

Article

A Bibliometric Analysis of the Use of Artificial Intelligence Technologies for Social Sciences

Tuba Bircan ^{1,*} and Almila Alkim Akdag Salah ²

¹ Interface Demography, Department of Sociology, Vrije Universiteit Brussel, Pleinlaan 5, 1050 Brussels, Belgium

² Human-Centered Computing Group, Department of Information and Computing Sciences, Utrecht University, Buys Ballotgebouw (BBG) 422, Princetonplein 5, 3584 CM Utrecht, The Netherlands

* Correspondence: tuba.bircan@vub.be

Abstract: The use of Artificial Intelligence (AI) and Big Data analysis algorithms is complementary to theory-driven analysis approaches and becoming more popular also in social sciences. This paper describes the use of Big Data and computational approaches in social sciences by bibliometric analyses of articles indexed between 2015 and 2020 in Social Sciences Citation Index (SSCI) of the Web of Science repository. We have analysed especially the recent research direction called Computational Social Sciences (CSS) that bridges computer analytical approaches with social science challenges, generating new methodologies of Big Data and AI analytics for social sciences. The results indicate that AI and Big Data practices are not confined to CSS only and are diffused in a wide variety of disciplines under Social Sciences and are made use of in many main research lines as well. Thus, the anticipated overlap between the Social Sciences & AI specialization and CSS has yet to be crystallised. Moreover, the impact of computational social science studies is not permeated to social science citation networks yet. Lastly, we demonstrate that the AI and Big Data publications that appear under the SSCI index are more oriented towards computational studies than addressing social science concepts, concerns, and challenges.



Citation: Bircan, T.; Salah, A.A.A. A Bibliometric Analysis of the Use of Artificial Intelligence Technologies for Social Sciences. *Mathematics* **2022**, *10*, 4398. <https://doi.org/10.3390/math10234398>

Academic Editors: Santiago Gutiérrez-Broncano, Benito Yáñez-Araque, Pedro Jiménez Estevez and Felipe Hernández-Perlines

Received: 11 October 2022

Accepted: 11 November 2022

Published: 22 November 2022

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



Copyright: © 2022 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/>).

Keywords: big data; artificial intelligence; computational social science; social sciences; bibliometrics

MSC: 00A06

1. Introduction

The concept of Big Data is gaining popularity at every aspect of life given the influence of technology on individuals and governance. The United Nations report Big Data for Development: Challenges and Opportunities points out Big Data as a data revolution and states that “these new data can provide snapshots of the well-being of populations at high frequency, high degrees of granularity, and from a wide range of angles, narrowing both time and knowledge gaps” [1] (p. 6). Hence, along with the recent efforts in explainable AI, it is assumed that the use of Big Data applications in social sciences will be accelerated. Nevertheless, to this date, no study thoroughly examined this assumption. Through this paper, we aim at answering the following questions:

- What is the evolving trend of Big Data and Artificial Intelligence (AI) in the field of (computational) social science? Within social science research, what are the main patterns of topic (keyword) distribution in the field of Big Data and AI?
- Which disciplines and journals are leading and promoting the utilisation of Big Data and AI in social sciences?
- What suggestions can be offered to improve future Big Data research in social sciences?

On one hand, increase in the amount of social Big Data attracted the attention of Social Science researchers to the promising uses of Big Data and AI methodologies in complementing traditional data sources and research methods for various applications such as text and

language analysis, network analysis, simulation and predictive analytics [2–5] as well as potential threats and ethical concerns [6–9]. On the other hand, AI researchers started using insights from social sciences for examining explainable AI [10], future social capacities of AI [11]. These paradigm shifts in scientific research methods prompts new directions for research [12] and emphasise the urgent need for engagement and collaboration of scholars from both AI and social sciences fields [13].

The diffusion of AI methodologies to social sciences is evident in the rise of the new research line called Computational Social Sciences, however, the impact of these studies and the scientific value and strength of social science research working with big data and AI methods have not yet been quantified or visualised. Building on the major goal of assess scholarly influence of big data and AI in social sciences, the novelty of this paper within the social sciences framework is threefold: (1) We will measure and evaluate big data and AI research output; (2) We will visualise the scholarly influence of specific subjects and (3) We will scrutinise the complexity of impact of computational social sciences through citation networks analysis.

Accordingly, the paper is organized as follows: in the next sub-sections, we offer a state-of-the-art overview of how Big Data and Artificial Intelligence related methodologies and concepts are used in social science academic literature. Section on Materials and Methods explains the data source and the methodology used for the study. To understand the diffusion of the concepts, as well as the application and theorisation of Artificial Intelligence and Big Data analytics in/for social sciences studies, we analyse articles that are indexed in the last five year in academic indices, using above all the Web of Science repository. The findings are discussed in the Results section, which starts with the analysis of computational social science literature with a focus on co-citation patterns and co-occurrence of author keywords over the years. Then we depict the distribution of most productive and influential disciplines and the journals for the publications of AI analytics and Big Data within social sciences. Conclusions are drawn in the last section.

1.1. Big Data and AI Applications in Social Sciences

The applications of advanced modelling in applied social science have increased, with a gradual shift towards data science with the growing availability of Big Data. Having said that, rather than the data availability, Big Data applications in social science, such as machine learning, enable a promising new “culture” of statistical modelling for the social scientist [14]. Statistical and computational methods and quantitative techniques are currently being fully exploited in numerous social science disciplines, including sociology, political science, and public administration (see [15–17]), as well as mathematical sciences (see [18–20]). Furthermore, the distinctiveness of both social science disciplines and variety of topics including Big Data and AI are also examined [21]. More importantly, computational social science (CSS) emerged as a new field focusing on how to incorporate computational approaches to social science methodologies, as well as on research ethics, interdisciplinary studies, data collection and visualisation [22]. CSS is an interdisciplinary approach to analyse the social dynamics of society by virtue of advanced computational systems from a data/information driven perspective [23]. CSS is not yet being accepted as a discipline on its own and still awaiting its potential to be realised.

Big Data applications have been used widely for commercial reasons and several methodologies are adopted for different disciplines to achieve high relevance and impact amid changes and transformations in how we study social science phenomena. Nevertheless, there is no consensus on the wide variety of Big Data conceptualisation [24,25].

Following Laney [26] (2001)’s definition of Big Data in terms of volume, variety, and velocity, Big Data is defined also as data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse [27] as unique datasets including “higher level of detail and refinement in the quality of observations, not just the number of data points or the amount of memory that their storage takes” [28] (p. 148); and as large amounts of different types of data produced from various types of sources,

such as people, machines or sensors [29]. Moreover, Iliadis and Russo [30] (2016) argue that when Big Data is considered as a “modern archive of data facts and data fictions”, the cultural, ethical, and critical perspectives should also be taken into consideration (p. 1). Taking varied approaches, in this study, within the social sciences framework, the working definition of Big Data is considered as extremely large data sets that may be analysed through advanced computational methods to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

1.2. Analysing Big Data and AI Literature with a Bibliometrics Approach

Bibliometrics refers to a process of evaluating and predicting the status quo and development trends of sciences and technology using mathematical, statistical, and network analysis as measurement methods [31]. The popularity of the emerging topics of the digital era, such as Big Data and Artificial Intelligence, can be different in various disciplines and previous research show that these topics benefit from both within- and outside-field citation links [32,33]. Several scholars employed bibliometrics for investigating the use and spread of Big Data and Artificial Intelligence in scientific works. In Figure 1, as of 2015 a drastic rise is visible in the number of research on Big Data and Artificial Intelligence for all WoS indexed publications from 1993 to 2020.

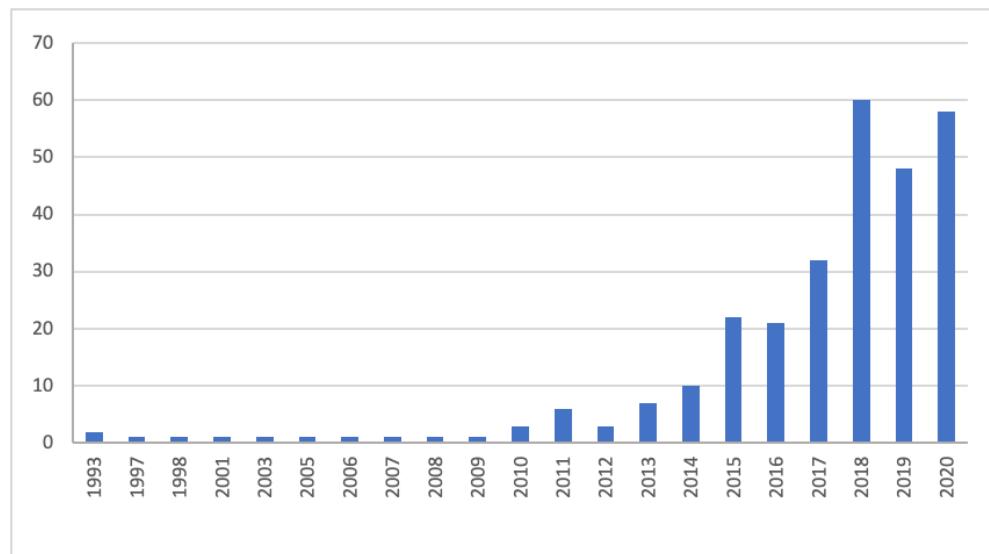


Figure 1. Bibliometric publications categorised with the keywords “Big Data”, “Artificial Intelligence”, “Machine Learning”, “Neural Networks”, “Natural Language Processing”. Source: WoS (543 publications).

Research on the bibliometrics of Big Data and AI publications either focus on earlier periods, such as Niu et al. [34] (2016)’s work on global research on artificial intelligence from 1990–2014 and Kalantari et al. [35] (2017)’s study on trends in Big Data research for 1980–2015, or they use limited number of keywords such as in Raban & Gordon [36] (2020) who used only “Big Data” or “Mega Data” when collecting their data sets in the format of published articles that are categorized with these keywords. Other studies focus points include but are not limited to author cooperations [37], interdisciplinarity [38], visualisation [39] and international collaborations [40]. A recent study [41] by Xu and Yu (2019) provides insights on a recent period (2009–2018), however it covers all Big Data research publications without any in-depth approach to the social sciences subject area.

In addition to these overall research, focused studies on bibliometric analysis of the academic literature on Big Data and/or AI ranges from explainable artificial intelligence [42], engineering applications [43], group decision-making [44]; business intelligence and analytics [45], sustainability [46]; circular economy [47], supply chain management [48] to higher education [49,50]. Despite this recent but rich literature, Big Data and AI applications aspect

has not been investigated from the social sciences perspective. Our study will contribute to the scientific knowledge by bringing forward the social sciences outlook on utilisation of Big Data and AI analytics between 2015 and 2020. Furthermore, we will analyse the position of computational social sciences in a bibliometric context and relate the results to the overall use of Big Data and AI applications in social sciences in general.

2. Materials and Methods

In this section, we detail our data sampling decisions and explain our study design. First, we will give an overview of bibliographic data sources, and second, we focus on the decisions guiding our study design and data collection, describing the nature and extend of the data we use for our analysis.

2.1. Bibliographic Repositories

The data in this study was drawn from the Web of Science (WoS). The WoS is an online subscription-based scientific citation indexing service originally produced by the Institute for Scientific Information (ISI), now maintained by Clarivate Analytics (previously the Intellectual Property and Science business of Thomson Reuters). As the WoS is the oldest indexing service for scientific publications, it is frequently compared to newer indexing services from different perspectives [51–53]. For bibliometric studies, WoS is still one of the most frequently used indexed database [54] for indexing only the highest-quality journals and its strength in representing the well interconnected core citation network components [55]. The WoS Core Collection consists of six online databases and for this study, search was done among the articles indexed by the Social Science Citation Index (SSCI) Expanded, which covers more than 8500 notable journals encompassing 150 disciplines. Consequently, for our study, detailed bibliometric data are extracted from the WoS.

2.2. Study Design

Given the dominance of scientific articles to the other forms of disseminating research findings (such as conference papers) in social sciences, we decided to limit our research to scientific peer-reviewed articles in social sciences. The search is conducted on topic, titles, abstracts, author keywords of the articles with the keywords: “Big Data”, “Artificial Intelligence”, “Machine Learning”, “Neural Networks”, “Natural Language Processing”. This combination of keywords will be referred as the Big Data and AI analytics keywords henceforth.

The selection of these keywords followed an iterative process where we used a set of AI-related keywords to retrieve data and compared the results. As illustrated in Table 1, the selected keywords provide distinct information as the overlap between the keywords is minimal. The time axis is not picked up in an arbitrary manner either. We have conducted different searches on the WoS, looking at the keyword sets, and the number of publications for each year. There is a steady and fast rise in the number of publications after 2015, and these cover the 79 percent of all (31,293) social sciences publications in the WoS if we examine publications from 2000 to 2020. Therefore, we conclude that the selection of 2015–2020 will provide sufficient information for the analysis. Figure 2 details our data collection steps.

We have furthermore used the following criteria for data collection: (i) scientific articles published in peer-reviewed journals; (ii) year of publication between 2015 and 2020; (iii) search descriptors appear in the title, abstract or keywords; (iv) published in English language and (v) in the Social Sciences Citation Index (SSCI) index and refined the results by choosing all the research areas that are listed under social sciences research area.

Based on our study design, the resulting search of articles in English for social sciences including Big Data and AI analytics keywords between 2015 and 2020 is 11,007 for the WoS. This makes our first dataset. Our second dataset aims at gathering publications with

the following keywords: “Computational social science” or “social computing” in social sciences research area. This dataset includes 396 articles in the WoS.

Table 1. Overlaps for 5 keywords (Big Data, Artificial Intelligence, machine learning, neural networks, natural language processing) for 2018 publications in social sciences. Source: WoS.

No of Keywords Out of 5	No of Articles
1	7210
2	1035
3	134
4	6
Grand Total	8385

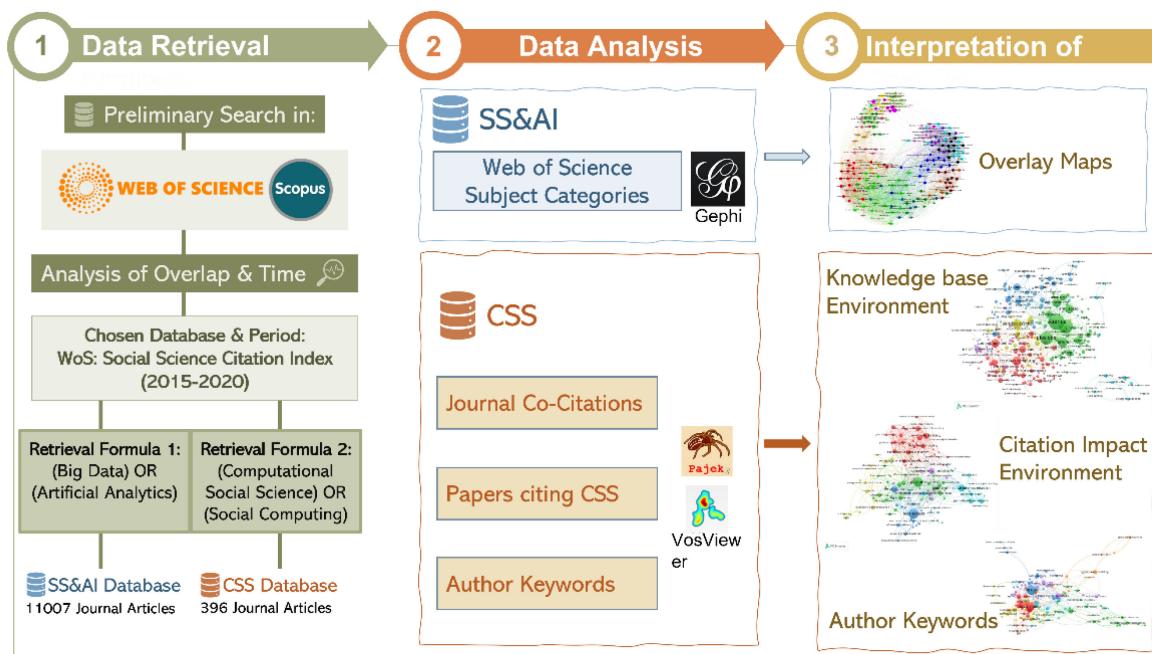


Figure 2. Flow chart of the bibliometric analysis.

Thus, in total we analyse two different data sets: Social Sciences and AI (SS&AI) Data and Computational Social Science (CSS) Data. Both include articles between 2015 and 2020 and each is collected from the WoS search results for Social Sciences. The retrieved set of articles was analysed to discover overall productivity, current research areas (subjects), influential journals and citation patterns.

During the data preparation, Scopus repository was also considered and compared with the WoS datasets. Besides the significant overlap in the datasets, subject categorisations of Scopus and WoS differs significantly and WoS allows more elaborate subject category breakdowns for particularly Social Sciences. Subject categories are not mutually exclusive for Scopus which allow multiple allocations for some publications, although WoS does not allow for that [55]. Therefore, the analyses are conducted on the WoS datasets.

Bibliometrics methodology [56] was adopted to map the time trend, the disciplinary/subject distribution, the high-frequency keywords, the topic evolutions, most influential journals, and citation influence of the related academic articles. Bibliometrics is an active research area that develops metrics and methodologies to measure the transformation of scientific disciplines. The most used unit of analysis in bibliometrics is the citations of academic articles. By looking at the graph structure build by the citations of a set of articles (published for example on the same subject category as our CSS data set) it is possible to find the most influential papers (the ones that receive the highest citations), prolific authors or author groups, most influential journals etc. These citations graphs can simply be built by defining

each article as a node, and the citations as links between these nodes. This approach would result in a directed graph. A more subtle and telling way of building a citation graph is by looking at the similarities of citation patterns between papers or aggregating this information and by generating similarity vectors of journals. This approach is called bibliographic coupling [57,58]. Thus, shared citations, or shared journals give an overview of which papers/journals are working on similar topics. Bibliographic coupling is based on the idea of co-citation which uses the frequency of a paper being cited by other papers as a (semantic) similarity measure [57,59].

For bibliographic coupling analysis an important distinction needs to be made in the data collection, which is best demonstrated with an example. Let us use our CSS data set for that purpose: When we generate a bibliographic coupling graph of all the citations in CSS data set, we analyse which key papers/journals are co-cited the most within this dataset. This shows the knowledge base, i.e., which articles are considered influential to research for computational social sciences. In order to understand the impact of the computational social science publications, we need to look at all the papers citing our CSS dataset, which results in a new dataset of 3820 papers. We will call this dataset simply as citing CSS data. The co-citation of this new set will show which authors, papers, and journals (depending on the aggregation level we prefer) followed the publications on computational social sciences and used the research in this area to further their own research.

Another bibliometrics method that we will make use of is the overlay maps prepared by [60]. Overlay maps are generated by looking at the citations of all journals in the WoS (SSCI and SCI data) and aggregating them at the level of subject categories. Journals can be categorised belonging to more than one subject category. Thus, the resulting network shows the relations between the subject categories of the WoS. This map can be used as a background of the scientific communication in general, over which one can project the distribution of citations on a specific research area or a research topic such as CSS and understand the flow of information of this dataset on top of the overall science map. The first overlay map was prepared in 2009, and it is updated again in 2012/2014. As the subject category structure of the WoS do not change drastically, these overlay maps can still be used to analyse the citation relations of our CSS dataset mapped on the overlay map of the subject categories in the WoS.

To support the bibliometric analysis and the graphical representation of the data two bibliometric analysis tools are employed: Gephi (Maison des Sciences de l'Homme, Paris, France) and VosViewer (Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands). Gephi is a software to visualise and analyse networks in general [61]. VosViewer is a free software developed by Van Eck and Waltman and offers an effective function in co-occurrence analysis and co-citation analysis [62]. For our study, we used VosViewer to make the co-citation and bibliographic coupling analysis of journals, and co-occurrence analysis of the keywords and Gephi for generating the overlay map network of the subject categories of Big Data and AI analytics in social sciences (We have furthermore made use of Levallois' macro to generate the overlay net files from WoS subject categories, available at <https://www.leydesdorff.net/overlaytoolkit/>, accessed on 11 October 2022).

To illustrate the bibliometric pipeline we used, see Figure 2. The pipeline demonstrates the steps taken from the data collection to analysis and results. For a bibliometric analysis pipeline, the choice of keywords and timeline along with the database source used is crucial. However, unlike the protocols developed for a systematic literature review, the bibliometric literature does not have a flow chart to render this step systematically similar. One reason for this is the difference between the two approaches: a systematic literature review is in its essence a qualitative and collective endeavour, undertaken by a group of researchers to locate essential literature and analyse it to generate an overview of a topic [63]. In contrast, bibliometric study is a quantitative approach that collects a wider net of publications to analyse the information flow between research areas, journals, authors.

3. Results

In-depth analyses for the WoS datasets for social science literature of AI and Big Data analytics and CSS are provided below.

3.1. Social Science Disciplines and Journals

Social science publications focusing on the use of Big Data and AI technologies are mapped based on their WoS subject categories on the overlay map. Figure 3 illustrates the subject areas of the SS & AI data. Here, links are citations between journals that fall under more than one subject category and cited together. This generates the overlay map which renders each node, i.e., each subject category as the same size.

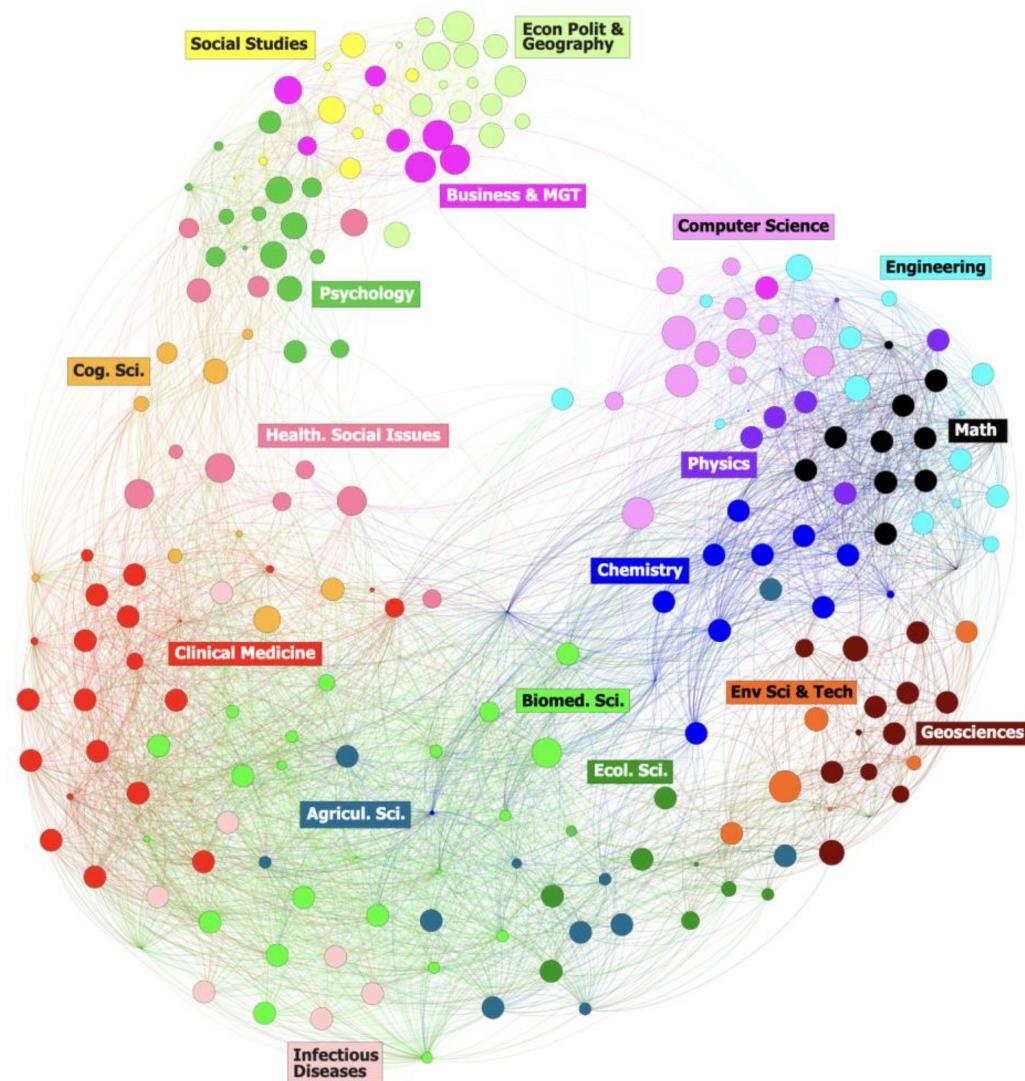


Figure 3. Subject Categories of social sciences articles between the years 2015–2020, based on the Big Data and AI analytics search terms. Source: WoS (11,007 articles).

When we project our own data set, i.e., citations in SS & AI dataset that link the subject categories, we see many subject categories' node sizes getting bigger. The size of the nodes is dependent on how many times they are cited in our SS & AI dataset. Even though the SS & AI dataset is collected only from the SSCI index, the subject categories which dominate the network are computer science, engineering, bio-medical sciences, and medicine. More importantly, the dataset contains citations that link a wide variety of subject categories, which shows that AI and Big Data approaches' diffusion to a wide area in social sciences.

When we take a closer look at the social science related subject categories (Figure 4), we see environmental science, geography, economics, mathematical methods in social sciences, political science, urban studies, public administration, planning and development and international relations belonging to the first cluster (light green). Second cluster (dark pink) includes communication, management, applied psychology, hospitality, leisure, and tourism. Interdisciplinary social sciences, social issues, demography, ethnic studies, anthropology, history, criminology/penology build the third cluster (yellow) up. Fourth cluster convenes women studies in multidisciplinary psychology, educational research, women's studies in social sciences, various sub disciplines of psychology such as social, experimental, developmental, clinical, mathematical, multidisciplinary psychology, development psychology and social work. Fifth cluster (light pink) which bridges to the medical sciences is composed of biomedical social sciences, ergonomics, transportation, services in health policy, healthcare sciences and public health. Next to psychology and substance abuse cluster (light orange), health care services and public, environmental, and occupational health bridges to the medical sciences.

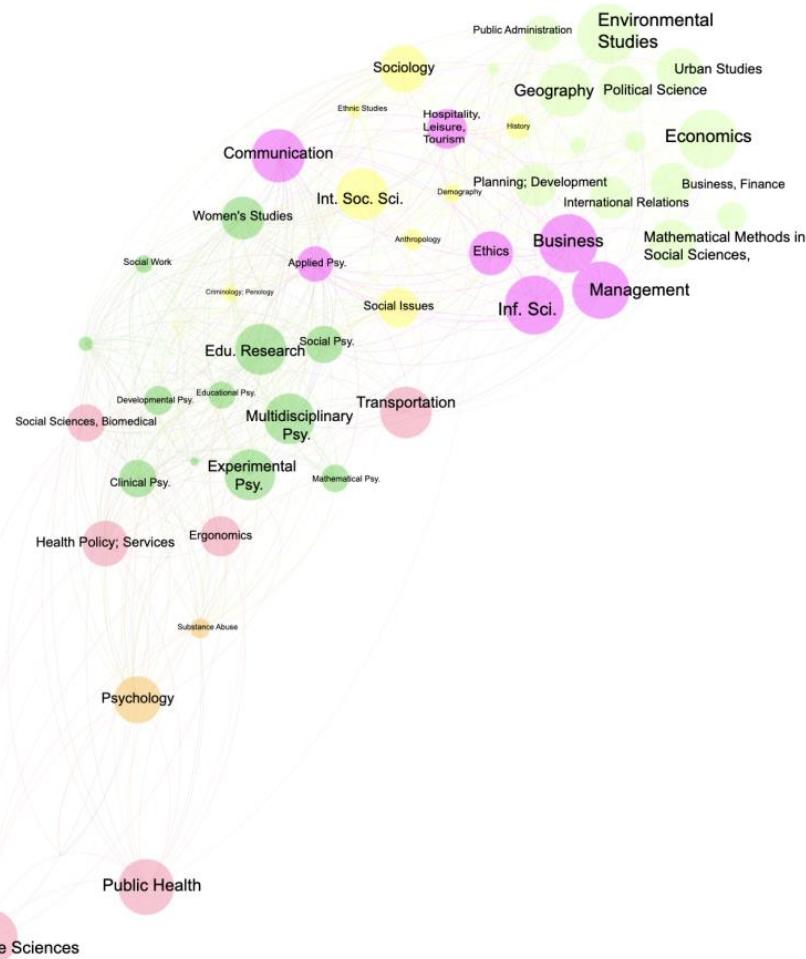


Figure 4. Focused view of Social Sciences related clusters in Subject Categories of social sciences articles between 2015–2020, based on the Big Data and AI analytics search terms. Source: WoS (11,007 articles).

3.2. Computational Social Science: Overarching or Underlying?

Although 2010s witnessed the birth of the (sub)discipline “Social Computing/Computational Social Sciences”, we have seen in the previous section that the AI practices are not only confined within the umbrella of this discipline and is permeated in many main research lines of social sciences as well as sciences in general.

In order to investigate the distribution of Big Data and AI analytics papers which are considered as a part of “computational social science”, we conducted a search with the keywords “computational social science” or “social computing”, which resulted in 396 articles collected from the WoS, i.e., the CSS data. In this section, we will give an in-depth analysis of this dataset, by looking at its knowledge base environment, its citation impact environment, and its author keyword co-occurrence network.

In Figure 5, we have the co-citation network of CSS dataset aggregated at the level of journals. The CSS dataset has a total number of 9040 journals that are cited, out of which 316 journals are co-cited at least 10 times. We can say that these 316 journals constitute the core knowledge base environment of CSS. We have used VosViewer’s clustering algorithm which rendered six distinct clusters.

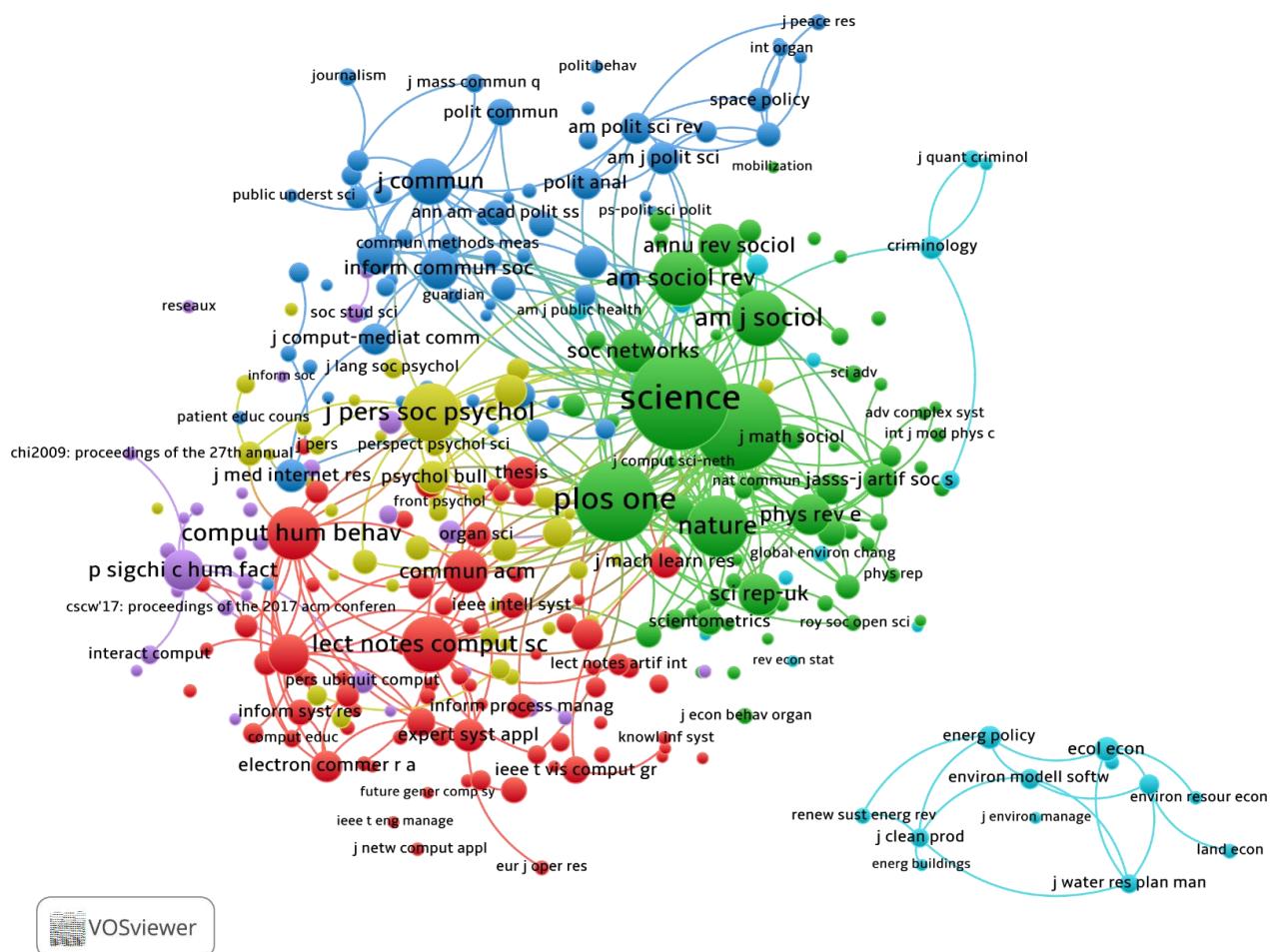


Figure 5. Knowledge base environment of CSS, based on WoS CSS dataset.

Existing clusters represented by different colours and the division is not fortuitous. In the first cluster, where nodes are in green, we see the notable interdisciplinary journals such as Science, PLoS ONE, Nature, Journal of Mathematical Sociology and Scientific Reports (which is also a journal of Nature), Social Networks and Physical Review E and Scientometrics as well as the fundamental sociological journals born in the USA, namely American Sociological Review, American Journal of Sociology and Annual Review of Sociology. Second cluster is represented in red and includes computer and data science-oriented journals from different disciplines other than social science, where the most influential outlets are the Lecture Notes in Computer Science, Computers in Human Behavior, Personal and Ubiquitous Computing, Expert Systems with Applications Journal and Journal of Machine Learning Research. Third cluster with blue nodes and links combines political science and communication science dissemination channels along with

a couple of interdisciplinary journals. Prominent journals whose co-citation network with CSS appears in this cluster are the Journal of Communication, Information Communication & Society, Journal of Computer-Mediated Communication, Communication Methods and Measures, Political Communication, American Journal of Political Science, The Annals of the American Academy of Political and Social Science, Big Data & Society, Space Policy and Mobilization journals. The resting three cluster are relatively smaller than the first three. Cluster 4, coloured in light green, is dominated by psychology-oriented journals such as Journal of Personality and Social Psychology and Psychological Bulletin. The fifth cluster represented in purple stands for human computer relation aspect and involves publication outlets as CHI Conference on Human Factors in Computing Systems, Interacting with Computers and Social Studies of Science. The sixth and the last cluster in light blue is a compound of environmental research outlets to wit Energy Policy, Environmental Modelling & Software, Ecological Economics, Environmental & Resource Economics with an addition of criminological approaches, where we see journals like the Criminology and the Journal of Quantitative Criminology. In short, the knowledge base environment of CSS is diverse, and the social science aspect is restricted within the interdisciplinary co-citation network.

When we scrutinise the co-citation network, and especially focus on leading social science journals like Annual Review of Sociology, American Journal of Sociology or American Sociological Review, we see that the majority are mostly confined to the cluster where journals that publish on either on a broad, interdisciplinary issues like Nature, Science or Plos One, and journals that are well known in Social Sciences. This is a general pattern; core social science journals do not permeate into the rest of the network. Figure 6 shows a focused view of the citation environment of Annual Review of Sociology, American Journal of Sociology or American Sociological Review and Big Data and Society. Annual Review of Sociology follows the general pattern whereas the other three journals are outliers.

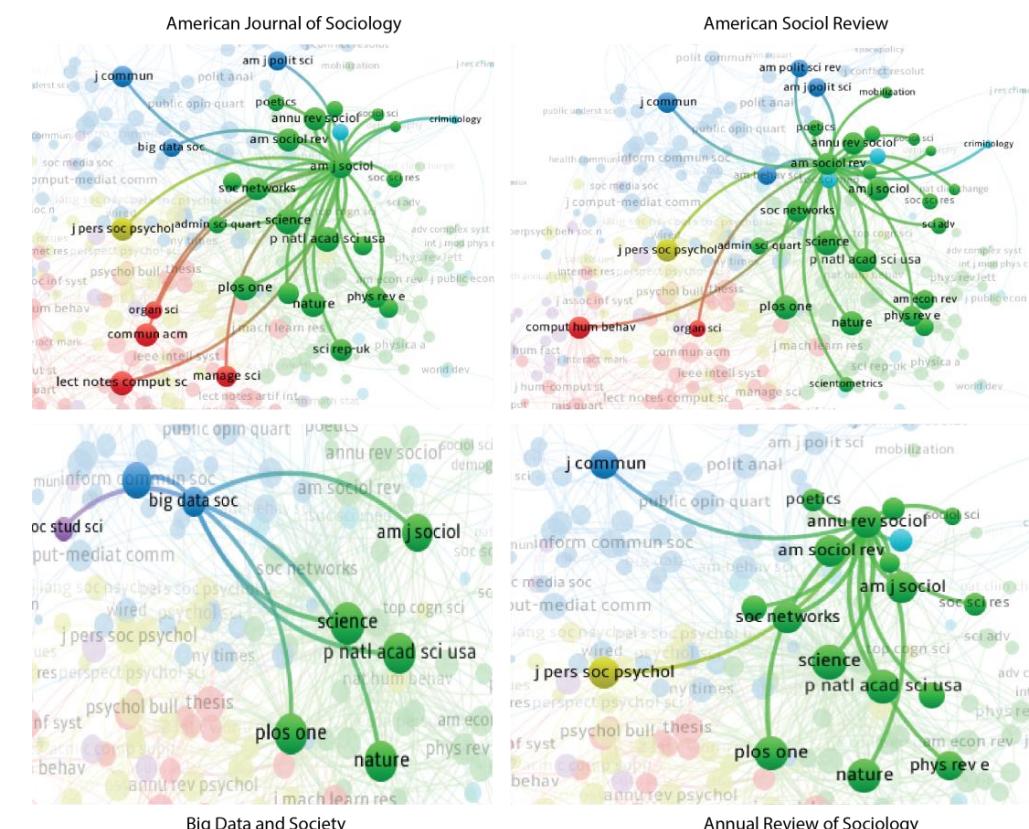


Figure 6. A focused view for American Journal of Sociology, American Sociological Review, Big Data & Society and Annual Review of Sociology.

Following the knowledge base environment, we furthermore looked at the citation impact environment of the CSS publications by generating a bibliographic coupling network of all articles (3820 in total) citing our CSS dataset. Figure 7 shows the resulting network, where we included journals that are co-cited at least 5 times. This resulted in a focus network that has 155 journals out of the 1502 in the original dataset. These journals build 6 clusters.

Clusters of the citation impact environment of CSS are namely computer science and engineering (in dark blue), communication (in red), interdisciplinary psychology and humanities (in dark green), health and medical research (in light blue) and energy and environment (in light green). Comparing the knowledge base networks with the citation impact environment, the composition of the clusters overlaps mostly with the exception of health and medical research publications and engineering science outputs (that accompany computer science).

The CSS articles have been cited by varying interdisciplinary outlets, most nourished of those are PLoS ONE, Electrical and Electronics Engineers (IEEE) Access, Sustainability, Journal of Medical Internet Research, Computers in Human Behavior. A remarkable lack in these networks is the social science journals. It appears that the CSS publications are fed by social science outlets, however, their scientific influence on social science articles and hence on journals is not evident.

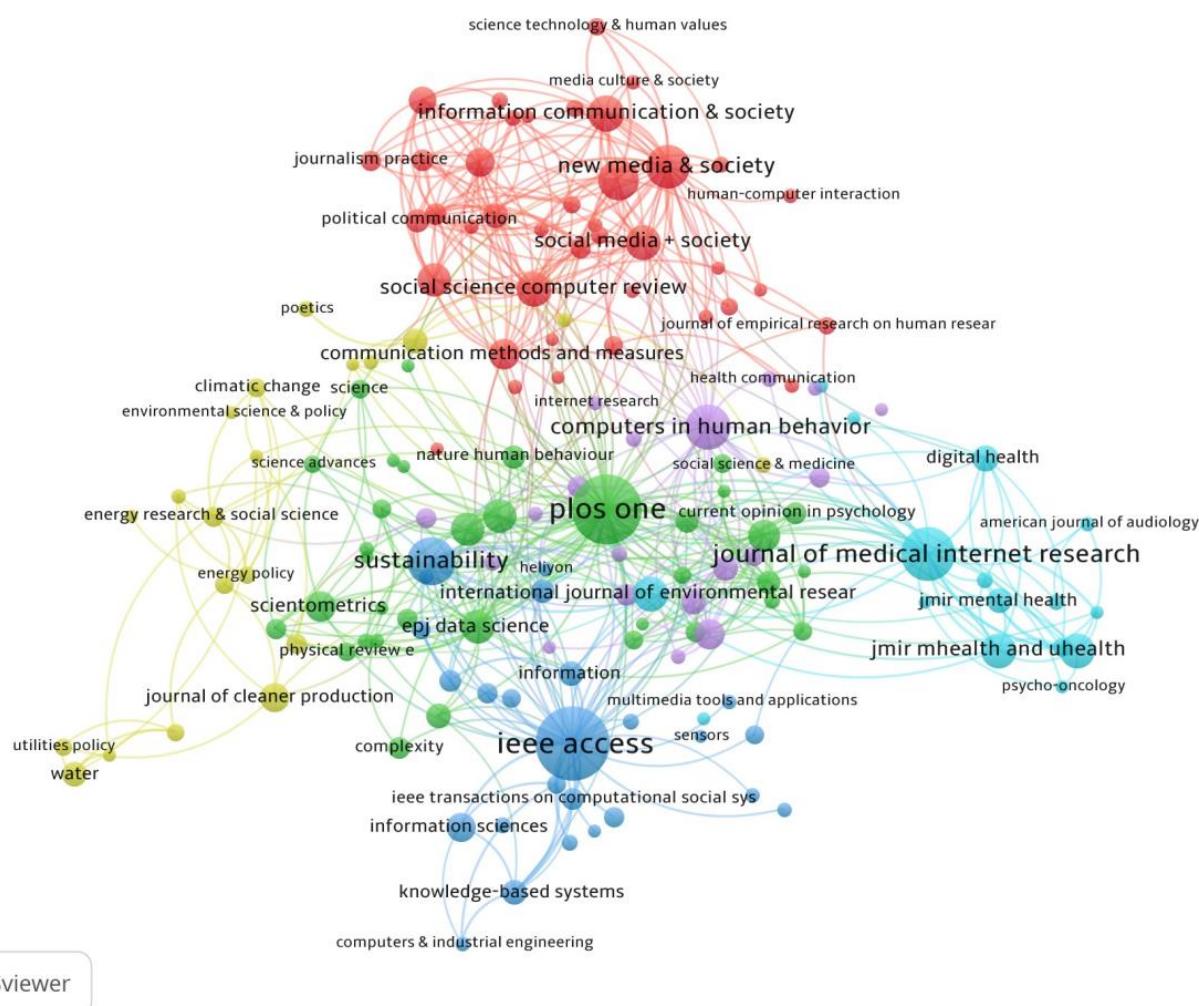


Figure 7. Citation impact environment of CSS, based on WoS, Citing CSS dataset.

To see how the social science journals that we observed as outliers in the knowledge environment network are cited in the citation impact environment, we prepared another

focused view (see Figure 8). As discussed above, our citation impact networks are drawn based on the 155 journals that are co-cited at least 5 times. This threshold leads to elimination of the American Journal of Sociology. Thus, Figure 8 represents Annual Review of Sociology, American Sociological Review and Big Data & Society. The focused view displays that PLoS ONE is the most inclusive journal that has citations from all three journals. Besides, amongst these social science journals, Big Data & Society's interdisciplinarity can be only within its cluster but also by journals from all other four clusters.

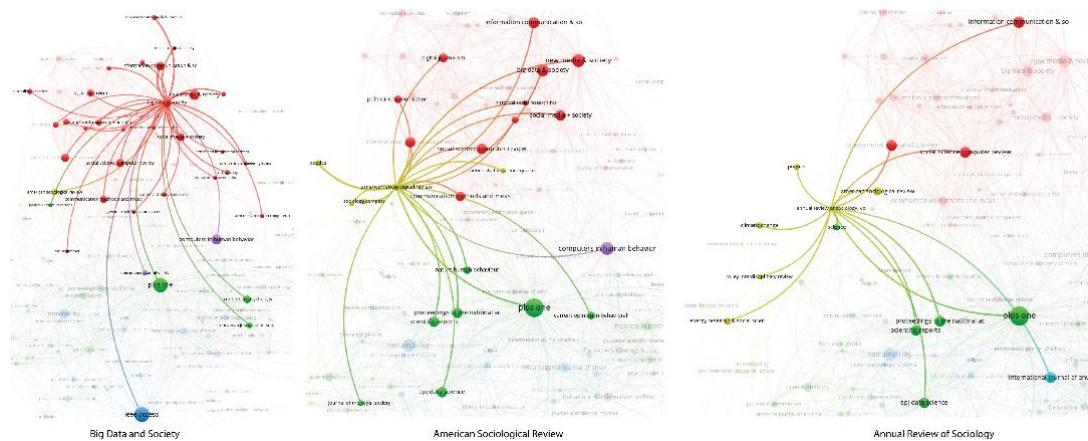


Figure 8. Focus view of Big Data & Society, Annual Review of Sociology and American Sociological Review in the Citation impact environment of CSS, based on WoS, Citing CSS dataset.

To understand the research focus of the authors who published as part of the CSS dataset, we need to visualise the network that is built by the co-occurrence of author keywords. In Figure 9, we see this network, which is generated by including all author keywords of the CSS dataset that at least had occurred twice in different papers. The resulting network has 173 nodes, and 8 clusters. The smallest clusters are the ones focused on a specific methodology and the related terms, such as complex networks and human behaviour (in brown, on the left bottom), or complexity science and agent-based modelling (in orange, on the right top). The main cluster with 31 keywords has social media at its heart (in red). Beside social media platforms such as Twitter and Facebook, here we see social media analysis related keywords (social media and identity, participation, public opinion etc.) and algorithm specific keywords (topic modelling, web 2.0, text analysis etc.). Keywords that build a bridge between these are sentiment analysis, misinformation, fake news, political polarisation. The second biggest cluster (in light green) is devoted to human computer interaction and related keywords such as affective computing, social and collaborative computing, user studies, etc. Here keywords such as privacy, social integration and emotion is also visible. However, overwhelmingly the keywords do come from the AI domain (fuzzy logic, mobile sensing, assistive technologies to name a few). The cluster in blue, where big data node dominates, has an interesting mix of keywords: from broken windows, ethics, policy analytics, research design and research ethics, e-business related terms such as e-business, e-commerce and data analytics are also part of this cluster. The purple cluster which is very much diffused into the other clusters have network theory related keywords such as complex systems, homophily, network science, social networks, mobility etc. The last cluster (in yellow) has artificial intelligence and social networks devoted keywords such as user identification, user profiling, visualisation, visual analytics, data mining and computer vision.

When we have a closer look to the author keywords of the CSS, in Figure 10, we decipher the networks of four most common keywords, namely Big Data, social media, social networks and privacy. The links between the clusters through these networks depict the differences and intersections of the clusters. Starting with the Big Data keyword network, it is seen that the network is compound of diverse and numerous keywords and inter-cluster

linkages. This can be interpreted as the “generalist approach” of the Big Data articles. Here the linked keywords are from varied disciplines and cover varied contexts such as content, theory, and methodology. In short, this network illustrates the articles following multivariate and multidisciplinary approach from a generalist perspective. Nevertheless, other three most common keywords’ networks are a good example of “specialised and convergent approaches”, for instance interlinked studies of social media, social networks, and privacy.

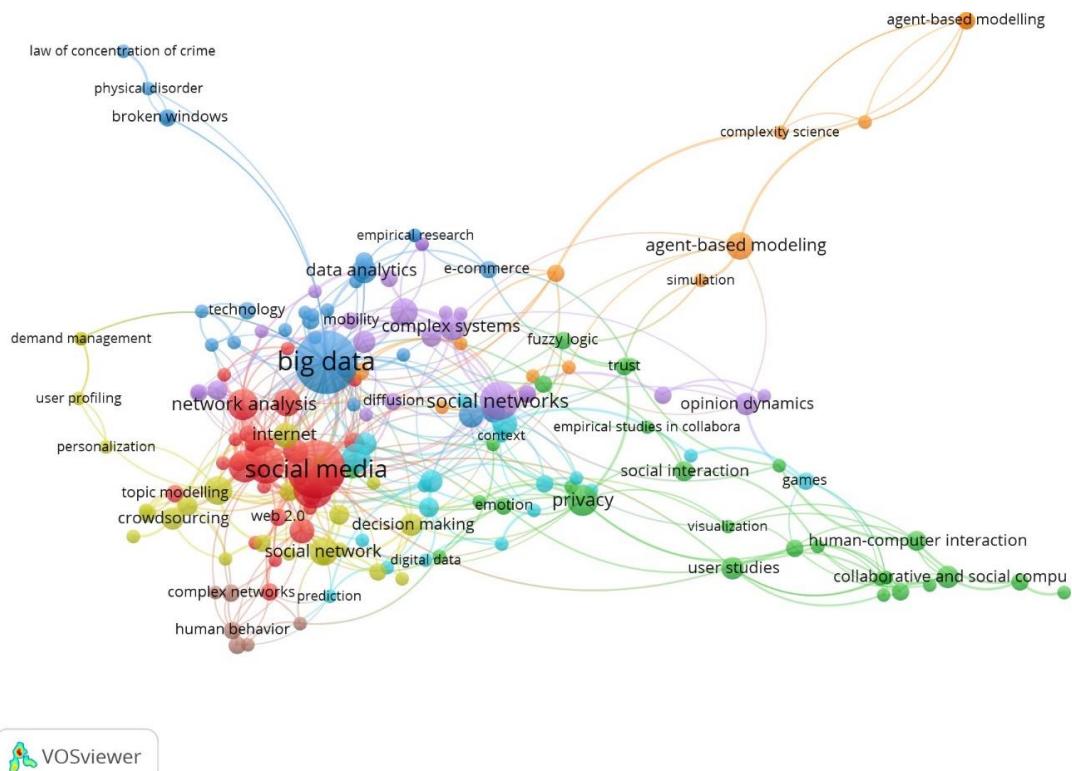


Figure 9. Author Keywords of CSS, based on WoS, CSS dataset.

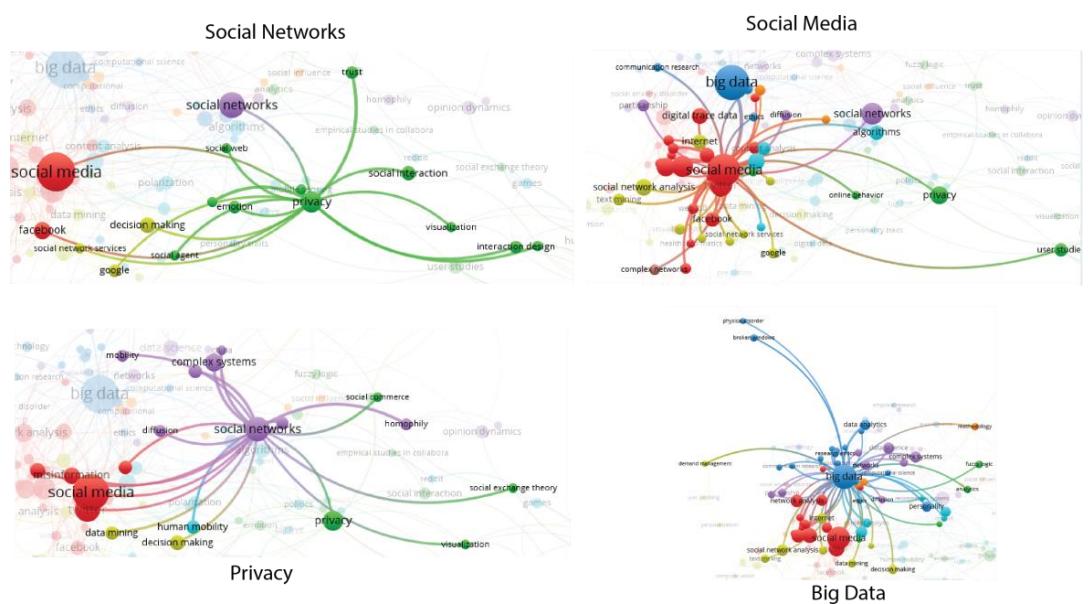


Figure 10. Focus view of Author Keywords of CSS: Big Data, Social Media, Social Networks and Privacy, based on WoS, CSS dataset.

4. Discussion

The recent advances in digital technologies do not transform only the societies but also the scientific spheres. The popularity of AI and Big Data does not discriminate any fields and reigns numerous scientific disciplines with its applications at different levels. Like other research areas, social sciences have been solicitous about Big Data and AI in the recent decade. As of 2015, there is a sharp increase in the number of SS & AI publications since AI and Big Data technologies are used across a very wide range in Social Sciences.

Computational applications and empirical studies occupy a significant share of the Big Data and AI theme in the social sciences. Interestingly, albeit 2010s witnessed the birth of the discipline “Social Computing/Computational Social Sciences”, we see that the AI practices are not only confined within the umbrella of this discipline and is permeated in many main research lines of social sciences. Hence, the anticipated overlap between the SS & AI specialisation and computational social science (CSS) has yet to be crystallised. Given its promising nature, in order to make use of its full potential, CSS needs to grow up and develop as one of the major social science disciplines. However, as our co-occurrence of keywords analysis’ results indicate, most keywords in CSS articles are technique-specific and unfortunately very few social science concepts are salient.

Considering the tendency towards Big Data and AI analytics within the social sciences, no one discipline in social sciences dominates over the others in SS & AI citation environment. Yet, the dissemination of the SS & AI articles and their citation impact sphere is restricted. This is not surprising given the fact that 32% of the overall articles published in the social sciences are never cited by another researcher within a five-year citation window [64].

On the other hand, our findings reveal that the AI and Big Data vs. social science balance is outweighed by the AI oriented studies. In-depth analysis of the publication outlets indicates that most of the AI-related approaches to Social Science research is carried out and published by data/computer scientists, and in related fields, but not in hard core social science journals. Nonetheless, in addition to the top three sociological journals (American Journal of Sociology, American Sociological Review, Annual Review of Sociology), new publication outlets strengthen the weak link between computational sciences and social sciences, which is evident in the citation networks of the CSS publications.

5. Conclusions

This paper assesses the scientific impact of Big Data and AI in social Sciences scholarly work sphere to provide a fundamental framework for future research. Our findings demonstrated that (1) There is a significant increase in the number of Big Data and AI research output topics or applications in different social sciences disciplines. (2) The citation networks of the social science related subject categories quantitatively and qualitatively demonstrate the connections between articles and authors by revealing significant subject areas. It is striking that the knowledge flow starts from social studies, economics, psychology, and business (major social science domains), goes through health and biomedical related disciplines and natural sciences towards mathematics, physics, engineering and computer science (major computational/data science domains). The relative influence of major social science domains on computational/data science domains (and vice versa) is weak. (3) When the publication oeuvre for CSS is scrutinised, six distinct clusters are identified for the publication outlets, which also illustrate that CSS knowledge base environment includes numerous interdisciplinary outlets with no significance presence of social sciences journals. All in all, the sphere of influence of the CSS papers is still limited due to their low diffusion into the social science citation networks.

The use of AI and Big Data analysis algorithms is complementary to theory-driven analysis approaches and relies on data-driven insights. It leverages the capacity to collect and analyse data at a scale that may reveal patterns of individual and group behaviours in finer granularity than traditional approaches can offer. In particular, approaches addressing new data sources are able to provide insights on an unprecedented scale and with reduced

time and investment requirements, once the required analysis tools are researched and communicated. Furthermore, through automated pipelines and well-documented best practices, open-sourced computational tools, and education to prepare interdisciplinary collaboration between computational scientists and social scientists it becomes possible to create timely interventions. For instance, the questions of societal challenges should be translated into algorithms that will be run on data from the industrial and governmental stakeholders, through well-defined and carefully regulated data collaborative, including systems that implement mechanisms of differential privacy.

For the future work, there are two major points we would like to rise. First, social science perspective of the AI and Big Data applications should also include the “ethics” aspect, nonetheless, it has not revealed itself in our keyword co-occurrence analysis. This should be interpreted with great caution and future research focusing on in-depth analyses of ethical aspects within the computational social science field will be beneficial. As the second point, our study was restricted to the peer-reviewed scientific articles written in English. Future research can expand our study through including reports, policy briefs and other types of publications and also in other languages (such as Chinese, Spanish, etc.) to study the dispersion and influence of the AI and Big Data in national and regional social science spheres.

Author Contributions: Conceptualization, T.B.; methodology, formal analysis, investigation, resources, data curation, writing—original draft preparation and writing—review and editing, T.B. and A.A.A.S.; visualization, A.A.A.S.; supervision, T.B.; project administration and funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the European Commission through the Horizon2020 European project: “HumMINGBird—Enhanced migration measures from a multidimensional perspective” (GA: 870661).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. UN Global Pulse. Big Data for Development: Opportunities and Challenges—White Paper. 2012. Available online: <https://www.unglobalpulse.org/wp-content/uploads/2012/05/BigDataforDevelopment-UNGlobalPulseMay2012.pdf> (accessed on 10 October 2022).
2. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
3. Kitchin, R. Big data, new epistemologies and paradigm shifts. *Big Data Soc.* **2014**, *1*, 2053951714528481. [CrossRef]
4. De Neufville, R.; Baum, S.D. Collective action on artificial intelligence: A primer and review. *Technol. Soc.* **2021**, *66*, 101649. [CrossRef]
5. Unver, H.A. Using social media to monitor conflict-related migration: A review of implications for AI Forecasting. *Soc. Sci.* **2022**, *11*, 395. [CrossRef]
6. Cath, C. Governing artificial intelligence: Ethical, legal and technical challenges and opportunities. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 20180080. [CrossRef] [PubMed]
7. Bircan, T.; Korkmaz, E.E. Big data for whose sake? Governing migration through artificial intelligence. *Nat. Humanit. Soc. Sci. Commun.* **2021**, *8*, 241. [CrossRef]
8. Gefen, A.; Saint-Raymond, L.; Venturini, T. AI for Digital Humanities and Computational Social Sciences. In *Reflections on Artificial Intelligence for Humanity*; Springer: Cham, Switzerland, 2021; pp. 191–202.
9. Kong, J.D.; Fevrier, K.; Effoduh, J.O.; Bragazzi, N.L. Artificial Intelligence, Law, and Vulnerabilities. In *AI and Society*; Chapman and Hall/CRC: London, UK, 2022; pp. 179–196.
10. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1–38. [CrossRef]
11. Gunning, D.; Stefk, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef] [PubMed]

12. Chang, R.M.; Kauffman, R.J.; Kwon, Y. Understanding the paradigm shift to computational social science in the presence of big data. *Decis. Support Syst.* **2014**, *63*, 67–80. [[CrossRef](#)]
13. Ligo, A.K.; Rand, K.; Bassett, J.; Galaitis, S.E.; Trump, B.D.; Jayabalasingham, B.; Linkov, I. Comparing the emergence of technical and social sciences research in artificial intelligence. *Front. Comput. Sci.* **2021**, *3*, 653235. [[CrossRef](#)]
14. Veltri, G.A. Big data is not only about data: The two cultures of modelling. *Big Data Soc.* **2017**, *4*, 2053951717703997. [[CrossRef](#)]
15. Anastasopoulos, L.J.; Whitford, A.B. Machine learning for public administration research, with application to organizational reputation. *J. Public Adm. Res. Theory* **2019**, *29*, 491–510. [[CrossRef](#)]
16. Edelmann, A.; Wolff, T.; Montagne, D.; Bail, C.A. Computational social science and sociology. *Annu. Rev. Sociol.* **2020**, *46*, 61–81. [[CrossRef](#)] [[PubMed](#)]
17. Schroeder, R. Big data and cumulation in the social sciences. *Inf. Commun. Soc.* **2020**, *23*, 1593–1607. [[CrossRef](#)]
18. Xiao, X.Y.; Jin, L.; Kateb, F.; Aldeeb HM, A. Modernisation of urban governance: An approach of ‘Blockchain+ Big Data’. *Appl. Math. Nonlinear Sci.* **2021**, *6*, 535–542. [[CrossRef](#)]
19. Li, Y.; Cheng, Q.; Khder, M.A. Nonlinear differential equations in computer-aided modeling of Big Data technology. *Appl. Math. Nonlinear Sci.* **2022**. ahead of print. [[CrossRef](#)]
20. Yang, X.; Liang, X.; Peng, L.; Liu, Y.; Elzefzafy, H. Research on urban landscape big data information processing system based on ordinary differential equations. *Appl. Math. Nonlinear Sci.* **2022**. ahead of print. [[CrossRef](#)]
21. Schroeder, R.; Taylor, L. Big data and Wikipedia research: Social science knowledge across disciplinary divides. *Inf. Commun. Soc.* **2015**, *18*, 1039–1056. [[CrossRef](#)]
22. Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. Social science. Computational Social Science. *Science* **2009**, *323*, 721–723.
23. Cioffi-Revilla, C. Computation and social science. In *Introduction to Computational Social Science*; Springer: Cham, Switzerland, 2017; pp. 35–102.
24. Harford, T. Big data: A big mistake? *Significance* **2014**, *11*, 14–19. [[CrossRef](#)]
25. Kitchin, R.; McArdle, G. What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data Soc.* **2016**, *3*, 2053951716631130. [[CrossRef](#)]
26. Laney, D. 3d data management: Controlling data volume, velocity and variety. *META Group Res. Note* **2001**, *6*, 1.
27. Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Hung Byers, A. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*; McKinsey Global Institute: Washington, DC, USA, 2011.
28. González-Bailón, S. Social science in the era of big data. *Policy Internet* **2013**, *5*, 147–160. [[CrossRef](#)]
29. European Commission. The EU Data Protection Reform and Big Data: Factsheet 2016. Available online: <http://ec.europa.eu/newsroom/just/document.cfm?docid=41523> (accessed on 10 October 2022).
30. Iliadis, A.; Russo, F. Critical data studies: An introduction. *Big Data Soc.* **2016**, *3*, 2053951716674238. [[CrossRef](#)]
31. Yu, D.; Xu, Z.; Pedrycz, W.; Wang, W. Information sciences 1968–2016: A retrospective analysis with text mining and bibliometric. *Inf. Sci.* **2017**, *418*, 619–634. [[CrossRef](#)]
32. Kwon, S.; Liu, X.; Porter, A.L.; Youtie, J. Research addressing emerging technological ideas has greater scientific impact. *Res. Policy* **2019**, *48*, 103834. [[CrossRef](#)]
33. Thelwall, M.; Sud, P. Do new research issues attract more citations? a comparison between 25 scopus subject categories. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 269–279. [[CrossRef](#)]
34. Niu, J.; Tang, W.; Xu, F.; Zhou, X.; Song, Y. Global research on artificial intelligence from 1990–2014: Spatially-explicit bibliometric analysis. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 66. [[CrossRef](#)]
35. Kalantari, A.; Kamsin, A.; Kamaruddin, H.S.; Ebrahim, N.A.; Gani, A.; Ebrahimi, A.; Shamshirband, S. A bibliometric approach to tracking big data research trends. *J. Big Data* **2017**, *4*, 30. [[CrossRef](#)]
36. Raban, D.R.; Gordon, A. The evolution of data science and big data research: A bibliometric analysis. *Scientometrics* **2020**, *122*, 1563–1581. [[CrossRef](#)]
37. Peng, Y.; Shi, J.; Fantinato, M.; Chen, J. A study on the author collaboration network in big data. *Inf. Syst. Front.* **2017**, *19*, 1329–1342. [[CrossRef](#)]
38. Hu, J.; Zhang, Y. Discovering the interdisciplinary nature of big data research through social network analysis and visualization. *Scientometrics* **2017**, *112*, 91–109. [[CrossRef](#)]
39. Liao, H.; Tang, M.; Luo, L.; Li, C.; Chiclana, F.; Zeng, X.-J. A bibliometric analysis and visualization of medical big data research. *Sustainability* **2018**, *10*, 166. [[CrossRef](#)]
40. Hu, H.; Wang, D.; Huang, S.Q. International collaboration in the field of artificial intelligence: Global trends and networks at the country and institution levels. In Proceedings of the 17th International Conference on Scientometrics and Informetrics, Rome, Italy, 2–5 September 2019; pp. 2501–2502.
41. Xu, Z.; Yu, D. A bibliometrics analysis on big data research (2009–2018). *J. Data Inf. Manag.* **2019**, *1*, 3–15. [[CrossRef](#)]
42. Alonso, J.M. Teaching explainable artificial intelligence to high school students. *Int. J. Comput. Intell. Syst.* **2020**, *13*, 974–987. [[CrossRef](#)]
43. Shukla, A.K.; Janmaiijaya, M.; Abraham, A.; Muhuri, P.K. Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Eng. Appl. Artif. Intell.* **2019**, *85*, 517–532. [[CrossRef](#)]

44. Heradio, R.; Fernandez-Amoros, D.; Cerrada, C.; Cobo, M.J. Group decision-making based on artificial intelligence: A bibliometric analysis. *Mathematics* **2020**, *8*, 1566. [[CrossRef](#)]
45. Liang, T.; Liu, Y. Research landscape of business intelligence and big data analytics: A bibliometrics study. *Expert Syst. Appl.* **2018**, *111*, 2–10. [[CrossRef](#)]
46. Lafuente-Lechuga, M.; Cifuentes-Faura, J.; Faura-Martínez, U. Sustainability, Big Data and Mathematical Techniques: A Bibliometric Review. *Mathematics* **2021**, *9*, 2557. [[CrossRef](#)]
47. Nobre, G.C.; Tavares, E. Scientific literature analysis on big data and internet of things applications on circular economy: A bibliometric study. *Scientometrics* **2017**, *111*, 463–492. [[CrossRef](#)]
48. Mishra, D.; Gunasekaran, A.; Papadopoulos, T.; Childe, S.J. Big data and supply chain management: A review and bibliometric analysis. *Ann. Oper. Res.* **2018**, *270*, 313–336. [[CrossRef](#)]
49. Hinojo-Lucena, F.-J.; Aznar-Díaz, I.; Cáceres-Reche, M.-P.; Romero-Rodríguez, J.-M. Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature. *Educ. Sci.* **2019**, *9*, 51. [[CrossRef](#)]
50. Marín-Marín, J.-A.; López-Belmonte, J.; Fernández-Campoy, J.-M.; Romero-Rodríguez, J.-M. Big data in education. a bibliometric review. *Soc. Sci.* **2019**, *8*, 223. [[CrossRef](#)]
51. Abdulhayoglu, M.A.; Thijss, B. Use of locality sensitive hashing (lsh) algorithm to match Web of Science and Scopus. *Scientometrics* **2017**, *116*, 1229–1245. [[CrossRef](#)]
52. Mongeon, P.; Paul-Hus, A. The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* **2015**, *106*, 213–228. [[CrossRef](#)]
53. Raban, D.R.; Gordon, A.; Geifman, D. The information society: The development of a scientific specialty. *Inf. Commun. Soc.* **2011**, *14*, 375–399. [[CrossRef](#)]
54. Norris, M.; Oppenheim, C. Comparing alternatives to the web of science for coverage of the social sciences' literature. *J. Informetr.* **2007**, *1*, 161–169. [[CrossRef](#)]
55. Stahlschmidt, S.; Stephen, D. *Comparison of Web of Science, Scopus and Dimensions Databases*; DZHW: Hannover, Germany, 2020.
56. Glänzel, W.; Schoepflin, U. A bibliometric study of reference literature in the sciences and social sciences. *Inf. Process. Manag.* **1999**, *35*, 31–44. [[CrossRef](#)]
57. Boyack, K.W.; Klavans, R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2389–2404. [[CrossRef](#)]
58. Kessler, M.M. *An Experimental Study of Bibliographic Coupling between Technical Papers*; Massachusetts Inst of Tech Lexington Lincoln Lab: Lexington, MA, USA, 1962.
59. Glänzel, W.; Czerwon, H. A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics* **1996**, *37*, 195–221. [[CrossRef](#)]
60. Rafols, I.; Porter, A.L.; Leydesdorff, L. Science overlay maps: A new tool for research policy and library management. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 1871–1887. [[CrossRef](#)]
61. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open-Source Software for Exploring and Manipulating Networks 2009. Available online: <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154> (accessed on 10 October 2022).
62. Van Eck, N.J.; Waltman, L. Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [[CrossRef](#)] [[PubMed](#)]
63. Petticrew, M.; Roberts, H. *Systematic Reviews in the Social Sciences: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
64. Alvesson, M.; Gabriel, Y.; Paulsen, R. *Return to Meaning: A Social Science with Something to Say*; Oxford University Press: Oxford, UK, 2017.