ARTICLE TEMPLATE

When driving becomes risky: Micro-scale variants of the lane-changing maneuver in highway traffic

Amna Qayyum[a], Bernard De Baets[b], Samuel Van Ackere[a], Frank Witlox[c], Guy De Tré[d] and Nico Van de Weghe[a]

[a]CartoGIS, Department of Geography, Ghent University, 9000 Ghent, Belgium;
[b]KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, 9000 Ghent, Belgium;
[c]SEG, Department of Geography, Ghent University, 9000 Ghent, Belgium;
[d]DDCM, Department of Telecommunications and Information Processing, Ghent University, 9000 Ghent, Belgium

ABSTRACT
Objective: Vehicular lane-changing is considered to be one of the riskiest driving maneuvers. Since vehicular automation is quickly becoming a reality, it is crucial to be able to identify when such a maneuver can turn into a risky situation. Recently, it has been shown that a qualitative approach: the Point Descriptor Precedence (PDP) representation, is able to do so. Therefore, this study aims to investigate whether the PDP representation can detect hazardous micro movements during lane-changing maneuvers in a situation of structural congestion in the morning and/or evening. *Method*: The approach involves analyzing a large real-world traffic dataset using the PDP representation and adding safety distance points to distinguish subtle movement patterns. *Results*: Based on these subtleties, we label four out of seven and five out of nine lane-change maneuvers as risky during the selected peak and the off-peak traffic hours respectively. *Conclusions*: The results show that the approach can identify risky movement patterns in traffic. The PDP representation can be used to check whether certain adjustments (e.g., changing the maximum speed) have a significant impact on the number of dangerous behaviors, which is important for improving road safety. This approach has practical applications in penalizing traffic violations, improving traffic flow, and providing valuable information for policymakers and transport experts. It can also be used to train autonomous vehicles in risky driving situations.

KEYWORDS
Pattern recognition; road safety; spatiotemporal modeling; vehicle interactions; data mining; intelligent transportation system

INTRODUCTION

Nowadays, video surveillance of traffic is becoming increasingly popular for a variety of security-related reasons. Over the past decade, video surveillance has been transformed by fundamental changes in the way digital data is captured, stored and shared, raising the need for methods allowing to unlock the information contained therein. Average

---

traffic flow metrics such as throughput, travel time and vehicle speed can be easily obtained from such data. What cannot be readily determined, however, is the point at which a safe driving maneuver may turn into a risky one or lead to an accident despite all safety precautions. It is therefore necessary to determine the circumstances under which recurring movement patterns can potentially become risky, especially as the era of autonomous vehicles becomes inevitable. Consequently, traffic authorities and policy makers are in high need of methods to analyze traffic movements for identifying potentially dangerous maneuvers. This requires methods that can comprehensively analyze data at the micro-scale.

From a micro-analysis perspective, traffic movement patterns have been mainly analyzed for some specific inter-vehicle crash types, e.g., rear-end, crossing, and merging (Huguenin, Torday, & Dumont, 2005;Mahmud, Ferreira, Hoque, & Tavassoli, 2019). Some studies also considered lane changing, frontal, run-off, hit pedestrian accidents (Dijkstra et al., 2010;Mak & Sicking, 2003;Stone, Chae, & Pillalamarri, 2002) and the risk of fatal accidents by increasing driver's average speed (Abdel-Aty & Pande, 2005;Mills, Freeman, Truelove, Davey, & Delhomme, 2021;Wu, Song, & Meng, 2021). Besides, different attempts have been made to extract traffic movement patterns by using quantitative approaches as in Mathew and Benekohal (2021); Sarkar, Bhaskar, Zheng, and Miska (2020); Zhang and Wang (2019), but due to insufficient prior knowledge, these approaches are suitable only in very specific circumstances. Mostly, the safety-related data is aggregated into summary statistics (such as the total number of crashes/conflict), which constitutes a major simplification of the real risk levels that change over time. As often not all numerical information about an event is sufficiently precise, hence, qualitative approaches enter the picture (Parsons & Parsons, 2001). A qualitative description of large amounts of quantitative data can even save memory and computational resources (Freksa, 1992), enabling the processing and analysis of more complex vehicle interactions. However, existing qualitative methods for studying large-scale motion data are often limited in scope and sophistication (Long & Nelson, 2013;Solanki, 2021).

In the context of micro-scale traffic analysis, several micro-simulation tools are available, such as PARAMICS (Fritzsche, 1994), AIMSUM (Huguenin et al., 2005), VISSIM (Dedes et al., 2011), MATSim (Horni, Nagel, & Axhausen, 2016), and SUMO (Silgu, Erdagi, Göksu, & Celikoglu, 2021). However, so far, none of these tools have been used to find recurring movement patterns in motion data. In TripVista (Guo, Wang, Yu, Zhao, & Yuan, 2011), an attempt has been made to discover traffic patterns at a road intersection; similar efforts have also been made by Liu, Liu, Ni, Fan, and Li (2010), Liu, Liu, Ni, Li, and Fan (2012), and Pu, Liu, Ding, Qu, and Ni (2013). However, handling of these tools requires substantial technical skills. Besides, they solely rely on quantitative techniques for data mining and visualization. In this work, we explore a qualitative approach for extracting traffic patterns, allowing for greater flexibility in the analysis (Queirós, Faria, & Almeida, 2017).

Some of the existing qualitative spatiotemporal calculi, like the Qualitative Trajectory Calculus (QTC) introduced by Van de Weghe, Kuijpers, Bogaert, and De Maeyer (2005) and the Qualitative Rectilinear Projection Calculus (QRPC) introduced by Glez-Cabrera, Álvarez-Bravo, and Díaz (2013), target traffic modeling at the microscale. However, these calculi are rather complex and have not yet been used for pattern extraction in traffic. Another qualitative approach is the Point-Descriptor-Precedence (PDP) representation introduced by Qayyum et al. (2021), which describes the relative motion of disconnected moving objects via a system of relational symbols ($<$, $=$, $>$). We adopt this representation as it is fairly easy to grasp and is far simpler than other

qualitative approaches used in the literature. It has already proven useful for extracting micro-movement patterns in curved movement directions on a simulated traffic dataset of a non-signalized T-intersection (Qayyum et al., 2022). However, the main objective of this work is to establish whether the PDP representation really works for real-world traffic dataset and up to which extent. Can this qualitative representation distinguish between risky and non-risky traffic movement patterns, and what are the merits of this representation over previous ones?

For ease of understanding, we illustrate our approach using a dataset extracted from a real-time traffic video stream of a multi-lane highway on a given Monday. Highways provide an excellent example of a confined motion environment where vehicles can only move in one direction and the degree of freedom is limited. Hence, vehicles can only change lanes, for example when overtaking. For this reason, we focus on detecting only one type of vehicle interaction, i.e., the lane-changing maneuver using an object detection algorithm based on deep learning. Note that various (quantitative) pattern recognition methods have already been proposed in the literature for identifying lane-changing maneuvers. One novel approach utilized a combination of Hidden Markov Model (HMM), Divisive Hierarchical Clustering (DHC), and Dynamic Time Wrapping (DTW) to accurately identify lane changes (Klitzke, Koch, & Köster, 2020). Mostly, these pattern recognition methods include Support Vector Machines (SVM) and Artificial Intelligence (AI) (Khelfa, Ba, & Tordeux, 2023). The mean detection accuracy using SVM ranges from 97% to 99%, while the detection time ranges from 20 to 22 ms under various road conditions. In contrast, the recognition accuracy of these AI models was only 80%. The computational complexity of these methods increases exponentially with larger datasets, so our qualitative approach is advantageous in terms of speed and scalability.

Since we obtained a 24-hour surveillance video, the computational cost for semantic segmentation of this video is quite high. Therefore, we select three morning hours (07:00–10:00) from the video. Then, based on the number of vehicles and apparent congestion times, we divide these hours into the two peak hours (7.30–9.30 am) and one off-peak hour composed of two half-hour windows (7:00–7:30 am and 9:30–10:00 am). The aim is to find out whether the driving behavior during the peak hours and that during the off-peak traffic hour differs, and whether there are unusual traffic interactions that can cause unsafe or risky situations. Existing studies commonly choose different traffic hours depending on the context of the analysis, e.g., (Andrienko et al., 2009;Lateef, 2011;Mandal et al., 2011;Pu et al., 2013), and (Poco et al., 2015). We use YOLOv4 (an object detection algorithm based on deep learning) Bochkovskiy, Wang, and Liao (2020) to extract the xyt-coordinates of the vehicles present in the selected hours to create a motion dataset. Again, the core here is not to dig deeper into the object detection algorithms, nevertheless interested readers are urged to check Bochkovskiy et al. (2020) for understanding the YOLOv4 model.

Then, we identify how many variants of the lane-changing maneuver are present in the dataset and which groups of vehicles show this variant. To do so, we first convert the xyt-coordinates of the vehicles into their equivalent PDP representations. Then, we compare these representations with one another to find out how vehicles undertake the lane-changing maneuver and how many variants of this maneuver exist. The whole process is carried out through a Python script, which is based on the procedure proposed by Qayyum et al. (2021). Since the vehicles detected in the video are automatically represented through their xyt-coordinates, it becomes possible to calculate their speed and safety distances. Hence, based on these parameters, we can label the identified variants as risky or non-risky. It should be noted that the proposed

method for identifying such variants is new and is not intended to be compared with conventional safety measures such as time to collision (TTC) and time after collision (PET), etc. This comparison requires a completely new study and is therefore beyond the scope of this paper.

The remainder of this paper is organized as follows. First, the Method Section describes the PDP representation and our approach to obtain a general view of the variants of the lane-changing maneuver detected during the selected hours, along with their frequency of occurrence. Then the Results Section describes the micro-scale subtleties found in these variants and provides our recommendations for categorizing them as risky or non-risky. Finally, the results, shortcomings, and potential applications of our approach are discussed in the Discussion Section.

METHOD

Since our approach is based on the Point-Descriptor-Precedence (PDP) representation introduced by Qayyum et al. (2021), it is important to first provide some insight into this representation.
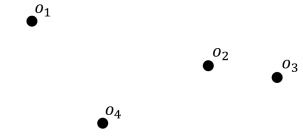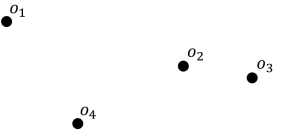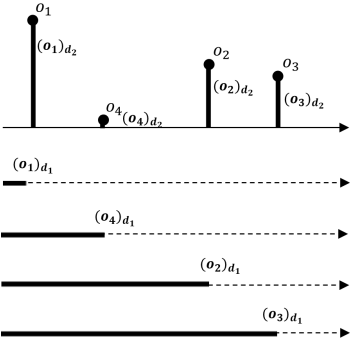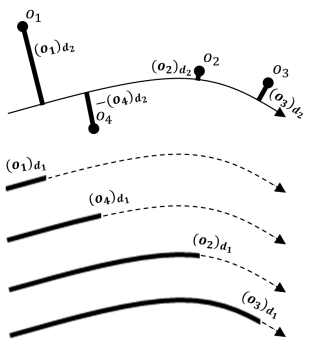
The Point-Descriptor-Precedence (PDP) Representation

The PDP representation expresses point configurations and their movements qualitatively as shown in Fig. 1. The ordering between the (moving) points is represented in terms of relational symbols ($<,=,>$) in the desired direction of interest. For instance, if the points are moving in a straight direction as in the case of vehicles on highways, then the precedence is derived for a straight directed line; or if the direction of interest is curved as in the case of intersections and roundabouts, then the precedence is derived for a curved directed line. The precedence is expressed using relational symbols through a number of distance descriptors. There can be multiple descriptors, but in this paper, we only use two descriptors d_1 and d_2. Usually, descriptor d_1 describes the distance of points traveled along the directed line and d_2 describes the signed orthogonal distance of the points to the directed line measured positive on the left, and negative on the right with respect to the direction. Figure 1 illustrates the PDP representation for a straight and a curved directed line. We show the same configuration of four points (o_1–o_4) in both cases. We use two descriptors d_1 and d_2 to describe the distance relations of these points w.r.t. the directed line in the form of precedences. Hence, the configurations are transformed into the PDP representations in terms of d_1- and d_2-precedences. It is important to note here that the same configuration of points results in different precedences as shown in Fig. 1, illustrating that the PDP representation is capable to highlight subtleties present in the movements.

Detecting Traffic Interactions

In this research, the following assumptions have been made in context of data collection and analysis:

(i) The moving objects are assumed as moving points since we are not interested in the mass or other dimensions of the objects.
(ii) The width of the lanes is assumed as constant.

	PDP	
	Straight Directed Line	**Curved Directed Line**
POINT	o_1, o_2, o_3, o_4	o_1, o_2, o_3, o_4
DESCRIPTOR	o_1, $(o_1)_{d_2}$, o_2, $(o_2)_{d_2}$, o_3, $(o_3)_{d_2}$, o_4, $(o_4)_{d_2}$; $(o_1)_{d_1}$, $(o_4)_{d_1}$, $(o_2)_{d_1}$, $(o_3)_{d_1}$	o_1, $(o_1)_{d_2}$, o_2, $(o_2)_{d_2}$, o_3, $(o_3)_{d_2}$, o_4, $(o_4)_{d_2}$; $(o_1)_{d_1}$, $(o_4)_{d_1}$, $(o_2)_{d_1}$, $(o_3)_{d_1}$
PRECEDENCE	$\begin{cases} d_1:\ o_1 < o_4 < o_2 < o_3 \\ d_2:\ o_4 < o_3 < o_2 < o_1 \end{cases}$	$\begin{cases} d_1:\ o_1 < o_4 < o_2 < o_3 \\ d_2:\ o_4 < o_2 < o_3 < o_1 \end{cases}$

$(o)_{d_1}$: Distance of point o along the directed line represented by d_1

$(o)_{d_2}$: Signed orthogonal distance of point o to the directed line represented by d_2

Distance of the point

Figure 1. A simple illustration of the PDP representation using straight and curved directed lines (Qayyum et al., 2021).

(iii) Movements in longer time periods can be described as a sequence of PDP-static representations.

(iv) Each PDP representation holds a particular position of the moving objects at a specific instant in time. No intermediate position is available in between the two PDP representations.

(v) For computational ease and analysis purposes, we discretized continuous time and performed our micro traffic-analyses all at a temporal resolution of 10 Hz.

(vi) In PDP, we define the precedence of points as a function of certain threshold settings to minimize the effect of noise.

Having illustrated the basic principles of the PDP representation, we now turn to the process underlying the creation of the dataset and the procedure used to identify traffic interactions based on the PDP representation. By 'interaction' we refer to a group of vehicles displaying a particular variant of the lane-changing maneuver. This will become clear in the following subsections.

Dataset

In order to realize the objectives of this research, it is important to choose a suitable location where the CCTV traffic camera clearly captures the flow of vehicles. The

camera should be placed in such a way that a movement dataset can be generated relatively easily by transforming the vehicles in the video into their xyt- coordinates. We assumed that a large difference in traffic interactions might be observed between the peak traffic hours and other moments of the day. As shown in Fig. 2, we chose the multi-lane dual E411-A4 highway in Jezus-Eik, Belgium, as our location. Although the image quality of the CCTV camera was not perfect, the vehicles were still visible. The camera was fixed and the surveillance video was recorded for 24 hours on a given Monday. The left-hand side (LHS) of this highway consists of four lanes: a merge lane (lane 4), one slow lane (lane 3) and two lanes for faster driving and overtaking other vehicles (lanes 1 and 2). The right-hand side (RHS) also consists of four lanes: one exit lane (lane 1) and three other lanes. We chose the left side of the motorway to study the interactions between vehicles because it contains a merge lane. Although merging seems to be a simple maneuver, it is one of the main causes of accidents. Side collisions, rear-end collisions, blind spot accidents and overtaking accidents occur frequently, making the merge lane on the motorway the most dangerous lane. Another reason for choosing the LHS was the occurrence of traffic jams in the morning, while the RHS neither had traffic jams nor busy traffic hours during the whole day.

Figure 2. The two-way multi-lane E411-A4 highway selected for the analysis, situated in Jezus-Eik, Belgium.

Since a 24h surveillance video contains thousands of vehicles and the computational cost of object detection for a full-day surveillance video is rather high, a vehicle count was performed between 12 am and 12 pm to identify the peak hours. Vehicles were counted in the first five minutes of each hour in each lane. A fixed crossing method was used to avoid double counting. Based on this counting process, the graph in Fig. 3 was created, showing that the peak traffic at LHS occurred between 7:30 am and 9:30 am. The video also clearly showed that there was extensive traffic congestion between 7:30 am and 9:30 am (Fig. 3). Based on these statistics, we selected 7:30 am to 9:30 am as the peak traffic hours for our analysis. By peak hours, we mean traffic congestion on the highway, apart from incidents such as traffic accidents, traffic light failures and construction works. The half hour before and the half hour after the peak hours were

jointly considered as the off-peak traffic hour (7:00-7:30 am and 9:30-10:00 am).

The segments covering the selected peak and off-peak traffic hours were extracted from the full-day video. To reduce the computation time for object detection, the temporal resolution of the video segments was reduced from 50 frames per second to 10 frames per second. The next step was to extract the xyt-coordinates (trajectories) of the vehicles present during these hours. For this purpose, we used YOLOv4, a well-established deep learning object detection algorithm Bochkovskiy et al. (2020). The vehicles detected were represented by their xyt-coordinates, which were automatically stored using Python's pickle file format with a resolution of 10 Hz (the pickle file format consumes 50% less memory compared to the CSV format Madhukar (2020)). A total of 75,590 vehicles were detected during the selected hours.
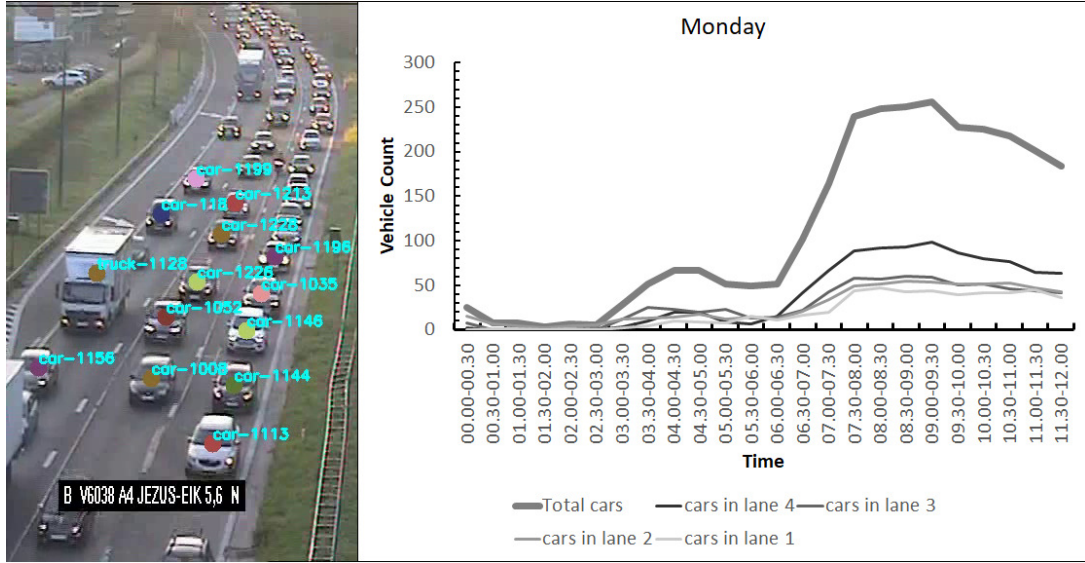
Figure 3. Left: A view of the traffic congestion at 7.30 am on the LHS of the highway. Right: Vehicle count graph showing a traffic peak between 7.30–9.30 am.

Retrieving Interactions and Their Instances

The object detection part generated a dataset of trajectories of 75,590 vehicles. The next step in the analysis is to detect recurring interactions between these vehicles and to see if variations thereof exist during the peak and off-peak traffic hours. The main purpose is to get an overall picture of the traffic in the selected hours, including "How many types of interactions are there?", "What are the most common interactions and how often do they occur?". For this purpose, we followed the procedure described in Qayyum et al. (2021) to extract movement patterns out of a motion dataset (with some slight modifications), which is reproduced in Appendix A as Procedure 1 for ease of reference. We briefly explain this procedure.

The first step of Procedure 1 is to denote our main dataset of 75,590 vehicles (O^T) as the target dataset T containing trajectories for 36,012 timestamps. The second step is the creation of the reference interactions that we wanted to find in the target dataset. Each reference interaction is represented by a reference dataset R. When reviewing the videos of the selected hours, we noticed nine interesting interactions

(Interactions A–I) in the off-peak traffic hour and seven interactions (Interactions J–P) in the peak traffic hours. These interactions are shown in Appendix B and Appendix C, respectively. All of these interactions are variants of the lane-changing maneuver. Given these interactions, we created nine reference datasets (R_A–R_I) for the off-peak traffic hour and seven reference datasets (R_J–R_P) for the peak hours. The main difference between the procedure described in Qayyum et al. (2021) and our approach is that we have used 16 reference interactions (datasets) instead of a single one. Each reference dataset contains an interaction among a different number of vehicles (O^R). For example, Interaction A involves three vehicles (O^R), Interaction E involves four vehicles, and Interaction O involves six vehicles. In the third step of the procedure, an empty list L is created for storing the groups of vehicles matching our reference datasets, to display the result at the end.

Steps 4–10 of the procedure involve a pre-processing of the target dataset, where the number of timestamps in the target dataset was reduced to avoid repetitive calculations. To do this, first, the target dataset needs to be represented in its PDP representation. Hence, we defined the first descriptor d_1 in the vertical driving direction (y-axis) and the second descriptor d_2 in the direction orthogonal to d_1 (x-axis). Then, based on these descriptors, for each timestamp, we represented the ordering of the vehicles present in the target dataset in a d_1- and d_2-precedence. Next, all these precedences are compared to remove those timestamps with redundant d_1- and d_2-precedences, creating a new dataset T^{new} with significantly fewer timestamps (18,000) than the original target dataset.

Steps 10–30 involve comparing the precedences of the reference datasets with those of the target dataset to find out which groups of vehicles matched with the 16 reference interactions that we have created. First, the target dataset is converted into a list of tuples according to the number of vehicles present in the reference dataset. For example, the reference dataset R_A of Interaction A has three vehicles (V_a, V_b, V_c). So the 75,590 vehicles of the target dataset are arranged in a set P of all possible 3-tuples: (V_1, V_2, V_3), (V_1, V_2, V_4), (V_1, V_2, V_5), and so on. Next, the precedences of the reference dataset (V_a, V_b, V_c) are compared with each of the target tuples. If the precedences of a reference dataset (say R_A) matched with those of any tuple of the target dataset irrespective of the time length involved, then that tuple is stored in the list L. Finally, the list L stores all tuples that matched with a given reference dataset. For instance, seven tuples of the target dataset matched with the reference dataset (R_A) of the Interaction A (see Appendix B): ($V_{5070}, V_{5049}, V_{5071}$), ($V_{6024}, V_{5971}, V_{5970}$), ($V_{6626}, V_{6572}, V_{6581}$), ($V_{7623}, V_{7717}, V_{7616}$), ($V_{8121}, V_{8132}, V_{8112}$), ($V_{8758}, V_{8747}, V_{8726}$), and ($V_{11424}, V_{11449}, V_{11447}$). Hence, we can say that there are seven instances of Interaction A in the target dataset, or, said otherwise, Interaction A occurred seven times during the selected hours. These instances are shown as "matches" in Appendix B and Appendix C in the form of tuples along with each reference interaction. Interested readers are further advised to consult Qayyum et al. (2021) for a detailed explanation of Procedure 1.

RESULTS

It is important to note that the target dataset contained inaccuracies due to the lower quality of the camera images. Therefore, deriving accurate PDP representations for all vehicles was not possible. The equality relation ($=$) between the coordinates of vehicles coordinates was difficult to obtain. To account for this, we added external points in the

range of 45 pixel units around the centroid of each vehicle as a buffer. This means that the positions of the vehicles were considered approximately equal ($\approx$) if the difference between their coordinates was found in the range $[-45, +45]$ while using the respective descriptor to derive precedences.

The final list L produced as a result of Procedure 1 also shows how often an interaction has occurred during that selected hour. For example, Interaction A occurs seven times in the off-peak traffic hour based on the instances found; Interaction B occurs 11 times and Interaction C four times. The same is calculated for the peak hours; for instance, Interaction J occurs 25 times, Interaction K 83 times, and so on. Based on these results, we created the graphical representation in Fig. 4, which gives an overall picture of the traffic interactions found in the peak and the off-peak traffic hours, along with their frequency of occurrence. Figure 4 also shows how many instances of these interactions are categorized as risky and non-risky, which is further elaborated in the upcoming Section.

From this graphical representation, one can see the top five interactions in the off-peak traffic hour, which are Interactions D, F, B, A, and H. All these five occurrences involved lane-changing either from the left or right lane (see Appendix B: Figs. B1–B4). Hence, in our case, we can assume that during the off-peak traffic hours, when there is usually less or no traffic, people tend to make speedy lane-changing maneuvers either from the left lane or the right lane. The top five interactions found during the peak hours are Interactions K, L, J, M, and N, shown in Appendix C. These interactions imply that during rush hour or in a traffic jam, people tend to change lanes when they see an opportunity to do so and often try to drive in that lane where traffic is moving the fastest. In the next section, we will describe how this lane-changing behavior can lead to risky driving conditions and how we can detect this with the PDP representation.
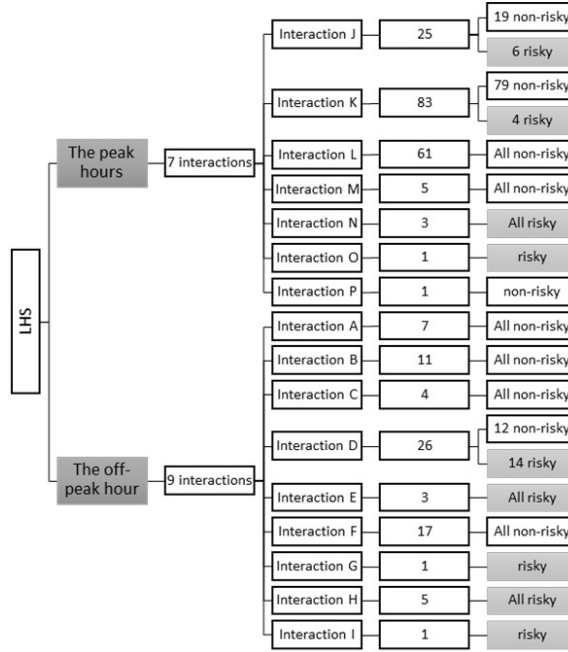
Figure 4. Graphical representation giving an overall view of the traffic interactions found in the peak and the off-peak traffic hours along with their frequency.

Categorizing the Interactions as Risky and Non-Risky

In the Method Section, we explained how relevant interactions between vehicles can be identified. To determine which of these interactions could be risky, we need to identify micro-scale variations between them. Again, the PDP representation is used since it allows for the detection of subtleties in the motion data. How this can be done is discussed in this section by examining the interactions during the peak and off-peak traffic hours separately.

The Peak Traffic Hours

During the selected peak hours, we found seven relevant interactions (Interactions J–P), as shown in Appendix C. To identify subtleties in the interactions pointing at risky conditions, we added an external point at the front and one at the rear of each vehicle as 'safety distance points'. These points can easily be calculated from the xyt-coordinates of the vehicles. By visualizing these safety distance points of the vehicles, it can be seen when a vehicle is maintaining a proper distance from the other ones and when there could be a dangerous situation. Situations involving sudden braking or improper lane-changing might be detected by adjusting the centroid and safety points of the second vehicle with the rear of the first vehicle. To understand this, we consider Interaction J found during the peak hours, as shown in Fig. C1 in Appendix C, along with its 25 instances. Here, vehicle b is changing lanes from left to right. To check for subtleties, we first recreated this interaction by adding two external safety distance points s^- and s^+ at the rear and front end of each vehicle. A glimpse of the recreated Interaction J is depicted in Fig. C4 for only two timestamps t_1 and t_2 and two lanes (lanes 2 and 3).

As a next step, we tried to find matches with our reference datasets in the target dataset by following Procedure 1. Interestingly, only 19 out of 25 instances matched with Interaction J. The remaining six instances were not detected because the d_1-precedences of these six instances showed a change in the ordering of vehicles b and e at timestamp t_2. This change is highlighted in red in Fig. C4. Here, vehicle b was too close to vehicle e during the lane change, which could have led to a collision. In the non-risky situation, these two vehicles had to maintain the relation $e^{s^+} < b^{s^-}$ in the direction of descriptor d_1, which was reversed to $b^{s^-} < e^{s^+}$ when vehicle b came too close to vehicle e. For this reason, all six instances were categorized risky and have been placed in the risky subcategory of Interaction J in Fig. 4.

Similarly, four out of 83 instances of Interaction K (Appendix C: Fig. C1) showed risky behavior, with vehicle d coming too close to vehicle a when changing lanes from right to left. This was also the most common interaction found during the peak hours. In Interaction N (Appendix C: Fig. C3), vehicles b and c changed lanes almost simultaneously, creating a risky situation. Moreover, Interaction P depicted in Fig. C3 was also considered risky in our opinion. Here, three vehicles (a, b and c) were stuck in the traffic jam on lane 2. Vehicle a saw free space in lane 3 and moved from lane 2 to lane 3, after which vehicle b changed lanes. The same pattern was followed by vehicle c. While this behavior is common in traffic jams, it can be annoying and risky for other vehicles entering other lanes. In summary, there were a total of four interactions (J, K, N, and P) that showed risky driving behavior during the selected peak hours.

The Off-Peak Traffic Hour

The nine interactions (Interactions A–I) found during the off-peak traffic hour are shown together with their instances in Appendix B: Figs. B1–B4. For identifying risky variations in these interactions, we followed the same strategy of adding safety distance points in the PDP representation as described for the peak hours in the previous Section. In the case of the off-peak traffic hour, 14 of the 26 instances of Interaction D showed risky behavior when the distance between vehicles b and c was alarmingly small. Consequently, their safety distance points are too close and the precedences changed accordingly. Similarly, in Interaction E, all three instances showed risky behavior with vehicle b rapidly approaching vehicle d without maintaining a safe distance. Interaction G was quite dangerous with vehicle b driving from the merging lane to lane 1 with the risk of collision with the front vehicle in lane 3 and the vehicles coming from lane 2 and lane 1. Although this type of interaction is less frequent, it still occurs in reality. In the off-peak traffic hour that we selected, this interaction happened only once, but if we had taken the full day video to analyze, this interaction would most likely have been found multiple times. Another risky interaction found was Interaction H. Here, vehicles a and c change lanes almost simultaneously, even though vehicle b is in between them. Also, vehicle a is changing lanes and overtaking vehicle b from the right, which is not allowed when there is less traffic. We found five instances of this interaction in the off-peak traffic hours. Similarly, in an instance of Interaction I, vehicle c changed lanes from right to left without maintaining a proper safety distance from the vehicle in front and behind. All these risky interactions were not obvious to the naked eye alone, nevertheless, the PDP representation made it possible to identify the variations in these interactions. We believe that traffic experts can see many more risky variations in driving maneuvers by exploiting this representation in their daily traffic assessment. In summary, five out of nine interactions were categorized as risky during the selected off-peak traffic hour.

DISCUSSION

Nowadays, it is common to have access to large volumes of traffic data, either from traffic video surveillance, GPS, or other innovative technologies. As these datasets continue to grow and autonomous vehicles become inevitable, a major challenge is to effectively represent and analyze large datasets to extract useful information, such as microscopic traffic patterns or abnormal driving behavior. Such information can be used not only as a guide for the upcoming autonomous vehicles, but also by traffic experts for checking risky driving behavior. On the one hand, there exist many quantitative techniques to perform this task, but their outcome is not always promising when handling missing information present in the motion data. On the other hand, various qualitative approaches to data mining have been proposed, but these approaches fail to capture several interesting patterns at once. Another issue is that 'live' traffic management requires fast data processing.

In this research, the main objective was to make it possible to automatically detect various – perhaps unknown – vehicle interactions based on a qualitative approach: the PDP representation. This representation was chosen to qualitatively represent large traffic data volumes to easily and quickly identify micro-scale traffic variations. Besides, we wanted to know if we could make recommendations on the safety of these variations based on the PDP representation.

To meet these objectives, we needed a significantly big real-world traffic surveillance dataset, which was hard to access due to the EU privacy (GDPR) policies. In fact, there are several high-quality open datasets extracted from drone videos, but many of them are intended for macro traffic analysis and do not include the xyt-coordinates of individual vehicles. The beauty of the PDP representation is that it only requires these xyt coordinates of the individual moving objects and can then detect subtle changes in their movements. Fortunately, the local traffic agency agreed to provide such coordinates data for research purposes. The traffic agency has over 1850 surveillance cameras at various locations. It was difficult to select a good quality camera at a location with complex traffic situations, as we were aiming for a location where there is a big difference between the driving maneuvers at the peak hours and at other times of the day. Moreover, the surveillance video should be of sufficient quality to clearly observe the vehicles such that these can be mapped to xyt-coordinates without major detection errors. However, not all of these requirements could be met at the same time. The locations having traffic jams did not have high quality surveillance video, whereas the locations with high-quality video did not have structural traffic jams throughout the day. Finally, we selected a highway location, where there was traffic congestion in the morning. We divided the surveillance video of this location into peak and off-peak traffic hours. By peak hours, we meant the specific moments of the day when there were clear signs of traffic congestion. By analyzing the peak and off-peak traffic hours separately, we expected to observe a variety of traffic interactions. The example of a multi-lane highway was selected to reduce the number of complex interactions for the analysis and to study only one driving maneuver, namely the lane-changing.

The xyt-coordinates generated by detecting the vehicles in the video also contained registration inaccuracies due to the low quality of the camera video. Note that we did not detect motorbikes as they were not included in the object-detection algorithm used. Besides, there were several detection issues since the camera was not able to notice traffic properly from above during the congestion period, possibly due to shadow effects. Consequently, it was hard to obtain accurate PDP representations of the vehicles, especially for finding the '=' relation between them. To overcome this issue, we used a buffer around the vehicles so that the vehicles can be treated as approximately equal ($\approx$) whenever required for deriving the PDP representations.

In our method, we used the PDP approach to obtain an overall view of traffic during the selected peak and off-peak hours. We identified different interactions separately for the peak and off-peak hours and also counted the number of times they occurred (instances). This information allowed us to see which interactions occur regularly during the selected hours and what general conclusions can be drawn about the driving behavior underlying these interactions.

In the Results Section, we tried to comment on the safety of the identified interactions during the peak and off-peak traffic hours. By adding two safety distance points around the vehicles in the PDP representation, we proposed a way to identify subtleties that can lead to risky interactions. Of course, this addition of points also requires a good estimation of the safety distance, and since a lane change maneuver has a certain duration, the start and end time of this maneuver must also be clearly defined. One could also take a smaller and a larger safety distance, resulting in the former case in less dangerous situations and in latter case in more dangerous situations. This is a nice example of how the PDP representation can represent a vehicle with multiple points instead of just one, i.e., centroid. The risky and non-risky categorization also depends on the speed of the cars doing the maneuver. The position of the safety distance point can be linked to the driving speed of the vehicles so that the local safety rules are

taken into account in the analysis. Identifying such subtle variations in the driving patterns could be useful for traffic experts, traffic monitoring agencies, policymakers, and people responsible for developing infrastructure for urban mobility since traffic safety is a part of urban mobility. Besides, the automotive industry could use this approach to train autonomous vehicles in risky driving situations in practice. In addition, using traffic surveillance data to gain different insights into driving behavior could also help to make the public aware that surveillance cameras are not only there to monitor the speed limit, but that they can also take note of dangerous and aggressive driving behavior.

As cameras have now become an important part of road safety, it is advisable to install high-quality surveillance cameras in locations with high traffic volumes during the day. In this way, these locations can provide a better insight into traffic events and driving behavior to improve road safety guidelines. With our approach, it is certainly possible for traffic experts to check the difference between traffic interactions in terms of vehicle speed limits. For example, going from 100 km/h to 90 km/h or from 90 km/h to 80 km/h, it is possible to use the PDP representation and see what happens to the number of interactions and the factors behind these interactions. Analyzing motorbikes' driving patterns with our approach could also reveal some interesting patterns that could be useful for traffic analysts. The same applies to trucks and larger vehicles, as they usually have different safety parameters (safety distances when braking, changing lanes, etc.).

Another limitation of this work is the definition of risky lane change in the Results Section. We have used the distance between two vehicles to determine safety for our ease. This seems to be the simplest approach, but distance is not the only factor we need to classify risky and non-risky situations. In the literature, TTC, PSD or other surrogate safety measures have been well defined to quantify safety. However, the added value of our approach could be seen by combining PDP and TTC/PET. For example, if the PDP analysis shows that a particular lane change maneuver is risky, the TTC calculation could be adjusted to take into account a shorter time window for collision avoidance. Similarly, if PDP identifies an risky lane change, PET can be used to calculate the time it takes for the vehicles to come to a safe stop after a collision. While TTC and PET are established measures of safety, they have their limitations. By using PDP as an additional measure, a more complete picture of safety can be obtained. For example, if TTC and PET show that a particular maneuver is safe, but PDP identifies a potential risk, further investigation can be carried out to determine the best course of action. Therefore, in the future work, the proposed method should be studied together with the existing methods and the advantages of the proposed method over the existing methods should be discussed in detail.

The PDP approach can be predictive to some extent, but its ability to predict future behavior and risk on a time scale of seconds or minutes is limited. This is because the PDP approach is based on analyzing past accident data and identifying patterns and risk factors associated with those accidents. While these patterns and risk factors can provide predictions of future accident risk, they cannot always take into account the unpredictability of human behavior and other factors that may contribute to accidents.

In addition, as currently applied, the PDP approach is typically used to identify high-risk locations or scenarios rather than to predict specific accident events over short periods. For example, the PDP approach can identify a high-risk, high-volume, low-visibility intersection, but cannot predict whether an accident will occur at that intersection in the next few minutes (Qayyum et al., 2022). It would still be interesting to explore whether the PDP approach can be merged with the other approaches and

technologies that can be used to predict accident risk in shorter time frames. Some interesting studies to compare with our approach are Munira, Sener, and Dai (2020); Yang, Tian, Wang, and Yuan (2022); Yang, Yin, Wang, Meng, and Yuan (2023).

Important to mention here is that most methods used in the literature have employed Support Vector Machines (SVM) and Artificial Intelligence (AI), and each method has its strengths and weaknesses. Since the PDP approach is not an AI or machine learning method, it is difficult to comment on its recognition accuracy. For the large dataset used in our study, the recognition accuracy of the PDP representation was 87% with a recognition time of 21 to 27 ms. The SVM approach, on the other hand, achieved high accuracy rates of over 90%, but required more than 30 ms to detect the lane change maneuvers. It should be noted that the accuracy of our approach is highly dependent on the quality of the dataset, which needs further investigation. In the present case, the dataset had occlusion problems, so the object detection needs to be significantly improved. Nevertheless, the PDP representation outperformed the SVM method in terms of speed and scalability, as our method can be scaled to a network of highways and intersections. Other methods such as template matching and machine learning algorithms showed promising results in the literature but had limitations in terms of robustness and scalability. In summary, our analysis shows that PDP representation is an effective method to detect lane change maneuvers for large, computationally intensive datasets. However, further research is needed to develop hybrid methods that combine the strengths of multiple methods to achieve better overall performance.

Lastly, by analyzing the risky vehicle interactions identified by the proposed approach and comparing them with actual/near accidents, it may be possible to determine the correlation between the two. For example, if the proposed PDP approach detects a high frequency of risky lane change maneuvers in a particular traffic situation, it would be important to assess whether this situation is associated with a higher risk of crashes. This should be the topic of another article, as the current article aims to show that the proposed PDP approach is a step towards identifying potentially risky vehicle interactions. However, it is important to note that its effectiveness in predicting crash risk cannot be fully evaluated without data from actual crashes and near misses. In order to draw meaningful conclusions about accident risk, the data must be carefully designed and analyzed. This entire process requires a considerable amount of time and effort and is therefore the subject of a separate article.

In summary, we have identified micro-scale traffic interactions in a sufficiently big traffic data set modeled with the PDP representation. This study is focused on analyzing traffic interactions during different times of day and identifying potentially risky and non-risky interactions. We started with a real traffic data stream from a multi-lane highway and selected the best data we could obtain and analyze during the peak and off-peak traffic hours. A future prospect is to analyze this data stream from weekends or holidays to see how it differs to our current analysis of a weekday data. Besides, the significance of this study is to provide insights into how traffic patterns and interactions can be improved to increase safety on the road. The practical conclusion is that the PDP representation can be used to check whether certain adjustments (e.g., changing the maximum speed) have a significant impact on the number of dangerous behaviors, which is important for improving road safety. It has been shown in Section 4 that a simple change in speed can result in the change of safety distance points. This change can then be translated and seen easily in the PDP representation. Such minute changes are not directly visible to check if the pattern detected is going to be risky in the future. But with the PDP representation, one can comment that based on these

speed adjustments, a certain detected pattern could turn into a risky situation.

Since all detected traffic interactions involved lane changing, we called them variants of the lane-changing maneuver. Our approach could be used in GIS applications for spatial pattern analysis of traffic. It could also serve as a useful tool for policy makers and transport experts to extract useful information from traffic data through data mining. It is also important to investigate hybrid methods that combine machine learning algorithms with feature extraction methods along with the PDP representation to improve accuracy and robustness in specific scenarios. Furthermore, since computation time is an issue for future applications, it would be interesting to see if it would be possible to perform live analysis of traffic safety using our approach. If so, what would be the requirements or limitations. Such research questions are very interesting for the future and we believe that big data processing techniques in combination with the PDP representation could be useful here.

DISCLOSURE STATEMENT

The authors declare that they are not aware of any competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DECLARATION ON THE AVAILABILITY OF DATA

The datasets and the code supporting the findings of this study are available at `https://doi.org/10.6084/m9.figshare.17033315`.

References

Abdel-Aty, M., & Pande, A. (2005). Identifying crash propensity using specific traffic speed conditions. *Journal of Safety Research*, *36*(1), 97–108.

Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., & Giannotti, F. (2009). Interactive visual clustering of large collections of trajectories. In *2009 ieee symposium on visual analytics science and technology* (pp. 3–10).

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv e-prints*, 1–17.

15

Dedes, G., Grejner-Brzezinska, D., Guenther, D., Heydinger, G., Mouskos, K., Park, B., & Toth, C. (2011). Integrated GNSS/INU, vehicle dynamics, and microscopic traffic flow simulator for automotive safety. In *14th international ieee conference on intelligent transportation systems (itsc)* (pp. 840–845).

Dijkstra, A., Marchesini, P., Bijleveld, F., Kars, V., Drolenga, H., & van Maarseveen, M. (2010). Do calculated conflicts in microsimulation model predict number of crashes? *Transportation Research Record*, *2147*(1), 105–112. Retrieved from `https://doi.org/10.3141/2147-13`

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In *Theories and methods of spatio-temporal reasoning in geographic space* (pp. 162–178). Springer.

Fritzsche, H.-T. (1994). A model for traffic simulation. *Traffic Engineering & Control*, *35(5)*, 317–321.

Glez-Cabrera, F. J., Álvarez-Bravo, J. V., & Díaz, F. (2013). QRPC: A new qualitative model for representing motion patterns. *Expert Systems with Applications*, *40(11)*(11), 4547–4561.

Guo, H., Wang, Z., Yu, B., Zhao, H., & Yuan, X. (2011). Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Ieee pacific visualization symposium* (pp. 163–170).

Horni, A., Nagel, K., & Axhausen, K. (Eds.). (2016). *Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press.

Huguenin, F., Torday, A., & Dumont, A. (2005). Evaluation of traffic safety using microsimulation. In *Proceedings of the 5th swiss transport research conference–strc, ascona, swiss* (pp. 1–19).

Khelfa, B., Ba, I., & Tordeux, A. (2023). Predicting highway lane-changing maneuvers: A benchmark analysis of machine and ensemble learning algorithms. *Physica A: Statistical Mechanics and its Applications*, 128471.

Klitzke, L., Koch, C., & Köster, F. (2020). Identification of lane-change maneuvers in real-world drivings with hidden markov model and dynamic time warping. In *2020 ieee 23rd international conference on intelligent transportation systems (itsc)* (pp. 1–7).

Lateef, M. U. (2011). Spatial patterns monitoring of road traffic injuries in Karachi metropolis. *International Journal of Injury Control and Safety Promotion*, *18(2)*(2), 97–105.

Liu, S., Liu, Y., Ni, L., Li, M., & Fan, J. (2012). Detecting crowdedness spot in city transportation. *IEEE Transactions on Vehicular Technology*, *62(4)*(4), 1527–1539.

Liu, S., Liu, Y., Ni, L. M., Fan, J., & Li, M. (2010). Towards mobility-based clustering. In *Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining* (pp. 919–928).

Long, J. A., & Nelson, T. A. (2013). Measuring dynamic interaction in movement data. *Transactions in GIS*, *17(1)*(1), 62–77.

Madhukar, B. (2020). *Complete guide to different persisting methods in Pandas.* Retrieved from `https://analyticsindiamag.com/complete-guide-to-different-persisting-methods-in-pandas/` (Accessed: 2021-12-21)

Mahmud, S. M. S., Ferreira, L., Hoque, M. S., & Tavassoli, A. (2019). Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment. *International Association of Traffic and Safety Sciences*, *43*(1), 27–36. Retrieved from `https://doi.org/10.1016/j.iatssr.2018.07.002`

Mak, K. K., & Sicking, D. L. (2003). *Roadside Safety Analysis Program (RSAP).* Transportation Research Board.

Mandal, K., Sen, A., Chakraborty, A., Roy, S., Batabyal, S., & Bandyopadhyay, S. (2011). Road traffic congestion monitoring and measurement using active RFID and GSM technology. In *14th international ieee conference on intelligent transportation systems (itsc)* (pp. 1375–1379).

Mathew, J., & Benekohal, R. F. (2021). Highway-rail grade crossings accident prediction using Zero Inflated Negative Binomial and Empirical Bayes method. *Journal of Safety Research*, *79*, 211–236. Retrieved from `https://doi.org/10.1016/j.jsr.2021.09.003`

Mills, L., Freeman, J., Truelove, V., Davey, J., & Delhomme, P. (2021). Comparative judge-

ments of crash risk and driving ability for speeding behaviours. *Journal of Safety Research*, *79*, 68–75.

Munira, S., Sener, I. N., & Dai, B. (2020). A bayesian spatial poisson-lognormal model to examine pedestrian crash severity at signalized intersections. *Accident Analysis & Prevention*, *144*, 105679.

Parsons, S., & Parsons, G. (2001). *Qualitative Methods For Reasoning Under Uncertainty* (Vol. 13). Mit Press.

Poco, J., Doraiswamy, H., Vo, H. T., Comba, J. L. D., Freire, J., & Silva, C. T. (2015). Exploring traffic dynamics in urban environments using vector-valued functions. In *Computer graphics forum* (Vol. 34(3), pp. 161–170).

Pu, J., Liu, S., Ding, Y., Qu, H., & Ni, L. (2013). T-Watcher: A new visual analytic system for effective traffic surveillance. In *14th ieee international conference on mobile data management* (Vol. 1, pp. 127–136).

Qayyum, A., De Baets, B., Baig, M. S., Witlox, F., De Tré, G., & Van de Weghe, N. (2021). The Point-Descriptor-Precedence representation for point configurations and movements. *International Journal of Geographical Information Science*, *35(7)*(7), 1374-1391. Retrieved from `https://doi.org/10.1080/13658816.2020.1864378`

Qayyum, A., De Baets, B., De Cock, L., Witlox, F., De Tré, G., & Van de Weghe, N. (2022). Application of the Point-Descriptor-Precedence representation for micro-scale traffic analysis at a non-signalized T-junction. *Geo-Spatial Information Science*, 1–25. Retrieved from `https://doi.org/10.1080/10095020.2022.2069520`

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, *3(9)*(9), 369–387. Retrieved from `https://doi.org/10.46827/ejes.v0i0.1017`

Sarkar, N. C., Bhaskar, A., Zheng, Z., & Miska, M. P. (2020). Microscopic modelling of area-based heterogeneous traffic flow: Area selection and vehicle movement. *Transportation Research Part C: Emerging Technologies*, *111*, 373-396. Retrieved from `https://doi.org/10.1016/j.trc.2019.12.013`

Silgu, M. A., Erdagi, I. G., Göksu, G., & Celikoglu, H. B. (2021). Combined control of freeway traffic involving cooperative adaptive cruise controlled and human driven vehicles using feedback control through SUMO. *IEEE Transactions on Intelligent Transportation Systems*, 1-15. Retrieved from `https://doi.org/10.1109/TITS.2021.3098640`

Solanki, K. (2021). *Quantitative techniques - application, importance, limitations.* Retrieved from `https://www.toppers4u.com/2021/01/quantitative-techniques-application.html` (Accessed: 2021-12-21)

Stone, J. R., Chae, K., & Pillalamarri, S. (2002). *The effects of roundabouts on pedestrian safety* (Tech. Rep.). Southeastern Transportation Center, University of Tennessee – Knoxville.

Van de Weghe, N., Kuijpers, B., Bogaert, P., & De Maeyer, P. (2005). A qualitative trajectory calculus and the composition of its relations. In M. A. Rodríguez, I. Cruz, S. Levashkin, & M. J. Egenhofer (Eds.), *Geospatial semantics* (pp. 60–76). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wu, P., Song, L., & Meng, X. (2021). Influence of built environment and roadway characteristics on the frequency of vehicle crashes caused by driver inattention: a comparison between rural roads and urban roads. *Journal of Safety Research*, *79*, 199–210.

Yang, Y., Tian, N., Wang, Y., & Yuan, Z. (2022). A parallel fp-growth mining algorithm with load balancing constraints for traffic crash data. *International Journal of Computers Communications & Control*, *17*(4).

Yang, Y., Yin, Y., Wang, Y., Meng, R., & Yuan, Z. (2023). Modeling of freeway real-time traffic crash risk based on dynamic traffic flow considering temporal effect difference. *Journal of Transportation Engineering, Part A: Systems*, *149*(7), 04023063.

Zhang, W., & Wang, W. (2019). Learning V2V interactive driving patterns at signalized intersections. *Transportation Research Part C: Emerging Technologies*, *108*, 151-166. Retrieved from `https://doi.org/10.1016/j.trc.2019.09.009`

17

Appendix A.

Procedure 1

1 Input a target dataset T of n point objects $O^T = \{o_1^T, o_2^T, \ldots, o_n^T\}$ whose location is tracked for u timestamps $t_1, t_2, \ldots, t_u$

2 Input a reference dataset R of r point objects $O^R = \{o_1^R, o_2^R, \ldots, o_r^R\}$ whose location is tracked for v timestamps $t_1', t_2', \ldots, t_v'$

3 Initialize L: A list of r-tuples from O^T matching the reference dataset R

4 $p_1 \leftarrow$ (d_1-precedence of O^T at t_1 of T, d_2-precedence of O^T at t_1 of T)

5 $T^{new} \leftarrow \{(t_1, p_1)\}$;   // initialize a new dataset

6 for every two successive timestamps t_i, t_{i+1} of T do

7 $p_i \leftarrow$ (d_1-precedence of O^T at t_i, d_2-precedence of O^T at t_i)

8 $p_{i+1} \leftarrow$ (d_1-precedence of O^T at t_{i+1}, d_2-precedence of O^T at t_{i+1})

9   if $p_i \neq p_{i+1}$ // if the precedences are not matched

10   add (t_{i+1}, p_{i+1}) to T^{new}

11 $P \leftarrow \emptyset$

12 for each (t_i, p_i) in T^{new} do

13 $P \leftarrow P \cup \{r$-tuples of O^T at t_i according to $p_i\}$

14 $L \leftarrow ()$

15 $w \leftarrow T^{new}$.size()   // timestamps in the target dataset

16 $i \leftarrow 1$

17 for each tuple k in P do

18 $M \leftarrow ()$

19 $j \leftarrow 1$

20   while $i \leq w$ do

21 $q_i =$ (d_1-precedence of k at t_i, d_2-precedence of k at t_i)

22   while $j \leq v$ do   // timestamps in the reference dataset

23 $q_j =$ (d_1-precedence of O^R at t_j', d_2-precedence of O^R at t_j')

24   if $q_i = q_j$

25   add t_j' to M

26 $j = j + 1$

27 $i = i + 1$

28   if M.sort()$=$ True

29   add k to L

30 return L

Appendix B. The off-peak traffic hour (7–7.30 am and 9.30–10 am)

18

Interaction A

INTERACTION-A MATCHES: 7

1. (5070,5049,5071)
2. (6024,5971,5970)
3. (6626,6572,6581)
4. (7623,7717,7616)
5. (8121,8132,8112)
6. (8758,8747,8726)
7. (11424,11449,11447)

t_1
$\begin{cases} d_1: & b < a < c \\ d_2: & a \approx c < b \end{cases}$

t_2
$\begin{cases} d_1: & a \approx b < c \\ d_2: & a \approx c < b \end{cases}$

t_3
$\begin{cases} d_1: & a < b < c \\ d_2: & a \approx c < b \end{cases}$

t_4
$\begin{cases} d_1: & a < b < c \\ d_2: & a \approx b \approx c \end{cases}$

Interaction B

INTERACTION-B MATCHES: 11

1. (870,1051,1101)
2. (15146,15333,15343)
3. (15751,15807,15715)
4. (16084,16070,16029)
5. (16167,16172,16160)
6. (17062,17068,16940)
7. (18284,18277,18232)
8. (18318,18162,18284)
9. (18608,18613,18611)
10. (22223,22215, 22219)
11. (22500,22498,22400)

t_1
$\begin{cases} d_1: & c < a < b \\ d_2: & a < b \approx c \end{cases}$

t_2
$\begin{cases} d_1: & c < a < b \\ d_2: & a < b < c \end{cases}$

t_3
$\begin{cases} d_1: & a \approx c < b \\ d_2: & a < b < c \end{cases}$

t_4
$\begin{cases} d_1: & a < c < b \\ d_2: & a \approx b < c \end{cases}$

Figure B1. Two non-risky interactions of the off-peak traffic hour along with their matches (instances) found in the target dataset.

Interaction C

INTERACTION-C MATCHES: 4

1. (5181,5166,5210)
2. (11154, 11162,11150)
3. (11942,11852,11963)
4. (15040,15054,15066)

t_1
$\begin{cases} d_1: & c < a \approx b \\ d_2: & c < a < b \end{cases}$

t_2
$\begin{cases} d_1: & c < a < b \\ d_2: & c < a < b \end{cases}$

t_3
$\begin{cases} d_1: & c < a < b \\ d_2: & c < a < b \end{cases}$

t_4
$\begin{cases} d_1: & c < a < b \\ d_2: & c < a < b \end{cases}$

t_5
$\begin{cases} d_1: & c < a < b \\ d_2: & a \approx c < b \end{cases}$

Interaction D

INTERACTION-D MATCHES: 26	
1. (68,411,393)	14.(12736,12700,12716)
2. (224,385,364)	15.(12842,12908,12756)
3. (666,643,409)	16.(14752,14726,14732)
4. (905,910,888)	17.(14853,14798,14787)
5. (3250,3177,3133)	18.(14636,14810,14752)
6. (4568,4648,4659)	19.(15415,15449,15193)
7. (6323,6366,6125)	20.(15416,15193,15415)
8. (7036,7246,7261)	21.(15532,15418,15491)
9. (7570,7351,7584)	22.(17123,16969,16991)
10.(7810,7797,7761)	23.(18232,18218,18067)
11.(10740,10724,10695)	24.(19751,19713,19756)

t_1
$\begin{cases} d_1: & a < b < c \\ d_2: & a \approx b \approx c \end{cases}$

t_2
$\begin{cases} d_1: & a < b < c \\ d_2: & a \approx c < b \end{cases}$

t_3
$\begin{cases} d_1: & a < b \approx c \\ d_2: & a \approx c < b \end{cases}$

t_4
$\begin{cases} d_1: & a < c \approx b \\ d_2: & a \approx c < b \end{cases}$

t_4
$\begin{cases} d_1: & a < c < b \\ d_2: & a \approx c < b \end{cases}$

Figure B2. Interactions C and D found in the off-peak traffic hour. Interaction D is the topmost occurring interaction in the off-peak traffic hour.

19

Interaction E

t_1	t_2	t_3	

INTERACTION-E MATCHES: 3
1. (905,910,902,916)
2. (6382,6303,6290,6125)
3. (12623,12445,12551,12606)

$\begin{cases} d_1: & a < b < d < c \\ d_2: & a \approx b \approx c < d \end{cases}$ $\begin{cases} d_1: & a < b < d < c \\ d_2: & a \approx c < b < d \end{cases}$ $\begin{cases} d_1: & a < b < d < c \\ d_2: & a \approx c < b \approx d \end{cases}$

Interaction F

t_1	t_2	t_3	t_4

INTERACTION-F MATCHES: 17

1. (585,532,479)
2. (636,585,601)
3. (863,797,794)
4. (902,867,818)
5. (4460,4468,4458)
6. (3541,3589,3507)
7. (4645,4579,4637)
8. (4782,4777,4743)
9. (5520,5501,5480)

10. (6418,6472,6450)
11. (7252,7213,7240)
12. (7796,7817,7821)
13. (7933,7898,7946)
14. (8532,8518,8447)
15. (9388,9367,9374)
16. (12061,12043,12014)
17. (12525,12524,12290)

$\begin{cases} d_1: & b < a < c \\ d_2: & a < b \approx c \end{cases}$ $\begin{cases} d_1: & a \approx b < c \\ d_2: & a < b \approx c \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & a < b < c \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & a \approx b < c \end{cases}$

Figure B3. Another two interactions were found in the off-peak traffic hour along with their total occurrences.

Interaction G

t_1	t_2	t_3	t_4	t_5

INTERACTION-G MATCHES: 1
1. (2840,2821,2818,2794,2830)

$\begin{cases} d_1: & a < c < e < b < d \\ d_2: & a \approx b < c \approx d < e \end{cases}$ $\begin{cases} d_1: & a < c < e < b < d \\ d_2: & a < b \approx c \approx d < e \end{cases}$ $\begin{cases} d_1: & a < c < e < b < d \\ d_2: & a < c \approx d < b < e \end{cases}$ $\begin{cases} d_1: & a < c < e < b < d \\ d_2: & a < c \approx d < b \approx e \end{cases}$ $\begin{cases} d_1: & a < c < e < b \approx d \\ d_2: & a < c \approx d < e \approx b \end{cases}$

Interaction H

t_1	t_2	t_3	t_4	

INTERACTION-H MATCHES: 5
1. (7491,7492,7489)
2. (7526,7503,7491)
3. (11002,10854,10978)
4. (10982,11002,10854)
5. (14904,14867,14853)

$\begin{cases} d_1: & a < b < c \\ d_2: & a < b \approx c \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & a < c < b \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & c < a < b \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & c < a \approx b \end{cases}$

Interaction I

t_1	t_2	t_3	t_4	t_5

INTERACTION-I MATCHES: 1
1. (10257,10310,10264)

$\begin{cases} d_1: & a < c < b \\ d_2: & a \approx b < c \end{cases}$ $\begin{cases} d_1: & a < c < b \\ d_2: & a \approx b \approx c \end{cases}$ $\begin{cases} d_1: & a < c < b \\ d_2: & c < a \approx b \end{cases}$ $\begin{cases} d_1: & a < b \approx c \\ d_2: & c < a \approx b \end{cases}$ $\begin{cases} d_1: & a < b < c \\ d_2: & c < a \approx b \end{cases}$

Figure B4. Three risky interactions found in the off-peak traffic hour. Vehicle b in Interaction G is changing lanes recklessly. In Interaction H, vehicles a and c are changing lanes simultaneously. In Interaction I, car c is changing lanes from right to left in hurry.

20
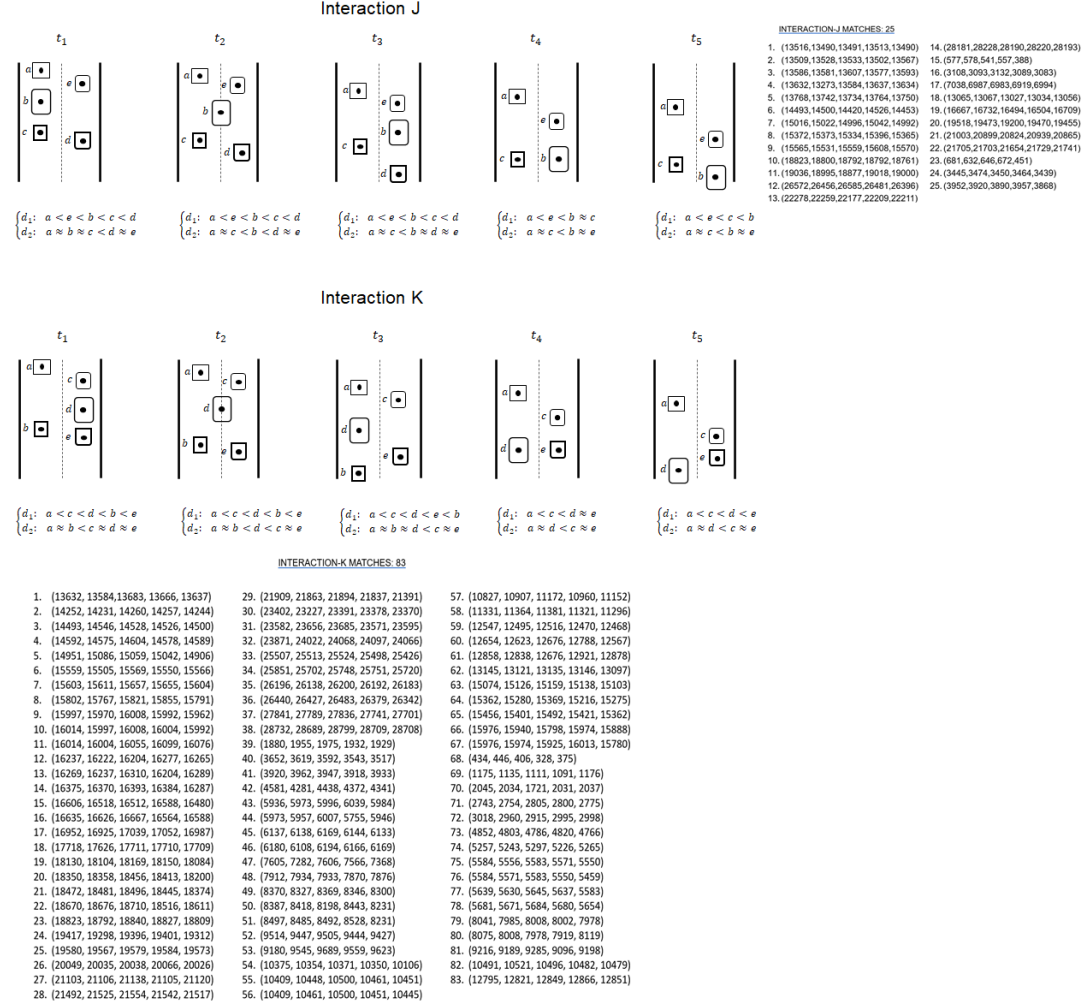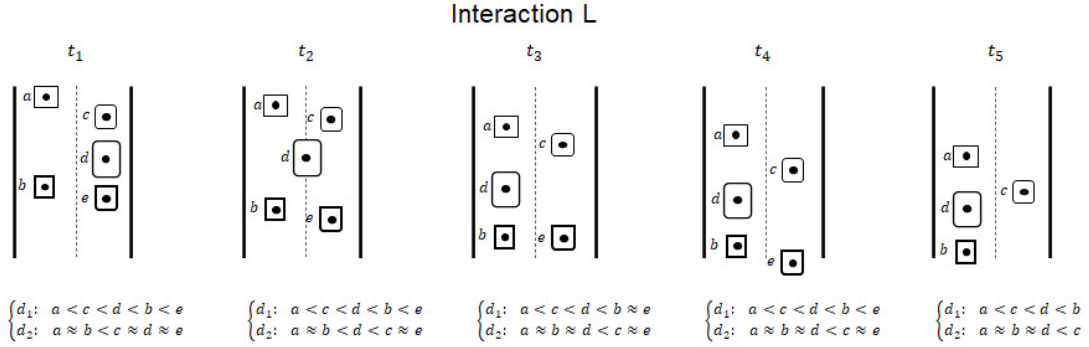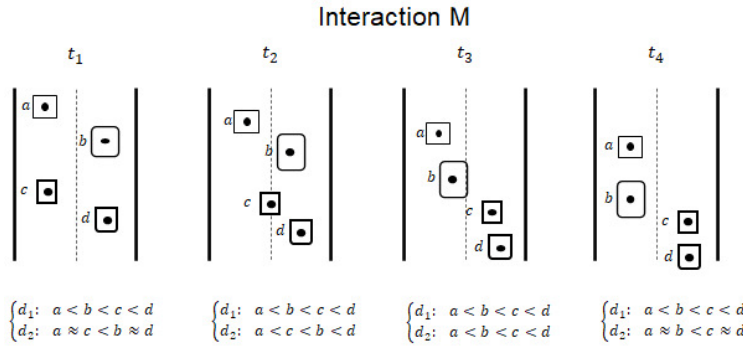
Appendix C. The peak traffic hours (7.30–9.30 am)

Figure C1. Interactions J and K were among the topmost occurring interactions in the peak traffic hours. Some of their variants were also considered risky.
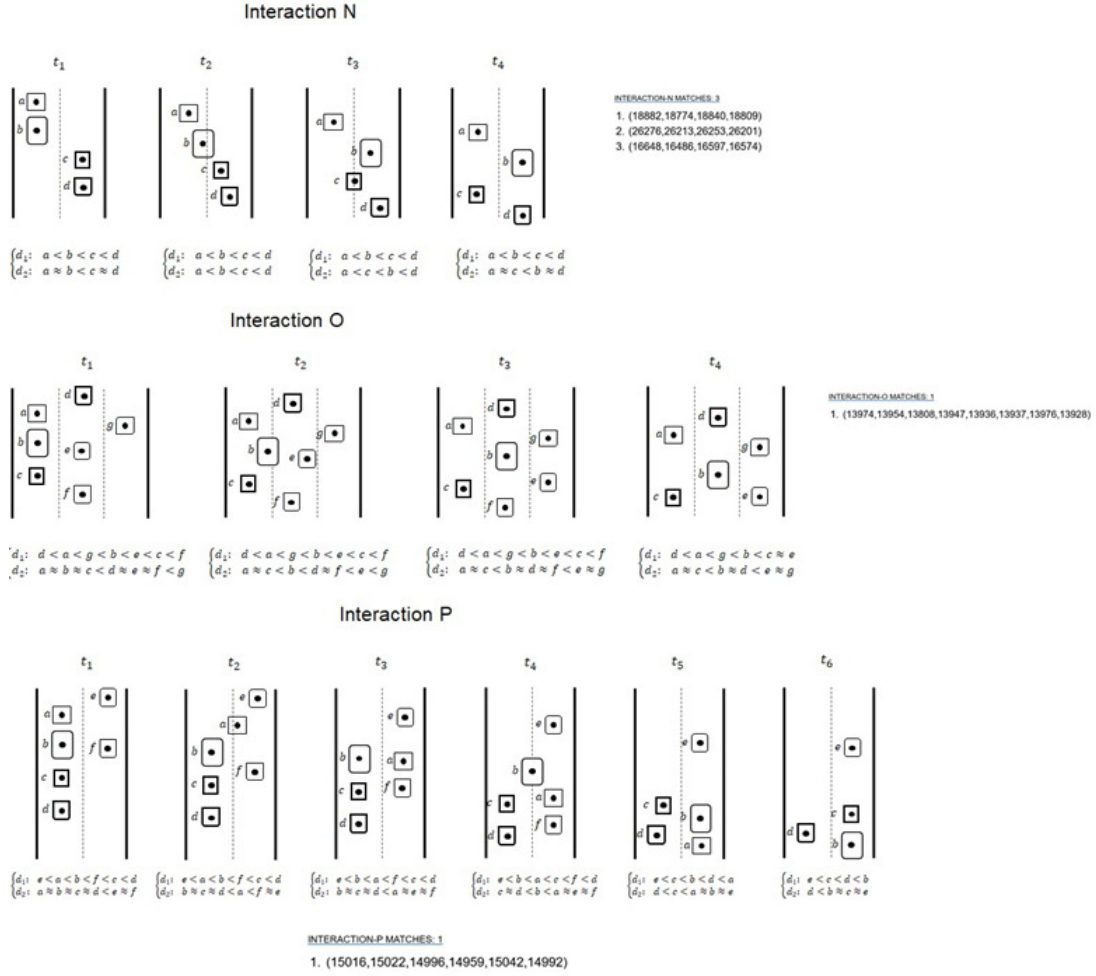
Interaction L

t_1 t_2 t_3 t_4 t_5

$\begin{cases} d_1: & a < c < d < b < e \\ d_2: & a \approx b < c \approx d \approx e \end{cases}$
$\begin{cases} d_1: & a < c < d < b < e \\ d_2: & a \approx b < d < c \approx e \end{cases}$
$\begin{cases} d_1: & a < c < d < b \approx e \\ d_2: & a \approx b \approx d < c \approx e \end{cases}$
$\begin{cases} d_1: & a < c < d < b < e \\ d_2: & a \approx b \approx d < c \approx e \end{cases}$
$\begin{cases} d_1: & a < c < d < b \\ d_2: & a \approx b \approx d < c \end{cases}$

INTERACTION L MATCHES: 61

1. (13608, 13753, 13831, 13798, 13697)
2. (15949, 15920, 15947, 15922, 15946)
3. (16375, 16370, 16393, 16384, 16287)
4. (16670, 16687, 16718, 16602, 16667)
5. (17764, 17718, 17766, 17711, 17709)
6. (17980, 17891, 17898, 17913, 17875)
7. (18168, 18149, 18154, 18173, 18169)
8. (18158, 18172, 18200, 18160, 18168)
9. (18370, 18338, 18358, 18348, 18333)
10. (19070, 19021, 19073, 19033, 188787)
11. (19669, 19682, 19715, 19690, 19677)
12. (19983, 20034, 20035, 20009, 19999)
13. (20772, 20786, 20779, 20800, 20732)
14. (21843, 21804, 21794, 21901, 21863)
15. (22478, 22468, 22440, 22386, 22450)
16. (23469, 23473, 23457, 23456, 23489)
17. (23723, 23640, 23603, 23227, 23662)
18. (24341, 24366, 24358, 24297, 24310)
19. (24380, 24384, 24386, 24401, 24403)
20. (24892, 24835, 25149, 25112, 24886)
21. (25741, 25729, 25702, 25730, 25704)
22. (1106, 1098, 1107, 1102, 1082)
23. (1440, 1464, 1489, 1463, 1481)
24. (2671, 2566, 2623, 2549, 2594)
25. (6862, 6837, 6959, 6859, 6963)
26. (7147, 7140, 7184, 7166, 7155)
27. (7289, 7237, 7250, 7133, 7235)
28. (7437, 7442, 7371, 7392, 7374)
29. (8872, 8804, 8896, 8549, 8820)
30. (9495, 9514, 9553, 9505, 9427)
31. (9180, 9545, 9689, 9559, 9623)
32. (10652,10625,10719,90688,10713)
33. (12740,12634,12579,12615,12567)
34. (13620,13594,13665,13642,13545)
35. (14344,14364,14360,14294,14310)
36. (16652,16654,16689,16623,16607)
37. (19632,19624,19651,19628,but 9600)
38. (19609,19654,19705,19683 19632)
39. (21083,21064,21031,21051 20949)
40. (21903,21854,21899,21849 21866)
41. (23553,23465,23549,23476 23497)
42. (118,140,186,132,115)
43. (767,782,818,783,743)
44. (1021,1146,1191,1147,1175)
45. (1021,1147,1015,1191,1175)
46. (2530, 2527,2529,2527,2496)
47. (2644, 2640,2655,2664,2516)
48. (2915, 2815,2872,2876,2810)
49. (3141,3127,3136,3110,3064)
50. (5444 5443,5394,5403,5439)
51. (5813,5838,5853,5781,5850)
52. (5911,5901,5917,5900,5896)
53. (6392 6349,6387,6381,6398)
54. (7373 7343,7374,7344,7324)
55. (8023,7999,7997,8041,8002)
56. (8220,8227,8234,8214,8129)
57. (8283,8254,8285,8262,7948)
58. (8290,8287,8287,8282,8285)
59. (9128,9083,9110,9080,9065)
60. (11340,11225,11338,11332,11356)
61. (13791,13750,13931,13896,13853)

Interaction M

t_1 t_2 t_3 t_4

INTERACTION-M MATCHES: 5

1. (13726,13717,13653,13683)
2. (1252,1218,1161,1015)
3. (4852,4766,4803,4743)
4. (7716,7707,7724,7706)
5. (8386,8376,8356,8324)

$\begin{cases} d_1: & a < b < c < d \\ d_2: & a \approx c < b \approx d \end{cases}$
$\begin{cases} d_1: & a < b < c < d \\ d_2: & a < c < b < d \end{cases}$
$\begin{cases} d_1: & a < b < c < d \\ d_2: & a < b < c < d \end{cases}$
$\begin{cases} d_1: & a < b < c < d \\ d_2: & a \approx b < c \approx d \end{cases}$

Figure C2. Two non-risky interactions found in the peak traffic hours.

22

Figure C3. Interaction N was considered as risky in the peak traffic hours as is evident from vehicles b and c. Interaction O was considered as non-risky. Interaction P was also considered risky during the peak traffic hours. Three vehicles (a, b, and c) stuck in the left lane annoyingly changed lanes.
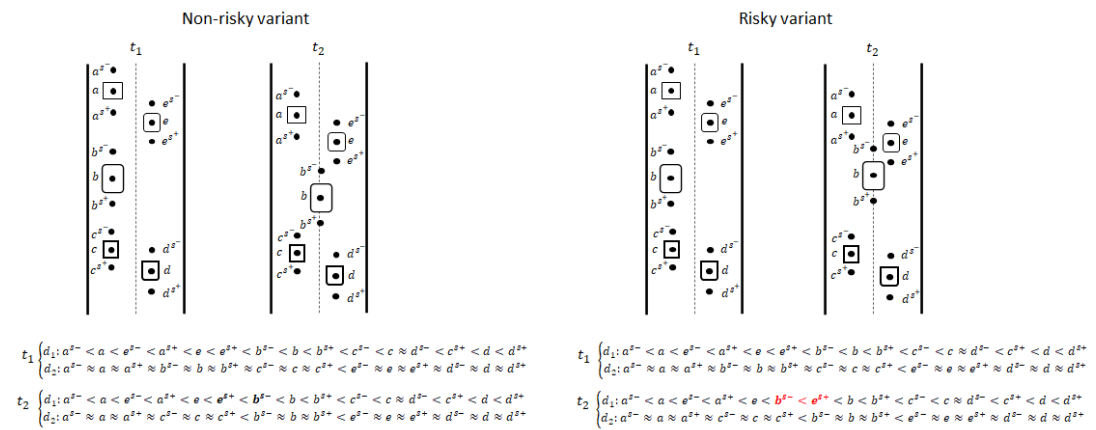
Figure C4. The non-risky and risky variants of Interaction J. Vehicle b is too close to vehicle e in the risky variant.