plotted these values for each station node (on top of a map of the LA region).

B. Traffic Data Analysis

Our second task was to understand the traffic data in our dataset, again using visuals and various computed statistics.

We began by computing the average speed of traffic at each time of day, averaging over the 44 days in the dataset. We plotted these values, marking the local extremas of highest- and lowest- average network velocity (Figure 7a). We then geographically plotted the average speeds of each station at these critical times of day (Figure 7b, 7c).

Going further, we created an GIF animation showing the evolution of speed of each station across time, including 3 full days of data. This video is uploaded to YouTube at https://www.youtube.com/watch?v=EWMMrWV6kVY. (You should watch it! We made it for you!)

C. Effect of Graph Structure on Traffic Flow

The third aim we had was to analyze the effect that the fixed structure of the road network has on the evolution of traffic on its nodes.

To do this, we computed the Pearson correlation coefficients between the speeds of each station and various centrality metrics. We did this across each time of day and plotted it, obtaining a curve describing the relationship between speed-centrality correlation and time of day (Figure 8a, 8b). (Each sensor's speed was again averaged over all 44 days in the dataset.) We also computed the p-values for each of these correlations at each time of day to determine their statistical significance, and plotted these as well (Figure 8c).

D. Machine Learning Models

Our final and most ambitious aim was to create our own machine learning model to predict traffic flow in our dataset. For this we considered two models: Our first model was simply a linear dynamical system on the network (see equation 1), parameterized by a weight matrix and a bias vector. Our second model was a modified version of this dynamical system, with an extra term imposing a "maximum velocity" and a third parameter vector of maximum velocities (see equation 2).

To determine the values for these parameters that best fit the data, we conducted a PyTorch learning experiment, training the parameters with gradient descent. We trained the model for 250K epochs, using the Adam optimizer and decreasing the learning rate every 50K epochs. We plotted each model's loss over time to compare the relative success of each model (Figure 9a, 9b).

After training our model, we performed visual and statistical tests on its learned parameters to verify that the model had learned a useful representation of the data. First, we geographically plotted the bias parameter and speed limit parameter for each station (Figure 10a, 10b), as well as the weight parameter for each edge between stations (Figure 10c). Furthermore, we computed the Pearson correlation coefficient between the bias parameter and speed of each station at 5:30 PM "rush-hour" (Figure 10d).

V. Theory

Our first machine learning model, a linear dynamical system on the network, can be expressed as a system of differential equations, one per node. For node i, the corresponding differential equation is

$$\frac{dv_i}{dt} = b_i + w_{ii}v_i + \sum_{j \in N(i)} w_{ij}v_j. \tag{1}$$

In matrix form, this is

$$\frac{d\mathbf{v}}{dt} = \mathbf{b} + \mathbf{W}\mathbf{v}$$

where we define $w_{ij} = 0$ whenever $j \notin N(i)$. However, to remove this restriction on the weights during training, we use the equivalent form

$$\frac{d\mathbf{v}}{dt} = \mathbf{b} + (\mathbf{A} \odot \mathbf{W})\mathbf{v}$$

where $\odot$ denotes the Hadamard product and $\mathbf{A}$ is the adjacency matrix.

Our second ML model is identical to the first except with the presence of an additional term in the differential equation:

$$\frac{dv_i}{dt} = b_i + w_{ii}v_i(c_i - v_i) + \sum_{j \in N(i)} w_{ij}v_j(c_j - v_j) \tag{2}$$

Intuitively, we expect this term to cause the speed at each station node to saturate at some upper value c_i instead of exponentially increasing to infinity (which is a problem with the linear model). In matrix form, we have

$$\frac{d\mathbf{v}}{dt} = \mathbf{b} + \mathbf{W}\mathbf{v} \odot (\mathbf{c} - \mathbf{v}).$$

In order to train the model, we compute the predicted velocity at the next time step, and compare it to the actual velocity at that time step. Mathematically, we define our loss function as

$$\mathscr{L}(b, W, c) = \sum_{t=0}^{N-1} \|\hat{\mathbf{v}}_{b,W,c}^{(t+1)} - \mathbf{v}^{(t+1)}\|^2$$

where

$$\begin{aligned}\hat{\mathbf{v}}^{(t+1)} &= \mathbf{v}^{(t)} + dt\frac{d\mathbf{v}}{dt} \\ &= \mathbf{v}^{(t)} + dt(\mathbf{b} + \mathbf{W}\mathbf{v}).\end{aligned}$$

After running gradient descent, we will obtain an approximation for the optimal values of the parameters (b,W and c) corresponding to minimizing the loss function

$$b^*, W^* = \text{argmin}_{b,W} L(b, W).$$

VI. Results

A. Graph Structure Analysis

From Figure 2, it's immediately clear that our road network has several holes that should be connected by freeways. This is a result of how the dataset was constructed: only 228 stations were randomly sampled from the full 39,000 traffic sensors, so some regions of the road network aren't represented. It's also clear from the figure that the edge connections are much "bushier" than a traditional road network, with a single node have many different neighbors from nearby nodes, whereas a node in a true road network would have at most 3 or 4 connected nodes. This again is a result of the way that the dataset was created, with edges being "guessed" after-the-fact using a distance threshold.

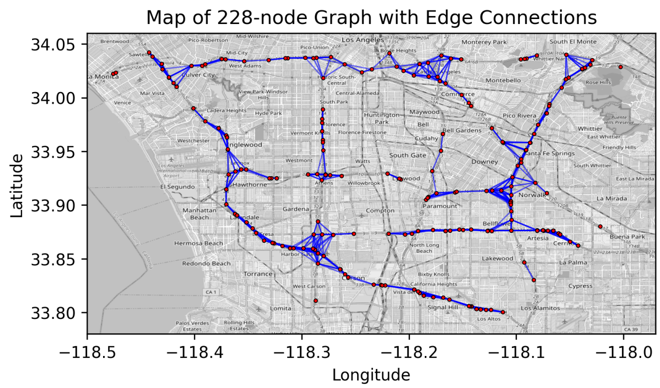


Fig. 2: Diagram of graph connections for the 228-node dataset.

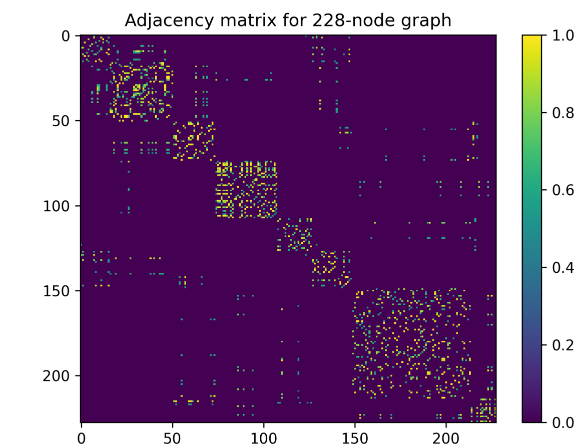
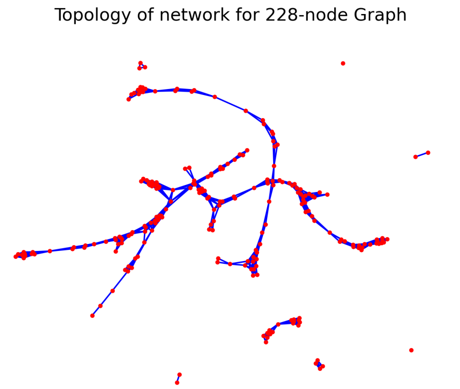
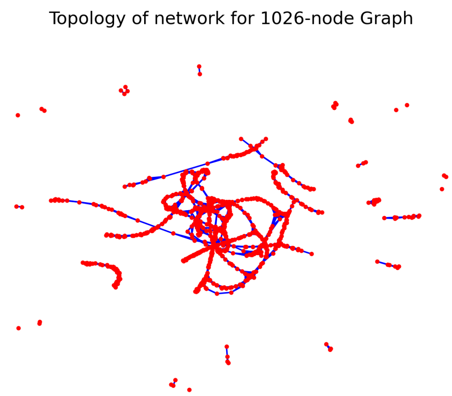


Fig. 3: Adjacency matrix for the 228-node dataset.



(a) Topology of road network in the 228-node dataset.



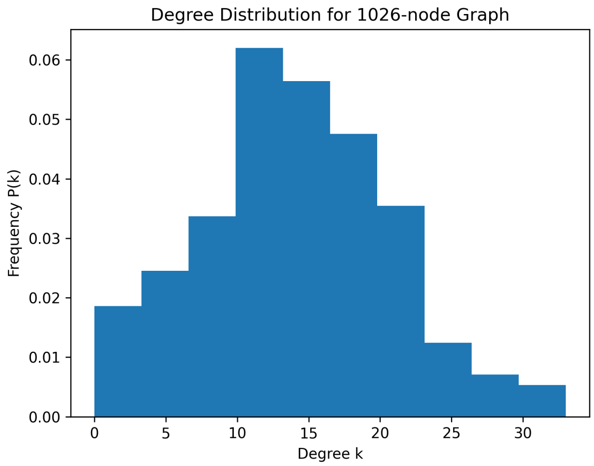
(b) Topology of road network in the 1026-node dataset.

Fig. 4: Topology of networks for each dataset.

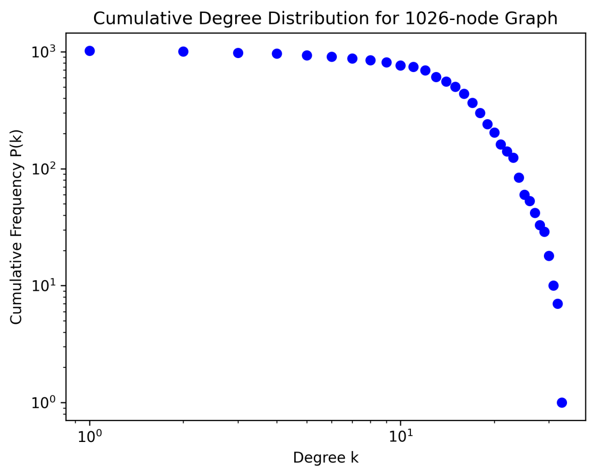
The NetworkX graph plots for the two datasets give us an intuitive understanding of the structure of the graphs. The PeMSD7(M) dataset's graph has 3 main components, each with one intersection. Each of these components likely corresponds to the major highways in the LA region. The remaining components are small, consisting of only a few nodes. The PeMSD7(L) dataset's graph has 1 main component, which is highly connected within itself, and again has many small components. The adjacency matrix for the PeMSD7(M) dataset (Figure 3) shows 6 main "groupings" of nodes, possibly corresponding to the 6 clusters of edges visible in Figure 2.

The degree distribution of the PeMSD7(L) dataset (Figure 5a) provides further evidence that our graph is "bushier" than a traditional network, since the degrees approximately follow a normal distribution centered around $k = 15$. The corresponding cumulative degree distribution (Figure 5b) is not linear when plotted on log-log axes, meaning that the road graph does not follow a power law distribution and so is not scale-free. These facts point possibly point to the artificial nature of our graph, since it was constructed from the distance-thresholding logic mentioned above and not from an actual road map. However, the approximate structure of the road map is captured sufficiently to suit our purposes.

When we computed the various centrality measures of the road graph, we obtained mixed results. Both eigenvector and Katz centrality failed to identify the meaningful intersection points along the road network, and were mostly zero-valued. We geographically plot eigenvector centrality in Figure 6a; Katz centrality was similar. However, both degree centrality and PageRank centrality were successful at identifying central nodes in the road network (Figure 6b, 6c). Between these two, PageRank centrality appears the most reliable at identifying road intersections.

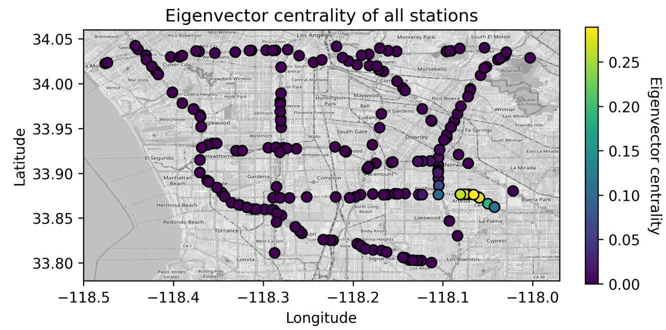


(a) Degree distribution histogram plot.

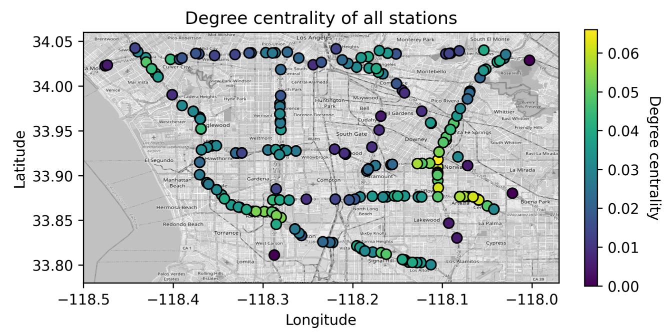


(b) Cumulative degree distribution, plotted in log-log scale.

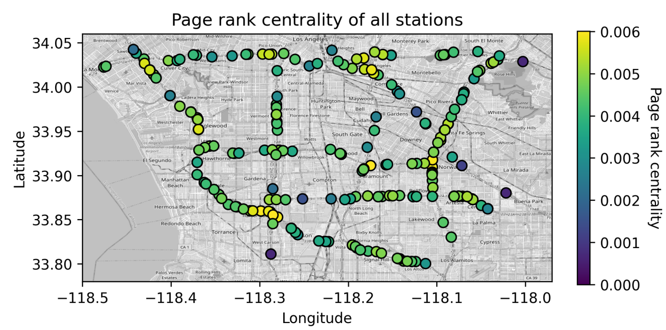
Fig. 5: Degree distribution plots for the 1026-node network.



(a) Geographic map of each node's eigenvector centrality.



(b) Geographic map of each node's degree centrality.



(c) Geographic map of each node's PageRank centrality.

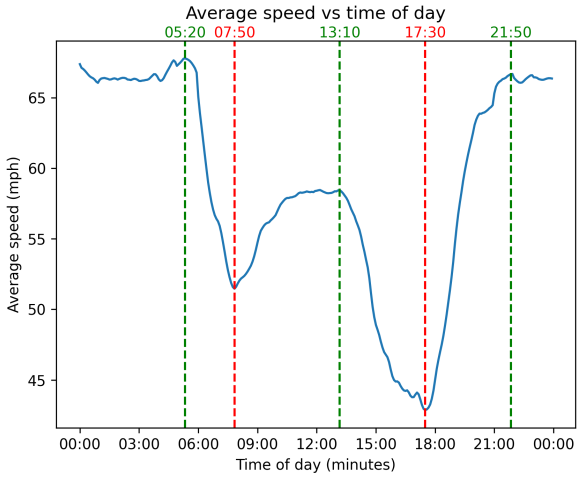
Fig. 6: Maps of various centrality measures for the 226-node network.

B. Traffic Data Analysis

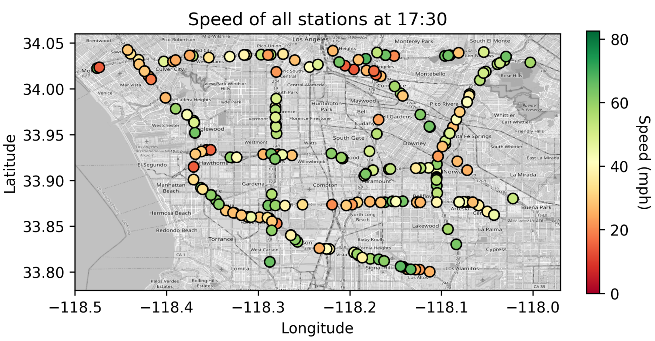
Our plot of average speed versus time of day reveals strong trends in the traffic data for the PeMSD7(M) dataset (Figure 7a). In particular, we see a clear downward spike in average speed around 7:50 AM, a local minimum of average speed (a.k.a. ”morning rush hour“) with average 52 mph. A second, deeper downward spike occurs around 5:30 PM (”evening rush hour“), the global minimum of average speed with average 42 mph. We also see three local maxima of average speed, 5:20 AM, 1:10 PM and 9:50 PM. The local minima correspond to maximal use of the road network by drivers (likely commuting to/from work), whereas the local maxima correspond to minimal use of the road (likely everyone is already at work or at home). Geometrically plotting the speeds of all stations at these critical times of day (Figure **??**, 7c) confirms our understanding that 5:30 PM is traffic-ridden while 5:20 AM is entirely traffic-free.

C. Effect of Graph Structure on Traffic Flow

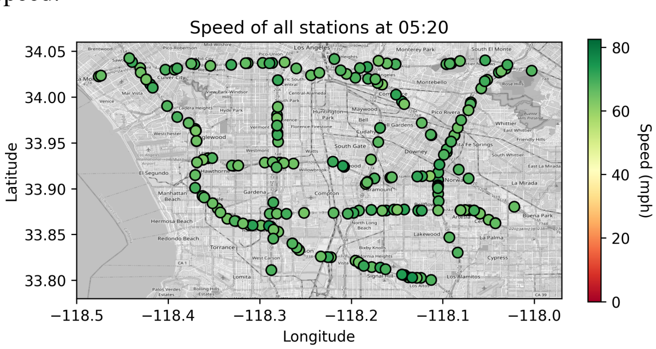
When we plot Pearson correlation coefficient of speed and centrality of each station, computed for each time of day, we see that the correlation between speed and centrality depends on the time of the day (Figure 8a 8b). Specifically, during ”rush hour“ there appears to be a slight negative correlation between centrality and average speed. Intuitively, this would mean that more central nodes (i.e. stations at highway intersections) encounter more traffic during rush hour than other nodes. Conversely, we see that during ”off-peak hours“ speed is slightly positively correlated with centrality. Intuitively,



(a) Average speed of all sensors in the network at a given time of day, averaged across all 44 days in the dataset.

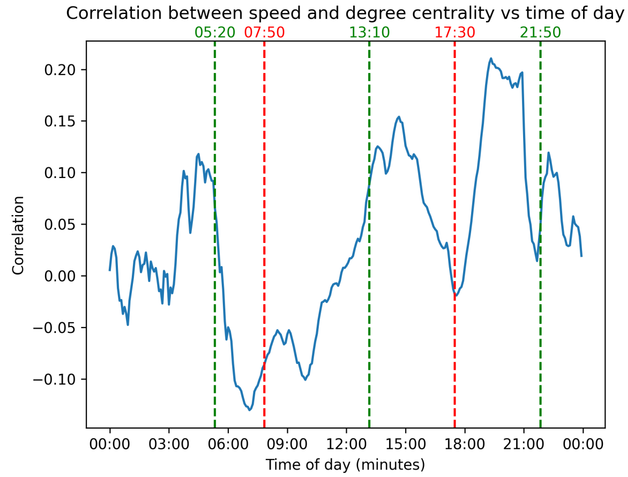


(b) Geographic map of speeds at all stations at 5:30 PM, the time of day corresponding to a global maximum in average speed.

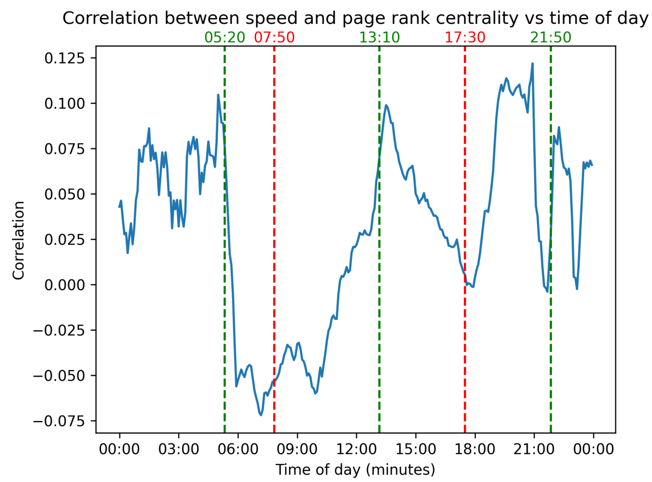


(c) Geographic map of speeds at all stations at 5:20 AM, the time of day corresponding to a global maximum in average speed.

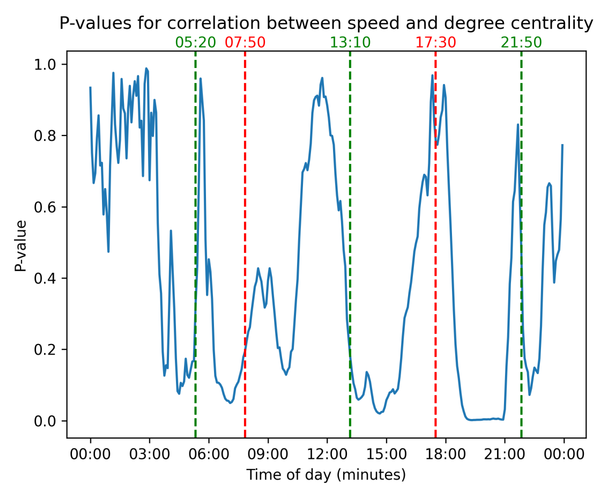
Fig. 7: Speed across time of day.



(a) Pearson correlation coefficient between each station's average speed and degree centrality, computed at each time of day.



(b) Pearson correlation coefficient between each station's average speed and PageRank centrality, computed at each time of day.



(c) The corresponding p-values for the Pearson correlations between speed and degree centrality in subfigure (a).

Fig. 8: Correlation between centrality and speed across time of day.

this means that when there is no traffic the stations at intersections have higher speed.

This is an interesting relationship, however in order to confirm its statistical significance we plotted the corresponding p-values for these correlations (Figure 8c), which varied wildly and were rarely below 0.05. This indicates that these correlations are not statistically significant and can be disgarded. We concluded that there is no correlation between centrality and speed, regardless of centrality measure chosen.

D. Predictive Model

After training our linear model and logistic model for 250K epochs each, they had attained a MSE loss of 2.86×10^3 and 1.22×10^1 respectively across the PeMSD7(M) dataset. Unfortunately, neither of these losses are on-par with the trained models presented in other papers, such as the Spatio-Temporal Graph Convolutional Network [7]. However, we may still gain insight from the two model's learning behavior: The fact that the logistic model was able to achieve such lower loss than the linear model indicates that the logistic model may be in some way a better representation of the "underlying mechanics" of traffic flow.

For the final logistic model, we geographically plot the learned bias for each station (Figure 10a), as well as the learned max speed parameters for each station (Figure 10b). We also geometrically plot the learned weight parameters by coloring the edges between the stations (Figure 10c). Although the learned max speed and weights are nearly constant across the stations and edges and are uninsightful, the biases are clustered together in groups of positive and negative bias. To gain insight in the meaning of these bias values, we plot the correlation between the network bias and the speed at 5:30 PM rush hour (Figure 10d), which reveals a significant positive correlation between bias and speed. When we compute the Pearson correlation coefficient between bias and average speed of the network, we get 0.1516, with a p-value of 0.0220. This represents a statistically-significant correlation, which gives us some confidence that our trained dynamical system learned some meaningful attributes for the traffic data.

VII. Conclusions

Through our research, we found many different properties and behaviors in the traffic network that is represented by our dataset. In terms of graph structure, we found that there was one large connected component with many other smaller components, the degree distribution was approximately normal and the cumulative degree distribution was not scale-free, and using degree and page rank centrality we were able to find the traffic sensors that had the highest connections in our network. For our analysis of traffic data across time of day, we found significant local minimums and maximums in speed that we hypothesize are explained by human and societal behavior. We performed statistical tests for the correlation between of aspects of the graph structure on traffic flow over time, which revealed no statistically significant correlations. Lastly, we created a dynamical systems model that was able to learn meaningful information about the underlying traffic data, although it failed to be an accurate predictor of network traffic flow.

We hope to do further research on the real-world patterns and behavior that this model represents such as understanding what human behavior and societal factors can play a role in causing traffic congestion. We also would like to check if our data is true for our assumptions since we assumed in our analysis that each traffic sensor represents a unique traffic connection. Improvement towards our own model and further research towards our competing research paper's predictive model and testing and comparison to current data and comparison with our results with the actual results.

VIII. Data Availablilty

The datasets we used, PeMSD7(M) and PeMSD7(L), are available online at https://github.com/VeritasYin/STGCN_IJCAI-18/tree/master/dataset.

IX. Contribution Statement

David computed all notable network measurements and resources and created and trained the dynamical systems model on the network. Michael selected the research topic, created the research question, analyzed the majority of the parts of the network visuals and properties, and oversaw the development of the project. Both teammates contributed equally.

References

[1] B. C. Tefft, S. Rosenbloom, R. Santos, and T. Triplett, *American driving survey: 2014 – 2015 - aaa foundation for traffic safety*, Jun. 2018. [Online]. Available: https://aaafoundation.org/american-driving-survey-2014-2015/.

[2] J. Liu, *Commuters in this city spend 119 hours a year stuck in traffic*, Sep. 2019. [Online]. Available: https://www.cnbc.com/2019/09/04/commuters-in-this-city-spend-119-hours-a-year-stuck-in-traffic.html.

[3] C. Gilligan, *The 10 u.s. cities with the worst traffic*. [Online]. Available: https://www.usnews.com/news/cities/articles/10-cities-with-the-worst-traffic-in-the-us.

[4] S. Dorger and D. Diongson, *These are the cities in the u.s. with the worst traffic*. [Online]. Available: https://www.thestreet.com/money/30-cities-with-the-worst-traffic-in-the-us.

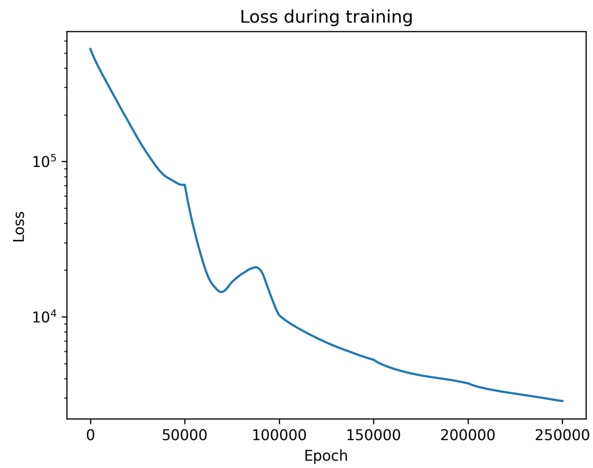
[5] C. Fernandez, *This city has the worst traffic in the u.s.-and it's actually a good thing: "congestion shows the economy is moving"*, Aug. 2023. [Online]. Available: https://www.cnbc.com/2023/08/17/north-america-cities-highest-traffic-delays-inrix-report.html.

[6] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Physical review E*, vol. 51, no. 2, p. 1035, 1995.

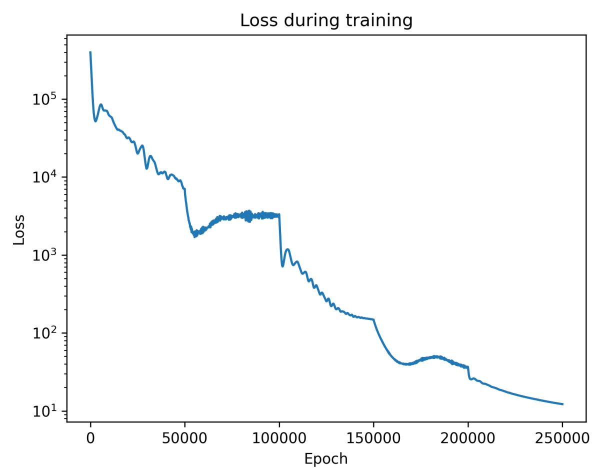
[7] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting," *CoRR*, vol. abs/1709.04875, 2017. arXiv: 1709.04875. [Online]. Available: http://arxiv.org/abs/1709.04875.

[8] *Caltrans performance measurement system (pems)*. [Online]. Available: http://pems.dot.ca.gov/.

X. Appendix: Machine Learning Figures

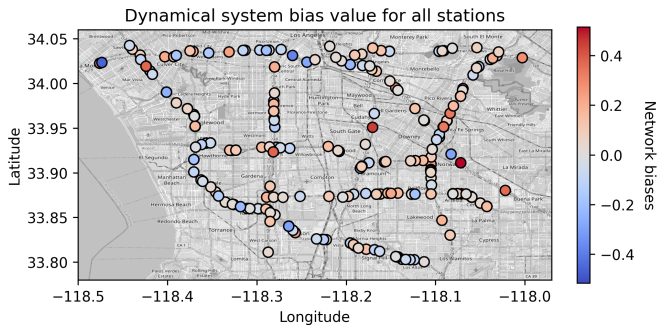


(a) Loss curve for training the linear dynamical system.

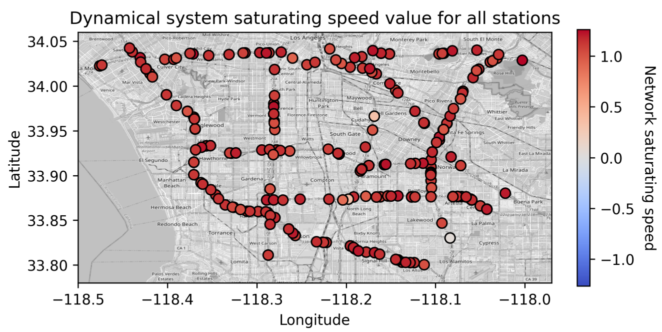


(b) Loss curve for training the logistic dynamical system.

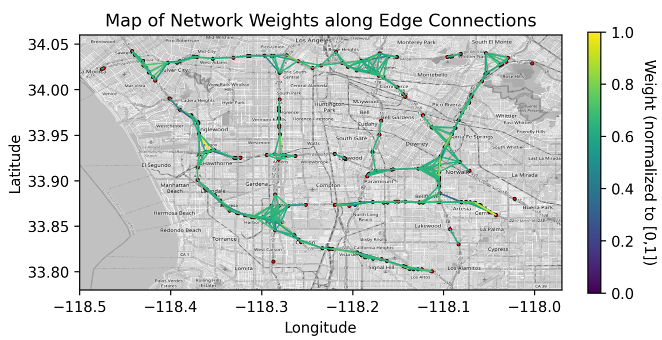
Fig. 9: Loss curves.



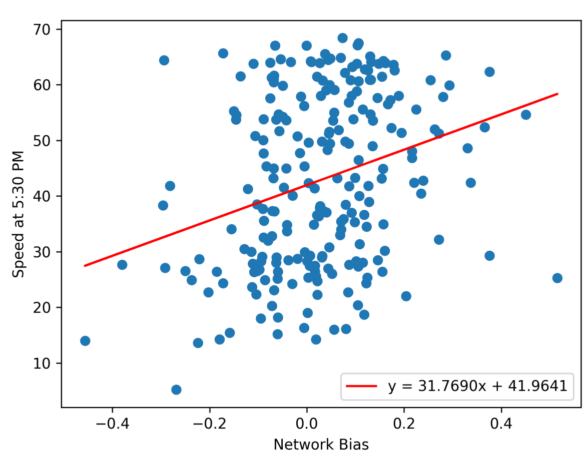
(a) Geographic map of bias parameter learned for each station.



(b) Geographic map of speed limit parameter learned for each station.



(c) Geographic map of weight parameter learned for each edge between stations.



(d) Correlation between speed at ”rush-hour“ (5:30 PM) and learned bias parameters for each station.

Fig. 10: Maps of the learned parameters for our logistic dynamical system.