A Data-Driven Approach for the Analysis of Ridership Fluctuations in Transit Systems

Jovan Pavlović
jovan.pavlovic@famnit.upr.si
University of Primorska
FAMNIT,
Koper, Slovenia

Miklós Krész
miklos.kresz@innorenew.eu
InnoRenew CoE,
Izola, Slovenia

László Hajdu
laszlo.hajdu@famnit.upr.si
University of Primorska
FAMNIT,
Koper, Slovenia

András Bóta
andras.bota@ltu.se
Luleå University of Technology
Luleå, Sweden

ABSTRACT

This study focuses on identifying critical components within urban public transportation networks, particularly in the context of fluctuating demand and potential pandemic scenarios. By employing advanced agent-based simulations, we analyzed passenger interactions and ridership patterns across the San Francisco Bay Area's transit system. Key findings reveal specific transit stops and routes that are highly sensitive to changes in demand, often serving as bottlenecks or high-risk areas for the spread of infectious diseases.

KEYWORDS

network modeling, public transportation, agent-based simulation, community detection, demand fluctuations

1 INTRODUCTION

Efficient public transportation is vital for urban mobility, economic productivity, and public health. During the COVID-19 pandemic, transit systems worldwide were dramatically affected, resulting in a significant decline in ridership due to lockdowns, social distancing measures, and the shift to remote work [1, 6]. Physical distancing, a widely used non-pharmaceutical intervention to prevent the spread of the virus, further reduced the capacity of public transportation services, limiting their ability to meet demand [13].

Factors such as population growth, economic conditions, and environmental policies can also cause fluctuations in public transportation usage. Understanding these changes is crucial for planning resilient and efficient transit systems that can adapt to the evolving needs of cities.

Adjusting service frequency during peak and off-peak hours allows for more efficient use of resources and helps maintain service levels that meet demand without overloading the system. Additionally, rerouting or introducing new transit lines in underserved areas can improve accessibility and attract more users, or conversely, in the case of a pandemic, these changes can discourage usage to help manage public health risks. Infrastructure updates, such as upgrading stations for better crowd flow can also help transit systems adapt to changes. Therefore, it's important to identify the parts of the transit system that are most affected by changes in ridership to develop these strategies effectively.

This research uses agent-based simulations to analyze passenger interactions within transit systems. While networks traditionally depict routes and stops, improved data collection now allows tracking individual passenger interactions. Smart card data [12] and activity-based travel models [10, 11] capture detailed passenger contact patterns. However, creating accurate real-world contact networks from this data poses challenges, including computational complexity and privacy issues [4, 5].

We used activity-based travel demand models to simulate probable traveler paths in transit networks, considering demand, supply, and service details. These models are complemented by schedule-based transit assignment models, which provide accurate estimates of travel time and waiting times. We analyzed the outputs of transit assignments, considering transit route usage, congestion, and waiting times at transit stops, to identify critical components of the transit network that could be potentially affected by changes in transit demand. Additionally, we processed this data to generate contact networks. We then applied a modularity-based community detection algorithm to extract non-overlapping communities of passengers from the contact network and used these communities to further analyze critical bus routes used by different communities.

2 BACKGROUND

This work is inspired by the methodologies used in previous studies [2, 3, 7]. However, rather than explicitly modeling the spread of disease to identify high-risk transit components, it focuses on examining the components most likely to be affected by changes in ridership trends due to a pandemic or other scenarios.

The contribution of this work is to develop a framework that identifies critical components in terms of factors like changes in transit demand, vehicle capacities, and transit schedules. The insights derived from this framework can be further utilized for modeling transit operations in these scenarios.

3 METHODOLOGY

3.1 Transit simulation model

We used a schedule-based transit assignment model, FAST-TrIPs [8], to simulate passenger movement within the transit network. This model's time-dependent structure captures daily service variability

and focuses on specific transit vehicle trips, which is crucial for accurately reflecting passengers' route choices based on the service schedule. FAST-TrIPs operates on a transit network composed of nodes that represent stops. Trips are connected to specific routes within this network, and transfer links connect nodes where passengers can change vehicles. This setup allows for precise modeling of both vehicle movements and passenger transfers across the transit network.

At the heart of FAST-TrIPs is the Transit Hyperpath Algorithm, which constructs a subnetwork of probable transit routes and assigns probabilities to these routes using a logit route choice model. The algorithm calculates hyperpaths by considering user-preferred arrival times and waiting time windows, allowing for the simulation of passenger journeys with a focus on real-time decision-making and path selection. Passenger movement is then modeled using a pre-estimated route choice model that incorporates factors such as in-vehicle time, waiting time, walking time (for access, egress, and transfers), and transfer penalties.

The transit assignment model generates detailed outputs, such as vehicle load profiles and passenger trajectories. The load profile provides information on the number of passengers boarding and alighting at each stop, along with timestamps, offering insight into passenger counts throughout the route. Passenger trajectories document each passenger's activities, including stop and vehicle IDs with timestamps, enabling the modeling of interactions between passengers.

3.2 Input data

FAST-TrIPs requires various input files, including transit system data stored in GTFS-PLUS format, and transit demand data that contains information about the trips individual passengers make throughout the day, including trip origins, destinations, and preferred arrival times. Additionally, path weights associated with in-vehicle time, waiting time, walking time, and transfer penalties must be specified as input.

In the current study, we used GTFS-PLUS data [1] from the San Francisco Bay Area in California from 2017, which includes 854 routes (covering bus, heavy rail, light rail, and ferry routes) and 36,058 trips serving 6,181 stops over a 24-hour weekday. On the demand side, we used data generated in the same year using the SF-CHAMP travel forecasting tool.

Since calibrated path weights were not available for the Bay Area network, we borrowed path weights from a previous study [9] corresponding to the Austin, Texas region.

3.3 Contact network

As mentioned previously, FAST-TrIPs outputs detailed passenger trajectories that can be further processed to produce a contact network. In this network, each passenger traveling within the transit system is considered a node, and edges connect any two passengers who share a vehicle trip for a positive time period. The vehicle trip refers to a specific route with a specific departure time and is unique to a single vehicle. Each edge is associated with three attributes: the contact start time, contact duration (in seconds), and the vehicle trip ID

[1] https://mtcdrive.app.box.com/s/3i3sjbzpsrbhxlwpl4v4vx9b0movferz

3.4 Community detection algorithm

We used the Clauset-Newman-Moore greedy modularity maximization algorithm [3] to find the community partition of the contact network with the highest modularity. This community detection algorithm is a hierarchical agglomeration method designed to efficiently identify community structures within large, sparse networks. Unlike traditional methods, which can be computationally expensive, this algorithm operates in a time complexity of $O(md \log n)$, where n is the number of vertices, m is the number of edges, and d is the depth of the dendrogram describing the community structure. For many real-world networks, which are sparse and hierarchical (with $m \sim n$ and $d \sim \log n$), the algorithm runs in nearly linear time, $O(n \log^2 n)$.

3.5 Limitations

The primary limitation of this study is the size of the demand data. Although the GTFS data originates from a transit network serving millions daily, computational constraints prevented us from simulating real-world demand accurately. Consequently, train routes were not filled beyond half capacity, making it impossible to realistically assess the effects of demand changes on the trains. Additionally, the dataset contains outdated transit system and demand information. However, the proposed method serves as a proof of concept and can be directly applied to more comprehensive travel datasets.

Another limitation is the lack of a detailed comparative study with state-of-the-art methodologies that aim to achieve similar objectives. This choice was due to space constraints, but future research will expand on this comparison, with findings to be published in a full-length journal paper.

4 RESULTS

4.1 Model outputs and contact network

Due to computational limits, the simulations used a reduced number of iterations to reassign passengers to alternative routes. Despite this, most passengers (41,845 out of 44,912) successfully reached their destinations, resulting in 83,280 completed trips.

Figure 1 presents a boxplot of average waiting times, aggregated by passenger, transit route, and transit stop.

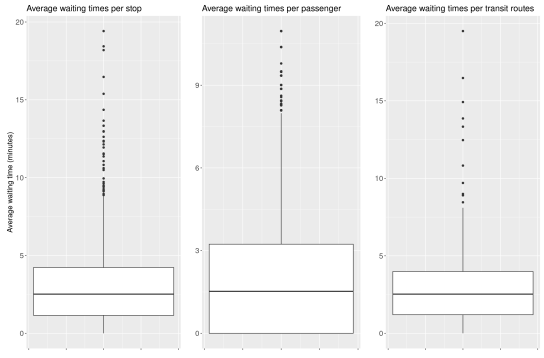

Figure 1: Boxplots of average waiting times

The derived contact network consisted of 41,845 passenger nodes and 3,530,995 contact links. The density plot of contact start times,

displayed in Figure 2, peaked at 7 AM and 5 PM, reflecting typical weekday commutes. The average contact duration was 18 minutes and 43 seconds. Figure 4 shows the density plot of contact durations.

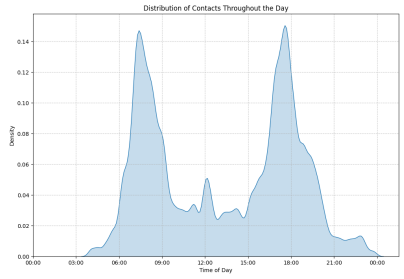

Figure 2: Density plot of contact stat times

The degree distribution of the contact network, shown in Figure 3, indicates an average of 134 contacts per person, with a maximum of 1,011, following a skewed power law distribution.

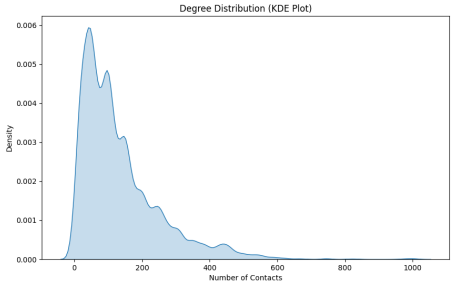

Figure 3: Degree distribution

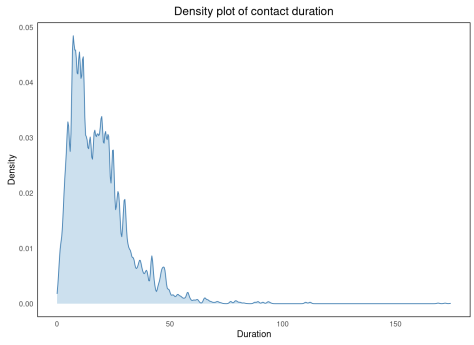

Figure 4: Density plot of contact duration

4.2 Identifying critical components

We first aimed to identify transit stops that may be sensitive to changes in demand. These stops are characterized by two key properties: they serve sufficiently large groups of people, and the average waiting times at these stops are longer than those at most other stops in the transit network. The idea is that such stops could become critical in scenarios where transit demand increases, potentially turning them into bottlenecks. Additionally, in epidemiological situations, passengers waiting at these stops might face an increased risk of infection. To focus on the most relevant stops, we filtered out those serving fewer than 100 people and sorted the remaining stops based on average waiting time. Table 1 provides information on the 10 stops with the longest average waiting times. Most of the listed stops are served by multiple bus routes and have between 100 and 200 passengers waiting at them throughout the day.

In order to identify critical transit routes we took two approaches. Firstly we identified routes whose vehicle trips are on average most congest. Due to limited transit demand here we focused only on bus routes. Table 2 shows 10 most critical bus routes identified in this way.

The second approach involved identifying critical trips with respect to the community structure of the contact network. Community detection algorithm divided the network into 627 communities, with the largest 10 containing 37% of all passengers in the network. We then identified transit routes used by passengers who appear in at least two of these ten communities and ranked the routes by the number of communities whose passengers travel on them. Table 3 shows the ten most critical routes identified in this way. As can be observed, all of the identified bus and trolleybus routes belong to the San Francisco Municipal Railway (SF Muni) system operating in San Francisco.

Figures 5 and 6 summarize the obtained results. Critical stops are marked in red, bus routes used by multiple communities are colored in green, and the most congested bus routes are marked in blue.

As observed, the majority of the most congested bus routes connect different cities within the Bay Area or link various cities to San Francisco. For example, several of these routes travel between Contra Costa and Alameda counties, as well as between San Mateo and Alameda counties. Additionally, some routes connect Berkeley and San Francisco, while many others link San Mateo County, Santa Clara County, Marin County, and Petaluma in Sonoma County to San Francisco. Most of the critical bus stops are concentrated in San Francisco, with several others located in the centers of various cities in the Bay Area, including Berkeley, Oakland, San Jose, and Palo Alto. As previously noted, the bus and trolleybus routes connecting different communities that commute in the Bay Area belong to the SF Muni system operating within San Francisco.

5 CONCLUSION

Using agent-based simulations and network analysis techniques, we identified transit stops and routes that are most vulnerable to changes in demand, whether due to a pandemic or other social and economic factors. Our findings show the importance of focusing on crowded routes and stops with long wait times, as these are likely to become bottlenecks when demand increases. The application of community detection to passenger contact networks further reveals how interconnected different transit routes are within major urban areas, emphasizing the importance of certain routes in keeping public transportation running smoothly.