

Article

Scientific Production on GPS Trajectory Clustering: A Bibliometric Analysis

Gary Reyes^{1,2,*}, Roberto Tolozano-Benites^{1,†}, Laura Lanzarini^{3,†}, César Estrebou^{3,†} and Aurelio F. Bariviera^{4,†}

¹ Carrera de Sistemas Inteligentes, Universidad Bolivariana del Ecuador, Campus Durán Km 5.5 vía Durán Yaguachi, Durán 092405, Ecuador; rtolozano@ube.edu.ec

² Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil 090514, Ecuador

³ Instituto de Investigación en Informática LIDI (Centro CICPBA), Facultad de Informática, Universidad Nacional de La Plata, Buenos Aires CP1900, Argentina; laural@lidi.info.unlp.edu.ar (L.L.); cesarest@lidi.info.unlp.edu.ar (C.E.)

⁴ Department of Business & ECO-SOS, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain

* Correspondence: gxyreysz@ube.edu.ec

† These authors contributed equally to this work.

Abstract: Clustering algorithms or methods for GPS trajectories are in constant evolution due to the interest aroused in part of the scientific community. With the development of clustering algorithms considered traditional, improvements to these algorithms and even unique methods considered as “novel” for science have emerged. This work aimed to analyze the scientific production that exists around the topic “GPS trajectories clustering” by means of bibliometrics. Therefore, a total of 559 articles from the main collection of Scopus were analyzed, initially filtering the generated sample to discard any articles that did not have a direct relationship with the topic to be analyzed. This analysis establishes an ideal environment for other disciplines and researchers since it provides a current state of the trend of the subject of study in their field of research.

Keywords: trajectory clustering; GPS trajectories; trajectory clustering algorithms; bibliometrics; bibliometry



Academic Editors: Wolfgang Kainz and Hartwig H. Hochmair

Received: 12 February 2025

Revised: 2 April 2025

Accepted: 6 April 2025

Published: 10 April 2025

Citation: Reyes, G.;

Tolozano-Benites, R.; Lanzarini, L.; Estrebou, C.; Bariviera, A.F. Scientific Production on GPS Trajectory Clustering: A Bibliometric Analysis. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 165. <https://doi.org/10.3390/ijgi14040165>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. 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/>).

1. Introduction

Vehicle trajectory analysis has become essential to addressing mobility problems in complex urban environments, where traffic and congestion present increasing challenges. The implementation of artificial intelligence and data mining in this field makes it possible to identify travel patterns from large volumes of data, facilitating the understanding of traffic flows and their relationship with road infrastructure [1,2]. This type of analysis supports traffic planners and managers in making data-driven decisions, helping to reduce congestion and optimize the use of road networks [3].

In addition, intelligent transportation systems, which employ advanced spatial analysis tools, enable real-time traffic monitoring, detecting critical points and assessing safety conditions in various areas of the city [4]. This constant monitoring enables a timely response to road infrastructure problems while improving road safety and maintenance [5]. In turn, spatial data processing techniques have applications that go beyond transportation, providing value in areas such as massive data analysis and the study of consumer behavior, extending the impact of these advances to various sectors [6,7].

The analysis of scientific production in emerging areas, such as intelligent transportation systems and the study of vehicle trajectories, provides a better understanding of the trends and impact of these fields in science and technology. Bibliometric studies provide a valuable framework for observing how research into these topics has evolved, revealing patterns of collaboration, citations, and relevance that reflect their growing importance in the scientific community.

Bibliometric study is a discipline that has had an important growth within the scientific community in recent years. Eugene Garfield, with the establishment of the Institute for Scientific Information (ISI) in the 1960s, initiated the measurement of articles, journals, researchers, and institutions [8]. Bibliometric research examines authorship, publication, citations, and content by applying quantitative measures to a body (corpus) of literature [9]. Currently, scientific articles are stored and indexed in large scientific databases, allowing one to measure the parameters they have, such as their keywords, numbers of citations, numbers of authors, author collaboration and impact, and annual scientific production, among others. The main idea is that having more citations in a scientific field indicates greater importance and quality and is more remarkable [10]. The reason for indexing articles is given by the following: authors cite other papers for their central idea, which is due to the connection they have with the central theme of their research or work. Since any author can select which article to cite, including only the most relevant and related to their own article, most of the articles that are cited could demonstrate the impact or importance they have had within their scientific field. The information that can be obtained can be leveraged by various institutions, as valuable information on both individual and aggregate impact is given. Therefore, it could help in the recruitment of teachers or in devising research strategies in universities and research councils. However, bibliometric studies can also help with information about the history of a certain topic, in addition to publicizing the scope of, or trend that led to, that research topic. This helps new researchers to have an idea of the impact that a research topic has on their scientific field [11]. This type of analysis is made possible through the availability of large bibliographic databases such as Scopus or Web of Science, among others. These indexing services are an important means for the evaluation process in academia.

Within this area of study, bibliometric analysis has been applied to various disciplines to identify trends, thematic evolution, or collaborative networks. However, the specific field of GPS trajectory clustering has received little attention from a bibliometric perspective. Some studies have addressed methodologies for processing vehicle trajectories and clustering them using different algorithms [12], but few have analyzed the scientific production in this field from a quantitative approach. This lack of systematic studies limits the understanding of the impact and development of research in this field.

Despite the growing interest in GPS trajectory clustering [13], there have been no bibliometric studies in recent years that have systematically analyzed its scientific output. Most previous studies have focused on the development of methodologies and algorithms without examining their impact on the scientific community. This study fills this gap by providing a quantitative view on the evolution of the field, as well as its main contributors and emerging trends, which will allow researchers and decision-makers to better understand its development and future projection.

Scopus is a bibliographic database that collects citations and abstracts from a wide variety of neutral sources. These resources are carefully selected by independent experts who are recognized leaders in their respective disciplinary fields. Scopus offers researchers a range of discovery and analysis tools. This platform not only facilitates the search and retrieval of relevant information, but also promotes collaboration and the exchange of ideas between individuals and institutions in the scientific community. With a broad scope,

Scopus indexes content from more than 7000 publishers, covering a diversity of disciplines. In addition, it hosts a vast data collection, with more than 91 million records, including more than 94,000 affiliation profiles and the contribution of more than 17 million authors.

From a macroscopic level, metrics can be determined that are common to many journals and are useful for different stakeholders. However, some characteristics change from one context or discipline to another. There are a number of researchers and journals that perform unevenly. In recent years, there has been an expansion in the number of journals and an increase in the periods in which they are published. This may be because of the expansion of the academic sector in several countries, which has increased gradually in the last decade in various countries. In addition, scientific disciplines have different parameters regarding the publication of an article. Therefore, it is important to study their characteristics and/or equivalent topics in order to provide a meaningful classification for bibliometric parameters.

The objective of this paper is to analyze the metadata of all articles indexed in the Scopus bibliographic database that performed “GPS trajectory clustering” using algorithms or methods. Unlike other studies, this research offers a comprehensive view of the evolution of scientific production in the area, identifying the main journals, the most influential authors, and the emerging thematic trends related to GPS trajectory clustering. It should also be noted that the samples generated by the bibliographic database were manually filtered to exclude all articles that were not part of the field of study. This article provides useful information on the main journals that have published articles on this particular topic, as well as the evolution of the scientific field over time. In addition, other aspects are discussed, such as the most cited authors, the areas in which these articles are most published, the number of publications per year, strategic diagrams on the impact of the topics, and the thematic evolution, among others.

The bibliometric analysis was performed graphically by the VOSviewer software, which is a software tool for creating maps based on network data and then visualizing and exploring these maps [14], including graphs of citations, sources, and authors. In addition, we made use of the bibliometrix package and its graphical interface, biblioshiny, based on the R programming language, which was developed by Aria and Cuccurullo [15] to perform analyses on the graphical distribution of the corresponding author, the most cited articles, the main keywords, the main publication sources, the strategic diagrams of the keywords, and the thematic evolution of the keywords. Both software tools are open source, which allows researchers to use all of their functionalities, such as determining the most cited article and co-authorship, among others.

The remainder of this paper is structured as follows. Section 2 describes the materials and methods used in the analysis methodology. Section 3 details the data under analysis, as well as the main findings of the study by means of bibliometrix and its graphical interface biblioshiny. In addition, the analysis of the selected indicators using VOSviewer is described. Finally, Section 4 presents the main conclusions and explains the possible lines of research that can be derived from the analysis.

2. Materials and Methods

To analyze scientific production in “GPS trajectory clustering”, a detailed methodological approach was developed that integrated bibliometric analysis and specialized data visualization tools. To this end, articles indexed exclusively in the Scopus database that address this topic were analyzed using specific algorithms or methods.

First, we filtered the metadata of all articles indexed in Scopus that addressed “GPS trajectory clustering”. For the collection of these articles, only publications from 2002 to 2023 were considered. A rigorous filtering was then applied to ensure the representativeness of

the dataset, and papers that were not directly linked to the field of study were manually discarded. This included selecting articles based on keywords such as “trajectory clustering”, “GPS trajectories”, “clustering methods”, and related terms. As a result, 559 articles were extracted, including general studies and specific developments in GPS trajectory clustering algorithms.

The visualization and analysis of bibliographic networks were performed using VOSviewer [14], a specialized tool that allows for mapping networks of co-citation, author collaboration, and thematic distribution. This software facilitated the creation of structural graphs showing the relationships between the most relevant articles, journals, and authors in the field of study. Likewise, the analysis was complemented with the use of bibliometrix and its interface biblioshiny, developed in R language, which enabled the evaluation of keywords, thematic evolution, and the identification of emerging trends. Both programs, freely available and open source, offer the advantage of replicability and allow other researchers to extend the application of this method to similar studies.

To quantify the concentration of variables, such as the distribution of authors, countries, and research areas, Shannon entropy was used to evaluate the homogeneity of data dispersion. The values obtained were used to interpret the patterns of concentration in authorship and international collaborations, as well as the diversification of topics in the literature on GPS trajectories.

3. Results

We worked with the bibliographic metadata (bibliographic metadata) of the articles that were indexed in the bibliographic database Scopus. Therefore, only the articles that developed or investigated the “GPS trajectories clustering” were selected. Scopus hosted a total of 559 sample papers published in 333 sources (journals, books, etc.) during the period of 2002–2023. These papers were (co-)authored by 1416 individuals, and the vast majority of papers were multi-authored. However, only 11 papers were single-authored. The average number of authors per document was 3.87. Upon analysis, it was observed that they were concentrated in two main research areas: computer science and engineering. Scopus assigns indexed articles to one or more research areas. The 559 articles in the sample were assigned to various research areas, giving a total of 1094, i.e., they belonged to more than one specific area. The five main research areas are shown in Table 1.

Table 1. Main areas of research assigned to the sample papers. Source: Scopus.

Research Areas	Records	% of 1094
Computer science	391	35.74%
Engineering	176	16.09%
Social sciences	125	11.43%
Mathematics	123	11.24%
Earth and planetary sciences	69	6.31%
Total of the 5 main research areas	884	80.80%

The results show that research on GPS trajectory clustering is mainly concentrated in computer science and engineering, reflecting a focus on the development of models and algorithms for spatial data processing. However, the presence of areas such as social sciences, mathematics, and earth sciences suggests that the topic also attracts interest in disciplines that address broader applications, from mobility and transportation analysis to geospatial studies and mathematical modeling. The difference in the number of publications between computer science and the rest of the areas could indicate that there is still room to explore

interdisciplinary approaches, integrating social or environmental perspectives that allow for a more complete understanding of the impact and applications of clustering trajectories.

The details of the annual publications of articles are shown in Table 2. For Scopus, it is observed that the early years had the reception of few articles related to GPS trajectory clustering, although in the last decade the number of articles published increased, possibly due to the reception of the scientific community. The total number of records in the sample has an average annual growth rate of 15.6% from 2002 to 2023.

Table 2. Number of articles published per year. Source: Scopus.

Years	Items	Annual Growth Rate
2002	2	-
2003	2	0.0%
2004	1	-50.00%
2005	0	-100.00%
2006	0	-
2007	1	-
2008	2	100.00%
2009	11	450.00%
2010	12	9.09%
2011	12	0.00%
2012	15	25.00%
2013	21	40.00%
2014	30	42.86%
2015	31	3.33%
2016	37	19.35%
2017	48	29.73%
2018	55	14.58%
2019	68	23.64%
2020	57	-16.18%
2021	52	-8.77%
2022	60	15.38%
2023	42	-30.00%
Total	559	15.6%

The sustained growth in the publication of articles on GPS trajectory clustering in the last two decades reflects the growing interest of the scientific community in this line of research. Although in the early years the production was limited, since 2009, there has been a steady increase, possibly driven by advances in spatial data processing and access to large volumes of georeferenced information. The average annual growth rate of 15.6% shows that the topic is becoming increasingly strong in the academic world, although variations in recent years could be due to changes in research trends, the emergence of new methodologies, or data availability. This behavior indicates that, although research in the area has gained relevance, it is still an evolving field, with opportunities to expand its impact in different scientific and technological applications.

3.1. Geographical Distribution of the Corresponding Authors

Table 3 shows China as the top country whose authors have published the most papers, followed by the USA as the second country with the second most published papers. The top ten countries accumulate 53.9% of the published papers related to “GPS trajectory clustering”. The acronyms SCP, MCP, and MCP Ratio correspond to “Single Country Publications”, “Multi-Country Publications”, and “Proportion of Multi-Country Publications”. Table 4 shows the top countries ordered by the total number of citations. The average number of citations for all articles is 21.92. China and USA, the two countries

with the most published articles and total citations, are above this figure, with an average of 19.30 and 34.60, respectively. Although China is the first country in terms of published articles, it has the second lowest average number of citations per article among the leading countries. It is also important to note that the USA is the country with the highest average number of citations per article, which can be used as a common denominator in the average scientific importance or quality of the articles. The countries that collaborate the least internationally with other countries are the Netherlands and Thailand, with a publication rate of 0.0%. The country that collaborates the most internationally with other countries is China, where 30.80% of the papers are of this type.

Table 3. Ten countries of corresponding authors. Source: Scopus.

Country	Articles	Frequency	SCP	MCP	MCP Ratio
China	203	36.3%	159	44	21.7%
USA	29	5.2%	17	12	41.4%
India	16	2.9%	14	2	12.5%
Italy	13	2.3%	12	1	7.7%
Korea	11	2.0%	8	3	27.3%
Portugal	8	1.4%	4	4	50.0%
Japan	6	1.1%	5	1	16.7%
Australia	5	0.9%	3	2	40.0%
France	5	0.9%	2	3	60.0%
Germany	5	0.9%	3	2	40.0%
Total 10 countries	301	53.9%	227	74	31.7%

Table 4. Top ten total citations by country. Source: Scopus.

Country	Total Citations	Average Citations of Articles
China	3908	19.30
USA	1004	34.60
Turkey	400	400.00
Italy	202	15.50
Hong Kong	173	34.60
Switzerland	173	34.60
Greece	167	33.40
Spain	156	39.00
France	127	25.40
Australia	118	23.60
Total (all countries)	7138	21.92

The data show that China and the US lead the scientific production on the topic of GPS trajectory clustering, with China leading in the number of publications and the US having the highest average number of citations per article. Although China contributes a large number of studies, the influence and visibility of articles published in the USA may be greater, possibly due to differences in perceived quality, access to high-impact journals, or the scientific collaboration network. In addition, the proportion of international publications varies significantly between countries, with China standing out for its high level of collaboration, while others, such as the Netherlands and Thailand, do not publish in this area. These differences could be due to national research strategies, the availability of funding, or the degree of internationalization of their scientific communities. The fact that more than half of the articles are concentrated in only ten countries is evidence that production in this area is not yet widely distributed globally, so opportunities could be taken to strengthen research in other regions through international collaboration networks.

3.2. Main Publication Sources

Table 5 shows the top ten sources that publish articles related to “path clustering algorithms”. The top three are *Lecture Notes in Computer Science* (LNCS) (including its subseries *Lecture Notes in Artificial Intelligence*, LNAI, and *Lecture Notes in Bioinformatics*, LNBI), which is a series of conference proceedings that publishes the latest research advances in all areas of computer science; *ISPRS International Journal of Geo-Information*, which is an international peer-reviewed open access journal on geo-information; *IEEE Access*, which is a leading multidisciplinary open access journal. *GIS: Proceedings of the ACM International Symposium On Advances In Geographic Information Systems* is from ACM SIGSPATIAL International Conferences on advances in interdisciplinary research in all aspects of geographic information systems. *ACM International Conference Proceeding Series* are a series of International Conference Proceedings (ICPS) that provide a mechanism for publishing the contents of high-quality conferences, technical symposia, and workshops. The *International Journal of Geographical Information Science* is a peer-reviewed journal that publishes topics related to fundamental and computational geographic information science, among others. *Cluster Computing—The Journal of Networks Software Tools and Applications* is a peer-reviewed scientific journal on parallel processing, distributed computing systems, and computer communication networks. *IEEE Transactions on Intelligent Transportation Systems* is a journal that is published through *IEEE Access*; among the scope of the topics published are communications (inter-vehicle and vehicle-to-road), computers (hardware, software), and information systems (databases, data fusion, security), among others. *International Archives Of The Photogrammetry, Remote Sensing And Spatial Information Sciences - Isprs Archives* is a series of peer-reviewed proceedings published by the International Society for Photogrammetry and Remote Sensing (ISPRS). The scientific journal *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/ Journal of Transportation Systems Engineering and Information Technology* is included in the Scopus database. Its main subject areas of published articles are computer science applications, systems and control engineering, modeling and simulation, and transportation. Finally, *Transactions in GIS* is an international peer-reviewed journal that publishes original research articles, review articles, and short technical notes on the latest advances and best practices in spatial sciences.

Table 5. The ten most relevant sources. Source: Scopus.

Sources	Articles	Type
<i>Lecture Notes in Computer Science</i> (including the subseries <i>Lecture Notes in Artificial Intelligence</i> and <i>Lecture Notes in Bioinformatics</i>)	38	Conference Proceedings
<i>ISPRS International Journal of Geo-Information</i>	17	Journal
<i>IEEE Access</i>	15	Journal
<i>GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems</i>	13	Conference Proceedings
<i>ACM International Conference Proceeding Series</i>	12	Conference Proceedings
<i>International Journal of Geographical Information Science</i>	11	Journal
<i>IEEE Transactions on Intelligent Transportation Systems</i>	8	Journal
<i>International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives</i>	8	Journal
<i>Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology</i>	7	Journal
<i>Transactions in GIS</i>	7	Journal

When analyzing the main sources of publication in the field of GPS trajectory clustering, there is a clear predominance of conference proceedings over scientific journals. This finding points to the fact that the dissemination of advances in this field takes place

mainly in academic events, where researchers present recent and developing results. In particular, the *Lecture Notes in Computer Science* series leads in the number of publications, and this is indicative of the close linkage of this subject with computer science and artificial intelligence. However, the presence of specialized publications in geoinformation, intelligent transportation systems, and distributed computing is evident, indicating a multidisciplinary approach. The combination of multiple publication sources evidences that GPS trajectory clustering is an evolving area of research, with a balance between the dissemination of preliminary work at conferences and the consolidation of findings in peer-reviewed journals.

3.3. Most Cited Articles

Table 6 shows the list of the top 10 papers categorized as highly cited papers in Scopus. According to González-Betancor and Dorta-González [16], the most highly cited papers are those that have received a number of citations equal to or greater than the q th percentile for their field and year of publication. A highly cited paper is recognized as possessing scientific excellence, setting the foundation for the field in which its context in the world is focused. Therefore, they serve to highlight important articles in different fields. These articles become avenues for research. The first most cited article was proposed by Yuan et al. [12] who designed a variance–entropy-based clustering approach to estimate the distribution of travel time between two reference points in different time slots. Abul et al. [17] proposed a novel concept of k -anonymity based on co-location that exploits the inherent uncertainty of a moving object’s whereabouts. Jing Yuan et al. [18] designed a variance–entropy-based clustering approach to estimate the travel time distribution between two landmarks at different time intervals. Tang et al. [19] used an observed matrix of the central area in Harbin city to model traffic distribution patterns based on the entropy maximization method, and the estimation performance verified its effectiveness. Schroedl et al. [20] presented an approach to induce high-accuracy maps from vehicle traces equipped with differential GPS receivers. Guo et al. [21] presented a new methodology for detecting the location of spatial patterns and structures embedded in the origin–destination of motions. Abul et al. [22] addressed the problem of anonymization of moving object databases and proposed the novel concept of co-location-based k -anonymization, which exploits the inherent uncertainty of the whereabouts of moving objects. Li et al. [23] proposed an incremental clustering framework for trajectories that contains two parts: online microcluster maintenance and offline macrocluster creation. Chen et al. [24] proposed a probabilistic framework for inferring trip purposes; it has one phase that identifies activity areas and computes probabilities using Bayes’ theorem, while the second phase clusters delivery points and matches activity areas for real-time responses. Finally, Monreale et al. [25] presented a method that guarantees anonymity in trajectory data using a transformation based on spatial generalization and k -anonymity, providing formal data protection with a theoretical upper bound on re-identification.

The most cited research demonstrates the consolidation of lines of research that have had a significant impact on the development of methodologies for mobility analysis. The high number of citations received by these works is indicative of the fact that they have served as a key reference in subsequent studies, either for their theoretical contribution or for the applicability of their approaches in real contexts. Moreover, the diversity of topics addressed, such as traffic modeling, data privacy, and improved map accuracy, highlights the interdisciplinarity of the field and its evolution towards increasingly sophisticated solutions. This phenomenon also indicates that trajectory research has not only advanced in methodological terms, but has also generated new questions and challenges that continue to be explored in recent studies.

Table 6. The ten most cited articles, arranged in descending order by number of citations. Source: Scopus.

Author (Year) and Title	Source	Citations
Yuan et al. (2010). T-drive: driving directions based on taxi trajectories [12].	<i>GIS: International Conference on Advances in Geographic Information Systems</i>	884
Abul et al. (2008). Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases [17].	2008 IEEE 24th International Conference on Data Engineering	400
Jing Yuan et al. (2013). T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence [18].	<i>IEEE Xplore</i>	348
Tang et al. (2015). Uncovering urban human mobility from large scale taxi GPS data [19].	<i>Physica A: Statistical Mechanics and its Applications</i>	232
Schroedl et al. (2004). Mining GPS Traces for Map Refinement [20].	<i>Data Mining and Knowledge Discovery</i>	197
Guo et al. (2012). Discovering Spatial Patterns in Origin-Destination Mobility Data [21].	<i>Transactions in GIS</i>	145
Abul et al. (2010). Anonymization of moving objects databases by clustering and perturbation [22].	<i>Information Systems</i>	144
Li et al. (2010). Incremental Clustering for Trajectories [23].	<i>Springer Berlin Heidelberg</i>	132
Chen et al. (2018). TripImputor: Real-Time Imputing Taxi Trip Purpose Leveraging Multi-Sourced Urban Data [24].	<i>IEEE Transactions on Intelligent Transportation Systems</i>	125
Monreale et al. (2010). Movement data anonymity through generalization [25].	<i>Transactions on Data Privacy</i>	125

Table 7 shows the most productive authors. The table was made from a manual search, since bibliometrix, when analyzing the parameter of the authors, was unable to differentiate between authors who had the same surname with the same initials of their other names. Therefore, the following results were obtained. In the first places are Wang Haoyu and Li Jinhong, with 16 and 15 published articles, respectively, followed by Li Xue and Liu Yinzh, with 12 articles, and finally, Li Qing, with 11 published articles.

The identification of the most productive authors allows us to recognize the main contributors in this field of study, as well as the institutions that have promoted greater scientific production. The fact that researchers are affiliated with certain universities is evidence of a concentration of scientific production in these universities, which could be related to specialized research clusters that promote advances in this area.

Table 7. Most productive authors. Source: Scopus.

Authors	Institution	Articles
Wang Haoyu	Yunnan University, Kunming, China	16
Li Jinhong	North China University of Technology, Beijing, China	15
Li Xue	Shandong University of Science and Technology, Qingdao, China	12
Liu Yizhi	Hunan University of Science and Technology, Xiangtan, China	12
Li Qing	Shandong University of Science and Technology, Qingdao, China	11

3.4. Main Keywords

Table 8 shows the ten most frequently used keywords in GPS trajectory clustering articles. Scopus provides two types of keywords: (a) Author Keywords, which are terms selected directly by the authors of the articles and reflect their perception of the key concepts of their works, and (b) Keywords-Plus, which are generated automatically from an algorithm that extracts frequent terms from the titles of the references cited in each

article, without intervention by the authors. Keyword-Plus allows for the identification of recurrent terms that may not have been explicitly mentioned in the authors' keywords, thus providing a broader view of the thematic connections in the analyzed literature. The two most frequent author keywords were "clustering" and "trajectory". The Keywords-Plus, in their first places, contained the words "trajectories" and "clustering algorithms", which are present in articles published by Reyes et al. [26]. It is observed that at least four of the main keywords in both types coincide, possibly because they encompass everything that has to do with trajectories and GPS data; in addition, they are used in the process of data mining.

Table 8. Main keywords. Source: Scopus.

Author Keywords	Articles	Keywords-Plus	Articles
clustering	67	trajectories	278
trajectory	34	clustering algorithms	169
trajectory clustering	27	global positioning system	119
gps	26	data mining	105
gps trajectory	24	taxicabs	88
dbscan	22	cluster analysis	75
data mining	21	roads and streets	70
gps data	21	gps trajectories	61
gps trajectories	20	trajectory clustering	61
big data	14	gps	57

3.5. Keyword Strategy Diagram

In the strategic diagram, it is possible to see which topics are emerging, are trending, are within, or have disappeared from a field of research by analyzing the keywords. When joint word analysis is used to map science, clusters of keywords (and their interconnections) are obtained. These clusters are considered themes. Each research topic obtained in this process is characterized by two parameters, "density" and "centrality", that are fundamental for assessing its degree of development and importance [27].

Density refers to the number of relationships between keywords within a specific topic, indicating how centralized and cohesive that topic is within a research field. A topic with a high density has a higher concentration of related keywords, indicating that it is a well-developed and consolidated area.

On the other hand, centrality measures the position of a topic within the network of interactions between keywords, thus reflecting how influential it is within the field. A topic with high centrality is considered key to the development of the discipline, since it is closely linked to other relevant topics.

These two parameters, together, make it possible to identify which topics are fundamental in the field of study, which are highly influential, and which are emerging as research trends.

The objective of this analysis is to identify which topics have been central to the field of GPS trajectory clustering, which have maintained their relevance over time, and which emerging trends might guide future research. The bibliometrix package through its biblioshiny interface allows for the creation of the thematic map or strategic diagram of keywords, titles, and abstracts. Given the interpretation of the strategic diagram of Cobo et al. [27], the diagram provided by bibliometrix is analyzed as follows:

1. The topics in the upper right quadrant are well developed and are important for the structuring of a research field. They are known as the driving themes of the specialty since they have a strong centrality and a high density. The location of the topics in this quadrant implies that they are externally related to concepts applicable to other topics that are closely related conceptually.

2. Topics in the upper left quadrant have well-developed internal linkages but unimportant external linkages and are, therefore, of marginal importance to the field. These topics are highly specialized and peripheral in nature.
3. Themes in the lower left quadrant are underdeveloped and marginal. The themes in this quadrant have a low density and low centrality, representing mainly emerging or missing themes.
4. The topics in the lower right quadrant are important for a research field but are not developed. So, this quadrant groups basic cross-cutting and general topics.

Figures 1 and 2 show the strategic diagrams pertaining to KeyWords Plus and Scopus author keywords. For Figure 1, the KeyWords Plus are shown; its upper right quadrant contains the topics “trajectories” and part of the topic “gps”, considered as a group of well-developed and important subtopics for the research field of “algorithms or methods for GPS trajectory clustering”. Its upper left quadrant partially contains the topics “gps” and “location”, i.e., they contain well-developed subtopics, although they are not of importance to the research field. Its lower left quadrant contains the keyword “trajectory data”, and the other half of the theme “gps” within these themes comprises sub-themes that are underdeveloped, not taken into account, emerging, or missing. Finally, the lower right quadrant has the “cluster analysis” theme, i.e., it contains important sub-themes, although they are not fully developed. The strategic diagram makes it possible to distinguish the consolidated and fundamental topics of the field of study, as well as those that have been losing relevance or have recently emerged. This provides an overview of the status and evolution of the research area.

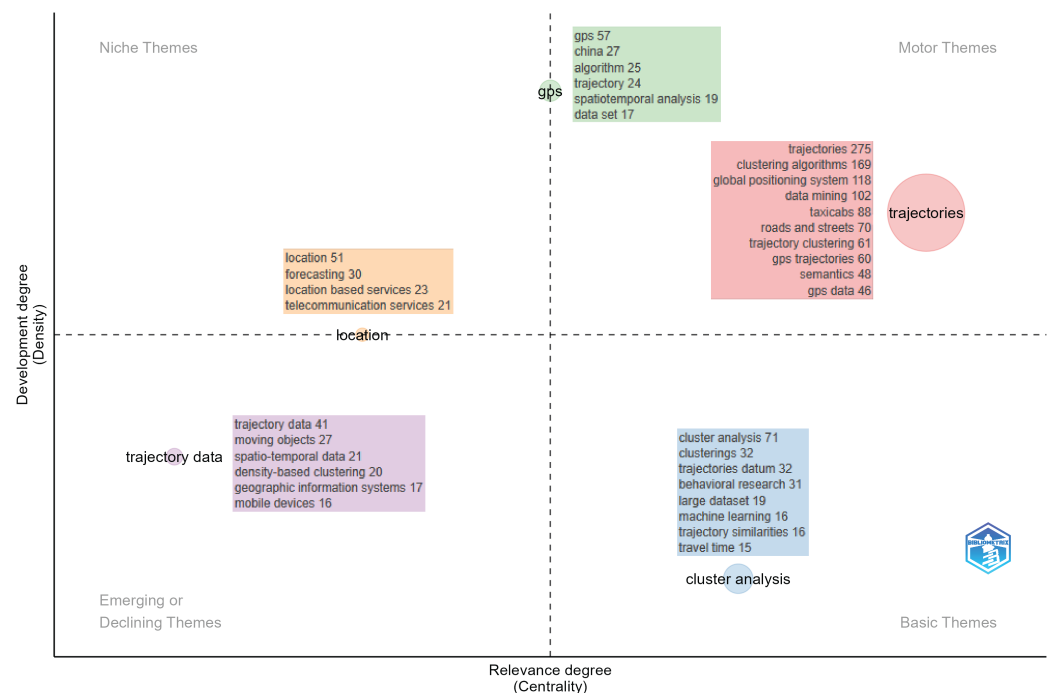


Figure 1. Strategic diagram of KeyWords Plus generated with bibliometrix. Source: Scopus.

In Figure 2, author keywords are shown. Its upper right quadrant contains part of the theme “mobility”, which is considered as a group of well-developed and important sub-themes for the research field of “algorithms or methods for GPS trajectory clustering”. Its upper left quadrant has the “vehicle trajectory” theme and partially the “mobility” and “urban computing” themes, i.e., they contain well-developed sub-themes, although they are not of importance to the research field. Its lower left quadrant contains the other half of the “urban computing” theme. Within this theme are sub-themes that are underdeveloped,

not taken into account, emerging, or missing. Finally, its lower right quadrant contains the “trajectory cluster” and “clustering” themes, i.e., they contain important sub-themes, although they are not fully developed.

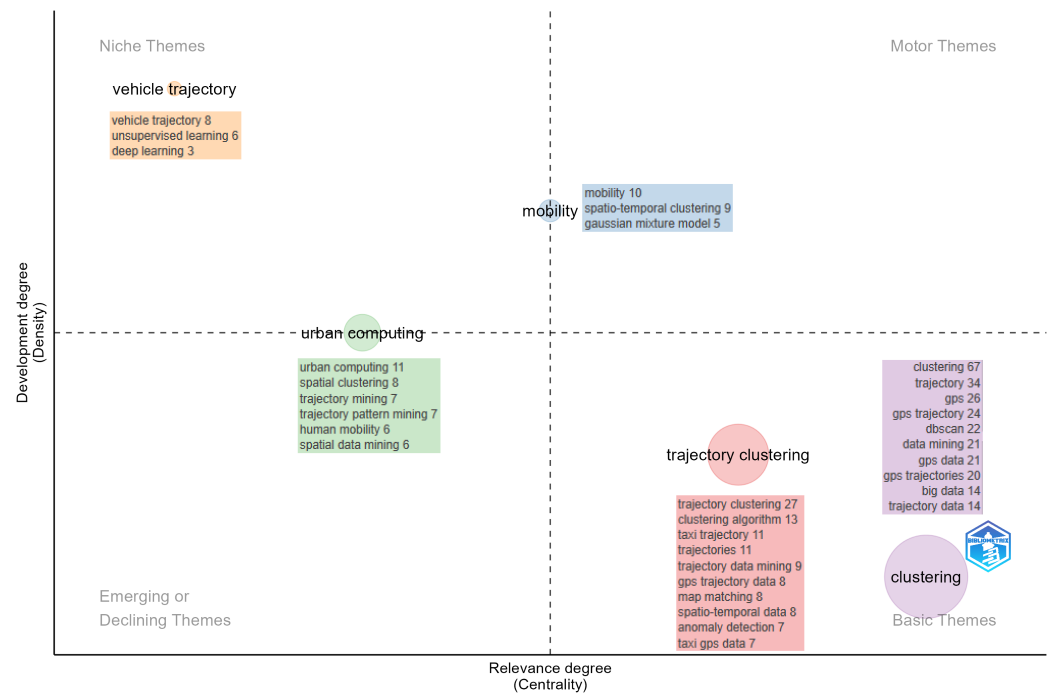


Figure 2. Strategic diagram of the authors’ keywords generated with bibliometrix. Source: Scopus.

Through these graphs, it is possible to understand how the thematic contributions have influenced the evolution of GPS trajectory clustering. The various studies have presented varied approaches such as the integration of data from multiple sensors, the use of artificial intelligence to improve accuracy, or the development of models based on deep learning, which has had a significant impact on the structuring of the topics. For example, the emergence of subtopics such as “trajectory clustering” and “spatio-temporal data” reflects the increasing application of advanced techniques for the analysis of large volumes of mobility data. This suggests that the evolution of the literature has been aligned with the development of new approaches, application of new computational techniques, and improvement in trajectory data processing capability.

3.6. Thematic Evolution of Keywords

For the thematic analysis of the evolution of the keywords, the R package bibliometrix (version 4.1.4) with its graphical interface biblioshiny was used, in which a range of years was established to observe the changes that exist between one thematic or another. Figures 3 and 4 show the thematic evolution of the authors’ keyWord Plus and keywords from the beginning of the studies in the research field to the present. In Figure 3, the theme trajectories is maintained, although it is integrated with some of the sub-themes that belong to data location, taxi cabs, and trajectory data. This forms a new group; however, the clustering theme prevails possibly because it maintains, in its entirety, the sub-themes that were present from 2002 to 2018. The themes “location”, “taxi cabs”, and “trajectory data” also became new themes that maintained certain sub-themes of the themes that existed before 2018. However, it is noted that “gps” has undergone minor changes in the subtopics that have been presented up to the present time.

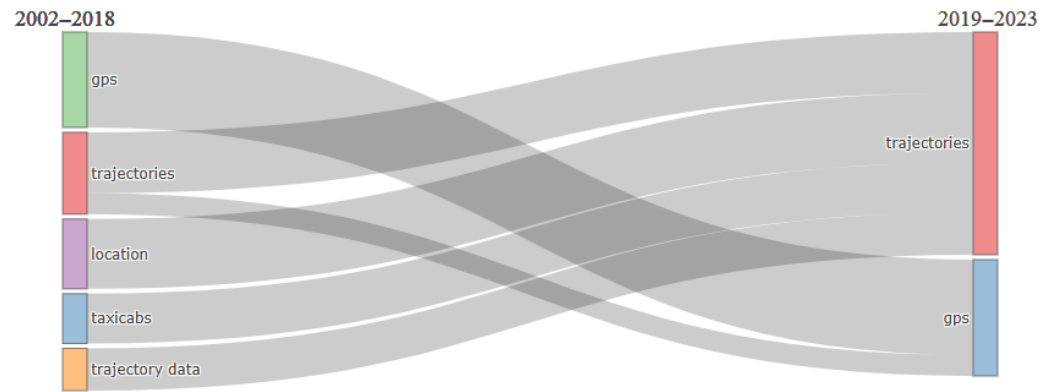


Figure 3. Strategic diagram of the authors' Keywords Plus generated with bibliometrix. Source: Scopus.

In Figure 4, the clustering theme is integrated with some of the sub-themes that belonged to trajectory, trajectory clustering, and big data. Likewise, the trajectory clustering theme is maintained, although some of its sub-themes became part of the clustering theme. Other topics, such as trajectory mining, are made up entirely of the subtopics that before 2018 belonged to the location prediction topic. At present, topics such as mobility, trajectories, and spatio-temporal data have emerged, whose subtopics have been derived from the clustering topic. Finally, it is observed that none of the current themes have retained the sub-themes in their entirety.

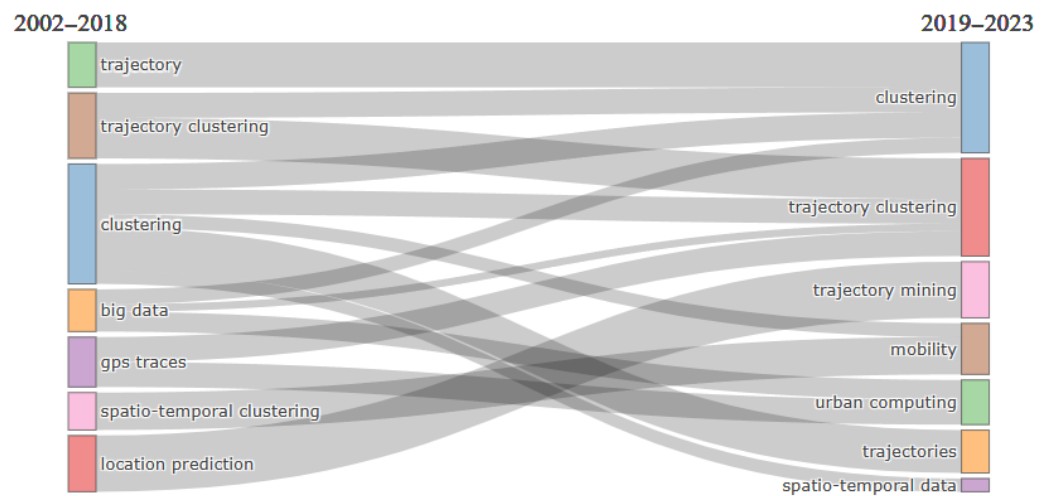


Figure 4. Strategic diagram of the authors' keywords generated with bibliometrix. Source: Scopus.

In both figures, it can be seen that there is a clear division into two periods (2002–2018 and 2019–2023), revealing a transition in methodological approaches and a specialization of the topics covered. During the 2002–2018 period, studies focused on general topics such as GPS, trajectories, and location, with initial data analysis methodologies. Starting in 2019, there was a shift to more specialized areas such as trajectory mining, urban computing, and big data, driven by technological and methodological advances. This shift reflects greater sophistication in trajectory analysis and the incorporation of more complex and specific urban mobility data.

The analysis of the evolution of keywords also reveals the emergence of new research foci in the field of GPS trajectory clustering. Topics such as “urban computing” and “mobility” have gained prominence in recent years, which seems to indicate a shift in focus from the development of purely mathematical algorithms towards the integration

of these methods in real applications, such as transportation planning and urban traffic management. In addition, the increased relevance of terms such as “trajectory mining” indicates a transition to more sophisticated approaches that leverage recent techniques such as machine learning to optimize trajectory analysis in large-scale, dynamic environments. These patterns suggest that the future of the field will be marked by the development of more efficient methodologies that allow data to be analyzed with greater accuracy and scalability.

These results highlight how the central topics in GPS trajectory clustering have evolved over time, showing the shift in research priorities. It is important to note that some emerging fields, such as the application of artificial intelligence in trajectory analysis, do not appear explicitly in the identified trends, despite their growing impact on the development of new clustering methods. The integration and transformation of sub-themes indicate the emergence of new lines of study, while the disappearance of some terms indicates a possible change in the focus of the scientific community.

The evolution of the topics covered in the literature on GPS trajectory clustering appears to be influenced by several factors. One is the increasing availability of positioning data, facilitated by the increased use of GPS-enabled devices and access to large volumes of mobility data. In addition, advances in data processing algorithms and machine learning techniques have led to the development of new approaches to trajectory analysis, resulting in the consolidation of certain lines of research and the emergence of new trends. These factors have redefined the field, directing research towards more efficient and accurate methods.

In recent years, the development of new methodologies has increasingly incorporated approaches based on artificial intelligence, which has improved the accuracy and flexibility of trajectory analysis. Although terms such as “deep learning” and “neural networks” do not appear very frequently in the identified thematic trends, their impact is reflected in the evolution of concepts such as “trajectory clustering” and “spatio-temporal data”. This suggests a transition towards more advanced models that integrate artificial intelligence techniques to optimize processes such as segmentation, noise filtering, and mobility pattern recognition, in line with the evolution of the field in the processing of large volumes of data.

3.7. Degree of Concentration of Selected Variables

In this subsection, some bibliometric variables are analyzed in order to show the degree to which they are concentrated. According to Stuart [28], bibliometric studies can be broadly classified as relational or evaluative, either providing information on the relationship between the units of analysis or assisting in the evaluation of the units of analysis. To perform this type of analysis, the information theory proposed by Shannon [29] is used. This theory provides different metrics that allow information to be obtained, such as standard deviation, skewness, and kurtosis. He also developed his own metric called the Shannon entropy. By means of a discrete probability distribution $P = \{p_j; j = 1, \dots, N\}$ with $\sum_{j=1}^N p_j = 1$, the Shannon entropy is defined as follows:

$$S[P] = - \sum_{j=1}^N p_j \ln(p_j) \quad (1)$$

The Shannon entropy can be interpreted or used in many ways in other scientific fields. Mejia-Barron et al. [30] made use of Shannon entropy and a fuzzy logic system to diagnose short-circuit faults. In another article, Babichev et al. [31] presented a gene expression profile reduction technology based on a complex use of fuzzy logic methods, statistical criteria, and Shannon entropy. On the other hand, Savakar and Hiremath [32] discussed

the detection of the falsification of an image using Shannon entropy and similarity and dissimilarity measures. Finally, it is used in bibliometric studies in order to study the equity/concentration distribution of different important variables such as research topics and authors, among others [33]. For a better interpretation of the information, Shannon entropy is used in its normalized form, dividing it by its maximum value. Therefore, the normalized concentration index is defined as follows:

$$H[P] = \frac{S[P]}{S_{MAX}} = \frac{-\sum_{j=1}^N p_j \ln(p_j)}{\ln N} \quad (2)$$

This is under the condition that $0 \leq H \leq 1$, where $H = 1$ means that all categories are uniformly represented, i.e., there is no concentration, and $H = 0$ means that the distribution is concentrated at a single point. The normalized entropic concentration index was calculated for the distribution of authors, sources, countries, research areas, and citations. The results are shown in Table 9, where it is observed that the authors are evenly distributed. The sources are also evenly distributed, as shown in Table 5. The countries publishing articles related to the topic of study are highly concentrated in a few countries, as shown in Table 3. However, taking into account the value of the index of authors and countries shows that the authors within these countries are evenly distributed. Similarly, in the research areas, a moderately low concentration is detected, as can be seen in Table 1, where 74.5% of publications are distributed among the areas of computer science, engineering, social sciences, and mathematics. Finally, the most cited articles are moderately concentrated, as can be seen in Table 6.

Table 9. Entropic concentration index (H) of the selected variables. Source: Scopus.

Variable	H
Authors	0.9665
Sources	0.9211
Countries	0.5375
Areas of research	0.6682
Article citations	0.8169

The results obtained show a remarkably homogeneous distribution both among authors and sources, suggesting that scientific production in this field is heterogeneous and is not dominated by a small cluster of authors. This situation could be interpreted as an indication of a field open to new contributions and collaborations. With regard to geographical distribution, although there is a concentration in certain nations, the analysis reveals a balanced distribution of authors within these regions. In terms of research areas, a moderate concentration is observed, indicating that, although the participation of several disciplines is evident, some key areas, such as computer science and engineering, exhibit greater production in this field. Finally, although the most cited articles show some concentration, this is not excessive, suggesting that there are multiple influential papers without centralization by a few.

An alternative measure to observe the distribution that authors follow according to their productivity is Lotka's law. According to the empirical finding made by Lotka [34], Lotka's law follows a form of Zipf's law. The original finding was based on a database restricted to physics and chemistry. Its equation based on this restriction is defined below:

$$a_n = \frac{a_1}{n^2}, n = 1, 2, \dots, N \quad (3)$$

where a_n is the number of authors publishing n articles, and a_1 is the number of authors publishing a single article. Lotka [34] derived his empirical law from a very specific sample; however, a generalization of his equation could be as follows:

$$a_n = \frac{a_1}{n^c}, n = 1, 2, \dots, N \quad (4)$$

where c is a parameter to be estimated to best fit the distribution data. The value of $c = 2.52$, with $R^2 = 0.96$. Table 10 summarizes the actual and fitted distributions of the number of authors publishing n articles. It is observed that the actual number of authors publishing only one article is lower than predicted by Lotka's law, confirming that authorship is not more widely and evenly distributed.

Table 10. Observed distribution of the number of authors who wrote a given number of articles and adjusted values of Lotka's law. Source: Scopus.

Number of Articles	Authors	Observed Frequency	Adjusted Frequency
1	1080	0.7627	0.7630
2	184	0.1299	0.1300
3	64	0.0452	0.0450
4	35	0.0247	0.0250
5	17	0.0120	0.0120
6	12	0.0085	0.0080
7	6	0.0042	0.0040
8	6	0.0042	0.0040
9	2	0.0014	0.0010
10	2	0.0014	0.0010
11	4	0.0028	0.0030
12	2	0.0014	0.0010
15	1	0.0007	0.0010
16	1	0.0007	0.0010

3.8. Charts of Citations, Sources, and Authors

The following figures were generated using the VOSviewer software tool (version 1.6.20) that allows for the creation of network-based maps, allowing for visualization and exploration. Developed by Van Eck and Waltman [14], it allows us to count the words that appear in the title, abstract, and keywords, obtaining the relationships that appear in the different documents that are published. Figure 5 represents the cloud map with the words that are relevant in the articles. The map shows how many times the words appear in the articles and how much of a relationship exists between them. The map is divided into groups: the blue part has a concentration of the word system, which, in turn, is related to the words analysis, research, technology, and evaluation. In the red part, there are words that are related to urban planning or urbanism, and among the words are study, cab, road, demand, and congestion. The green, yellow, and purple parts allude to concepts associated with the movement of objects and their different applications. The words study, city, analysis, system, and movement stand out because they create the links between the whole set of words, this has allowed for the detection of new perspectives of analysis towards emerging applications such as the one proposed by Reyes et al. [35].

Figure 6 is a representation almost similar to Figure 5, with the difference that words are counted in binary. This means that when a word appears, VOSviewer will only count it once regardless of the number of times it appears in the document. This slight difference can change the results that were obtained with the previous graphs because if a word is repeated a lot, it does not enter into the count of the final result. In the cloud map, it can be seen that the yellow parts of Figure 6 are merged with the words that have to

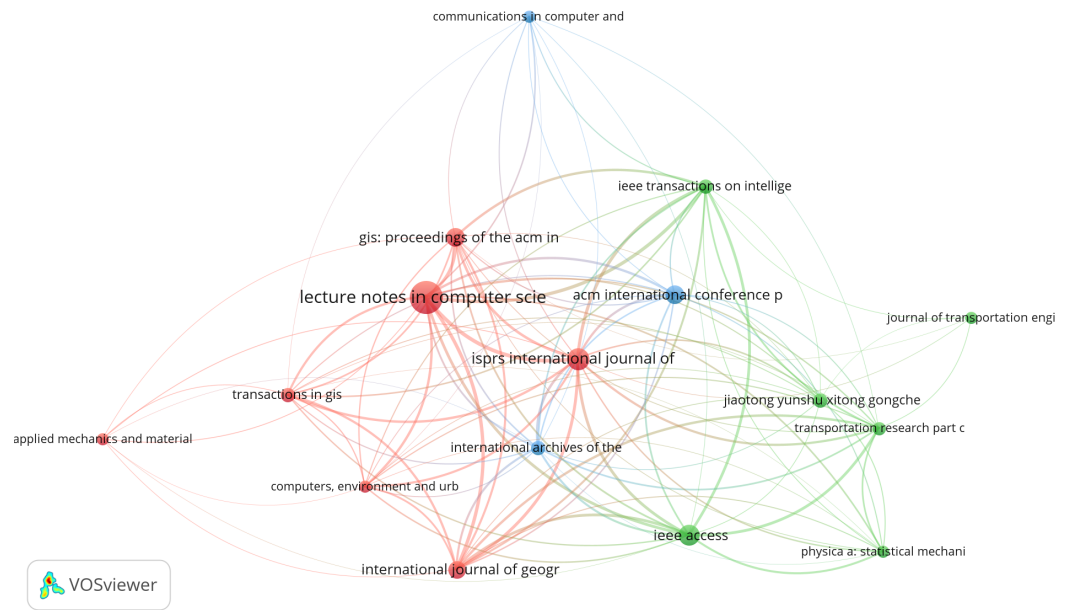


Figure 7. Cloud map of journals where articles on “GPS trajectory clustering” are published, generated with VOSviewer. Source: Scopus.

Figure 8 shows all the articles that belong to the sample, and the size of the node that is created depends on the number of citations they have. This result can be seen in Table 6 in the Most Cited Articles Subsection. In Figure 8, it can be seen that the two nodes that stand out the most are Jing Yuan et al. [18], published in *IEEE Xplore*, on the design of a variance–entropy-based clustering approach for the estimation of the distribution of travel times between two different points, and Schroedl et al. [20], published in *Data Mining and Knowledge Discovery*, that present an approach to induce high-precision maps from vehicle traces equipped with differential GPS receivers.

Other authors also stand out for their number of citations, such as Tang et al. [19], published in *Physica A: Statistical Mechanics and its Applications*, and Guo et al. [21], published in *Transactions in GIS*.

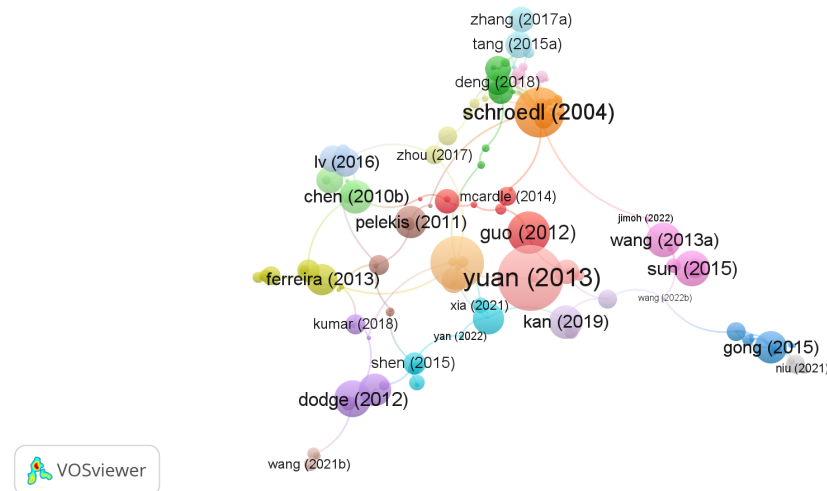


Figure 8. Cloud map created from authors with journal papers on “GPS trajectory clustering”, generated with VOSviewer. Source: Scopus.

4. Conclusions

This analysis shows that clustering for GPS trajectories comprises a combination of urban planning and the effects that vehicles have on streets, roads, or carts. It should be noted that this would not be possible without Global Positioning Systems or GPS, in addition to the integration of correct trajectory clustering algorithms such as TraClus, Kmeans, Tra-DbSCAN, and others. This study makes a significant contribution to the bibliometric analysis of clustering algorithms or methods for GPS trajectories. It examined 559 articles published in Web of Science, and these records allowed for finding significant results in the relationships between keywords, authors, and citations, among others. It was found that there were important articles that were not found in Scopus, for example, "Time-focused clustering of trajectories of moving objects" by Nanni and Pedreschi [36], which is considered in other bibliographic sources as a highly cited article. Table 9 shows a high concentration of authors from China, although the diversity of countries does predominate. In addition, it can be seen in Figure 8 that the citations between articles are closely related, possibly indicating that the topic of study is consolidating.

Although there is a wide variety of clustering algorithms for trajectories, there are very few studies or literature reviews about how they work or what fields of research they can target. The only one that slightly touches on these is Yuan et al. [37], with an analysis of clustering algorithms for trajectories. However, the study adapted it to a general context of the topic of study. According to the bibliometric review, the study of GPS data obtained from vehicles can help solve both road and urban problems. Therefore, in this line, there is still a lack of studies that provide starting guidelines for new researchers who wish to enter the field of GPS trajectory clustering, for example, to identify which roads, air spaces, or sea spaces are the most suitable for the rapid mobility of multimodal means of transport; planning the routes of modern urbanization or those under construction to reduce vehicular traffic; analyzing which patterns cause traffic accidents in order to try to avoid them; determining which routes are the most feasible for autonomous vehicles to circulate; and establishing safe roads, streets, or highways for people using a different means of transportation such as bicycles, scooters, and skateboards, among others. Finally, one can also explore a review of trajectory clustering algorithms focused on other areas such as the analysis of the mobility or migration of animals or people; trajectories of robots and unmanned aerial vehicles; and hurricane trajectory analysis, among others. In relation to this aspect, there are almost no papers indicating the trend of clustering algorithms or methods for GPS trajectories in this field of research.

Author Contributions: Conceptualization, Gary Reyes; methodology, Gary Reyes; validation, Laura Lanzarini and César Estrebow; formal analysis, Laura Lanzarini and César Estrebow; investigation, Gary Reyes; data curation, Gary Reyes; supervision, Roberto Tolozano-Benites and Aurelio F. Bariviera. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used and reproducibility instructions for this study are openly available on GitHub at <https://github.com/gary-reyes-zambrano/Bibliometric-Analysis-for-GPS-Trajectory-Clustering>. These data are from Scopus, accessed 25 January 2024.

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

References

1. Reyes, G.; Crespo, C.; León-Granizo, O.; Bazán, W.; Horta, R. Propuesta de método de extracción de ubicaciones georreferenciales de una red de carreteras para el análisis de trayectorias GPS. *Investig. Tecnol. Innov.* **2022**, *14*, 1–15. [[CrossRef](#)]
2. Reyes, G.; Lanzarini, L.C.; Estrebow, C.A.; Maquilón, V. Vehicular Flow Analysis Using Clusters. In Proceedings of the XXVII Congreso Argentino de Ciencias de La Computación (CACIC), Virtual Modality, 4–8 October 2021.

3. Moreira, J.S.; León, C.C.; Zambrano, G.R.; Joel, C.M.J. Parámetros que influyen en el congestionamiento vehicular [Parameters influencing in the vehicular overcrowding]. *Int. J. Innov. Appl. Stud.* **2018**, *24*, 1440–1455.
4. Reyes, G.; Vera, L. Reference Architecture for an Intelligent Transportation System. *Int. J. Innov. Appl. Stud.* **2016**, *15*, 2028–9324.
5. Reyes, G.; Lanzarini, L.; Estrebow, C.; Fernandez Bariviera, A. Dynamic Grouping of Vehicle Trajectories. *J. Comput. Sci. Technol.* **2022**, *22*, e11. [[CrossRef](#)]
6. Lanzarini, L.C.; Hasperu, W.; Villa Monte, A.; Jimbo Santana, P.; Reyes Zambrano, G.; Corvi, J.P.; Fernández Bariviera, A.; Olivas Varela, J.Á. Minería de Datos, Minería de Textos y Big Data. In Proceedings of the XXI Workshop de Investigadores En Ciencias de La Computación (WICC 2019, Universidad Nacional de San Juan), San Juan, Argentina, 25–26 April 2019.
7. Zambrano, G.R.; Banchón, J.M. Computación afectiva y análisis del comportamiento del consumidor [Affective computing and analysis of consumer behavior]. *Int. J. Innov. Appl. Stud.* **2017**, *20*, 551–559.
8. Merediz-Solá, I.; Bariviera, A.F. A Bibliometric Analysis of Bitcoin Scientific Production. *Res. Int. Bus. Financ.* **2019**, *50*, 294–305. [[CrossRef](#)]
9. Haddow, G. Bibliometric Research. In *Research Methods*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 241–266. [[CrossRef](#)]
10. Dede, E.; Ozdemir, E. Mapping and Performance Evaluation of Mathematics Education Research in Turkey: A Bibliometric Analysis from 2005 to 2021. *J. Pedagog. Res.* **2022**, *4*, 1–19. [[CrossRef](#)]
11. Singh, N.; Gupta, A.; Kapur, B. A Bibliometric Analysis of *IJQRM* Journal (2002–2022). *Int. J. Qual. Reliab. Manag.* **2023**, *40*, 1647–1666. [[CrossRef](#)]
12. Yuan, J.; Zheng, Y.; Zhang, C.; Xie, W.; Xie, X.; Sun, G.; Huang, Y. T-Drive: Driving Directions Based on Taxi Trajectories. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 99–108. [[CrossRef](#)]
13. Wang, K.; Pang, L.; Li, X. Identification of Stopping Points in GPS Trajectories by Two-Step Clustering Based on DPCC with Temporal and Entropy Constraints. *Sensors* **2023**, *23*, 3749. [[CrossRef](#)]
14. Van Eck, N.J.; Waltman, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics* **2010**, *84*, 523–538. [[CrossRef](#)]
15. Aria, M.; Cuccurullo, C. Bibliometrix: An R-tool for Comprehensive Science Mapping Analysis. *J. Inf.* **2017**, *11*, 959–975. [[CrossRef](#)]
16. González-Betancor, S.M.; Dorta-González, P. Porcentaje de Artículos Altamente Citados: Una Medida Comparable Del Impacto de Revistas Entre Campos Científicos. *Rev. Esp. Doc. Cient.* **2015**, *38*, e092. [[CrossRef](#)]
17. Abul, O.; Bonchi, F.; Nanni, M. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 376–385. [[CrossRef](#)]
18. Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 220–232. [[CrossRef](#)]
19. Tang, J.; Liu, F.; Wang, Y.; Wang, H. Uncovering Urban Human Mobility from Large Scale Taxi GPS Data. *Phys. A Stat. Mech. Its Appl.* **2015**, *438*, 140–153. [[CrossRef](#)]
20. Schroedl, S.; Wagstaff, K.; Rogers, S.; Langley, P.; Wilson, C. Mining GPS Traces for Map Refinement. *Data Min. Knowl. Discov.* **2004**, *9*, 59–87. [[CrossRef](#)]
21. Guo, D.; Zhu, X.; Jin, H.; Gao, P.; Andris, C. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Trans. GIS* **2012**, *16*, 411–429. [[CrossRef](#)]
22. Abul, O.; Bonchi, F.; Nanni, M. Anonymization of Moving Objects Databases by Clustering and Perturbation. *Inf. Syst.* **2010**, *35*, 884–910. [[CrossRef](#)]
23. Li, Z.; Lee, J.G.; Li, X.; Han, J. Incremental Clustering for Trajectories. In *Database Systems for Advanced Applications*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 5982, pp. 32–46. [[CrossRef](#)]
24. Chen, C.; Jiao, S.; Zhang, S.; Liu, W.; Feng, L.; Wang, Y. TripImputor: Real-Time Imputing Taxi Trip Purpose Leveraging Multi-Sourced Urban Data. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3292–3304. [[CrossRef](#)]
25. Monreale, A.; Andrienko, G.; Andrienko, N.; Giannotti, F.; Pedreschi, D.; Rinzivillo, S.; Wrobel, S. Movement Data Anonymity through Generalization. *Trans. Data Privacy* **2010**, *3*, 91–121.
26. Reyes, G.; Estrada, V.; Tolozano-Benites, R.; Maquilón, V. Batch Simplification Algorithm for Trajectories over Road Networks. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 399. [[CrossRef](#)]
27. Cobo, M.; López-Herrera, A.; Herrera-Viedma, E.; Herrera, F. An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field. *J. Inf.* **2011**, *5*, 146–166. [[CrossRef](#)]
28. Stuart, D. Open Bibliometrics and Undiscovered Public Knowledge. *Online Inf. Rev.* **2018**, *42*, 412–418. [[CrossRef](#)]
29. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]

30. Mejia-Barron, A.; De Santiago-Perez, J.; Granados-Lieberman, D.; Amezcua-Sanchez, J.; Valtierra-Rodriguez, M. Shannon Entropy Index and a Fuzzy Logic System for the Assessment of Stator Winding Short-Circuit Faults in Induction Motors. *Electronics* **2019**, *8*, 90. [[CrossRef](#)]
31. Babichev, S.; Barilla, J.; Fišer, J.; Škvor, J. A Hybrid Model of Gene Expression Profiles Reducing Based on the Complex Use of Fuzzy Inference System and Clustering Quality Criteria. In Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019), Prague, Czech Republic, 9–13 September 2019. [[CrossRef](#)]
32. Savakar, D.G.; Hiremath, R. Copy-Move Image Forgery Detection Using Shannon Entropy. In *Applied Computer Vision and Image Processing*; Springer: Singapore, 2020; Volume 1155, pp. 76–90. [[CrossRef](#)]
33. Polyakov, M.; Polyakov, S.; Iftekhar, M.S. Does Academic Collaboration Equally Benefit Impact of Research across Topics? The Case of Agricultural, Resource, Environmental and Ecological Economics. *Scientometrics* **2017**, *113*, 1385–1405. [[CrossRef](#)]
34. Lotka, A.J. The Frequency Distribution of Scientific Productivity. *J. Wash. Acad. Sci.* **1926**, *16*, 317–323.
35. Reyes, G.; Lanzarini, L.; Hasperué, W.; Bariviera, A.F. Proposal for a Pivot-Based Vehicle Trajectory Clustering Method. *Transp. Res. Rec.* **2022**, *2676*, 281–295. [[CrossRef](#)]
36. Nanni, M.; Pedreschi, D. Time-Focused Clustering of Trajectories of Moving Objects. *J. Intell. Inf. Syst.* **2006**, *27*, 267–289. [[CrossRef](#)]
37. Yuan, G.; Sun, P.; Zhao, J.; Li, D.; Wang, C. A Review of Moving Object Trajectory Clustering Algorithms. *Artif. Intell. Rev.* **2017**, *47*, 123–144. [[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.