

## Research Article

# Paths Study on Knowledge Convergence and Development in Computational Social Science: Data Metric Analysis Based on Web of Science

Yuxi Liu <sup>1</sup>, Xin Feng <sup>2,3</sup>, Yue Zhang <sup>3</sup>, Ying Kong <sup>4</sup> and Rongyao Yang <sup>2</sup>

<sup>1</sup>Computational Communication Research Center, Beijing Normal University, Zhuhai 519087, China

<sup>2</sup>School of Economics and Management, Yanshan University, Qinhuangdao 066004, China

<sup>3</sup>School of Urban Geology and Engineering, Hebei GEO University, Shijiazhuang 050031, China

<sup>4</sup>School of Foreign Languages and Culture, Hebei GEO University, Shijiazhuang 050031, China

Correspondence should be addressed to Ying Kong; kong78@126.com

Received 20 January 2022; Accepted 12 April 2022; Published 29 April 2022

Academic Editor: Fujun Lai

Copyright © 2022 Yuxi Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computational social science, as an emerging interdisciplinary discipline, is a field ushered in by long-term development of traditional social science. It is committed to supplying data thinking, resources, and analytics to study human social behavior and social operation laws to accurately grasp and judge the developing path of the discipline, which is of great significance to promote the innovation and development of social sciences. This study is to conduct a systematic quantitative analysis from a bibliometric perspective, aiming to provide a reference for scholars to explore the paths and changing rules in the field. We use the relevant literature in Web of Science as the dataset. After eliminating journal calls and irrelevant literature, R language and SciMAT tools are used to visualize and analyze the number of articles, keyword clustering, keyword cooccurrence network, and theme evolution, so as to summarize and sort out the paths of computational social science research. The study found that the annual volume of publications has been gradually increasing and will probably remain active in the next few years with high productivity. Subject themes in different periods are diversified, and the evolutionary relationship is found complex as well. Besides, as a cross discipline, scientific knowledge from different fields cross collides and couples with each other in the big data environment, changing the traditional concept of computational social science and forming a new development path. Recently, the emergence of “big data+” has promoted the rise of new subject areas, making the development of new disciplines a reality.

## 1. Introduction

In recent years, computers, the Internet, and the successive emergence of large-scale datasets about human behavior have substantially improved people’s ability to collect and analyze information and promoted the rise of computational social science [1]. And once the discipline was introduced, it was rapidly accepted and widely disseminated by the academic community and increasingly became a hot topic in the field of humanities and social sciences research. Computational social science is a product of the penetration, integration, and innovation of modern science and emerging technological tools such as natural science, computer science, big data, and social science [2], and the discipline is

profoundly influencing the development of many disciplinary fields and the change of related research paradigms. Therefore, the emergence of computational social science is not only the inevitable choice of the development of social science research itself, but also the development of modern science and technology provides favorable instrumental support and is the urgent requirement for social science research by the development of human society. It is for the above reasons that the discipline has been widely of interest to the academic community for its introduction and development. Therefore, as an emerging interdisciplinary field, it is beneficial to promote further research in this discipline by sorting out and analyzing its development path, prospects, and relationship with related disciplines.

*1.1. Literature Review of Computational Social Science Research.* In 2009, David Lazer and 15 other scholars published computational social science (CSS) in science, which provided an in-depth exposition of computational social science and became a symbol of the establishment of computational social science [3], triggering a wave of research boom. Computational social science applies algorithms and computational tools to complex data and studies social phenomena at multiple scales through modeling and computation [4]. Among them, the societal agent-based modeling (ABM) simulation, in particular, faces the challenge of modeling and analyzing complex adaptive social systems [5]. In addition, computational social science facilitates more systematic testing of theories and increases the possibility of research replication. These two factors facilitate the field of social sciences to reach a higher scientific status. In 2013, Laubichler et al. [6] briefly reviewed the computational research system in the history of science, discussing its impact on research, education, etc., and its connection with similar shifts in the natural sciences and social sciences emphasizing big data. Laubichler et al. argue that computational methods help to reconnect the history of science with individual scientific disciplines. In 2014, Sohn [7] developed a rational-choice-theory-driven framework for computational social science in the Strategic Foundation of Computational Social Science, constantly paying attention to social interaction on the Internet. But in 2017, Bravo and Farjam analyzed the prospects and challenges of computational social science [8]. In 2018, H. Wallach argued that computational social science was not a simple addition of computer science and social data and that machine learning in social science should be viewed differently from machine learning in other fields [9]. In 2019, Peng et al. argue that computational social science has led to a paradigm shift in social science research, especially in communication studies [10]. In 2020, Tao et al. argued that online social media was an important source of information for computational social science and network intelligence research and that computational social science may be transformed into a continuous learning process when social phenomena are considered and transformed into data for analysis [11]. In 2020, Lazer et al. defined computational social science as the development and application of computational methods to complex large-scale human (sometimes simulated) behavioral data to analyze its opportunities and challenges [9]. In 2020, Zhang et al. analyzed the differences and connections, areas involved, and application methods of social science, computational social science, and computer science [12]. In 2021, Tornberg and Uitermark [13] criticized the complexity view of computational social science and advocated an unorthodox computational social science based on critical realism metatheories.

*1.1.1. Literature Review of Scientometric Studies.* Scientometric analysis uses quantitative methods to explore the intrinsic laws of disciplinary development with the help of econometric analysis tools and to describe the development process, present the intrinsic association of literature,

predict trends, evaluate scientific systems, and promote scientific and technological progress through visualization. In 2017, Leung et al. applied multiple bibliometric analysis methods (citation, cocitation, keyword, and cword analysis) to explore the theoretical basis and thematic evolution of hotel, theoretical foundations, and thematic evolution of research in the field of tourism and social media [14]. In 2018, Zou et al. used literature cocitation, keyword cooccurrence, and burst detection analysis to visually explore the knowledge base, thematic distribution, research frontiers, and trends in road safety research [15]. In 2020, Marques relied on SSCI database for authorship analysis and citation analysis and focused on the interrelationship between public service motivation and leadership for all public management articles published before January 2018 [16]. In 2021, Ching-Jie Wen and Zi-Jian Ren used cocitation, cword, and cluster analysis to study the progress and trends of BIM [17]. In 2021, Abad-Segura et al. used scientometric methods to analyze 1,130 papers in the Scopus database to summarize the current state of blockchain research within the last five years and future research directions for scholars and investors in research projects [18].

*1.1.2. Research Questions.* Computational social science is undergoing rapid changes in the big data environment, and research hotspots are constantly evolving. Unfortunately, there are few studies in the academic community that have conducted systematic quantitative analysis from a bibliometric perspective, for example, for research hotspots in the field of computational social sciences since the 21st century. Therefore, this paper intends to use R language Biblioshiny, SciMAT and other bibliometric tools to comprehensively quantify the research system in the field of computational social science, to visualize and discuss its development trends. To help relevant researchers understand the disciplinary pathways, the main research questions to be addressed in this paper are as follows:

- (1) What has been the dynamic evolution of scientific knowledge production in the field of computational social science since its inception as a field of study?
- (2) How have the main research directions in the field changed during that period?
- (3) What are the directions of research in the discipline in the coming years?

## 2. Data Sources and Research Methods

*2.1. Data Sources.* Web of Science (WoS) database includes three well-known citation index databases (SCI, SSCI, A&HCI), which collect authoritative and influential journals in various disciplines. Because of its strict journal selection standards and citation indexing mechanism, WOS has become one of the most important basic evaluation tools in bibliometrics and scientometrics, as well as a literature retrieval tool.

Therefore, we use the Web of Science core database as data source in the paper; through searching of

“computational social science,” “computational economics,” “computational sociology,” “scientific metrology,” “bibliometrics,” and “big data social analysis,” 7665 articles were selected. To ensure the timeliness of the literature, the interval was set from 2001 to 2020. After excluding journal calls, conference announcements, news, and irrelevant literature, a total of 5856 articles were incorporated into the data of this study.

**2.2. Research Methodology.** Standard bibliometric analysis includes five steps: research design, data collection, data analysis, data visualization, and analysis results (as shown in Figure 1). Biblioshiny bibliometric software package can be used for full-process bibliometric analysis and visual display, including data sets, resources, authors, conceptual structure, knowledge structure, and social structure; SciMAT, as a knowledge mapping analysis tool, is used to explore the evolution of topics in the field by drawing topic coverage maps, evolutionary path maps, and strategic coordinate maps. In this paper, we use Biblioshiny and SciMAT to visualize and network, analyze the literature, and study the knowledge integration and development path of computational social science.

### 3. Analysis of Evolutionary Dynamics

#### 3.1. Statistical Description of the Literature on Computational Social Sciences

**3.1.1. Analysis of the Volume of Articles Issued.** This section will give the answer to the first question: what has been the dynamic evolution of scientific knowledge production in the field of computational social science since its inception as a field of study? The analysis of developmental evolution can be divided into year-by-year or phased time series analysis. Bibliometrics considers that the annual number of publications in a research field is one of the important indicators for evaluating the development status of the field, while phased time series analysis shows the overall trend characteristics by describing different development stages. This paper combines two approaches together.

From 2001 to 2020, there is an overall increasing trend despite slight fluctuations in the number of articles (as shown in Figure 2 and Table 1). In this study, the collected literature spans a large time period and the number of early articles is small, so the literature is divided into four phases, i.e., 2001–2005, 2006–2010, 2011–2015, and 2016–2020, by combining the number of articles and a fixed time window. 2001–2005 is in the low-yield exploration period, the concept of computational social science has not yet been introduced, and the relevant literature grew slowly and did not attract much attention from scholars; 2006–2010 was the budding stage of development, and the number of papers gradually increased; in 2009, the concept of computational social science was formally proposed, which caused a research boom among relevant scholars; 2011–2015 entered the rapid development stage, and especially in 2014–2015 the number of papers grew rapidly; 2016–2020 reached the high-yield active period, and the number of literature continues to increase until the peak in

2020. From the regression trend, it is predictable that the number of literature in the field of computational social science research will continue to grow in the future period.

**3.1.2. Analysis of Annual Citation Trends.** Figure 3 shows that the average number of citations of papers peaked around 2012, and the overall trend of slow growth has been observed from 2014 to the present. At the same time, the average annual citation rate of articles in computational social science did not maintain the same growth rate but experienced several fluctuations, but the number of citations in general still showed an increasing trend, the highest of which was in 2012, when the average number of citations per paper reached 7.8. In that year, 14 European and American scholars, including R. Conte from the Italian National Council for Scientific Research [19], published in the European Journal of Physics. Special Issue No. 1 published a “Manifesto for Computational Social Science,” which discussed the current state of development and future prospects of social science in five aspects: opportunities of the times, technological developments, methodological innovations, challenges of the time, and expected impacts. The average citation rate of articles increased from 2008 to 2014, and most of the research focused on the field of integration and development of social science and algorithmic models. Therefore, combined with the analysis of the number of published papers, this paper believes that while scientific progress is made, especially when a certain technology makes breakthrough progress, research in this field will still show a high growth trend.

**3.1.3. The Highest Cited Articles.** Figure 4 shows that, among 5856 articles, clustering by fast search and finding of density peaks, with up to 2180 citations, proposes an analysis method based on the idea of clustering that can effectively eliminate outliers [20]. Next, private traits and attributes are predictable from digital records of human behavior, with 942 citations. Kosinski et al. [21] found that digital behavioral records like “Facebook Likes” can be used to predict a range of highly sensitive personal attributes, including sexual orientation, ethnicity, religious and political views, and personality traits, thus proposing a demographic model that can predict individual psychology from preferences. As mentioned, computational social science is an emerging interdisciplinary discipline that is the result of long-term knowledge accumulation in traditional social sciences and is a multidisciplinary and problem-solving oriented research.

**3.1.4. Analysis of the Most Relevant Authors.** Among the authors involved in the field of computational social science, Moat et al. had H-index, G-index, and citation frequency of 10, 13, and 444, respectively (Table 2). Meanwhile, many author collaboration subnetworks were generated in the author collaboration graph (Figure 5), with six subnetworks centered on Moat, Preis, and Lazer forming a relatively complex relationship. The results suggest that in computational social science, networks are networks of

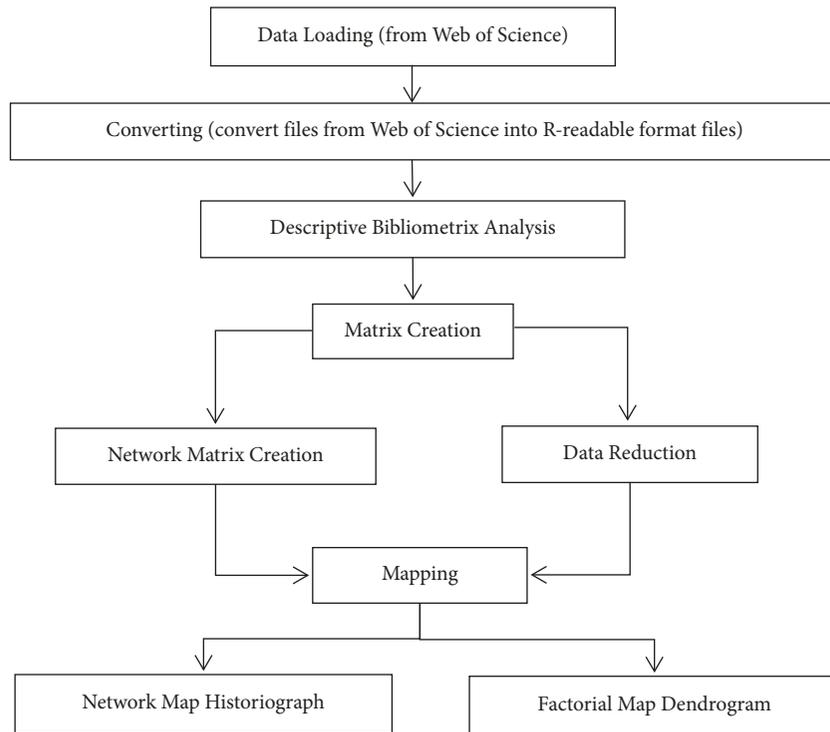


FIGURE 1: Bibliometric analysis process.

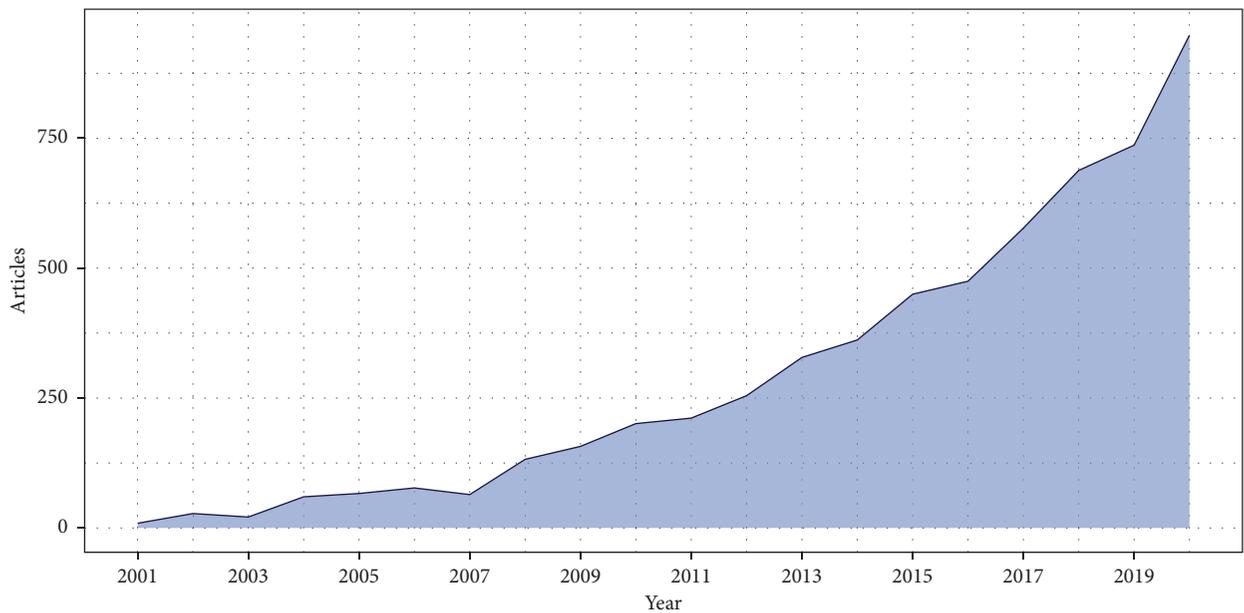


FIGURE 2: Academic papers published related to computational social science, 2001–2020.

TABLE 1: Valid literature and high frequency keywords in different periods.

Periods	2001–2005	2006–2010	2011–2015	2016–2020	Total
Valid literature	185	632	1608	4331	5856
High-frequency keywords	7	29	97	204	337

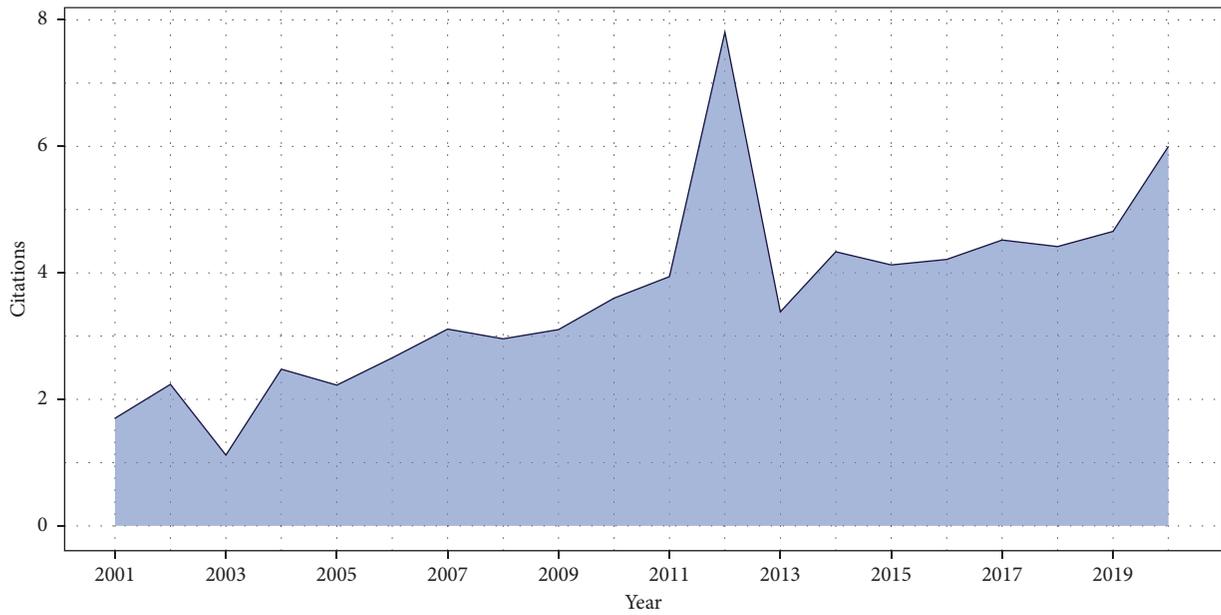


FIGURE 3: Average annual citations in computational social science, 2001–2020.

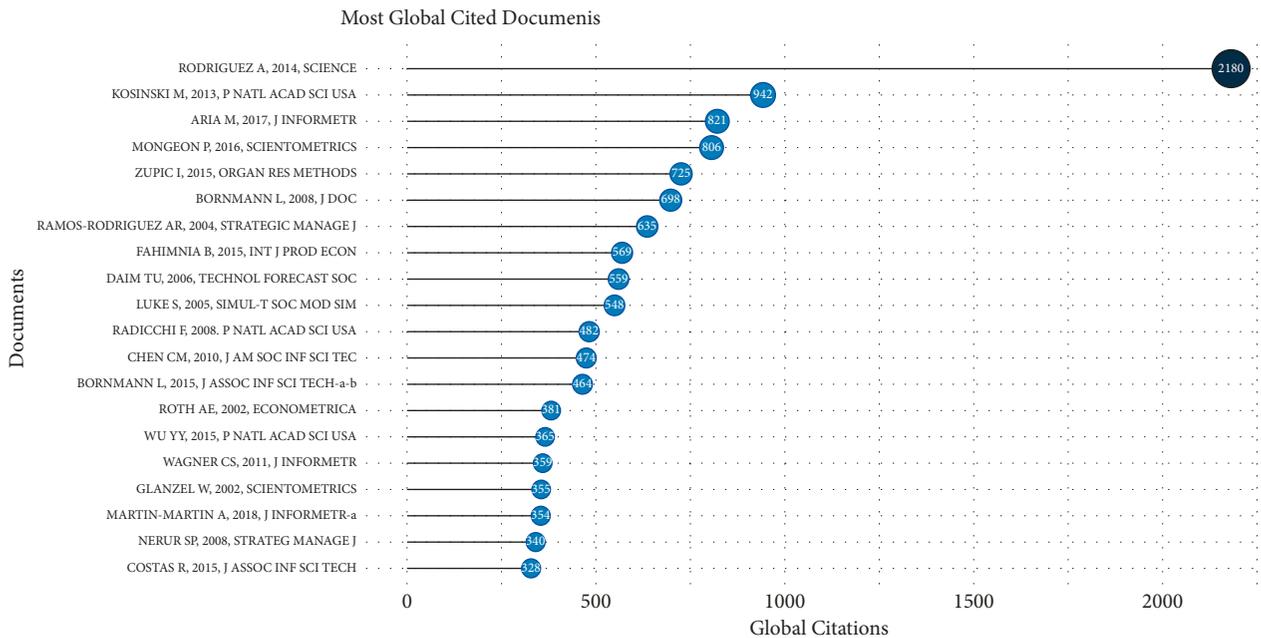


FIGURE 4: Top 20 cited literature.

relationships between nodes, new network structures are formed between subnetworks as different nodes may form subnetworks, and the existence of subnetworks increases the complexity of the network structure to some extent. Moat et al. used social media behavioral data to simulate human mobility patterns [22] and big data to quantify crowd size [23]. In 2009, Lazer et al. published a paper entitled Computational Social Science, which marked the birth of this intersection; in 2020, they published Opportunities and Challenges in Computational Social Science, while the paper Opportunities and Challenges for Computational Social

Science published in 2020 reflected on the development and current state of research in the field of computational social science. Although Lazer et al. are not ranked high in the number of publications, they have made indelible contributions to the generation and development of computational social science.

Because the field is actually very widely distributed and highly interdisciplinary, the knowledge and thinking used come from many other fields as well. Among the front-line scholars in this field, there are Albert-Laszlo Barabasi from a physics background, Duncan Watts from an engineering

TABLE 2: Top 20 authors with significant impact in computational social science.

Author	h_index	g_index	m_index	TC	NP	PY_start
Moat HS	10	13	1.111	444	13	2013
Preist	8	11	0.889	407	11	2013
Cloff-Revillac	5	8	0.294	626	8	2005
Bail CA	6	7	1.000	279	7	2016
Stanley HE	6	7	0.545	392	7	2011
Burnap P	6	6	0.857	240	6	2015
Carley KM	5	6	0.250	459	6	2002
Cheny	6	6	0.375	319	6	2006
Juergens P	6	6	0.600	237	6	2012
Jungherr A	5	6	0.500	229	6	2012
Squazzoni F	5	6	0.357	114	6	2008
Brandt F	4	5	0.308	64	5	2009
Kosinski M	5	5	0.556	1402	5	2013
Lepri B	2	3	0.500	10	5	2018
O'Brien DT	4	5	0.571	94	5	2015
Shah DV	3	5	0.429	127	5	2015
Conte R	4	4	0.400	202	4	2012
Ferrarae	4	4	0.400	189	4	2012
Garcia D	4	4	0.444	77	4	2013

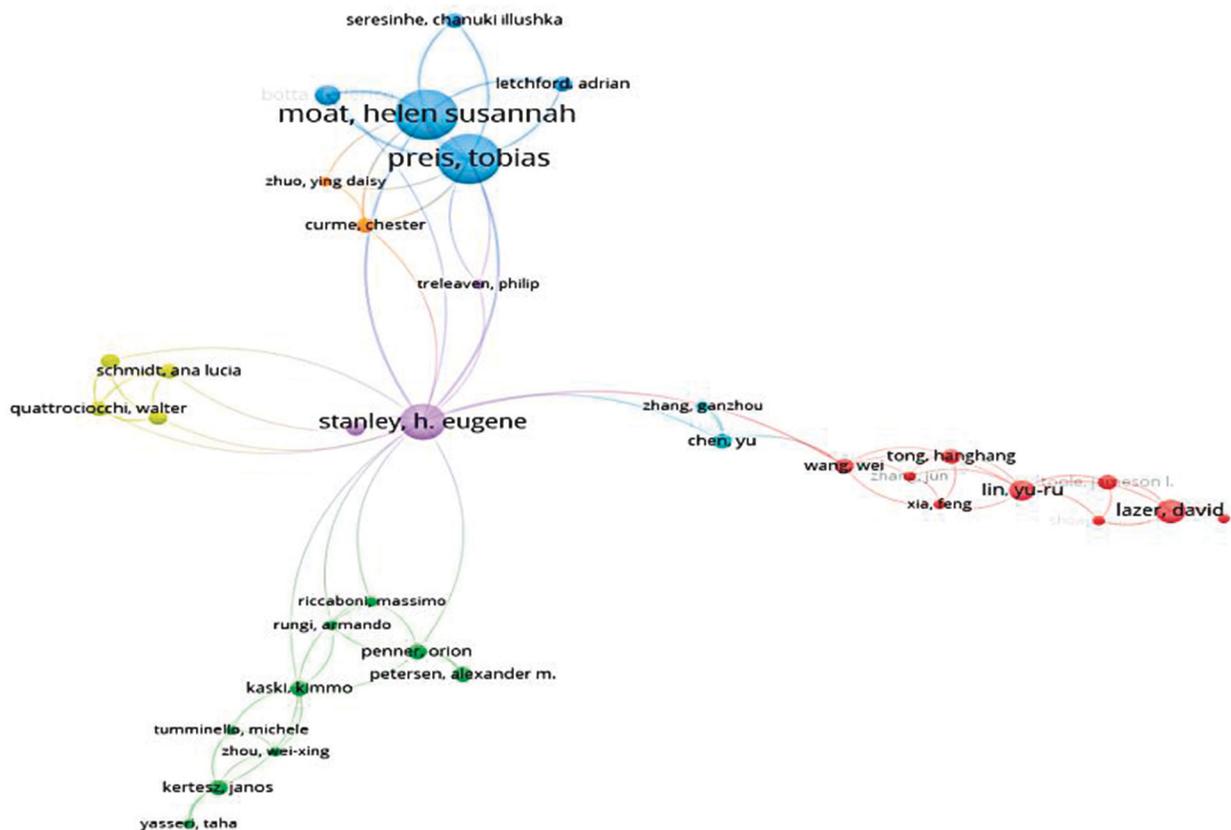


FIGURE 5: Collaborative network of scholars related to computational social science.

background, Nicholas Christakis from a medical background, Joshua Blumenstock from a computer background, Sinan Aral from an information systems background, and Michael Macy from a sociology background (David Lazer by political science). It can be seen that in the process of studying computational social sciences, mathematics, statistical mechanics, complexity science, network science,

natural language processing, artificial intelligence, and social science theories are all knowledge that may be applied.

*3.1.5. Time Analysis of Top Authors' Publications.* In this study, the top 20 most prolific scholars in the field are reflected in terms of the number of publications per year to

explore the research output pattern of scholars (Figure 6). It can be seen that after the first publication in 2009, Abramo G. has maintained at least 4 publications per year in the field, and as many as 15 in 2011, including *A Heuristic Approach to Author Name Disambiguation in Bibliometrics Databases for Large-Scale Study Evaluation* (Figure 6) [24] which attracted wide attention and was cited 128 times for proposing a heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. As can be seen in Figure 6, many prolific scholars began to pay attention to the development of this field only after the formal introduction of “computational social science” in 2009. However, as a cross-discipline, there are more and more high-quality authors involved, and the number of related publications has been on the rise since then.

**3.2. Keywords Analysis.** This section aims to answer the second question: how have the main research directions in the field changed since the advent of computational social science? Keywords, which are the core and essence of the literature, are highly condensed and summarized. Among the 5856 scientific documents analyzed, this paper identifies 337 high-frequency keywords for word cloud analysis and cooccurrence matrix visualization.

**3.2.1. Analysis of the Application of High-Frequency Keywords.** The word cloud in Figure 7 shows the most frequently appearing words in papers of computational social science, showing different sizes according to their frequency of appearance, among which the most common keywords include science, big data, model, and bibliometrics.

Computational social science belongs to the field of social science, integrating theories from various disciplines, and is a theoretical and methodological system for human beings to deeply understand society, transform society, and solve complex social problems in the fields of politics, economy, culture, etc. [25]. Data are the cornerstone of human civilization, and data models are simulations and abstractions of the real world, which make it easy for us to understand the essence behind the phenomena; for example, some emerging well-resourced administrative data research models that can both protect privacy and analyze microlevel data [26].

The era of big data challenges traditional disciplinary boundaries and research paradigms. In the era of big data, the volume of data has proliferated and the data collected are often unstructured. Traditional social sciences are gradually showing symptoms of weakness in explaining and solving such social problems, and more and more problems need to be studied using data-driven approaches. Computational social science enables the application of algorithms and computational tools to complex data and even unstructured data, providing us with the means to model and perform high-performance computation to solve real-world problems.

**3.2.2. Keywords Cooccurrence Network Analysis.** High-frequency words in different time spans were selected to construct keyword cooccurrence matrices, and the Cooccurrence Network of Biblioshiny was used to visualize and analyze the cooccurrence matrices in different time series to discover the correlations among theme words and reveal the evolution of keywords in the process to provide support for related research in this field [27]. In this paper, the keyword cooccurrence network was constructed based on four different time spans of literature issued: low-yield exploration period (2001–2005), budding development period (2006–2010), rapid development period (2011–2015), and high-yield active period (2016–2020). The cooccurrence analysis was performed on the basis of the time to form the cooccurrence mapping shown in Figures 8–11. The size of nodes and fonts in the view is determined by the density of the point. Larger nodes and fonts imply a higher density of the node, indicating that the keyword has a higher frequency of cooccurrence with other keywords; i.e., researchers in the field are paying more attention to it and it is a hot topic, and vice versa. The line between a node and another represents the appearance of two nodes in the same literature, and the color of the line indicates the frequency of the two nodes in the same article.

From 2001 to 2005, computational social science was not yet explicitly proposed, and the research was mainly bibliometric, scientific, and impact-oriented. With each keyword cited infrequently, even some keywords appeared separately. During this period, research focused on social research and science measurement, which laid a solid theoretical and technical foundation for the birth of computational social science in terms of basic theory, professional knowledge and technical methods.

In 2009, Lazer et al. introduced the concept of “computational social science” for the first time, marking the birth of computational social science, and at the same time, big data computing was incorporated into the system. Since then, research on computational social science has received extensive attention from academia and has produced a large number of relevant research. It is foreseeable that the future computational model will become an important infrastructure of society; the information technology of intelligent interconnection of everything will describe the real world in a richer way; experts and scholars in sociology and other fields will have a better understanding of the social system on which human beings live.

The impact of computational social science was becoming increasingly widespread in 2011–2015, and the new methods it offers are applicable to research based on big data. Through big data analysis, simulation studies are conducted to improve management and enhance the science of policy decision-making and evaluation, thus driving economic growth. It can be said that computational social science is a product of the development of social science knowledge, economic and social development, data collection and analysis, network and computing infrastructure, and algorithmic models.

From 2016–2020, big data emerged with increasing frequency and impact, and related research became a central



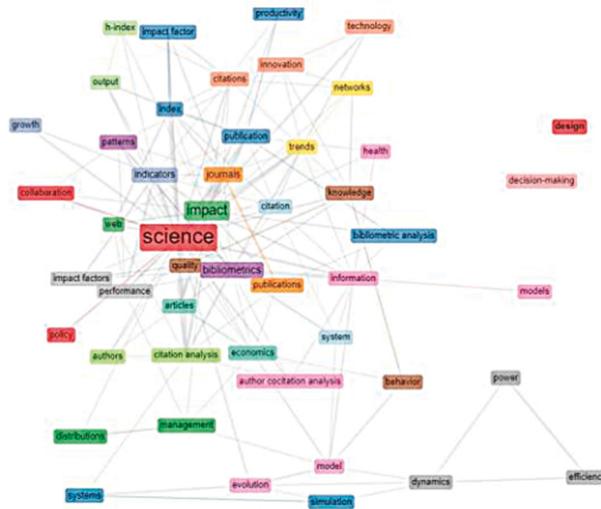


FIGURE 9: 2006–2010 cooccurrence network.

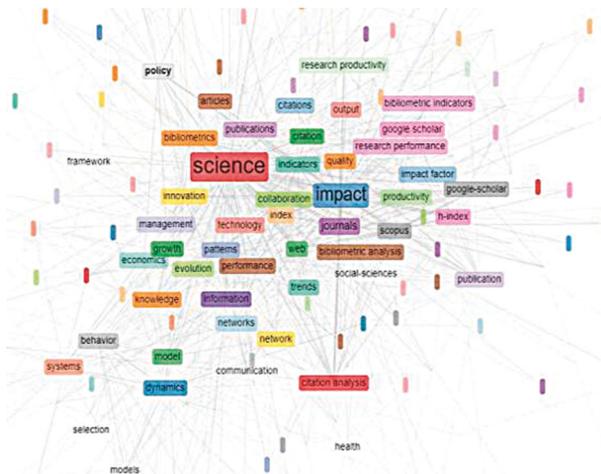


FIGURE 10: 2011–2015 cooccurrence network.

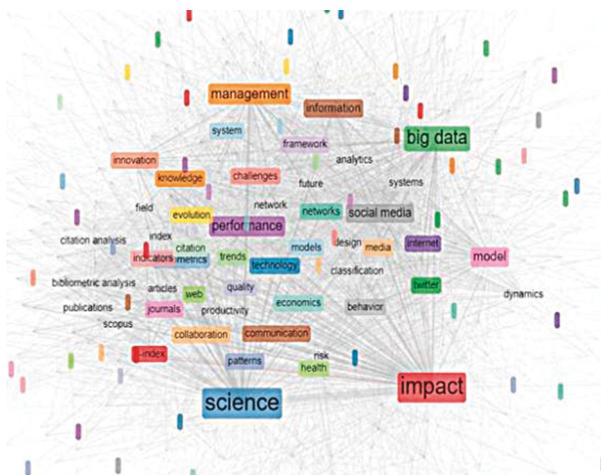


FIGURE 11: 2016–2020 cooccurrence network.





the close or distant relationships between keywords in the field of computational social science, as shown in Figure 13.

Multiple correspondence analysis (MCA) is a common method that compresses large data with multiple variables into a low-dimensional space to form an intuitive two-dimensional graph that uses planar distances to reflect the similarity between keywords. Keywords close to the center indicate researching focus in recent years, as shown in Figure 14.

Theme evolution analysis studies the changing rules, evolutionary relationships, evolutionary paths, and evolutionary trends of theme, intensity, and structure over time, which played an important role in demonstrating the development of the field while grasping the direction of development and predicting the trends of the field. In this paper, with the help of SciMAT, a knowledge graph visualization tool for topic evolution analysis, we can visually observe the hot topics and the relationships between topics in the field of computational social science at different time, so as to understand the process of topic words evolution and provide theoretical support for subsequent research.

*3.2.4. Keywords Clustering and Multiple Correspondence Analysis.* According to the keyword hierarchical clustering analysis, the high-frequency keywords are divided into three clusters by the dividing line with a height of 1, as shown in Table 3.

Firstly, cluster 1 is constructed with “social media” and “big data” as the core keywords. We find that the emergence of social media platforms in recent years has sparked an explosion of social data and thereby attracted a growing desire and demand from academia for research in the field of social media and social data analysis. However, research of this scale requires a high level of expertise in computational and data science domains, which limits the researchers who can undertake social media data-driven research to those with computational expertise or those who have contact with these experts for their research (people who are part of the team) [30]. But scientific methods such as the Social Media Macroscopic (SMM) provide computational social science solutions that aim to remove this limitation and make social media data, analysis, and visualization tools available to researchers and students of all professional levels.

Secondly, cluster 2 is formed with keywords such as “bibliometrics”. With the continuous increase of computer system functions and the continuous proliferation of literature information, bibliometric research is increasingly dependent on computers, and the research on quantitative analysis tools in computers plays a pivotal role. Computer-aided information measurement analysis research has become an important content and a new trend in the development of information measurement research [31].

Cluster 3 is formed with “model” and “network” as the core keywords. Since the establishment of computational social science, researchers have never stopped innovating and optimizing various models that can solve social problems. This year, the introduction of complex network theory has injected new vitality into computational social science and has quickly become a field of research direction.

Meanwhile, in the keywords multiple analysis, Cluster 1 (red) is the largest cluster, indicating that computational social science research as a whole is biased toward social science, focusing on areas such as social media, big data, network science, and human behavior analysis, meanwhile emphasizing the practical application of research methods such as modeling, social network analysis, topic modeling, text mining, and simulation. In addition to producing meaningful results in the academic community, problem-solving-centered disciplinary research also contributes to a more replicable, cumulative, and coherent science [32]. For example, social science plays an important role in the field of machine learning, which, unlike machine learning in other fields, requires not only applying machine learning methods, but also working with social scientists to perform error analysis and demonstrate uncertainty [9]. Jackson [33] proposed an approach to build framework software in the social science, including software frameworks for solving social science computational problems, and validation of models. In addition, several scholars have discussed how social systems simulation should evolve to have a greater impact on the field of social science by documenting the progress of several key approaches to social systems’ modeling, including game theory, statistics, and computer modeling [34]. Bosnjak et al. [35] used computational and network techniques to improve the accuracy of identifying rare phenomena. Compared with qualitative, crude quantitative methods, the technique can be applied to other research topics, and it well illustrates how the proper implementation of computational methods can effectively identify rare events and bridge the gap between deductive and inductive approaches to scientific inquiry. In fact, as the results of the new round of technological revolution, such as big data, cloud computing, and artificial intelligence, continue to be introduced into the field of social science research, the natural and thinking sciences will achieve a high degree of integration and innovation with the social sciences at a greater span and deeper level [2].

*3.2.5. Theme Evolution Path Analysis.* Based on the 2001–2020 literature in fields related to computational social science, a visual analysis of topic evolution over time is performed, and the results help analyze the flow conditions of different topics in the field and elucidate information such as flow directions and transition relationships. Each node in Figure 15 represents a topic, the node size is proportional to the number of keywords contained in the topic, the flow between nodes represents the evolutionary direction of the topic, and topics in adjacent study time zones are connected to represent the continuity between them. The visual characteristics of the lines are width and color. The width is used to indicate the number of shared keywords: the thicker the line, the higher the relevance, and the color helps to distinguish different topics.

From the evolutionary path diagram (Figure 15) and the evolutionary state of each period, it is clear that research in the field of computational social science is in a developmental stage and has not yet matured. The research themes

TABLE 3: Keywords clustering.

Cluster category	Keywords
Cluster 1	Twitter, information, social media, big data, media, challenges, communication, Internet.
Cluster 2	Collaboration, field, h index, publications, journals, citation analysis, articles, publication, citation, indicators, index, bibliometrics, quality, metric analysis, impact, science, citation.
Cluster 3	Framework, classification, behavior, systems, model, models, system, design, patterns, technology, innovation, network, performance, knowledge, risk, health, management, networks, dynamics, trends, evolution, economics.

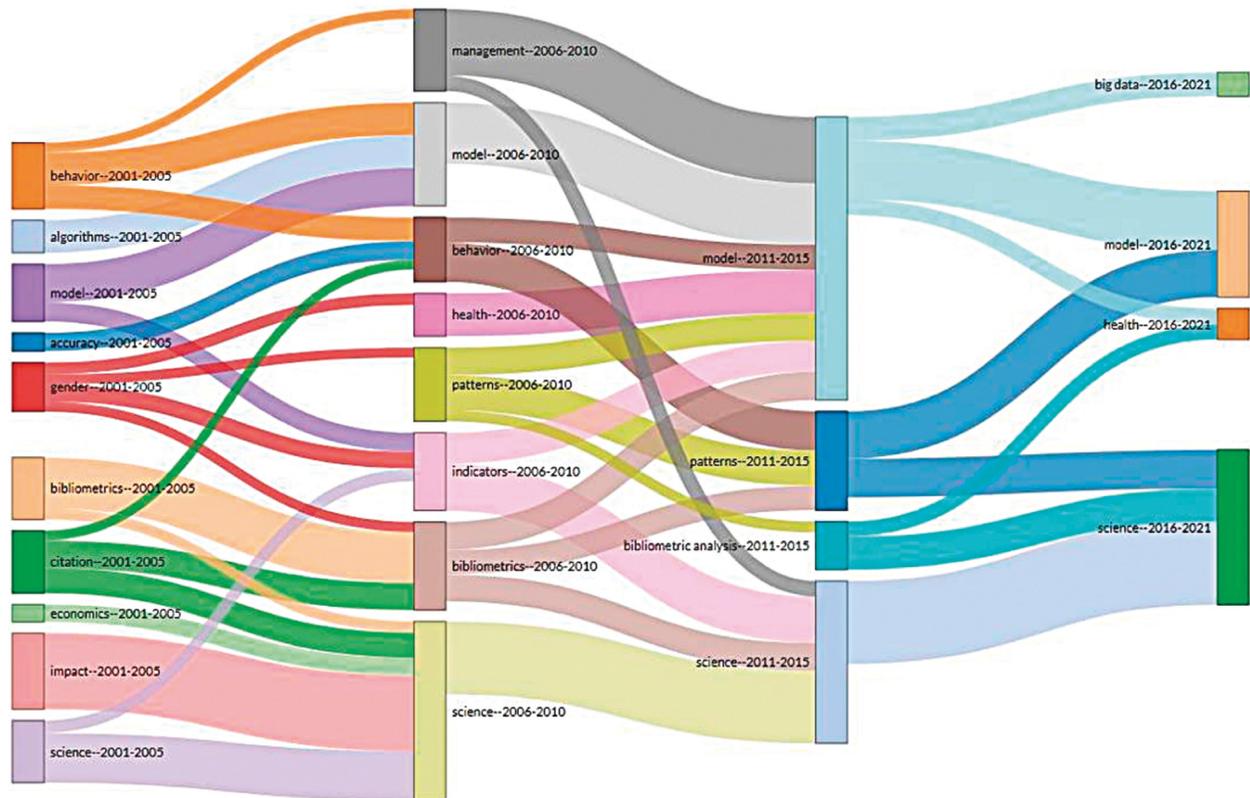


FIGURE 15: Thematic evolution of computational social science.

vary in different periods, with complex evolutionary relationships and unstable processes; meanwhile the differentiation, integration, transfer, and regeneration of themes are evident. Since its development, three evolutionary paths have been formed: (1) behavior  $\rightarrow$  management, model, behavior  $\rightarrow$  big data, science (2) bibliometric  $\rightarrow$  bibliometric  $\rightarrow$  science (3) impact  $\rightarrow$  science  $\rightarrow$  science.

Society is a complex system that combines the development of emerging disciplines and has characteristics that can be modeled or reconstructed by digital data [13]. Social science can help to understand social phenomena, and algorithms can help to make decisions, while the former need to open the black box of data mining algorithms [36] in order to “make a big difference”. In the era of big data, more and more human activities will exist in various databases, generating large-scale data on human behavior and providing the possibility to obtain patterns of human behavior and social processes. In a way, computational communication science, social media, and big data are reshaping

social impact. That is, based on the scale of interpersonal data, scholars have had to reassess existing assumptions about connectivity, exposure, and social influence [37]. Computational social science has led to a paradigm shift in traditional social science, particularly communication research [10], by narrowing social science to a narrower field that relies on knowledge of computational statistics to solve problems [38]. However, the development of computational social science has some shortcomings, and many institutional structures are still in their infancy. Thus collaboration should be strengthened, new data infrastructures should be improved, and ethical legal and social implications should be focused, and scientific institutions should be reorganized [12].

**3.2.6. Theme Evolution State Analysis.** In this paper, we analyze the evolutionary state of clustering themes by using strategic coordinate diagrams from the degree of relevance

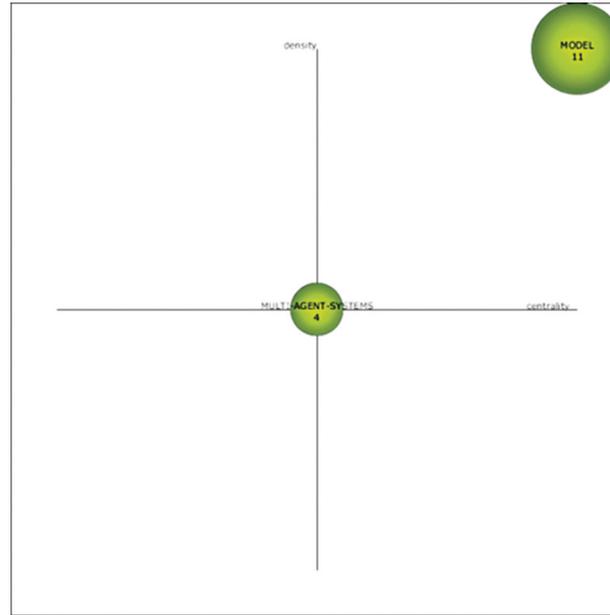


FIGURE 16: 2001–2005 strategic coordinate chart.

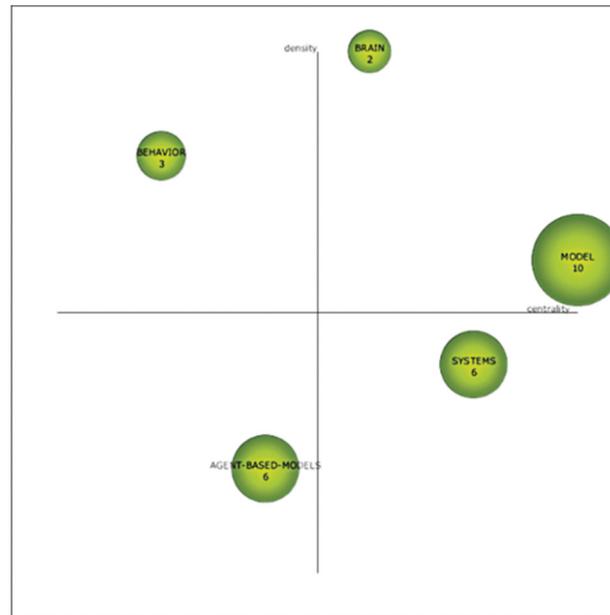


FIGURE 17: 2006–2010 strategic coordinate chart.

and developable requirements of the subject terms (Figures 16–19). In the figure, the nodes represent the clustering themes, and the size of the nodes is proportional to the literature volume, i.e., the more literature volume, the higher the degree of attention and the hotter the research. Density index is used as the vertical coordinate and centrality index as the horizontal coordinate in the figure. Density represents the strength of connection of basic knowledge units within a single topic, and centrality represents the strength of connection between a topic and other

topics. The right-angle coordinate system is divided into four quadrants according to the density and centrality values, the first quadrant is motor themes, which are highly mature core themes; the second quadrant is developed and isolated themes, which are highly mature isolated themes. The third quadrant is for emerging or disappearing themes, which are new or disappearing themes; the fourth quadrant is for basic and transversal themes, which are basic themes with low maturity and are most likely to become research hotspots or trends in the future.

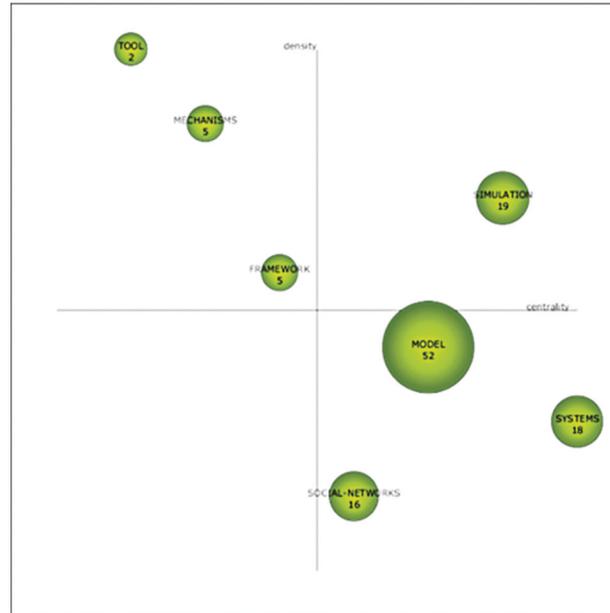


FIGURE 18: 2011–2015 strategic coordinate chart.

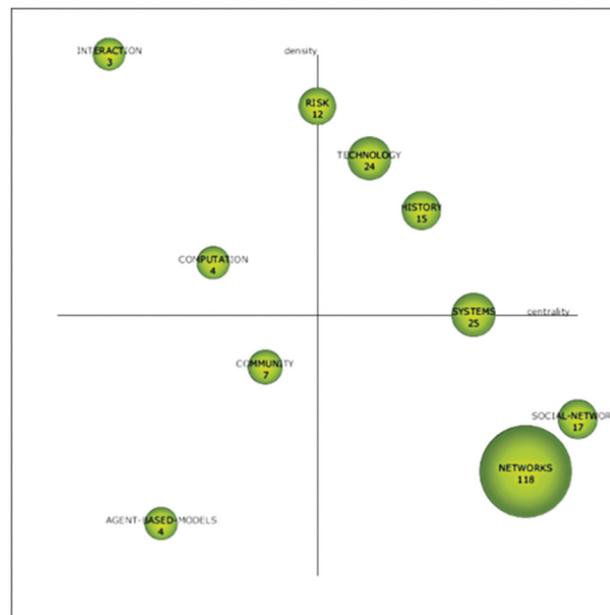


FIGURE 19: 2016–2020 strategic coordinate chart.

This section is also divided according to four different time periods of posting volume, as a way to observe the evolutionary status of the theme terms. In the low-yield exploration stage, because the concept of computational social science has not yet been proposed, there are fewer relevant studies, and no system has been formed, with single topic focusing on aspects such as the construction and optimization of model algorithms, which aim to provide objective basic conditions for computational social science in terms of technical methods. At the budding stage, the concept of computational social science was formally

proposed, attracting the attention of a large number of scholars in related fields. In the process of model algorithm construction and optimization, subject-based models were introduced and widely used in fields such as social networks and economics. Although not yet mature, system simulation and social simulation based on bibliometrics and scientometrics are highly connected and have become powerful tools to promote computational social science. After 2010, the influence of computational social science is growing and the research topics are rapidly increasing, so the fields of social networks and social simulation in the context of digital

networks have received the attention of many scholars. Scientific knowledge, theoretical methods, and organizational models from different fields cross collide and couple with each other in the process of solving social problems, and the knowledge fusion among them has changed the traditional concept of computational social science and gradually formed a new development path.

In 2016–2020, human-computer interaction, risk analysis, and community began to receive academic attention, but most of the topics were not well connected with other topics. After the system model was proposed, it is still the focus of current research after nearly 20 years of development. In addition, areas such as social media and government governance in the context of big data have gradually become research hotspots. Folgado and Sanz collected data on the tweets of leaders on Twitter and used various data science tools to perform neural network calculations in which the accuracy of the political leanings of the tweets was able to reach about 90% [35]. From the evolution of the subject terms in the four periods in Figures 16–19, it is obvious that computational social science, as an interdisciplinary discipline, has begun to diversify its research topics and interact more frequently with each other, promoting the integration of knowledge among related disciplines.

As seen in Figure 19, the subject terms social networks and networks have extremely high centripetal degrees of 1.0 and 0.9, respectively, indicating frequent connections between social networks and other fields, but low density of 0.3 and 0.2, respectively, judging from the size of the two nodes, shows that the volume of literature on this topic has been high but not yet mature. It is inferred that this field has great development prospects and is most likely a research focus for the future development of computational social science.

#### 4. Conclusion and Outlook

This paper collects 5856 papers in research areas related to computational social sciences and uses descriptive analysis, keyword clustering analysis, and topic evolution analysis with the help of bibliometric analysis and knowledge graph visualization to conduct a comprehensive analysis and to summarize computational social science research from the perspectives of annual publication, keyword cooccurrence, research trends, and topic evolution. This essentially defines the direction of current and future research in the coming years.

The study shows that the number of papers in computational social science has gradually increased in the past 20 years and can be divided into four stages with time, namely, low-production exploration period, budding development period, rapid development period, and high-production active period. The trend of annual increase in the number of publications confirms the academic interest in the field, which leads to the judgment that the literature production in the field of computational social sciences is likely to remain in an active period of high production in the coming years.

Since the introduction of the discipline, the main research directions include big data, model algorithms, and

scientometrics. Computational social science is a typical interdisciplinary discipline, which emerged from the rapid progress of computing technology and in-depth research on social science. It collects and analyzes data in large quantities with unprecedented breadth, depth, and scale and implements computational modeling methods to predict the behavior of sociotechnical systems such as human-computer interaction. It is a discipline made possible by the joint development of social science knowledge, economic and social needs, data collection and analysis, network and computational infrastructure, and algorithmic models and is a product of the evolution of the scientific research paradigm from experimental and theoretical science to computational and exploratory science. Thus, computational social science has fused different disciplinary paradigms, promoted research in related fields previously neglected by a single discipline, increased communication and integration across disciplines, and formed many emerging frontier disciplines. With “problem-solving” research as the focus, computational social science has changed the traditional conception of social science, gradually formed a new development path for the discipline, and will surely promote the solution of many important practical problems in the future.

At the same time, big data, modeling algorithms, and social networks are high-frequency keywords in the field of computational social science in recent years, and they are also the main research directions of computational social science in the coming years due to their large research potential. Due to the paradigm shift of computational social science research and science technology in recent years, the discipline has emerged more in the form of “big data+”, promoting the emergence of big data sociology, big data political science and other disciplines, and making the development of new disciplines happen. In addition, there are many themes in the field of computational social science at different times with complex evolution of the relationship. From the analysis of keyword hotspots, the main focus in recent years is on big data, science, social impact, and other fields. From the analysis of the evolution of the theme, social computational science will carry out research in the fields of bibliometry, social behavior, influence, and so on.

This study has some limitations, which mainly include the following aspects: (1) the subject vocabulary of the retrieved articles is different, leading to discrepancies with other literature reviews. (2) The data come from WoS databases, and literatures from databases such as WoS or Google Scholar are not included. (3) This paper mainly applies the Biblioshiny package in R language and SciMAT to conduct the study, and different tools may have identified different clusters and lines of research. In future, with the development of databases, bibliometric software, and other scientific technologies, scholars can conduct more comprehensive and systematic quantitative research on computational social sciences.

Computational social science, as an emerging interdisciplinary research field, has had a wide impact in sociology, economics, political science, management, psychology, and other social sciences. However, the research in this field has emerged in a relatively short period of time, and its

application in social science research is still in its initial stage, with factors such as fragmented theories, extreme lack of comprehensive talents with multidisciplinary backgrounds, knowledge constraints of the original disciplinary framework, and many uncertainties in the future. In the later research, scholars should draw on the achievements of scientific paradigms in the field of computational social science, continue to study “big data+”, model simulation, social network analysis and other directions, promote the integration and development of disciplines, and devote themselves to solving the real problems of society. Therefore, the development of computational social science still needs the joint efforts of more like-minded people.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Authors' Contributions

The authors Yuxi Liu and Xin Feng contributed equally to this work.

### Acknowledgments

The authors acknowledge the financial support from the National Natural Science Foundation of China (Grand No. 11905042,11875005); Natural Science Foundation of Hebei Province (Grand No. G2021203011); and Association of Social Science Development of Hebei Province (Grand No.20210501003).

### References

- [1] J. H. Han, Y. Zhang, and Z. F. Li, “Computational social science—origins, changes and challenges,” *Tsinghua Social Science*, vol. 1, no. 2, pp. 118–132, 2019.
- [2] X. Z. Ding, “Theoretical exploration and development prospects of computational social science,” *Journal of Zhaoqing College*, vol. 42, no. 04, pp. 1–6, 2021.
- [3] D. Lazer, A. Pentland, and L. Adamic, “Computational social science,” *Science*, vol. 323, no. 5915, pp. 721–723, 2009.
- [4] J. H. Han and Z. F. Li, “Computational social science: Meaning, characteristics and prospects—an interview with experts in computational social science in the United States,” *Scientology Research*, vol. 36, no. 10, pp. 1729–1736+, 2018.
- [5] K. D. Jong, C. Cioffirevilla, and J. Bassett, *Evolutionary Computation and Agent-Based Modeling: Biologically-Inspired Approaches for Understanding Complex Social systems[M]*, Kluwer Academic Publishers, Amsterdam, Netherlands, 2012.
- [6] M. D. Laubichler, J. Maienschein, and A. J. Renn, “Computational perspectives in the history of science: to the memory of Peter Damerow,” *Isis*, vol. 104, no. 1, pp. 119–130, 2013.
- [7] Y. Sohn, “Strategic foundation of computational social science,” in *Proceedings of the International Conference on World Wide Web*, ACM, Seoul Korea, April 2014.
- [8] G. Bravo and M. Farjam, “Prospects and challenges for the computational social sciences,” *Journal of Universal Computer Science*, vol. 23, no. 11, pp. 1057–1069, 2017.
- [9] H. Wallach, “Computational social science ≠ computer science + social data,” *Communications of the ACM*, vol. 61, no. 3, pp. 42–44, 2018.
- [10] T.-Q. Peng, H. Liang, J. Jonathan, and H. Zhu, “Introducing computational social science for Asia-Pacific communication research,” *Asian Journal of Communication*, vol. 29, no. 3, pp. 205–216, 2019.
- [11] X. Tao, J. D. Velasquez-Silva, J. Liu, and N. T. Zhong, “Editorial: computational social science as the ultimate web intelligence,” *World Wide Web*, vol. 23, no. 3, pp. 1743–1745, 2020.
- [12] J. Zhang, W. Wang, and F. Xia, “Data-driven computational social science: a survey,” *Big Data Research*, vol. 21, 2020.
- [13] P. Tornberg and J. Uitermark, “For a heterodox computational social science,” *Big Data & Society*, vol. 8, no. 2, 2021.
- [14] X. Y. Leung, J. Sun, and B. Bai, “Bibliometrics of social media research: a co-citation and co-word analysis,” *International Journal of Hospitality Management*, vol. 66, pp. 35–45, 2017.
- [15] X. Zou, W. L. Yue, and H. L. Vu, “Visualization and analysis of mapping knowledge domain of road safety studies,” *Accident Analysis & Prevention*, vol. 118, no. SEP, pp. 131–145, 2018.
- [16] M. G. Marques Tânia, “Research on public service motivation and leadership: a bibliometric study,” *International Journal of Public Administration*, vol. 44, no. 7, pp. 591–606, 2021.
- [17] A. Qjw, A. Zjr, and L. B. Hui, “The progress and trend of BIM research: a bibliometrics-based visualization analysis,” *Automation in Construction*, vol. 124, 2021.
- [18] E. Abad-Segura, A. Infante-Moro, M. D. Gonzalez-Zamar, and E. Lopez-Meneses, “Blockchain technology for secure accounting management: research trends analysis,” *Mathematics*, vol. 9, no. 14, 2021.
- [19] R. Conte, N. Gilbert, G. Bonelli et al., “Manifesto of computational social science,” *The European Physical Journal - Special Topics*, vol. 214, no. 1, pp. 325–346, 2012.
- [20] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [21] K. Michal, S. David, and G. Thore, “Private traits and attributes are predictable from digital records of human behavior,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 5802–5805, 2013.
- [22] D. Barchiesi, T. Preis, S. Bishop, and H. S. Moat, “Modelling human mobility patterns using photographic data shared online,” *Royal Society Open Science*, vol. 2, no. 8, 2015.
- [23] F. Botta, H. S. Moat, and T. Preis, “Quantifying crowd size with mobile phone and Twitter data,” *Social Science Electronic Publishing*, vol. 2, no. 5, 2015.
- [24] C. A. D’Angelo, C. Giuffrida, and G. Abramo, “A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 2, pp. 257–269, 2011.
- [25] S. F. Meng, Y. Li, and J. H. Zhu, “Social computing: opportunities and challenges in the era of big data,” *Computer Research and Development*, vol. 50, no. 12, pp. 2483–2491, 2013.
- [26] H. Margetts, “Political behaviour and the acoustics of social media,” *Nature Human Behaviour*, vol. 1, no. 4, 2017.
- [27] H. Li, H. Y. Chen, M. Li, and L. Y. Wang, “Analysis of research hotspots and theme evolution in the field of machine learning

- based on scientometrics,” *Journal of Shenyang University (Social Science Edition)*, vol. 23, no. 01, pp. 19–25, 2021.
- [28] B. Fahimnia, C. S. Tang, H. Davarzani, and J. Sarkis, “Quantitative models for managing supply chain risks: a review,” *European Journal of Operational Research*, vol. 247, no. 1, pp. 1–15, 2015.
- [29] V. Krishnamurthy, R. Luo, and B. Nettasinghe, “Segregation in social networks: markov bridge models and estimation [M],” in *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5484–5488, IEEE, Toronto, Canada, June 2021.
- [30] J. T. Yun, N. Vance, C. Wang et al., “The Social Media Macroscopic: a science gateway for research using social media data,” *Future Generation Computer Systems*, no. 111, pp. 819–828, 2020.
- [31] R. Zhao and L. Xu, “Visual analysis of transition from bibliometrics to webometrics,” *Information Science*, vol. 29, no. 07, pp. 975–983, 2011.
- [32] M. J. Lazer David, A. Pentland, J. Watts Duncan et al., “Computational social science: obstacles and opportunities,” *Science (New York, N.Y.)*, vol. 369, no. 6507, pp. 1060–1062, 2020.
- [33] P. J. Jackson, “Software solutions for computational modelling in the social sciences[M],” *Intelligent Systems Reference Library*, vol. 52, 2014.
- [34] D. C. Earnest and E. Frydenlund, “Flipping coins and coding turtles[M],” *The Evolution of M & S in the Social Sciences*, vol. 1, pp. 237–259, 2017.
- [35] M. Ring-Ramirez, T. Elliott, and J. Earl, “Spillover as movement agenda setting: using computational and network techniques for improved rare event identification,” *Social Science Computer Review*, vol. 39, no. 5, pp. 981–1002, 2021.
- [36] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, vol. 366, no. 6464, pp. 447–453, 2019.
- [37] JN. Cappella, “Vectors into the future of mass and interpersonal communication research: big data, social media, and computational social science,” *Human Communication Research*, vol. 43, no. 4, pp. 545–558, 2017.
- [38] M. G. Folgado and V. Sanz, “Exploring the political pulse of a country using data science tools,” *Mathematics*, vol. 18, no. 3, pp. 356–373, 2020.