



Article Measuring Trajectory Similarity Based on the Spatio-Temporal Properties of Moving Objects in Road Networks

Ali Dorosti¹, Ali Asghar Alesheikh¹ and Mohammad Sharif^{2,*}

- ¹ Department of Geospatial Information Systems, K. N. Toosi University of Technology, Tehran 19967-15433, Iran; alidorosti@email.kntu.ac.ir (A.D.); alesheikh@kntu.ac.ir (A.A.A.)
- ² Institute of Mobility and Urban Planning, University of Duisburg-Essen, 45127 Essen, Germany

Correspondence: mohammad.sharif@uni-due.de

Abstract: Advancements in navigation and tracking technologies have resulted in a significant increase in movement data within road networks. Analyzing the trajectories of network-constrained moving objects makes a profound contribution to transportation and urban planning. In this context, the trajectory similarity measure enables the discovery of inherent patterns in moving object data. Existing methods for measuring trajectory similarity in network space are relatively slow and neglect the temporal characteristics of trajectories. Moreover, these methods focus on relatively small volumes of data. This study proposes a method that maps trajectories onto a network-based space to overcome these limitations. This mapping considers geographical coordinates, travel time, and the temporal order of trajectory segments in the similarity measure. Spatial similarity is measured using the Jaccard coefficient, quantifying the overlap between trajectory segments in space. Temporal similarity, on the other hand, incorporates time differences, including common trajectory segments, start time variation and trajectory duration. The method is evaluated using real-world taxi trajectory data. The processing time is one-quarter of that required by existing methods in the literature. This improvement allows for spatio-temporal analyses of a large number of trajectories, revealing the underlying behavior of moving objects in network space.

Keywords: spatio-temporal similarity; movement pattern; network space; graph; taxi trajectory

1. Introduction

Movement, an essential aspect of human civilization, has become increasingly significant as societies advance. Various forms of movement impact people's lives, necessitating comprehensive investigation and analysis. The advent of science and technology, particularly the proliferation of data-gathering instruments like GPS, has resulted in a substantial accumulation of movement-related information [1]. Within this context, trajectories provide valuable insights into the movement of point objects, such as vehicles and bicycles, over time. Each trajectory comprises a series of coordinates representing subsequent locations and timestamps, enabling critical information about object movement [2].

Raw trajectory data pose inherent challenges, including issues related to storage, computational complexity, and the presence of undesirable errors. As a result, the need arises to assess and extract meaningful information from moving object trajectories. One crucial criterion for such analysis is the notion of similarity, which refers to the existence of comparable patterns and characteristics within movement data [3]. It is important to note that similarity is not confined solely to the spatial dimension, but also extends to the temporal, semantic, and contextual dimensions [4].

Measuring the similarity of trajectories holds paramount importance across diverse domains. This analysis enables us to gain deeper insights into the movement patterns of objects or individuals, offering valuable applications. In transportation and urban planning, it contributes by optimizing routes, reducing traffic congestion, and informing infrastructure



Citation: Dorosti, A.; Alesheikh, A.A.; Sharif, M. Measuring Trajectory Similarity Based on the Spatio-Temporal Properties of Moving Objects in Road Networks. *Information* 2024, *15*, 51. https://doi.org/ 10.3390/info15010051

Academic Editors: Jianbo Li, Junjie Pang and Antonio Comi

Received: 12 December 2023 Revised: 10 January 2024 Accepted: 16 January 2024 Published: 17 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). development. Ecologists rely on trajectory similarity to track animal migrations and habitat preferences [5]. Criminal investigations benefit from tracking suspects' movements [6], while healthcare providers improve patient care by studying mobility within healthcare facilities [7]. Additionally measuring trajectory similarity supports recommendation systems, anomaly detection, predictive modeling [8], and environmental conservation. It aids decision making, pattern recognition, data visualization, model training, and emergency response, enabling more efficient and informed actions in various fields.

Examining and calculating similarity in a unidimensional manner is impractical due to the inherent interplay between the spatial and temporal dimensions of moving objects. The spatial and temporal dimensions of trajectory similarity are intricately linked, exerting mutual influence on each other. Spatial factors, such as geographic distances, significantly impact the temporal aspect by determining travel time. Longer spatial distances often correspond to lengthier temporal durations, making spatial information crucial for accurate temporal assessments. Conversely, temporal details, like the time of day or the duration spent at specific locations, can influence the spatial dimension. Temporal variations, such as rush hours, can lead to spatial congestion or divergence in trajectory patterns. The integration of both dimensions in the analysis is vital, as it enables the detection of patterns in time and space, offering a more comprehensive understanding of movement behavior. Therefore, a comprehensive approach is necessary to account for the combined influence of both dimensions on trajectory similarity [2,8,9].

Traditional similarity measure functions often calculate spatial similarity within the Euclidean space, involving x and y coordinates [10]. Euclidean space's oversimplification, assuming straight-line distances, can lead to inaccuracies when assessing similarity, especially in network space. Measuring trajectory similarity in network space presents distinct advantages over using Euclidean space due to its capacity to account for real-world constraints, complex structures, and semantic information. Network-based approaches reflect the actual connectivity between locations, making them more suitable for tasks like route planning and traffic analysis. Moreover, network-based methods offer computational efficiency and adaptability, allowing for tailored solutions in various domains, from transportation to urban planning, where understanding movement within intricate networks is essential for more precise and context-aware analyses. Consequently, the need arises for a more efficient technique that can assess the spatio-temporal similarity of trajectories with reduced computational requirements [11].

The temporal dimension of trajectory similarity is of paramount importance because it holds the key to understanding how objects or individuals move and behave over time. It provides crucial insights into patterns, trends, and irregularities within trajectories, helping in the identification of both regular behaviors and anomalies. Temporal data also plays a pivotal role in predictive modeling, event detection, and resource allocation, influencing decisions in fields like transportation, healthcare, and security. It enhances contextual understanding by revealing the timing of events, enabling efficient resource management, and supporting the detection of deviations from expected patterns. In essence, the temporal dimension enriches our ability to interpret and harness trajectory data, making it a fundamental aspect of analysis across diverse domains [10].

In this research, we introduce a spatio-temporal similarity measurement method that addresses multiple challenges in trajectory analysis including computation cost, considering temporal parameters, and efficiency improvement. Our method involves mapping trajectory datasets onto the network space as a graph. The contributions of this research include: (1) By incorporating network constraints and employing more parameters for temporal similarity (i.e., duration of common segments, occurrence time, and duration of entire trajectories), our method not only measures the spatio-temporal similarity of network-constrained trajectories but also provides a more realistic representation of movement behavior in real-world road networks. (2) Overcoming the computational challenges associated with traditional spatial similarity calculations, our method offers faster and more scalable calculations. Through empirical evaluations and case studies, we demonstrate

the effectiveness and applicability of our method in calculating spatio-temporal similarity for network-constrained of several taxi trajectories. The findings of this research can be used in data-driven applications such as traffic prediction, recommender systems, and resource allocation.

2. Review of Literature

Measuring the similarity of trajectories is normally conducted in Euclidean space [3,12] or network space [13]. When it comes to network space, a variety of distance functions have been used including the Euclidean and network distance functions [14,15]. Given the differences in the definition of distance in those spaces, similarity measurement should be network-centric and cannot depend on metrics distance.

Calculating similarity based on network parameters (nodes and edges) has led to the use of set theory-based methods like Jaccard [16] and the longest common subsequence (LCS) [17]. These methods primarily focus on assessing the spatial dimension of similarity by identifying common segments between trajectories. However, the existing methods based on set theory still have room for improvement, particularly in addressing the challenge of accurately measuring similarity in scenarios involving intricate network structures. The complexity of real-world networks, such as those with non-linear patterns or uncertain information, poses a significant limitation to the effectiveness of current set theory-based measurement approaches [18].

Addressing the spatio-temporal dimension of similarity within the network space is a critical consideration. While some previous research efforts have attempted to quantify spatio-temporal similarity, they have often focused on limited criteria, such as differences in the timing of trajectory beginnings or ends, without fully exploring the wealth of available temporal parameters. Additionally, the lack of a comprehensive examination of diverse temporal aspects may hinder the ability of existing methods to capture the nuanced patterns present in real-world spatio-temporal networks [16,19]. This highlights the need for a more comprehensive approach to temporal similarity measurement considering more temporal metrics in the context of network-constrained trajectories.

Previous research has illuminated the challenges associated with calculating trajectory similarity in network space, emphasizing the need for data preprocessing such as noise removal, mapping positional data onto the network, and dealing with outlier data. These pre-processing steps are necessary to align the raw trajectory data with network criteria, so that, if noise remains in the data, in the methods that use distance metrics to measure similarity, the validity of the similarity values is questioned. On the other hand, the presence of noise causes the trajectory data to be incorrectly mapped on the network and the accuracy of the results obtained for similarity to be questioned, especially in methods based on a set theory, which depend on the results of mapping lines on the network space [20]. Regarding the outlier data, it should be mentioned that, if these data remain in the similarity calculation process, it will cause heterogeneous changes in the similarity results and will make errors in the distance metrics used to determine the similarity [21]. However, they also contribute to the computational complexity and processing time, underscoring the demand for more efficient methods that can alleviate these challenges [17].

In conclusion, the quest to calculate trajectory similarity within network spaces necessitates the development of a robust method that excels in both spatial and temporal dimensions. Such a method should, while maintaining efficiency, include the complexities of the network space in the real world and manage the pre-processing needs in such a way that the errors resulting from the inaccuracy of pre-processing are not reflected in the similarity results. This is necessary to ensure that it can efficiently and quickly compare large amounts of trajectory data in various applications, from transportation planning to urban development and beyond. It is clear that addressing these complexities is pivotal for advancing our understanding of movement patterns within network-constrained environments. According to the cases mentioned this research, a method for measuring similarity has been presented that can answer the complexities of the network space and pre-processing and the efficiency and high speed of measuring similarity.

3. Methodology

This section proposes the network-constraint spatio-temporal similarity measure procedure. The methodology of the study consists of three main stages, shown graphically in Figure 1: (i) data pre-processing, (ii) similarity measure, and (iii) performance evaluation.



Figure 1. Methodology for spatio-temporal similarity measure of trajectories in network space.

3.1. Trajectory Pre-Processing

3.1.1. Noise Detection

Noise is typically present in the raw movement data collected by tracking devices. In this study, the raw trajectories including noises (outliers) are identified based on the average speed of moving objects and eliminated from the dataset [22,23]. The moving object travels a certain distance based on an average speed recorded by GPS receivers based on a constant sampling rate (e.g., 15 s). If the distance between two consecutive points in a trajectory exceeds three times the particular distance estimated for each trajectory, it is concluded that the coordinates of some points in this trajectory include noise, and the trajectory is considered an outlier and is removed from the dataset [24].

3.1.2. Map Matching

The process of matching a series of point coordinates to a logical representation of the real world, such as a road network, is known as map matching. This section aims to assign unique ID numbers belonging to each part of the road network to the points in every route. In this regard, a method based on the Hidden Markov model (HMM) is used for matching trajectory data to the road network [24]. HMM determines the most likely path taken by a moving object, such as a vehicle or person, based on a sequence of observed GPS data points [25,26]. The basic idea behind HMM map matching is to find the path through the HMM states (segments of the road network) that maximizes the likelihood of the observed GPS data. In this study, the result of matching for each point of a trajectory on a piece of the road network is represented by a unique id of that piece of the road network [27,28].

3.1.3. Filtering

The filtering process reduces the volume of data, leading to significant savings in processing costs in similarity measurement. In the first step, we eliminate distant and irrelevant trajectories that do not intersect with the selected target trajectory space, which is integral for measuring the similarity of other trajectories. Initially, a 150 m buffer zone around the target trajectory (the trajectory where the similarity of the rest of the trajectories compared to that trajectory is calculated) is established. The distance is chosen based on the complexity and density of road networks obtained through network analysis. This buffer zone effectively represents a reasonable range within which we expect moving objects to generate their trajectories through the given network. By excluding any trajectories that fail to intersect with this buffer zone, we ensure that irrelevant data are omitted. This strategic filtering allows us to focus more on trajectories that are more likely to reflect actual travel patterns similar to the target trajectory. This leads to diminishing the data involved in the similarity measurement and consequently reducing computational costs [29].

A trajectory may have an intersection with the 150 m range of the target trajectory without any common parts (e.g., perpendicular lines) [30]. To overcome this issue, the second filtering step focuses on the average azimuth of the first and last five segments of each trajectory, serving to exclude trajectories that exhibit a substantially divergent direction of movement compared to the target trajectory. This involves calculating and comparing the average azimuth of these segments in each trajectory against the target trajectory. Any trajectory with an azimuth difference exceeding 0.2 radians is systematically excluded from the similarity calculation process. The selection of this 0.2 radians threshold is meticulously based on data complexity and network density [31]. By applying this filter, our approach streamlines similarity calculations to focus on trajectories which are more likely to align with similar travel patterns, ultimately enhancing the precision of our similarity measurements [32].

3.2. Spatial Similarity Measure

After performing map matching to align the trajectories with the network space, each trajectory is divided into a number of segments, which are the links between consecutive intersections within the network. To measure spatial similarity between each pair of trajectories, we use the Jaccard Similarity coefficient, which calculates the similarity between two trajectories A and B by dividing the size of their intersected segments by the size of their union (Equation (1)) [33]. This measure is sensitive to both the number and distribution of segments in the trajectories and can be used to compare the similarity of different travel patterns. For example, for two sample trajectories A and B in Figure 2, the spatial similarity is equal to 2/(5 + 5 - 2) = 0.25.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$
(1)

3.3. Temporal Similarity Measure

To calculate temporal similarity, a multivariate equation is defined that considers three factors: (1) the difference in travel time between the common segments of the two trajectories, (2) the start time of each trajectory, and (3) the total duration of each trajectory. These variables are combined together as shown in Equation (2).

$$TS = 1 - \frac{\left[\sum_{i=1}^{m} |\Delta Tt_i - \Delta Tq_i|\right] * e^{(T_{S_T} - T_{S_Q})}}{\max(\Delta T_T, \Delta T_O)}$$
(2)

where m represents the number of common segments between the two trajectories, $T_{S(T)}$ and $T_{S(Q)}$ represent the start time of the first and second trajectories, respectively, calculated from time 00:00 every day. $\Delta T_{t(i)}$ and $\Delta T_{q(i)}$ represent the duration of the ith common



segment of the first and second trajectories, respectively. ΔT_T and ΔT_Q represent the total duration of the first and second trajectories, respectively.

Figure 2. Two random trajectory examples with common segments and timestamps.

In this regard, due to the fact that the result of the fraction is reduced from the number 1, the larger the form of the fraction becomes, it reduces the amount of temporal similarity. In the first term of the fraction, if the travel time difference for the same sections is greater, the value of this term increases and causes the fraction to become larger, so the temporal similarity between the two trajectories becomes less. In the second term of the fraction, if the starting times of the trajectories are further apart, the power of the number e becomes larger and the value of this term increases and the temporal similarity decreases. In the sentence used in the denominator of the fraction, the result of the fraction is reduced by considering the maximum amount of the total travel time of both trajectories, and the longer the two trajectories are, the more normal the value of the fraction becomes. The maximum time taken for the two trajectories normalizes the fraction and reduces the effect of large temporal differences of common segments in cases where the number of common segments is large. This ensures that longer trajectories are not necessarily less temporally similar because they have more segments or longer shared segments. For two sample trajectories A and B in Figure 2, the temporal similarity is equal to:

$$1 - \frac{((10 - 10) + (10 - 10)) * e^{(0 - 0)}}{\max(50, 50)} = 1.$$

3.4. Spatio-Temporal Similarity Measure

To calculate the final similarity between two trajectories, a weighted average of the calculated spatial and temporal similarities is used. The weights assigned to each similarity metric have been obtained through experimental iterations. By using a weighted average, we are able to take into account the relative importance of each similarity metric in determining the overall similarity between two trajectories [16,19]. After evaluating different weight values and taking into account that, in applications such as path prediction [5], the influence of the spatial term of similarity is slightly greater than that of the temporal term, and also, according to previous researches, the coefficients of 0.6 and 0.4 are considered for spatial and temporal similarity results. By using these coefficients, we have tried to

consider the effect of spatial similarity a little more than temporal similarity. Equation (3) shows the weighted spatio-temporal similarity measure. According to the example in Figure 2, the spatio-temporal similarity between trajectories A and B is equal to $(0.6 \times 0.25) + (0.4 \times 1) = 0.55$

Spatio-temporal similarity = $(0.6 \times \text{Spatial similarity}) + (0.4 \times \text{Temporal similarity})$ (3)

Finally, to make the final similarity score more interpretable and comparable, it is normalized to the range of 0 to 1. This normalization ensures that the final similarity score can be easily interpreted as a percentage, where a score of 1 indicates a perfect match and a score of 0 indicates no similarity between the two trajectories. The integrated results are stored as a similarity matrix, where element ij represents the similarity between trajectory number i and trajectory number j.

4. Implementation

4.1. Data

The dataset used in this study consists of over 13 million taxi trajectories in Porto City from 1 October 2013 to 30 September 2014 (Figure 3) [34]. The trajectories were collected using GPS tracking devices installed in the taxis, which recorded the latitude and longitude of the taxis at 15 s intervals. Moreover, the dataset includes the start and end time of each trajectory and other information such as the fare and the distance traveled. Due to the large size of the dataset, a sample of 15,000 trajectories are considered for the experiments from central area of Porto city; after noise reduction, 13,309 trajectories remained.



Figure 3. Study area: top left, geographic location of Porto city, Portugal; bottom, Porto City; and right, geographic location of the study area in Porto.

In addition, the shapefile of the Porto city roads provided by the OpenStreetMap (OSM) database is considered as the road network, and the trajectories are matched to this network based on the unique OSMid of each road segment. In this dataset, the ID of the outgoing route and the return route of two-way roads are considered to be different; thus the network graph created from this dataset is a directed graph.

4.2. Results

Based on the noise reduction approach, 1691 out of 15,000 trajectories (reduced by 11%) were detected as outliers. Among the remaining 13,309 trajectories, the spatio-temporal similarity calculation process was repeated 177,129,481 times. After distance filtering, the number of repetitions of the process was reduced to 48,636,646 times (reduced by 72%). After filtering based on azimuth, the number of repetitions in the similarity calculation process was reduced by 94%). Finally, by using both of the filters, the repeating similarity measure was reduced to 4,191,878 (reduced by 98%). Most of the iterations removed from the process have a value of zero for calculated similarity.

The spatial similarity values obtained for the trajectories are normalized between 0 and 1. The closer this value is to 1, the greater the similarity between the two trajectories is. The distribution of spatial similarity values is shown in Figure 4. As can be seen, the distribution of non-zero values for spatial similarity in all different filtering modes and the no-filter mode are equal to each other.



Figure 4. Non-zero values of spatial similarity distribution.

The weighted average spatio-temporal similarity results, which are between 0 and 1, are shown in Figure 5. Due to the equal non-zero similarity values in multiple methods with filtering and without filtering, the shapes of the four plots overlap and cannot be distinguished from each other.

Moreover, our methodology incorporates the use of network space and input data filtering to optimize the computation cost of similarity calculations by reducing the number of iterations in the process. The significant reduction in execution time when employing these filters underscores the efficiency of our method. As shown in Table 1, without any filtering, our method required 310 min to process completely, but with the application of a distance filter, this time was reduced to 180 min. Further, the azimuth filter achieved a remarkable reduction to just 12 min. The combined usage of both filtering modes resulted in an impressive execution time of only 2 min and 10 s. The experiments were conducted on a system equipped with an Intel Core i9 12900k processor with 64 GB of RAM, Santa Clara, CA, USA. Due to the limitations of the shared system, only one core out of the 24 available processor cores and 16 GB of the total 64 GB system RAM was utilized in the calculation process. It is evident that, with access to a system boasting enhanced specifications, even more remarkable results could be achieved.



Figure 5. Non-zero values of spatio-temporal similarity distribution.

Table 1. Comparing process time and iteration of process in each method of filtering.

| Filtering Method | Algorithm Iteration | Reduction Percentage | Process Time |
|----------------------|---------------------|-----------------------------|---------------------|
| Without filter | 177,129,481 | 0 | 310' |
| Distance | 48,636,646 | 72.5 | 180' |
| Azimuth | 11,753,469 | 93.3 | 12' |
| Distance and azimuth | 4,191,878 | 97.6 | 2'10" |

5. Discussion

In this research, we introduced a fast, non-metric, and multi-characteristic approach to calculate spatio-temporal similarity between trajectories. The Jaccard method, which is used to calculate spatial similarity, is able to calculate the commonalities between the trajectories as a set of segments from the road network graph quickly and without restrictions on the size and equality of the number of road segments. At the same time, this method is very strict and the calculated spatial similarity values are mostly between 0 and 0.2. In addition, it is limited to trajectories that have a segment in common with each other.

Similarity is a relative criterion and the more realistic the criteria involved in the similarity calculation process are, the closer the similarity values will be to reality. Our temporal similarity method (Equation (2)), which uses three key temporal criteria to calculate similarity in the time dimension, calculates similarity values which more closely match what happened in reality. The use of temporal similarity and its combination with spatial similarity has been able to distribute the final results of spatio-temporal similarity in a wider range and with a more normal distribution. This wider range for the distribution of similarity values indicates a higher resolution of the calculated similarity values.

The taxis in the city of Porto have different patterns of behavior. The filtering methods ensure that taxi trajectories that have no similarity are prevented from entering the similarity calculation process before the start of the similarity calculation process. As is clear from the results given in Table 1, this increases the speed of the similarity calculation process significantly. Considering that the plots of the results of the similarities calculated using different filtering methods overlap in Figures 4 and 5, it is clear that the filtering methods exclude only those trajectories that are not similar to the target trajectory from the similarity calculation process

To compare the performance of the developed spatio-temporal trajectory similarity measurement method with the methods available in the literature, we conducted a com-

parative analysis with the LCS method [17], MSM method [35], and the Convolutional neural network (CNN) method [36]. The results of this analysis, as shown in Figure 6, show that the similarity values obtained through the LCS method have a biased distribution between 0 and 0.2, while the similarity values obtained by our method have a more normal distribution around the value of 0.4. This demonstrates the effectiveness and reliability of our approach in capturing similarity in trajectory data. Unlike the LCS method, in the MSM method, the calculated similarity values for the trajectories have been distributed between 0.4 and 1, which indicates that this method exhibits lower stringency. The results of the CNN method also show a biased distribution around a value of 1, signifying the comparatively lower stringency of this method compared to others.



Figure 6. Implementing similarity functions in the literature.

The MSM method uses the average Euclidean distance of the corresponding points of two trajectories with a threshold for maximum distance to calculate the spatial similarity. In addition, the use of Euclidean distance to calculate the similarity means that, when the threshold value of the distance is considered large, a non-zero similarity value is calculated for all the trajectories located at that distance from the base trajectory, and practically all of the characteristics of the network space are not affected in calculating the similarity. For this reason, as can be seen in Table 2, this method has a very high computational cost and execution time.

Table 2. Iteration, processing time, mean value and standard deviation of results of the developed spatio-temporal similarity measure method comparing with LCS, MSM and CNN method (mean value and standard deviation is calculated for non-zero values as the same as Figure 6).

| Method | Processing Time | Algorithm Iteration | $\mathbf{Mean} \pm \mathbf{SD}$ |
|---------------------------------------|-----------------|---------------------|---------------------------------|
| Developed method without filtering | 310′ | 177,129,481 | 0.384 ± 0.056 |
| Developed method with filtering | 2'10" | 4,191,878 | 0.384 ± 0.056 |
| LCS method | 660′ | 177,129,481 | 0.067 ± 0.072 |
| MSM method | 1100′ | 177,129,481 | 0.705 ± 0.210 |
| CNN method | 1650' | 177,129,481 | 0.935 ± 0.085 |

To calculate the similarity between trajectories, the CNN method first divides them into smaller sub-trajectories using temporal features and depicts them in pixel space. Then, according to the main network graph, it trains a CNN model for each trajectory and finally calculates their similarity by comparing pairs of trajectories models. Considering the number of trajectories in the dataset and the fact that in this method a model is trained for each trajectory, as shown in Table 2, this method has a very high computational cost.

In terms of execution time, when using the developed method without applying filtering, the similarity calculation for the used data set takes 310 min; however, when using filtering methods, this time is reduced to 2 min and 10 s. The executions of the LCS, MSM, and CNN methods for the same data set takes 660, 1100, and 1650 min, respectively (Table 2).

The standard deviation and mean similarity values calculated using the stated methods show that the developed method has the lowest standard deviation value among other methods. This shows that the calculated similarity values have a more normal distribution around their mean and are more consistent than the results of other methods, which, while being rigorous, can be more suitable for applications such as trajectory prediction and clustering of trajectories.

The practical implications of spatio-temporal similarity in trajectory are significant, including applications in path prediction, route optimization, and data-driven decision-making. By significantly reducing computation time while maintaining accuracy, our method holds great promise in advancing the field of trajectory analysis and enhancing the efficiency of various transportation and spatial data-related tasks.

6. Conclusions

This study aims to provide an efficient and fast method for calculating the spatiotemporal similarity of trajectories in the network space. For this purpose, a spatio-temporal method has been developed based on the network characteristics and time parameters of trajectories. Spatial similarity is calculated based on the terms of the sets and the Jaccard relation. Temporal similarity is proposed based on the relationships among the time parameters affecting the trajectory. This includes the time difference in occurrence, the time difference in the total course, and the length of the time difference in common segments to present a more realistic expression of similarity. The final spatio-temporal similarity is obtained from the linear combination of these two similarity values. The measured similarity results have a wider distribution than previous methods, indicating the high resolution of the method due to the use of more parameters in the similarity measurement. Additionally, the innovative pre-processing framework applied to the data, along with the low complexity of the proposed method, has resulted in a high speed of the similarity measurement process. The execution time of the process has been reduced from 310 min to 2 min and 10 s by saving the abundance of non-zero values. Due to the high resolution of the calculated similarities, these similarity values can be used in applications such as route prediction and data clustering to measure the efficiency of the measured similarities.

Author Contributions: Conceptualization, M.S. and A.A.A.; methodology, A.D. and M.S.; software, A.D.; validation, A.D.; formal analysis, A.D. and M.S.; writing—original draft preparation, A.D. and M.S.; writing—review and editing, M.S. and A.A.A.; supervision, M.S. and A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found at https://www.kaggle.com/datasets/crailtap/taxi-trajectory/data and https://figshare. com/articles/dataset/train_csv_zip/24796053 (accessed on 13 June 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Zhang, R.; Rong, Y.; Wu, Z.; Zhuo, Y. Trajectory Similarity Assessment On Road Networks Via Embedding Learning. In Proceedings of the 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), New Delhi, India, 24–26 September 2020; pp. 1–8.
- Yao, D.; Zhang, C.; Zhu, Z.; Huang, J.; Bi, J. Trajectory clustering via deep representation learning. In Proceedings of the 2017 international joint conference on neural networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3880–3887.

- 3. Sharif, M.; Alesheikh, A.A. Context-awareness in similarity measures and pattern discoveries of trajectories: A context-based dynamic time warping method. *GIScience Remote Sens.* **2017**, *54*, 426–452. [CrossRef]
- Sharif, M.; Alesheikh, A.A.; Tashayo, B. CaFIRST: A context-aware hybrid fuzzy inference system for the similarity measure of multivariate trajectories. J. Intell. Fuzzy Syst. 2019, 36, 5383–5395. [CrossRef]
- Alizadeh, D.; Alesheikh, A.A.; Sharif, M. Prediction of vessels locations and maritime traffic using similarity measurement of trajectory. *Ann. GIS* 2020, 27, 151–162. [CrossRef]
- Cleasby, I.R.; Wakefield, E.D.; Morrissey, B.J.; Bodey, T.W.; Votier, S.C.; Bearhop, S.; Hamer, K.C. Using time-series similarity measures to compare animal movement trajectories in ecology. *Behav. Ecol. Sociobiol.* 2019, 73, 151. [CrossRef]
- 7. Keatley, D.A.; Clarke, D.D. Crime Linkage: Finding a Behavioral Fingerprint Using the "Path Similarity Metric". J. Police Crim. Psychol. 2020, 35, 240–246. [CrossRef]
- 8. Bayat, S.; Roe, C.M. Driving assessment in preclinical Alzheimer's disease: Progress to date and the path forward. *Alzheimer's Res. Ther.* **2022**, 14, 168. [CrossRef] [PubMed]
- Krogh, B.; Jensen, C.S.; Torp, K. Efficient in-memory indexing of network-constrained trajectories. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016; pp. 1–10.
- 10. Grossi, R.; Marino, A.; Moghtasedi, S. Finding structurally and temporally similar trajectories in graphs. In Proceedings of the 18th International Symposium on Experimental Algorithms (SEA 2020), Catania, Italy, 16–18 June 2020.
- 11. Tao, Y.; Both, A.; Silveira, R.I.; Buchin, K.; Sijben, S.; Purves, R.S.; Laube, P.; Peng, D.; Toohey, K.; Duckham, M. A comparative analysis of trajectory similarity measures. *GIScience Remote Sens.* **2021**, *58*, 643–669. [CrossRef]
- 12. Shang, S.; Ding, R.; Zheng, K.; Jensen, C.S.; Kalnis, P.; Zhou, X. Personalized trajectory matching in spatial networks. *VLDB J.* **2014**, 23, 449–468. [CrossRef]
- Qiu, M.; Pi, D. Mining frequent trajectory patterns in road network based on similar trajectory. In Proceedings of the Intelligent Data Engineering and Automated Learning–IDEAL 2016: 17th International Conference, Yangzhou, China, 12–14 October 2016; Proceedings 17. pp. 46–57.
- 14. Won, J.-I.; Kim, S.-W.; Baek, J.-H.; Lee, J. Trajectory clustering in road network environment. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining, Nashville, TN, USA, 30 March–2 April 2009; pp. 299–305.
- 15. Mao, Y.; Zhong, H.; Xiao, X.; Li, X. A segment-based trajectory similarity measure in the urban transportation systems. *Sensors* **2017**, *17*, 524. [CrossRef]
- 16. Xia, Y.; Wang, G.-Y.; Zhang, X.; Kim, G.-B.; Bae, H.-Y. Spatio-temporal similarity measure for network constrained trajectory data. *Int. J. Comput. Intell. Syst.* **2011**, *4*, 1070–1079.
- 17. Yuan, H.; Li, G. Distributed in-memory trajectory similarity search and join on road network. In Proceedings of the 2019 IEEE 35th international conference on data engineering (ICDE), Macao, China, 8–11 April 2019; pp. 1262–1273.
- Kim, J.; Mahmassani, H.S. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transp. Res. Procedia* 2015, 9, 164–184. [CrossRef]
- Chang, J.-W.; Bista, R.; Kim, Y.-C.; Kim, Y.-K. Spatio-temporal similarity measure algorithm for moving objects on spatial networks. In Proceedings of the Computational Science and Its Applications–ICCSA 2007: International Conference, Kuala Lumpur, Malaysia, 26–29 August 2007; Proceedings. Part III 7. pp. 1165–1178.
- 20. Yuan, L.; Li, D.; Hu, S. A map-matching algorithm with low-frequency floating car data based on matching path. *EURASIP J. Wirel. Commun. Netw.* **2018**, 2018, 146. [CrossRef]
- 21. Wan, Z.; Dodge, S.; Bohrer, G. Leveraging similarity analysis to understand variability in movement behavior. *Trans. GIS* 2023, 27, 1441–1466. [CrossRef]
- Custers, B.; Kerkhof, M.V.D.; Meulemans, W.; Speckmann, B.; Staals, F. Maximum physically consistent trajectories. ACM Trans. Spat. Algorithms Syst. 2021, 7, 1–33. [CrossRef]
- 23. Haidri, S.; Haranwala, Y.J.; Bogorny, V.; Renso, C.; da Fonseca, V.P.; Soares, A. PTRAIL—A python package for parallel trajectory data preprocessing. *SoftwareX* 2022, *19*, 101176. [CrossRef]
- Zheng, L.; Xia, D.; Zhao, X.; Tan, L.; Li, H.; Chen, L.; Liu, W. Spatial-temporal travel pattern mining using massive taxi trajectory data. *Phys. A Stat. Mech. Its Appl.* 2018, 501, 24–41. [CrossRef]
- 25. Yang, C.; Gidofalvi, G. Fast map matching, an algorithm integrating hidden Markov model with precomputation. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 547–570. [CrossRef]
- Goh, C.Y.; Dauwels, J.; Mitrovic, N.; Asif, M.T.; Oran, A.; Jaillet, P. Online map-matching based on hidden markov model for real-time traffic sensing applications. In Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 16–19 September 2012; pp. 776–781.
- 27. Chao, P.; Xu, Y.; Hua, W.; Zhou, X. A survey on map-matching algorithms. In Proceedings of the Databases Theory and Applications: 31st Australasian Database Conference, ADC 2020, Melbourne, Australia, 3–7 February 2020; pp. 121–133.
- Raymond, R.; Morimura, T.; Osogami, T.; Hirosue, N. Map matching with hidden Markov model on sampled road network. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 2242–2245.
- 29. Lee, W.-C.; Krumm, J. Trajectory preprocessing. In *Computing with Spatial Trajectories*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 3–33.

- 30. Li, F.; Shi, W.; Zhang, H. A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3695–3705. [CrossRef]
- 31. Moayedi, A.; Ali Abbaspour, R.; Chehreghan, A.; Mojtabaee, P. The credibility evaluation of the trajectory clustering results using a user-defined similarity. *Earth Obs. Geomat. Eng.* **2021**, *5*, 132–144.
- 32. Besse, P.C.; Guillouet, B.; Loubes, J.-M.; Royer, F. Review and perspective for distance-based clustering of vehicle trajectories. *IEEE Trans. Intell. Transp. Syst.* 2016, 17, 3306–3317. [CrossRef]
- 33. Koutra, D.; Parikh, A.; Ramdas, A.; Xiang, J. *Algorithms for Graph Similarity and Subgraph Matching*; Technical Report; Carnegie-Mellon-University: Pittsburg, PA, USA, 2011.
- 34. Chris, C. Taxi Trajectory Data from ECML/PKDD 15: Taxi Trip Time Prediction (II) Competition. Available online: https://www.kaggle.com/datasets/crailtap/taxi-trajectory/data (accessed on 24 April 2015).
- 35. Furtado, A.S.; Kopanaki, D.; Alvares, L.O.; Bogorny, V. Multidimensional similarity measuring for semantic trajectories. *Trans. GIS* **2016**, *20*, 280–298. [CrossRef]
- 36. Liang, M.; Liu, R.W.; Li, S.; Xiao, Z.; Liu, X.; Lu, F. An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation. *Ocean. Eng.* **2021**, 225, 108803. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.