Kufa Journal of Engineering
Vol. 16, No. 1, January 2025, P.P. 57-79
Article history: Received 27 April 2024, last revised 26 May 2024, accepted 13 June 2024

GPS TRAJECTORY CLUSTERING FOR SPATIO – TEMPORAL BEHAVIOR ANALYSIS: THE APPLICATION OF HEATMAP TECHNIQUES AND SPATIO- TEMPORAL DUAL GRAPH NEURAL NETWORK

Mais Muhanad[1] and Wadhah R.Baiee [2]

[1] Department of Software/ College of Information Technology/ University of Babylon/ Iraq Email maismm.sw.msc@student.uobabylon.edu.iq

[2] Department of Software/ College of Information Technology/ University of Babylon/ Iraq Email wadhah.baiee@uobabylon.edu.iq

https://doi.org/10.30572/2018/KJE/160105

ABSTRACT

The introduction of GPS technology has led to the creation of vast amounts of spatio-temporal data, which captures the movement patterns of different things. Efficient allocation of resources to ensure user satisfaction is a crucial factor in shaping the future of urban planning and development. It is required to comprehend the factors that can contribute to the creation of methods for studying user behaviours using a substantial number of persons within a brief timeframe. It is essential to employ appropriate clustering approaches to analyze this data in order to comprehend spatio-temporal behaviours.Heatmaps offer a graphical display of changes in density across both location and Time, making them a user-friendly tool for initial data analysis and identifying areas of high activity. The Spatio-Temporal Dynamic Graph Neural Network (ST-DGNN) utilizes graph neural networks to represent the intricate connections present in spatio-temporal data, encompassing both spatial interdependencies and temporal changes. Our methodology improves the accuracy and interpretability of trajectory clustering by integrating different methods. The suggested method has been shown to identify relevant clusters effectively and reveal noteworthy spatio-temporal characteristics through experimental analysis on real-world GPS datasets. The research utilizes a dataset comprising 182 users for analysis. Numerous measures are taken to boost the clustering accuracy of the applied techniques, including addressing missing values and outliers.

Additionally, this thesis introduces a framework for time estimation based on graph-based deep learning, termed Spatio-Temporal Dual Graph Neural II Networks (STDGNN). The method entails constructing node-level and edge-level graphs that depict the adjacency connections between intersections and road segments. The results showed a number of cluster changes in each period of time dependent on move users and period; for example, the (2592) cluster of period one hour.

KEYWORDS

GPS trajectory, heatMap techniques, deep clustering, Spatio-Temporal, Dual Graph Neural Network.

1. INTRODUCTION

Gadgets often come equipped with GPS devices, which help gather spatiotemporal data. These devices position items, giving rise to a wealth of location data and trajectories that show the many moving objects' movement patterns. The insights into the movements and behaviours of moving objects provided by these trajectories are useful for applications such as traffic flow analysis and navigation.(Dutta, Das, and Patra, 2022). These days, mobile devices are utilized extensively and include many applications that enable user agreements. The case study timeline is represented by rating matrices showing user behaviours (Alasadi and Baiee, 2016). Trajectory data mining has become increasingly popular, drawing interest from academics in various fields, including computer science, sociology, geography, and more. The objective is to extract meaningful information from trajectory data. (Chaker, Aghbari, and Junejo, 2017). Preprocessing methods include data point calibration, noise removal from lost or weak GPS signals, and sampling missing data points to clean up trajectory data. The next data modelling step creates the foundational structure for trajectory data mining (Moreira-Matias et al., 2016). A clustering algorithm based on machine learning, designed to accommodate the sequential nature of activity data, will be employed on Clustering-technic Global Positioning System (GPS) trajectory data.

Nowadays, the most popular methods for segment clustering inside trajectory clustering start with developing a distance similarity measuring method, followed by employing a clustering method to complete trajectory clustering (Q. Yu et al., 2019). The fluctuating nature of road sections and trail transitions necessitates the adaptation of dynamic properties in a networked and interactive manner. Modelling any of these aspects independently restricts the accuracy enhancement in estimating intersection and connection issues(Jin, Sha, et al. 2021). A graph-based deep learning platform for street intersections and connections, namely the Dual Space-Time Graph Neural Networks (STDGNN), has been introduced to address these challenges. Specifically, graphs at the nodes and edges are initially established to aptly characterize the relationships concerning the proximity of intersections and road sections. The objective is to extract common space-time correlations between intersections and road segments (Jin, Yan, et al. 2021).

The research's primary contribution is as follows:

- Modelling the conversion of GPS trajectory data to a heat map, following the separation of the spatial and temporal properties, transforming the dynamic node graph, and entering them into an algorithm (STDGNN) to identify The location of the user for each period is more accurate and also helps us to find out the relationship between the trajectory of users.

The subsequent sections of this study are structured as follows: The next section reviews the relevant literature. The integrated approach for mining spatiotemporal travel patterns is then presented in Section 3, including sequence pattern mining, spatial relation mining among trajectory clusters, and spatiotemporal trajectory clustering. Section 4 analyses the spatiotemporal trajectory patterns produced by the datasets. Section 5 wraps up by summarizing the results of this study and suggesting potential avenues for future research.

2. PREVIOUS LITERATURE

In this section, we present relevant studies on the clustering of GPS trajectories to analyze spatio-temporal behaviour. Deep learning-based techniques have gained importance recently in the GPS trajectory clustering process for spatiotemporal behaviour analysis. Using GPS data, activity sequences for participants are created. The study characterizes the spatiotemporal activity participation behaviour of 1461 participants at AirVenture, a planned special event held in Oshkosh. Dataset Methodologies Clustering customer trips is a convolutional neural network approach called CSRNet. (Agglomerative (bottom-up) Hierarchical Clustering using Ward's technique). The results of six clusters and statistical tests verify notable mobility and time utilization variations(Abkarian et al. 2022). To determine the demand for bus travel dispersion patterns using a framework for journey time prediction called Spatio-Temporal Dual Graph Neural Networks. One of the main responsibilities for creating intelligent transportation systems is calculating travel time and information. Datasets Beijing, Shanghai, and Porto Heat maps provide an analysis of the spatiotemporal trajectory patterns produced by the datasets. The results demonstrate that STDGNN performs noticeably better than a number of cutting-edge basics (Jin, Yan, et al. 2021). Forecasting urban gatherings can assist in keeping an eye on a range of odd group activities, which is crucial for traffic management and public safety in smart cities with complex spatiotemporal connections. Techniques CgNet for urban congregation prediction results demonstrate the advantages of our model beyond several baselines(T. Chen et al. 2020). Deep Clustering Optimization Method for Graph Neural Networks (DCOM-GNN) data analysis relies heavily on deep clustering. As graph data is so common these days, new deep clustering models are always being proposed for graphs. The results enhance the performance of several deep clustering models on graphs(Yang et al. 2023). It shows how to reconsider GPS trajectory mining methods in the context of big data behavioural analysis from a geographical, semantic, and quantitative perspective. Studies of datasets employ real GPS trajectory data from visitors to China's Palace Museum. Sequential-based clustering algorithms like k-means, BIRCH, and DBSCAN (like PrefixSpan) could be used to identify comparable trajectories. Outcomes The geographic perspective can provide patterns of users' temporal and spatial distribution, which can be useful in preventing resource or energy waste. Additionally, grouping users based on movement metrics can facilitate providing efficient, eco-friendly, and energy-efficient services to various user groups (Huang and Wang 2022). GPS trajectory data shows strong spatiotemporal association when restricted by city function zoning, travel preference, and road network structure. Cab trajectory datasets were analyzed in Wuhan, China. Spatiotemporal DBSCAN techniques. As a result, we isolate 21,416 sluggish trajectories during the workday and 34,156 slow trajectories during the weekend (Liu et al. 2020). Mobile phone network operators' call detail records (CDR) have been widely used to model and analyze human-centric mobility. Datasets that incorporated USB and GPS data DBSCAN-based clustering techniques. Transistor performance results (80% accuracy and 96% recall) (Bonnetain et al. 2021). The paper examines datasets derived from actual GPS trajectory data collected from visitors to the Palace Museum in China. Using algorithms k-means, BIRCH, and DBSCAN. The results have direct implications for the efficient distribution of public resources and can be used to the intelligent management of building energy services. Mining trajectories from both a geographic and quantitative standpoint is quite valuable (C. Yu and He 2017).

3. METHODOLOGY

The methodology presented in this section combines travel sequence pattern mining inside trajectory clusters with individual-based spatiotemporal trajectory pattern mining. Preprocessing, feature selection, dataset updating, HeatMap, and the Spatial-Temporal Graph Neural Network (STDGNN) are some of the components of the suggested system. The proposed system is shown in Fig. 1.

3.1. Dataset

This GPS trajectory data was collected over more than five years (April 2007 to August 2012) by 182 people as part of the Geolife project (Microsoft Research Asia). A sequence of timestamped dots with latitude, longitude, and altitude information identifies the GPS track of this dataset. This collection comprises 17,621 trajectories totaling 1,292,951 kilometers and 50,176 hours of travel time. These trajectories were taken using a variety of GPS recorders and GPS phones, and they show a range of sampling rates. For instance, every 1 to 5 seconds or every 5 to 10 meters per point, a dense representation of 91.5 per cent of the trajectories is captured("Microsoft Research – Emerging Technology, Computer, and Software Research" n.d.). Training is one-shot and it should be repeated training of to update the dataset.

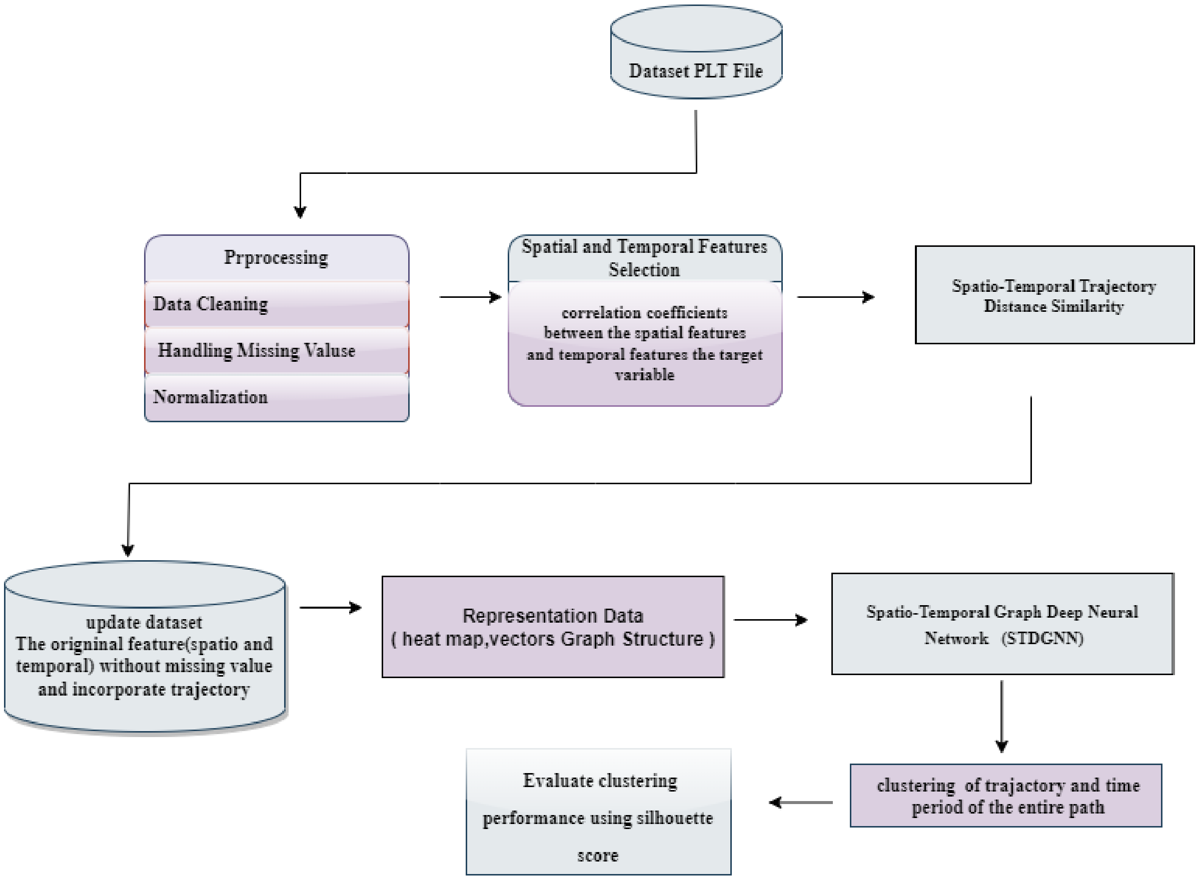


Fig. 1 demonstrates the proposed system

Table 1: The details of the dataset's features

NO	Feature	Description
1	X	Latitude in decimal degrees
2	Y	Longitude in decimal degrees
3	All set to 0	All set to 0 for this dataset
4	Altitude in feet	Altitude in feet (-777 if not valid).
5	Date number of days	Date - number of days (with fractional part) that have passed since 12/30/1899
6	Date	Date as a string
7	Time	Time as a string

3.2. Preprocessing

Techniques for preprocessing data are essential for getting datasets ready. These methods can be broadly divided into two groups: adding and changing attributes or choosing data items and attributes for analysis. These techniques cover a range of approaches to common problems with datasets, including noise, missing values, and inconsistent data (Abkarian et al. 2022). A significant number of data items may have missing attribute values, and the performance of clustering models may suffer as a result. Discusses a number of methods for dealing with missing values. By ignoring missing values during analysis, any data objects that have missing values are not included in the analysis. The selection of a strategy is contingent upon the particular features of the dataset and the objectives of the study (Pancasila, Haryono, and Sulistyo 2020).

Normalization is a method used to deal with this inconsistent issue. Ensuring that each feature in the dataset is represented in the same measurement unit. Normalization helps prevent the dominance of large values over smaller ones and ensures a fair comparison between them by doing this. It also helps reduce differences from differing scales of distinct features (Young and Johnson 2015).

There are several ways to normalize data, such as z-score normalization and min-max normalization. Equation 1 is utilized in min-max normalization to determine the value (Pancasila, Haryono, and Sulistyo 2020).

$$X' = \frac{x - min_x}{max_x - min_x} \quad (1)$$

where X' represents the normalized value corresponding to each feature. x is the original value for a feature. min_x refers to the minimum original value for a feature. max_x represents the maximum original value for a feature.

3.3. Feature selection

Feature selection (FS) the process of choosing the most significant and pertinent collection of features for a specific problem from the initial feature set is done. It is essential for many reasons, including lowering computing demands, boosting prediction performance, improving data comprehension, and lessening the difficulties caused by the curse of dimensionality. Features can generally be classified as irrelevant, weakly relevant, or extremely relevant. The main objective of FS approaches is finding the highly relevant traits that are most instructive for the target class. FS techniques can be divided into several groups: filter, wrapper, embedding, and hybrid methods (Abkarian et al., 2022).

We use the correlation method, which evaluates each feature with the target by considering its predictive capacity, with the main goal being to assess the correlation between features and the target class. The feature's relevance lies in its range from (1) to (-1) so that an attribute's weight equal to zero indicates no correlation, whilst a weight around ±1 indicates a strong correlation (Tiwari, 2013).

Equation (2) computes the correlation between any feature and the target class.

$$Cor(x,y) = \frac{\sum(Xi-\bar{X})(Yi-\bar{Y})}{\sqrt{\sum(Xi-\bar{X})^2}\sqrt{\sum(Yi-\bar{Y})^2}} \tag{2}$$

Where X is the feature, Y is the target class, $\bar{Y}$ is the average of the target class, and $\bar{X}$ Is the average of the feature.

3.4. Spatio-temporal trajectory distance similarity

As the amount and accessibility of trajectory data grow, the significance of storing, retrieving, and examining trajectories is rising. A crucial aspect of trajectory analysis involves the calculation of trajectory similarity(Buchin, Dodge, and Speckmann 2014). A similar number of the spatio-temporal distance function's components, including directional, temporal, and spatial distance between two line segments, were determined, and the weighted approach was used to integrate the various distance components(Zhang, Lee, and Lee 2018).

Spatio-temporal trajectory similarity incorporates considerations of spatial distance, temporal distance, and additional directional distance. The prescribed distance functions are outlined as follows(Y. Tian et al. 2021):

Looking for Comparable Paths on Road Networks To obtain comparable paths on road networks, one could utilize one of the subsequent techniques (Hwang, Kang, and Li 2005).

Method 1: Using the spatiotemporal distance between trajectories as a basis, find related trajectories.

Method 2: Refine similar paths by considering their spatial distance and filter off paths based on their temporal similarity.

Method 3: Refine comparable trajectories based on temporal distance and filter trajectories based on spatial similarity.

Trajectories' Similarity in Space on the Road Network Let P be a set of points of interest on a particular route network. Then, according to definitions, the spatial resemblance between two paths, TR_A and TR_B, is

$$\text{SimPOI}(TR_A, TR_B, P) = \begin{cases} 1, \text{ if } \forall p \in P, \text{ p is on } TR_A \text{ and } TR_B \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

We require a temporal similarity measure and a spatial similarity to use Method 3 for finding similar trajectories. The inverse of temporal distance is known as temporal similarity. Unlike the description of temporal distance in Method 2, temporal distance can be defined as the difference between the times two objects pass the same point in space when a Point of Interest (POI) is specified.

Temporal Distance between Trajectories for one POI Suppose that p ∈ P, and P is the set of POI. Then the temporal distance between two trajectories TR_A and TR_B is (Hwang, Kang, and Li, 2005) .

$$dist_T\ (TR_A,TR_B,\ p) = |t(TR_A,\ p) - t(TR_B,\ p)| \quad (4)$$

When neither TRB nor TRA crosses p, the temporal distance is considered infinite. Each trajectory, TR, is represented as a point t(TR)=(t(TR,p1), t(TR,p2),.., t(TR,pk)) in a kdimensional space, where k is the number of POIs if we take t(TR, pi) as the time the i-th POI was passed. Then, for a collection of points of interest, The temporal distance between two trajectories is defined as the LP distance in this k-dimensional space.

Temporal Distance between Trajectories for a Set of POIs Suppose that P is a set of POIs and TRA and TRB are two trajectories. Then, the temporal distance between TR_A and TR_B is (Hwang, Kang, and Li 2005).

$$distT\ (TRA,TRB,\ P) = Lp(TRA,TRB,\ p)= \left(\sum_{i=1}^{k} |(p_i(TR_A) - p_i(TR_B)|\right)^{\frac{1}{p}} \quad (5)$$

3.5. Representation data

The process of converting unprocessed data into a format appropriate for a particular job or analysis is known as representation data. Typically, this transformation entails lowering dimensionality, identifying pertinent features, and storing the data in a structured format, highlighting significant traits or trends.

3.5.1. Heat maps

Defining a metric for comparing datasets across different periods is challenging due to its complexity. Therefore, heat maps are employed in various periods. To streamline the process and enable clustering, feature extraction is necessary to reduce the variable size. Initially, the original heat map dataset undergoes feature extraction to decrease its dimensions. Subsequently, the derived features from the heat maps are utilized for clustering and analyzing spatial-temporal patterns (Gu et al. 2012).

Analysis of hours of the day: Photographing GPS points for different hours of the day using heat maps, it is likely that some distinctive features will appear differently, and this helps the user to analyze the distribution of GPS points during specific hours, as well as by obtaining algorithm more accurate results.

Algorithm 1 . Heat maps GPS trajectory each period time dataset

---

Algorithm: Heat maps GPS trajectory each period time dataset

Input: GPS trajectory user each period of Time

Output: heat maps each GPS trajectory user each period of Time

---

Proses:

```
Begin:
  # Create a grid
 “1.  grid = np.zeros(grid_size)
    # Fill the grid with data values
  2. for coord, value in data.items():
  3.  x, y = coord
  4. if 0 <= x < grid_size[0] and 0 <= y < grid_size[1]:
  5. grid[x, y] = value  # Plot the heatmap
   6. plt.imshow(grid,cmap=cmap,interpolation='nearest')
 7. plt.colorbar()
 8. plt.show()"
End
```

3.5.2. Representation vector

Transforming 3D heatmap data into a 2D vector format is the first step in converting a dataset heatmap to a vector. This procedure is frequently required to use the data in machine learning algorithms that require vector inputs or additional analysis. There are different ways to flatten a heatmap into a vector. There are two common strategies: Row-wise flattening and concatenating each row of the heatmap matrix to create a single vector. Column-wise Flattening: Concatenate each column of the heatmap matrix to create a single vector. Check that the heatmap's dimensions and values meet your expectations after converting to a vector. Verify that there are no new problems or unexpected changes as a result of the conversion.

3.5.3. Representation graph

Interpreting the vector's constituents into graph qualities, such as nodes, edges, or other characteristics, is necessary to convert a vector representation into a graph. Once the vector items have been assigned to the graph attributes, check to see if the final graph structure is what you expected. Verify that all other graph features are adequately represented, nodes are connected correctly, and attribute values are assigned correctly. After creating the graph, it can be utilized for various graph-based tasks, including machine learning, network analysis, and visualization as illustrated in Fig. 2.

3.6. Spatio-Temporal dual graph neural network

A Spatio-Temporal Graph Deep Neural Network (ST-GNN or ST-DGNN) is a neural network specifically crafted to model and analyze data featuring spatial and temporal dependencies, typically represented in the form of a graph. This network is widely employed in diverse domains such as transportation, environmental science, social network analysis, and others, where the data exhibits both geographical (spatial) and time-dependent (temporal) characteristics (Jin, Yan, et al. 2021). Let's dissect the essential elements:

- Spatio-Temporal data types:

Spatial Component: Encompasses the geographical or spatial details within the data, such as locations, coordinates, or relationships among entities in a spatial domain.

Temporal Component: Encompasses the time-related details within the data, including timestamps, sequences, or the chronological order of events (Jin, Yan, et al. 2021).

Due to the substantial variations in data collecting and representation methods in real-world applications, it is possible to classify ST data into distinct kinds. Varying application contexts and ST (Spatio-Temporal) data types give rise to different sorts of data mining tasks and issue formulations, necessitating the use of diverse deep learning models; classify the spatio-temporal data into the following categories: event, trajectory, point reference, raster, and video Fig. 3 depicts an illustration of the ST data types (Wang, Cao, and Yu 2022).

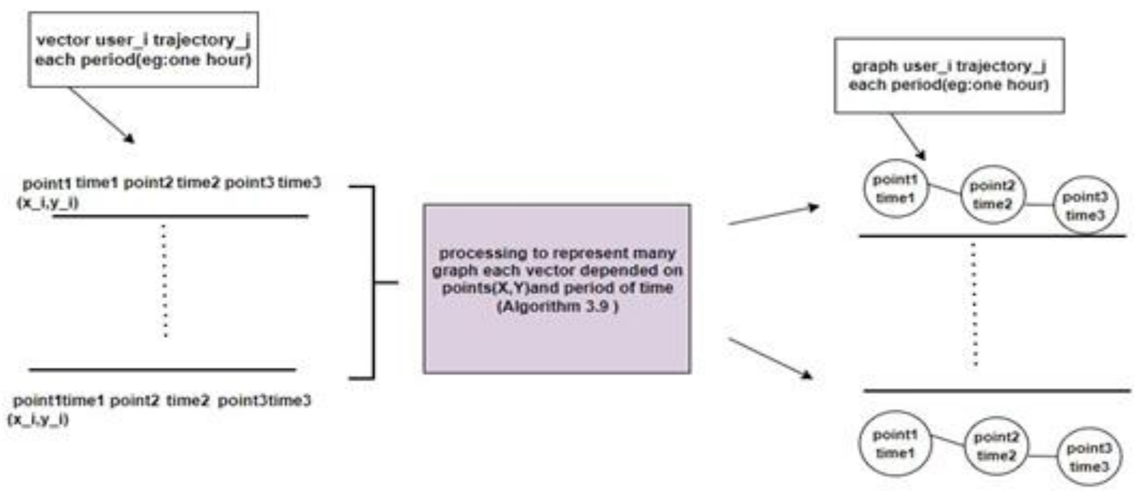


Fig. 2 shows Representation of Graph

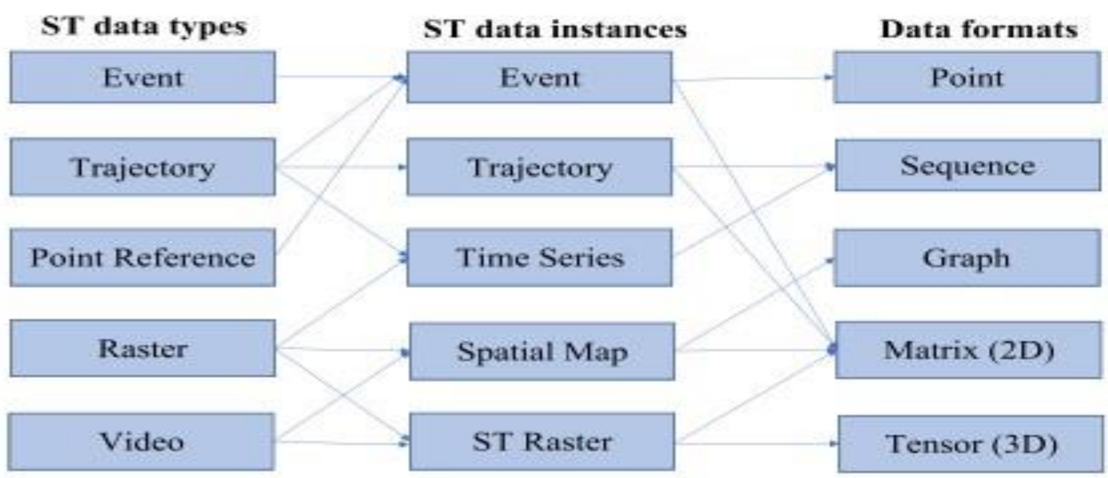


Fig. 3: representations of different ST data types

- Graph Representation:

The data is commonly organized in a graph format, where entities (such as locations or individuals) are depicted as nodes, and edges represent the connections or interactions between these entities. This graph framework is employed to capture the spatial dependencies among entities. Utilizing graph-structured data has the potential to enhance current methodologies. Initially, we create an attribute spatio-temporal graph, wherein each node corresponds to a region, edges signify mobility between regions, and node attributes indicate the distribution of Points of Interest (PoI) within an area (Hou et al. 2022).

- Deep Neural Network

Deep neural networks are utilized to capture intricate connections and hierarchical features inherent in spatio-temporal data. These networks may include diverse layers, including convolutional layers for extracting spatial features, recurrent layers for capturing temporal dependencies, and graph neural network layers to manage the inherent graph structure. The initial module employs a graph convolutional neural network to grasp spatial dependencies, while the second module adopts an encoding-decoding temporal learning structure incorporating a self-attention mechanism(Hou et al. 2022). Our research focuses primarily on dynamic social networks, which change over Time due to frequent node and link additions and deletions. While some previous research has examined learning dynamic networks, we model time-evolving networks more effectively by utilizing the STGNN idea and attention mechanism(Min et al. 2021).

In order to capture the spatial dependencies, we opt for the simple graph convolution strategy to gather information from neighbouring nodes that are within a 1-hop distance(Jin, Yan, et al. 2021).

$$\mathrm{Z(l+1)} = \sigma\left(\mathrm{L}.\,\mathrm{Z}^{(1)}.\,\mathrm{W}^{(1)}\right) \ldots \ldots \ldots \text{(eq 6)}$$

Where $\mathrm{L} = \widetilde{\mathrm{D}}^{-1/2}\widetilde{\mathrm{A}}\widetilde{\mathrm{D}}^{-1/2}$, $\widetilde{\mathrm{A}} = \mathrm{A} + \mathrm{I}, \widetilde{\mathrm{D}} = \mathrm{diag}(\sum_{\mathrm{j}} \widetilde{\mathrm{A}}_{1\mathrm{j}} \ldots\ldots \sum_{\mathrm{j}} \widetilde{\mathrm{A}}_{\mathrm{Nj}})$.

A is a graph that is symmetrical in terms of its adjacency, while I is a graph that represents identity. The symbol σ Represents a non-linear activation function, and we utilize the ReLU function for graph convolution. Z(l) represents the input graph representation, while Z(l+1) represents the output graph representation.

The temporal convolution operation is formulated as follows (Jin, Yan, et al. 2021):

$$Z_T = \|_0^N \sigma\left(z_i * \Gamma_1(\theta_1)\right) \odot \sigma_2\left(z_i * \Gamma_2(\theta_2)\right) \qquad (7)$$

Where θ_1, θ_2 are the learnable parameters of temporal convolution , $\odot$ is element-wise product operation, σ_1 (.) and σ_2 (.) are activation functions of two different temporal convolution models respectively. Empirically, Tanh function can be selected as σ_1 (.) and Sigmoid function is usually selected as σ_2 (.) to control the ratio of information passed. z_i represents the latent representation of node i. The temporal convolution network is applied to each node in a node-wise graph or edge-wise graph. Thus, the output is the result of integrating all N nodes of a node-wise graph or edge-wise graph.

- Graph Neural Networks (GNNs)

GNNs represent a distinct category of neural networks tailored for processing graph-structured data. They excel at capturing inter-node relationships in a graph, making them well-suited for tasks with spatial dependencies. GNNs typically feature graph convolutional layers that consolidate information from adjacent nodes (Zhao et al. 2022). A Spatial-Temporal Graph Neural Network (SGN) has emerged, blending conventional GNN and temporal learning modules. SGN is a framework that can record spatial and temporal information simultaneously. Correlations for non-Euclidean data. In the realm of intelligent transportation, STGNN is commonly employed in numerous studies (W. Chen et al. 2020). Current graph deep learning models use Graph Convolutional Networks (GCN) and a weighted adjacency matrix to comprehend correlations between links. However, these models often overlook intersections that connect different road segments (Fang et al. 2020). Fig. 4 The detailed architecture of Spatio-Temporal Graph Deep Neural Network. Now that a layer has been added to STGNN, we can enter all GPS points anytime. This layer operates on a dynamic node number at various times.

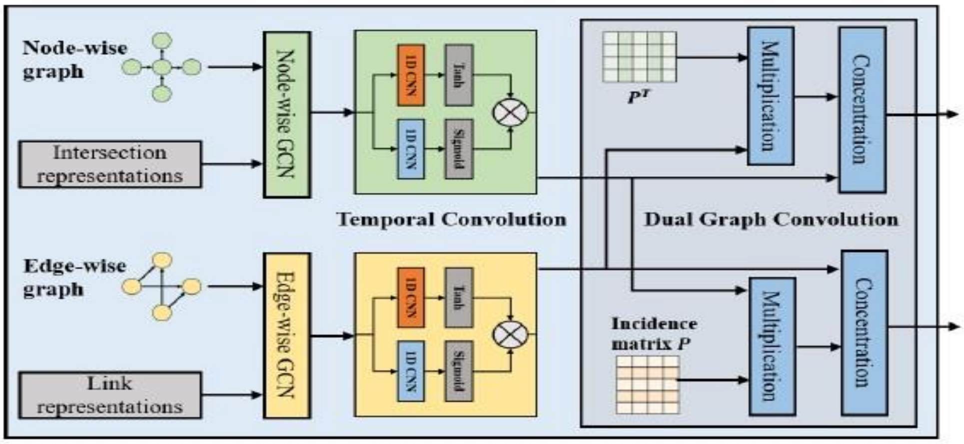


Fig. 4. The detailed architecture of Spatio-Temporal Graph Deep Neural Network

- Deep Clustering Graph Neural Network

The Deep Clustering Optimization Method for Graph Neural Networks is a simple add-on that makes it possible for pre-existing deep clustering models on graphs to take a more appropriate clustering-oriented direction. GNN comprises two distance optimization modules: one for within-cluster distances and another for between-cluster distances. The first module's goal is to more logically modify the inter-cluster distance of the initial model output, which allocates weight coefficients to clusters based on their similarities. Throughout this procedure, we consider both structural and attribute similarity to determine the relative importance of various clusters (Yang et al. 2023). Our suggested framework is based on graph neural networks (GNNs) and group nodes based on their features and graph topology (similar features and strong connections between nodes in the same cluster). The underlying premise of our approach is that node attributes provide a useful starting point for determining cluster assignments. Because many real-world networks have the homophily property, this is a reasonable assumption. Furthermore, because of the smoothing impact of (ML) multi-layer operation, the properties of nodes in strongly connected communities tend to Become similar even in disassortative networks. Let X¯ be the node representation matrix that one or more Players have given. We use a multi-layer (ML) with softmax on the output layer to calculate a cluster assignment for the nodes. This ML maps each node feature (xi) to the ith row of a soft cluster assignment matrix S:

$$\bar{X} = GNN(X, \tilde{A}; \Theta\text{GNN}) \tag{8}$$

$$S = MP(\bar{X}\,; \Theta\text{MP}) \tag{9}$$

where ΘGNN and ΘMP are trainable parameters(Maria Bianchi, Grattarola, and Alippi 2020).

- Evaluate clustering

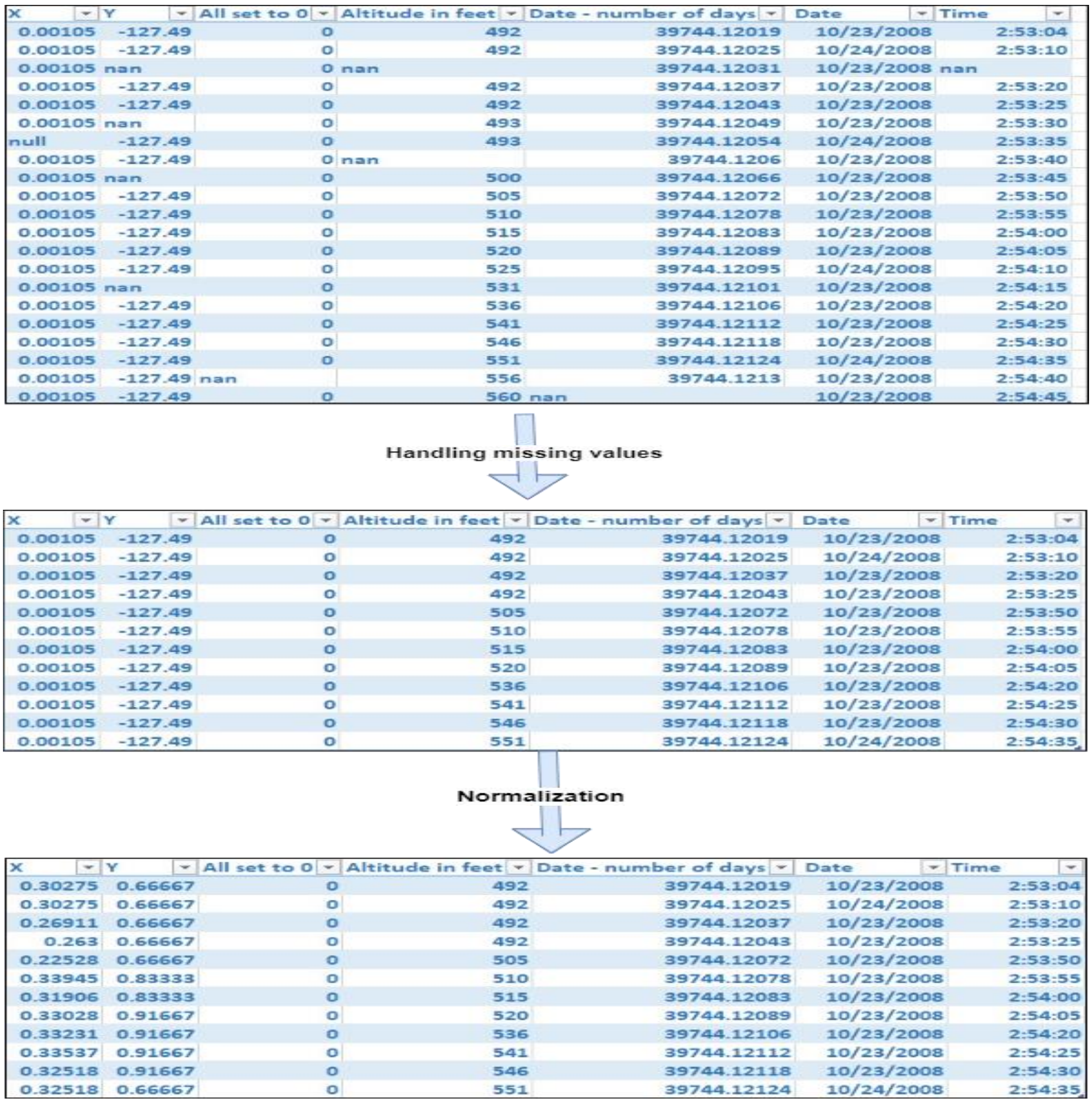
The cikit-learn silhouette score function is used to determine each sample's mean silhouette coefficient. The silhouette coefficient is computed by taking into account the mean nearest-cluster distance (b) and mean intra-cluster distance (a) of each data point. The silhouette coefficient for a given sample is (b − a)/max(a,b).

• The data point is in the correct cluster if the silhouette score is close to + 1.
• If the data point's silhouette score is close to 0, it may be a part of another cluster.
• A silhouette score that is close to -1 indicates that the data point is in the incorrect cluster (F. Tian et al. 2014).

4. RESULTS AND DISCUSSIONS

The study seeks to cluster user activities based on GPS data, emphasizing both spatial and temporal aspects. Various enhancements were implemented to enhance clustering accuracy, and the model underwent testing using data from 182 users. The Evaluation involved preprocessing, feature selection, similarity data analysis, heat maps data examination, and applying a Graph Deep Neural Network.

The sole preprocessing measure applied was addressing missing values and localization. The results are briefly presented in Fig. 3. The steps are on a small sample of the dataset.After the data cleaning step, the missing values are processed in each feature. The mean method was used to delete the missing values (see the features in Fig. 3). Normalization is used to lessen the effects of large differences in data. The method of min-max localization, involving scale translation, is utilized to adjust feature values within the range of (-1) to (1), given the presence of negative values in the data, as illustrated in Fig. 5.



X	Y	All set to 0	Altitude in feet	Date - number of days	Date	Time
0.00105	-127.49	0	492	39744.12019	10/23/2008	2:53:04
0.00105	-127.49	0	492	39744.12025	10/24/2008	2:53:10
0.00105	nan	0	nan	39744.12031	10/23/2008	nan
0.00105	-127.49	0	492	39744.12037	10/23/2008	2:53:20
0.00105	-127.49	0	492	39744.12043	10/23/2008	2:53:25
0.00105	nan	0	493	39744.12049	10/23/2008	2:53:30
null	-127.49	0	493	39744.12054	10/24/2008	2:53:35
0.00105	-127.49	0	nan	39744.1206	10/23/2008	2:53:40
0.00105	nan	0	500	39744.12066	10/23/2008	2:53:45
0.00105	-127.49	0	505	39744.12072	10/23/2008	2:53:50
0.00105	-127.49	0	510	39744.12078	10/23/2008	2:53:55
0.00105	-127.49	0	515	39744.12083	10/23/2008	2:54:00
0.00105	-127.49	0	520	39744.12089	10/23/2008	2:54:05
0.00105	-127.49	0	525	39744.12095	10/24/2008	2:54:10
0.00105	nan	0	531	39744.12101	10/23/2008	2:54:15
0.00105	-127.49	0	536	39744.12106	10/23/2008	2:54:20
0.00105	-127.49	0	541	39744.12112	10/23/2008	2:54:25
0.00105	-127.49	0	546	39744.12118	10/23/2008	2:54:30
0.00105	-127.49	0	551	39744.12124	10/24/2008	2:54:35
0.00105	-127.49	nan	556	39744.1213	10/23/2008	2:54:40
0.00105	-127.49	0	560	nan	10/23/2008	2:54:45

Handling missing values

X	Y	All set to 0	Altitude in feet	Date - number of days	Date	Time
0.00105	-127.49	0	492	39744.12019	10/23/2008	2:53:04
0.00105	-127.49	0	492	39744.12025	10/24/2008	2:53:10
0.00105	-127.49	0	492	39744.12037	10/23/2008	2:53:20
0.00105	-127.49	0	492	39744.12043	10/23/2008	2:53:25
0.00105	-127.49	0	505	39744.12072	10/23/2008	2:53:50
0.00105	-127.49	0	510	39744.12078	10/23/2008	2:53:55
0.00105	-127.49	0	515	39744.12083	10/23/2008	2:54:00
0.00105	-127.49	0	520	39744.12089	10/23/2008	2:54:05
0.00105	-127.49	0	536	39744.12106	10/23/2008	2:54:20
0.00105	-127.49	0	541	39744.12112	10/23/2008	2:54:25
0.00105	-127.49	0	546	39744.12118	10/23/2008	2:54:30
0.00105	-127.49	0	551	39744.12124	10/24/2008	2:54:35

Normalization

X	Y	All set to 0	Altitude in feet	Date - number of days	Date	Time
0.30275	0.66667	0	492	39744.12019	10/23/2008	2:53:04
0.30275	0.66667	0	492	39744.12025	10/24/2008	2:53:10
0.26911	0.66667	0	492	39744.12037	10/23/2008	2:53:20
0.263	0.66667	0	492	39744.12043	10/23/2008	2:53:25
0.22528	0.66667	0	505	39744.12072	10/23/2008	2:53:50
0.33945	0.83333	0	510	39744.12078	10/23/2008	2:53:55
0.31906	0.83333	0	515	39744.12083	10/23/2008	2:54:00
0.33028	0.91667	0	520	39744.12089	10/23/2008	2:54:05
0.33231	0.91667	0	536	39744.12106	10/23/2008	2:54:20
0.33537	0.91667	0	541	39744.12112	10/23/2008	2:54:25
0.32518	0.91667	0	546	39744.12118	10/23/2008	2:54:30
0.32518	0.66667	0	551	39744.12124	10/24/2008	2:54:35

Fig. 5. Preprocessing steps

The objective of feature selection is to reveal the most relevant and non-redundant features from the feature set and compute the correlation coefficients between the features, as indicated in Table 2. Exclude the "Altitude in feet" feature because features have values over 50% similar to the target variable(features X, Y).

Table 2: features selection steps

X	Y	Date - number of days	Date	Time
0.045023	0.974296	0	10/23/2008	2:53:04
0.04502	0.97429	7.05E-	10/23/2008	2:53:2

After this step, data is represented in the number of heat maps depending on the user's trajectories and period. Heat maps help the user analyze the distribution of GPS points during specific hours, as illustrated in Fig. 6. Thus, an enhanced trajectory clustering technique**.**

The results of Heat maps are represented in a vector format. These advantages enhance the efficiency of visualizing and interpreting spatial patterns and trends obtained from GPS data.

Finally, The results of vector representations of dynamics graph-based GPS user's data enable more comprehensive analysis, modelling, and visualization of spatial relationships and connectivity within the geographical domain.

The impact of similarity on the representation data and Spatial-Temporal Graph Neural Network (ST-DGNN) can be significant, affecting the accuracy, resilience, and interpretability of the model's results. Fig. 7 below illustrates the different heat maps after applying Spatio-Temporal Trajectory Distance Similarity.

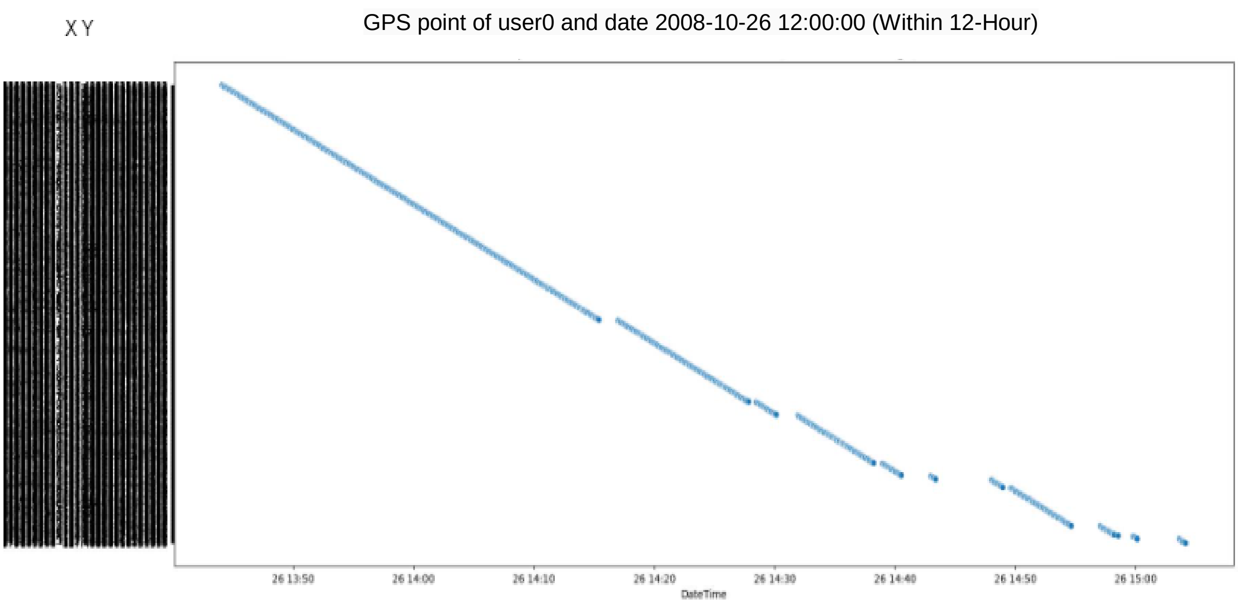


Fig. 6. The details of the trajectory of the user for the period of Time (without similarity).

After preprocessing and representing the data on GPS datasets, the results may include cluster assignments for each GPS graph point. Each graph point is assigned to a specific cluster based on its spatial-temporal features and characteristics of the data using the ST-DGNN algorithm.

The results of the one-hour Time generated (2592) were that the clusters Fig. 8 shows a chart of the top Twelve clusters of the period in an hour. Table 3 illustrates a sample of the details of the C917 cluster. Fig. 9 heat map is a visual representation of cluster C917.

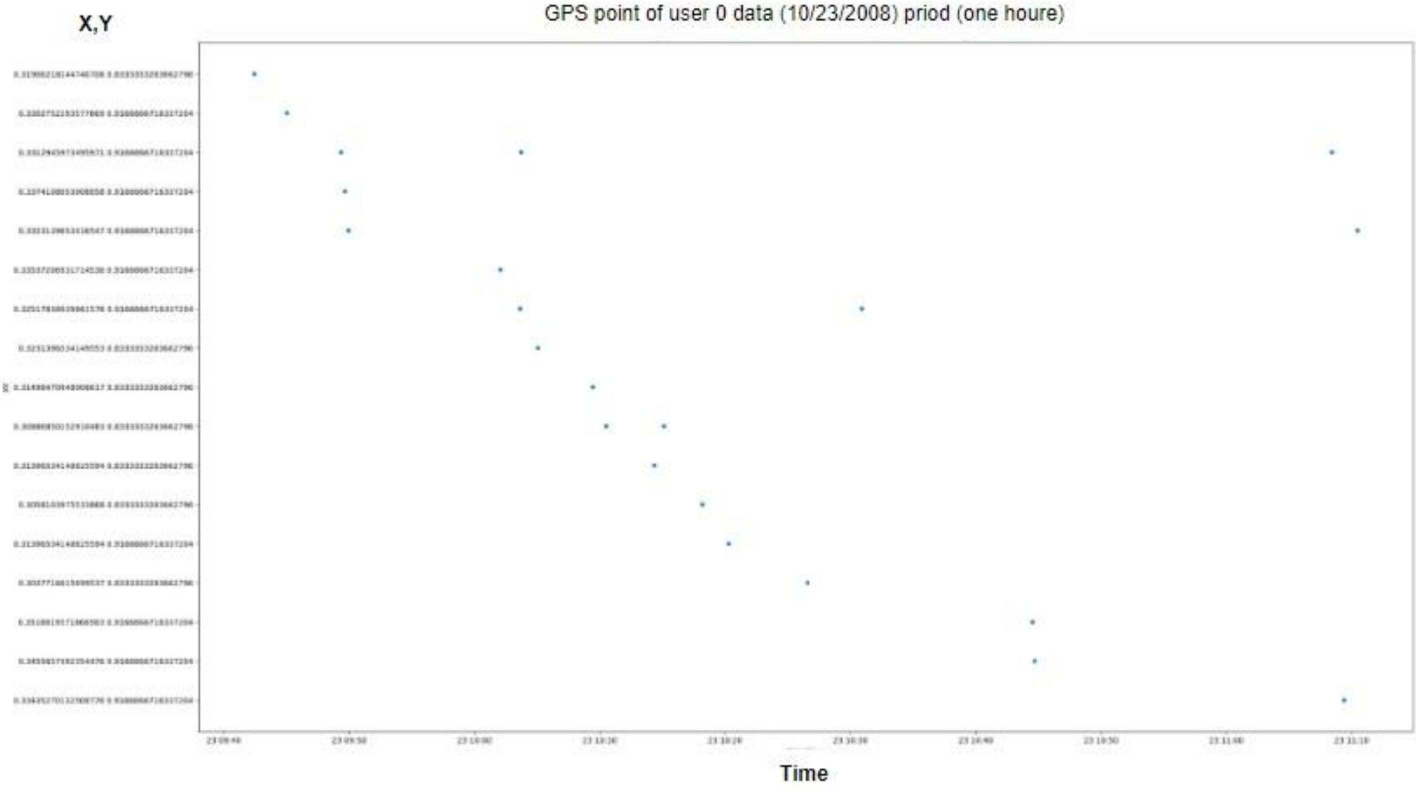


Fig. 7. The details of the trajectory of the user for the period of Time (with spatial-temporal trajectory distance similarity).

Table 3 shows a sample of the details of cluster C917 .

Event	feature_x	feature_y	Timestamp	Cluster
User0Graph8N1	0.505606524	0.072727	10/26/2008 14:52	917
User0Graph8N3	0.495412844	0.072727	10/26/2008 14:52	917
User0Graph8N4	0.484199796	0.072727	10/26/2008 14:53	917
User0Graph8N5	0.472986748	0.072727	10/26/2008 14:53	917
User0Graph8N6	0.4617737	0.072727	10/26/2008 14:53	917
User4Graph5N1	0.505606524	0.166667	10/26/2008 14:52	917
User4Graph5N2	0.495412844	0.166667	10/26/2008 14:52	917
User4Graph5N3	0.484199796	0.166667	10/26/2008 14:53	917
User4Graph5N4	0.472986748	0.166667	10/26/2008 14:53	917
User4Graph5N5	0.4617737	0.166667	10/26/2008 14:53	917

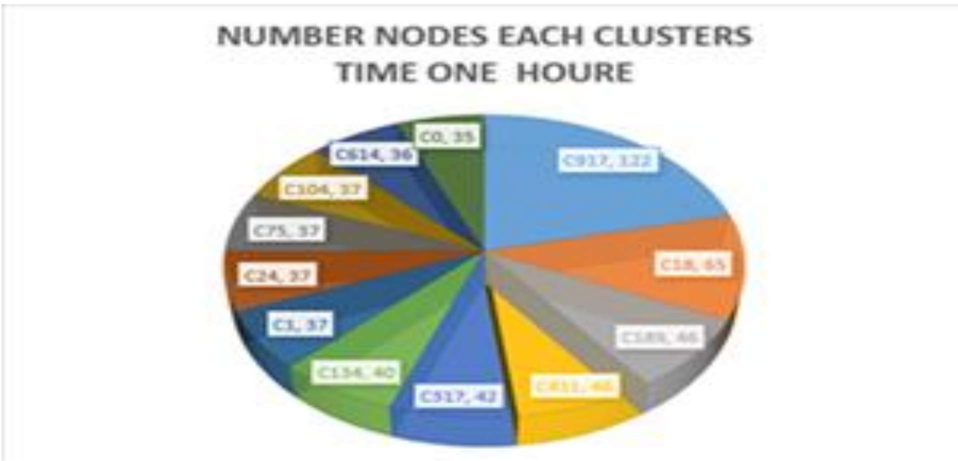


Fig. 8. The top clusters of period a one-hour

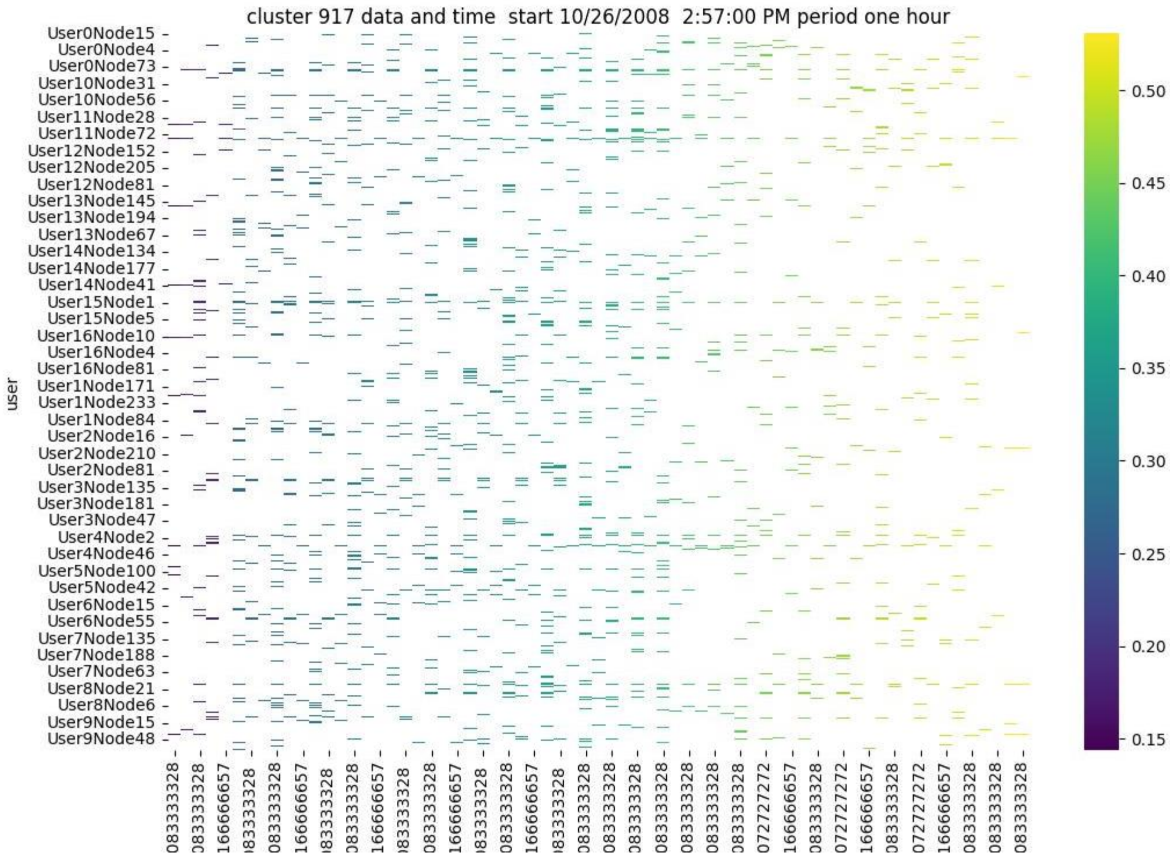


Fig. 9. The details of the cluster 917 content top trajectories users(heat map convert vector after graph) represent (user node) the spatial feature (X, Y) and temporal feature time start 10/26/2008 2:57:00 PM period one hour.

After viewing the sample results, you can find the busy site located within the C917 cluster on 10/26/2008 at 14:52, a period of one hour, and on the latitude and longitude line (feature_x, feature_y) (0.1- 0.2 – 0.3 -0.4 – 0.5,0.07- 0.08- 0.1).

The results of the six-hour time generated (1327) were that the clusters. Fig. 10 shows of the top Twelve clusters of the period in six hours. Table 4 illustrates a sample of the details of clusters.

Table 4: content on the results of (1327) six-hours Time

User	feature_x	Feature_y	timestamp1	Cluster
User0Graph1N1	0.38634	0.66666667	11/10/2008 3:35	1327
User0Graph1N2	0.383282	0.66666667	11/10/2008 3:38	1327
User0Graph1N3	0.381244	0.83333333	11/10/2008 3:41	1327
User6Graph1N1	0.004989	0.05454545	1/16/2009 16:30	1326
User6Graph1N2	0.004852	0.05454545	1/16/2009 16:31	1326
User9Graphq1N1	0.004989	0.05454545	1/16/2009 16:30	1326

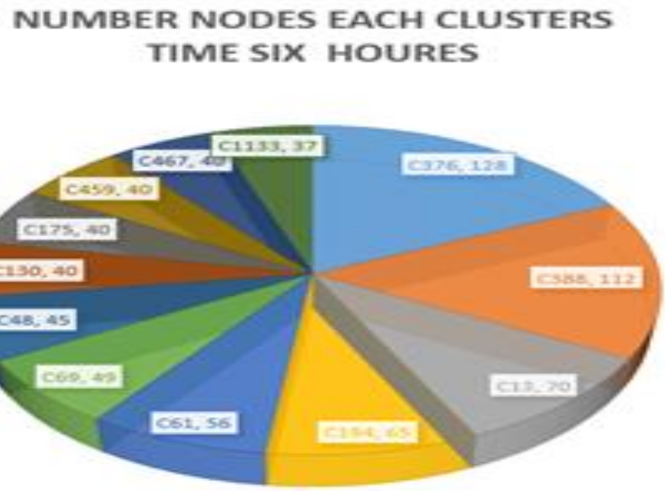


Fig. 10. The top clusters of period a six-hour

5. EVALUATION MODEL

Following the initialization of the DC-STDGNN model with the designated number of clusters. The model is trained using the spatial and temporal data on (2592) cluster period of one hour and (1327) cluster period of six hours.

After training, the model predicts the cluster assignments. All final clustering results used to convert the final_labels. Using the Silhouette Score, which ranges from -1 to 1, a high value indicates that objects are well-matched to their own cluster and poorly matched to neighbouring clusters. The clusters evaluated the Silhouette Score: 0.0.9025799 for one hour time, and the assessed clusters the Silhouette Score: 0.87257 for six hours.

6. CONCLUSION

This research introduces a new framework for processing Clustering-technic Global Positioning System (GPS) trajectory data, called Spatio-Temporal Dual Graph Neural Network (STDGNN). Due to this framework, the issues with dual graph modelling that arise from not taking into account road segments and intersections simultaneously have been minimized. We ran thorough experiments using real-world datasets to evaluate the performance of our model. The results indicate that STDGNN achieves higher accuracy in estimating user behaviour

analysis Now that a layer has been added to STGNN, let's be able to enter all GPS points at any time. This layer operates on a dynamic node number at various times.
Several frequent travel trajectories may be identified when comparing this study to earlier ones that used standard methods (DBSCAN, BC-DBSCAN) with a cluster. However, these methods have some drawbacks. The experimental dataset in this research has a two-week duration, which may be too short to extract frequent travel patterns for some individuals due to low data coverage and traditional algorithms do not show results for all road intersections (Y. Tian et al. 2021). Also, Certain research uses CSRNet, a convolutional neural network approach, which has some limits, particularly in terms of how many clusters it can produce. It shows the prevalence of six behavioural groups with statistical tests verifying substantial variations connected to movement and time consumption (Abkarian et al. 2022).

7. REFERENCES

Abkarian, Hoseb, Divyakant Tahlyan, Hani Mahmassani, and Karen Smilowitz. 2022. “Characterizing Visitor Engagement Behavior at Large-Scale Events: Activity Sequence Clustering and Ranking Using GPS Tracking Data.” Tourism Management 88 (August 2021): 104421. https://doi.org/10.1016/j.tourman.2021.104421.

Alasadi, Tawfiq A., and R. Wadhah Baiee. 2016. “Design 3D-SVDs Algorithm for Location Based Recommendation System.” Research Journal of Applied Sciences 11 (10): 1143–50. https://doi.org/10.3923/rjasci.2016.1143.1150.

Bonnetain, Loïc, Angelo Furno, Nour Eddin El Faouzi, Marco Fiore, Razvan Stanica, Zbigniew Smoreda, and Cezary Ziemlicki. 2021. “TRANSIT: Fine-Grained Human Mobility Trajectory Inference at Scale with Mobile Network Signaling Data.” Transportation Research Part C: Emerging Technologies 130 (September): 103257. https://doi.org/10.1016/J.TRC.2021.103257.

Buchin, Maike, Somayeh Dodge, and Bettina Speckmann. 2014. “Similarity of Trajectories Taking into Account Geographic Context.” Journal of Spatial Information Science 9 (9): 101–24. https://doi.org/10.5311/JOSIS.2014.9.179.

Chaker, Rima, Zaher Al Aghbari, and Imran N. Junejo. 2017. “Social Network Model for Crowd Anomaly Detection and Localization.” Pattern Recognition 61 (January): 266–81. https://doi.org/10.1016/j.patcog.2016.06.016.

Chen, Tianran, Yongzheng Zhang, Yupeng Tuo, and Weiguang Wang. 2020. “CgNet:Predicting Urban Congregations from Spatio-Temporal Data Using Deep Neural

Networks.” 2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings, December. https://doi.org/10.1109/GLOBECOM42002.2020.9322280.

Chen, Weiqi, Ling Chen, Yu Xie, Wei Cao, Yusong Gao, and Xiaojie Feng. 2020. “Multi-Range Attentive Bicomponent Graph Convolutional Network for Traffic Forecasting.” Proceedings of the AAAI Conference on Artificial Intelligence 34 (04): 3529–36. https://doi.org/10.1609/AAAI.V34I04.5758.

Dutta, Sumanto, Animesh Das, and Bidyut Kr Patra. 2022. “CLUSTMOSA: Clustering for GPS Trajectory Data Based on Multi-Objective Simulated Annealing to Develop Mobility Application.” Applied Soft Computing 130 (November): 109655. https://doi.org/10.1016/j.asoc.2022.109655.

Fang, Xiaomin, Jizhou Huang, Fan Wang, Lingke Zeng, Haijin Liang, and Haifeng Wang. 2020. “ConSTGAT: Contextual Spatial-Temporal Graph Attention Network for Travel Time Estimation at Baidu Maps.” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August, 2697–2705. https://doi.org/10.1145/3394486.3403320.

Gu, Jianwei, Mike Pitz, Susanne Breitner, Wolfram Birmili, Stephanie von Klot, Alexandra Schneider, Jens Soentgen, Armin Reller, Annette Peters, and Josef Cyrys. 2012. “Selection of Key Ambient Particulate Variables for Epidemiological Studies — Applying Cluster and Heatmap Analyses as Tools for Data Reduction.” Science of The Total Environment 435–436 (October): 541–50. https://doi.org/10.1016/J.SCITOTENV.2012.07.040.

Hou, Mingliang, Feng Xia, Haoran Gao, Xin Chen, and Honglong Chen. 2022. “Urban Region Profiling with Spatio-Temporal Graph Neural Networks.” IEEE Transactions on Computational Social Systems 9 (6): 1736–47. https://doi.org/10.1109/TCSS.2022.3183570.

Huang, Weixin, and Luying Wang. 2022. “Towards Big Data Behavioral Analysis: Rethinking GPS Trajectory Mining Approaches from Geographic, Semantic, and Quantitative Perspectives.” Architectural Intelligence 1 (1): 1–15. https://doi.org/10.1007/s44223-022-00011-y.

Hwang, Jung Rae, Hye Young Kang, and Ki Joune Li. 2005. “Spatio-Temporal Similarity Analysis between Trajectories on Road Networks.” Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3770 LNCS: 280–89. https://doi.org/10.1007/11568346_30.

Jin, Guangyin, Hengyu Sha, Yanghe Feng, Qing Cheng, and Jincai Huang. 2021. “GSEN: An Ensemble Deep Learning Benchmark Model for Urban Hotspots Spatiotemporal Prediction.” Neurocomputing 455: 353–67. https://doi.org/10.1016/j.neucom.2021.05.008.

Jin, Guangyin, Huan Yan, Fuxian Li, Jincai Huang, and Yong Li. 2021. “Spatio-Temporal Dual Graph Neural Networks for Travel Time Estimation,” 1–12. http://arxiv.org/abs/2105.13591.

Liu, C. K., K. Qin, K. Chen, and R. Ma. 2020. “UNCOVERING THE AGGREGATION PATTERN OF GPS TRAJECTORY BASED ON SPATIOTEMPORAL CLUSTERING AND 3D VISUALIZATION.” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3-W10 (3/W10): 255–60. https://doi.org/10.5194/ISPRS-ARCHIVES-XLII-3-W10-255-2020.

Maria Bianchi, Filippo, Daniele Grattarola, and Cesare Alippi. 2020. “Spectral Clustering with Graph Neural Networks for Graph Pooling.”

“Microsoft Research – Emerging Technology, Computer, and Software Research.” n.d. Accessed January 30, 2024. https://www.microsoft.com/en-us/research/?from=https%3A%2F%2Fresearch.microsoft.com%2Fapps%2Fpubs%2F%3Fid%3D131038&type=no-match.

Min, Shengjie, Zhan Gao, Jing Peng, Liang Wang, Ke Qin, and Bo Fang. 2021. “STGSN — A Spatial–Temporal Graph Neural Network Framework for Time-Evolving Social Networks.” Knowledge-Based Systems 214: 106746. https://doi.org/10.1016/j.knosys.2021.106746.

Moreira-Matias, Luís, João Gama, Michel Ferreira, João Mendes-Moreira, and Luis Damas. 2016. “Time-Evolving O-D Matrix Estimation Using High-Speed GPS Data Streams.” Expert Systems with Applications 44 (February): 275–88. https://doi.org/10.1016/J.ESWA.2015.08.048.

Pancasila, Irwan, Siswoyo Haryono, and Beni Agus Sulistyo. 2020. “Effects of Work Motivation and Leadership toward Work Satisfaction and Employee Performance: Evidence from Indonesia.” Journal of Asian Finance, Economics and Business 7 (6): 387–97. https://doi.org/10.13106/jafeb.2020.vol7.no6.387.

Tian, Fei, Bin Gao, Qing Cui, Enhong Chen, and Tie Yan Liu. 2014. “Learning Deep Representations for Graph Clustering.” Proceedings of the National Conference on Artificial Intelligence 2: 1293–99. https://doi.org/10.1609/aaai.v28i1.8916.

Tian, Yuan, Xinming Zhang, Binyu Yang, Jian Wang, and Shi An. 2021. “An Individual-Based Spatio-Temporal Travel Demand Mining Method and Its Application in Improving Rebalancing for Free-Floating Bike-Sharing System.” Advanced Engineering Informatics 50 (September 2020): 101365. https://doi.org/10.1016/j.aei.2021.101365.

Tiwari, Rajdev. 2013. "Correlation-Based Attribute Selection Using Genetic Algorithm Correlation-Based Attribute Selection Using Genetic Algorithm," no. December. https://doi.org/10.5120/847-1182.

Wang, Senzhang, Jiannong Cao, and Philip S. Yu. 2022. "Deep Learning for Spatio-Temporal Data Mining: A Survey." IEEE Transactions on Knowledge and Data Engineering 34 (8): 3681–3700. https://doi.org/10.1109/TKDE.2020.3025580.

Yang, Haoran, Junli Wang, Rui Duan, and Chungang Yan. 2023. "DCOM-GNN: A Deep Clustering Optimization Method for Graph Neural Networks." Knowledge-Based Systems 279 (November): 110961. https://doi.org/10.1016/J.KNOSYS.2023.110961.

Young, Rebekah, and David R. Johnson. 2015. "Handling Missing Values in Longitudinal Panel Data With Multiple Imputation." Journal of Marriage and Family 77 (1): 277–94. https://doi.org/10.1111/JOMF.12144.

Yu, Chang, and Zhao Cheng He. 2017. "Analysing the Spatial-Temporal Characteristics of Bus Travel Demand Using the Heat Map." Journal of Transport Geography 58: 247–55. https://doi.org/10.1016/j.jtrangeo.2016.11.009.

Yu, Qingying, Yonglong Luo, Chuanming Chen, and Shigang Chen. 2019. "Trajectory Similarity Clustering Based on Multi-Feature Distance Measurement." Applied Intelligence 49 (6): 2315–38. https://doi.org/10.1007/S10489-018-1385-X/METRICS.

Zhang, Dongzhi, Kyungmi Lee, and Ickjai Lee. 2018. "Hierarchical Trajectory Clustering for Spatio-Temporal Periodic Pattern Mining." Expert Systems with Applications 92 (February): 1–11. https://doi.org/10.1016/J.ESWA.2017.09.040.

Zhao, Shengjie, Kai Zhao, Yusen Xia, and Wenzhen Jia. 2022. "Hyper-Clustering Enhanced Spatio-Temporal Deep Learning for Traffic and Demand Prediction in Bike-Sharing Systems." Information Sciences 612 (October): 626–37. https://doi.org/10.1016/J.INS.2022.07.054.