From Digital Traces to AI-Driven Inquiry: The Evolution of Computational Social Science (2009–2025)

Yuqi Zhu[1], Chenyang Wang[1], and Jianxun Chu[2*]

[1] School of Humanities and Social Sciences, University of Science and Technology of China, Hefei 230026, China
[2] School of Marxism, University of Science and Technology of China, Hefei 230026, China
chujx@ustc.edu.cn

Abstract. Computational social science (CSS) is increasingly regarded as a transformative development that may signal a paradigm shift in contemporary social science research. To empirically assess this shift, this study conducts a systematic bibliometric analysis of CSS as a self-identified field. Based on literature retrieved from the Web of Science database, this study analyzes descriptive statistics, co-citation and co-word analyses, as well as thematic mapping, to explore publication trends, major contributing sources, influential authors, and core research themes. It further constructs co-citation networks, author–keyword–source linkages, and thematic structures to identify the field's intellectual foundations, primary research domains, and promising trends. The findings show that CSS has evolved from an early focus on digital traces and social network data to a platform-centered phase of social media, misinformation, and public opinion research, and more recently to an AI-driven phase shaped by NLP, machine learning, and large language models. By mapping this trajectory, the study contributes to social computing by clarifying how computational methods, digital platforms, and AI systems jointly reshape the production of social scientific knowledge.

Keywords: Computational social science · Social computing · Research methods · Paradigm shift · Bibliometric analysis

1 Introduction

Debates over how society can be studied scientifically have long shaped the development of social science. During the Scientific Revolution, natural philosophers demonstrated that the world could be investigated through systematic observation and mathematically formulated principles, establishing a paradigm for modern science. In the nineteenth century, Auguste Comte proposed sociology as a distinct empirical discipline, emphasizing the systematic study of social phenomena. These developments consolidated the boundary between natural and social sciences while leaving open the question of whether social inquiry could undergo its own paradigm shifts in methods, evidence, and explanation [1].

The rise of computational approaches has renewed this question. Large-scale digital traces, platform data, networked interactions, and AI-based analytical tools have reshaped how social behavior can be observed, measured, modeled, and explained. The publication of Lazer et al.'s seminal article in 2009 marked the emergence of computational social science (CSS), and related work has flourished since then [2]. In this study, we treat CSS as a transformative development that bridges long-standing methodological divides in social science and offers a new way to connect empirical observation, computational modeling, and social explanation.

Yet, despite its rapid growth, CSS remains conceptually fragmented and methodologically diverse. Scholars across sociology, political science, communication, as well as applied physics and computer science contribute to the field, but they do so with different perspectives, focuses, and research traditions [3,4,5,6]. This disciplinary dispersion raises fundamental questions about how CSS has actually developed, what research themes have emerged, and how different social science subjects have engaged with the field [7]. Moreover, the development of CSS has been strongly driven by advances in computational techniques, which have also enabled CSS to evolve rapidly. Therefore, we see a critical need for a systematic overview of CSS as a self-identified subject to map its themes, evolution, and trends, as well as its position among associated subjects.

2 Research Methods

2.1 Sample Selection

A key challenge in reviewing literature around CSS is defining the scope of the literature, as both "computational" and "social science" encompass wide and heterogeneous domains. To ensure that the sample reflects the scholarly community that explicitly identifies with CSS, rather than the much broader set of computationally oriented or engineering-driven studies, we adopted a targeted keyword strategy. Specifically, we used the exact terms "computational social science" and "computational social sciences" as the search keywords.

This approach was applied for several reasons. First, our objective was to analyze the scholarly conversation around CSS as a self-identified research domain, rather than to exhaustively capture all computational techniques applied to all social problems. Searching for the precise field label allows us to focus on literature in which authors explicitly position their work within CSS. Second, expanding the search into separate concept clusters would largely expand the scope without guaranteeing completeness, which would introduce many biases by capturing large volumes of purely technical papers that do not engage with CSS. Third, the term "computational social science" is widely used in program descriptions, journal scopes, and by specialized authors, so it provides a clear and consistent label for defining the field.

Based on this strategy, we conducted a search in the Web of Science database, which is known for its comprehensive coverage of multidisciplinary science and its

rich set of downloadable indicators for analysis [8,9]. The search query used was `TS=("computational social science" OR "computational social sciences")`, with the index date set from 2009-01-01 to 2025-12-31, yielding 881 results from the SCIE, SSCI, A&HCI, and ESCI editions. To ensure consistency, we selected only English-language documents, resulting in 859 items as the final sample.

2.2 Data Analysis

To examine the structural and thematic characteristics of the CSS literature, we used mainstream bibliometric tools for bibliometric analysis, namely bibliometrix [10], CiteSpace [11], and VOSviewer [12]. Specifically, CiteSpace and VOSviewer were used for co-citation analysis and keyword co-occurrence analysis, while the bibliometrix R package was used for descriptive statistics, source and affiliation analyses, and visualized keyword plots. These tools were applied for different analytical purposes. Their combined use allows for methodological diversity and enhances the richness of the findings. Moreover, they have been widely adopted across social sciences [13,14], making them well-established and reliable for analyzing scientific literature.

3 Results

3.1 Descriptive Analysis of Retrieved Literature of CSS

To provide an overview of the structural characteristics of the dataset, this section begins with descriptive indicators of the sample. Figure 1 displays the annual count and trend of CSS articles from 2009 to 2025. The trend shows a steady increase, with particularly notable growth from 2021, indicating rising attention to CSS in more recent years. The CSS literature is concentrated in a few key journals, including EPJ Data Science, PNAS, and Social Science Computer Review, Journal of Computational Social Science, and others.

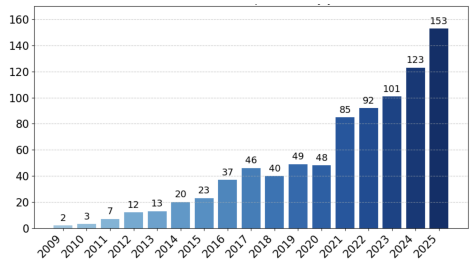

Fig. 1. Annual CSS publication output from 2009 to 2025.

3.2 Keyword Co-occurrence Network and Development

The keyword co-occurrence analysis reveals a highly interconnected intellectual structure within the field of CSS. As shown in Fig. 2, "computational social science" occupies the central and most influential position, indicating its role as the

conceptual hub linking diverse methodological, technological, and application-oriented domains. Surrounding the core node, six coherent thematic clusters emerge, each representing a major research strand within the field.

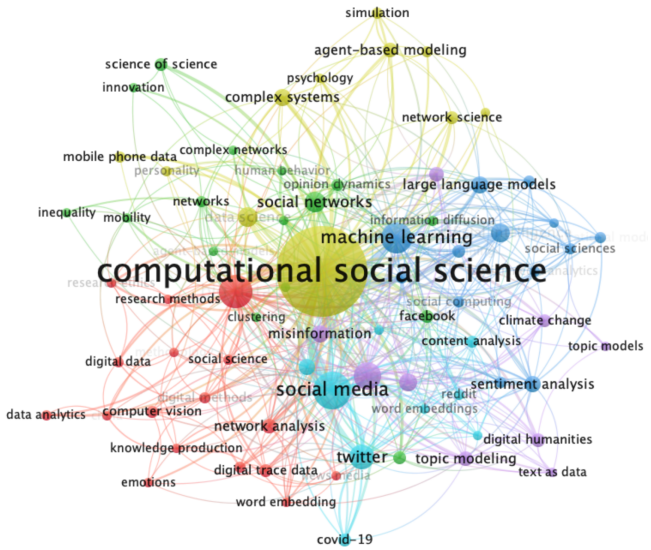

Fig. 2. Keyword co-occurrence map of CSS.

Cluster 1 (in red): Methodological foundations, digital data, and research ethics. It captures the methodological backbone of CSS, centered on research methods, digital data, big data, digital trace data, network analysis, and data analytics. The presence of research ethics also indicates the field's growing attention to the normative challenges raised.

Cluster 2 (in green): Social networks, social simulations, and complex systems. It focuses on networked and complex social structures, with keywords such as social networks, complex networks, agent-based modeling, simulation, information diffusion, and opinion dynamics. It represents a modeling-oriented strand that uses computational approaches to examine emergent social processes.

Cluster 3 (in dark blue): AI and machine learning for social computing. It highlights the growing role of AI, machine learning, large language models, data mining, sentiment analysis, and social media analytics in CSS. It reflects the application of advanced computational techniques to detect, interpret, and model social phenomena in digital environments.

Cluster 4 (in yellow): Disciplinary identity and theoretical core. It represents the conceptual core of CSS, linking computational social science with complex systems, network science, social simulation, experiments, psychology, and data

science. It shows how CSS integrates theories and methods from multiple disciplines to connect individual behavior with macro-level social outcomes.

Cluster 5 (in purple): Text-as-data and issue-focused analysis. It is centered on text-based computational methods, including topic modeling, text-as-data, natural language processing, and text analysis. It demonstrates how CSS uses large-scale textual data to study public discourse, misinformation, polarization, digital humanities, and other issue-oriented topics.

Cluster 6 (in light blue): Social media and social issues. It focuses on social media platforms and online communities, with keywords such as social media, Twitter, Reddit, social network analysis, algorithms, COVID-19, and electric vehicles. It reflects the use of platform data to study communication patterns, information diffusion, and socially relevant events.

3.3 Keyword Development and Evolution of CSS

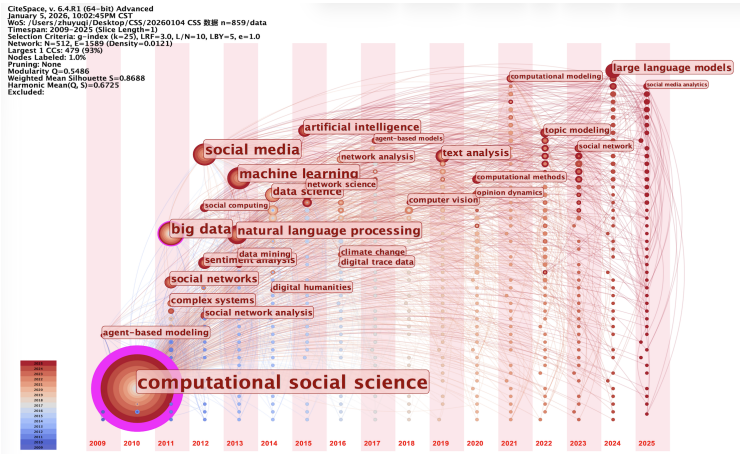

Fig. 3. Keyword timezone view of CSS from 2009 to 2025.

Figure 3 illustrates the temporal development and methodological evolution of CSS from 2009 to 2025, showing a highly interconnected research field in which the largest connected component accounts for 93% of the network.

It reveal a coherent trajectory from modeling-oriented approaches to data-intensive methods and, more recently, to AI-driven analytical techniques. In the early stage, keywords such as computational social science, agent-based modeling, complex systems, and social network analysis formed the conceptual and methodological foundation of the field. From the mid-2010s onward, CSS expanded around large-scale digital data and computational techniques, with big data, social media, machine learning, data science, NLP, and sentiment analysis

becoming increasingly frequent and interconnected. In the recent period, particularly after 2022, AI and LLMs emerged as prominent keywords, indicating the rapid uptake of AI-based methods in CSS research.

This temporal progression suggests that the evolution of CSS is characterized by both continuity and innovation. Earlier network-based and simulation-oriented approaches did not disappear but became connected with newer data-intensive and AI-enabled methods. The field has therefore evolved through the gradual accumulation and integration of computational tools, data sources, and analytical frameworks.

3.4 Co-citation Structure and Influential Studies

The co-citation network highlights several clusters of influential publications that constitute the intellectual foundations of CSS. Early works such as Lazer et al. [2] and Moat et al. [15] established the field's initial focus on digital trace data and large-scale behavioral measurement. Lazer et al. [2], in particular, occupies a central and bridging position in the network, consistent with its status as the most globally cited work in the dataset. Subsequent studies by Kosinski et al. [18] and Youyou et al. [19] extended this foundation by demonstrating the predictive power of digital records for inferring individual traits, thereby consolidating computational prediction as a key methodological pillar of CSS.

A second major cluster centers on social media platforms as social systems. Studies such as Bail et al. [20], Grinberg et al. [17], Vosoughi et al. [16], Vargo et al. [21], Farrell [22], and Stella et al. [23] apply computational methods to polarization, misinformation, political communication, and social bots. Their dense co-citation ties suggest that these works have formed a consolidated research tradition linking large-scale platform data with substantive social and political issues.

More recent methodological advances are represented by studies on text-as-data, NLP, transformer models, and LLM-based social simulation. Grimmer et al. [24] provide an important framework for machine learning and text analysis in the social sciences, while Bail [25] and Argyle et al. [26] illustrate the emerging use of generative AI and language models for social-scientific inquiry. At the same time, critical perspectives such as Bender et al. [27] remind the field of the limitations and risks of large language models. Although these newer works show high normalized citation rates, their more peripheral positions in the co-citation network suggest that their long-term integration into the core intellectual structure of CSS is still unfolding.

3.5 Geographic and Institutional Distribution of CSS

To examine the geographic and institutional distribution of CSS, we analyze publication trends by country (Figure 5) and the most relevant affiliations (Table 1). Together, these results show that CSS research has become increasingly international, but remains concentrated in a limited number of high-capacity countries and institutions.

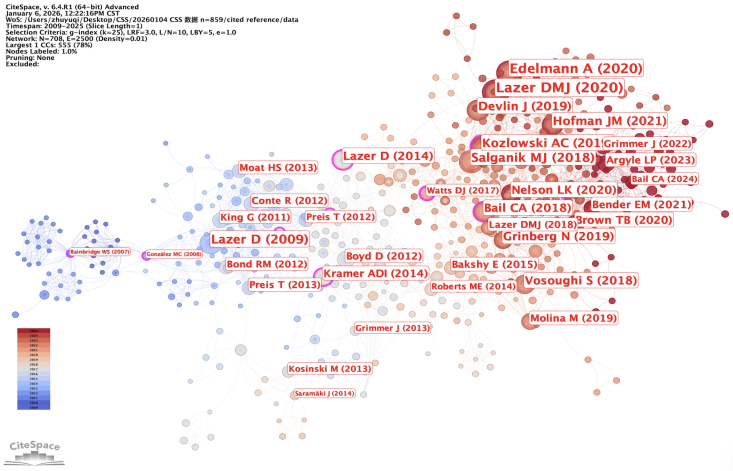

Fig. 4. Co-citation network of CSS.

Figure 5 shows the cumulative number of CSS publications by country for the top 10 contributing nations. The United States exhibits the strongest and most continuous growth throughout the period, followed by the United Kingdom, China, Germany, Italy, and other contributors. While these patterns indicate the gradual internationalization of CSS research, the distribution remains uneven, with a small number of countries accounting for the majority of publications.

The institutional distribution further confirms this concentration. As shown in Table 1, leading universities such as Northwestern University, the University of California System, the University of Oxford, MIT, and the University of Michigan account for a substantial share of CSS publications. Overall, these patterns suggest that the development of CSS is shaped not only by the diffusion of computational methods across countries, but also by the concentration of research capacity within a relatively small group of institutions.

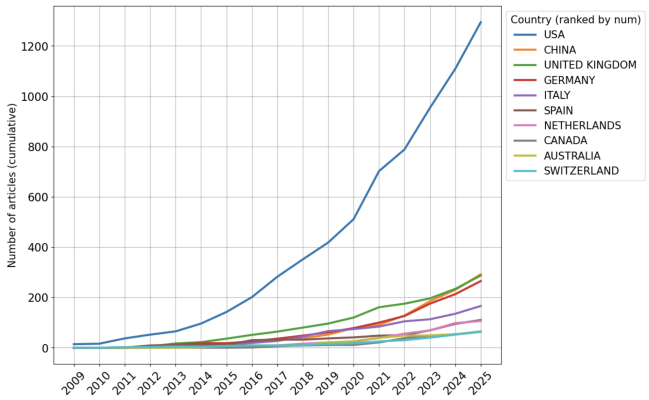

Fig. 5. Trend of article publication over years by countries (Top 10).

Table 1. Most relevant affiliations in the CSS dataset

Affiliation	Articles	Country
Northwestern University	58	USA
University of California System	43	USA
University of Oxford	40	UK
MIT	37	USA
University of Michigan	34	USA
Northeastern University	32	USA
Harvard University	30	USA
Stanford University	28	USA
University of London	27	UK
University of Texas System	26	USA
Duke University	25	USA
University of Amsterdam	23	Netherlands
New York University	21	USA
University of Warwick	21	UK
Princeton University	20	USA

The institutional-level comparison provides additional context for the geographic distribution of CSS research at the global level. Table 1 highlights a high degree of concentration in CSS. A small number of universities account for a disproportionate share of publications, and many of these affiliations are located in countries with the highest national output, mainly the United States and the United Kingdom. This pattern mirrors the author- and country-level results and underscores the role of core institutions in CSS research.

3.6 Strategic Thematic Analysis

As shown in Fig. 6, computational social science, big data, and machine learning emerge as motor themes. These themes are strongly interconnected and occupy a central position within the thematic network, suggesting that they constitute the main methodological and conceptual core of the literature. Several niche themes, including large language models, sentiment analysis, social computing, and digital methods, exhibit high internal coherence but relatively limited connections to other research themes, indicating their niche role in CSS research. Basic themes such as natural language processing, text analysis, network analysis, social media, Twitter, and misinformation display high centrality but low density, suggesting that they function as foundational elements across studies. Finally, themes such as artificial intelligence, data mining, and data-driven evaluation appear in the emerging or declining part, reflecting their comparatively weak integration and development within the analyzed literature.

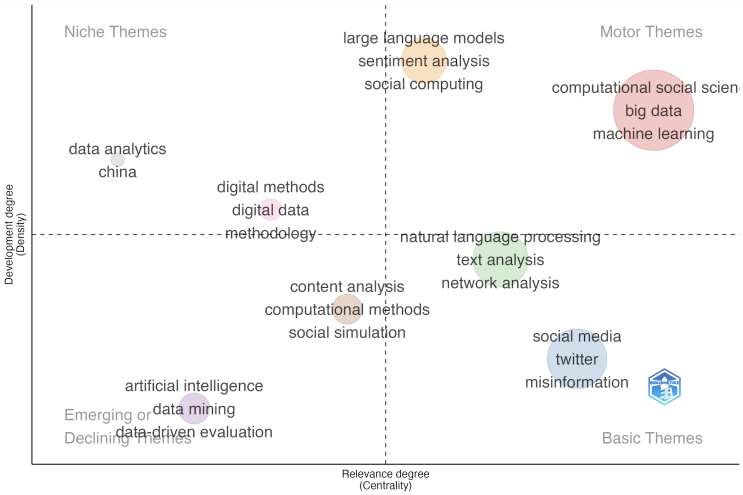

Fig. 6. Strategic thematic map of CSS.

4 Discussion

4.1 Interdisciplinary and Applied Nature of CSS: Tool Migration, Problem Formation, and Intellectual Convergence

Many evidence above could indicate that CSS is an interdisciplinary field, with intellectual foundations spanning social sciences, such as sociology, political science, communication, psychology; and natural sciences, such as computer science and network science. These contributions converge around a shared commitment to computationally enabled social inquiry. This convergence is visible in the co-citation structure: early foundational work foregrounded the promise and challenges of large-scale behavioral measurement using digital trace data, while later clusters became more centered on platform-mediated social processes such as information diffusion, polarization, misinformation, etc.

Importantly, interdisciplinarity in CSS is not only a matter of the involvement of multiple disciplines, but also a mechanism shaping how knowledge is produced. Methodological innovations, including scalable text analysis, network modeling, and transformer-based NLP, are rapidly translated into empirical studies of critical social issues. Meanwhile, recurring application domains, such as political communication and online behavior, exert a feedback effect by defining what kinds of data, validation practices, and interpretive standards become dominant. In this sense, CSS exhibits a dual movement: it uses computational methods from adjacent fields and reorients them toward social questions that require domain understanding, measurement validity, and interpretability, which could enrich multiple fields [28,29].

This also helps explain why certain themes occupy central positions in the thematic map. Topics such as social media, misinformation, machine learning, NLP, and LLMs function as connectors linking methodology to social inquiry,

while themes such as sentiment analysis and content analysis become motor themes because they are simultaneously method-rich and application-dense. The overall picture is therefore not simply one of topic diversity, but a pattern of intellectual convergence around reusable analytic pipelines, from data collection and representation to modeling and inference, that can be applied across multiple social domains.

4.2 Geographic Concentration and Its Implications for the Development of CSS

The results reveal a geographic concentration in the development of computational social science. At both the country and institutional levels, CSS research output is dominated by a relatively small number of countries, most notably the USA and the UK, and by a limited set of leading universities. This pattern suggests that the growth of the field has been driven primarily by research hubs.

This concentration can be partly explained by structural advantages in data access, computational infrastructure, and interdisciplinary training environments. Many of the leading institutions identified in this study are embedded in research ecosystems that combine strengths in social science, computer science, and large-scale data analytics, enabling sustained engagement with data-intensive research and methodological innovation. In this sense, geographic concentration reflects not only publication productivity but also the unequal distribution of the conditions required for computationally intensive research.

Such concentration has important implications for the intellectual orientation of CSS. When access to large-scale digital data is unevenly distributed across regions, the resulting research landscape may become skewed toward specific platforms, languages, populations, and regulatory contexts [30]. This may shape which social problems are most frequently studied, what types of data are regarded as valid evidence, and whose social realities are represented in the field. For example, the frequent use of social media data and English-language platforms may reflect both data availability and the geographic location of major research hubs.

Although the increasing contributions from countries such as China, Germany, and Italy indicate a gradual internationalization of CSS, geographic diversification of authorship does not necessarily translate into an equal diversification of empirical contexts or theoretical perspectives. From a developmental perspective, strong core institutions can accelerate methodological innovation and field consolidation, but a more diverse research landscape is needed to strengthen external validity and broaden the relevance of CSS. Encouraging cross-national collaboration and lowering barriers to data and computational resources may therefore play an key role in shaping the next phase of CSS research.

4.3 Developmental Stages of CSS and the Paradigm-Shift Claim

By synthesizing the descriptive trends, thematic mapping, and intellectual links, the evolution of CSS can be interpreted as a staged process.

The first phase (2009 – 2014), reflects the formative period of CSS. Foundational work such as Lazer et al. [2] and early work on digital trace data, social networks, and opinion dynamics established the conceptual basis of the field. The thematic analyses show that these topics were among the earliest to appear, yet they occupied relatively peripheral or emerging positions within the broader topic structure.

The second phase (2015 – 2019) is the rapid expansion of research on social media, misinformation, and computational modeling. During this period, CSS became more clearly defined through sustained empirical engagement with large-scale behavioral data from online platforms. The thematic map places these topics within the set of basic themes with high centrality, indicating their increasing importance across the field. This period also aligns with a diversification of disciplinary work, such as sociology, political science, psychology, etc.

The third phase (2020 – 2025), marks a transition toward an AI-driven and methodologically intensified era of CSS. The trend topics show a rise in machine learning, NLP, and LLMs. In the thematic map, these topics form central clusters alongside big data, data science, and AI, suggesting that computational techniques have become the primary engines driving new research directions. This phase also coincides with changes in the geography of research production: while the United States remains the dominant hub, China has emerged as an increasingly active contributor.

Such a staged view also implies that a paradigm shift is not guaranteed by technological adoption alone. The extent to which CSS constitutes a paradigm transformation depends on whether methodological advances are integrated with robust social-scientific standards: construct validity, transparent measurement, interpretability, theoretical integration, and replicability. The field's rapid methodological acceleration therefore makes reflection on research logics and evaluation standards more important.

4.4 From Digital Traces to AI-Driven Inquiry: Balancing Method-Driven and Problem-Driven CSS

The developmental trajectory identified in this study suggests that CSS is moving from the large-scale observation of digital traces toward a more active mode of AI-driven inquiry. In the early stage, digital trace data, social networks, and platform-based behavioral records provided the empirical foundation for CSS by making social behavior measurable at unprecedented scale and speed [31,32]. More recently, AI, NLP, machine learning, and LLMs have become increasingly central, enabling researchers not only to detect patterns in existing data, but also to model, simulate, and explain social processes. This shift changes the relationship among data, method, and theory: computational systems are no longer merely instruments for processing social data, but increasingly participate in the construction of social-scientific knowledge.

A central tension in this transition concerns the relationship between method-driven and problem-driven development. The method-driven trajectory is propelled by new models, architectures, and data pipelines that expand the capac-

ity for measurement, prediction, and simulation. Transformer-based approaches and LLMs exemplify this dynamic by enabling large-scale text analysis, multimodal inference, and generative modeling. Yet such advances may also risk prioritizing technical novelty over explanatory coherence, measurement validity, and substantive social relevance. By contrast, problem-driven inquiry begins from substantive questions such as misinformation, polarization, public opinion, behavioral dynamics, and platform governance, and treats computation as a set of instruments that must be justified by social-scientific inference goals. The paradigm-shift interpretation of CSS can therefore be strengthened by understanding the field as one that must balance methodological innovation with theoretically grounded and socially consequential inquiry.

Building on this synthesis, we highlight three agendas for the future development of CSS. First, AI-enabled measurement requires explicit validation protocols that go beyond predictive performance. As LLM-based and representation-learning methods become increasingly central, CSS needs transparent benchmarks, cross-domain generalization tests, uncertainty reporting, and construct-validity audits to align computational outputs with social-scientific concepts.

Second, governance, ethics, and reproducibility should become necessary components of AI-driven CSS. The growth of large-scale data collection and AI-based analysis amplifies concerns regarding data access, privacy, consent, replicability, bias, misinformation, and misuse [25,26,27]. Reproducible workflows, including data documentation, transparent modeling procedures, and independent audits where feasible, will be crucial for maintaining credibility and enabling cumulative research.

Third, CSS should strengthen interdisciplinary coherence and shared standards. Its hybrid nature is a major intellectual strength, but it also complicates the establishment of common evaluative criteria across computational and social-scientific traditions. Future research should more explicitly articulate how computational techniques, social theory, and domain knowledge interact in the research process, especially in validating computational measures, integrating explanation with prediction, and aligning model-driven analysis with substantive social questions.

4.5 Limitations

This study has several limitations. First, the dataset was constructed using exact keywords. While this strategy ensures that the included publications explicitly identify with CSS, it may exclude relevant studies that use computational methods to examine social phenomena but do not adopt the CSS label. Second, although the WoS database provides strong coverage of high-quality journal publications, it may not fully capture interdisciplinary work published in other databases, conference proceedings, books, or domain-specific outlets. Third, bibliometric techniques rely on metadata consistency and clustering algorithms, which inevitably simplify complex intellectual structures and may affect the interpretation of thematic boundaries.

5 Conclusion

This study examines how CSS has developed as a self-identified field and how it reflects a broader shift in social science research. Using journal articles explicitly labeled as CSS, we conducted a bibliometric analysis based on descriptive indicators, co-citation networks, three-field plots, and thematic mapping. The analysis identifies the core nodes, themes, topics, and methods through which CSS has evolved since 2009.

The findings reveal a pattern of rapid but uneven growth. CSS research output has expanded substantially, but it remains concentrated within a limited set of interdisciplinary journals, institutions, and countries. Intellectually, the field links computational methods with social-scientific questions concerning online behavior, social networks, political communication, misinformation, and public opinion. Its thematic evolution follows a staged trajectory: from an initial phase centered on digital trace data and networked social processes, to an expansion phase driven by platform-based social issues, and finally to a recent phase shaped by AI-enabled methods such as NLP, machine learning, and LLMs. In this sense, CSS is moving from using digital traces to observe social life toward using AI-driven systems to model, simulate, and explain social phenomena.

These patterns suggest that CSS represents more than a new research label. It reflects an ongoing shift toward data-intensive, model-based, and AI-enabled modes of social inquiry. In the AI era, this transformation creates new opportunities for linking empirical observation, computational modeling, and social explanation, but it also raises challenges of disciplinary integration, methodological validation, and ethical governance. The future development of CSS will therefore depend on how effectively the field balances method-driven innovation with problem-driven inquiry, and how it establishes shared standards for theory building, measurement validity, model evaluation, and interpretation.

By mapping the intellectual structure and developmental trajectory of CSS, this study provides a systematic reference for researchers seeking to situate their work within the evolving landscape of social computing. As computational techniques continue to advance, clarifying the epistemic foundations and research agendas of CSS will be essential for ensuring that its rapid innovation remains aligned with robust social-scientific understanding and societal relevance.

Acknowledgements This work was supported by the National Social Science Fund of China (23BXW056). Jianxun Chu is the corresponding author.

References

1. Kuhn, T.S.: The Structure of Scientific Revolutions, vol. 962. University of Chicago Press, Chicago (1997)
2. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al.: Computational social science. Science **323**(5915), 721–723 (2009). https://doi.org/10.1126/science.1167742

3. Luo, J.D., Liu, J., Yang, K., Fu, X.: Big data research guided by sociological theory: a triadic dialogue among big data analysis, theory, and predictive models. The Journal of Chinese Sociology **6**(1), 11 (2019)
4. Edelmann, A., Wolff, T., Montagne, D., Bail, C.A.: Computational social science and sociology. Annual Review of Sociology **46**, 61–81 (2020). https://doi.org/10.1146/annurev-soc-121919-054621
5. Theocharis, Y., Jungherr, A.: Computational social science and the study of political communication. Political Communication **38**(1–2), 1–22 (2021). https://doi.org/10.1080/10584609.2020.1833121
6. Cushman, F.: Computational social psychology. Annual Review of Psychology **75**, 625–652 (2024). https://doi.org/10.1146/annurev-psych-021323-040420
7. Hofman, J.M., et al.: Integrating explanation and prediction in computational social science. Nature **595**(7866), 181–188 (2021). https://doi.org/10.1038/s41586-021-03659-0
8. Singh, V.K., Singh, P., Karmakar, M., Leta, J., Mayr, P.: The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. Scientometrics **126**(6), 5113–5142 (2021)
9. Huang, Y.: Bibliometric analysis of GIS applications in heritage studies based on Web of Science from 1994 to 2023. Heritage Science **12**(1), 57 (2024). https://doi.org/10.1186/s40494-024-01163-y
10. Aria, M., Cuccurullo, C.: bibliometrix: an R-tool for comprehensive science mapping analysis. Journal of Informetrics **11**(4), 959–975 (2017). https://doi.org/10.1016/j.joi.2017.08.007
11. Chen, C.: How to Use CiteSpace (c) 2015–2020 (2020)
12. Van Eck, N.J., Waltman, L.: Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics **84**(2), 523–538 (2010). https://doi.org/10.1007/s11192-009-0146-3
13. Kemec, A., Altinay, A.T.: Sustainable energy research trend: a bibliometric analysis using VOSviewer, RStudio bibliometrix, and CiteSpace software tools. Sustainability **15**(4), 3618 (2023)
14. Xu, T., Wang, M.: Research progress of wearable devices in textile field: a bibliometric analysis using VOSviewer, RStudio bibliometrix and CiteSpace software tools. International Journal of Clothing Science and Technology **37**(1), 166–179 (2025)
15. Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Stanley, H.E., Preis, T.: Quantifying Wikipedia usage patterns before stock market moves. Scientific Reports **3**(1), 1801 (2013)
16. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018). https://doi.org/10.1126/science.aap9559
17. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D.: Fake news on Twitter during the 2016 U.S. presidential election. Science **363**(6425), 374–378 (2019). https://doi.org/10.1126/science.aau2706
18. Kosinski, M., Stillwell, D., Graepel, T.: 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 **110**(15), 5802–5805 (2013). https://doi.org/10.1073/pnas.1218772110
19. Youyou, W., Kosinski, M., Stillwell, D.: Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences of the United States of America **112**(4), 1036–1040 (2015). https://doi.org/10.1073/pnas.1418680112

20. Bail, C.A., et al.: Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences of the United States of America **115**(37), 9216–9221 (2018). https://doi.org/10.1073/pnas.1804840115
21. Vargo, C.J., Guo, L., Amazeen, M.A.: The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016. New Media & Society **20**(5), 2028–2049 (2018). https://doi.org/10.1177/1461444817712086
22. Farrell, J.: Corporate funding and ideological polarization about climate change. Proceedings of the National Academy of Sciences of the United States of America **113**(1), 92–97 (2016). https://doi.org/10.1073/pnas.1509433112
23. Stella, M., Ferrara, E., De Domenico, M.: Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences of the United States of America **115**(49), 12435–12440 (2018). https://doi.org/10.1073/pnas.1803470115
24. Grimmer, J., Roberts, M.E., Stewart, B.M.: Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press, Princeton (2022)
25. Bail, C.A.: Can generative AI improve social science? Proceedings of the National Academy of Sciences of the United States of America **121**(21), e2314021121 (2024). https://doi.org/10.1073/pnas.2314021121
26. Argyle, L.P., Busby, E.C., Fulda, N., Gubler, J.R., Rytting, C., Wingate, D.: Out of one, many: using language models to simulate human samples. Political Analysis **31**(3), 337–351 (2023). https://doi.org/10.1017/pan.2023.2
27. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623. ACM, Virtual Event Canada (2021). https://doi.org/10.1145/3442188.3445922
28. Mason, W., Vaughan, J.W., Wallach, H.: Computational social science and social computing. Machine Learning **95**(3), 257–260 (2014). https://doi.org/10.1007/s10994-013-5426-8
29. Wang, S., Corcoran, J.: Spatially integrated social sciences: opportunities for computational urban science. Computational Urban Science **5**(1), 11 (2025). https://doi.org/10.1007/s43762-025-00170-4
30. Freiling, I., Scheufele, D.A.: Reply to Gonzalez-Bailon and Lazer: industry control and conflicts of interest in social media research. Proceedings of the National Academy of Sciences of the United States of America **122**(47), e2522924122 (2025). https://doi.org/10.1073/pnas.2522924122
31. Evans, J.: Social computing unhinged. Journal of Social Computing **1**(1), 1–13 (2020). https://doi.org/10.23919/JSC.2020.0002
32. Chen, H., Yang, C., Zhang, X., Liu, Z., Sun, M., Jin, J.: From symbols to embeddings: a tale of two representations in computational social science. Journal of Social Computing **2**(2), 103–156 (2021). https://doi.org/10.23919/JSC.2021.0011