Review

Toward Trust-Based Recommender Systems for Open Data: A Literature Review

Chenhao Li [ID]**, Jiyin Zhang, Amruta Kale, Xiang Que, Sanaz Salati and Xiaogang Ma *** [ID]

Department of Computer Science, University of Idaho, Moscow, ID 83844-1010, USA;
li7809@vandals.uidaho.edu (C.L.); zhan4354@vandals.uidaho.edu (J.Z.); kale3781@vandals.uidaho.edu (A.K.);
xiangq@uidaho.edu (X.Q.); ssalati@uidaho.edu (S.S.)
* Correspondence: max@uidaho.edu; Tel.: +1-208-885-1547

Abstract: In recent years, the concept of "open data" has received increasing attention among data providers and publishers. For some data portals in public sectors, such as data.gov, the openness enables public oversight of governmental proceedings. For many other data portals, especially those in academia, open data has shown its potential for driving new scientific discoveries and creating opportunities for multidisciplinary collaboration. While the number of open data portals and the volume of shared data have increased significantly, most open data portals still use keywords and faceted models as their primary methods for data search and discovery. There should be opportunities to incorporate more intelligent functions to facilitate the data flow between data portals and end-users. To find more theoretical and empirical evidence for that proposition, in this paper, we conduct a systematic literature review of open data, social trust, and recommender systems to explain the fundamental concepts and illustrate the potential of using trust-based recommender systems for open data portals. We hope this literature review can benefit practitioners in the field of open data and facilitate the discussion of future work.

1. Introduction

Over the past decades, "open data" has been widely discussed by researchers and practitioners in various disciplines and sectors. For instance, we can see trends that government data has been increasingly made open and used [1]. In academia, big data, artificial intelligence (AI), machine learning, and data science has recently drawn a lot of attention in many disciplines. Most of them, if not all, have data as the foundation. Many domain-specific studies, such as those in biology, geology, socioeconomics, and space science, are deploying those technologies together with open data to accelerate scientific discoveries. For example, the Landsat images were made free and open access in 2008, which has led to a huge increase in the number of scientific publications in recent years [2].

Nevertheless, compared with the methods and technologies in big data, artificial intelligence, machine learning, and data science, open data is treated more like a campaign to shift the culture of data sharing and then build better accessibility for data. In a report on open data released by the International Science Council [3], it was argued that data must be "intelligently open", so they can be thoroughly scrutinized and appropriately re-used. While data providers are starting to adopt the culture shift of openness and scientific communities are making recommendations on the best practices of open data [4], most data portals are still using keywords or faceted search models as their main approach for data search and discovery. Given the trend that open data will play an increasingly important role in science and society, there is a lot of room for developing more intelligent and efficient methods to help researchers find and access data of interest.

In recent years, recommender systems have been introduced into many platforms, such as those in social media and e-commerce. The goal of recommender systems is to

provide potential information or products that might be of interest to a specific consumer. Platforms such as Amazon, Netflix, and YouTube all have their own recommender systems and associated algorithms. Intuitively, we would say that using recommender systems in open data portals is a potential solution to improve the efficiency of data discovery. For example, recommender systems can help researchers receive newest information in their discipline even they have not actively conducted searching, just like Mendeley sending users feeds of new publications related to their reading and searching history. Moreover, social trust has a strong relationship with the research of recommender systems, as it is used as an important metric in drawing recommendations. We propose that social trust can also be considered within the work of recommender systems for open data. Nevertheless, we still need to clarify the detailed interconnections between those three concepts before we draw work plans for technical development. A survey of existing publications seems to be a good way to meet that need.

In this paper, we perform a literature review of existing research on open data, recommender systems, and social trust, with the intention to illustrate both the relationships and gaps between these three domains and discuss directions for future work. We collected articles on Scopus by using a combination of keywords to search their title, abstract, and authors' keyword. In total, we obtained 1161 articles that were published between 2007 and early 2022. We only collected articles published after 2007 because that was when the study of social trust on social networks started to appear. We conducted bibliometric analyses of the collected articles to illustrate the patterns and trends of the research reported in them, and we also incorporated the review of a few other publications in the discussion for future work.

The remainder of this paper is organized as follows. Section 2 briefly explains the concepts of open data, social trust, and recommender system. Section 3 presents the steps and results of the bibliometric analyses. Section 4 discusses the patterns and trends by synthesizing the results of bibliometric analyses and a few other publications, and then gives a vision on topics for future research. Finally, Section 5 concludes the paper.

2. Open Data, Social Trust, and Recommender System

2.1. Open Data and Associated Concepts

The phrase "open data" first appeared in the early 1990s. For example, in a 1992 report released by NASA, NOAA, and USGS for the Global Change Data and Information System [5], a list of data management policy statements was drafted, and the purpose was to "facilitate full and open access to quality data for global change research". Since the mid-2000s, open data has gained more attention and action. In 2007, OECD released the Principles and Guidelines on Access to Research Data from Public Funding [6]. In 2013, the G8 leaders signed the Open Data Charter, which establishes five principles that all G8 members will implement [7]. A general definition of open data in those publications is that a part of data should be made open to everyone to use, re-use, and redistribute. A more comprehensive understanding is that open data must be considered from both technical and legal/ethical aspects [8–10]. The legal/ethical aspect means that there are legal and ethical frameworks to enable users to obtain the data, use it, and share the derived result. The technical aspect means that there should be no technical barriers for accessing and using the data, such as the common transmission system (e.g., the Internet), non-proprietary format, and standard terminologies. The FAIR (Findable, Accessible, Interoperable, and Reusable) open data principles [4] are a good representation of those aspects and have been well received among the open data practitioners. Findable means that data can be found using their assigned globally unique and persistent identifiers. Accessible means that the user can easily access the data. Interoperable means that data are formed in easily understandable language. Last, reusable means that users can easily use the data for their specific needs.

Among the many technical approaches for open data, the work on semantic web and knowledge graphs is specifically noteworthy. A central idea of semantic web and knowledge graph [11–14] is to add machine-readable structures (i.e., semantics) to data. The associated studies, such as linked open data [15,16] and five-star open data (1: on the web, 2: machine-readable, 3: non-proprietary format, 4: RDF (Resource Description Framework) standards, and 5: linked RDF) [17], provide many building blocks toward the foundation of the above-mentioned FAIR open data, such as clearly defined objects and relationships, unique identifiers, rich metadata, standard vocabularies, non-proprietary data formats, and more. The recently released Google Dataset Search engine [18] also has a strong relationship to semantic web and knowledge graphs as its foundation is the Schema.org, which provides metadata schemas to markup datasets of different subjects on the Web.

The open data movement has been thriving across different sectors for the past two decades. For example, the United States launched data.gov in 2009 and United Kingdom launched data.gov.uk in 2010, respectively, to publish open governmental data. The report released by the International Science Council in 2015 listed the progress and best practices of open data in several regions and countries, such as South America, Africa, China, and India [3]. The web portal Open Data Barometer [19] actively monitors the open data actions in 30 countries that have adopted the Open Data Charter [7], and gives scores based on several metrics, such as readiness, implementation, and emerging impact. In academia, there have also been many remarkable progresses in open data across different disciplines. A working group in the World Wide Web Consortium has summarized the best practices of publishing and using data on the Web [20], where a list of examples can be accessed. Besides the governmental and academic sectors, there are also crowd-sourcing open data movements among the general public. For example, during the Haiti earthquake in 2010, over 600 volunteers from the global OpenStreetMap community quickly enriched the map of Haiti to help local organizations respond to the crisis [21].

Although the open data movement has achieved impressive achievements across various sectors and organizations, the methods for data discovery and access on many open data portals have limited functionality, and there is room for improvement. For example, most open data portals, including those mentioned in Lóscio et al. [20], still only provide keyword and faceted search-ability on their user interfaces. While they can quickly return a large number of data search results to a user, there is still uncertainty in the match between the results and the user's specific needs. Many interesting research topics can arise from here, such as trustworthiness of search results, ranking of the search results based on multiple metrics, and personalized recommendation.

2.2. Social Trust

Ever since Myspace, the first social media website that reached a million monthly active users in 2004, human society has begun a new era where people use social media for their daily interactions with others [22]. For instance, according to Statista [23], Facebook has reached 2.93 billion monthly active users as of the first quarter of 2022. As social media websites become increasingly popular, many organizations also use them as platforms for advertising and recommending their products. Many users have the question on what and who they can trust on the social media websites as well as other platforms. Here is where the research topic of social trust computation arises. The term "social trust" generally means one person's expectation that another will behave in a particular way [24]. In the context of social media and the Internet, social trust is understood as a group of metrics to measure the trustworthiness of a certain user, a product, or a piece of information [25]. For the computation of social trust, many scientists have developed models or algorithms that calculate a person's trust score based on several aspects, such as relationship, common interest, and social status. Similar to the trust score of people, scientists have also developed models to determine whether a website, a product or a piece of information on the Internet can be trusted or not. A good example is the platform scamadvisor.com. It calculates the

trust score of a certain website by using both positive and negative indicators. The former includes popularity, social media activity, positive reviews, performance, and security of the website. The latter includes high-risk country of the website location, website ownership, website age, high-risk server, e-commerce platform, and more.

2.3. Recommender System

With the rise of online services, such as the shopping website Amazon, video streaming site YouTube, and many more during the past two decades, recommender systems have increasingly affected people's web browsing experience. In short, the goal of a recommender system is to determine a user's preference and recommend contents that the user may potentially be interested in. Developing a good recommender system is crucial for a website to increase its customer stickiness. According to MacKenzie et al. [26], product suggestions account for 35% of what customers buy on Amazon and 75% of what they watch on Netflix. In a real-world situation, customers can go to a store and tell a salesperson their preference. In turn, the salesperson can give recommendations based on his/her knowledge and the customers' preference. If the customers like the recommendation and enjoy the shopping experience, there will be a higher chance that they will come back for shopping again. Similar to that situation, the recommender system's purpose is to imitate this kind of interaction in an online environment.

In detail, recommender systems have three major paradigms: collaborative filter-based, content-based, and a hybrid approach [27,28]. The collaborative filter-based (CF) method only takes past records of user-item interaction as input. Usually, the user-item interaction records can be transformed into a matrix (Table 1). Then, systems can use the matrix to determine similar users and items, and then recommend new items based on those findings. The CF method can be further divided into two sub-methods called memory-based and model-based CF. The memory-based CF relies heavily on the user-item matrix, and it includes user-user and item-item methods. The user–user method will first calculate the similarity of users based on the rating they give on the same items and then divide those users into different groups based on similarity. For each group, the method will recommend popular items that are new to some of the group members. In comparison, the item–item method will use items as its main input. First, it will find an item that a certain user has given the highest rating. Second, it will find the rating scores of this item from all users. Third, it will find a list of other items with similar performance in rating scores. Fourth, it will recommend this list of items to the user in the first step. The model-based CF, as Rocca [29] mentioned, assumes that a latent model will explain the interaction between users and items. The advantage of the CF method is that it requires no information about the users or items because it is solely based on user–item interactions. The limitation of this method is that it suffers from a "cold start", when there is no user–item interaction recorded. Nevertheless, there are some ways to bypass the cold start stage, such as assigning random recommendations to new users. Thorat et al. [30] also discussed other limitations of the CF method.

Table 1. An exemplar user–item interaction matrix. In the table, u_i represents the user and i_i is for the item. The values in the matrix in the table are the user rating for each item (e.g., user1 gives item1 a rating score of 5).

	i_1	i_2	i_3
u_1	5	3	2
u_2	4	5	4
u_3	4	2	1

Unlike the CF method, which only relies on the user–item interaction matrix, the content-based method uses more information about users and items to develop recommendations. For example, people of different ages tend to buy different products when they visit their local mall. Kids are more likely to buy toys and candy, while adults are more likely to buy clothes. The content-based method can be further divided into two

approaches: item-centered and user-centered [29]. The main purpose of the item-centered approach is to train a model for a given item based on the attributes of users who have had interactions with it. Then, for a new user, this model can make predictions on his rating of this item. Similarly, the user-centered method will train a model for a certain user based on the attributes of items that the user has interacted with. Then, for a new item, this model can make predictions on the user's rating. Compared with the CF method, the content-based method also suffers from the issue of cold start, but it has significant improvement due to the incorporation of user and item attributes.

The hybrid approach is a combination of more than one filtering method, with the purpose to address some limitations of the other methods, such as cold start, overspecialization, and sparsity [30].

3. Bibliometric Analyses of Recent Publications

3.1. Data Source and Tools for Analysis

Bibliometric analysis is a useful method for assessing the impact of publications in a certain field of study. In our work, the objective of the bibliometric analysis is to illustrate both the relationships and gaps between open data, social trust, and recommender systems in existing publications, and discuss directions for the future work. We chose Scopus as the main database of literature in this bibliometric analysis as it covers a wide range of scientific articles across different sources and gives formatted metadata about articles (e.g., indexed keywords).

We conducted several rounds of queries to Scopus, using different combinations of keywords to search the title, abstract, and keywords of existing articles. During the initial Scopus query, we found out that there are very few results that include all the three keywords "open data", "social trust", and "recommender systems" in the same article. Due to this insufficiency, we chose to use alternative words and combinations of those three keywords to expand the scope of the query. Additionally, we focused on articles published in or after 2007 because that was when the study of social trust computation started to appear. The following string (Listing 1) shows the exact query used in our work. We ran the query on 10 March 2022 and obtained records of 1661 articles from Scopus. A copy of the retrieved literature records was stored at this GitHub repository [31].

Listing 1. Query Codes.

```
(TITLE-ABS-KEY ("open data") AND TITLE-ABS-KEY ("recommender system"))
OR (TITLE-ABS-KEY ("open data")
AND TITLE-ABS-KEY ("trustworthy"))
OR (TITLE-ABS-KEY ("trust")
AND TITLE-ABS-KEY ("recommender system"))
AND PUBYEAR > 2006
```

For the retrieved literature records, we chose VOSviewer [32] and Bibliometrix [33] as the main analysis and visualization tools. VOSviewer is a program for creating and visualizing bibliometric networks. These networks can be built via citation, bibliographic coupling, co-citation, or co-authorship relationships, and the networks can be further extended to include records of journals, researchers, or individual articles. Text mining capabilities are also included in VOSviewer, which may be used to create and visualize cooccurrence networks of other relevant terms retrieved from a corpus of scientific literature [32]. Bibliometrix is an open-source application for quantitative research of scientometrics and bibliometrics, which contains all the common methods of bibliometric analysis [33]. It can import bibliographic data from websites, such as Scopus, and construct data matrices for analyses of co-citation, coupling, co-word, scientific collaboration, and more.

3.2. Results of Bibliometric Analysis

In the data cleansing and pre-processing, we discovered that there are some duplicate terms in the authors' keywords. For example, there are many occurrences of "recommender system", "recommender systems", and "recommendation system". As they mean the same concept, we reconciled those terms into a single keyword, "recommender system", for the convenience of our analysis. Similar operations were also taken to several other keywords.

We conducted several analyses to the cleansed datasets, including keyword frequency, density and centrality, timeline, and keyword co-relationship. The following sections will illustrate the most representative results.

3.2.1. Timeline Analysis

Figure 1 shows the linear plot for annual article production from 2007 to the present based on the 1661 articles we retrieved from Scopus. The graph illustrates that over the last one and a half decades, the number of articles relevant to "recommender system", "trust", and "open data" has steadily increased. The drop in 2022 is mainly because we only had a partial record for that year. The diagram in Figure 2 shows the cumulative growth of authors' keywords among the 1661 articles. In this figure, recommender system, CF and trust are ranked top three, and linked open data is ranked at the sixth place. The keyword "open data" is not shown in Figure 2 as it is ranked low at the 14th place (26 records by 2022). From the diagram, we can see the rapid growth of articles relevant to "recommender systems" and "trust" over the past one and a half decades, but the growth articles relevant to "linked open data" and "open data" is significantly lower.
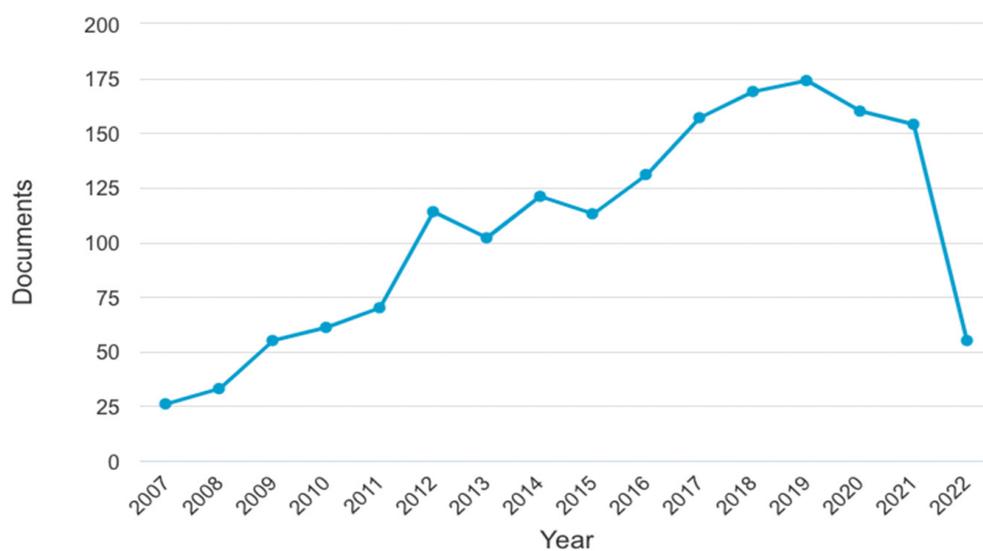

Figure 1. Linear plot for annual article numbers between 2007 and 2022.
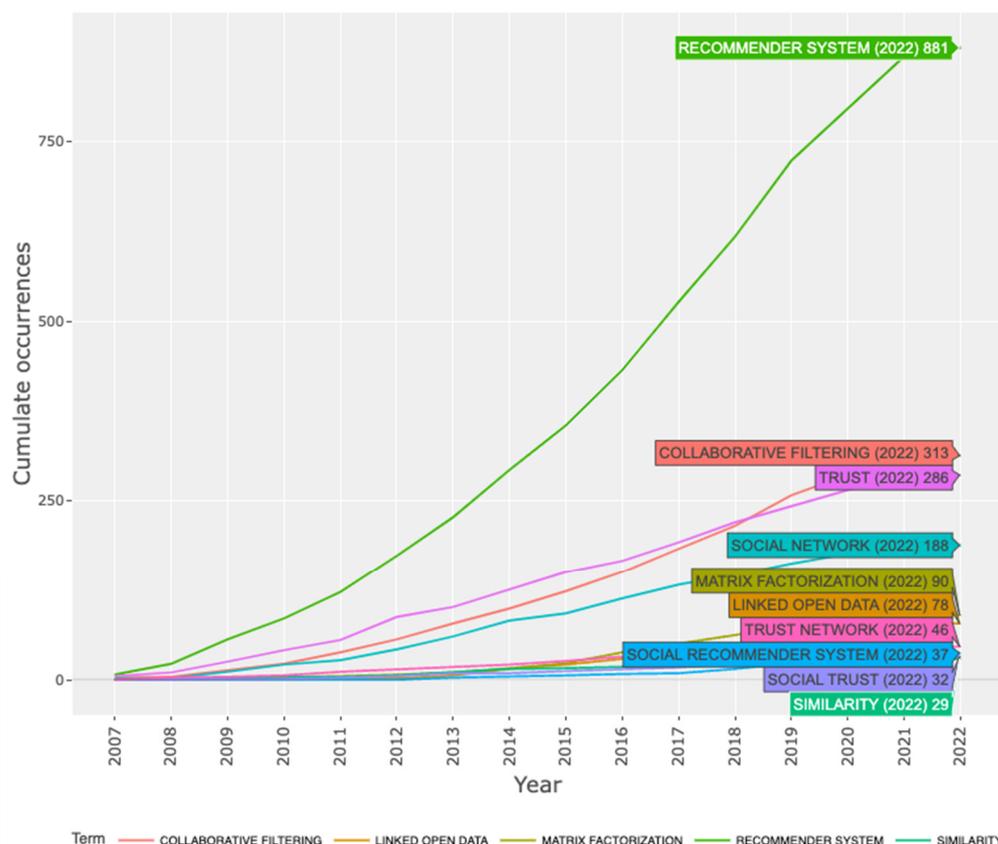
Figure 2. Cumulative trend of authors' keywords among the 1661 articles.

3.2.2. Keyword Co-Relationship Analysis

Co-word analysis is a method for analyzing keyword co-occurrences, identifying linkages and interactions between the topics under study, and exploring potential research trends [34]. Figure 3 is the keyword co-relationship map in our result. The nodes represent the top 23 authors' keywords from the 1661 articles, in which the lowest count of keyword occurrence is 20. Those 23 keywords were grouped into five major clusters on the map, as depicted by the color of the nodes. The width of edges on the map represents the frequency of co-occurrence between two keywords. It is apparent that there are strong relationships between "recommender system", "trust", "collaborative filtering", and "social network". In contrast, the relationship between "open data" (on the right of the map) and the other keywords is much weaker.

The results of the bibliometric analysis show an increasing trend of studies on social trust and recommender systems. They also illustrate that among the existing publications there are limited studies on using social trust and recommender systems for open data. Nevertheless, this gap may also mean there is a big potential to explore in that direction. In the next section we will investigate more details about the technical approaches of social trust, recommender systems, and open data, and discuss the emerging research topics.
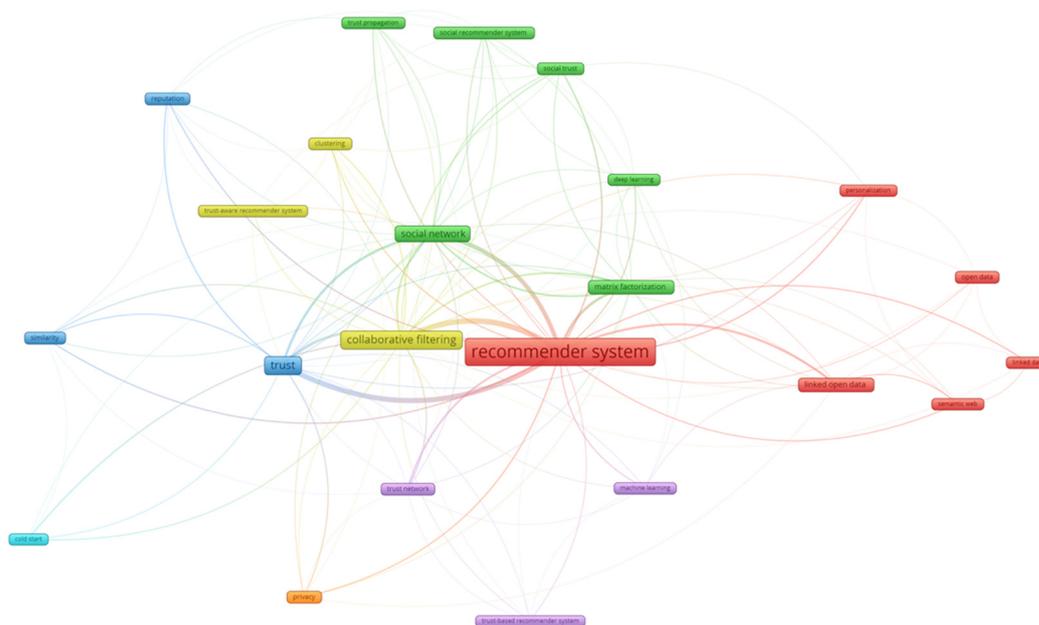
Figure 3. Analysis of keyword co-relationship based on authors' keywords.

4. Discussion of Trends, Challenges, and Future Works

After conducting the bibliometric analysis and reviewing the collected articles in detail, we realized that the general topics from articles about "open data", "recommender system", and "social trust" can fall into two major categories: "trust and recommender system" and "open data and recommender system". Nevertheless, from Figure 3, we also noticed that even though the recommender system and open data can be put in a same cluster (i.e., nodes in red color), their relevance is minimal (i.e., the width of edges between the red nodes is very narrow). In comparison, we found strong connections between the recommender system and trust (i.e., wide edges in Figure 3). This might be caused by the massive research on trust-based recommender systems in the literature. In the sections below, we will analyze those patterns in more detail and offer a vision on research trends and future works.

4.1. Social Trust and Recommender System

As Figure 3 illustrates, most keywords in the left part of the graph, such as "trust", "social network", "social trust", and "trust network", are related to the theme of social trust. Due to the co-relationship between those keywords in the 1661 articles, they are shown in several clusters in the graph (i.e., the nodes in blue, green, and purple colors). Two other keywords, "cold start" and "privacy", are rendered in different colors, due to the unique co-relationships, but they are also within the scope of social trust and recommender system research. Another noteworthy cluster in Figure 3 is the nodes in yellow color, including the keywords "trust-aware recommender system", "collaborative filtering", and "clustering". They represent the core research topics in the intersection of social trust and recommender systems in the collected literature. After further reading and analyses of the literature, we realized that social trust in recommender systems can further fall into two categories: trust-based recommender system and trustworthiness of recommender system. The latter is not directly shown in Figure 3. The two subsections below will give more details about them. While the literature search results show that there are limited intersections between those two topics and open data, they have a big potential to be incorporated into open data portals, and thus, it is necessary to analyze the latest studies in each topic.

4.1.1. Trust-Based Recommender System

As shown in the result of bibliometric analysis, research and interest in trust-based recommender systems have been increasing in recent years. Trust-based recommender system is a recommender system based on trust metrics. The trust metrics in those recommender systems can vary case by case, but they tend to follow the concepts and mechanism of social trust. For example, Ozsoy and Polat [35] suggested using a trust network to enhance the accuracy and efficiency of the recommender system based on the theory that a person tends to trust the recommendations from other people that they already know and trust. Similarly, Shokeen and Rana [36] suggested using social rating networks to determine the outcome of a recommender system, but their work went deeper into the layers of the social network. For instance, direct and indirect friends (i.e., friends of friends) have different degrees of trust in the recommender system. Other researchers have also applied user behavior or features as a trust metric in a recommender system. For example, Rrmoku et al. [37] suggested using a user's social profile (i.e., gender, age, hobbies) and touristic preferences as a trust metric. A common feature of those studies is that even though different ideas, definitions, and configurations of trust metrics are used, their attempts are similar, which is using some aspects of social trust as the foundation in the design of their recommender systems. In real-world applications, trust-based recommender systems have shown the potential to increase the accuracy and efficiency of the standard recommending method. For example, according to Peng and Chou [38], trust-based recommender systems can help alleviate the "cold-start" issue in standard CF methods and generate better accuracy.

4.1.2. Trustworthiness of Recommender System

Another topic of interest that arose from our literature reading is the trustworthiness of recommender system. Recommender systems have also long struggled to receive consumers' trust due to many reasons, such as faulty information generated by profile injection or human misbehavior [39]. As Jha et al. [40] pointed out, the growth of the recommender system since its conception shows that a reliable, efficient, and effective recommender system is a pressing requirement. The literature showed that there is active research on improving the robustness of recommender systems against biased data or intentional human-made faulty data. For example, Xue et al. [41] suggested an iterative methodology for calculating the total trustworthiness of all reviewers in a system and using it to predict the possibility of someone being a review spammer (e.g., people writing fake reviews to either promote or demote certain products or services). Similarly, Stitini et al. [42] proposed innovative studies on improving trust and transparency in recommendation systems by detecting fake news on social networks.

Besides filtering biased and faulty data in a recommender system to improve its accuracy and trustworthiness, other approaches have also been studied to improve users' confidence in the recommender system. As Torkamaan et al. [43] pointed out, in addition to classic rating-based preference elicitation, recommender systems frequently incorporate implicit user preferences gathered from behavioral and contextual data to improve the quality and accuracy of personalized recommendations. These tactics, on the other hand, may detract from the user experience by eliciting mixed feelings such as dread, anxiety, surprise, discomfort, or creepiness. Accordingly, there have been studies to reduce users' anxiety and improve their comfortableness and stickiness with a recommender system. For example, Zarzour et al. [44] suggested creating an explainable recommender system to increase the transparency of the workflow. Giving a brief explanation to the user of what and how certain information is recommended to them will help them gain confidence. To address users' unwillingness to provide personal information due to privacy concerns, Parvathy et al. [45] suggested an efficient privacy protection method by fusing Principal Component Analysis and Rotation Transformation in a trust-based recommender system.

4.2. Open Data and Recommender System

In the bibliometric analysis, we noticed that while the co-relationship between the recommender system and social trust is strong, the connection between the recommender system and open data is weak (Figure 3). We looked through the literature and found that, among the 1661 articles we collected, there are only 26 occurrences of the term "open data" in the authors' keywords. Half of those occurrences are in the articles about open data and trust. For example, Wiencierz and Lünich [46] suggested guidelines for transparent communication for open data applications to overcome the concern about privacy violations. The other half of the 26 occurrences are about open data and recommender systems. Most of them are about using linked open data to solve different issues and needs in recommender systems. As shown in the graph of Figure 3, five keywords, "open data", "linked data", "semantic web", "linked open data", and "personalization", are in the same cluster for open data (i.e., nodes in red color). Among those five keywords, "linked open data", which is five-star open data [17], has a relatively stronger co-relationship with "recommender system". Linked open data is a field of specific interest to many researchers as it incorporates many state-of-the-art technologies in data representation and sharing. According to Bizer [16], the term "Linked Open Data" refers to all data that is published on the Web in accordance with the Linked Data Principles. The goal underlying these principles is to realize both standardized representation of data and linkages between data sources on the Web. Similar to how hyperlinks connect all webpages into a single global information space on the conventional Web, these linkages connect all linked data into a single global data graph. In our understanding, the structured representation of linked data is a big factor as to why there are relatively more studies between it and recommender systems, compared with other keywords of open data. For example, Yochum et al. [47] presented an overview of applications using linked open data to create a location-based recommender system.

However, we noticed that there are very few articles about building and using recommender systems for open data, although this field has caught the attention from some researchers. For example, Devaraju and Berkovsky [48] suggested using user features as weights on open data recommendation. In another study, Sornkongdang et al. [49] created a data category recommendation framework called DataCat to help data providers publish their data in the correct category on an open data portal. Obviously, more studies can be conducted to incorporate social trust and recommender systems into the open data ecosystem, to extend the methods of data discovery and access from keyword and faceted search to more intelligent approaches. For example, the metadata and provenance in the FAIR data principles [4] and the quality information in data documentation [50] might be potential input for a recommender system.

4.3. Discussion of Potential Future Works

As indicated in our bibliometric analyses and detailed review, there have been active studies between social trust and recommender systems but limited work between open data and recommender systems or between open data and social trust. However, the gap also means a big potential to apply the mature work of social trust and recommender systems in the emerging open data ecosystem. Below is a list of potential studies based on our understanding.

Open data is closely related to various objects, such as data provider, data portal, broker, user, search engine, data analysis tool, scientific domain, and more. Those objects and the interactions between them form the so-called open data ecosystem [51,52]. Although recently the number of open data portals has been increasing quickly, most data portals are still not able to provide advanced functionalities, such as context analysis, user preference prediction, and personalized recommendations. Despite the popularity of open data, many users still rely on using keywords and faceted models as the major method for data searching. We suggest studies on creating recommender systems to be used in open data portals. This will help users discover datasets more suitable to their interests and needs. For a simple example, a researcher types in keyword, "tick", on a data portal. The

recommender system on the portal finds a dataset named "climate change and disease", in which (1) there are records about tick-borne diseases, (2) several keywords of the dataset match with the search history of this researcher, and (3) the data citation records show that several collaborators of this researcher have used this dataset before. Then, the system will give a higher rank to this dataset in the search results returned to the user. In another mechanism, a recommender system will also be able to analyze the profile and research interests of a certain user and suggest other data that match the user's background. For example, a data portal can send a notification to the user when a new dataset matching the user's search history is published. In this way, the recommender system can act as a powerful assistant to show users the broad information of data available in their discipline, and it can even spark innovative ideas or opportunities for multi-disciplinary collaboration that are not easily discoverable to users.

Due to the characteristics of the open data, such as openness, reusability, and abundant metadata, there is a potential for creating and implementing both trust- and content-based recommender systems for open data. Here, an intuitive approach is to check whether the metadata elements of open data can be used as input metrics for recommender systems. In our work, we compared some of the metadata elements in Schema.org/Dataset and the Google Dataset Search engine [18] with the needs of trust- and content-based recommender systems, and listed their utility in Table 2.

Table 2. Utility of Schema.org/Dataset metadata elements in trust- or content-based recommender systems.

Metadata Element	Utility in Trust-Based Recommender System	Utility in Content-Based Recommender System
description	Y	Y
name	N	Y
creator	Y	Y
citation	Y	Y
funder	Y	Y
hasPart	N	N
identifier	Y	N
isAccessibleForFree	N	Y
keywords	N	Y
license	Y	Y
measurementTechnique	Y	Y
sameAs	N	N
spatialCoverage	Y	Y
temporalCoverage	Y	Y
variableMeasured	Y	Y
version	Y	N
url	Y	N

As demonstrated in Table 2, several metadata elements such as creator, citation, funder, spatialCoverage, temporalCoverage, and identifier can be used directly as trust metrics for trust-based recommender systems. Moreover, some metadata elements such as keywords, and description can be used as item features in content-based recommender systems. The analyses presented in this paper show that the current technical framework of open data makes it feasible to develop trust- and content-based recommender systems. Those systems, once established, will be a valuable addition to the current open data portals and will facilitate more fluent and efficient communications and workflows between data providers, portals, and users in the open data ecosystem.

The technical framework of open data is also under active discussion and extension, among which many components can be further leveraged in recommender systems. For example, provenance can help solve issues such as accountability and authenticity of data [53]. The provenance information, if documented in detail, can also be used as

trust metrics in recommender systems. In a recent article, Peng et al. [50] discussed the community guidelines for documenting quality information of open data. Such information is currently missing in most data portals. If such guidelines are widely implemented, the documented information can also be additional input to recommender systems.

5. Conclusions

As open data is increasingly accepted and implemented across different sectors, there are also needs for more intelligent and efficient technologies in data discovery and access. This study presents a systematic literature review of existing works on recommender systems, social trust, and open data. Records of 1661 publications were collected from Scopus. The bibliometric analyses show that there are very active studies between social trust and recommender systems, but there is limited work between open data and recommender systems or between open data and social trust. That gap also means there are opportunities, and this has been a major driving force for us to write this paper to call attention from the community. In the discussion, we analyzed the trends of studies among those three domains and gave more details on the comparison of technologies. Our general understanding is that the abundant and mature studies on recommender systems and social trust can be adapted to address the needs of intelligent technologies for open data. At the end of the discussion, we also gave a few suggestions for future work. We hope this literature review illustrates the landscape of studies on open data, social trust, and recommender systems, and we expect to see more works on trust- and content-based recommender systems to be created for open data.

References

1. Jäger, B.; Bartenberger, M.; Leitner, P. A framework for semantic business process management in e-government. In Proceedings of the IADIS International Conference WWW/INTERNET 2013, Fort Worth, TX, USA, 22–25 October 2013; pp. 363–367.
2. Zhu, Z.; Wulder, M.A.; Roy, D.P.; Woodcock, C.E.; Hansen, M.C.; Radeloff, V.C.; Healey, S.P.; Schaaf, C.; Hostert, P.; Strobl, P.; et al. Benefits of the free and open Landsat data policy. *Remote Sens. Environ.* **2019**, *224*, 382–385. [CrossRef]
3. Science International. *Open Data in a Big Data World*; Science International: Paris, France, 2015.
4. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 1–9. [CrossRef] [PubMed]
5. NASA; NOAA; USGS. *Global Change Data and Information System (GCDIS): A Draft Tri-Agency Implementation Plan*; National Aeronautics and Space Administration (NASA); National Oceanic and Atmospheric Administration (NOAA); U.S. Geological Survey (USGS): Washington, DC, USA, 1992.
6. OECD. *OECD Principles and Guidelines for Access to Research Data from Public Funding*; OECD Publishing: Paris, France, 2007.
7. G8. G8 Open Data Charter and Technical Annex. 2013. Available online: https://opendatacharter.net/g8-open-data-charter/ (accessed on 19 May 2022).
8. European Union. Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data. 2010. Available online: https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data (accessed on 19 May 2022).
9. Ma, X.; Asch, K.; Laxton, J.L.; Richard, S.M.; Asato, C.G.; Carranza, E.J.M.; van der Meer, F.D.; Wu, C.; Duclaux, G.; Wakita, K. Data exchange facilitated. *Nat. Geosci.* **2011**, *4*, 814. [CrossRef]

10. International Science Council. Open Science for the 21st Century. 2020. Available online: https://council.science/publications/open-science-for-the-21st-century/ (accessed on 19 May 2022).

11. Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]

12. Gutiérrez, C.; Sequeda, J.F. Knowledge graphs. *Commun. ACM* **2021**, *64*, 96–104. [CrossRef]

13. Hitzler, P. A review of the semantic web field. *Commun. ACM* **2021**, *64*, 76–83. [CrossRef]

14. Chaudhri, V.; Baru, C.; Chittar, N.; Dong, X.; Genesereth, M.; Hendler, J.; Kalyanpur, A.; Lenat, D.; Sequeda, J.; Vrandečić, D.; et al. Knowledge Graphs: Introduction, History and, Perspectives. *AI Mag.* **2022**, *43*, 17–29.

15. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*; IGI Global: Hershey, PA, USA, 2011; pp. 205–227.

16. Bizer, C.; Vidal, M.E.; Skaf-Molli, H. Linked open data. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: New York, NY, USA, 2018. [CrossRef]

17. Berners-Lee, T. Linked Data Design Issues. Available online: https://www.w3.org/DesignIssues/LinkedData.html (accessed on 19 May 2022).

18. Brickley, D.; Burgess, M.; Noy, N. Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In Proceedings of the The World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 1365–1375.

19. Open Data Barometer. The Open Data Barometer: A Global Measure of How Governments are Publishing and Using Open Data for Accountability, Innovation and Social Impact. 2022. Available online: https://opendatabarometer.org (accessed on 19 May 2022).

20. Loscio, B.F.; Burle, C.; Calegari, N. (Eds.) Data on the Web Best Practices. 2017. Available online: https://www.w3.org/TR/dwbp/ (accessed on 19 May 2022).

21. Radford, T. Haiti 10 Years Later: Growth of a Humanitarian Mapping Community. 2020. Available online: https://www.hotosm.org/updates/haiti-10-years-later-growth-of-a-crisis-mapping-community/ (accessed on 19 May 2022).

22. Ortiz-Ospina. The Rise of Social Media. 2019. Available online: https://ourworldindata.org/rise-of-social-media (accessed on 19 May 2022).

23. Statista. Number of Monthly Active Facebook Users Worldwide as of 1st Quarter 2022. Available online: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ (accessed on 19 May 2022).

24. Verducci, S.; Schröe, A. Social Trust. In *Encyclopedia of Database Systems*; Springer: New York, NY, USA, 2010. [CrossRef]

25. Golbeck, J. *Computing with Social Trust*; Springer: London, UK, 2009.

26. MacKenzie, I.; Meyer, C.; Noble, S. How Retailers Can Keep Up with Consumers. 2018. Available online: https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers (accessed on 19 May 2022).

27. Isinkaye, F.; Folajimi, Y.; Ojokoh, B. Recommendation systems: Principles, methods and evaluation. *Egypt. Inform. J.* **2015**, *16*, 261–273. [CrossRef]

28. Althbiti, A.; Ma, X. Collaborative Filtering. In *Proceedings of the Encyclopedia of Big Data*; Schintler, L.A., McNeely, C.L., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 1–4. [CrossRef]

29. Rocca, B. Introduction to Recommender Systems. 2019. Available online: https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada (accessed on 19 May 2022).

30. Thorat, P.B.; Goudar, R.M.; Barve, S. Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System. *Int. J. Comput. Appl.* **2015**, *110*, 31–36.

31. Li, C. Scopus Publication Records for a Literature Review on Recommender System, Social Trust, and Open Data. 2022. Available online: https://github.com/CHenhao-lI1995/lit-record-2022 (accessed on 2 June 2022).

32. Perianes-Rodriguez, A.; Waltman, L.; Van Eck, N.J. Constructing bibliometric networks: A comparison between full and fractional counting. *J. Informetr.* **2016**, *10*, 1178–1195. [CrossRef]

33. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **2017**, *11*, 959–975. [CrossRef]

34. Chen, C. *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

35. Ozsoy, M.G.; Polat, F. Trust based recommendation systems. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Niagara, ON, Canada, 25–28 August 2013. [CrossRef]

36. Shokeen, J.; Rana, C. A trust and semantic based approach for social recommendation. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 10289–10303. [CrossRef]

37. Rrmoku, K.; Selimi, B.; Ahmedi, L. An Approach of Utilizing Exponential Rank and In-Inverse Closeness Centrality on Recommender Systems. In Proceedings of the 2021 International Conference on Information Technologies (InfoTech), Varna, Bulgaria, 16–17 September 2021. [CrossRef]

38. Peng, T.C.; Chou, S.-c.T. iTrustU: A blog recommender system based on multi-faceted trust and collaborative filtering. In Proceedings of the 2009 ACM Symposium on Applied Computing, Honolulu, HI, USA, 8–12 March 2009. [CrossRef]

39. Dong, M.; Yao, L.; Wang, X.; Xu, X.; Zhu, L. Adversarial dual autoencoders for trust-aware recommendation. *Neural Comput. Appl.* **2021**. [CrossRef]

40. Jha, G.K.; Gaur, M.; Ranjan, P.; Thakur, H.K. A survey on trustworthy model of recommender system. *Int. J. Syst. Assur. Eng. Manag.* **2021**. [CrossRef]

41. Xue, H.; Li, F.; Seo, H.; Pluretti, R. Trust-Aware Review Spam Detection. In Proceedings of the 2015 IEEE Trust-com/BigDataSE/ISPA, Helsinki, Finland, 20–22 August 2015. [CrossRef]

42. Stitini, O.; Kaloun, S.; Bencharef, O. Towards the Detection of Fake News on Social Networks Contributing to the Improvement of Trust and Transparency in Recommendation Systems: Trends and Challenges. *Information* **2022**, *13*, 128. [CrossRef]

43. Torkamaan, H.; Barbu, C.M.; Ziegler, J. How can they know that? A study of factors affecting the creepiness of recommendations. In Proceedings of the 13th ACM Conference on Recommender Systems, Copenhagen, Denmark, 16–20 September 2019. [CrossRef]

44. Zarzour, H.; Jararweh, Y.; Al-Sharif, Z.A. An Effective Model-Based Trust Collaborative Filtering for Explainable Recommendations. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 238–242. [CrossRef]

45. Parvathy, M.; Sundarakantham, K.; Shalinie, S.M.; Dhivya, C. An efficient privacy protection mechanism for recommendation using hybrid transformation technique. In Proceedings of the 2014 Sixth International Conference on Advanced Computing (ICoAC), Chennai, India, 17–19 December 2014. [CrossRef]

46. Wiencierz, C.; Lünich, M. Trust in open data applications through transparency. *New Media Soc.* **2020**. [CrossRef]

47. Yochum, P.; Chang, L.; Gu, T.; Zhu, M. Linked Open Data in Location-Based Recommendation System on Tourism Domain: A Survey. *IEEE Access* **2020**, *8*, 16409–16439. [CrossRef]

48. Devaraju, A.; Berkovsky, S. Do users matter?: The contribution of user-driven feature weights to open dataset recommendations. In Proceedings of the Poster Track of the 11th ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, 27–31 August 2018.

49. Sornkongdang, N.; Sanglerdsinlapachai, N.; Anutariya, C. DataCat: Attention-based Open Government Data (OGD) Category Recommendation Framework. In Proceedings of the 2021 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Ayutthaya, Thailand, 21–23 December 2021. [CrossRef]

50. Peng, G.; Lacagnina, C.; Downs, R.; Ganske, A.; Ramapriyan, H.; Ivánová, I.; Wyborn, L.; Jones, D.; Bastin, L.; Shie, C.; et al. Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets. *Data Sci. J.* **2022**, *21*. [CrossRef]

51. Lindman, J.; Kinnari, T.; Rossi, M. Business roles in the emerging open-data ecosystem. *IEEE Softw.* **2015**, *33*, 54–59. [CrossRef]

52. Welle Donker, F.; van Loenen, B. How to assess the success of the open data ecosystem? *Int. J. Digit. Earth* **2017**, *10*, 284–306. [CrossRef]

53. Kale, A.; Nguyen, T.; Harris, F.C.; Li, C.; Zhang, J.; Ma, X. Provenance documentation to enable explainable and trustworthy AI: 568 A literature review. *Data Intell.* **2022**, 1–41. [CrossRef]