SAMO: A Sequential Pattern Mining Model for Evaluating Road Criticality in Urban Traffic Networks

Nourhan Bachir*, Hassan Harb†, Chamseddine Zaki†, and Roland Billen*
*Geospatial Data Science and City Infor. Modelling (GeoScITY), Faculty of Sciences, University of Liège, Belgium
†College of Engineering and Technology, American University of the Middle East, Kuwait
Emails: nbachir@uliege.be, hassan.harb@aum.edu.kw, chamseddine.zaki@aum.edu.kw, rbillen@uliege.be

***Abstract*—This paper proposed a Sequential pAttern Mining mOdel (SAMO) for discovering critical links and evaluating road importance in transportation systems. SAMO introduces novel criticality indices derived from association rule mining and data analysis of vehicular trajectory data. These indices are designed to assess the criticality of road links within a network by leveraging mining frequent patterns and extract meaningful associations from the trajectory data. SAMO prioritizes links that feature prominently in frequent trajectory patterns, with additional weight given to patterns with high confidence levels. We evaluated the performance of our indices within machine learning prediction models alongside with the traditional indices used in the literature. Our results demonstrate that integrating our proposed indices improves prediction accuracy across various models, with the indices consistently ranking among the top features. This indicates the promising potential of our indices for evaluating and assessing critical links in transportation networks.**

***Index Terms*—Criticality Indices, Association Rule Mining, Vehicular Trajectory Data, Machine Learning, Traffic Flow Analysis**

I. INTRODUCTION

With the ever-growing complexity of urban traffic networks, the need for effective methods to assess the criticality of road links has become increasingly paramount. Understanding the criticality of specific links is fundamental for urban planners, transportation engineers, and emergency responders in ensuring the resilience and efficiency of transportation systems. Critical links are those of utmost importance, whose blockage or disruption could significantly impact the entire network, potentially leading to severe congestion, route diversions, or even widespread panic.

In the literature, one can find a significant number of indices and measures that have been proposed to evaluate the criticality of road links. Betweenness Centrality (BC) stands out as one of the most commonly studied indices [1], [2], [3], [4]. Researchers have investigated both unweighted and weighted forms of the BC index, incorporating different weight types such as traffic flow, link length, travel-time, and congestion [2], [3]. Some studies have combined BC with other non-graphical indices, such as link length, clustering coefficient, degree, and road network connectivity [1], [5], [4]. Conversely, certain studies did not consider BC at all, focusing instead on indices like flow and demand [6], [7], [8].

Our literature review indicates a lack of research on applying sequential data mining algorithms for critical link analysis in vehicle trajectory data. While previous studies have utilized these techniques to analyze taxi movement patterns [9], [10] and public transportation movement patterns [11], [12], they have not focused on identifying critical links prone to disruptions or their impact on the transportation network.

This paper presents novel criticality indices derived from association rule mining and data analysis of vehicular trajectory data. These indices aim to provide a comprehensive evaluation of link criticality within the context of the entire network. The underlying concept is to assign higher criticality values to links that are frequently traversed in common vehicle movement patterns. Moreover, links found in longer and more confidently identified patterns are ranked higher, indicating their increased importance in the network. Our new model is called SAMO and it is specifically designed for mining frequent trajectory patterns and extracting meaningful associations from the trajectory data. By incorporating both support and confidence metrics from association rule mining, SAMO can better evaluate a link importance and its connections to critical patterns. The performance of the proposed indices has been evaluated through comparative analysis with traditional indices commonly used in the literature, employing various machine learning models. The integration of these new indices into the models has yielded improved results, with the indices emerging as top features across different models. This underscores their potential for effectively evaluating and prioritizing critical links in urban traffic networks.

The rest of the paper is organized as follows: Section II outlines the methodology employed in this study, including the data collection and analysis techniques. Section III presents the results and analysis, demonstrating the effectiveness of the proposed indices. Finally, Section IV concludes the paper and suggests directions for future research.

II. METHODOLOGY

This section outlines the methodology used to mine sequential patterns from vehicular trajectory data and determine the most commonly traversed road patterns. The methodology introduces a new version of Apriori algorithm adapted to the unique characteristics of trajectory data. The main idea behind this new version is to mine sequential patterns and identify the longest and most commonly used road patterns; more a link appears in such patterns, more important it will considered. Additionally, longer frequent patterns signify higher link criticality. We use the "support" and "confidence" metrics in association rules to measure the likelihood of a link being part of a pattern, thus connecting frequent links and determining their criticality. We combine these metrics into a single index called the "Sequential Impact Score" (SIS).

A. Preliminary Definitions

- *Trajectory Sequence (Tr):* $V = \{V_1, V_2, ..., V_n\}$ represents a group of vehicles that travel for a certain period in a given geographical area. $Tr_i = \{e_x, e_y, ..., e_n\}$ represents the trajectory sequence of vehicle i, and $TrDB$ represents the set of sequences of all vehicles.
- *Sequence length:* The length of a sequence is the number of all links within the sequence, denoted as $length(Tr_k)$.
- *Sub-sequence:* $Tr_i = \{e_x, e_y, \ldots, e_n\}$, the trajectory sequence of vehicle i, is a sub-sequence of the trajectory sequence of vehicle j, $Tr_j = \{\ldots, e_x, e_y, \ldots, e_n, e_m, \ldots\}$, where $length(Tr_i) \leq length(Tr_j)$ and Tr_j contains the whole sequence of Tr_i in the same order.
- *Movement pattern (Mp):* A movement pattern M represents a specific pattern (sub-sequence) to be detected in the trajectory data, denoted as $M = \{e_x, e_y\}$ or $M' = \{e_z\}$.
- *Movement pattern order:* The order of a movement pattern is the length of the respective sequence, denoted as $order(M)$.
- *Movement rule:* A movement rule R is defined as an association rule between two movement patterns expressed as $M \rightarrow M'$.
- *Support:* The support of movement pattern M is the number of appearances of this movement pattern as a sub-sequence in all trajectory sequences S. The support of the rule $R = M \rightarrow M'$ is the support of movement pattern MM' in the mobile database.
- *Confidence:* The confidence of movement rule R is defined as $confidence(R) = support(MM')/support(M)$.
- *Frequent Movement Pattern Set (FqM):* For a given support threshold *minsup*, a frequent movement pattern is a pattern whose support is not lower than *minsup*. A Frequent Movement Pattern Set denoted as FqM_k is the set of movement patterns of order k whose support is not lower than the *minsup*.
- *Confident Rules Set (CR):* For a given confidence threshold *minconf*, a confident rule is a association rule whose confidence is greater than or equal to *minconf*.

B. SAMO Description

An interesting approach for discovering critical links in transportation networks involves extracting meaningful patterns from vehicle movement sequential data. Existing algorithms for sequence pattern discovery, such as association rules algorithms like Apriori, suffer from low accuracy when applied to vehicle trajectory data. To address this issue, this paper presents a new vehicle movement sequential data mining model for sequence pattern mining abbreviated as *SAMO* based on the vehicle trajectory data structure (Figure 1).

1) Vertical Projection of Trajectory Data: In order to efficiently generate frequent rules with high confidence, a vertical projection of trajectory data is performed. Using this approach:

- The database is queried only once, and each unit movement's presence in sequences is projected into one list called *Trajectory Identity List* (TIL).
- To explore the support of combinations, it is enough then to query these lists.

2) New Definitions: New definitions and concepts are proposed, including the concept of outgoing edges and the use of a Positioning Table (PT) to capture the order and possible recurrence of unit movements in trajectory sequences.

- **Outgoing Edges (Ox):** This defines a set of edges that can follow a specific edge. For instance, if edge x can be followed by edges $y_1, y_2, \ldots, y_n$, then $O_x = \{y_1, y_2, \ldots, y_n\}$. This property significantly reduces the search space and complexity when generating frequent sequences.
- **Positioning Table (PT):** This is a data structure used to store and manage the positions of occurrence of movement patterns within trajectory sequences. It consists of Ordered-Positioning Lists (OPLs) for each sequence where a particular pattern occurs. This allows for efficient querying and manipulation of pattern occurrences.
- **Ordered-Positioning List (OPL):** An OPL records the positions of occurrence of a movement pattern within a specific trajectory sequence. It maintains the order of occurrences, enabling accurate tracking of pattern occurrences and extensions.

3) Position Tables Extension Approach (PT-Ext): The PT-Ext algorithm is used to extend the positioning tables of movement patterns. It works by examining each row in the PT of a given pattern X and identifying sequences where X occurs. Then, it checks if an outgoing edge y follows the last occurrence of X in each sequence. If these conditions are met, it extends the pattern X by y and updates the corresponding OPLs in the extended PT. This ensures accurate tracking of pattern extensions while considering the order and recurrence of patterns.

C. SAMO Steps

SAMO employs a path Positioning Table (PT) for data storage, mining, and pattern expansion. The model includes steps for generating outgoing edges, constructing positioning tables, mining frequent movement patterns, extending patterns using PT-Ext, and pruning unnecessary patterns.

D. Position Tables Extension Approach (PT-Ext)

1) **Generate Outgoing Edges:** Explore the graphical representation of the map and determine the set of outgoing edges for each edge. This step establishes the potential

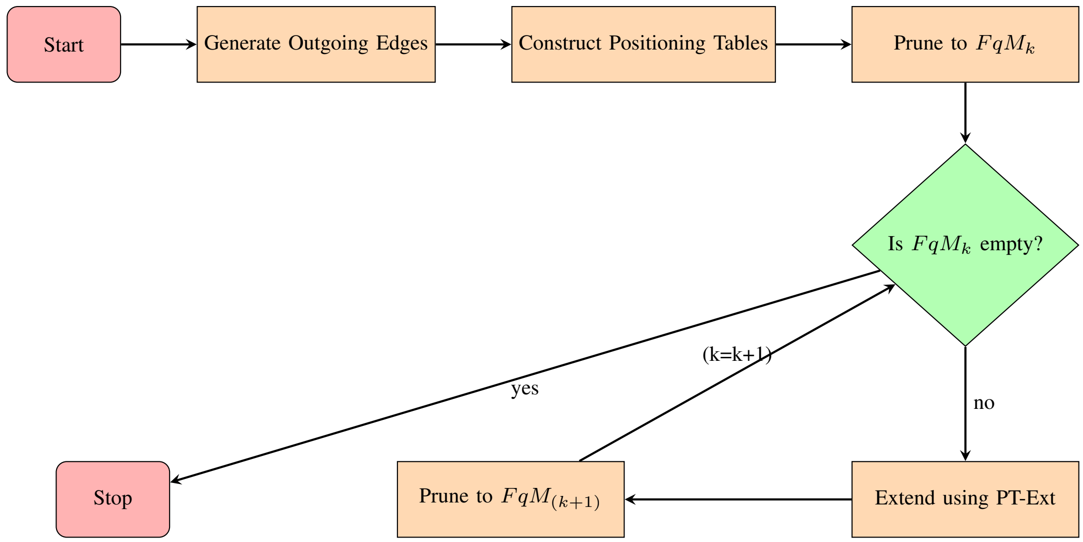

Fig. 1. SAMO

subsequent edges for each edge in the trajectory network.

2) **Construct Positioning Tables (PTs):** Scan the vehicle driving trajectories database to construct PTs for movement patterns of order k. These PTs store the positions of occurrence of each pattern within the trajectory sequences.
3) **Prune to FqM_k:** Prune the generated PTs to identify the set of frequent movement patterns FqM_k of order k. This step eliminates non-frequent patterns and focuses on patterns with sufficient support in the dataset.
4) **Extend using PT-Ext:** Extend the frequent patterns FqM_k into order $k+1$ using the PT-Ext algorithm. This involves identifying sequences where the last pattern in FqM_k is followed by an outgoing edge and updating the PTs accordingly.
5) **Prune to $FqM_{(k+1)}$:** Prune the extended PTs to generate the set of frequent movement patterns $FqM_{(k+1)}$ of order $k+1$. This step further refines the patterns by considering their extensions and support in the dataset.

A final step is then applied **Generate Confident Rules Set** in order to form rules meeting the minimum confidence threshold from *Frequent Movement Pattern Sets*.

E. Indices Calculation

Once all the frequent movement patterns and confident rules are generated, we calculate each edge's indices as follows: let n denote the order of the highest non-empty set of frequent movement patterns with a specific movement pattern M under examination. Then, let $SFI(M)$ (defined in equation (1) be the *Support Frequency Score* of movement pattern M indicating the index calculated using the *support* metric, and $CIS(M)$ (defined in equation (2) be the *Confidence Impact Score* of pattern M indicating the index calculated using the $confidence$ metric as follows:

$$SFI(M) = 1 \cdot x_1 + (i) \cdot x_i + \ldots + (n) \cdot x_n \qquad (1)$$

$$CIS(M) = 1 \cdot y_1 + (i) \cdot y_i + \ldots + (n) \cdot x_n \qquad (2)$$

where:

- The terms $1 \cdot x_1$ and $1 \cdot y_1$ signify the contribution of occurrences of pattern M in the first-order set of frequent movement patterns (FqM_1) and the first-order set of confident rule (CR_1) respectively.
- The terms $(i) \cdot x_i$ and $(i) \cdot y_i$ encapsulates the importance of pattern M in the i^{th} order set of frequent movement patterns (FqM_i) and confident rules (CR_i) respectively. Here, the factor i is utilized to weigh patterns by their order, assigning higher importance to appearances in higher-order patterns.
- Similarly, the terms $(n) \cdot x_n$ and $(n) \cdot y_n$ take into account the occurrences of pattern M in the last non-empty set of frequent movement patterns (FqM_n) and confident rules (CR_n), with a weight proportional to n.

The *Sequential Impact Score* of movement pattern M is defined in the following equation (3):

$$SIS(M) = SFI(M) + CIS(M) \qquad (3)$$

Essentially, it aggregates the occurrences of pattern M across various orders, with each order weighted according to its significance. This approach ensures that the most frequent patterns and confident rules are identified, and higher importance is assigned to links based on their consistent presence in higher-order frequent movement patterns and confident rules. These indices offer a thorough evaluation of the sustained importance of a movement pattern across different levels of sequence complexity.

By calculating the Sequential Impact Scores for each movement pattern, we can gain valuable insights into patterns that consistently show significance across different orders of frequent movement patterns and confident rules. This information is crucial for decision-making and understanding the evolving relevance of these patterns.

III. Results

In this section, we present our experimentation and evaluation of our proposed model. We compared SAMO with the model proposed in [13]. Then, we integrated both models into various machine learning (ML) prediction models. The evaluation focused on predicting the criticality of links using both static and dynamic features. We compared model performance with and without the new indices and analyzed their rankings within different models.

A. SUMO Scenarios and Data Generation

In our study, we decided to make use of LuST [14] and MoST scenarios [15] based on the cities of Luxembourg and Monaco respectively. We ran the SUMO simulations to generate dynamic and static indices. The dynamic features were obtained through SUMO simulations, which provided specific indices for each edge (link). The assessment of the adverse impact following link disruptions in microscopic simulations varies across different studies depending on the specific use case. In order to evaluate link criticality, we conducted extensive simulations by systematically removing one link at a time and recording the change in total round trip time (TRTT). The magnitude of this change indicates the link's importance, i.e.criticality.

B. Parameters, Values, and Justification

In our model, the frequency of pattern occurrences serves as a quantitative measure of the prevalence of specific movement patterns within the dataset. Counting the occurrences of patterns provides us with a foundation for prioritizing patterns for further analysis. By focusing on patterns with higher occurrence frequencies, we refine our understanding of dominant vehicular behaviors while maintaining a data-driven perspective. To do so, we have identified two key parameters, each with specific values and justifications. These parameters play a crucial role in fine-tuning our approach to extract meaningful patterns and relationships from vehicle trajectory data.

1) Minimum Support Threshold: This threshold serves as a fundamental parameter in our methodology. It determines the threshold frequency a pattern must satisfy to be considered for further analysis. To comprehensively explore patterns across various popularity orders, we have chosen a spectrum of threshold values: 3%, 5%, 8%, 10%, and 12%, whose respective minimum number of occurrences is $(6, 465), (10, 776), (17, 242), (21, 552)$, and $(25, 863)$, respectively. This range enables us to strike a balance between capturing rare patterns that might offer unique insights and identifying frequently occurring patterns that could indicate important trends.

As the minimum support threshold is manipulated, a notable phenomenon arises: the frequency of identified patterns changes. Higher thresholds lead to a decrease in the number of identified patterns, as patterns must surpass a higher popularity bar to be considered. This relationship is critical as it ensures that the method remains sensitive to the minimum support threshold while carefully curating patterns with substantive implications.

2) Minimum Confidence Threshold: This threshold parameter is central also for our criticality link analysis process. It quantifies the strength of the rules derived from patterns, indicating the reliability of the associations between different trajectory events. The confidence threshold is expressed as a variable fraction, allowing us to adapt its value according to the characteristics of the dataset under investigation. This adaptability ensures that the rules generated accurately reflect the inherent uncertainty present in real-world vehicle movement data. In our work, we explored different confidence values: $0.6, 0.7, 0.8, 0.9$, and 1. As this value increases only edges in more "confident" rules and corresponding patterns are assigned higher weight.

Indeed, selecting the minimum support and confidence thresholds requires balancing less common patterns and excluding exceedingly frequent ones. Striking this balance ensures that we capture both the long-tail patterns that might provide unique insights and the highly frequent patterns that may underscore critical vehicular interactions. Studying different minimum support thresholds, we observed a variation in the max order reached. As the threshold increases, the max order reached decreases. Since the confidence threshold is also central in our algorithm, we studied the variation of the confident rule (pattern) counts generated as different levels across different minimum support and confidence thresholds. In order to visualize this variation we used a heatmap shown in Figure 2.

In our experimentation, we use 5% and 0.8 as our minimum support and confidence thresholds, respectively. These minimum support and confidence thresholds ensure that we are exploring a sufficient amount of frequent movement patterns and a sufficient number of confident movement patterns are generated.

C. SAMO Results

To assess our model performance, we initially compared the execution time of SAMO with an optimized version of Apriori, a widely-used method in the literature. We enhanced the efficiency of this implementation by utilizing the set of outgoing edges during initialization. However, our findings, illustrated in Figure 3, reveal a significant disparity in execution times. Notably, as the minimum support threshold decreases and the number of filtered edges diminishes, the existing implementation struggles to cope with the growing complexity of combinations, whereas SAMO demonstrates robust performance even under these challenging conditions.

When visualizing SFI, as depicted in Figure 4, we observe the most frequently traveled edges and sequences, with a heightened emphasis on edges within frequent sequences. This is why it highlights the primary roads in the city center while also capturing the popularity of highways. Contrarily, CIS, shown in Figure 5, reflects the frequent edges that are part of the most confident patterns. With a minimum confidence level of 0.8, there is an 80% probability of accessing these edges within a frequent pattern. Consequently, it prioritizes highways and connecting edges. Notably, links in the city center that are highlighted by FqM scores receive a lower score in CM because there are more alternative routes in the city center.

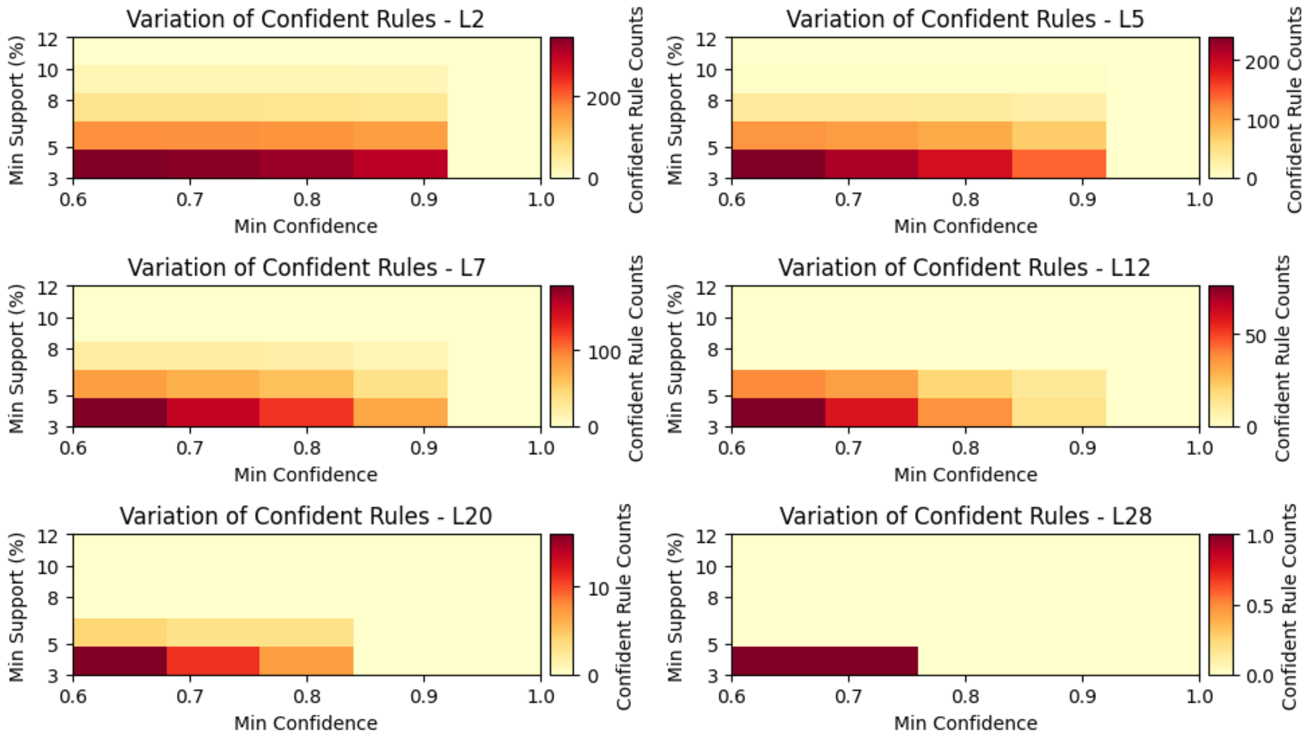

Fig. 2. Confident Rule Counts for Different Orders

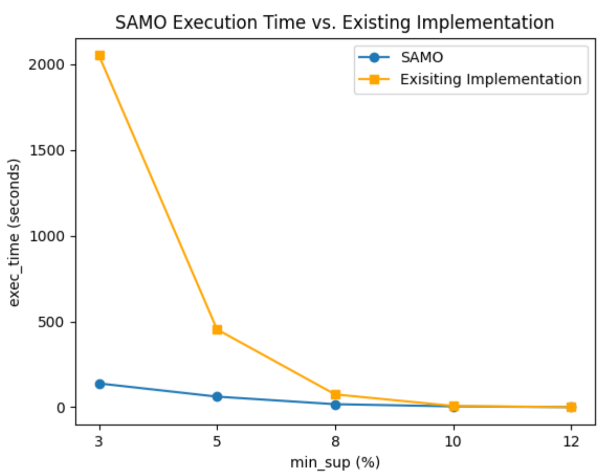

Fig. 3. SAMO Execution Time vs. Enhanced Apriori Algorithm

D. Additional Indices for Prediction Model

Beside the proposed indices, we utilized a set of static and dynamic indices derived from the SUMO simulation output. These indices provide valuable insights into various aspects of traffic flow and network characteristics. Below is a concise explanation of the key indices used:

- **Static Indices:** Length, Width, Max Speed, Cost, EBC, Type.
- **Dynamic Indices:** Support, Relative Support, Sampled Seconds (in sec.), Travel Time (in sec.), Overlap Travel Time (in sec.), Density (in #veh/km), Occupancy (in %), Speed (in m/s), Speed Relative.

The above indices, along with the proposed indices, contribute to a comprehensive understanding of traffic dynamics and network behavior, enhancing the accuracy of criticality assessment and prediction model performance evaluation.

E. Model Results

We employed the following machine learning models for prediction: Random Forest, Gradient Boosting, Linear Regression, K-Nearest Neighbors, Ridge Regression, MLP, Support Vector Regression, Lasso Regression, Decision Tree, and Gaussian Process Regression. The mean squared error (MSE) results for each model, both with and without the proposed indices, are presented in Table I. The results show that integrating the proposed indices into the prediction models consistently improves their performance. The reduction in MSE across various models indicates that the new indices provide valuable information about link criticality, leading to more accurate predictions.

F. Model Top Features Results

We further analyzed the performance of models trained on top features. Table II presents the updated rankings of CIS, SFI, and SIS for selected models within the top features. The updated rankings reveal that our proposed indices, particularly CIS, continue to demonstrate significant importance across different machine learning models. CIS's high ranking in several models highlights its robustness in capturing link criticality effectively. The varying performance of SFI and SIS across

Fig. 4. SFI Score Weighted Links on city of Luxembourg

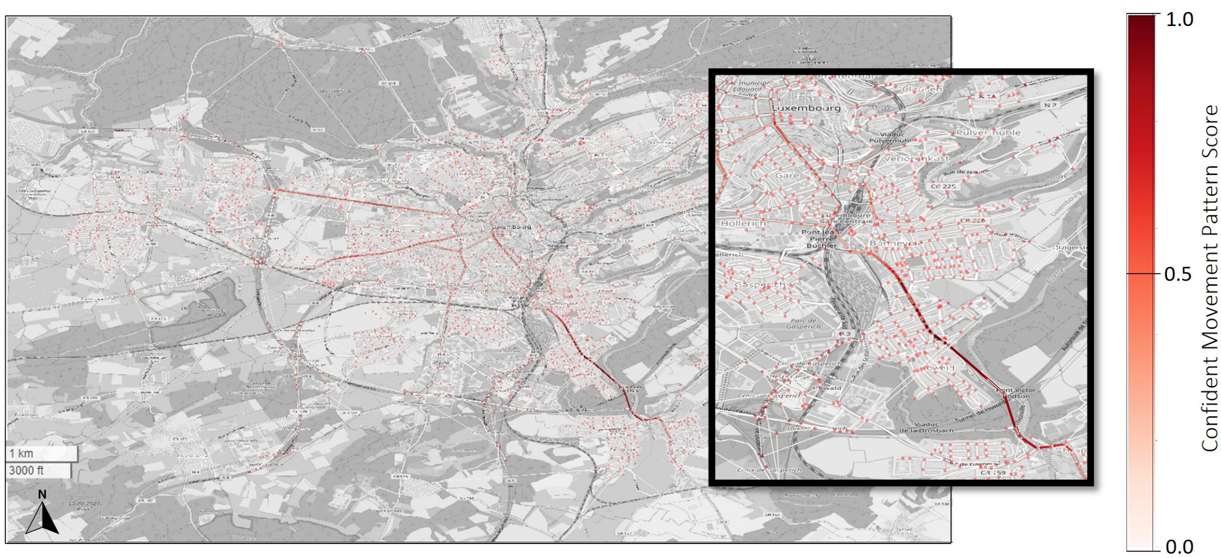

Fig. 5. CIS Score Weighted Links on city of Luxembourg

models suggests their sensitivity to specific model architectures and feature selections. Particularly, Random Forest and Gradient Boosting emerged as the best-performing models, with the lowest MSE values both before and after the inclusion of the proposed indices. These models also placed CIS in the 4th and 2nd positions respectively, reflecting the index's high relevance and contribution to the model's predictive accuracy.

G. Analysis and Insights

The inclusion of the proposed indices led to significant improvements in model performance, as evidenced by the reduction in MSE values across all models. This indicates that the proposed indices capture critical aspects of link importance that are not fully represented by traditional features. The analysis of top features in various models further highlights the significance of our proposed indices:

- **CIS** ranked within the top 5 features in several models, demonstrating its robustness in capturing link criticality.
- **SFI** and **SIS** showed more variability in their rankings, indicating that their effectiveness might depend on the specific model and context. For instance, SFI did not appear in the top features for some models but ranked 8th and 18th in others, showing moderate importance.

The most frequently top-ranked features alongside our indices include occupancy, density, ebc (edge betweenness centrality), and various traffic volume measures. This suggests that both dynamic traffic conditions and static network characteristics are crucial for accurately predicting link criticality. Overall, our proposed indices provide a valuable addition to

TABLE I
MSE RESULTS FOR MACHINE LEARNING MODELS.

Model	MSE without Indices	MSE with Indices	MSE with Indices and Top Features
Random Forest	0.4238	0.3897	0.3782
Gradient Boosting	0.4849	0.4143	0.3977
K-Nearest Neighbors	0.4896	0.4258	N/A
Linear Regression	0.5116	0.4462	0.4499
Ridge Regression	0.5172	0.4565	0.4501
MLP	0.5732	0.4663	N/A
Decision Tree	0.8654	0.4665	0.4665
Support Vector Regression	0.5068	0.4912	N/A
Lasso Regression	0.7609	0.5698	0.5683
Gaussian Process Regression	170.1954	138.9047	N/A

TABLE II
RANKINGS OF PROPOSED INDICES IN TOP FEATURES

Model	CIS	SFI	SIS
Random Forest	4th	18th	27th
Gradient Boosting	2nd	N/A	N/A
Linear Regression	23rd	22nd	39th
Ridge Regression	12th	24th	27th
Decision Tree	3rd	8th	36th
Lasso Regression	4th	N/A	N/A

traditional metrics, offering a more comprehensive evaluation of link criticality. The improvement in prediction accuracy and the consistent presence of our indices among top features across different models underscore their potential for practical applications in traffic management and network optimization.

IV. CONCLUSION

In this paper, we proposed novel criticality indices based on association rule mining and data mining of vehicular trajectory data. Our indices aim to evaluate the criticality of links (roads) within a network by considering their frequency and confidence in vehicle trajectory patterns. Through extensive experiments using various machine learning (ML) prediction models, we demonstrated the effectiveness of our proposed indices in predicting link criticality. Integration of the indices into the models consistently led to improvements in prediction accuracy, as evidenced by the reduction in mean squared error (MSE) values across different models. Moreover, the analysis of top features within the models revealed the significant relevance of our proposed indices, particularly the Criticality Impact Score (CIS), which consistently ranked among the top features across various ML models. The high ranking of CIS indicates its robustness in capturing link criticality effectively. Furthermore, our indices complement traditional static and dynamic features used in traffic analysis, providing a more comprehensive assessment of link criticality. By considering both dynamic traffic conditions and static network characteristics, our indices offer valuable insights for traffic management and network optimization. Additionally, the comparative analysis showcased the robustness and efficiency of our SAMO model in handling large-scale trajectory datasets. SAMO outperformed the existing implementation of Apriori, particularly when the minimum support threshold decreased, indicating its suitability for trajectory mining tasks in transportation research. The results of our study underscore the potential of our proposed indices for practical applications in transportation engineering and urban planning.

REFERENCES

[1] Y. Feng, H. Wang, C. Chang, and H. Lu, "Intrinsic correlation with betweenness centrality and distribution of shortest paths," *Mathematics*, vol. 10, no. 14, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/14/2521

[2] M. Akbarzadeh, S. Memarmontazerin, S. Derrible, and F. Salehi, "The role of travel demand and network centrality on the connectivity and resilience of an urban street system," *Transportation*, vol. 46, 08 2019.

[3] P. Gauthier, A. Furno, and N.-E. El Faouzi, "Road network resilience: How to identify critical links subject to day-to-day disruptions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, p. 036119811879211, 08 2018.

[4] R. el rashidy and S. Grant-Muller, "An assessment method for highway network vulnerability," *Journal of Transport Geography*, vol. 34, p. 34–43, 01 2014.

[5] F. Li, H. Jia, Q. Luo, Y. Li, and L. Yang, "Identification of critical links in a large-scale road network considering the traffic flow betweenness index," *PLOS ONE*, vol. 15, p. e0227474, 04 2020.

[6] E. Jenelius, "Redundancy importance: Links as rerouting alternatives during road network disruptions," *Procedia Engineering*, vol. 3, pp. 129–137, 01 2010.

[7] D. Scott, D. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, pp. 215–227, 05 2006.

[8] E. Jenelius, T. Petersen, and L.-G. Mattsson, "Road network vulnerability: Identifying important links and exposed regions," vol. 20, 06 2006.

[9] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 855–866, 03 2019.

[10] R. Ibrahim and M. O. Shafiq, "Detecting taxi movements using random swap clustering and sequential pattern mining," *Journal of Big Data*, vol. 6, no. 1, p. 39, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0203-6

[11] Y. Wang, Y. Tian, B. Yang, J. Wang, X. Hu, and S. An, "Planning flexible bus service as an alternative to suspended bicycle-sharing service: A data-driven approach," *Journal of Advanced Transportation*, vol. 2023, pp. 1–15, 01 2023.

[12] S. Hu, Q. Liang, H. Qian, J. Weng, W. Zhou, and P. Lin, "Frequent-pattern growth algorithm based association rule mining method of public transport travel stability," *International Journal of Sustainable Transportation*, vol. 15, 09 2020.

[13] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," *Proc. 20th Int. Conf. Very Large Data Bases VLDB*, vol. 1215, 08 2000.

[14] L. Codecá, R. Frank, S. Faye, and T. Engel, "Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52–63, 2017.

[15] L. Codeca and J. Härri, "Monaco SUMO Traffic (MoST) Scenario: A 3D Mobility Scenario for Cooperative ITS," in *SUMO 2018, SUMO User Conference, Simulating Autonomous and Intermodal Transport Systems, May 14-16, 2018, Berlin, Germany*, Berlin, GERMANY, 05 2018.