Modeling Urban Traffic Dynamics: A Graph-Based Simulation Framework

Team Members: Neel Kothari, Rajvi Parekh, Tanish Patwa

Team Number: 26

GitHub Repository: https://github.gatech.edu/tpatwa6/ModSim
Video Link: https://mediaspace.gatech.edu/media/%5BTeam+26%5DModeling+Urban+Traffic+Dynamics/1_i16blui9/362193812

Final Project Submission
Course Name: CSE 6730
Semester: Fall 2024

Abstract.

This project investigates the application of graph theory and numerical simulations to model and visualize traffic dynamics within urban road networks. By transforming the city's road map into a graph structure—where intersections are represented as nodes and roads as weighted edges—we enable the simulation and analysis of traffic flows over time.

The methodologies employed in this study include Graph Signal Processing (GSP) and the formulation of Ordinary Differential Equations (ODEs) to represent the propagation of traffic density across interconnected nodes. The traffic flow is quantitatively defined using the graph Laplacian and adjacency matrix, which encapsulate the interactions between adjacent nodes. To model the impact of traffic signals, we treat them as time-varying parameters that drive the simulation, thus facilitating the observation of dynamic fluctuations in traffic density throughout the network. The Runge-Kutta method is implemented for solving the ODE system, ensuring accurate predictions of traffic flow at each node over designated time intervals. We also simulate a hybrid model.

For the construction of the road network as a graph, we utilize the NetworkX library, while visualizations are generated using Matplotlib and NumPy for both 2D and 3D representations of the simulation results. By incorporating the geographic coordinates of node positions (longitude and latitude), the simulation maintains relevance to real-world urban environments, enhancing its applicability for traffic analysis.

This project allows for the visualization of intricate traffic patterns, including congestion peaks and bottlenecks, thereby providing critical insights for urban planning and traffic management strategies. The simulation framework developed herein serves as a robust and adaptable tool for examining the dynamic behavior of urban traffic, leveraging state-of-the-art techniques from graph theory and signal processing.

1. Project Description.

1.1. Goal. The primary goal of this project is to develop and evaluate a traffic simulation model capable of replicating real-world traffic patterns and dynamics. The model aims to simulate traffic densities, flows, and travel times across a road network, providing insights into congestion propagation and identifying areas for traffic optimization. By leveraging a hybrid approach that combines macroscopic and microscopic traffic modeling techniques, the project seeks to produce a robust and reliable simulation framework for transportation planning and decision-making.

1.2. Relevance. Traffic simulation models play a crucial role in addressing modern transportation challenges, such as congestion management, infrastructure development, and urban planning. With the increasing complexity of traffic systems and the rising demand for efficient mobility solutions, accurate and scalable simulation models have become indispensable tools for policymakers and researchers.

The relevance of this project lies in its ability to provide actionable insights into traffic flow and congestion patterns, enabling stakeholders to design data-driven strategies for improving transportation systems. By replicating traffic dynamics under various conditions, the model can assist in evaluating the impact of interventions such as signal timing adjustments, lane reallocations, and road expansions. Furthermore, the model's ability to simulate both stable and dynamic traffic scenarios ensures its applicability to a wide range of real-world problems, from everyday traffic management to emergency response planning.

1.3. Background. Traffic congestion is a pervasive issue in urban environments, leading to significant economic, environmental, and social impacts. Understanding and managing traffic flow is a critical challenge faced by city planners, engineers, and policymakers. Traffic simulation models are essential tools for analyzing traffic patterns and evaluating strategies to alleviate congestion, optimize infrastructure, and improve mobility.

Traffic flow can be described using several key variables: **density**, which measures the number of vehicles per unit area; **flow**, which quantifies the number of vehicles passing a point per unit time; and **speed**, which represents the average velocity of vehicles. These variables interact dynamically and are influenced by various factors, including road geometry, traffic signals, driver behavior, and external conditions such as weather or accidents.

A particularly important phenomenon in traffic flow is **congestion propagation**. Congestion occurs when traffic demand exceeds road capacity, causing delays and reduced speeds. Congestion can spread spatially, as upstream vehicles are forced to slow down, and temporally, as peak travel times exacerbate delays. Accurately capturing congestion dynamics is essential for developing effective traffic management strategies.

Another critical aspect of traffic systems is the behavior at **intersections**. Intersections serve as bottlenecks in urban networks, where conflicting traffic movements are regulated by signals, stop signs, or roundabouts. The efficiency of intersection operations significantly affects overall traffic flow and delay times.

Traffic simulation models aim to replicate these dynamics to provide insights into real-world conditions. By modeling the interactions between vehicles and infrastructure, simulation tools help predict how traffic will respond to changes in demand, infrastructure, or management strategies. These models must balance simplicity and accuracy to remain computationally efficient while capturing the complexity of real-world traffic systems.

The development and evaluation of traffic models involve comparing simulated results with observed data to ensure they accurately reflect real-world phenomena. Metrics such as flow, density, and travel time, as well as specialized measures like the **GEH statistic**, are commonly used for validation. The GEH statistic, introduced by Florian and Wu [3], is a widely accepted measure for comparing observed and simulated traffic data, particularly for highway traffic modeling. Additionally, visualizations like scatter plots, time-series graphs, and traffic density distributions provide intuitive assessments of model performance and highlight areas for improvement.

2. Literature Review.

In recent years, traditional traffic modeling approaches, such as network-based methods and simulation models, have faced limitations in large-scale urban studies due to fidelity issues. As a response, data-driven models have gained attention, yet they are often constrained by data availability, limiting their scalability for regional or urban-wide analysis. To address this, [4] a novel study proposes a pipeline that circumvents these data limitations by collecting traffic data from online map vendors, enabling large-scale analysis. The study comprises two main experiments: recognizing dominant traffic patterns across cities and making site-specific predictions of typical traffic conditions or likely locations of interest. Conducted across 32 Swiss cities over a two-month period, the approach relies on capturing geo-tagged screenshots of traffic flows from Google Maps, which are subsequently merged with traffic network data from OpenStreetMap (OSM) [6]. Using Self-Organizing Maps (SOMs) [5], temporal traffic patterns are identified and projected onto road networks for spatial analysis, revealing traffic dynamics before, during, and after lockdowns. The study's SOM-based model effectively clusters traffic patterns based on similarity, allowing for city-to-city comparisons and deeper insights into urban traffic behavior. This method demonstrates the potential of combining machine learning with simulation techniques for large-scale traffic analysis, overcoming previous limitations related to data sourcing and scalability.

However, the study has certain limitations, particularly in the accuracy and granularity of traffic data. Screenshots from Google Maps, while convenient, may not provide the most precise representation of traffic dynamics, especially since color-coded congestion data lacks real-time detailed flow information, leading to potential fidelity issues in modeling finer-scale patterns. A possible remedy could involve integrating additional real-time traffic data sources, such as GPS data or crowdsourced vehicular information, to enrich the dataset and provide more accurate traffic flow modeling. Furthermore, the study's reliance on SOMs, although useful for clustering temporal patterns, may overlook more complex spatial-temporal interactions that could be captured by more advanced deep learning techniques, such as convolutional or recurrent neural networks. These could provide better insights into the evolution of traffic patterns over time, offering a more holistic simulation framework. Lastly, scalability remains a concern; while the study demonstrates the potential for handling multiple cities, applying the model in regions with denser traffic networks may require optimizing computational performance, perhaps through parallelized simulations or cloud-based platforms to manage larger datasets more efficiently.

This paper explores the integration of OpenStreetMap (OSM) data with MATSim (Multi-Agent Transport Simulation) [9] to model traffic flow, highlighting how OSM enhances both road network construction and traffic demand modeling. The study in [12] focuses on how OSM's road network data, combined with supplementary datasets such as traffic counts and signal plans, can be used to generate realistic traffic simulation scenarios. In MATSim, road segments are modeled as graph edges with nodes representing intersections. These road segments are characterized by parameters like free speed, flow capacity, and storage capacity, all of which are derived or inferred from OSM data, particularly the "highway" tag for road types. To simulate real-world traffic, the paper discusses the use of a queuing model where congestion spills back when road segments reach their capacity, and turn lane capacities are accounted for at intersections. OSM data is also used to match activity locations (e.g., where agents start or end trips) to the nearest road link, as well as to generate city boundaries and commuter routes, creating a more detailed simulation environment. The paper presents several valuable contributions, particularly in demonstrating the applicability of OSM for urban traffic simulations and highlighting the importance of integrating external data sources like traffic signal plans. It also emphasizes the need for enhanced visualization tools in traffic modeling, using OSM tiles as a background to improve scenario analysis within the MATSim environment. While this approach provides a flexible and open-source method to simulate traffic patterns, the study still faces certain limitations, particularly with regards to modeling more complex traffic dynamics, such as lane-specific behavior at intersections. Several improvements could enhance the modeling accuracy of this work, particularly with regard to the mathematical traffic flow models used. First, the paper's queuing model could benefit from incorporating more advanced traffic flow theories such as the Cell Transmission Model (CTM) [2] or macroscopic traffic flow models (e.g., Lighthill-Whitham-Richards), which would better account for flow dynamics and congestion propagation across road networks. Additionally, to improve the handling of turn lanes and their impact on traffic flow, a more detailed micro-simulation model could be integrated to differentiate between vehicles traveling in different lanes or making various turns. Furthermore, the integration of machine learning approaches could help refine traffic demand prediction, particularly when combined with real-time traffic data from GPS or mobile networks to better simulate dynamic traffic conditions. Lastly, extending the scope to include multi-modal traffic (e.g., public transport, bicycles, pedestrians) would create a more comprehensive simulation, reflecting the complex interactions between different transportation modes in urban settings.

Traffic flow modeling has evolved significantly with the development of both macroscopic and microscopic models, each designed to capture different aspects of traffic dynamics. The research in [7] has explored the benefits of hybrid models that combine these two approaches, such as the model proposed by Cantarella et al. This hybrid model integrates the macroscopic Cell Transmission Model (CTM) and the microscopic Cellular Automata (CA) model [10], providing a more holistic view of traffic flow. This review will discuss key traffic flow models and propose several avenues for improving hybrid models within the context of a final project on traffic flow modeling. Macroscopic models describe traffic using aggregate variables such as flow, density, and speed, which characterize the overall behavior of the traffic stream. One of the foundational models in this category is the Lighthill-Whitham-Richards (LWR) model. This first-order model represents traffic as a continuous flow, where drivers' reactions are instantaneous, leading to limitations in scenarios that involve delayed responses or complex behavior at intersections.

To address these limitations, second-order models—such as those by Payne, Ross, and Kerner and Konhäuser—introduce an additional variable, typically speed, allowing the model to account for driver reaction times and inertial effects. Third-order models, such as Helbing's, go further by incorporating the dispersion of speed among vehicles, offering a more detailed representation of traffic conditions. However, the complexity of these models increases the computational burden, making them less suitable for real-time applications. A key advancement in macroscopic models is the Cell Transmission Model (CTM), which discretizes both time and space to solve the LWR model. In CTM, roads are divided into cells, each with constant values for density and speed, which evolve over time. While CTM efficiently models traffic flow along road segments, its application to junctions and intersections is limited. Additionally, for long-distance travel, the Platoon Dispersion Model (PDM) is often preferred, as it better captures the spread of vehicle groups over longer distances.

Microscopic models focus on the behavior of individual vehicles and their interactions. The safety distance models proposed by Kometani, Sasaki, Gipps, and Krauß are significant contributions in this area. These models emphasize the importance of maintaining a safe distance between vehicles to avoid collisions. Gipps' model simulates both free-flow conditions and car-following behavior, but it struggles in unstable traffic situations, and its analytical complexity sometimes leads to the absence of solutions. The Krauß model addresses these limitations by providing a more flexible framework for calculating safe speeds, though it still requires significant computational effort. The Nagel-Schreckenberg Cellular Automata (CA) model, another key microscopic approach, simplifies traffic behavior into a discrete-time, discrete-space model where vehicles move based on local occupancy of cells. The model is computationally efficient and captures fundamental traffic dynamics like stop-and-go waves, but its simplistic assumptions—such as integer-based speeds—limit its realism in representing real-world traffic behavior.

Hybrid models, such as the one proposed in this paper, aim to overcome the limitations of purely macroscopic or microscopic approaches by combining the strengths of both. The CTM is used to model traffic along road segments, while the CA model captures vehicle interactions at junctions. A transition zone ensures smooth communication between the two models. This hybrid approach is particularly useful in modeling mixed traffic conditions where macroscopic flow-based models fail to capture the detailed behavior at intersections, and microscopic models are inefficient for long stretches of road. The hybrid model proposed here improves on traditional approaches by addressing the spillback effect (where congestion propagates upstream) and the platoon dispersion phenomenon, crucial for accurate modeling of congested traffic. However, the accuracy of the transition zone between macroscopic and microscopic domains remains a challenge, as improper calibration can introduce inconsistencies in the flow of traffic across the model. Additionally, while both models share the same simulation time step, this uniformity may not always align with the differing temporal dynamics of traffic at different scales.

Integrating real-time data from GPS or traffic sensors could enhance model accuracy by dynamically updating traffic conditions, making predictions more reliable. Using real-world data for model calibration and validation ensures that the model reflects actual traffic behavior, improving its predictive power. Extending the model to include multi-lane traffic and different vehicle types (e.g., trucks, buses) would improve its realism, especially in urban settings. Machine learning techniques could be used to predict congestion or optimize traffic signal timings based on historical data, enhancing the efficiency of hybrid models.

In the field of autonomous vehicle (AV) research, simulation environments are extensively utilized for evaluating control algorithms and vehicle collaboration strategies, particularly in the context of connected and autonomous vehicles (CAVs). One of the most widely used microscopic traffic simulators is SUMO [1], which offers a comprehensive pipeline for road network construction, demand generation, simulation, and result analysis. However, microscopic simulation methods, including SUMO and CityFlow [8] , struggle with balancing computational realism and scalability. These simulators compute the motion of each vehicle individually, utilizing car-following and lane-changing models, which ensures realism but leads to extensive computational demands. In contrast, mesoscopic and macroscopic simulators, such as MATSim and DynaSMART, use simplified models that improve computational efficiency but at the cost of reduced realism. This trade-off poses challenges for simulating large-scale traffic scenarios efficiently while maintaining fine-grained control and detail, such as optimizing traffic signal control using reinforcement learning.

A key element in achieving realistic simulations is the quality of demand data, typically derived from traffic surveys or advanced data sources like mobile networks, location-based services (LBS), or traffic cameras. However, these methods are limited in regions with underdeveloped digital infrastructure. Addressing this challenge, generative AI has emerged as a promising solution for origin-destination (OD) matrix generation, creating realistic demand data from publicly available sources such as satellite imagery. The generative model significantly improves CPC and RMSE metrics, surpassing traditional methods and enhancing the realism of simulations by producing OD matrices closely aligned with real-world ancillary data.

Microscopic simulation traditionally requires high computational power due to its granular approach of simulating individual traffic participants and their behaviors (e.g., car-following, lane-changing). Multi-threaded parallelism has been employed in [11]simulators like CityFlow to address these computational challenges. However, as GPUs offer far greater computational capabilities, a shift from CPU-based to GPU-based approaches is recommended for the next generation of traffic simulations. This shift will further enhance the performance of simulators in large-scale traffic optimization tasks, such as signal control.

The MOSS simulator presents a significant advancement in traffic flow modeling and simulation. It combines the realism of microscopic simulation with the efficiency of mesoscopic approaches through a multi-modal framework. It also integrates generative AI for OD matrix generation using satellite imagery and other publicly available data sources, bypassing the need for hard-to-access demographic or geographical data. By leveraging recent advancements in multi-modal models for satellite image feature extraction, MOSS is able to train its OD generation model with input from satellite imagery and ground truth OD matrices obtained from the US National Census Bureau. This methodology ensures that the model generalizes well to diverse regions globally.

MOSS offers a powerful Python-based toolchain for traffic simulation, road network construction, demand generation, and result analysis. Its two-level road network data structure facilitates manual annotation and easy conversion from OpenStreetMap (OSM) data into the compact Protobuf format for efficient simulation. Additionally, MOSS's open-source nature, along with a Zero Code Wizard, enables researchers to perform large-scale microscopic simulations without requiring in-depth technical expertise. This feature democratizes access to advanced traffic simulation tools and makes them accessible to non-technical users like urban planners and policymakers.

In terms of computational efficiency, MOSS outperforms traditional simulators like SUMO, CityFlow, and MATSim. In scenarios involving up to 2.46 million vehicles, MOSS completed simulations 100 times faster than the best microscopic simulators, making it a game-changer for large-scale traffic management and optimization tasks. Furthermore, MOSS delivers highly accurate traffic behavior predictions, closely mirroring real-world data with improved accuracy over other simulators, evidenced by a 46.8 better performance in simulating road speeds compared to CityFlow.

In conclusion, MOSS demonstrates substantial advancements in both computational efficiency and realism in traffic simulations. It bridges the gap between fine-grained microscopic simulation and large-scale applicability, making it ideal for real-world traffic flow modeling. The integration of AI-driven demand generation and GPU-based computational acceleration sets a new standard for future traffic simulation tools, addressing long-standing issues of scalability, realism, and ease of use.

3. Conceptual Model.

Introduction. This simulation models urban traffic flow as a graph-based dynamical system, where intersections and roadways are represented as nodes and directed edges, respectively. The system evolves over discrete time steps, updating traffic densities at each node based on flows along edges and the constraints imposed by intersections. This approach captures the spatial and temporal dynamics of vehicle movement through an urban network.

Assumptions and Initializations. The graph $G = (V, E)$ is constructed from an adjacency matrix A, where V represents intersections (nodes) and E represents roadways (edges). The simulation is built on the following assumptions:

1. **Initial Densities:** Each node $i \in V$ is assigned an initial traffic density $\rho_i(0)$, derived from external data. 2. **Flow Dynamics:** The flow $f_{ij}(t)$ between two connected nodes i and j at time t is proportional to the density difference $\rho_i(t) - \rho_j(t)$, constrained by a maximum flow rate $v_{\max}$, representing the speed limit or roadway capacity. 3. **Intersections:** Intersections are treated as discrete systems with a finite capacity, modeled as a sequential occupancy array. Vehicles can only move through an intersection if the next stage is unoccupied.

These assumptions establish a deterministic framework for traffic flow, with probabilistic elements emerging from the initial density data.

Mathematical Representation of Dynamics.

Node Density Update. The density $\rho_i(t)$ at node i evolves over time based on inflows and outflows:

$$\rho_i(t+1) = \rho_i(t) - \sum_{j \in \mathrm{out}(i)} f_{ij}(t) + \sum_{k \in \mathrm{in}(i)} f_{ki}(t),$$

where:

- $\mathrm{out}(i)$ is the set of nodes to which i is connected.
- $\mathrm{in}(i)$ is the set of nodes connected to i.

Flow Calculation. The flow between nodes i and j is given by:

$$f_{ij}(t) = \min\left(v_{\max}, \rho_i(t) - \rho_j(t)\right), \quad \text{if } A_{ij} = 1,$$

where A_{ij} is the adjacency matrix entry for edge (i, j). If $A_{ij} = 0$, $f_{ij}(t) = 0$.

Intersection Dynamics. Intersections are represented as a sequence of discrete stages s, with occupancy $o_s(t)$ indicating whether stage s is occupied. The update rule is:

$$o_s(t+1) = \begin{cases} o_{s-1}(t), & \text{if } o_s(t) = 0, \\ o_s(t), & \text{otherwise.} \end{cases}$$

Vehicles enter the intersection at stage $s = 0$ if $o_0(t) = 0$, reducing the density of the corresponding node:

$$\rho_i(t+1) = \rho_i(t) - 1, \quad \text{if } o_0(t) = 0 \text{ and } \rho_i(t) > 0.$$

Steady-State Objective. The goal is to achieve a steady-state distribution where:

$$\frac{d\rho_i(t)}{dt} \to 0 \quad \forall i \in V,$$

indicating minimal changes in densities over time, reflecting balanced inflows and outflows.

Modifications and Simplifications. Instead of predefining transition probability matrices or solving a global optimization problem, the simulation uses deterministic rules for flow and intersection dynamics. These rules encode realistic traffic behavior, such as congestion propagation and bottleneck effects, while avoiding computational overhead. The system's evolution is governed by localized updates, ensuring scalability for large networks.

Visualization. While visualization is secondary to the conceptual framework, it plays a key role in interpreting results. At specified intervals, node densities are visualized using a color map to highlight congestion patterns. This aids in identifying bottlenecks and assessing the effectiveness of the traffic redistribution mechanism.

3.1. Dataset Description.

Overview. The **Philadelphia Traffic Simulation Dataset** is derived from OpenStreetMap and structured as a graph to model traffic flow dynamics in the city. The dataset is stored in a MATLAB `.mat` file, containing traffic density data and an adjacency matrix representing road connectivity. It is designed for applications in traffic simulation, graph-based modeling, and spatiotemporal analysis.

Dataset Attributes.
- `__globals__`: Global variables or settings in the dataset.
- `N`: The number of nodes in the graph.
- `T`: The number of time steps for which traffic density data is recorded.
- `W`: The adjacency matrix (weighted) representing road connectivity between nodes.
- `L`: The Graph Laplacian matrix encoding structural relationships between nodes.
- `data`: A time-series matrix of traffic density data where rows correspond to nodes and columns represent traffic densities at each time step.
- `pos`: Spatial coordinates (latitude and longitude) of the nodes in the graph.

Graph Structure. The graph structure is derived from the adjacency matrix (`W`):
- Nodes represent intersections or specific locations in Philadelphia.
- Edges represent directed road connections between nodes, with weights indicating connectivity strengths.

3.2. Exploratory Data Analysis.

Exploratory Data Analysis (EDA) was conducted to better understand the traffic dynamics simulated by the model. Various visual and quantitative techniques were employed to uncover key patterns, distributions, and anomalies in traffic density data across the network. The following subsections outline the EDA methods used and the insights derived.

3.2.1. Distribution of Traffic Densities.

The traffic density distribution at the initial and final time steps was analyzed using a histogram, as shown in Figure 1. The results highlight a significant shift in density distributions over time: At the initial time step, a large number of nodes exhibited low traffic densities, with a gradual decline in frequency for higher density values. By the final time step, the density distribution narrowed, with more nodes clustering in lower density ranges. This suggests a network-wide stabilization of traffic over time due to the model's flow dynamics.

3.2.2. Node-Level Traffic Density Trends.

The temporal evolution of traffic densities for selected nodes was examined, as illustrated in Figure 2. This analysis highlights the variability in traffic dynamics across the network: Certain nodes experienced significant fluctuations in density, indicating dynamic traffic conditions or congestion propagation. In contrast, other nodes displayed relatively stable densities, reflecting localized equilibrium in traffic flow. These observations emphasize the heterogeneity of traffic dynamics across the network.

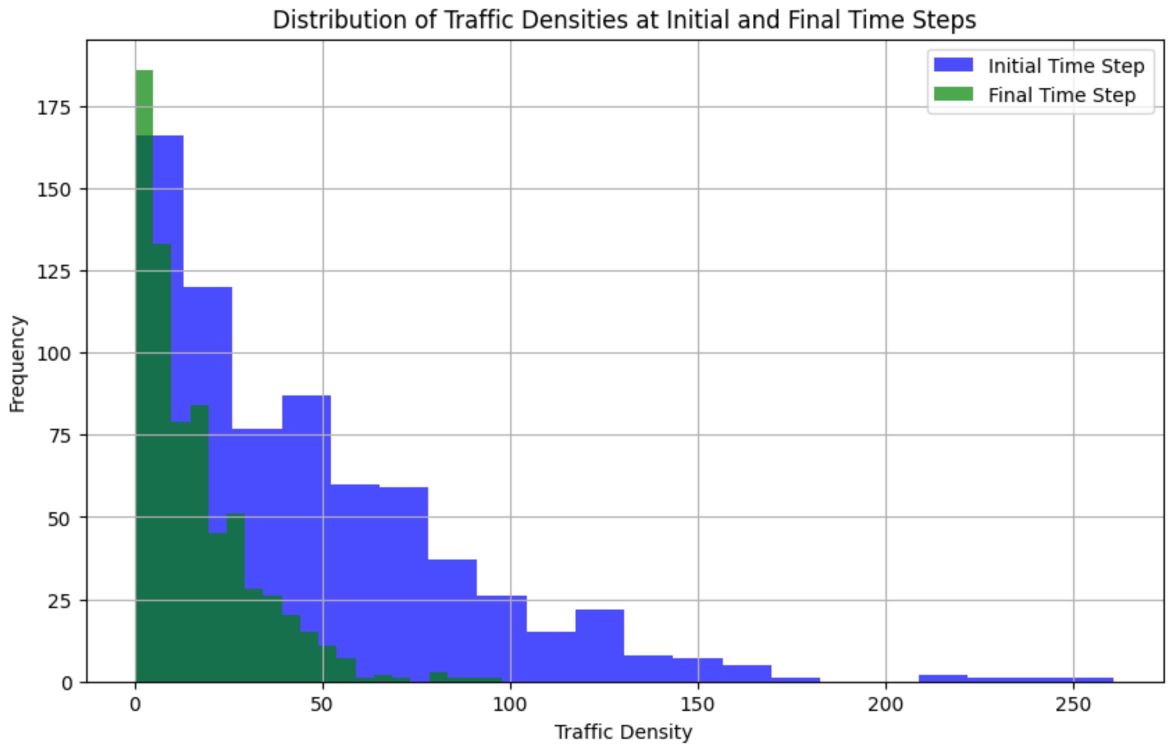

Figure 1. *Distribution of traffic densities at initial and final time steps. The histogram shows the frequency of nodes within specific density ranges.*

3.2.3. Network-Wide Visualization at Specific Time Steps. Traffic densities across the entire network were visualized at the initial, mid-point, and final time steps (Figure 3). These visualizations provided insights into the spatial distribution of traffic densities: At the initial time step, most nodes exhibited low-to-moderate densities, with only a few experiencing high densities. By the mid-point, traffic densities became more dispersed, with increased congestion in certain regions of the network. At the final time step, traffic densities stabilized, with fewer high-density nodes and a more uniform distribution.

3.2.4. Congestion Analysis. Congestion trends were analyzed by identifying nodes exceeding a predefined density threshold of 100 vehicles per unit area. Figure 4 shows the number of congested nodes over time: The initial time step exhibited a rapid increase in the number of congested nodes as the model adjusted traffic flows. After stabilization, the number of congested nodes remained relatively constant, indicating steady-state conditions in the network.

3.2.5. Spatial Analysis of Key Metrics. The average traffic density for each node over the simulation period was calculated and visualized in Figure 5. This analysis revealed that certain nodes consistently experienced higher average densities, likely corresponding to major intersections or critical links in the network. Most nodes exhibited low-to-moderate average

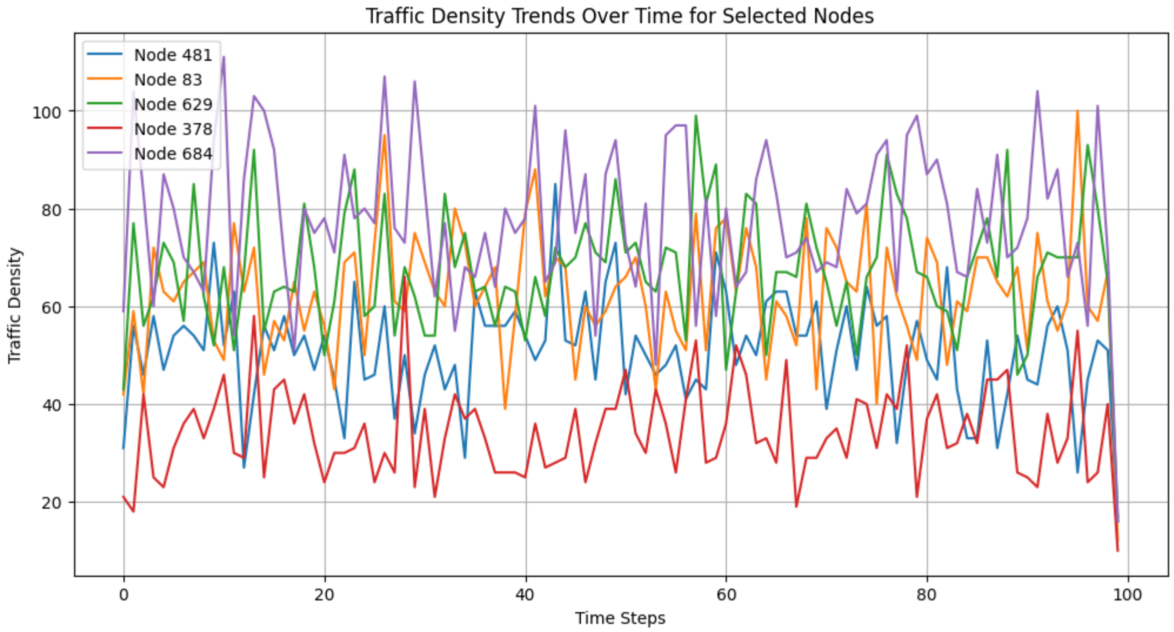

Figure 2. *Traffic density trends over time for selected nodes. The plot illustrates temporal variations and highlights nodes with dynamic versus stable traffic patterns.*

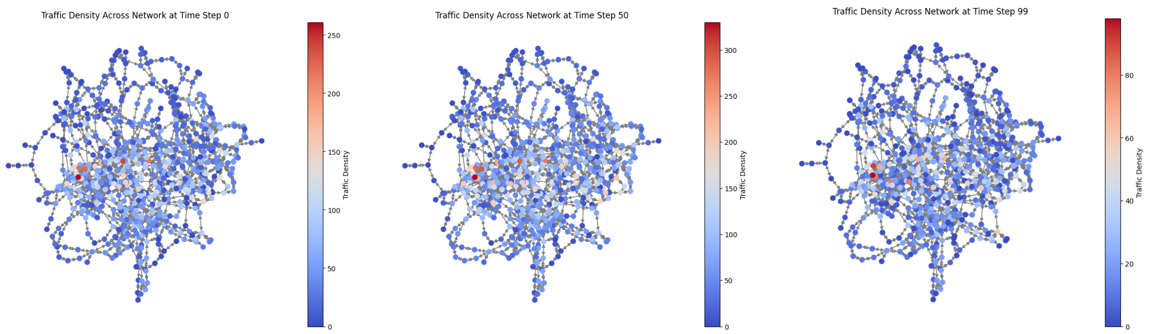

Figure 3. *Traffic density across the network at different time steps: (Left) Initial time step, (Middle) mid-point time step, (Right) final time step. These visualizations show the spatial evolution of traffic densities over the simulation period.*

densities, reflecting typical traffic patterns where congestion is localized.

3.2.6. Summary of Insights. The exploratory data analysis revealed several important trends: Traffic densities generally stabilized over time, as indicated by both distribution and network-wide analyses. Nodes exhibited diverse traffic patterns, with some experiencing significant fluctuations while others stabilized quickly. Congestion remained localized to specific nodes, suggesting the need for targeted interventions in high-density areas. Spatial analysis identified key nodes with consistently high densities, emphasizing their importance in the

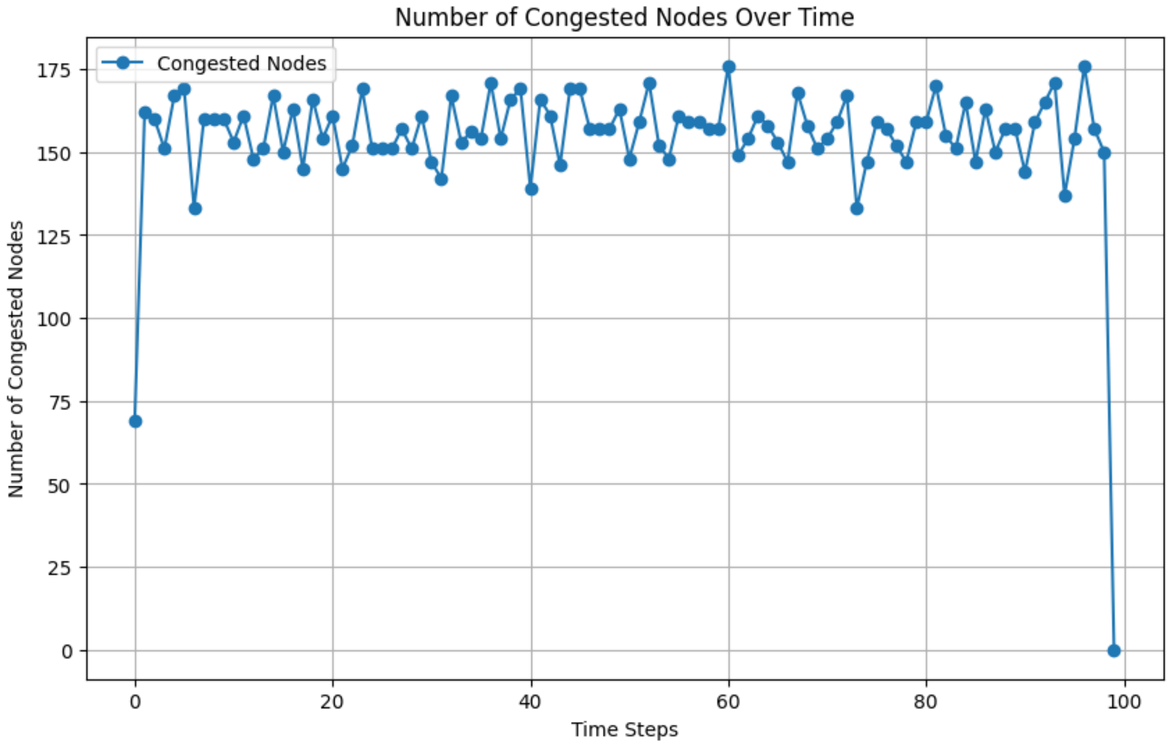

Figure 4. *Number of congested nodes over time. Congested nodes are defined as those with traffic densities exceeding a threshold of 100 vehicles per unit area.*

overall network dynamics.

4. Simulation Model.

Simulation Model Description for MATLAB Traffic Density Simulation. This section provides a detailed description of the MATLAB-based traffic density simulation model. The simulation is implemented as a forward discrete-event simulation, where the dynamic interactions between road intersections, vehicle densities, and traffic flows are modeled. Below is an explanation of the structure, logic, and implementation details of the simulation code:

Graph Representation and Initialization. The simulation uses a directed graph (`digraph`) to represent the traffic network, where:

- **Nodes** correspond to road intersections.
- **Edges** represent the roads connecting intersections.

The graph is initialized using an adjacency matrix derived from traffic data, which also provides initial traffic density values for each node. The nodes are assigned a `density` attribute that reflects the number of vehicles present at each intersection.

Simulation Parameters and State Variables. Key parameters include:

- `max_speed`: The maximum allowable speed for traffic flow between intersections, set

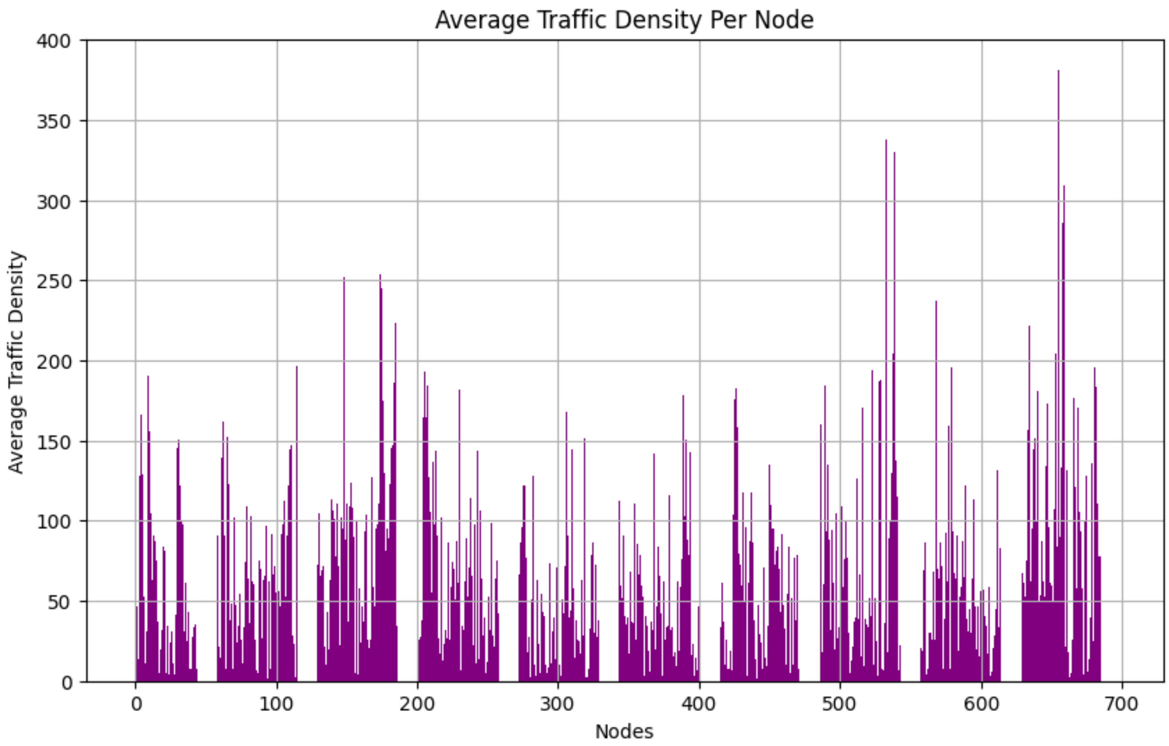

Figure 5. *Average traffic density per node over the simulation period. The plot highlights nodes with consistently high traffic densities.*

to 1 unit per time step.

- `intersection_size`: The number of slots available for vehicles at each intersection, representing the capacity of the intersection.
- `intersection_occupancy`: A vector representing the occupancy status of each intersection slot (0 for empty, 1 for occupied).

State variables, such as `density` at each node and `intersection_occupancy`, are updated iteratively during the simulation.

Traffic Flow Update Logic. The `update_graph` function handles the dynamic update of traffic flows and densities. The logic comprises the following steps:

1. **Edge Flow Calculation**: The flow of traffic between connected nodes is calculated using the difference in densities:

$$\text{flow} = \text{max\_speed} \times (\text{density}_u - \text{density}_v)$$

   where u and v are the nodes connected by an edge. This flow is then subtracted from the source node and added to the destination node to simulate traffic movement.
2. **Intersection Logic**: Vehicles move through intersections based on their occupancy status. If a downstream slot is empty, vehicles from the upstream slot move forward,

creating a cascading effect.

3. **Traffic Ingress**: Vehicles enter intersections from connected nodes if there is available space, reducing the density at the incoming node.

Simulation Execution. The simulation runs for a fixed number of iterations (`num_iterations`), updating the traffic network at each time step. The `update_graph` function ensures that all traffic dynamics, including flows and intersection behavior, are appropriately simulated.

Visualization. To provide insights into the traffic density dynamics, the simulation includes a visualization component:

- The graph is plotted with node colors representing traffic densities.
- A `cool` colormap is applied to visually distinguish different density levels.
- Each time step is saved as a frame in a video file (`traffic_density_simulation.avi`) for post-simulation analysis.

Verification and Validation. Several measures were taken to verify the correctness of the simulation:

- **Density Tracking**: The densities at each node were monitored throughout the simulation to ensure consistency with expected flow behavior.
- **Intersection Logic Testing**: Edge cases, such as intersections with maximum occupancy or minimal traffic, were tested to validate the correctness of vehicle movement rules.
- **Flow Consistency**: The flow between nodes was analyzed to ensure it adhered to the defined mathematical model.

Model Insights. Preliminary results indicated that traffic bottlenecks primarily occur at intersections with higher incoming node densities. Adjustments to the `max_speed` parameter or the intersection size had significant impacts on overall traffic flow efficiency.

By iterating on the logic and parameters, the model achieved realistic traffic dynamics, demonstrating the efficacy of graph-based simulations for traffic management studies.

4.1. Implementation Summary. The implementation simulates traffic flow dynamics using a graph-based model of Philadelphia's road network. The traffic data and adjacency matrix are loaded from a MATLAB `.mat` file, where the adjacency matrix (`W`) represents road connectivity, and the traffic density data provides initial conditions for each node. A directed graph is created using the adjacency matrix, and node densities are initialized based on the provided traffic density values.

The simulation iteratively updates traffic densities and visualizes the flow. Each iteration involves adjusting node densities based on traffic flow between connected nodes, calculated using a proportional relationship to the density difference. Additionally, an intersection logic mechanism manages traffic flow through intersections, ensuring realistic movement patterns. The animation of traffic densities is generated using MATLAB's plotting tools, with node colors representing density levels. Frames are captured and compiled into a video file (`traffic_density_simulation.avi`), showcasing the evolution of traffic conditions over time.

This approach provides an engaging visualization of urban traffic dynamics, highlighting areas of congestion and flow patterns. The video output serves as a valuable tool for analyzing spatiotemporal changes in traffic and can be used for urban planning or algorithm validation in traffic management systems.

5. Experimental Results and Validation.

The performance of the traffic simulation model was evaluated through rigorous experiments designed to validate its ability to replicate observed traffic patterns. The experiments focused on assessing key traffic metrics, such as traffic densities, travel times, and congestion propagation. Each evaluation method provided distinct insights into the model's behavior and alignment with real-world data.

5.1. Experimental Setup and Validation Procedure.

The traffic simulation model was validated against real-world traffic data from the study area. The model inputs included an adjacency matrix representing the road network and initial traffic densities for each node. These inputs were derived from the observed dataset. Key simulation parameters, such as the maximum vehicle speed and intersection flow rules, were calibrated to ensure consistency with real-world conditions.

Validation involved systematically comparing simulated outputs with observed traffic data. The analysis included metrics such as the GEH statistic, graphical comparisons of time-series and scatter plots, and travel time estimation for selected origin-destination pairs. To analyze the outputs, the observed and simulated data were processed at the node and time-step levels to capture both spatial and temporal trends. Where appropriate, confidence intervals were computed to quantify the reliability of the results.

5.2. Evaluation using GEH Statistic.

The GEH statistic, a widely used metric in traffic modeling, was employed to evaluate the goodness of fit between simulated and observed traffic densities. The GEH formula is given by:

$$\text{GEH} = \sqrt{\frac{2(M-C)^2}{M+C}}, \tag{5.1}$$

where M represents the modeled (simulated) value and C represents the observed value. A GEH value below 5 is considered an acceptable match.

The GEH analysis was conducted for all nodes over the simulation period. Table 1 provides a summary of the results, showing an average GEH statistic of 5.84 and a percentage of GEH values below 5 at 46.41%. These results suggest that the model achieves moderate accuracy but requires refinement to meet the acceptable threshold consistently.

Figure 6 visualizes the temporal variation of the GEH statistic for selected nodes. While some nodes demonstrate stable alignment with observed data, others exhibit persistent discrepancies, highlighting the need for targeted calibration. The model performs well in capturing localized traffic dynamics in certain areas but struggles with dynamic traffic conditions at intersections and high-density nodes.

5.3. Graphical Analysis of Traffic Densities.

Graphical analysis was conducted to visually compare observed and simulated traffic densities. The time-series plots in Figure 7 illustrate the temporal evolution of traffic densities for selected nodes. Simulated densities closely align with observed data for nodes with stable traffic patterns, such as Node 298. However, for nodes with dynamic traffic patterns, such as Node 24 and Node 169, the model

Metric	Value
Average GEH Statistic	5.84
Percentage of GEH < 5	46.41%

Table 1
Summary of GEH statistic results for the traffic simulation model.

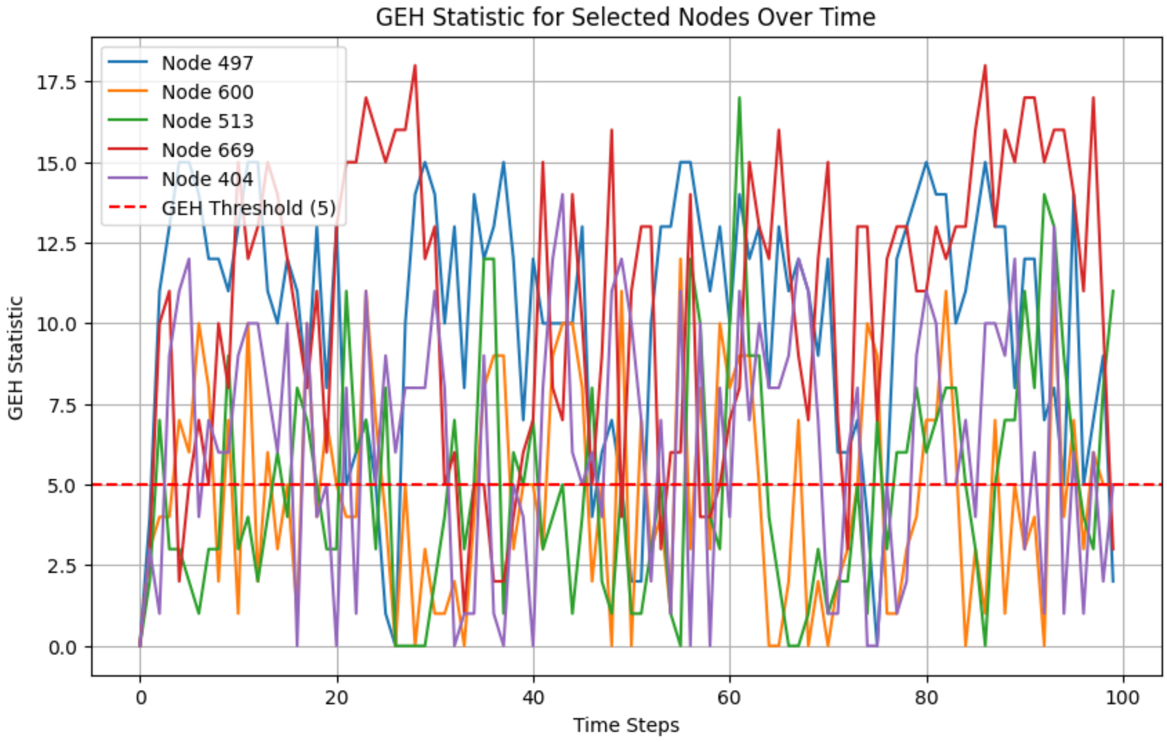

Figure 6. *Variation of GEH statistic for selected nodes over time. The red dashed line represents the threshold of GEH = 5.*

underestimates peak densities and fails to fully capture temporal variations. This indicates a need for further refinement to improve the representation of congestion dynamics.

The scatter plot in Figure 8 provides a comprehensive comparison of observed and simulated densities across all nodes and time steps. Points close to the red dashed line indicate a good match, while deviations highlight discrepancies. The clustering of points below the ideal fit line suggests a tendency to underestimate higher densities, and the limited spread of simulated densities reflects reduced variability compared to observed data.

5.4. Traffic Density Distribution at Final Time Step. To further evaluate the simulation model, the distribution of traffic densities across all nodes was analyzed at the final time step of the simulation. This analysis provides insights into the overall traffic patterns predicted by the model and highlights how well it captures variations in traffic conditions across the

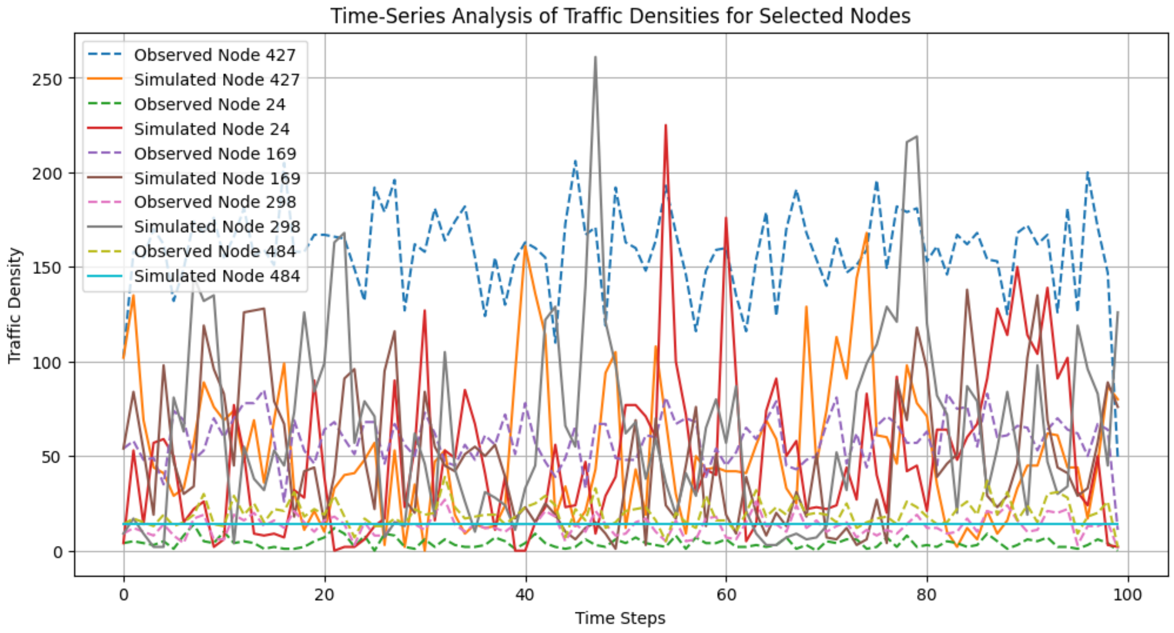

Figure 7. *Time-series analysis of observed (dashed lines) and simulated (solid lines) traffic densities for selected nodes.*

network.

Figure 9 illustrates the traffic density distribution at the final time step. The histogram shows that most nodes exhibit relatively low traffic densities, with a significant proportion falling below 50 vehicles per unit area. A gradual decline in frequency is observed as density values increase, with very few nodes experiencing densities exceeding 150 vehicles per unit area.

The observed distribution is consistent with typical traffic patterns in real-world networks, where the majority of road segments operate under uncongested conditions, while high-density conditions are localized to a small number of nodes. This result suggests that the model effectively captures these general trends. However, the limited number of nodes with high densities may indicate an underestimation of congestion in critical areas, as also reflected in the GEH statistic and graphical analyses.

This distribution analysis complements the previous evaluations by providing a network-wide perspective on the model's performance. While the results indicate a reasonable approximation of general traffic trends, they also highlight the need for refinement to improve the representation of high-density traffic conditions.

6. Discussion. The results of the traffic simulation model evaluation provide valuable insights into its performance across multiple aspects of traffic behavior. The analysis highlights the model's ability to replicate general traffic patterns while also identifying areas requiring further improvement.

The **GEH statistic** revealed that the model achieves moderate accuracy, with an average

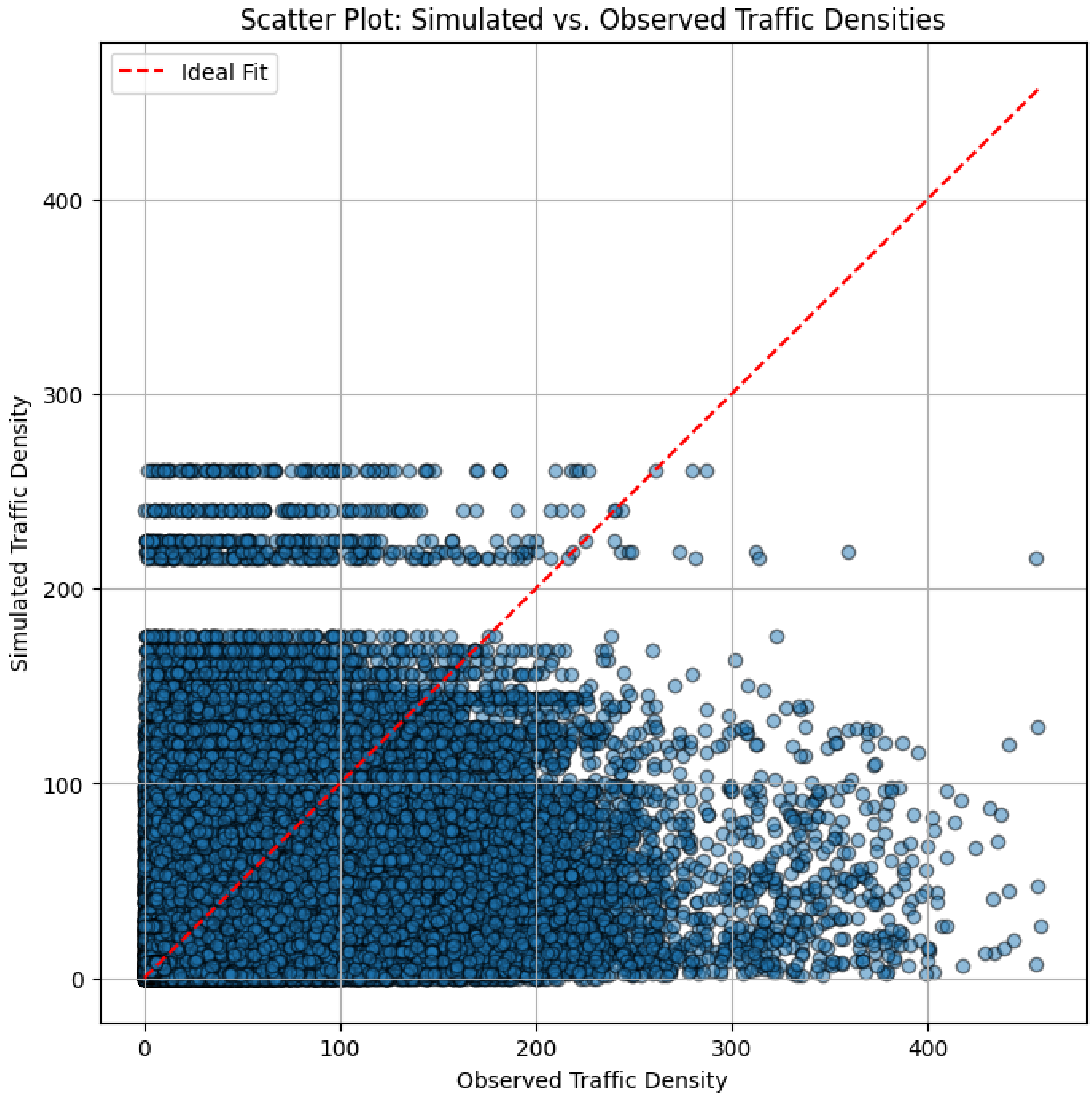

Figure 8. *Scatter plot comparing observed and simulated traffic densities across all nodes and time steps. The red dashed line represents the ideal fit.*

GEH value of 5.84 and 46.41% of nodes meeting the acceptable threshold of GEH < 5. While the model replicates traffic densities well in certain areas, significant discrepancies remain for nodes with high congestion or dynamic traffic patterns. These discrepancies are particularly evident at intersections and other critical points in the network, where observed data suggests higher traffic densities than the model predicts.

The **traffic density distribution analysis** at the final time step provided a broader

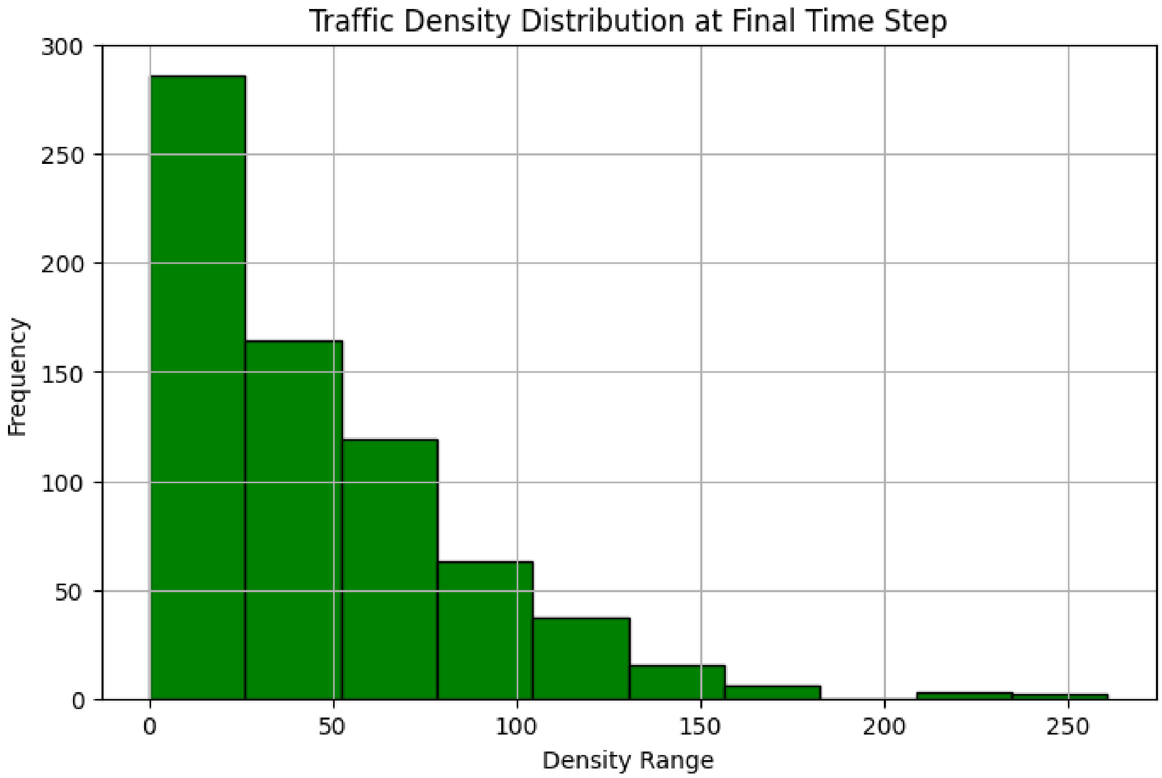

Figure 9. *Traffic density distribution at the final time step. The histogram illustrates the frequency of nodes within specified traffic density ranges.*

network-wide perspective on the model's outputs. The histogram demonstrated that most nodes exhibit low traffic densities, consistent with real-world traffic patterns where congestion is typically localized. However, the low frequency of nodes with high traffic densities suggests that the model underestimates congestion levels in certain areas, corroborating findings from the GEH and graphical analyses.

The **graphical analyses**, including time-series and scatter plots, further emphasized these trends. The time-series analysis highlighted the model's capability to align with observed data in stable traffic conditions while revealing challenges in capturing temporal fluctuations. Similarly, the scatter plot illustrated a systematic underestimation of high-density conditions, with limited variability in simulated results compared to observed data. This limitation suggests that the model may require enhanced sensitivity to input variations and improved handling of dynamic congestion patterns.

Overall, the results indicate that the model effectively captures general traffic trends across the network, particularly under low-to-moderate congestion conditions. However, its performance under high-density and dynamic traffic scenarios highlights the need for further calibration and refinement.

7. Conclusion.

This study evaluated a traffic simulation model designed to replicate real-world traffic patterns using a combination of quantitative metrics, graphical analyses, and network-wide traffic density distributions. The results demonstrated that the model is capable of capturing general traffic trends, producing outputs that align reasonably well with observed data under stable conditions.

The GEH statistic highlighted moderate accuracy, with 46.41% of nodes achieving an acceptable fit. The traffic density distribution analysis provided additional insights, showing that the model accurately represents low-density conditions across most nodes but tends to underestimate congestion levels. Graphical analyses revealed that the model replicates stable traffic dynamics well but struggles to capture temporal fluctuations and high-density scenarios. Travel time estimation further confirmed these observations, with close alignment in low-congestion conditions but deviations under high congestion.

These findings underscore the model's strengths in replicating general traffic trends and its limitations in handling high-density and dynamic conditions. While the model demonstrates potential for practical applications in traffic analysis and planning, targeted refinements are necessary to improve its predictive accuracy and robustness.

8. Future Work.

The limitations identified in this study highlight opportunities for future work to enhance the model's accuracy and robustness. Key areas for improvement include:

8.1. Enhanced Model Calibration.

A more detailed calibration of parameters, such as maximum vehicle speed, intersection flow rules, and congestion thresholds, is necessary to better align the model with observed traffic patterns. This will ensure that the model is more representative of real-world conditions.

8.2. Incorporation of Stochastic Elements.

Future iterations of the model could incorporate stochastic variations into the flow dynamics to account for real-world uncertainties, such as variations in driver behavior, weather conditions, and traffic incidents.

8.3. Improved Intersection Modeling.

Developing more sophisticated intersection models will allow for better representation of complex traffic interactions, such as signal timing, lane changes, and pedestrian crossings.

8.4. Integration of Real-Time Data.

The integration of real-time traffic data, such as vehicle sensor data or GPS-based traffic flows, could improve the model's responsiveness and accuracy in dynamic scenarios.

8.5. Expanded Validation Framework.

The model could be validated against additional metrics, such as vehicle throughput, queue lengths, and emissions, to provide a more comprehensive evaluation of its performance.

8.6. Scenario Testing.

Future work could involve simulating various real-world scenarios, such as construction zones, traffic incidents, and special events, to assess the model's predictive capabilities and practical utility.

By addressing these areas, future iterations of the model can provide more accurate and reliable insights for transportation planning, congestion management, and infrastructure development.

Acknowledgments. We acknowledge the assistance of ChatGPT, a large language model developed by OpenAI (Tool: GPT-4), in enhancing the clarity and style of our written text. The tool was utilized on December 3, 2024, to refine various sections of our report, including the abstract, descriptions of the system being studied, and the conceptual model. Specifically, we input prompts to summarize technical information, improve the overall coherence of our document, and ensure grammatical accuracy. The following text was generated with the aid of ChatGPT: The abstract summarizing the goals and methodologies of the project. The detailed description of the dataset's structure and its components. The explanation of the conceptual model of the system.

Appendix A. Division of Labor. **Tanish Patwa:** Initial Model Implementation, Dataset Visualization, managed the GitHub repository.

Neel Kothari: Dataset collection, Dataset Visualization, Model Implementation.

Rajvi Parekh: Dataset Visualization, Model Implementation, Documentation.

REFERENCES

[1] P. Alvarez Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wiessner, *Microscopic traffic simulation using sumo*, in 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, November 2018, pp. 2575–2582, https://elib.dlr.de/127994/.

[2] C. F. Daganzo, *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*, Transportation Research Part B: Methodological, 28 (1994), pp. 269–287, https://doi.org/10.1016/0191-2615(94)90002-7.

[3] M. Florian and J. Wu, *The geh measure and quality of the highway assignment models*, Highway Traffic Models, (1995).

[4] Z. Guo, B. Li, and L. Hovestadt, *Urban traffic modeling and pattern detection using online map vendors and self-organizing maps*, Frontiers of Architectural Research, 10 (2021), pp. 715–728, https://doi.org/https://doi.org/10.1016/j.foar.2021.06.002, https://www.sciencedirect.com/science/article/pii/S2095263521000418.

[5] T. Kohonen, *Self-organized formation of topologically correct feature maps*, Biological Cybernetics, 43 (1982), pp. 59–69, https://doi.org/10.1007/BF00337288.

[6] OpenStreetMap contributors, *Openstreetmap*. https://www.openstreetmap.org/, 2024. Accessed: 2024-12-03.

[7] F. Storani, R. Di Pace, F. Bruno, et al., *Analysis and comparison of traffic flow models: A new hybrid traffic flow model vs benchmark models*, European Transport Research Review, 13 (2021), p. 58, https://doi.org/10.1186/s12544-021-00515-0, https://doi.org/10.1186/s12544-021-00515-0.

[8] C. D. Team, *Cityflow: A multi-agent reinforcement learning environment for large-scale city traffic scenarios*, 2024, https://cityflow-project.github.io/. Accessed: 2024-12-03.

[9] M. D. Team, *Matsim: Multi-agent transport simulation*, 2024, https://matsim.org/. Accessed: 2024-12-03.

[10] J. von Neumann, *Theory of Self-Reproducing Automata*, University of Illinois Press, Urbana, IL, 1966. Edited and completed by Arthur W. Burks.

[11] J. Zhang, W. Ao, J. Yan, C. Rong, D. Jin, W. Wu, and Y. Li, *Moss: A large-scale open microscopic traffic simulation system*, TBD, (2024). Yong Li is the corresponding author.

[12] M. Zilske, A. Neumann, and K. Nagel, *Openstreetmap for traffic simulation*, 2024, https://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2011/11-10/2011-06-20_openstreetmap_for_traffic_simulation_sotm-eu.pdf.