

Review

Query Processing of Geosocial Data in Location-Based Social Networks

Arianna D'Ulizia , Patrizia Grifoni  and Fernando Ferri *

Consiglio Nazionale delle Ricerche—IRPPS, 00185 Rome, Italy; arianna.dulizia@irpps.cnr.it (A.D.); patrizia.grifoni@irpps.cnr.it (P.G.)

* Correspondence: fernando.ferri@irpps.cnr.it; Tel.: +39-06492724216

Abstract: The increasing use of social media and the recent advances in geo-positioning technologies have produced a great amount of geosocial data, consisting of spatial, textual, and social information, to be managed and queried. In this paper, we focus on the issue of query processing by providing a systematic literature review of geosocial data representations, query processing methods, and evaluation approaches published over the last two decades (2000–2020). The result of our analysis shows the categories of geosocial queries proposed by the surveyed studies, the query primitives and the kind of access method used to retrieve the result of the queries, the common evaluation metrics and datasets used to evaluate the performance of the query processing methods, and the main open challenges that should be faced in the near future. Due to the ongoing interest in this research topic, the results of this survey are valuable to many researchers and practitioners by gaining an in-depth understanding of the geosocial querying process and its applications and possible future perspectives.

Keywords: location-based social networks; query processing; geosocial data



Citation: D'Ulizia, A.; Grifoni, P.; Ferri, F. Query Processing of Geosocial Data in Location-Based Social Networks. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 19. <https://doi.org/10.3390/ijgi11010019>

Academic Editor: Wolfgang Kainz

Received: 5 November 2021

Accepted: 23 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increasing use of social media, which has reached over 3.8 billion people in 2020 worldwide [1], along with the recent advances in geo-positioning technologies, has produced a great amount of geosocial data, consisting of spatial, textual, and social information, to be managed and queried. Geosocial networks, also known as location-based social networks, have gained a relevant interest in the last decade, both from the users and the scientific community. Two examples of the most popular geosocial networks are Foursquare (www.foursquare.com, accessed on 22 December 2021) and Flickr (www.flickr.com, accessed on 22 December 2021), which couple social network functionalities with geographical information. To show this interest in numbers, we searched for “geosocial networking” OR “geosocial networks” OR “location-based social networks” in the title of the scientific articles indexed in the search engine Web of Science (WoS), in order to also investigate the scientific interest of the topic. The results in Figure 1 show a growing trend that reached its peak in 2018 by demonstrating that the scientific community has been interested in the topic of geosocial networking in the period 2010–2020 (no results were returned from 2000 to 2009).

Specifically, the scientific interest of the researchers in geosocial networking was mainly addressed to the following research topics, as analysed by Armenatzoglou and Papadias [2]: social and spatial data management, query processing, link prediction, recommendations, metrics and properties, and privacy issues. To show the interest for each research topic in numbers, we searched again the scientific articles indexed in WoS by restricting the previous search by adding a further keyword, corresponding to Armenatzoglou and Papadias' research topics, logically joined to the previous three keywords (“geosocial networking” OR “geosocial networks” OR “location-based social networks”) using the AND operator. The details of each search are provided in Table 1.

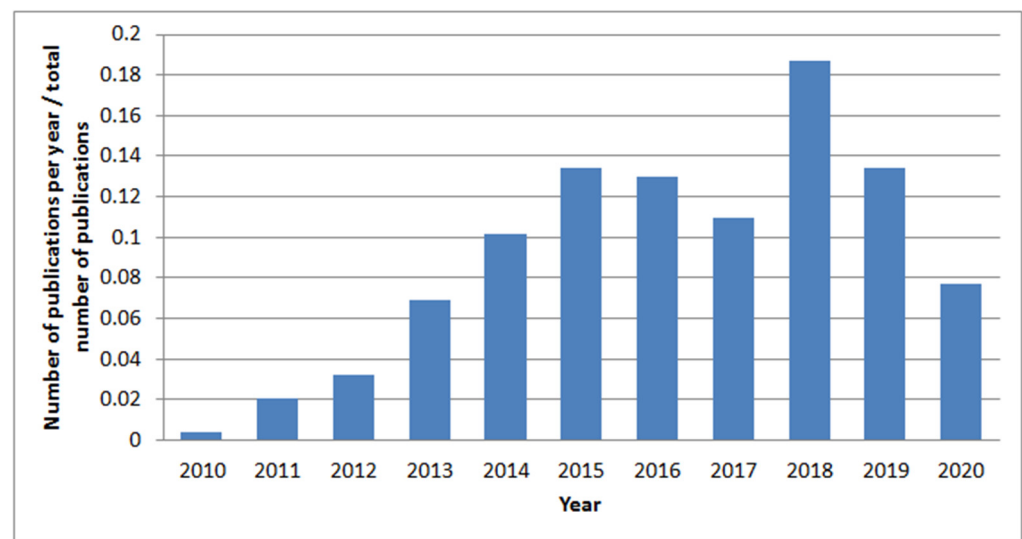


Figure 1. Number of scientific articles on “geosocial networking” OR “geosocial networks” OR “location-based social networks” in WoS by year (retrieved on March 2021).

Table 1. Scientific articles published in WoS dealing with the geosocial networking topics surveyed by Armenatzoglou and Papadias [2]. The asterisk (*) in the query allows finding all words that start with the same letters (e.g. network* finds network, networks, networking, etc.).

Armenatzoglou and Papadias' Geosocial Networking Topics	Search Keywords	Number of Published Articles Retrieved from WoS
Social and spatial data management	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “data management”)	1
Query processing	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “quer*”)	11
Link prediction	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “predict*”)	7
Recommendations	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “recommend*”)	71
Metrics	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “metric*”)	2
Privacy	((“geosocial networking” OR “geosocial network*” OR “location-based social network*”) AND “privacy”)	33

Therefore, the trend provided in Figure 1 and Table 1 shows us that geosocial networking is a popular topic that attracts the interest of the scientific community. The main addressed research issues within this topic are the recommendations of geosocial data that facilitate users to find relevant places and friends, the privacy of the users' sensitive geosocial data, and the query processing that allows extracting meaningful data from geosocial databases.

In this paper, we focus on the issue of query processing by providing a survey of the geosocial data representations, querying methods, applications, and evaluation methods, subsequently providing a systematic literature review of 57 scientific articles published over the two last decades (2000–2020) in major journals, conferences, and workshops and indexed by three major scientific search engines (WoS, Scopus, and Google Scholar).

Although several surveys have been proposed in the last few years, dealing with the various geosocial networking topics surveyed by Armenatzoglou and Papadias (recommendations [3], privacy issues [4,5], social and spatial data management [6]), to the best of our knowledge, none of these surveys focuses on the query processing topic. Due

to the ongoing interest in this research topic, the results of this survey are valuable to many researchers and practitioners by gaining an in-depth understanding of the geosocial querying process and its applications and possible future perspectives.

Aiming to identify the trends and opportunities of the research about geosocial query processing, the main research objectives of this article can be detailed as follows:

1. To study how query processing methods are applied to geosocial data by researchers and practitioners, categorising them according to the kinds of geosocial queries, the kind of method(s) used to retrieve the result of the query, the kind of access method, and the opportunity to provide an approximate solution;
2. To summarise the metrics and datasets used to evaluate geosocial queries in location-based social networks;
3. To point out the primary research challenges in this field that emerged from analysing the literature.

The remainder of the paper is organised as follows. A brief overview of the existing definitions of LBSN or geosocial networks and an overview of the process of querying geosocial data is provided in Section 2. Section 3 introduces the research methodology adopted to conduct the literature search and the analyses performed. The results of the quantitative analysis are presented in Section 4. In Section 5, we discuss the study results according to the four review questions defined in the study. Finally, in Section 6, we provide some concluding remarks.

2. Preliminary Concepts

2.1. Definitions of LBSN or Geosocial Networks

There are several definitions for “geosocial network” or “location-based social network”: the first formal definition was given by Quercia et al. [7] in 2010, who defined it as “a type of social networking in which geographic services and capabilities such as geocoding and geotagging are used to enable additional social dynamics”. One year later, Zheng [8] refined this definition by stating that “a location-based social network (LBSN) does not only mean adding a location to an existing social network so that people in the social structure can share location embedded information but also consists of the new social structure made up of individuals connected by the interdependency derived from their locations in the physical world as well as their location-tagged media content, such as photos, video, and texts”. In 2013, Roick and Heuser [9] defined LBSNs simply as “social network sites that include location information into shared contents”. Finally, one most recent definition is given by Armenatzoglou and Papadias [10] and is the following: “geosocial network (GeoSN) is an online social network augmented by geographical information”.

From the above definitions, it is evident that the peculiarity of LBSNs is the coupling of geographical information/services with social network sites that allow LBNS users to benefit from the communication and sharing functionalities provided by social networks, enhanced with geographic positions of users to locate contents, people, and activities in a physical space.

To model both the social and geographical relationships in it, a LBSN is often represented through a multilevel geosocial model, with a geosocial graph $G(V, E)$; i.e., an undirected graph with vertex set V and edge set E . Each vertex $v \in V$ represents a user and has one or more spatial locations $(v.x_i, v.y_i)$ with $1 \leq i \leq n$ in the two-dimensional space associated with the n locations visited by the corresponding user, and has one or more geo-located media content $m_j(v.x_i, v.y_i)$ with $1 \leq j \leq p$ associated to the i th location visited by the corresponding user. Each edge $e = (u, v) \in E$ denotes a relationship (e.g., friendship, common interest, shared knowledge, etc.) between two users v and $u \in V$. A graphical representation of a geosocial graph $G(V, E)$ representing an LBSN is given in Figure 2. Three layers can be differentiated, as also suggested by Gao and Liu [11]. The first layer, named *social layer*, contains the users of the LBSN and the relationships among them. The second layer, named *location* or *geographical layer*, consists of the geographical information in the two-dimensional space associated with the locations visited by the users.

The last layer, named *media content layer*, contains information about the media contents produced/shared by the users when visiting the locations.

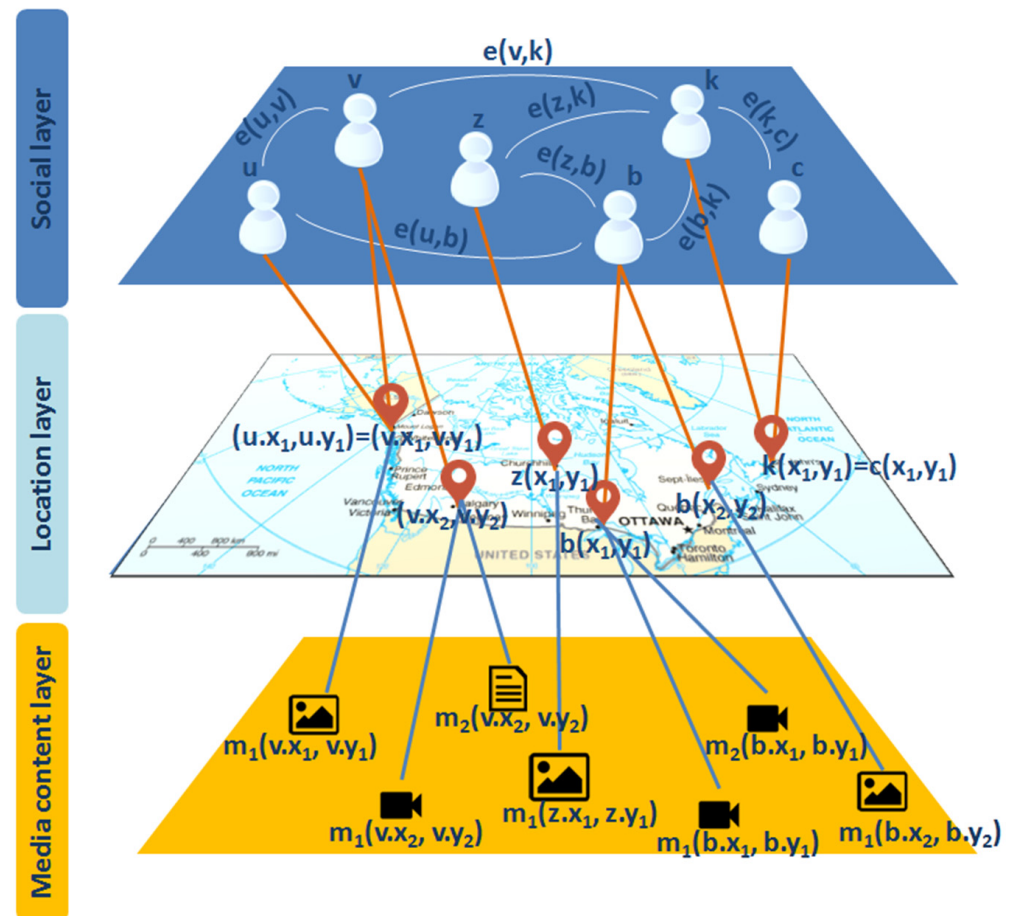


Figure 2. Multilevel geosocial model representing an LBSN with the three layers associated.

2.2. The Process of Querying Geosocial Data

To process the geosocial queries, different kinds of query primitives are defined in the literature as fundamental operations that can be further combined to answer a wide range of general-purpose geosocial queries. As suggested in [12], these kinds of query primitives can be grouped in three categories according to the layer of the geosocial graph that is exploited by the query primitive: social query primitives that exploit the data over the social graph, spatial query primitives that exploit the data over the spatial graph, and activity query primitives that exploit the data over the media content graph. A brief description of the query primitives used in geosocial query processing literature is provided in Table 2.

In addition to the query primitives, several basic heuristics or algorithms are applied to retrieve the geosocial data. Some examples found in the literature on geosocial querying are:

- Best-first search algorithm: it allows to explore paths to search in the geosocial graphs by using an evaluation function to decide which among the various available nodes is the most promising to explore [13];
- Depth-first search algorithm: it allows to explore paths to search in the geosocial graphs by starting at a given node and exploring as far as possible along each branch before backtracking [14];
- Dijkstra search algorithm: it allows to find, for a given source node in the geosocial graph, the shortest path between that node and every other node [15];

- Branch and bound algorithm: it allows to explore branches of the geosocial graphs, which represent subsets of the solution set, by checking against upper and lower estimated bounds on the optimal solution and then enumerates only the candidate solutions of a branch that can produce a better solution [16];
- Measure and conquer algorithm: it allows to explore branches of the geosocial graphs, by using a (standard) measure of the size of the subsets of the solution set (e.g., number of vertices or edges of graphs, etc.) to lower bound the progress made by the algorithm at each branching step [17].

Table 2. Query primitives.

Primitive	Description
Filter	Removes some vertices or edges from the graph that do not satisfy a selection condition.
Partitioning	Compute a partition of the vertex set into n parts of size c .
Scoring/Ranking	Ranks the vertices based on a scoring function to predict the values associated with each vertex.
Sorting	Re-arrange the vertices on the graph according to one or more keys.
Join	Compute the join between two vertex sets if a condition defined on their features is satisfied.
Clustering	Partition the vertex set into a certain number of clusters so that vertices in the same cluster should be similar to each other,
Pruning	Simplify a graph by reducing the number of edges while preserving the maximum path quality metric for any pair of vertices in the graph.

Several query indexing approaches have also been developed in the literature to optimise the processing of geosocial queries and quickly retrieve all of the data that a query requires. Existing indexing methods can be roughly categorised into three classes: the spatial-first, the social-first, and the hybrid indexing methods. The spatial-first indexing methods prioritise the spatial factor for the index construction and then improve it with the social factor. For example, MR-Tree [18], GIM-tree [19], TaR-tree [20], and SIL-Quadtree [21] employ a spatial index (e.g., R-tree, Quad-tree, G-tree) and integrate it with the textual and social information of objects. The social-first indexing methods prioritise social relationships among objects for the index construction and then improve it with the spatial information of objects. Representatives of these methods are the Social R-tree [22], B-tree [23], and 3D Friends Check-Ins R-tree [24], which index each user along with their social relationships and then integrate the spatial information. Finally, hybrid indices are developed to store both the spatial and social information of objects giving them the same priority. For example, NETR-tree [25], CD-tree [26], and SaR-tree [27,28] encode both social information and spatial information into two major pieces of information that are used to prune the search space during the query time.

3. Research Methodology

This section illustrates the methodology used to conduct an objective and replicable literature search to systematically analyse the published research knowledge and answer our research questions. To this end, we have chosen the scientific method called systematic literature review (SLR). Specifically, we have followed the SLR process described in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations [29]. The steps of the SLR process, as adapted to this study, can be summarised as follows: (1) identifying the review focus; (2) specifying the review question(s); (3) identifying studies to include in the review; (4) data extraction and study quality appraisal; (5) synthesising the findings; and (6) reporting the results.

3.1. Identifying the Review Focus

Considering the first step, the review focuses on analysing and systematising scientific knowledge related to geosocial query processing in location-based social networks. Specifically, we aim to study the query processing methods, the evaluation methodologies, and the open challenges envisaged by researchers and practitioners in their scientific works.

3.2. Specifying the Review Questions

This research objective is addressed by trying to define the following review questions (RQ), as required by Step 2 of the SLR protocol:

RQ 1: What kinds of geosocial queries are proposed in the literature? This question aims to identify the main categories of geosocial queries;

RQ 2: What are the query processing methods applied to geosocial data? This question aims to identify the methodologies and query patterns and trends;

RQ 3: How geosocial query processing methods are evaluated? This question aims to identify the metrics and datasets used to evaluate geosocial queries in LBSN;

RQ 4: Which open challenges in geosocial querying have been envisaged? This question aims to identify the challenges and future research directions in the area of study.

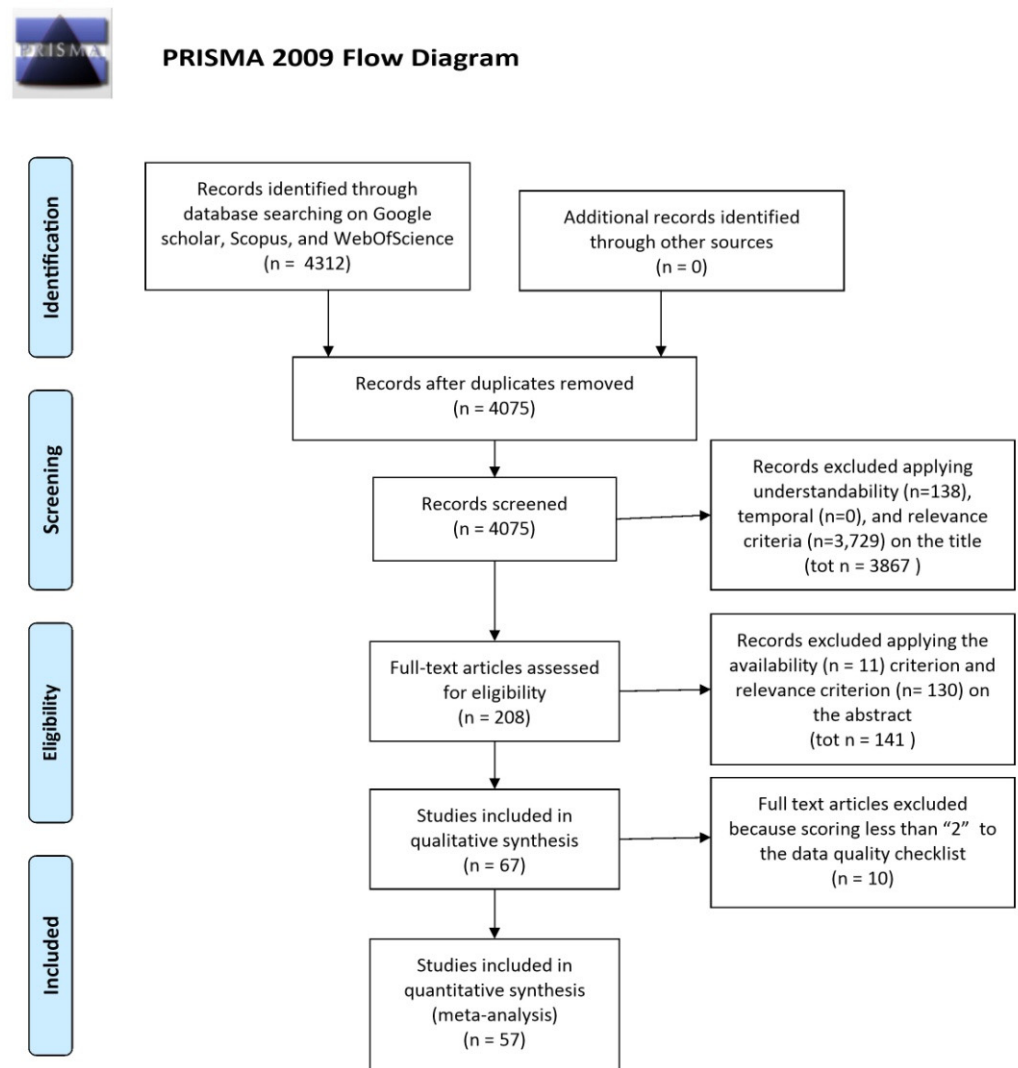
3.3. Identifying Studies to Include in the Review

Once we identified the review focus and review questions of the study, the next step of the SLR process is identifying studies to include in the review. This step includes the following four phases recommended by the PRISMA statement [29], as shown in the flow diagram of Figure 3: (1) identify records through database searching and other sources (identification phase); (2) screen and exclude records (screening phase); (3) assess full-text articles for eligibility (eligibility phase); and (4) include studies for qualitative analysis (included phase).

To identify the initial set of scientific papers, we defined the following search strings: ("location-based social network*" OR "geosocial network*" OR "geographic social network*" OR "LBSN*" OR "geosocial networking" OR "location-based social networking" OR "geosocial networking") AND "quer*".

These terms were chosen from the research questions to represent the scientific knowledge we want to search for. Moreover, we included the synonyms and related terms found in the scientific literature. For instance, "location-based social network" is also referred to as "geographic social network" or "geosocial network". Moreover, related terms to "location-based social network" are "location-based social networking" and "geosocial networking". Therefore, we included all these terms in the search strings.

The sources we used in our search for identifying the scientific works are twofold: (i) indexed scientific databases containing formally published literature (e.g., published journal papers, conference proceedings, books); and (ii) non-indexed databases containing grey literature (e.g., theses and dissertations, research and committee reports, government reports, preprints, etc.). We chose to include also grey literature in our systematic review because several studies highlighted the importance to consider it to avoid missing significant evidence [30,31]. Considering the first kind of source, Scopus and the Web of Science (WoS) core collection were identified as the most comprehensive of the published scientific research. The choice of using them was motivated by their multidisciplinary nature that allows a wider domain coverage of the retrieved literature concerning more domain-oriented databases. Moreover, Scopus is among the largest databases containing over 76 million publication records, and WoS provides a greater depth of coverage containing published literature of over 15 years. Therefore, they complement each other. Attending to the second kind of database, Google Scholar was used in this review for retrieving the grey literature since several studies have proved its effectiveness in searches for grey literature in systematic reviews [32,33].



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 3. The PRISMA four-phase flow diagram. Adapted from [29].

The search results during the screening phase were filtered according to the inclusion and exclusion criteria described in Table 3. Specifically, the duplication (e1) and understandability (e3) exclusion criteria and the temporal (i2) and relevance (i1) inclusion criteria based on the studies' titles were applied. The understandability criterion was formulated for the difficulties to examine the content of articles that are not written in English.

Table 3. Exclusion and inclusion criteria formulated for the study.

Exclusion Criteria	
	Duplication criterion:
e1	<ul style="list-style-type: none"> • same articles retrieved from two different search engines; • articles retrieved from the same search engine with the same title and authors but published in different sources.
e2	Availability criterion: <ul style="list-style-type: none"> • articles that are not accessible in full text.
e3	Understandability criterion: <ul style="list-style-type: none"> • articles that are written not in English.
Inclusion Criteria	
	Relevance criterion:
i1	<ul style="list-style-type: none"> • studies that are relevant to the review focus, i.e., they describe geosocial query processing in location-based social networks; • studies that are relevant to answer our research questions, i.e., they describe: (i) the query processing methods applied to geosocial data, or (ii) the evaluation process of geosocial query processing, or (iii) the open challenges in geosocial querying.
i2	Temporal criterion: <ul style="list-style-type: none"> • articles published in the period 2000–2020.

In the eligibility phase, the availability (e2) exclusion criterion and the relevance inclusion criteria (i1) were applied based on the studies' abstract. The availability criterion was formulated for the impossibility to analyse the content of articles that are not accessible in full text. Applying these criteria allows identifying eligible publications to establish evidence on the different geosocial query processing methods and data representation schemes.

3.4. Data Extraction and Study Quality Appraisal

The full text of the eligible articles was then analysed by two reviewers that assessed them according to a quality evaluation checklist composed of four questions, as shown in Table 4. The possible answers (with their related scores) for each quality assessment question are defined, as shown in the second column of Table 4. In case of disagreement, the “disagreed” articles were examined by a moderator that evaluated them again and provided the final scores.

Studies that scored less than “2” were excluded from the qualitative analysis, while articles that scored “2” or more were included in the systematic review.

Finally, the full texts of the included articles were analysed, and the following information was extracted from them (if any):

- Kind of geosocial query;
- Geosocial query processing method;
- Indexing method;
- Approximate solution (if available);
- Evaluation method(s);
- Evaluation metric(s);
- Evaluation dataset(s);
- Future/open challenges.

The last two phases of the SLR process, i.e., synthesising the findings and reporting the results, will be detailed in the following sections.

Table 4. Quality assessment questions and scores formulated for the study.

Quality Assessment Questions		Scores
QA1	Does the article describe a geosocial query processing method?	1—yes, the geosocial query processing method is fully described. 0.5—partially, the geosocial query processing method is only summarised without describing in detail some steps. 0—no, the geosocial query processing method is only cited, without describing it.
QA2	Does the article describe the geosocial data representation schema?	1—yes, the geosocial data representation schema is fully described. 0.5—partially, the geosocial data representation schema is only summarised without describing it in detail. 0—no, the geosocial data representation schema is not described.
QA3	Does the article provide an evaluation of the geosocial query processing method?	1—yes, the geosocial query processing method is evaluated. 0—no, the geosocial query processing method is not evaluated.
QA4	Does the article state the open/future challenges?	1—yes, the open/future challenges are clearly stated. 0—no, the open/future challenges are not stated.

4. Results of the SLR and Quantitative Analysis

During the identification phase, described in Section 3.3 and depicted in Figure 3, a total of 4312 articles were returned using the three search engines (retrieved on March 2021): 4054 from Google Scholar, 172 from Scopus, and 86 from Web of Science, respectively.

As required by the duplication criterion, removing duplicate records resulted in 4075 papers. Excluding also the articles that are not written in English (understandability criterion), a total of 3943 articles was screened for the inclusion criteria. Applying the temporal criterion resulted in no articles being excluded because all retrieved papers were published in the period 2000–2020. The relevance criterion was applied by searching for the term “quer*” in the articles’ titles, resulting in 208 articles at the end of the screening phase.

Removing the articles that are not accessible in full text (11 studies for the availability criterion) and the articles that are not relevant (130 studies for the relevance criterion) by applying the relevance criterion to the articles’ abstract, a total of 67 articles were retained for a full evaluation of eligibility. Specifically, the articles that do not talk about geosocial queries in the abstract were excluded.

Two reviewers assessed these 67 studies according to the quality evaluation checklist shown in Table 4. Seven studies that scored less than “2” were excluded, while the remaining 57 studies were included in the qualitative synthesis and the information listed in Section 3.4 were extracted from their full texts. Table 5 provides an overview of the selected studies, where the reference, publication type, publication year, publisher, and citation count (from Google Scholar) for each study are provided.

Table 5. Overview of the selected studies.

ID	Reference	Kind of Source	Year of Publication	Publisher	Citation Count
S1	[34]	Conference	2017	Springer	15
S2	[2]	Conference	2013	ACM	92
S3	[35]	Conference	2019	Springer	0
S4	[36]	Journal	2019	Elsevier	2
S5	[18]	Conference	2017	IEEE	8
S6	[37]	Conference	2012	Springer	50
S7	[38]	Journal	2020	IEEE	1
S8	[39]	Journal	2020	IEEE	0
S9	[40]	Conference	2013	ACM	132
S10	[41]	Journal	2018	Springer	1
S11	[42]	Conference	2020	ACM	1
S12	[43]	Conference	2020	Springer	2
S13	[28]	Thesis	2016	repository.hkbu.edu.hk	0

Table 5. Cont.

ID	Reference	Kind of Source	Year of Publication	Publisher	Citation Count
S14	[44]	Journal	2020	IEEE	2
S15	[45]	Journal	2016	ACM	3
S16	[46]	Journal	2020	Elsevier	1
S17	[47]	Conference	2019	Springer	1
S18	[48]	Journal	2020	Taiwan Academic Network Management Committee	0
S19	[49]	Journal	2017	Springer	3
S20	[50]	Journal	2016	w.bncss.org	3
S21	[51]	Journal	2016	Elsevier	11
S22	[52]	Journal	2017	Springer	52
S23	[27]	Journal	2015	IEEE	29
S24	[53]	Conference	2017	Springer	1
S25	[54]	Journal	2015	Springer	29
S26	[55]	Conference	2014	Springer	23
S27	[24]	Conference	2020	Springer	2
S28	[23]	Journal	2020	mdpi.com	2
S29	[56]	Conference	2019	IEEE	2
S30	[57]	Conference	2017	Springer	5
S31	[58]	Conference	2013	Springer	54
S32	[59]	Conference	2017	Springer	10
S33	[60]	Journal	2018	IEEE	3
S34	[20]	Conference	2015	microsoft.com	10
S35	[61]	Journal	2020	Elsevier	0
S36	[62]	Conference	2018	Springer	4
S37	[63]	Journal	2016	mdpi	5
S38	[64]	Journal	2019	Elsevier	0
S39	[25]	arxiv	2019	arxiv.org	0
S40	[19]	Journal	2020	Elsevier	2
S41	[65]	Journal	2018	IEEE	19
S42	[22]	Conference	2012	ACM	107
S43	[66]	Conference	2018	Springer	2
S44	[67]	Journal	2017	Springer	12
S45	[68]	Conference	2018	search.ieice.org	2
S46	[69]	Thesis	2015	etda.libraries.psu.edu	0
S47	[70]	Journal	2020	IEEE	0
S48	[12]	Conference	2016	IEEE	0
S49	[26]	Conference	2020	IEEE	5
S50	[21]	Journal	2020	Elsevier	1
S51	[71]	Journal	2018	academic.oup.com	13
S52	[72]	Journal	2015	IEEE	20
S53	[73]	Journal	2016	ACM	8
S54	[74]	Journal	2014	Elsevier	41
S55	[75]	Journal	2018	Springer	5
S56	[76]	Conference	2018	IEEE	8
S57	[77]	Conference	2010	ACM	65

The selected studies have been published mainly in journals (50.88%—29 studies), followed by conference proceedings (43.86%—25 studies), theses (3.51%—2 studies), and only 1 preprint (1.75%). Therefore, the majority of the studies (94.74%) are formally published studies (journal and conference papers), while only 5.26% are composed of grey literature (thesis and preprint).

The temporal distribution of the selected publications, shown in Figure 4, underscores the increasing interest of the scientific community in the topic of geosocial querying, which started growing in 2010 and continues to grow in 2020.

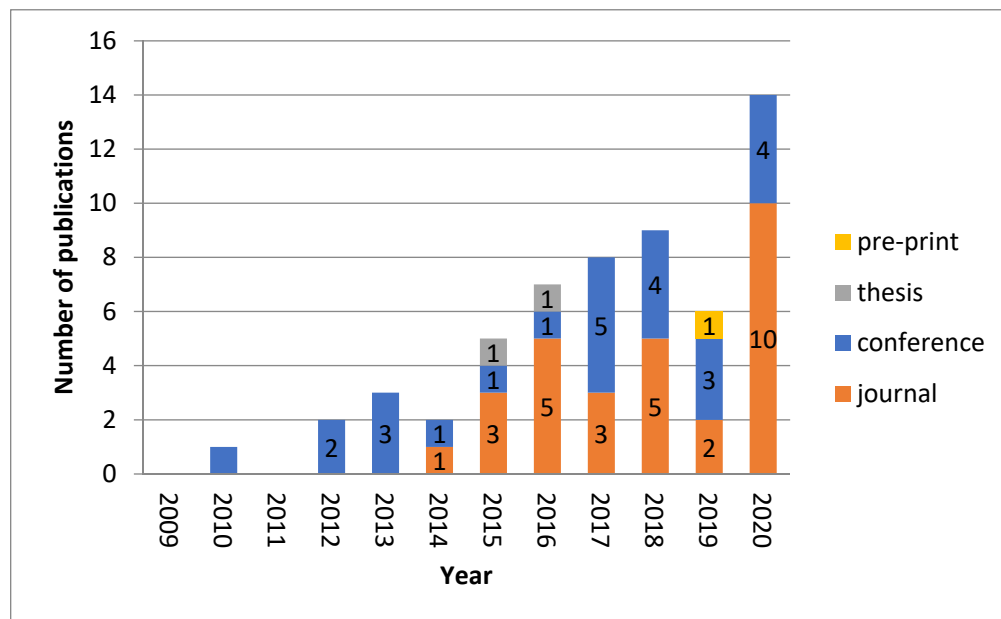


Figure 4. Temporal distribution of the selected publications.

5. Findings and Discussion

This section analyses how the 57 selected studies answered our four review questions introduced in Section 3.2. Specifically, to deal with RQ1, we start by analysing and classifying the kinds of geosocial queries. With respect to RQ2, the query processing methods applied to the geosocial network data are extracted and classified. Addressing RQ3, the metrics and datasets used to evaluate the geosocial queries in LBSN are analysed. Finally, as part of RQ4, the open challenges in geosocial querying proposed in these studies are analysed.

5.1. RQ 1: What Kinds of Geosocial Queries Are Proposed in the Literature?

To answer the first RQ, we look first at the kinds of queries proposed by the selected studies, and then at the constraints (social, spatial, temporal) considered.

Based on our analysis, we identified seven categories of geosocial queries (as presented in Figure 5) that consider both social and spatial relations: geosocial group queries, geosocial keyword queries, geosocial top-k queries, geosocial skyline queries, geosocial moving queries, geosocial fuzzy queries, and geosocial nearest neighbor queries. Moreover, among the selected studies, there were three frameworks providing a collection of query primitives essential for geosocial queries.

In the following paragraphs, we briefly discuss each category of the geosocial queries defined above.

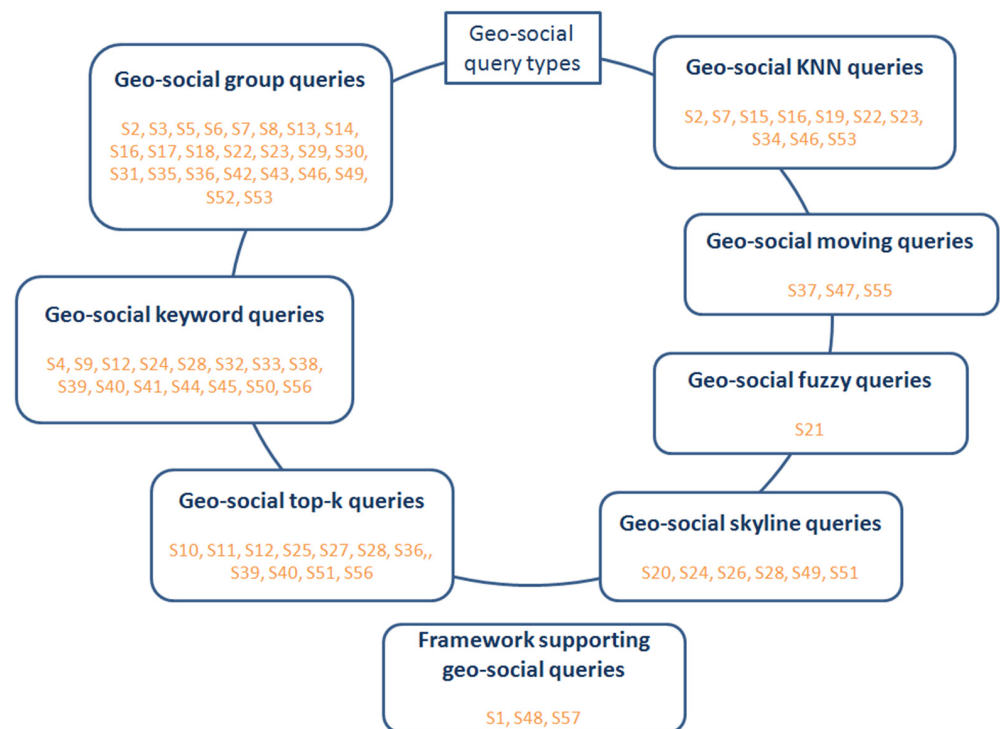


Figure 5. Categories of geosocial queries identified by analysing the surveyed studies, along with the IDs of the studies (defined in Table 4) belonging to each category.

5.1.1. Geosocial Group Queries

The most numerous category of geosocial queries is the *group query* with 25 studies (43.85%), which allows finding a group of users close to each other both socially and geographically. Generally, the studies addressing this kind of query start from spatial queries (e.g., range, k nearest neighbour, spatial join) to find geographically close users and integrate them by considering grouping concepts to find also socially close users. That results in several kinds of queries (see Table 6) that we have grouped here in the class of *geosocial group queries*. An example of a geosocial group query, inspired by the work in [74], is depicted in Figure 6.

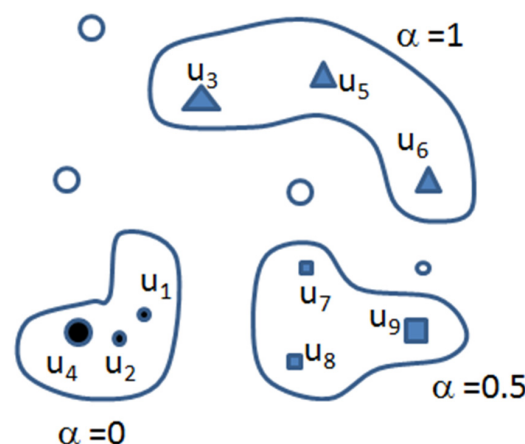


Figure 6. An example of geosocial group query that considers a set of users $\{u_1, u_2, \dots, u_9\}$ located in the places depicted by circles, squares, and triangles. The sizes of those shapes indicate the user's interests in the query keywords. Query q requests a user group of size 3 that maximizes the ranking function. The query returns the set of users $\{u_1, u_2, u_4\}$ when $\alpha = 0$ (i.e., only the group diameter is considered), the set of users $\{u_3, u_5, u_6\}$ when $\alpha = 0.5$, and the set of users $\{u_7, u_8, u_9\}$ for $\alpha = 1$ (i.e., only the group interest is considered). $\alpha \in [0, 1]$ is a parameter used to balance the group interest and the group diameter.

Table 6. Geosocial group queries.

ID	Name of the Query	Description
S2	Range Friends (RF)	returns the friends of a user within a given range
	Nearest Friends (NF)	returns the nearest friends of a user to a given location
	Nearest Star Group (NSG)	returns a user group, which (i) forms a star subgraph of the social network, and (ii) minimises the aggregate (Euclidean) distance of its members to a given location
S3 S18	Minimum user spatial-aware interest group query (MUSIGQ)	returns a group of users that have the common interests and stay in the near spots
S5	Multiple Userdefined Spatial Query (MUSQ)	returns the best answers for a group of users considering both their locations and non location preferences
S6	Circle of Friend Query (CoFQ)	finds a group of friends who are close to each other both socially and geographically
S7	Cohesive group nearest neighbor (CGNN)	return a group of attendees such that the travel cost of each attendee is within a range, and the total travel cost of all attendees is minimised
	Cohesive group nearest neighbor queries under multi-criteria (MCGNN)	return a group of attendees and a set of locations such that the travel cost of each attendee is within a range, and the overall scores of locations are maximised under multi-criteria
S8	l-cohesive m-ridesharing group (lm-CRG)	retrieves a cohesive ridesharing group by considering spatial, social, and temporal information
S13 S54	Spatial-aware Interest Group (SIG)	retrieves a user group where each user is interested in the query keywords and the users are close to each other in the Euclidean space
	Geo-Social K-Cover Group (GSKCG)	finds a minimum user group in which the members satisfy certain social relationship and their associated regions can jointly cover all the query points
	Social-aware Ridesharing Group (SaRG)	retrieves a group of riders by taking into account their social connections besides traditional spatial proximities
S14	Group planning query over spatial-social networks (GP-SSN)	retrieves a group of friends with common interests on social networks and a number of spatially close points of interest (POIs) that best match group's preferences and have the smallest traveling distances to the group.
S16	Reverse nearest neighborhood (RNH)	discovers the neighborhoods that find a query facility as their nearest facility among other facilities in the dataset
S17 S30	Spatial Group Preference (SGP)	returns top-k POIs that are much likely to satisfy the group's preferences for POI categories
S22	Geosocial group query	retrieves k users that satisfy the minimum acquaintance constraint and has the minimum spatial distance to the query issuer
S23	Geo-Social K-Cover Group (GSKCG)	retrieves a minimum user group in which each user is socially related to at least k other users and the users' associated regions can jointly cover all the query points
S29	Group nearest compact POI set (GNCS)	finds a compact set of POIs that is close to all users
S31	Group trip planning (GTP)	returns for each type of data points those locations that minimize the total travel distance for the entire group
S35	User community preference query	return satisfied POIs based on semantic spatial information and semantic category preference weights
S36	Geo-Social Group preference Top-k (SG-Topk)	returns top-k places that are most likely to satisfy the needs of users based on spatial and social relevance
S42 S52	Socio-Spatial Group Query (SSGQ)	select a group of nearby attendees with tight social relation
S43	Personalised geosocial group (PGSG)	find a venue and a user group, where each user is socially connected with at least c other users, and the maximum distance of all the users in the group to the venue is minimised
S46	Reverse Nearest Social Group (RNSG)	finds all social groups that satisfy k-core constraint and have their farthest member (individual with maximum euclidean distance to the query point) as a reverse nearest neighbor of the query point
S49	Skyline cohesive group query	finds a group of users, which are strongly connected and closely co-located
S52	Multiple Rally-Point Social Spatial Group Query (MRGQ)	selects an appropriate activity location for a group of nearby attendees with tight social relationships
S53	Consensus query	finds a meeting place that minimises the travel distance for at least a specified number of group members

The main types of spatial constraints that have been applied in these studies are the following:

- Distance: typical distance functions are Euclidean distance for items that are located in a small area; network distance, which is the length of the shortest path between the items on the road network of the search area; and Haversine formula, which is the distance between the items on the surface of a sphere [43].
- Range: the locations of the retrieved items (users/objects/PoIs) are within the query region.
- Coverage: the coverage of a set of query points is the minimum rectangle containing all query points.
- Travel cost, which is the expected cost of a direct travel from one item to the other.

More than half of the studies use distance (mainly Euclidean) to measure the spatial distance between two points in the space. Eight studies apply the travel cost constraint, only 3 works use the range constraint, and 2 studies the coverage (see Table 7).

Table 7. Main types of spatial, social, and temporal constraints applied in geosocial group queries.

Constraints		Paper ID	Total	
Spatial	Distance	Euclidean	S3, S5, S6, S13, S16, S18, S54	7
		No-Euclidean	S2, S17, S22, S23, S30, S35, S36, S42, S52, S43, S46	11
	Range	S2, S7, S23	3	
	Coverage	S13, S54	2	
	Travel cost	S7, S8, S13, S14, S29, S49, S53, S54	8	
Social	Friendship	S2, S29, S31, S36, S53	5	
	Interest/preference score	S3, S5, S13, S14, S17, S18S30, S35, S54	9	
	Closeness	S6, S7, S16	3	
	Acquaintance	S8, S13, S22, S23, S42, S43, S46, S49, S52, S54	10	
Temporal		S8	1	

The social constraints that have been applied in these studies are the following:

1. Friendship: in a geosocial network, friendship relations correspond to the edges between two nodes representing users.
2. Interest/preference score: considers the interest(s)/preference(s) of a user or a group of users in spatial objects annotated by one or more keywords and can be computed by its/their check-ins on these spatial objects.
3. Closeness: it restricts the users in a social group considering the proximity of candidate attendees to corresponding locations in the physical world, and sometimes even the ratings of assembly points as additional references [38].
4. Acquaintance: it imposes a minimum degree on the familiarity of group members (which may include q); i.e., every user in the group should be familiar with at least k other users [52]. It is a measure of group cohesiveness. The value of k can be defined according to a minimum social distance that should be less than or equal to an acceptable social boundary.

The majority of the studies (10 studies) apply the acquaintance constraint, while 9 works use the interests/preferences constraint, 5 studies apply the friendship constraint, and 3 studies the closeness (see Table 7). The acquaintance constraint allows avoiding finding a group with mutually unfamiliar members by retrieving a cohesive subgroup in the geosocial network.

Finally, only one study [39] proposing geosocial group queries incorporates temporal constraints, in addition to spatial and social ones, to retrieve a cohesive ridesharing group.

5.1.2. Geosocial Keyword Queries

Generally, the studies addressing this kind of query start from conventional spatial keyword queries to find objects that are spatially and textually relevant to the user-supplied keywords, and integrate them by considering also collective and social criteria to find these objects. The number of surveyed studies that belong to this class of geosocial query is 15 (26.31%) (see Table 8). An example of a geosocial keyword query, inspired by the work in [40], is depicted in Figure 7.

The type of spatial constraints that has been applied in these studies is twofold: (i) the distance, already defined in the previous sub-section on “Geosocial group queries”; and (ii) the cost, which is calculated according to two kinds of cost functions, the maximum sum cost and the diameter cost. The maximum sum cost is defined as the linear combination of the maximum distance between the query and a node in the POI set [40], while the diameter cost is defined as the maximum distance between any pair of nodes in the POI set [64]. Similarly to the geosocial group query, the majority of the studies (9 studies) use the distance to measure the spatial distance, while 6 studies use the cost.

Table 8. Geosocial keyword queries.

ID	Name of the Query	Description
S9 S32 S41	collective spatial keyword query (CoSKQ)	finds a set of objects in the database such that it covers a set of given keywords collectively and has the smallest cost
S12	Multiple Reverse Top-k Geo-Social Keyword Query (RkGSKQ)	aims to find all the users who have multiple geosocial objects in their top-k geosocial keyword query results
S24	Geo-Social Keyword Skyline Query (GSKSQ)	returns the skyline of a set of POIs based on a query point, the social relationships of the query owner, and query keywords
S28	geo-social top-k keyword (GSTK)	retrieves the k best data objects based on spatial, textual and social relevance
S28	geosocial skyline keyword (GSSK)	returns every object within range which is not dominated by any other object in terms of distance to the query location and aggregated score of social and keyword relevance
S4 S33 S45	multiple-user location-based keyword (MULK) query	returns a set of POIs that are ‘close’ to the locations of the users in a group and can provide them with potential options at the lowest expense (e.g., minimising travel distance)
S38	multiple-user closest keyword- set (MCKS) query	searches a set of Points of Interest (POIs) that cover the query keyword-set, are close to the locations of multiple users, and are close to each other
S39	Social-based Time-aware Spatial Keyword Query (STSKQ)	returns the top-k objects by taking geo-spatial score, keywords similarity, visiting time score, and social relationship into consideration
S40	diversified top-k geosocial keyword (D k GSK) query	returns the top- k objects based on their spatial and textual proximity to q as well as the check-in counts of u ‘s friends at such objects
S44	Popularity-aware collective keyword (PAC-K) query	finds a group of popular POIs that cover the query’s keywords and satisfy the distance requirements from each node to the query node and between each pair of nodes, such that the sum of rating scores over these nodes for the query keywords is maximized
S50	Social space Keyword Query	returns the top-k semantic trajectory for users has higher social relevance and shorter distance while satisfying spatial and keyword constraints
S56	why-not top-k geosocial keyword (WNGSK) query	returns the top-k objects based on their spatial and textual proximity to the query location as well as the check-in counts of user’s friends at such objects

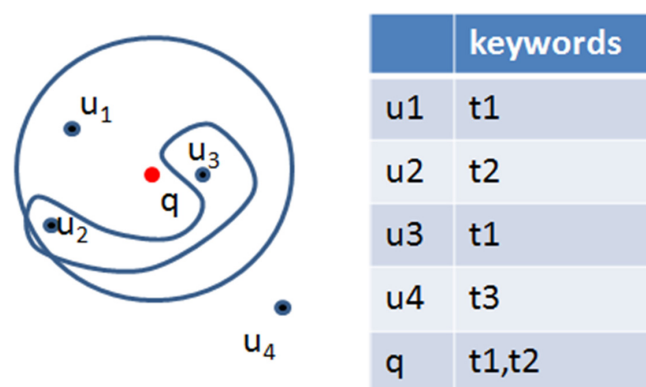


Figure 7. An example of a geosocial keyword query that considers a set of objects $\{u_1, u_2, \dots, u_4\}$ located in the places depicted by circles and associated with keywords shown in the table on the right. Query q requests a location (red circle) and a set of keywords. The query returns the set of objects $\{u_2, u_3\}$ that minimizes the distance and contains the required keywords.

Considering the social constraints, besides the friendship relationships among the nodes of the network, further social constraints that have been applied in these studies are the following:

- Relevance: it is obtained from the number of fans and the relationship between these fans and the query user, where a fan is a user who exhibits positive behavior towards an object (e.g., check-in, like, share, etc.) [23];
- Relationship effect: it can be measured by the similarity of embedding vectors between users and their neighbors with all users' check-in records [25].

The majority of these studies (4 studies) apply the relevance constraint, while 2 studies apply the friendship constraint, and only 1 work uses the relationship effect constraint (see Table 9).

Table 9. Main types of spatial, social, and collective constraints applied in geosocial keyword queries.

Constraints		Paper ID	Total
Spatial	Cost	S9, S32, S41, S38, S44, S50	6
	Distance	S12, S24, S28, S4, S33, S39, S40, S45, S56	9
Social	Friendship	S12, S24	2
	Relevance	S28, S40, S50, S56	4
	Relationship effect	S39	1
Collective		S4, S9, S32, S33, S38, S41, S44, S45	8

In addition to these social constraints, several geosocial keyword queries (8 studies) apply a collective constraint, meaning that the group's keywords collectively cover the query keywords.

5.1.3. Geosocial Top-k Queries

The third most numerous class of geosocial queries is the geosocial top-k query with 11 studies (19.3%) (see Table 10). Generally, the studies addressing this kind of query rely on the conventional top-k queries that retrieve the top-k objects based on a user-defined scoring function, and enrich the top-k query semantics by considering both spatial and social relevance components to compute the scoring function. An example of a geosocial top-k query, inspired by the work of [71], is shown in Figure 8.

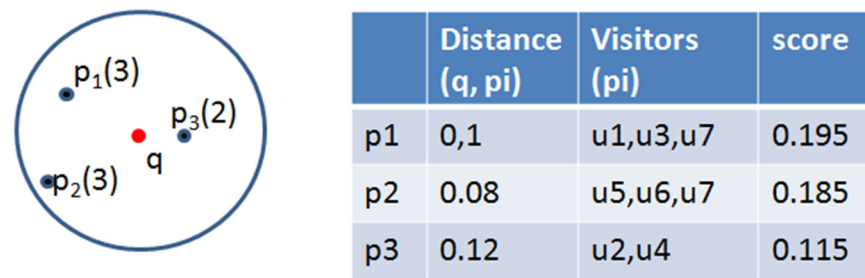


Figure 8. An example of a geosocial top-k query that considers the query location q , a set of places $\{p_1, p_2, p_3\}$, and a set of users $\{u_1, u_2, \dots, u_7\}$. The table on the right side shows the spatial distances between the query location and places, the number of visitors of each place, and the score of each place, according to the scoring function.

Table 10. Geosocial top-k queries.

ID	Name of the Query	Description
S10	Top-k join queries	compute the k combinations of several query search results over geospatial and social data sources with the highest score
S11	Top-k spatio-social Point-of-Interest Queries	rank POIs by a weighted sum of their popularity and proximity
S12	Multiple Reverse Top-k Geo-Social Keyword Query (RkGSKQ)	aims to find all the users who have multiple geosocial objects in their top-k geosocial keyword query results
S25	Geo-Social Ranking top-k query	ranks the k users with the highest scores computed on their distance to a location, the number of their friends in the vicinity of the location, and possibly the connectivity of those friends
S27	Geo-Social Temporal Top-k (GSTTk)	retrieves top-k places (points of interest) ranked according to their spatial, social, and temporal relevance to the query user
S28	Geo-social top-k keyword (GSTK)	retrieves the k best data objects based on spatial, textual and social relevance
S36	Geo-Social Group preference Top-k (SG-Topk)	returns top-k places that are most likely to satisfy the needs of users based on spatial and social relevance
S39	Social-based Time-aware Spatial Keyword Query (STSKQ)	returns the top-k objects by taking geo-spatial score, keywords similarity, visiting time score, and social relationship into consideration
S40	Diversified top-k geosocial keyword (D k GSK) query	returns the top- k objects based on their spatial and textual proximity to q as well as the check-in counts of u 's friends at such objects
S51	Top-k famous places (TkFP)	retrieves top-k places (points of interest) ranked according to their spatial and social relevance to the query user
S56	Why-not top-k geosocial keyword (WNGSK) query	returns the top-k objects based on their spatial and textual proximity to the query location as well as the check-in counts of user's friends at such objects

All the studies apply the distance, defined in the previous sub-section on “Geosocial group queries”, as a spatial constraint of the query.

Considering the social constraints, besides the friendship, relevance, and relationship effect, already mentioned and described in the previous classes of queries, further social constraints that have been applied in these studies are the following:

- Popularity: it is obtained by quantifying how many users have the location in their k nearest neighbours results [42];
- Social connectivity: the social connectivity of a geosocial graph can be defined as the graph density and can be measured by a formula provided [78].

The majority of these studies (7 studies) apply the relevance constraint, while 4 studies apply the friendship constraint, and only 1 work uses the relationship effect, the popularity, or the connectivity constraint (see Table 11).

Table 11. Main types of spatial, social, and temporal constraints applied in geosocial top-k queries.

	Constraints	Paper ID	Total
Spatial	Distance	S10, S11, S12, S25, S27, S28, S36, S39, S40, S51, S56	11
	Friendship	S12, S25, S27, S51	4
	Popularity	S11	1
Social	Relationship effect	S39	1
	Relevance	S10, S27, S28, S36, S40, S51, S56	7
	Connectivity	S25	1
Temporal		S27, S39	2

Finally, two studies [24,25] proposing geosocial top-k queries incorporate temporal constraints, in addition to spatial and social ones.

5.1.4. Geosocial Skyline Queries

The skyline operator was introduced by Borzsony et al. [79] for retrieving a set of data objects O that are not dominated by others, meaning that any other set of object O' is worse than O for all the attributes of the query. The category of geosocial skyline query enriches the semantics of the skyline operator by considering also the social relationships of the query owner for retrieving the set of data objects O . Six of the surveyed studies (10.5%) belong to this class of geosocial query (see Table 12). An example of a geosocial skyline query, inspired by the work in [55], is shown in Figure 9.

Table 12. Geosocial skyline queries.

ID	Name of the Query	Description
S20	LBSNs friend recommendation skyline query (LFRSQ)	returns the friend recommendation list by considering three factors: (a) common friend, (b) distance influence, and (c) similarity score, which is calculated from location similarity and friend influence between user and candidate friends
S26	Geosocial skyline query	reports for a given user and a given location the pareto-optimal set of persons who are close to the location and closely connected to the user
S24	Geo-Social Keyword Skyline Query (GSKSQ)	returns the skyline of a set of PoIs based on a query point, the social relationships of the query owner, and query keywords
S28	Geosocial skyline keyword (GSSK)	returns every object within range which is not dominated by any other object in terms of distance to the query location and aggregated score of social and keyword relevance
S49	Skyline cohesive group query	finds a group of users, which are strongly connected and closely co-located
S51	Socio-Spatial Skyline Query (SSSQ) query	returns every place for which there does not exist any other place that has a better social score and better spatial score

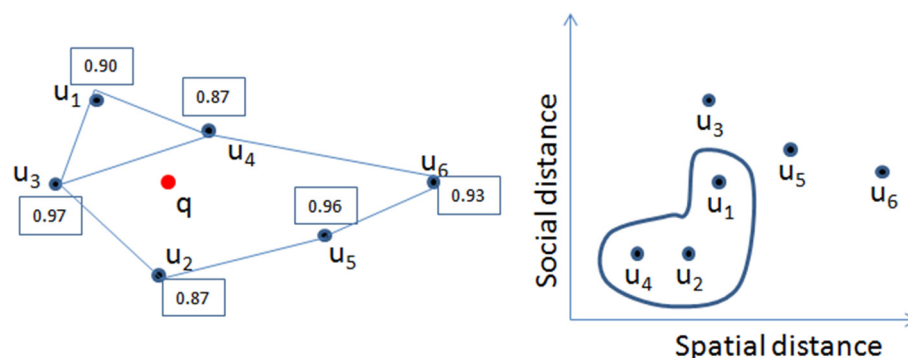


Figure 9. An example of a geosocial skyline query that considers the query location q and a set of users $\{u_1, u_2, \dots, u_6\}$ with the social distance of each user from q indicated in the labels. The query returns the set of users $\{u_1, u_2, u_4\}$ according to the social and spatial distance in the skyline space.

Similarly to the category of geosocial top-k queries, all the studies proposing geosocial skyline queries apply the distance as a spatial constraint of the query.

Attending to the social constraints, in addition to the friendship, relevance, and acquaintance, already mentioned and described in the previous categories of queries, further social constraints that have been applied in these studies are the following:

- Social influence: it is applied to retrieve friends who have closer social ties and it is computed based on both the social connections and similarity of the check-in activities [50].
- Social similarity: it measures how socially close people are. Several methods for measuring this proximity have been proposed in the literature, and the most adopted are the Random Walks with Restart method and the Bookmark Coloring Algorithm, which considers all walks between two users [55].

In terms of numbers, the most applied social constraint in this category is the friendship constraint (2 studies), followed by social influence, social similarity, relevance, and acquaintance constraints with one study each (see Table 13).

Table 13. Main types of spatial and social constraints applied in geosocial skyline queries.

Constraints		Paper ID	Total
Spatial	Distance	S20, S24, S26, S28, S49, S51	6
	Friendship	S24, S51	2
	Influence	S20	1
Social	Similarity	S26	1
	Relevance	S28	1
	Acquaintance	S49	1

5.1.5. Geosocial Nearest Neighbor Queries

Chen and Lu [80] define a nearest neighbour (NN) query as a query aimed to find the set of nearest items (users/objects/PoIs) to the query point in terms of spatial distance. The most popular variant of NN query is the k-nearest neighbor (k-NN) query that retrieves the k-nearest points to the query point. An example of a k-NN query, extracted from [46], is provided in Figure 10. The geosocial NN query extends the computation of the nearest items by considering not only the spatial distance but also social criteria to find these objects. Ten of the surveyed studies (17. 5%) belong to this class of geosocial query (see Table 14).

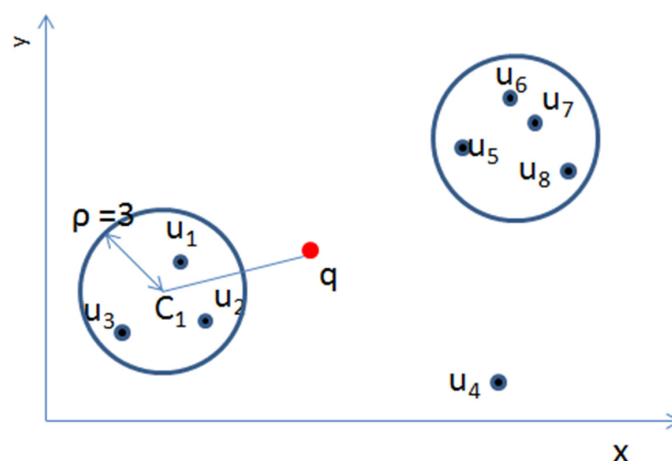


Figure 10. An example of a geosocial nearest neighbor query that considers a set of users $\{u_1, u_2, \dots, u_8\}$ and the query location q . The query returns C_1 with radius constraint $p = 3$, which is the nearest neighborhood to q .

Table 14. Geosocial nearest neighbor queries.

ID	Name of the Query	Description
S2	Nearest Friends (NF)	returns the nearest friends of a user to a given location
S7	Cohesive group nearest neighbor (CGNN)	returns a group of attendees such that the travel cost of each attendee is within a range, and the total travel cost of all attendees is minimised
	Cohesive group nearest neighbor queries under multi-criteria (MCGNN)	return a group of attendees and a set of locations such that the travel cost of each attendee is within a range, and the overall scores of locations are maximised under multi-criteria
S15	k-Relevant nearest neighbor (k-RNN)	retrieves close-by and relevant (as judged by the crowd) POIs
S16	Reverse nearest neighborhood (RNH)	discovers the neighborhoods that find a query facility as their nearest facility among other facilities in the dataset
S19	kNN and range queries	discover the hot zones (highly populated areas) based on users' spatial movement patterns and incorporate them into the construction of watchtowers
S22	Geosocial group queries	retrieve k users that satisfy the minimum acquaintance constraint and has the minimum spatial distance to the query issuer
S23	Geo-Social K-Cover Group (GSKCG)	retrieves a minimum user group in which each user is socially related to at least k other users, and the users' associated regions can jointly cover all the query points
S34	k-nearest neighbor temporal aggregate (kNNTA) query	returns the top-k locations that have the smallest weighted sums of (i) the spatial distance to the query point and (ii) a temporal aggregate on a certain attribute over the time interval
S46	Reverse Nearest Social Group (RNSG)	finds all social groups that satisfy k-core constraint and have their farthest member (individual with maximum euclidean distance to the query point) as a reverse nearest neighbor of the query point
S53	Consensus query	finds a meeting place that minimises the travel distance for at least a specified number of group members

The spatial constraints that have been applied in these studies are the distance and travel costs, already defined in the sub-section on “Geosocial group queries”. Specifically, 8 studies apply the distance, while only 2 studies apply the travel cost (see Table 15).

Table 15. Main types of spatial, social, and temporal constraints applied in geosocial nearest neighbor queries.

	Constraints	Paper ID	Total
Spatial	Distance	S2, S15, S16, S19, S22, S23, S34, S46	8
	Travel cost	S7, S53	2
Social	Relevance	S15	1
	Popularity	S19, S34	2
	Closeness	S7, S16	2
	Friendship	S2, S46, S53	3
	Acquaintance	S22, S23	2
Temporal		S34	1

Attending to the social constraints, five different kinds of social constraints have been applied in these studies: the friendship constraint, which is the most applied in this category with 3 studies, followed by popularity, closeness, and acquaintance constraints with 2 studies, and the relevance with 1 study.

Finally, one study [20] proposing geosocial NN queries incorporates also temporal constraints, in addition to spatial and social ones.

5.1.6. Geosocial Moving Queries

Moving queries are an important type of query of moving objects, asking for a set of objects that satisfy the spatial query constraints in a given time interval. The geosocial moving queries enlarge the query requests also to the variation in social relationships, in addition to the movements with spatial and temporal characteristics [63]. Three of the surveyed studies (5.26%) belong to this category of geosocial query (see Table 16).

Table 16. Geosocial moving queries.

ID	Name of the Query	Description	Constraints					
			Spatial	Spatio-Temporal		Social		
			Distance	Trajectories	Route	Relationships	Similarity	Trust
S37	Geosocial moving query	retrieves trajectories, underlying geographical space and social relationships for mass moving objects	✓	✓	X	✓	X	X
S47	Moving reverse nearest neighbour (RNN) query	retrieves neighbourhoods that consider the moving query point as the nearest of all the other facilities	X	✓	X	X	✓	X
S55	Social trust aware personalised route query (STPRQ)	finds a proper route R from the starting venue to the destination that should pass through several venues of the respective categories and be credible and popular in the social circle of the query user	✓	X	✓	X	X	✓

Similarly to the category of geosocial top-k queries, all the studies proposing geosocial moving queries consider distance as a spatial constraint of the query.

Considering the spatio-temporal constraints, the surveyed studies apply two different kinds of movement constraints: trajectory and route constraints. The former defines constructs for retrieving the trajectories of the moving object, while the latter allows searching for the optimal route that passes through the locations specified in the query.

Attending to the social constraints, in addition to the friendship and social similarity, already mentioned and described in the previous categories of queries, a further social constraint that has been applied in these studies is social trust. It measures the credibility between two persons and can be computed considering features that exploit social information and user behavioural patterns, including user profiles, social structure, and user behaviors in the geosocial network [75].

5.1.7. Geosocial Fuzzy Queries

Fuzzy queries have been defined by Hassine et al. [81] as queries with imprecision in the preferences about the desired items that are expressed usually using fuzzy conditions. Therefore, the terms in the queries do not have to be an exact match with the retrieved terms but within the maximum distance specified in the fuzziness.

Only one surveyed work [51] proposes fuzzy queries for geosocial networks. Specifically, in the work of Chen et al. [51], fuzzy queries are defined over a social relational network model, called an intuitionistic fuzzy social relational network (IFSRN) model,

representing and reasoning with negative, positive, and neutral relationships between actors, and can get the degrees of truth and the degrees of false of the fuzzy queries.

5.1.8. Frameworks Supporting Geosocial Query Processing

In addition to the 54 studies proposing the geosocial queries classified in the seven categories described above, 3 of the surveyed studies propose the following frameworks providing a collection of query primitives essential for geosocial queries:

1. J-CO framework [34] that provides a data model, an execution model, and a pool of operators (basic and spatial), which constitute the query language for querying heterogeneous collections of geo-referenced data and social network information.
2. GeoSocial-GraphX platform [12] that incorporates several query primitives (social, spatial and activity) essential for LBSN queries.
3. Socio-Spatial Network Algebra [77] that is composed of a set of seven operators that serve as the building blocks of a socio-spatial query language over a joined socio-spatial graph.

5.2. RQ 2: What Are the Query Processing Methods Applied to Geosocial Data by Selected Studies?

We addressed the second research question by analysing the kind of method(s) used to retrieve the result of the query, the kind of access method (if index-based or not), and whether or not they provide an approximate solution [82,83].

Considering the kind of query processing method, we checked the algorithms of the query processing proposed in the selected studies and we searched for the query primitives or algorithms described in Section 2.2. Based on our analysis, the most applied primitive in geosocial queries is pruning with 31 studies (57.4%), followed by sorting (15 studies—27.8%), scoring (14 studies—25.9%), clustering (8 studies—14.8%), filtering (6 studies—11.1%), and join and partitioning (1 study—1.8%). Considering the query algorithms, the most applied are the best first search algorithm and branch and bound with 6 studies each (11.1%), followed by measure and conquer (2 studies—3.7%), Dijkstra search, and depth-first search (1 study—1.8%).

Considering the kind of access method, the majority of the selected studies used an index-based approach (47 studies—87%) and only 7 studies (13%) do not use an index. The most applied class of indexing method is the spatial-first with 30 studies (63.8%), followed by the hybrid approach with 14 studies (29.8%) and the social-first with 3 studies (6.4%).

Finally, the majority of the selected studies do not provide an approximate solution (37—68.5%).

Table 17 summarises the selected studies with respect to the kind of query primitives/algorithms, access method, and indexing method they utilised.

Table 17. Query processing methods applied to the geosocial data in the selected studies.

ID	Kind of Query Primitives/Algorithms	Approximate Solution	Access Method	Index Name	Kind of Indexing Method
S1	-	-	-	-	-
S2	NA	no	non-index	-	-
S3	measure and conquer	no	non-index	-	-
S4	sorting, pruning	no	index	MRS-tree	hybrid
S5	sorting	no	index	MR-tree	spatial-first
S6	sorting, pruning	yes, ϵ -approximate Algorithm	index	R-tree	spatial-first
S7	filter	no	index	road network index IRN	hybrid

Table 17. Cont.

ID	Kind of Query Primitives/Algorithms	Approximate Solution	Access Method	Index Name	Kind of Indexing Method
S8	filter, incremental proximity search	no	index	Social-Equipped R-tree	spatial-first
S9	best-first search, pruning	yes, $\sqrt{3}$ -factor approximate algorithm	index	IR-tree	spatial-first
S10	join, sorting	teta-approximation algorithm	non-index	-	-
S11	scoring, filter	no	index	R-tree	spatial-first
S12	partitioning, filter	no	index	GIM-Tree	hybrid
S13	filter, branch and bound	no	index	SaRtree	hybrid
S14	pruning	no	index	IR and IS	spatial-first
S15	filter, scoring, pruning	yes, approximate shortest-path methods	index	spatial grid	spatial-first
S16	pruning	greedy solutions for approximation	index	R-tree	spatial-first
S17	scoring, pruning	no	index	CR-tree	spatial-first
S18	branch and bound/measure and conquer	no	non-index	-	-
S19	clustering, Dijkstra search	no	index	Watchtower	spatial-first
S20	sorting	no	non-index	-	-
S21	-	-	-	-	-
S22	clustering, pruning	no	index	SaR-tree	hybrid
S23	branch and bound, pruning	no	index	SaR-tree	hybrid
S24	scoring, pruning	no	index	SKR-tree	spatial-first
S25	branch and bound	no	non-index	-	-
S26	pruning	yes, social distance approximation	index	R-tree	spatial-first
S27	scoring, pruning	no	index	3D Friends Check-Ins R-tree	social-first
S28	scoring	no	index	B-tree	social-first
S29	pruning	no	non-index	-	-
S30	scoring, pruning	no	index	R-tree	spatial-first
S31	best-first search, pruning	no	index	R*-trees	spatial-first
S32	best-first search, pruning	yes, $\ln q.\psi $ -factor approximation	Index	IR-tree	spatial-first
S33	clustering, depth- first search	no	index	HI index	hybrid
S34	clustering, best-first search	no	index	TaR-tree	spatial-first
S35	scoring, pruning	no	index	tR-tree	spatial-first
S36	branch and bound	no	index	B+-Tree, Check-In R-Tree, Facility R-Tree	spatial-first

Table 17. Cont.

ID	Kind of Query Primitives/Algorithms	Approximate Solution	Access Method	Index Name	Kind of Indexing Method
S37	NA	no	index	R-tree	spatial-first
S38	scoring, pruning	yes, 3-approximation feasible result search algorithm	index	shortest-path tree	spatial-first
S39	best first search, pruning	no	index	NETR-tree	hybrid
S40	clustering, sorting, pruning	yes	index	GIM-tree	spatial-first
S41	scoring, pruning	yes, the approximate algorithm Unified-A	index	IR-tree	spatial-first
S42	branch and bound, sorting, pruning	no	index	Social R-Tree	social-first
S43	pruning	no	index	enhanced SaR-tree	hybrid
S44	clustering, scoring, sorting, pruning	no	index	I ³ ndex and nkIndex	hybrid
S45	clustering, best-first search	yes	index	IR-tree	spatial-first
S46	sorting, pruning	no	index	R*-tree	spatial-first
S47	sorting, pruning	no	index	R-tree	spatial-first
S48	NA	no	index	k-d tree and quadtree	spatial-first
S49	sorting, pruning	no	index	cd-tree	hybrid
S50	sorting, pruning	no	index	SIL-Quadtree	spatial-first
S51	scoring, pruning, sorting	no	index	FCRTree	hybrid
S52	sorting, pruning	no	index	BallTree	spatial-first
S53	clustering	no	index	R-tree	spatial-first
S54	scoring, pruning, sorting	no	index	IR-tree	spatial-first
S55	sorting, scoring	no	index	category-oracle inverted index	hybrid
S56	sorting, pruning	no	index	PIM-tree	hybrid
S57	-	-	-	-	-

5.3. RQ 3: How Are Geosocial Query Processing Methods Evaluated?

To answer this RQ, we identified 55 (96.5%) studies out of the selected studies that evaluated the proposed geosocial query processing methods, while two studies [34,51] do not provide any evaluation.

In the following sub-sections, we analyse both some important evaluation metrics used to assess the performance of geosocial query processing methods and the evaluation datasets.

5.3.1. Metrics

From the selected studies, we identified the following measures used to evaluate the performance of the query processing methods:

- Query response time, also named the query elapsed time or query processing time, which measures the time elapsed from the instant a query is issued to its result retrieval;
- Running time, also called the computation time, which is the length of time required to perform the query computational process;

- CPU time, which is the amount of time for which a central processing unit (CPU) is used for processing query instructions. According to what exactly the CPU is processing, this metric can be distinguished in client CPU time, which is the amount of time the CPU is busy executing client instructions, and the server CPU time, which is the amount of time the CPU is busy executing server instructions;
- Communication overhead, which is defined as the number of encrypted records sent as the result of an issued query [84];
- Correctness, which is the ratio between the number of the correct answers and the number of total queries;
- Accuracy, which is computed as the ratio between the cost functions of the result set obtained by the proposed query and the baseline solution [60];
- Index construction time, which can be defined as the time elapsed to construct the index structures [85];
- Approximation ratio, which is the usual way of measuring the performance of the query processing methods that provide approximate solutions and is computed as the ratio of the radius of approximate solution returned over that of the exact solution;
- I/O cost, which corresponds to the number of page/blocks accessed (I/O) to retrieve the data from the disk for each query;
- Pruning rate, which is computed as the ratio of the pruned PoIs to all the PoIs in the query range;
- Memory space, which is the total amount of memory used by the algorithm for query processing.

The most applied metric is the running time (43.9%), followed by I/O cost (26.3%), query response time (24.5%), and server CPU time (19.3%), as shown in Table 18.

Table 18. Metrics used by the selected studies.

Metrics	Paper ID	Total
Query response time/processing time	S2, S7, S8, S11, S16, S17, S19, S24, S31, S35, S39, S40, S53, S57	14
Running time	S3, S9, S10, S12, S13, S18, S22, S23, S25, S26, S28, S29, S30, S32, S38, S41, S42, S43, S44, S45, S46, S48, S49, S50, S52, S55	25
Server CPU time	S4, S5, S6, S14, S15, S27, S33, S34, S37, S47, S51	11
Client CPU time	S4, S5	2
Communication overhead	S4, S5	2
Correctness	S6	1
Accuracy	S7, S11, S33	3
Index construction time	S4, S19, S55, S56	4
Approximation ratio	S9, S32, S38, S41, S45	5
I/O cost	S12, S13, S14, S15, S22, S27, S28, S31, S34, S36, S39, S50, S51, S53, S54	15
Pruning rate	S17, S24, S30, S35	4
Memory space	S47, S55, S56	3

None of these metrics alone provides the perfect way to evaluate the query processing performance since each of them has limitations. This fact justifies the use of multiple metrics by the majority of the surveyed studies (62.5%).

5.3.2. Evaluation Datasets

As discussed by Brinkhoff [86], preparation and use of well-defined evaluation datasets are fundamental for enabling a systematic evaluation of the performance of query pro-

cessing algorithms and data structures. To achieve that, real-world and synthetic datasets have been used in the literature. The former are collected from real applications. The latter are generated by constructing a model that learns the statistical properties of the real data and using the model to produce the synthetic data, as well explained by Dankar and Ibrahim [87].

The selected studies used predominantly real-world datasets (56.1%—32 studies) to perform the evaluation of the geosocial query process, while 19 studies (33.3%) used both real-world and synthetic datasets and 2 studies (3.5%) used synthetic datasets only. The two remaining studies (S4 and S19) do not specify the datasets used for the evaluation. The predominant use of real-world datasets is probably due to the fact that they provide more realistic benchmarking results, even if the effort to record them can be very high compared to synthetic datasets.

Table 19 provides a summary of the real-world datasets used by the selected studies, along with their main characteristics; i.e. the size, which is the number of items (users, locations, vertices, objects, PoIs, etc.) collected in the dataset, and the sources, which are the location-based social network or the road network used to acquire the data. The most popular real-world dataset (with 23 studies or 41%) is the Gowalla dataset [88], which is available at the Stanford Large Network Dataset Collection (<http://snap.stanford.edu/data/index.html>, accessed on 22 December 2021) and contains 6,442,892 check-ins generated by 196,591 users at 1,280,969 locations worldwide from February 2009 to October 2010. The next most applied dataset is the Brightkite dataset with 10 studies (around 18%), followed by the Foursquare dataset with 7 studies (12.5%). Brightkite is available at the Stanford Large Network Dataset Collection (<http://snap.stanford.edu/data/index.html>, accessed on 22 December 2021) and contains 4,491,143 check-ins generated by 58,228 users at 772,789 locations. The Foursquare dataset is collected via the Foursquare API (<https://developer.foursquare.com/>, accessed on 22 December 2021) and, unlike the previous two datasets, it is not standardised as each study considers a different size (number of users) in their evaluation.

Table 19. Real-world datasets used by the selected studies.

Dataset	Paper ID	Size	Sources
Foursquare dataset	S2, S6, S43, S46, S48, S49, S55	12,652 users [S2] 20,550 users [S6, S48] 76,503 users [S43, S55] 87,229 users [S46] 2,153,471 users [S49]	Foursquare
Twitter dataset	S2, S22	2,220,627 users	Twitter
Gowalla dataset	S4, S5, S7, S12, S19, S20, S22, S23, S24, S25, S26, S27, S28, S34, S36, S37, S43, S48, S49, S50, S51, S55, S56	6,442,892 check-ins 1,280,969 locations 196,591 users	Gowalla, Stanford large network dataset collection
FB dataset	S7, S42	4039 vertices	Facebook
TW dataset	S7	17,069,982 vertices	Twitter
Brightkite dataset	S7, S12, S23, S24, S37, S43, S48, S49, S50, S55	4,491,143 check-ins 58,228 users	Brightkite
Orkut dataset	S7	3,072,441 vertices	Orkut
California road network dataset	S7, S31, S49, S53, S19	21,048 vertices 62,556 PoIs	California road network
San Francisco road network dataset	S7, S19	174,956 vertices	San Francisco road network
Florida road network dataset	S7, S29, S38	1,070,376 vertices	Florida road network
Western USA road network dataset	S7	6,262,104 vertices	Western USA road network
BE dataset	S8, S13	11,036 vertices	Brightkite in Europe
GE dataset	S8, S13	38,983 vertices	Gowalla in Europe

Table 19. Cont.

Dataset	Paper ID	Size	Sources
BA dataset	S8, S13	32,228 vertices	Brightkite in America
GA dataset	S8, S13	49,613 vertices	Gowalla in America
Hotel dataset	S9, S32, S40	20,790 objects	www.allstays.com , (accessed on 22 December 2021)
Web dataset	S9, S32, S40	579,727 objects	WEBSPAMUK2007 and TigerCensusBlock
GN dataset	S9, S32, S40, S45	1,868,821 objects	geonames.usgs.gov , (accessed on 22 December 2021)
Yahoo! Local Data Set	S10	909 locations	Yahoo! Local
Twitter + Instagram Data Set	S10	45,000,000 tweets and posts	Twitter + Instagram
LAS dataset	S11, S12	27,000 points	Yelp in Las Vegas
Yelp dataset	S39, S40, S56	99,798 objects 527,532 users	Yelp Dataset Challenge
Bri + Cal dataset	S14	61,000 vertices	Brightkite + California road network
Gow + Col dataset	S14	70,000 vertices	Gowalla + Colorado road network
NE dataset	S16	123,593 PoIs	TIGER project at the US Census Bureau
RR dataset	S16	257,942 PoIs	TIGER project at the US Census Bureau
CAS dataset	S16	196,902 PoIs	TIGER project at the US Census Bureau
Beijing dataset	S17, S35	607,307 PoIs	Beijing
Guangzhou dataset	S17	551,595 PoIs	Guangzhou
Dianping dataset	S22, S54	2,673,970 users	https://goo.gl/uUV4Wg , (accessed on 22 December 2021)
Twitter-2010	S22	41,652,098 users	Twitter
Flickr dataset	S29, S38, S44	68,776 users	Flickr
OpenStreetMap dataset	S33, S56	41,905 objects	OpenStreetMap
Weeplaces dataset	S39	99,378 objects 16,021 users	Weeplaces
NA dataset	S19, S40	175,813 vertices 58,228 users	North America road network
USA dataset	S40	3,598,623 vertices 81,306 users	United States road network
Large dataset	S42	153,577 users	Foursquare
Whrrl dataset	S46	4871 users	Whrrl
New York road network dataset	S49	264,346 vertices	New York road network
Northeast USA road network dataset	S49	1,524,453 vertices	Northeast USA road network
DataSet_4SQ	S52	153,577 users	Foursquare
Jiebang dataset	S54	353,493 users	Jiebang
New York City (NYC) dataset	S34	72,626 locations	Foursquare
Los Angeles (LA) dataset	S34	45,591 locations	Foursquare
GS dataset	S34	182,968 locations	Foursquare

5.4. RQ 4: Which Open Challenges in Geosocial Querying Have Been Envisaged?

The definition of the query processing methods applied to geosocial data brings many opportunities for research; however, there are also several open challenges that should be faced in the near future. Table 20 provides a summary of these issues that we have

extracted from the surveyed studies and opportunely divided into three main categories: technological challenges, privacy-related challenges, and social challenges.

Table 20. Open challenges envisaged by the selected studies.

	Open Challenges	ID
Technological	use of the shortest route, the interest of riders, obstacles on the road, and location uncertainty to enhance the query ridesharing system	S8
	use of the historical information of each user in the group to automatically setting the group preference and its weight	S17
	to allow each user to specify the minimum number of attendees with each attribute value required to be selected	S42
	empirical “relevance” assessment of the query results involving real-world data collected from the Web	S15
	to adopt deep learning technologies to train knowledge graphs of users, so as to intelligently perceive the preference information of a user community and choose the best POI	S35
	development of a corresponding index structure and various query algorithms, and the distributed implementation of a data model using a large-scale graph	S37
	to incorporate more sophisticated spatial queries such as skyline and distance-based joins	S22
	integration of methods to favor users whose friends are concentrated near the query and to investigate the adaptation of these methods to related application domains, such as spatial-keyword search	S25
Privacy-related	to study geo-social top-k collective keyword queries	S28
	to protect the location privacy of users while evaluating GTP queries	S31
	group planning over privacy-preserved or inconsistent spatial-social networks	S14
	to consider a user location as a region instead of a point that is desirable from the standpoint of privacy	S53
Social	to investigate the issue of social trust and how to integrate social trust into geo-social group query	S43
	to incorporate social relationships as an important criterion in group formation and develop novel query processing techniques	S54
	to study the evaluation of social trust in location-based social networks and to seek other approximate algorithms for solving this new problem	S55
	to investigate how other social information, such as social relationships between mobile users, can be utilized to speed up spatial query processing	S19

With respect to the technological challenges, the results of the SLR reveal a need to explore new kinds of social and spatial data to include in the query processing for refining the results of the geosocial queries. For instance, Shim et al. [39] suggested the use of the shortest route or the interest of riders to enhance the query ridesharing processing and to apply this kind of query also to environments with obstacles on the road and location uncertainty. Zhang et al. [47] proposed the use of the historical information of each user in the group to automatically set the group preference and its weight in the social graph. Furthermore, several works suggested to focus future research on the development of new approaches for (i) assessing the relevance of the query results, for instance, by using real-world data collected from the Web [45]; and (ii) training knowledge graphs, for instance, by using deep learning technologies to intelligently perceive the user community preference information and choose the best POI to retrieve [61]. In addition, a look at new kinds of geosocial queries is also suggested by the surveyed works. In particular, more sophisticated spatial queries, such as skyline and distance-based joins [52] and geosocial top-k collective keyword queries [23], are proposed.

Regarding the privacy-related challenges, some surveyed works highlighted the need for solutions to protect the users’ location privacy. Hashem et al. [58], for example, suggested to study scenarios where the group of users does not reveal their locations among

each other, and Ali et al. [73] proposed to consider a user location as a region instead of a point to avoid to disclose the precise location.

Finally, attending to the social challenges, future research needs to focus on the concept of social trust by investigating how social trust can be evaluated in location-based social networks [75] and how it can be integrated into geosocial query processing [66]. Moreover, future studies may even investigate how to incorporate other social information, such as the social relationships between mobile users, to develop novel query processing methods and speed up spatial query processing [49,74].

6. Conclusions

This study has examined the geosocial query processing in location-based social networks through a systematic literature review of the scientific knowledge extracted from indexed scientific databases, containing formally published literature, and from non-indexed databases, containing grey literature. Out of the 4312 papers returned from the initial search on these databases, 67 studies were retained after the application of the inclusion and exclusion criteria defined in the methodology, of which 57 were selected for the qualitative synthesis according to the scores obtained in the quality evaluation checklist.

We have found that the scientific community's interest in the topic of geosocial querying has started growing in 2012 and continued to grow till 2020. Furthermore, the result of our analysis shows that seven categories of geosocial queries can be identified: geosocial group queries proposed by 43.85% of the selected studies, followed by geosocial keyword queries (26.31%), geosocial top-k queries (19.3%), geosocial nearest neighbor queries (17.5%), geosocial skyline queries (10.5%), geosocial moving queries (5.26%), and geosocial fuzzy queries (1.75%). Moreover, three of the surveyed studies (5.26%) propose frameworks supporting a collection of query primitives essential for geosocial queries.

Regarding the query processing methods, we have observed that the kind of query primitive predominantly applied in the geosocial query process is pruning (57.4%), followed by sorting (27.8%), scoring (25.9%), clustering (14.8%), filtering (11.1%), join (1.8%), and partitioning (1.8%), while the most frequently used query algorithms are the best-first search algorithm (11.1%) and branch and bound (11.1%), followed by measure and conquer (3.7%), Dijkstra search (1.8%), and depth-first search (1.8%). Moreover, we found out that the majority of the selected studies used an index-based approach to optimize the retrieval of the geosocial data, and the spatial-first indexing method is the most common class of indexing methods (63.8%). Another key finding is that most of the selected studies (68.5%) do not provide an approximate solution, probably because it is preferable to have a completely accurate answer, even if through a more time-consuming process, instead of faster but not accurate approximate results.

Concerning the evaluation methodologies, we found out that one of the most common measures used to evaluate the performance of the query processing methods is running time (43.9%), followed by I/O cost (26.3%), the query response time (24.5%) and server CPU time (19.3%). Moreover, to perform the evaluation of the geosocial query process, real-world datasets are mainly used (56.1%), followed by both real-world and synthetic datasets (33.3%). The Gowalla dataset is the most popular real-world dataset applied by 41% of the selected studies.

Finally, the findings of the study highlight the need to explore (i) new kinds of social and spatial data to include in the query processing for refining the results of the geosocial queries; (ii) solutions to protect the location privacy of users; and (iii) methods for evaluating and integrating social trust into geosocial query processing.

Author Contributions: Conceptualization, Arianna D'Ulizia, Fernando Ferri and Patrizia Grifoni; methodology, Arianna D'Ulizia, Fernando Ferri and Patrizia Grifoni; validation, Fernando Ferri; formal analysis, Arianna D'Ulizia; investigation, Arianna D'Ulizia, Fernando Ferri and Patrizia Grifoni; data curation, Arianna D'Ulizia; writing—original draft preparation, Arianna D'Ulizia, Fernando Ferri and Patrizia Grifoni; writing—review and editing, Arianna D'Ulizia, Fernando Ferri. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data supporting reported results are available on the publicly archived dataset created on 4TU.ResearchData with the following DOI: 10.4121/17693705.

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

References

1. Kemp, S. Digital 2020 global overview report. Retrieved May 2020, 21, 2020.
2. Armenatzoglou, N.; Papadopoulos, S.; Papadias, D. A general framework for geo-social query processing. *Proc. VLDB Endow.* **2013**, *6*, 913–924. [\[CrossRef\]](#)
3. Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in location-based social networks: A survey. *Geoinformatica* **2015**, *19*, 525–565. [\[CrossRef\]](#)
4. Sahnoune, Z.; Yep, C.Y.; Aïmeur, E. Privacy Issues in Geosocial Networks. In *Risks and Security of Internet and Systems. CRiSIS 2014*; Lecture Notes in Computer Science; Lopez, J., Ray, I., Crispo, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; Volume 8924, pp. 67–82. [\[CrossRef\]](#)
5. Bilogrevic, I. Privacy in Geospatial Applications and Location-Based Social Networks. In *Handbook of Mobile Data Privacy*; Gkoulalas-Divanis, A., Bettini, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 195–228. [\[CrossRef\]](#)
6. Gunturi, V.M.V.; Brugere, I.; Shekhar, S. Modeling and Analysis of Spatiotemporal Social Networks. In *Encyclopedia of Social Network Analysis and Mining*; Alhajj, R., Rokne, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1–12. [\[CrossRef\]](#)
7. Quercia, D.; Lathia, N.; Calabrese, F.; Di Lorenzo, G.; Crowcroft, J. Recommending social events from mobile phone location data. In Proceedings of the International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 971–976.
8. Zheng, Y. Location-based social networks: Users. In *Computing with Spatial Trajectories*; Springer: New York, NY, USA, 2011; pp. 243–276.
9. Roick, O.; Heuser, S. Location Based Social Networks—Definition, Current State of the Art and Research Agenda. *Trans. GIS* **2013**, *17*, 763–784. [\[CrossRef\]](#)
10. Armenatzoglou, N.; Papadias, D. Geo-Social Networks. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: New York, NY, USA, 2018; pp. 1620–1623. [\[CrossRef\]](#)
11. Gao, H.; Liu, H. Data analysis on location-based social networks. In *Mobile Social Networking*; Springer: New York, NY, USA, 2013; pp. 165–194. [\[CrossRef\]](#)
12. Saleem, M.A.; Xie, X.; Pedersen, T.B. Scalable processing of location-based social networking queries. In Proceedings of the 17th IEEE International Conference on Mobile Data Management (MDM), Porto, Portugal, 13–16 June 2016; Volume 1, pp. 132–141.
13. Pearl, J. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*; Addison-Wesley: Boston, MA, USA, 1984; p. 48.
14. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*, 2nd ed.; Section 22.3: Depth-first search; MIT Press: Cambridge, MA, USA; McGraw-Hill: London, UK, 2001; pp. 540–549. ISBN 0-262-03293-7.
15. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [\[CrossRef\]](#)
16. Land, A.H.; Doig, A.G. An automatic method of solving discrete programming problems. *Econometrica* **1960**, *28*, 497–520. [\[CrossRef\]](#)
17. Fomin, F.V.; Grandoni, F.; Kratsch, D. Measure and Conquer: Domination—A Case Study. In Proceedings of the 32nd International Colloquium on Automata, Languages and Programming, Lisbon, Portugal, 11–15 July 2005; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3580, pp. 191–203.
18. Duan, X.; Wang, Y.; Chen, J.; Zhang, J. Authenticating preference-oriented multiple users spatial queries. In Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Torino, Italy, 4–8 July 2017; Volume 1, pp. 602–607.
19. Zhao, J.; Gao, Y.; Ma, C.; Jin, P.; Wen, S. On efficiently diversified top-k geo-social keyword query processing in road networks. *Inf. Sci.* **2019**, *512*, 813–829. [\[CrossRef\]](#)
20. Sun, Y.; Qi, J.; Zheng, Y.; Zhang, R. K-Nearest Neighbor Temporal Aggregate Queries. In Proceedings of the 18th International Conference on Extending Database Technology, Brussels, Belgium, 23–27 March 2015. [\[CrossRef\]](#)
21. Cao, K.; Sun, Q.; Liu, H.; Liu, Y.; Meng, G.; Guo, J. Social space keyword query based on semantic trajectory. *Neurocomputing* **2020**, *428*, 340–351. [\[CrossRef\]](#)
22. Yang, D.N.; Shen, C.Y.; Lee, W.C.; Chen, M.S. On socio-spatial group query for location-based social networks. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 949–957.
23. Attique, M.; Afzal, M.; Ali, F.; Mehmood, I.; Ijaz, M.F.; Cho, H.-J. Geo-Social Top-k and Skyline Keyword Queries on Road Networks. *Sensors* **2020**, *20*, 798. [\[CrossRef\]](#)

24. Sohail, A.; Cheema, M.A.; Taniar, D. Geo-Social Temporal Top-k Queries in Location-Based Social Networks. In Proceedings of the Australasian Database Conference, Melbourne, Australia, 3–7 February 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 147–160. [\[CrossRef\]](#)
25. Yang, Z.; Gao, Y.; Gao, X.; Chen, G. NETR-Tree: An Efficient Framework for Social-Based Time-Aware Spatial Keyword Query. *arXiv* **2019**, arXiv:1908.09520.
26. Li, Q.; Zhu, Y.; Yu, J.X. Skyline Cohesive Group Queries in Large Road-social Networks. In Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 20–24 April 2020; pp. 397–408.
27. Li, Y.; Chen, R.; Xu, J.; Huang, Q.; Hu, H.; Choi, B. Geo-Social K-Cover Group Queries for Collaborative Spatial Computing. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 2729–2742. [\[CrossRef\]](#)
28. Li, Y. *Efficient Group Queries in Location-Based Social Networks*. Semantic Scholar. 2016. Available online: <https://www.semanticscholar.org/paper/Efficient-group-queries-in-location-based-social-Li/edd525bbaed1aa4ae97066364e84298e2327f087> (accessed on 22 December 2021).
29. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [\[CrossRef\]](#)
30. Mahood, Q.; Van Eerd, D.; Irvin, E. Searching for grey literature for systematic reviews: Challenges and benefits. *Res. Synth. Methods* **2013**, *5*, 221–234. [\[CrossRef\]](#)
31. Paez, A. Grey literature: An important resource in systematic reviews. *J. Evid. Based Med.* **2017**, *10*, 233–240. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Haddaway, N.R.; Collins, A.; Coughlin, D.; Kirk, S.A. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS ONE* **2015**, *10*, e0138237. [\[CrossRef\]](#)
33. Yasin, A.; Fatima, R.; Wen, L.; Afzal, W.; Azhar, M.; Torkar, R. On Using Grey Literature and Google Scholar in Systematic Literature Reviews in Software Engineering. *IEEE Access* **2020**, *8*, 36226–36243. [\[CrossRef\]](#)
34. Bordogna, G.; Capelli, S.; Psaila, G. A Big Geo Data Query Framework to Correlate Open Data with Social Network Geotagged Posts. In *The Annual International Conference on Geographic Information Science*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 185–203. [\[CrossRef\]](#)
35. Huang, C.-Y.; Chien, P.-C.; Chen, Y.H. A Measure and Conquer Algorithm for the Minimum User Spatial-Aware Interest Group Query Problem. In *International Computer Symposium*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 440–448. [\[CrossRef\]](#)
36. Wang, Y.; Hassan, A.; Duan, X.; Zhang, X. An efficient multiple-user location-based query authentication approach for social networking. *J. Inf. Secur. Appl.* **2019**, *47*, 284–294. [\[CrossRef\]](#)
37. Liu, W.; Sun, W.; Chen, C.; Huang, Y.; Jing, Y.; Chen, K. Circle of friend query in geo-social networks. In *International Conference on Database Systems for Advanced Applications*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 126–137.
38. Guo, F.; Yuan, Y.; Wang, G.; Chen, L.; Lian, X.; Wang, Z. Cohesive Group Nearest Neighbor Queries on Road-Social Networks under Multi-Criteria. *IEEE Trans. Knowl. Data Eng.* **2020**, *33*, 3520–3536. [\[CrossRef\]](#)
39. Shim, C.; Sim, G.; Chung, Y.D. Cohesive Ridesharing Group Queries in Geo-Social Networks. *IEEE Access* **2020**, *8*, 97418–97436. [\[CrossRef\]](#)
40. Long, C.; Wong, R.C.W.; Wang, K.; Fu, A.W.C. Collective spatial keyword queries: A distance owner-driven approach. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 22–27 June 2013; pp. 689–700.
41. Kanza, Y.; Shalem, M. Combined geo-social search: Computing top-k join queries over incomplete information. *GeoInformatica* **2017**, *22*, 615–660. [\[CrossRef\]](#)
42. Maropaki, S.; Chester, S.; Doukeridis, C.; Nøravåg, K. Diversifying Top-k Point-of-Interest Queries via Collective Social Reach. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, Ireland, 19–23 October 2020; pp. 2149–2152.
43. Jin, P.; Gao, Y.; Chen, L.; Zhao, J. Efficient Group Processing for Multiple Reverse Top-k Geo-Social Keyword Queries. In *International Conference on Database Systems for Advanced Application*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 279–287. [\[CrossRef\]](#)
44. Al-Baghdadi, A.; Sharma, G.; Lian, X. Efficient Processing of Group Planning Queries Over Spatial-Social Networks. *IEEE Trans. Knowl. Data Eng.* **2020**, 2093–2098. [\[CrossRef\]](#)
45. Efstathiades, C.; Efentakis, A.; Pfoser, D. Efficient Processing of Relevant Nearest-Neighbor Queries. *ACM Trans. Spat. Algorithms Syst.* **2016**, *2*, 1–28. [\[CrossRef\]](#)
46. Islam, S.; Shen, B.; Wang, C.; Taniar, D.; Wang, J. Efficient processing of reverse nearest neighborhood queries in spatial databases. *Inf. Syst.* **2020**, *92*, 101530. [\[CrossRef\]](#)
47. Zhang, Z.; Jin, P.; Tian, Y.; Wan, S.; Yue, L. Efficient Processing of Spatial Group Preference Queries. In Proceedings of the International Conference on Database Systems for Advanced Applications, Chiang Mai, Thailand, 22–25 April 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 642–659.
48. Huang, C.Y.; Chien, P.C.; Chen, Y.H. Exact and Heuristic Algorithms for Some Spatial-aware Interest Group Query Problems. *J. Internet Technol.* **2020**, *21*, 1199–1205.
49. Tang, L.; Chen, H.; Ku, W.-S.; Sun, M.-T. Exploiting location-aware social networks for efficient spatial query processing. *GeoInformatica* **2017**, *21*, 33–55. [\[CrossRef\]](#)

50. Zheng, S.; Zaman, A.; Morimoto, Y. Friend Recommendation by Using Skyline Query and Location Information. *Bull. Netw. Comput. Syst. Softw.* **2016**, *5*, 68–72.
51. Chen, S.-M.; Randyanto, Y.; Cheng, S.-H. Fuzzy queries processing based on intuitionistic fuzzy social relational networks. *Inf. Sci.* **2016**, *327*, 110–124. [[CrossRef](#)]
52. Zhu, Q.; Hu, H.; Xu, C.; Xu, J.; Lee, W.-C. Geo-social group queries with minimum acquaintance constraints. *VLDB J.* **2017**, *26*, 709–727. [[CrossRef](#)]
53. Taguchi, N.; Amagata, D.; Hara, T. Geo-social keyword Skyline queries. In Proceedings of the International Conference on Database and Expert Systems Applications, Lyon, France, 20–31 August 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 425–435.
54. Armenatzoglou, N.; Ahuja, R.; Papadias, D. Geo-Social Ranking: Functions and query processing. *VLDB J.* **2015**, *24*, 783–799. [[CrossRef](#)]
55. Emrich, T.; Franzke, M.; Mamoulis, N.; Renz, M.; Züfle, A. Geo-social skyline queries. In Proceedings of the International Conference on Database Systems for Advanced Applications, Bali, Indonesia, 21–24 April 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 77–91.
56. Zhao, S.; Xiong, L. Group nearest compact POI set queries in road networks. In Proceedings of the 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, China, 10–13 June 2019; pp. 106–111.
57. Tian, Y.; Jin, P.; Wan, S.; Yue, L. Group preference queries for location-based social networks. In Proceedings of the Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, Beijing, China, 7–9 July 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 556–564.
58. Hashem, T.; Hashem, T.; Ali, M.E.; Kulik, L. Group trip planning queries in spatial databases. In Proceedings of the International Symposium on Spatial and Temporal Databases, Munich, Germany, 21–23 August 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 259–276.
59. Chan, H.K.H.; Long, C.; Wong, R.C.W. Inherent-cost aware collective spatial keyword queries. In Proceedings of the International Symposium on Spatial and Temporal Databases, Arlington, VA, USA, 21–23 August 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 357–375.
60. Wang, Y.; Duan, X.; Yang, X.; Zhang, Y.; Zhang, X. Interactive Multiple-User Location-Based Keyword Queries on Road Networks. *IEEE Access* **2018**, *6*, 51401–51418. [[CrossRef](#)]
61. Wang, Y.; Zhu, L.; Ma, J.; Hu, G.; Liu, J.; Qiao, Y. Knowledge Graph-Based Spatial-Aware User Community Preference Query Algorithm for LBSNs. *Big Data Res.* **2020**, *23*, 100169. [[CrossRef](#)]
62. Sohail, A.; Hidayat, A.; Cheema, M.A.; Taniar, D. Location-Aware Group Preference Queries in Social-Networks. In Proceedings of the Australasian Database Conference, Goald Coast, Australia, 24–27 May 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 53–67. [[CrossRef](#)]
63. Zhang, H.; Lu, F.; Xu, J. Modeling and Querying Moving Objects with Social Relationships. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 121. [[CrossRef](#)]
64. Zhao, S.; Cao, X. Multiple-user closest keyword-set querying in road networks. *Inf. Sci.* **2019**, *509*, 133–149. [[CrossRef](#)]
65. Chan, H.K.H.; Long, C.; Wong, R.C.W. On generalizing collective spatial keyword queries. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 1712–1726. [[CrossRef](#)]
66. Ma, Y.; Yuan, Y.; Wang, G.; Bi, X.; Wang, Y. Personalized geo-social group queries in location-based social networks. In Proceedings of the International Conference on Database Systems for Advanced Applications, Goald Coast, Australia, 21–24 May 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 388–405.
67. Zhao, S.; Cheng, X.; Su, S.; Shuang, K. Popularity-aware collective keyword queries in road networks. *GeoInformatica* **2017**, *21*, 485–518. [[CrossRef](#)]
68. Wang, Y.; Duan, X.; Yang, X.; Zhang, Y.; Zhang, X. Processing Multiple-User Location-Based Keyword Queries. *IEICE Trans. Inf. Syst.* **2018**, *101*, 1552–1561. [[CrossRef](#)]
69. Upreti, N. Reverse Nearest Social Group Query. Master's Thesis, Electronic Theses and Dissertations for Graduate School, Pennsylvania State University, State College, PA, USA, 2015.
70. Allheeib, N.; Taniar, D.; Al-Khalidi, H.; Islam, S.; Adhinugraha, K.M. Safe Regions for Moving Reverse Neighbourhood Queries in a Peer-to-Peer Environment. *IEEE Access* **2020**, *8*, 50285–50298. [[CrossRef](#)]
71. Sohail, A.; Cheema, M.A.; Taniar, D. Social-Aware Spatial Top-k and Skyline Queries. *Comput. J.* **2018**, *61*, 1620–1638. [[CrossRef](#)]
72. Shen, C.-Y.; Yang, D.-N.; Huang, L.-H.; Lee, W.-C.; Chen, M.-S. Socio-Spatial Group Queries for Impromptu Activity Planning. *IEEE Trans. Knowl. Data Eng.* **2015**, *28*, 196–210. [[CrossRef](#)]
73. Ali, M.E.; Tanin, E.; Scheuermann, P.; Nutanong, S.; Kulik, L. Spatial consensus queries in a collaborative environment. *ACM Trans. Spat. Algorithms Syst.* **2016**, *2*, 1–37. [[CrossRef](#)]
74. Li, Y.; Wu, D.; Xu, J.; Choi, B.; Su, W. Spatial-aware interest group queries in location-based social networks. *Data Knowl. Eng.* **2014**, *92*, 20–38. [[CrossRef](#)]
75. Ma, Y.; Yuan, Y.; Wang, G.; Bi, X.; Qin, H. Trust-Aware Personalized Route Query Using Extreme Learning Machine in Location-Based Social Networks. *Cogn. Comput.* **2018**, *10*, 965–979. [[CrossRef](#)]
76. Zhao, J.; Gao, Y.; Chen, G.; Chen, R. Why-not questions on top-k geo-social keyword queries in road networks. In Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 16–19 April 2018; pp. 965–976.

77. Doytsher, Y.; Galon, B.; Kanza, Y. Querying geo-social data by bridging spatial networks and social networks. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, San Jose, CA, USA, 2 November 2010; pp. 39–46.
78. Apon, S.H.; Ali, M.E.; Ghosh, B.; Sellis, T. Social-Spatial Group Queries with Keywords. *ACM Trans. Spat. Algorithms Syst.* **2021**, *8*, 1–32. [\[CrossRef\]](#)
79. Borzsony, S.; Kossmann, D.; Stocker, K. The skyline operator. In Proceedings of the 17th international conference on data engineering, Heidelberg, Germany, 2–6 April 2001; pp. 421–430.
80. Chen, F.; Lu, C.-T. Nearest Neighbor Query, Definition. In *Encyclopedia of GIS*; Shekhar, S., Xiong, H., Eds.; Springer: Boston, MA, USA, 2008; pp. 782–783. [\[CrossRef\]](#)
81. Ben Hassine, M.A.; Touzi, A.G.; Galindo, J.; Ounelli, H. How to Achieve Fuzzy Relational Databases Managing Fuzzy Data and Metadata. In *Handbook of Research on Fuzzy Information Processing in Databases*; IGI Global: Hershey, PA, USA, 2008; pp. 351–380. [\[CrossRef\]](#)
82. D’Ulizia, A.; Ferri, F.; Formica, A.; Grifoni, P. Approximating Geographical Queries. *J. Comput. Sci. Technol.* **2009**, *24*, 1109–1124. [\[CrossRef\]](#)
83. D’Ulizia, A.; Ferri, F.; Grifoni, P.; Rafanelli, M. Relaxing constraints on GeoPQL operators for improving query answering. In Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA’06), Krakow, Poland, 4–8 September 2006; Lecture Notes in Computer Science 4080. Springer: Berlin/Heidelberg, Germany, 2006; pp. 728–737.
84. Moghadam, S.S.; Fayoumi, A. Toward Securing Cloud-Based Data Analytics: A Discussion on Current Solutions and Open Issues. *IEEE Access* **2019**, *7*, 45632–45650. [\[CrossRef\]](#)
85. Thoombayil Asokan, U. Methods for Evaluating Query Auto Completion Systems. Ph.D. Thesis, Minerva Access, University of Melbourne, Parkville, Australia, 2021.
86. Brinkhoff, T. Real and Synthetic Test Datasets. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer Science+Business Media LCC: New York, NY, USA, 2009; pp. 2339–2344. [\[CrossRef\]](#)
87. Dankar, F.K.; Ibrahim, M. Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation. *Appl. Sci.* **2021**, *11*, 2158. [\[CrossRef\]](#)
88. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.