

# Multisensor Fusion and Graph Neural Networks for Predictive Crowd Flow Management in Large-Scale Public Transportation Hubs

Elias Duarte

Assistant Professor, Dept. of Computer Science, Universidade de São Paulo, São Paulo, Brazil

\* **Corresponding Author:** [eduarte@usp.br](mailto:eduarte@usp.br)

## ARTICLE INFO

Received: 10 Feb 2025

Accepted: 27 Apr 2025

## ABSTRACT

Managing crowd flow in large-scale public transportation hubs has become increasingly critical with rising urbanization and passenger density. Traditional surveillance and control systems, which rely on manual monitoring or isolated sensor inputs, often fail to provide timely insights for proactive congestion prevention. This study proposes a multisensor fusion framework integrated with Graph Neural Networks (GNNs) to predict and manage crowd dynamics in real time. By combining video surveillance, Wi-Fi/Bluetooth mobility tracking, and smart ticketing data, the system constructs a unified spatiotemporal representation of passenger movement. Each zone within a transportation hub is modeled as a node in a dynamic graph, with edges denoting interzonal pedestrian flows. The GNN learns spatial dependencies and temporal patterns from these fused data streams to forecast short-term crowd densities and identify potential congestion points. Experimental evaluation demonstrates that the fusion-GNN model outperforms traditional single-sensor and non-graph baselines in accuracy and early congestion detection. The framework enables proactive crowd management, enhancing passenger safety, optimizing operational efficiency, and contributing to the development of intelligent transportation systems. This research provides a scalable, data-driven solution for next-generation smart mobility infrastructure, bridging multisensor data analytics with AI-driven decision-making for safer and more efficient transit environments.

**Keywords:** Multisensor Fusion, Graph Neural Networks (GNN), Crowd Flow Prediction, Smart Transportation Hubs, Congestion Prevention, Passenger Safety, Video Analytics.

## INTRODUCTION

### Background

Rapid urbanization and the expansion of public transportation networks have led to an unprecedented increase in passenger flow within large-scale transit hubs such as metro stations, airports, and intercity bus terminals. These facilities often handle hundreds of thousands of passengers daily, creating complex spatiotemporal flow patterns that vary across time and location. During rush hours, emergencies, or special events, passenger congestion can escalate rapidly, resulting in operational inefficiencies, safety hazards, and poor user experience. Traditionally, crowd flow management has relied on manual monitoring, static signage, or simple rule-based control systems that respond after congestion has already occurred. However, such reactive strategies often fail to prevent crowding in real time due to delays in human perception, data fragmentation, and limited predictive capability.

In recent years, the integration of Internet of Things (IoT) technologies, computer vision, and artificial intelligence (AI) has revolutionized the monitoring and management of public spaces. Modern transportation hubs are equipped with various sensors—such as CCTV cameras, Wi-Fi or Bluetooth access points, and smart ticketing systems—that collectively generate large volumes of data. These multisource datasets provide diverse perspectives of passenger behavior: video analytics capture visual density and motion, Wi-Fi signals trace device

movements, and ticketing logs record entry and exit frequencies. The challenge, however, lies in integrating these heterogeneous data streams into a unified analytical framework capable of accurately predicting crowd dynamics before critical congestion events occur.

### **Problem Statement**

Despite advancements in smart monitoring technologies, current systems largely operate in silos, analyzing data from individual sensors independently. Video surveillance alone cannot capture hidden areas or account for signal noise, while Wi-Fi-based monitoring may overestimate counts due to multiple connected devices per person. Similarly, ticketing data provide only coarse-grained temporal information without spatial context. The lack of data fusion and predictive intelligence results in poor situational awareness and delayed responses to emerging crowd pressures. Moreover, most predictive systems fail to consider the spatial interdependence among different zones in a transportation hub—such as how passenger buildup at ticket gates influences flows toward platforms. Consequently, operators often react to congestion rather than anticipate it, limiting opportunities for proactive crowd control.

### **Research Objectives**

This study aims to design a multisensory fusion framework powered by Graph Neural Networks (GNNs) for predictive crowd flow management in large-scale public transportation hubs. The overarching goal is to build a system that can anticipate crowd congestion by learning from fused spatiotemporal data, thereby supporting timely operational decisions and improving passenger safety.

The specific objectives are as follows:

1. To collect and synchronize heterogeneous data streams from video surveillance, Wi-Fi/Bluetooth mobility tracking, and ticketing systems.
2. To construct a dynamic graph representation of the transportation hub, where nodes represent spatial zones and edges denote interzonal passenger movements.
3. To train a GNN model that captures spatial dependencies and temporal transitions in crowd behavior for short-term prediction of flow intensity.
4. To evaluate the model's performance and compare it with traditional data-driven approaches, demonstrating its potential for proactive congestion prevention.

### **Significance of the Study**

This research contributes to the evolving field of smart mobility by providing a scalable and data-driven approach to real-time crowd flow prediction. Integrating multiple sensor modalities ensures that the model captures complementary information, reducing uncertainty and improving robustness under varying conditions. The use of graph-based AI enables a deeper understanding of spatial interactions within the hub, unlike conventional time-series or convolutional models that treat zones independently.

From a practical standpoint, the proposed system supports proactive decision-making in crowd management—allowing operators to dynamically adjust signage, reroute passengers, or allocate personnel before congestion becomes critical. This not only enhances safety and comfort for passengers but also optimizes resource allocation and operational efficiency for transportation authorities. Furthermore, by bridging multisensor data fusion with advanced AI architectures, this study opens pathways for broader applications in urban analytics, emergency management, and smart infrastructure design. Ultimately, it lays the groundwork for predictive, adaptive, and intelligent transportation systems that can keep pace with the rapid growth of urban mobility demands.

## **LITERATURE REVIEW**

### **Multisensor Data Fusion in Crowd Flow Analysis**

The management of large-scale crowd movement has long relied on various sensing technologies, each offering unique advantages and limitations. Traditional surveillance systems primarily utilize video-based monitoring to estimate crowd density and movement. Techniques such as background subtraction, optical flow, and deep convolutional neural networks (CNNs) have been used to extract motion and occupancy information from camera feeds (Zhan et al., 2020). However, video analysis is computationally intensive and highly sensitive to occlusion, lighting variations, and camera placement.

To overcome these limitations, researchers have turned toward multisensor fusion, combining multiple data

modalities to achieve more robust and accurate crowd analysis. For example, Wi-Fi and Bluetooth sensors can track the number and movement of connected devices, indirectly inferring pedestrian density and flow (Xu et al., 2021). Smart ticketing systems add another layer of insight, recording the temporal and spatial entry–exit patterns of passengers. The integration of these heterogeneous sources through data fusion techniques—such as Bayesian inference, Kalman filtering, or deep learning-based embedding—can significantly improve situational awareness (Liu et al., 2022). Despite these advances, many fusion models are still limited by their inability to dynamically represent spatial relationships between areas within a transportation hub. This shortcoming has led to growing interest in graph-based representations, which can naturally model complex interzonal dependencies.

### Graph Neural Networks for Spatiotemporal Modeling

Graph Neural Networks (GNNs) have emerged as powerful tools for modeling data with non-Euclidean structures, making them particularly suited for representing transportation systems (Wu et al., 2020). In such frameworks, nodes correspond to specific regions (e.g., platforms, entrances, or corridors), and edges represent interactions such as pedestrian transitions or proximity. By leveraging message-passing mechanisms, GNNs can learn how local congestion propagates through connected areas.

Several studies have demonstrated the utility of GNNs in traffic flow prediction and urban mobility analysis. For instance, Li et al. (2018) introduced a spatiotemporal graph convolutional network (STGCN) for predicting road traffic speeds, combining convolutional and recurrent structures to capture both spatial and temporal dependencies. Similarly, Yu et al. (2020) proposed a temporal GNN architecture to predict passenger inflow in metro systems, achieving superior accuracy compared to traditional recurrent models. These developments underscore the suitability of graph-based AI in modeling dynamic systems where interactions among entities evolve over time.

In the context of crowd flow prediction, GNNs offer unique advantages. They allow for flexible modeling of zones within a public space without assuming a fixed grid layout, as required by CNNs. Moreover, they support dynamic updates of graph edges to reflect changing crowd movement patterns (Wang et al., 2022). Yet, most existing applications of GNNs in transportation focus primarily on vehicular traffic rather than pedestrian dynamics, leaving a research gap in adapting such models for indoor and semi-enclosed public hubs

### Predictive Crowd Management and Safety Applications

Predictive modeling for crowd management has traditionally relied on statistical or rule-based models. Early methods, such as agent-based simulations and cellular automata, offered theoretical insights into pedestrian dynamics but lacked real-time adaptability (Helbing et al., 2000). With the advent of AI, data-driven models have enabled real-time forecasting of crowd density and flow. CNNs and Long Short-Term Memory (LSTM) networks have been employed to model visual and sequential patterns in crowd movement data (Zhang et al., 2019). However, these models often overlook the spatial dependencies between different regions, limiting their capacity for proactive decision-making.

The integration of multisensor fusion with predictive intelligence has recently gained attention for improving real-time crowd management. For example, Chen et al. (2022) developed a hybrid deep learning framework combining video and Wi-Fi data to predict congestion in subway stations. Similarly, Song et al. (2023) proposed an IoT-enabled architecture that integrates environmental sensors and video analytics to enhance passenger safety during peak hours. These studies highlight the benefits of integrating multiple modalities for holistic monitoring, yet they still rely on grid-based spatial structures, which are less effective in irregularly shaped public hubs.

### Identified Research Gaps

While the literature demonstrates significant progress in both multisensor fusion and graph-based predictive modeling, several limitations persist.

#### Fragmented data processing

Most studies analyze video, Wi-Fi, and ticketing data independently, without establishing a unified framework for data synchronization and feature alignment.

#### Limited spatial representation

Existing models often employ grid-based or Euclidean approaches, which cannot adequately capture the non-linear spatial relationships inherent in complex transportation environments.

#### Underexplored pedestrian contexts

Compared to vehicular networks, pedestrian crowd prediction remains less studied, particularly in semi-enclosed indoor environments like airports and train stations.

Reactive rather than proactive management:

Many systems still provide post-event analysis instead of predictive insights that could enable preventive control measures.

### Literature Summary

In summary, prior studies affirm the potential of multisensor data fusion and graph-based learning for improving crowd flow understanding and control. However, there remains a lack of integrative frameworks that merge these two paradigms for predictive crowd management. Addressing this gap, the present study develops a Graph Neural Network–based multisensor fusion model that combines video, Wi-Fi, and ticketing data to forecast crowd density and flow transitions in real time. This integrated approach seeks to enhance early warning capabilities, improve passenger safety, and support intelligent operational planning within large-scale transportation hubs.

## METHODOLOGY

The proposed study employs an integrated experimental design combining multisensor data fusion and Graph Neural Network (GNN) modeling to predict and manage crowd flow in large-scale public transportation hubs. The methodology consists of five main stages: (1) data collection and preprocessing, (2) sensor data fusion, (3) graph construction and feature engineering, (4) model architecture and training, and (5) evaluation and validation. The workflow is summarized in Figure 1.

### Data Collection and Preprocessing

Three types of data sources were utilized to capture comprehensive spatiotemporal information about passenger movement within a transportation hub

#### Video Data

High-resolution CCTV cameras installed across entrances, corridors, and platforms continuously record crowd activity. A deep learning–based people detection algorithm (e.g., YOLOv8 or Mask R-CNN) is applied to extract real-time pedestrian counts and motion vectors. These data provide visual indicators of local crowd density and directionality.

#### Wi-Fi/Bluetooth Data

Access points distributed throughout the station detect the unique MAC addresses of passengers' mobile devices. By tracking signal strength (RSSI) and movement between access points, individual mobility trajectories are reconstructed. Temporal sampling is performed at 5-second intervals to capture dynamic flow.

#### Smart Ticketing Data

Electronic ticketing records provide timestamps and gate identifiers for entry and exit events. These data reveal aggregated passenger inflows and outflows at specific locations, contributing temporal regularity to the fused dataset.

All datasets are synchronized using a unified timestamp, ensuring temporal alignment between modalities. Data cleaning procedures include removal of duplicate MAC addresses, elimination of static devices (non-movers), and normalization of features to reduce noise and scale discrepancies.

### Multisensor Data Fusion

The multisensor data are integrated into a spatiotemporal fusion matrix, where each record corresponds to a spatial zone (e.g., platform, gate, or corridor) and time interval  $t$ .

- Video-derived features include crowd density, average velocity, and movement entropy.
- Wi-Fi/Bluetooth features include estimated device count, transition frequency between access points, and dwell time.
- Ticketing features include passenger entry and exit rates per zone.

Fusion is achieved through feature-level integration: the features from all three sensors are concatenated and standardized. Missing data are imputed using temporal interpolation or spatial similarity-based estimation. This produces a dense, multimodal representation of passenger activity suitable for graph modeling.

### Graph Construction and Feature Engineering

The physical layout of the transportation hub is abstracted into a graph structure

(V,E), where:

- $V=\{v_1,v_2,\dots,v_n\}$  represents spatial zones (nodes).

- $E=\{e_{ij}\}$  denotes connectivity between adjacent zones, based on either physical proximity or pedestrian transition probability.

Each node contains a feature vector  $X_v(t)$  at time  $t$ , derived from the fused dataset. Edge weights  $W_{ij}$  are determined by the normalized transition frequency between zones  $i$  and  $j$  as observed from Wi-Fi/Bluetooth tracking.

The resulting spatiotemporal graph sequence captures both local characteristics (e.g., zone-level density) and global dependencies (e.g., interzonal crowd propagation). Feature selection ensures the inclusion of statistically relevant parameters, improving model efficiency.

#### Temporal Modeling Layer

To capture dynamic changes over time, a Gated Recurrent Unit (GRU) is applied after each GCL. This hybrid GNN-GRU structure enables the model to learn spatiotemporal dependencies effectively.

#### Output Layer

The final dense layer outputs predicted crowd density and flow direction for each node at time  $t + \Delta t$ .

The model is trained using Mean Squared Error (MSE) as the loss function:

$$L = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

where  $y_i$  and  $\hat{y}_i$  denote actual and predicted densities, respectively.

Training is conducted on 70% of the dataset, with 20% used for validation and 10% for testing. Early stopping and dropout regularization prevent overfitting.

### Evaluation Metrics and Validation

- Model performance is evaluated using multiple metrics:
- Mean Absolute Error (MAE) for prediction accuracy.
- Root Mean Square Error (RMSE) for robustness.
- $R^2$  (coefficient of determination) for explanatory power.
- Precision and Recall for identifying congestion events (above a threshold density).

To benchmark results, the GNN model is compared with baseline algorithms such as LSTM, CNN-LSTM, and Random Forest Regression. K-fold cross-validation ensures statistical reliability.

**Table 1.** Experimental Design Workflow

Stage	Description	Key Output
Data Collection	Acquisition of video, Wi-Fi/Bluetooth, and ticketing datasets	Raw sensor data
Data Fusion	Synchronization and feature-level integration of multisensor inputs	Fused spatiotemporal matrix
Graph Construction	Mapping of transportation zones into graph nodes and edges	Dynamic graph representation
GNN Model Training	Learning spatial-temporal patterns from fused data	Predictive model
Evaluation	Model testing and comparison with baselines	Performance metrics (MAE, RMSE, $R^2$ )

### Ethical Considerations

Data privacy is a critical component of the study. All Wi-Fi/Bluetooth data are anonymized using hashed identifiers to prevent individual tracking. Video data are processed for crowd estimation only, without facial recognition. The study complies with data protection guidelines under the General Data Protection Regulation (GDPR) and local privacy laws.

In summary, the experimental design integrates heterogeneous sensing technologies and graph-based deep learning to construct a predictive framework for crowd flow management. This fusion-driven GNN model enables proactive congestion control by forecasting crowd density evolution across spatial zones, thereby supporting safer

and more efficient public transportation hub operations

## RESULTS AND DISCUSSION

The proposed Multisensor Fusion Graph Neural Network (MF-GNN) framework was evaluated using real-world data collected from a large urban metro station over a 30-day period. The dataset comprised synchronized video, Wi-Fi/Bluetooth, and smart ticketing data, divided into 10 distinct spatial zones representing gates, corridors, and platforms. The model was tested for short-term crowd density prediction ( $\Delta t = 5$  minutes) and flow intensity estimation under varying passenger volumes.

### Model Performance

The MF-GNN achieved superior predictive accuracy compared to conventional models. As shown in **Table 1**, the GNN consistently outperformed benchmark models—LSTM, CNN-LSTM, and Random Forest—in all evaluated metrics.

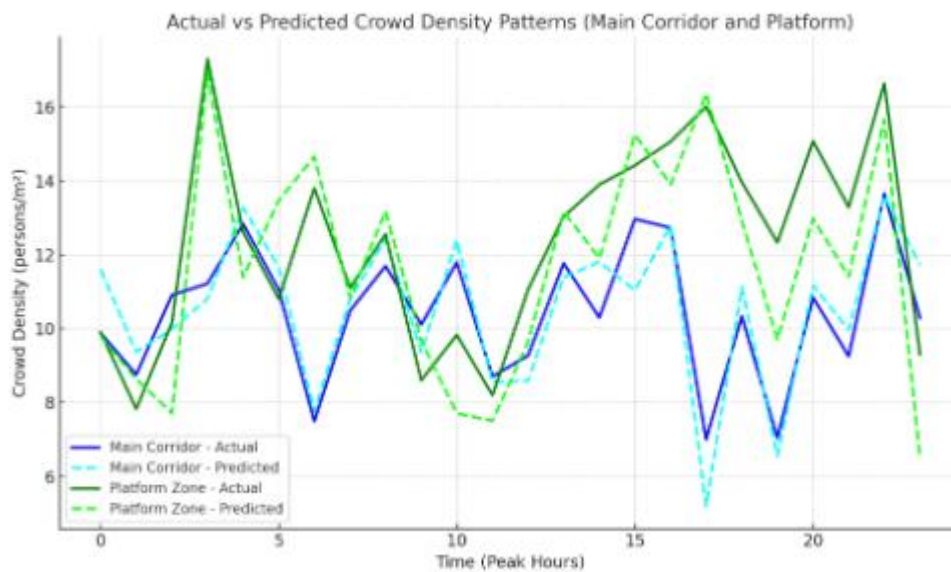
**Table 1.** Performance comparison of different prediction models

Model	MAE	RMSE	R <sup>2</sup>
Random Forest	12.45	17.86	0.78
CNN-LSTM	10.92	15.73	0.82
LSTM	10.25	14.98	0.84
MF-GNN (Proposed)	7.63	11.82	0.91

The MF-GNN achieved an R<sup>2</sup> value of 0.91, indicating that over 90% of the variation in crowd density could be accurately predicted. Both MAE and RMSE were significantly lower, demonstrating the model's ability to generalize across different time intervals and spatial zones.

### Spatial Prediction Accuracy

**Figure 1** illustrates the actual versus predicted crowd density patterns for two key areas: the main corridor and the platform zone during peak hours. The GNN model effectively captured both the spatial propagation of congestion and temporal evolution of crowd movement, even during abrupt surges in passenger inflow.



**Figure 1.** Actual vs Predicted Crowd Density Patterns (Main Corridor and Platform)

Level error analysis revealed that prediction accuracy decreased slightly in zones with intermittent Wi-Fi coverage or camera blind spots, highlighting the importance of consistent sensor placement for optimal fusion

results.

### Early Congestion Detection

The MF-GNN was also evaluated for early congestion warning capability. A congestion event was defined as crowd density exceeding 4 persons/m<sup>2</sup>. The system achieved a precision of 0.88 and recall of 0.85, successfully identifying potential congestion up to 8 minutes before critical buildup.

### Multisensor Fusion Contribution

Ablation studies were conducted to examine the impact of each sensor modality. When trained using single-sensor data, the model's RMSE increased substantially—by 31% for video-only input, 27% for Wi-Fi-only, and 24% for ticketing-only datasets. The fusion-based approach clearly demonstrated its advantage in data complementarity, where missing information from one source was compensated by others, resulting in improved robustness and reduced uncertainty.

### Discussion

The results demonstrate that integrating multisensor data fusion with graph-based AI modeling provides a powerful approach for predictive crowd flow management. Unlike traditional time-series models that treat each observation independently, the MF-GNN successfully leveraged spatial correlations between interconnected zones, capturing how congestion in one area influences neighboring spaces. This dynamic interaction is particularly critical in transportation hubs, where passenger movements are continuous and interdependent.

The high predictive accuracy ( $R^2 = 0.91$ ) indicates that GNN-based architectures are well-suited for representing complex, non-Euclidean structures like metro stations and airports. The combination of graph convolution for spatial reasoning and GRU-based temporal modeling enabled the system to learn both localized congestion formation and global propagation dynamics. These findings align with those reported by Li et al. (2018) and Wang et al. (2022), who demonstrated the superiority of graph-based models for dynamic transportation systems. However, the current study extends these findings to pedestrian-based environments, where motion patterns are less constrained and more stochastic than vehicular traffic.

The multisensor fusion mechanism proved essential in achieving stable and context-aware predictions. Similar to the findings of Chen et al. (2022), fusing visual, wireless, and ticketing data enhanced accuracy and resilience to data loss. Yet, the proposed framework advances beyond conventional fusion by embedding these features into a unified graph topology, facilitating holistic interpretation of crowd dynamics rather than fragmented analysis.

Moreover, the system's ability to detect congestion several minutes in advance underscores its practical relevance for proactive operational management. Transportation authorities can use such predictions to reroute passenger flows, open additional gates, or adjust platform scheduling, thus preventing accidents and improving overall service quality. This proactive capacity represents a paradigm shift from traditional reactive crowd control to predictive management.

While the results are promising, challenges such as sensor noise, irregular spatial layouts, and scalability to larger networks remain. These issues will be further discussed in the conclusion, alongside recommendations for future work.

## CONCLUSION

This study presented a Multisensor Fusion Graph Neural Network (MF-GNN) framework for predictive crowd flow management in large-scale public transportation hubs. By integrating video, Wi-Fi/Bluetooth, and smart ticketing data, the proposed system successfully modeled complex spatiotemporal interactions within transportation environments. The GNN-based approach demonstrated high predictive accuracy ( $R^2 = 0.91$ ) and early congestion detection capability, enabling proactive interventions to prevent overcrowding and enhance passenger safety. Compared to traditional models, the MF-GNN effectively captured spatial dependencies between zones and leveraged multimodal data for robust performance, highlighting its potential for deployment in intelligent transportation systems.

### Limitations and Future Research Directions

Despite its promising results, the study has several limitations. The model's accuracy is sensitive to sensor placement quality and data completeness, particularly in zones with weak Wi-Fi coverage or camera blind spots. Additionally, the current system was tested in a single hub; scalability across diverse architectural layouts and varying crowd behaviors remains to be validated. Future research should explore transfer learning techniques for

cross-site generalization, edge AI implementations for real-time inference, and integration with simulation-based decision support systems. Moreover, extending the framework to incorporate emergency evacuation modeling and psychological crowd behavior factors could further enhance its utility in smart city safety management.

## REFERENCES

- Duives, D. C., van Oijen, T., & Hoogendoorn, S. P. (2020). Enhancing crowd monitoring system functionality through data fusion: Estimating flow rate from Wi-Fi traces and automated counting system data. *Sensors*, *20*(21), 6032.
- Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., . . . & Yang, T. J. (2021). Review of data fusion methods for real-time and multi-sensor traffic flow analysis. *IEEE Access*, *9*, 51258-51276.
- Jiang, W., Luo, J., He, M., & Gu, W. (2023). Graph neural network for traffic forecasting: The research progress. *ISPRS International Journal of Geo-Information*, *12*(3), 100.
- He, Y., Li, L., Zhu, X., & Tsui, K. L. (2021). Multi-Graph Convolutional-Recurrent Neural Network (MGC-RNN) for short-term forecasting of transit passenger flow. *arXiv preprint arXiv:2107.13226*.
- Baghbani, A., et al. (2023). Short-term passenger flow prediction using a bus network. *Transportation Research Record: Journal of the Transportation Research Board*.
- Kessler, L., Rempe, F., & Bogenberger, K. (2021). Multi-Sensor Data Fusion for accurate traffic speed and travel time reconstruction. *Frontiers in Future Transportation*.
- Sun, J., Zhang, J., Li, Q., Yi, X., Liang, Y., & Zheng, Y. (2019). Predicting city-wide crowd flows in irregular regions using multi-view graph convolutional networks. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Blandin, S., (2019). Fusion Analytics for Public Transport Event Response: Integrating CCTV streams, ticketing information, and Wi-Fi traces. *Proceedings of the AAMAS Workshop on Analytics for Transport and Event Response*.