

Bus stop congestion monitoring based on process discovery algorithms

Seon Kim¹[0009-0006-9722-6699], Jungtak Oh¹ [0009-0001-9129-0351] and Jongchan Kim^{1*}[0000-0003-3279-7440]

¹ Yonsei University, Wonju 26493, South Korea
{kso323, bluetaks, jongchan.kim}@yonsei.ac.kr

Abstract. Bus stop congestion is a key factor that influences the service coverage and operational efficiency of bus transportation systems, directly impacting traffic flow, passenger experience, and environmental sustainability. This study proposes a novel approach to quantitatively evaluate the congestion of bus stops by applying process discovery algorithms. Specifically, congestion levels were analyzed by dividing the data into two distinct periods—commuting and non-commuting hours—to capture temporal variations at the bus stop. The experimental results demonstrated that metrics capturing the relational characteristics between nodes in the process model—particularly Modified Density and Entropy-based metrics—were more effective at detecting congestion phenomena than traditional structure-oriented metrics such as simplicity, size, the Extended Cardoso Metric, and the Extended Cyclomatic Metric. This analytical approach presents a methodology for analyzing congestion levels at bus stops, offering valuable insights for optimizing bus transit infrastructure.

Keywords: bus stop, congestion, process discovery, process model complexity

1 Introduction

The level of congestion at bus stops is a critical factor that directly affects the punctuality, efficiency, and user satisfaction of bus transportation services [1,2]. High congestion levels often result in a range of negative outcomes, including delays in boarding and alighting, insufficient waiting space within the stop, and diminished service reliability. These challenges may further escalate, contributing to citywide traffic congestion and environmental deterioration.

Process mining (PM) enables the discovery of actual business process flows by analyzing execution events within real information systems [3]. It also provides methodologies for in-depth process analysis across various dimensions, including conformance checking to verify the alignment between ongoing processes and their expected counterparts, as well as the analysis of various performance metrics such as remaining time.

Process discovery in PM offers dynamic information and visualization. Additionally, it enables an intuitive understanding of individual process case trajectories and variations within a process. Leveraging these capabilities, PM techniques can be explored for conceptualizing various social phenomena as processes. For instance, this

study conceptualized each bus passing through a specific bus stop as a distinct process. By analyzing the operational history of all buses traversing that stop, we measured the process complexity associated with that particular bus stop.

Complexity metrics commonly used for process models include simplicity, size, Extended Cardoso Metric, and Extended Cyclomatic Metric [16,17]. These traditional metrics primarily focus on quantifying fundamental structural attributes of the process, such as the number of elements. To complement these, we also employ Modified Density and Entropy-based metrics, which capture the relational and behavioral complexity inherent in the process. Modified Density indicates the degree of interconnectedness among nodes within the process model, offering insights into the intensity of interactions between operational components. Entropy-based metrics quantify the degree of randomness or uncertainty in process execution, enabling a deeper understanding of variability and irregularity in passenger flow and bus stop congestion patterns.

This study aims to propose a framework for evaluating, comparing, and analyzing bus stop congestion through the application of these complexity metrics. By concentrating on metrics that capture the relational and behavioral aspects of congestion, we provide a framework for objective, event log-driven data. Additionally, by analyzing congestion variations between morning commuting hours and non-commuting hours, this research aims to support the development of customized operational strategies that enhance transit service efficiency and passenger satisfaction.

The remainder of the paper is structured as follows. Section 2 summarizes related work. Section 3 details the experimental procedure used. Section 4 presents and discusses the results. Finally, Section 5 concludes the paper.

2 Related Work

Urban traffic congestion has traditionally been assessed using various quantitative measures, such as average travel speed, total travel time, and traffic volume indices [4]. However, recent studies on traffic congestion have focused on pattern analysis using GPS-based data instead of relying on aggregate statistics. Kohan et al. [5] analyzed traffic flow patterns to identify congestion segments, and Zhu et al. [6] modeled vehicle movement and traffic jam characteristics using a deep learning-based attention mechanism. He et al. [7] conducted a dynamic analysis of congestion zones, onset times, and underlying causes, utilizing taxi GPS data. These studies have overcome the limitations of conventional single indicators by employing advanced data and methodologies.

Buses and bus stops are critical congestion points within a transportation system, and consequently, studies are being conducted to measure their congestion levels. Carli et al. [8] proposed a method for automatically assessing urban traffic congestion by utilizing buses as probes and integrating their positional data with information on surrounding traffic conditions for analysis. Kanamitsu et al. [9] contributed to understanding bus and bus stop congestion by automatically estimating passenger congestion levels within buses using Bluetooth Low Energy (BLE) signals. These studies also utilized GPS data.

Bus punctuality serves as an indicator for evaluating the quality of bus transportation services, and reflects the degree of congruence between actual operating times and the published schedule [10]. Liu et al. [11] analyzed the factors influencing bus punctuality across different time periods, bus route segments, and climatic conditions. The analysis revealed that bus delays and deviations from scheduled arrival times were more frequent during peak hours, in the later segments of the bus route, and under adverse weather conditions. Chepuri et al. [12] and Ait-Ali et al. [13] assessed changes in traffic patterns over time and identified a relative reduction in congestion during off-peak hours. Cats et al. [14] analyzed the influence of driving patterns and traffic condition variables on travel time variability.

Furthermore, in the transportation sector, PM is applied to analyze real-world operational flows and behavioral patterns based on event logs. Delgado et al. [15] analyzed the characteristics of bus routes and passenger movement patterns using publicly available data collected from the public transportation system in Montevideo, Uruguay. This study identified discrepancies between theoretical models and real-world operations and proposed that PM is an effective tool for diagnosing the efficiency of transportation systems.

However, previous studies have primarily concentrated on analyzing congestion levels and patterns across entire road networks or at the bus route level, rather than quantitatively evaluating congestion at individual bus stops. Furthermore, most studies utilize extensive GPS-derived data. This study analyzes temporal changes extracted from event logs, quantifying complexity at each station level through various metrics.

3 Experimental Procedure

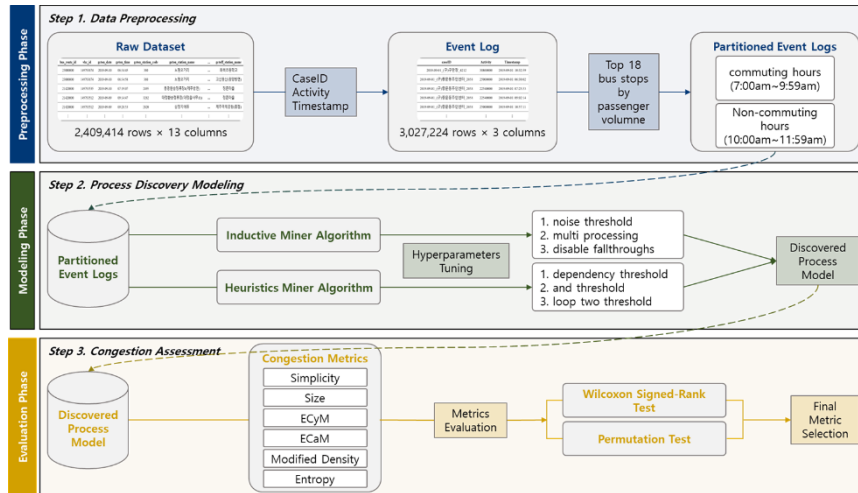


Fig. 1. Overall framework of the research

The analytical framework of this study begins with preprocessing the raw data to prepare event logs suitable for PM. Subsequently, process discovery techniques are employed to extract and analyze congestion patterns at bus stops utilizing various complexity metrics. Finally, statistical tests are conducted to validate the significance of observed congestion differences. The entire workflow is summarized in Fig. 1.

Event logs are generated from raw data related to bus boarding and alighting, and the data are classified by two time periods. Subsequently, a process discovery algorithm is applied to the event logs to derive process models for each bus stop. Various congestion metrics are then calculated for the discovered process models, and the results are compared and evaluated using statistical methods.

The specifics of each phase are detailed in Subsections 3.1 to 3.5.

3.1 Raw Dataset

This study utilized the publicly available dataset¹ for Jeju Island, South Korea, which was released through the DACON competition. The dataset comprises bus card-specific boarding and alighting events recorded over a period of approximately two months, from September 1 to October 31, 2019. The data were provided in CSV file format. Each record includes the information presented in Table 1.

Table 1. Excerpt of Raw Dataset used in the research

bus_route_ id	vhc_ id	geton_ date	geton_ time	geton_ station_ _code	geton_ station_ _name	getoff_ station_ _name
23000000	149793 674	2019- 09-10	06:34:4 5	360	Nohyeong 5-way Intersection	... Hwabuk ele- mentary School
21420000	149793 535	2019- 09-10	07:19:0 7	2495	Donggwang Transit Station 4	... Jeongjon Vil- lage
:	:	:	:	:	:	:

The dataset includes only boarding and alighting events that occurred between 6:00 AM and 12:00 PM and contains passenger transaction records linked to individual transportation cards, making it suitable for analyzing usage patterns by time period and bus stop.

3.2 Event Log Preprocessing

In this study, data preprocessing and event log transformation were conducted to enhance analytical accuracy and facilitate congestion analysis through PM. The dataset

¹ <https://dacon.io/competitions/official/229255/data>

was restructured to represent the chronological sequence of bus arrivals, as well as passenger boarding and alighting events, at each bus stop. This transformation enabled us to model the order in which buses arrive at stops on each operating day.

Initially, missing values were detected only in alighting events, and these records were excluded from the analysis. The data were then converted into an event log format appropriate for PM, with each event consisting of three essential components: Case ID, Activity, and Timestamp (see Table 2).

Table 2. Excerpt of Event Log used in the research

caseID	Activity	Timestamp
2019-09-01_(Old) Union Market_4212	30860000	2019-09-01 10:32:59
2019-09-01_(Old)Jungmun-dong Community Service Center_2058	25000000	2019-09-01 06:38:02
⋮	⋮	⋮

- caseID : Represents each bus stop unit by day of the week. To distinguish stops with the same name but different directions, both the stop name and the unique stop ID are used together (format: date_bus station name_bus station code).
- Activity : Indicates the bus route ID that stops at the respective bus stop.
- Timestamp : Refers to the exact date and time at which a boarding or alighting event occurs at the respective bus stop.

Subsequently, the event log was partitioned into two time periods: morning commuting hours (7:00:00 AM–9:59:59 AM) and non-commuting hours (10:00:00 AM–11:59:59 AM). Based on Chepuri et al. [12], who define peak commuting hours as the morning period from 7:00 AM to 10:00 AM, we designed our morning commuting hours accordingly. Eighteen bus stops with the highest passenger volumes were selected for analysis from the entire set of stops. The refined event logs were subsequently used to assess congestion and complexity across different time intervals and bus stops.

3.3 Process discovery algorithms

Process discovery is a technique in PM that automatically derives the actual flow and structure of processes based on event logs. By analyzing the sequence and interrelationships of events recorded in information systems, it reconstructs real-world business processes in a visual and structured form without relying on predefined models.

To quantitatively assess the structural characteristics and complexity of process models, only Petri net-based process discovery algorithms were employed [16]. Petri nets provide a precise representation of various behaviors observed in real-world business and service processes, including concurrency, branching, looping, and synchronization. These characteristics facilitate the quantification of complexity in intricate processes and effectively capture diverse patterns within real-world workflows. Moreover, Petri nets are a standard modeling language in the field of PM, having been reliably validated through decades of research and practical application.

Inductive Miner [3]. This algorithm detects various patterns in logs—including sequential, parallel, concurrent, and loop structures—by employing a recursive partitioning approach. It is robust to noise and ensures that the resulting model is always a sound Petri net. A primary advantage of this algorithm is its ability to effectively capture the structural characteristics of real-world business processes, particularly those involving complex branching structures or repetitive patterns. The main parameters presented for tuning are listed in Table 3, with default values indicated in bold.

Table 3. Hyperparameters, Tuned Values, Descriptions, and Effects of Inductive Miner

Hyperparameter	Tuning value	Description	Effect
noise_threshold	{ 0.0 , 0.1, 0.2, 0.3}	Ignores infrequent paths as the value increases.	lower = more complex model
multi_processing	{True, False }	Determines whether to use multiple CPU cores.	True = faster processing
disable_fallthroughs	{True, False }	Excludes certain exceptional paths when enabled.	True = some paths may be excluded from the model

Heuristic Miner [17,19]. This algorithm produces realistic process models by analyzing the direct frequency and causal relationships between events. It prioritizes the inclusion of frequently occurring paths and activity connections within the log, enabling the generation of concise and interpretable models despite the presence of noisy real-world data. This approach is useful in situations where event logs are incomplete or contain a large number of exceptional paths. The parameters adjusted in this study are presented in Table 4, with default values indicated in bold.

Table 4. Hyperparameters, Tuned Values, Descriptions, and Effects of Heuristic Miner

Hyperparameter	Tuning value	Description	Effect
dependency_thresh	{0.3, 0.5 , 0.7, 0.9}	Causality thresholds between activities	Higher = More significant paths included
and_threshold	{0.5, 0.65 , 0.8}	Parallel Relationship Recognition Threshold	Higher = stricter recognition of parallel (AND) structures.
loop_two_threshold	{0.3, 0.5 , 0.7}	2-Activity Loop Recognition Threshold	Higher = stricter detection of two-activity loop structures.

3.4 Process Model Complexity Metrics

To quantitatively evaluate congestion within each process model, we calculated complexity metrics. The primary congestion indicators utilized in this study are detailed as below, encompassing various structural and behavioral aspects of bus stop congestion.

(1) Simplicity² [20]. It is a metric that quantifies the structural simplicity of a process model, where a higher value indicates a simpler model. In this study, the Simplicity function provided by the PM4Py library was used.

(2) Size [17]. It refers to the total number of places(P) included in the process model and represents its basic structural complexity. A larger value indicates a more complex model. (PN indicates the Petri net)

$$Size(PN) = P \quad (1)$$

(3) Extended Cyclomatic Metric (ECyM) [16]. It measures the number of linearly independent paths such as loops, branches, and cycles in the control-flow graph of a process model. A higher value reflects greater structural complexity, especially in terms of repeated process flows, and is effective for capturing complex behaviors.

$$ECyM(PN) = |E| - |V| + p \quad (2)$$

(4) Extended Cardoso Metric (ECaM) [16,17]. It evaluates branching complexity by considering both the number and structural characteristics of decision points (e.g., AND, OR, XOR) in the model. A higher value results from more diverse branching structures, enabling the quantification of complex patterns in which multiple routes diverge or converge at a single bus stop.

$$ECFC_p(p) = |\{t \bullet \mid t \in p \bullet\}| \quad (3)$$

$$ECaM(PN) = \sum_{p \in P} ECFC_p(p) \quad (4)$$

(5) Modified Density. It is derived from the original Density metric [18], which represents the degree of connectivity between nodes in a Petri net, including activities that were not observed in the event log. In contrast, Modified Density that we propose calculates the connectivity only among activities that actually appear in the log, thus providing a more accurate representation of the real-world process complexity by focusing on observable behavior.

$$MD(PN) = \frac{E}{N \times (N-1)} \quad (5)$$

(6) Entropy [3]. It is a metric that quantifies the uncertainty and variability in the distribution of paths or events within a process model. A higher entropy value signifies increased diversity and distinctive flow patterns, reflecting the complexity at a given bus stop.

² <https://pm4py-source.readthedocs.io/en/stable/pm4py.evaluation.simplicity.html>

$$\text{Entropy} = - \sum_{i \in \{T, P, A\}} P_i \times \log_2 P_i \quad (6)$$

3.5 Validation

This study employed the Wilcoxon signed-rank test and the permutation test to evaluate the statistical significance of differences in bus stop congestion indicators between morning commuting and non-commuting hours. The Wilcoxon signed-rank test is a non-parametric method used to determine whether the median difference between paired data, measured repeatedly under two conditions (in this study, two time periods at the same bus stop), is statistically significant. The permutation test is a non-parametric method that assesses the significance of observed statistics by repeatedly and randomly rearranging the observed values between two conditions, without assuming any specific data distribution. Given that the results of the metrics do not assume normality, these two non-parametric tests were employed instead of the paired-samples t-test.

4 Experimental Results

In this study, the hyperparameters of two process discovery algorithms were tuned to evaluate congestion metrics within bus stops between commuting and non-commuting hours. Due to higher passenger volumes and bus frequencies during commuting hours, the event log was partitioned into these two periods to reflect differences in congestion and complexity. Eighteen major bus stops in Jeju Island, previously selected, were analyzed using six metrics. Statistical significance tests were used to determine whether specific complexity metrics could reliably distinguish between the two periods at the bus stop level.

4.1 Statistical Significance Analysis of the Heuristics Miner

The results of the combinations, with each hyperparameter set to its default value, are shown in Fig. 2 and Table 5. Fig. 2 illustrates the values of each metric by bus stop and time period, with red indicating commuting hours and blue indicating non-commuting hours. For ECaM, a logarithmic scale was applied to represent the overall trend due to its wide range of values.

As demonstrated in Fig. 2, Simplicity and Size metrics showed no consistent patterns in their average value changes or relationships across bus stops between the two time periods. For instance, a bus stop with a high Simplicity value was not necessarily associated with a low Size, and in some stops, the Size was observed to be larger during non-commuting hours than during commuting periods. Furthermore, within the same stop, Simplicity and Size tended to vary inversely, suggesting a weak correlation between structural simplicity/complexity and actual congestion levels. In other words, these structural metrics do not capture real congestion within bus stops.

In contrast, the Modified Density metric consistently exhibited higher values during commuting hours compared to non-commuting hours across most bus stops. Notably,

both the Wilcoxon signed-rank test and permutation test results (default parameter combination: $p = 0.000/0.000$) confirmed statistically significant differences (see Table 5). This indicates that the Modified Density metric is highly sensitive to congestion differences between time periods regardless of hyperparameter variations and aligns well with the subjective perception of congestion in real-world public transportation settings.

Both ECyM and ECaM showed some statistical significance in the Wilcoxon test under both default and minimum parameter settings; however, the permutation test results were inconsistent, and the differences in average values between time periods were not as pronounced as those observed with Modified Density. Similarly, Entropy exhibited partial significance only in the Wilcoxon test depending on parameter settings, while the permutation test did not reveal clear differences.

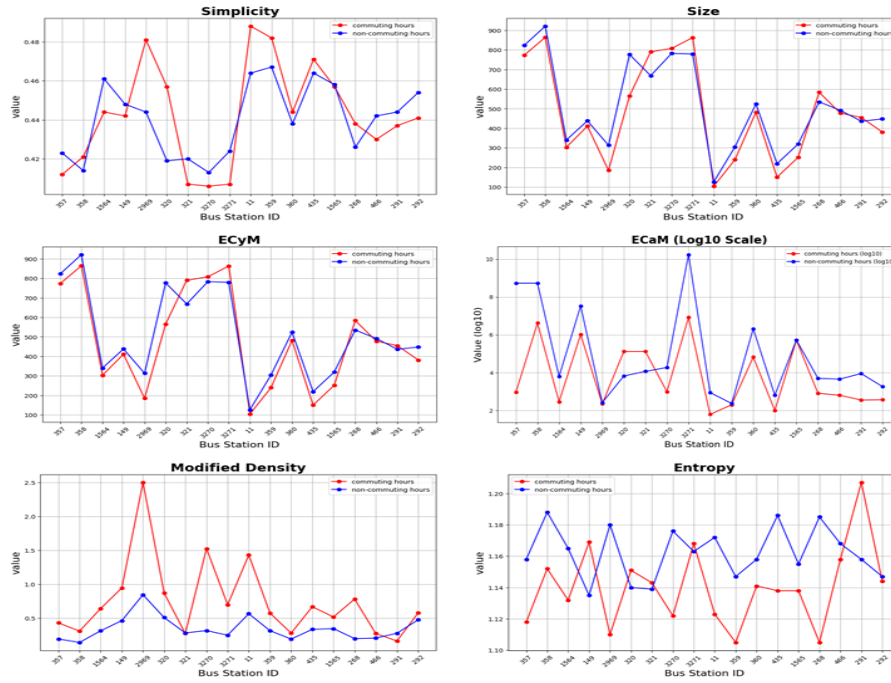


Fig. 2. Values for Simplicity, Size, ECyM, ECaM, Modified Density, and Entropy metrics at the bus stop: commuting hours vs. non-commuting hours

Table 5. Statistical test results for Simplicity, Size, ECyM, ECaM, Modified Density, and Entropy metrics between two time periods

S	Simplicity	Size	ECyM	ECaM	Modified Density	Entropy
Wilcoxon signed-rank test	1.000	0.001 *	0.081	0.007 *	0.000 *	0.008 *

Permutation test	0.784	0.197	0.695	0.084	0.000 *	0.167
-------------------------	-------	-------	-------	-------	----------------	-------

4.2 Statistical Significance Analysis by Metrics

The results of statistical significance assessments using the Wilcoxon signed-rank test and permutation test for all combinations generated by varying hyperparameters across the two process discovery algorithms and each applied metric are shown in Table 6. The best metric is the one with bold and underline, and the next best metric is the one with only bold.

Table 6. Summary of statistical test results for different congestion metrics

Algorithm	Evaluation	Metric	Proportion of significant p-values
Inductive miner	wilcoxon signed-rank test	Simplicity	0 / 16
		Size	4 / 16
		<u>ECyM</u>	<u>14 / 16</u>
		ECaM	4 / 16
		Modified Density	12 / 16
	permutation test	Entropy	2 / 16
		Simplicity	0 / 16
		Size	4 / 16
		<u>ECyM</u>	<u>10 / 16</u>
		ECaM	2 / 16
Heuristics miner	wilcoxon signed-rank test	Modified Density	6 / 16
		Entropy	3 / 16
	permutation test	Simplicity	6 / 36
		Size	19 / 36
		ECyM	13 / 36
		ECaM	15 / 36
		<u>Modified Density</u>	<u>34 / 36</u>
		Entropy	20 / 36
	permutation test	Simplicity	0 / 36
		Size	0 / 36
		ECyM	0 / 36
		ECaM	0 / 36
		<u>Modified Density</u>	<u>17 / 36</u>
		Entropy	0 / 36

For the Inductive Miner, the Wilcoxon signed-rank test showed that ECyM (14 out of 16) and Modified Density (12 out of 16) most frequently exhibited significant differences between commuting and non-commuting hours, while Size (4 out of 16), ECaM (4 out of 16), and Entropy (2 out of 16) showed significance only in some parameter combinations. The permutation test also revealed significant differences in some cases for ECyM (10 out of 16), Modified Density (6 out of 16), Size (4 out of 16), Entropy (3 out of 16), and ECaM (2 out of 16). However, Simplicity showed no significant differences in either test (0 out of 16).

For the Heuristics Miner, the Wilcoxon signed-rank test showed that Modified Density (34 out of 36) consistently exhibited significant differences, followed by Entropy

(20 out of 36), Size (19 out of 36), ECaM (15 out of 36), ECyM (13 out of 36), and Simplicity (6 out of 36), which showed statistical significance at some bus stops. In contrast, the permutation test revealed statistically significant differences only for Modified Density (17 out of 36), while no significant differences were observed for the other metrics across all parameter combinations (0 out of 36).

Overall, Modified Density demonstrated the greatest consistency and sensitivity in detecting congestion changes across time periods. This suggests that the metric can be effectively utilized for statistical discrimination and practical assessment of congestion variations in real-world public transportation operations.

5 Conclusion

This study conducted a quantitative analysis of bus stop congestion based on actual boarding and alighting event log data from bus passengers in Jeju Island, South Korea. Accordingly, process discovery algorithms were introduced to propose and apply congestion and complexity metrics. While existing research on bus stop congestion has primarily analyzed static indicators—such as the number of passengers boarding and alighting per stop, vehicle dwell counts, and physical tools—this study aimed to quantitatively reflect dynamic behaviors occurring at each stop, such as branching and flows.

Previous research on traffic, including buses and bus stops, has necessitated location-based data, such as GPS data. However, this study demonstrates the feasibility of quantitatively analyzing changes in congestion and complexity at the bus stop level using public transportation card data. This approach could enhance practical applications by improving data accessibility and reducing costs. Furthermore, it presents a novel approach by indicating that temporal congestion patterns can be captured using only simple data.

A limitation of this study is that its analysis was confined to a specific region and eighteen major bus stops, which limits the generalizability of its findings. Therefore, future research should aim to broaden the scope by including diverse urban areas, various transportation routes, and data covering extended time periods.

Another limitation is that each event in the preprocessed data was treated as representing a single passenger, without accounting for the number of passengers boarding or alighting per transportation card. This simplification limits the current analysis. Future work could consider passenger counts as an additional attribute to provide realistic analysis of bus stop congestion dynamics.

This study offers a quantitative framework for evaluating congestion at bus stops, aiding in the enhancement of operational efficiency and service quality in public transportation systems. By identifying areas of congestion, transit agencies can allocate resources effectively and respond dynamically to demand in real time. Continuous congestion monitoring also helps ensure adherence to service standards and enhances the passenger experience. Overall, these insights contribute to optimized management aligned with public transportation objectives.

References

1. Li, G., Nie, L., Gao, F., & He, Z.: Optimization of bus stop layout considering multiple factors including passenger flow direction. *PloS one*, 19(11), e0313040. (2024)
2. Kodirov, T., Togaev, G., & Kenjaeva, B.: Evaluation of complexity of urban bus routes. In *E3S Web of Conferences* (Vol. 365, p. 05005). EDP Sciences. (2023)
3. Van Der Aalst, W.: Data science in action. In *Process mining: Data science in action* (pp. 3-23). Berlin, Heidelberg: Springer Berlin Heidelberg. (2016)
4. MA, Yafeng, et al.: Traffic Congestion Measurement Based on Accumulating Volume. (2016)
5. Kohan, M., & Ale, J. M.: Discovering traffic congestion through traffic flow patterns generated by moving object trajectories. *Computers, Environment and Urban Systems*, 80, 101426. (2020)
6. Zhu, S., Ding, R., Zhang, M., Van Hentenryck, P., & Xie, Y.: Spatio-temporal point processes with attention for traffic congestion event modeling. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 7298-7309. (2021)
7. He, Y., Hofer, B., Sheng, Y., Yin, Y., & Lin, H.: Processes and events in the center: a taxi trajectory-based approach to detecting traffic congestion and analyzing its causes. *International Journal of Digital Earth*, 16(1), 509-531. (2023)
8. Carli, R., Dotoli, M., Epicoco, N., Angelico, B., & Vinciullo, A.: Automated evaluation of urban traffic congestion using bus as a probe. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 967-972). IEEE. (2015)
9. Kanamitsu, Y., Taya, E., Tachibana, K., Nakamura, Y., Matsuda, Y., Suwa, H., & Yasumoto, K.: Estimating congestion in a fixed-route bus by using ble signals. *Sensors*, 22(3), 881. (2022)
10. Kho, S. Y., Park, J. S., Kim, Y. H., & Kim, E. H.: A development of punctuality index for bus operation. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 492-504. (2005)
11. Liu, Gang, Ling Shi, and Tony Z. Qiu. Evaluation of Factors Affecting Bus On-Time Performance in Edmonton, Canada. No. 16-3076. 2016.
12. Chepuri, A., Ramakrishnan, J., Arkatkar, S., Joshi, G., & Pulugurtha, S. S.: Examining travel time reliability-based performance indicators for bus routes using GPS-based bus trajectory data in India. *Journal of Transportation Engineering, Part A: Systems*, 144(5), 04018012. (2018)
13. Ait-Ali, A.: Exploring measures to monitor passenger-centric punctuality in public transport: a pre study. *K2-Nationellt kunskapscentrum för kollektivtrafik*. (2024).
14. Cats, O.: Determinants of bus riding time deviations: Relationship between driving patterns and transit performance. *Journal of Transportation Engineering, Part A: Systems*, 145(1), 04018078. (2019)
15. Delgado, A., & Calegari, D.: Process mining for improving urban mobility in smart cities: Challenges and application with open data. (2023).
16. Lassen, K. B., & van der Aalst, W. M.: Complexity metrics for workflow nets. *Information and Software Technology*, 51(3), 610-626. (2009)
17. Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F. M., Marrella, A., ... & Soo, A.: Automated discovery of process models from event logs: Review and benchmark. *IEEE transactions on knowledge and data engineering*, 31(4), 686-705. (2018)
18. Fajardo, S., Kozowyk, P. R., & Langejans, G. H.: Measuring ancient technological complexity and its cognitive implications using Petri nets. *Scientific Reports*, 13(1), 14961. (2023)

19. Weijters, A. J., van Der Aalst, W. M., & De Medeiros, A. A.: Process mining with the HeuristicsMiner algorithm. (2006)
20. Lieben, J., Jouck, T., Depaire, B., & Jans, M.: An improved way for measuring simplicity during process discovery. In Workshop on Enterprise and Organizational Modeling and Simulation (pp. 49-62). Cham: Springer International Publishing. (2018)