WILEY | Hindawi

Research Article
Anatomy of Complex System Research

Yang Wang [ID][1] **and Xinyue Wang**[2]

[1]*School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an 710049, China*
[2]*School of Humanities and Social Science, Xi'an Jiaotong University, Xi'an 710049, China*

Correspondence should be addressed to Yang Wang; yang.wang@xjtu.edu.cn

Complex systems play important roles in science and economy and have become one of the major intellectual and scientific challenges of the twenty-first century. However, the way complex system research connects to other academic fields is much less well understood, with only anecdotal evidence. In this work, we present an anatomy of complex system research by leveraging a large-scale dataset that contains more than 100 million digitalized publications. First, we find that complex system research shows a steady growth relative to the whole science in the last 60 years, with a sudden burst after 2000, which might be related to the development of computational technologies. Although early complex system study shows strong referencing behaviors to mathematics and physics, it couples significantly with computer science in the twenty-first century and affects engineering strongly. Moreover, we find empirical evidence that complex system research tends to have multidisciplinary nature, as it is often inspired by or affects a diverse set of disciplines. Finally, we find significant positive correlations between fields' reference broadness and future scientific impacts. Overall, our findings are consistent with several characteristics of complex system research, i.e., its multidisciplinary, quantitative, and computational nature, and may have broad policy implications for supporting and nurturing multidisciplinary research.

1. Introduction

"I think the next century will be the century of complexity" said Stephen Hawking, who is a famous theoretical physicist and cosmologist. Indeed, we are surrounded by systems that are unprecedentedly complex, containing many individual components together with interactions among them, possibly nonlinear ones [1]. Consider, for example, the survival of human society needs cooperation between billions of individuals or communication infrastructures that connect and integrate human efforts. Another example is science. As a complex system, science contains millions of digitalized papers, patents, grant proposals, white papers, together with citation or referencing relationships among them, catalyzing the rapid development of modern science and technology [2, 3]. The above examples vividly illustrate that today's complex system research has implications for our society, from human mobilities [4], to economics [5], world trades [6], and human interactions or collaborations [7, 8].

Particularly with the emergence of network science at the dawn of the twenty-first century, complex system study generates a substantial amount of original research [9]. Also, it is related to many revolutionary technologies of the twenty-first century, including Google, Facebook, and Twitter [10].

The applied values and substantial implications of complex system study raise very interesting and important questions: where does complex system research come from? How tightly connected are other academic areas and the cutting-edge complex system research? Anecdotal evidence indicates that complex system study may be rooted in chaos theory, and it is also believed to be tightly connected to self-organization, agent-based modeling, and artificial intelligence. Despite of that, we lack a systematic and empirical understanding on how complex system research affects or is affected by other academic fields. Thanks to the availability of large-scale bibliometric datasets and the development of science of science research, the above-mentioned questions

can be empirically answered for now. Science of science offers us unprecedented opportunities to explore and uncover the citation dynamics [11–13], knowledge space [14], collaboration networks [7, 8, 15], and their association with societal impact [16, 17]. Indeed, recent empirical work regarding knowledge flow and scientist mobility among academic fields have attracted much attention, including the field of artificial intelligence [18], subfields of physics [19–21], and science in general [22, 23], together with the diffusion model [24].

Here we leverage the Microsoft Academic Graph (MAG) dataset to study the reference and citation patterns between complex system research and other academic fields, exploring their relationships. Consistent with the anecdotal evidence, we find that complex system research is inspired by mathematics and physics before 2000. Moreover, due to the rapid development of the computational technologies and the availability of large-scale datasets, complex system study has become largely connected with computer science afterwards. Finally, we show empirically that complex system study is inspired by a diverse set of disciplines compared with many other fields such as nuclear physics, and this broadness in the reference list tends to be associated with field's future scientific impact. Together, our study provides among the first empirical evidences that complex system research has multidisciplinary and computational nature. With the very nature of complex system research that may transfer to future scientific impact, our findings may have broad policy implications.

2. Results

The MAG dataset is from Microsoft Academic Services [25, 26], which contains more than 100 million papers from 1800 to 2018 together with their citation relationships. Moreover, the MAG dataset uses the natural language processing to identify paper academic fields, resulting in a field tree with 19 coarse-grained (level 0) and more than 80,000 subfields (level 1 and level 2) spanning from biology to computer science. The citation relationships and paper field information, together with the sufficient time span, allow us to understand the evolution of scientific fields. Finally, as will be shown below, the MAG data is also useful to our study since it contains sufficient computer science papers, which are well related to complex systems (there are 70,335 complex system papers in the MAG dataset).

2.1. Knowledge Production and Citation Networks. Throughout the past 60 years, major technological changes and paradigm shifts such as the development of chaos theory and computer science result in a substantial knowledge production in the field of complex systems, including network science, data science, and computational social science [3, 9, 27]. To quantify the knowledge production, Figure 1(a) shows the yearly number of publications from 1960 to 2018. Consistent with prior studies [19, 23], we find that science grows at an exponential rate, doubling every 12.3 years (Figure 1(a)). Complex system research also shows an

exponential growth trend (Figure 1(a)). Comparing the knowledge production of complex systems and the whole science, we find that complex system research shows a steady growth relative to science. Specifically, the percentage of complex system paper grows from less than 0.01% to 0.06% from 1970 to the end of 2010 with a sudden burst between 2000 and 2010, representing the golden age of complex systems (Figure 1(b)). Hence, the field's exponential growth is not only driven by social needs, but also by paradigm changes, such as the development of computer science and artificial intelligence. The results also agree with the Stephen Hawking's famous quote mentioned in the beginning of the paper.

Following the substantial growth of the core complex system field, it is also interesting to study the growth of its related fields. The next question is as follows: How to detect fields that are related to complex systems? Sam Edwards remarked the definition of physics as "*what physicists do*" [28]. Following in his definition, we expand the field of complex systems to a broad perspective by including papers that cite or were cited by complex system papers. We find academic papers related to complex system research also showing a similar exponential growth pattern with a doubling time of 10.33 years (Figure 1(a)). One may argue that papers that cite or were cited by complex system papers may not belong to its related fields, since such citation relationships might come from noise or random connections in the whole citation network. To address this issue, we compare the citation patterns of all MAG papers to a null model in which each paper's references are assigned randomly given their publication time, regardless of a paper's journal or research field [16]. Specifically, in the randomized MAG citation network, we switch all citation links between papers but preserve the total number of references of each paper and the year of referencing and referenced papers [16]. After that, we define papers related to complex systems if its relationships (references or citations) to complex system field are significantly higher than expected by chance [19]. Followed by the whole procedure, we end up with more than 1.2 million papers that are related to complex systems, showing consistent results with Figure 1(a). The steady growth of complex systems and its related fields raises an interesting question: Where does complex system research locate in the scientific knowledge space?

To visualize the knowledge graph, we extracted citation relationships among different fields to construct a field citation network (here, we use the MAG level-2 fields since complex system research belongs to this level). In the network, each node represents an academic field, and links between different fields are referencing or citation behaviors. Additionally, we consider the number of citations/references between two fields as the link weight in the citation network. To eliminate noise and insignificant links, we applied a backbone extraction method to the whole field citation network [29]. While the citation network often shows strong community structure, with subfields belonging to the same ancestor cite each other much more frequently than those with different ancestor fields (Figures 1(c)–1(e)), many fields show strong
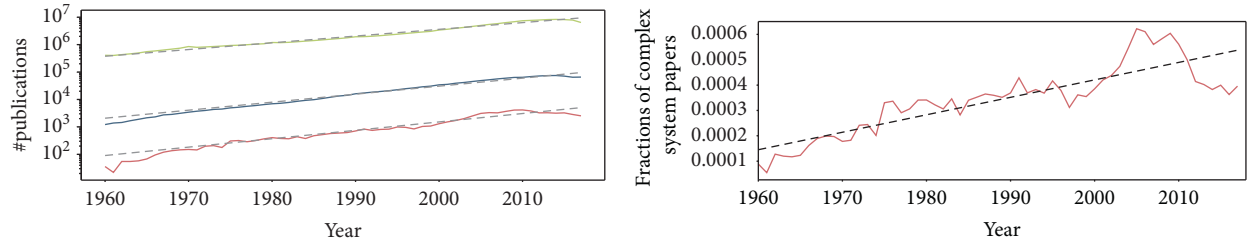
(a)

(b)

Complex systems

Complex systems
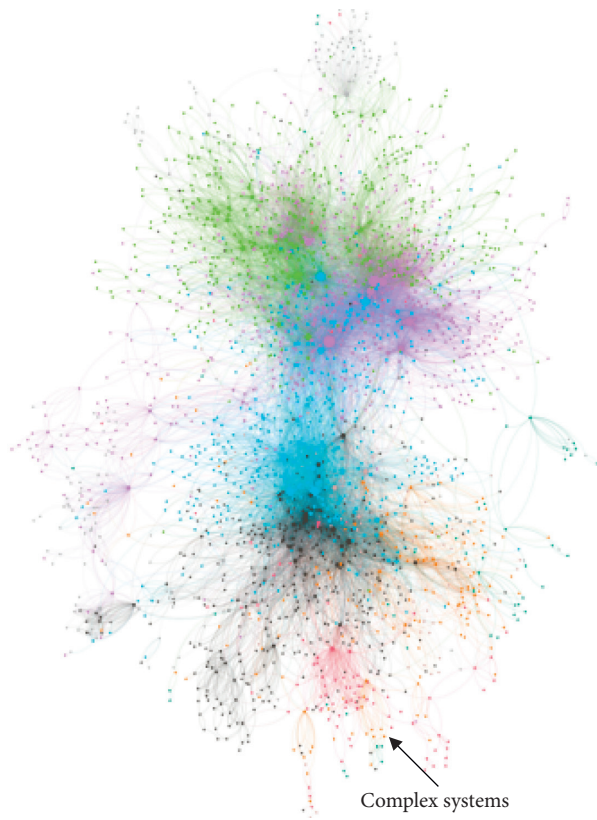
Biology
Medicine
Chemistry
Physics

Mathematics
Engineering
Computer science

Biology
Medicine
Chemistry
Physics

Mathematics
Engineering
Computer science
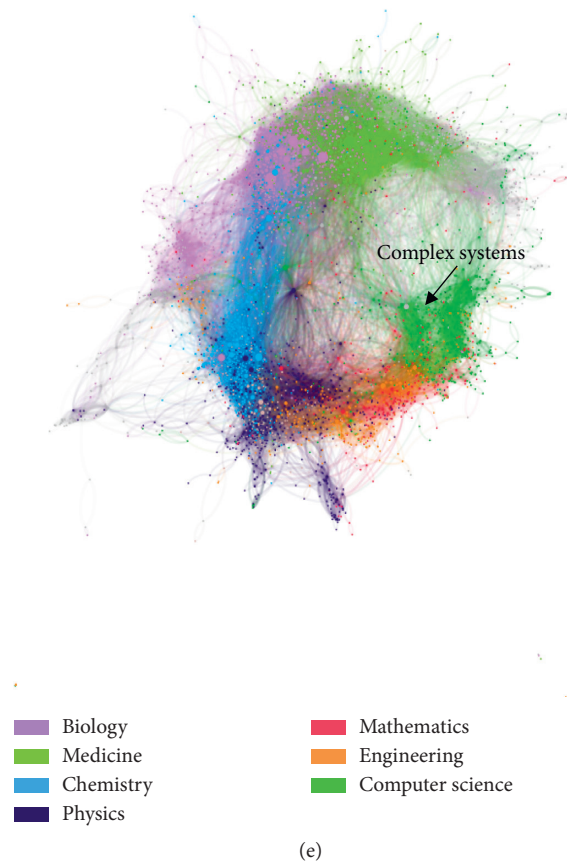
(c)

(d)

FIGURE 1: Continued.

(e)

FIGURE 1: The knowledge production and citation networks. (a) The yearly number of publications as a function of time from 1960 to 2018, for the whole Microsoft Academic Graph (green) and the field of complex systems (red) and its related fields (blue). We fit each growth pattern using an exponential function, shown in dashed lines. (b) The fraction of yearly number of papers of complex system relative to the whole MAG, together with a linear fit. Field citation networks in the (c) 1980s, (d) 1990s, and (e) 2000s, with nodes being MAG level-2 fields and links of the citation/reference relationships between two fields. In order to eliminate insignificant links, we applied a backbone extraction method. Different colors represent ancestor fields of these level-2 fields. We also annotate the field of complex system in the citation network.

interdisciplinary features. For example, chemistry lies between physics and biology, while engineering stands in the intersection of physics, mathematics, and computer science. Looking across different periods, we find complex system research moved from mathematics in the 1980s to computer science in the 2000s (Figures 1(c)–1(e)), indicating a substantial paradigm shift towards computational research [30]. Our findings are consistent with the fact that the development of computer science as well as the availability of large-scale datasets has catalyzed complex system research, possibly the emergence of network science as well. The knowledge production by the field of complex system, together with its position in the whole knowledge graph, prompts us to ask the following: Which specific field is closely related to complex systems and how to quantify the relationships between various fields rather than inspections?

2.2. Knowledge Flow and Its Related Fields. What are the origins of complex system research? Where does the knowledge created by complex system research go? The "linear model" of science demonstrates the importance of basic science to the development of applied research and technological development [31]. From the history of complex system field, its study of nonlinear dynamics or chaos originated from mathematics, atmospheric science, and physics deepens our understanding of its underlying mechanisms. Also, the development of complex system study contributes to the emergence of network science, data science, and computational social science. Therefore, we hope to understand the relevant knowledge flow. Does complex system theory emerge from mathematics or physics? Do computer science and engineering apply methodologies developed by the field of complex systems?

To investigate, we study the association between complex system research and various scientific fields through the relationship of references and citations. Specifically, we follow the definition of related papers to complex system research, and Figure 2(a) shows the distribution of their academic fields. We find that computer science and mathematics are the top two most frequent fields related to complex system research, followed by the fields of biology, chemistry, and physics. Other fields like philosophy or art are less likely to have connections with complex system study (Figure 2(a)).

To eliminate the effect of exponential growth of the paper production (Figure 1(a)), we calculate the share of references made from complex system study to other academic fields, as well as papers in other fields to complex system study. Following previous research [18, 19], the reference share from field A to field B is defined as

$$\text{share}_t (A, B) = \frac{\text{\#refs from A papers to B papers in year } t}{\text{\#refs made by A papers in year } t}.$$

(1)

The reference share measurement controls the number of references made by field A, thus controlling for its knowledge production (Figures 2(b) and 2(d)). Additionally, in order to control the paper production in field B that is the referenced field, we also compare the real number of references between fields A and B with the same quantity in the randomized citation network as defined in the previous section. Specifically, we focus on the z-score between the two fields as follows:

$$z_{A, B} = \frac{N_{\text{real}} - \mu_{A, B}}{\sigma_{A, B}},$$

(2)

where $\mu_{A, B}$ and $\sigma_{A, B}$ represent the average number and the standard deviation of connections between fields A and B generated from multiple randomizations, respectively (Figures 2(c) and 2(e)). A large z-score means fields A and B are more likely to be connected with respect to the null model.

Before 2000, complex system research frequently cites papers from mathematics and physics (Figure 2(b)); we find similar results after controlling for the paper production in referenced field (Figure 2(c)), suggesting early complex system study was shaped by mathematics and physics. The result is also consistent with our intuition that complex system study is closely related to the discovery of dynamical system theory or nonlinear system and nonequilibrium thermodynamics [32]. After 2000, however, complex system research strongly relies on computer science, indicating the paradigm shifts to computational research, such as data science and computational social science [33]. Interestingly, the references from complex system to chemistry and biology are largely driven by the dominance fraction of chemistry/biology papers with respect to the whole science. In fact, complex system research is less likely to cite biology and chemistry compared to the null model (Figure 2(c)).

How does complex system research affect other fields? We repeat the above analysis by calculating the reference share and z-score from other fields to complex system research. Several insightful patterns emerge. First, complex system research also affects mathematics and physics significantly before 2000, while computer science substantially cites complex system research after 2000 with steady growth pattern (Figures 2(d) and 2(e)). The results suggest that, in terms of knowledge flow, complex system research does not only act as a target but also an important input. Interestingly, engineering depends strongly on the development of complex system research after 2005 (Figure 2(e)), suggesting the applied value of complex system research.

The above analysis allows us to uncover the relationships between complex system study and other academic fields, prompting us to aggregate multiple years in order to investigate whether complex system study affects other fields more or vice versa. Figure 3 shows the z-score defined above for three different periods, i.e., 1980s, 1990s, and 2000s. By comparing z-score from complex system field to other fields (Figure 3, red bars) with the same quantity from other fields to complex system research (Figure 3, blue bars), we find the z-score from complex system to the field of mathematics is higher than the other way around, suggesting that complex system research consistently affects mathematics more from the year of 1980 (Figure 3(a)). Computer science, however, shows comparable knowledge flow from and to complex system research (Figure 3(b)), suggesting an extensive convolution structure between these two fields. Interestingly, engineering is more affected by complex system research since 2000, indicating the applied value of complex system research (Figure 3(e)). Overall, the framework of knowledge flow allows us to uncover the statistical regularities of interactions between various scientific fields.

2.3. Reference and Impact Broadness. Inspired by the results above, we discuss, in this section, the major characteristics of complex system research, and questions we would like to answer are as follows: Does complex system research make references from a diverse set of fields? What is the impact of complex system research?

In order to answer these questions, we quantify the broadness of disciplines reflected in the references and citations for every paper that belongs to the same academic field using the entropy measurement [34]. Specifically, given a paper with a reference or citation list, we embed each reference/citation to a vector of academic fields $d_\alpha = (d_{\alpha 1}, d_{\alpha 2}, \ldots, d_{\alpha k})$ based on the MAG level-1 fields. Then, the ith element of the paper is given by

$$a_i = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{k} \frac{\delta(d_{\alpha\beta}, i)}{k},$$

(3)

where δ is the delta function with 1 when $d_{\alpha\beta} = i$ and 0 otherwise, k is the number of MAG level-1 fields a reference/citation belongs to, n is the number of references for a given paper, and α and β are the index looping from the paper's reference list and field information of each reference, respectively. After embedding every paper to a vector, we perform L1 normalization to each vector, denoted by p. We then calculate the normalized entropy for each paper as
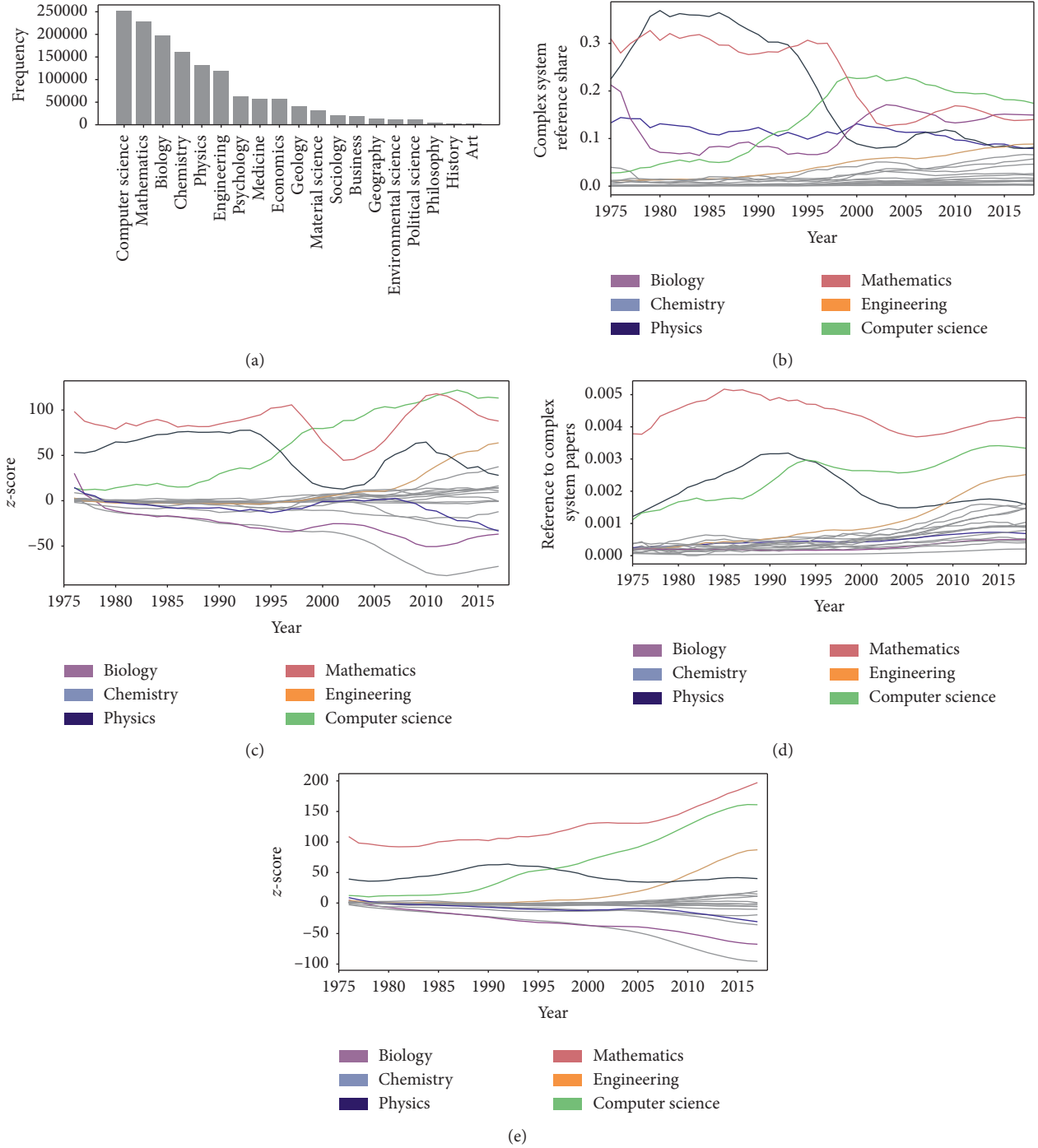
(a)

(b)

(c)

(d)

(e)

FIGURE 2: Knowledge between complex system study and other academic fields. (a) The distribution of academic fields for papers that cite or are cited by complex system research at least once. (b, c) The reference share and z-score (compared to the null model) from complex system research to other academic fields. (d, e) The same as (b) and (c) but from other academic fields to complex system research.

$$E(p) = -\frac{1}{\log_2 N} \sum_i^N p_i \log_2 p_i, \qquad (4)$$

where N is the total number of level-1 fields in the MAG dataset. A paper with $E = 1$ reflects its references/citations' equal distribution across all fields (broad referencing and impact); a paper with $E = 0$ indicates the

paper's references/citations are solely from one field (deeply disciplinary). Finally, we average the entropy over all papers within the same academic field to get the field reference or impact broadness.

First, Figures 4(a) and 4(b) show the distribution of field reference and impact broadness at 1995, documenting the narrow range of both referencing and impact broadness that are in contrast to the fat-tail citations distribution [35].
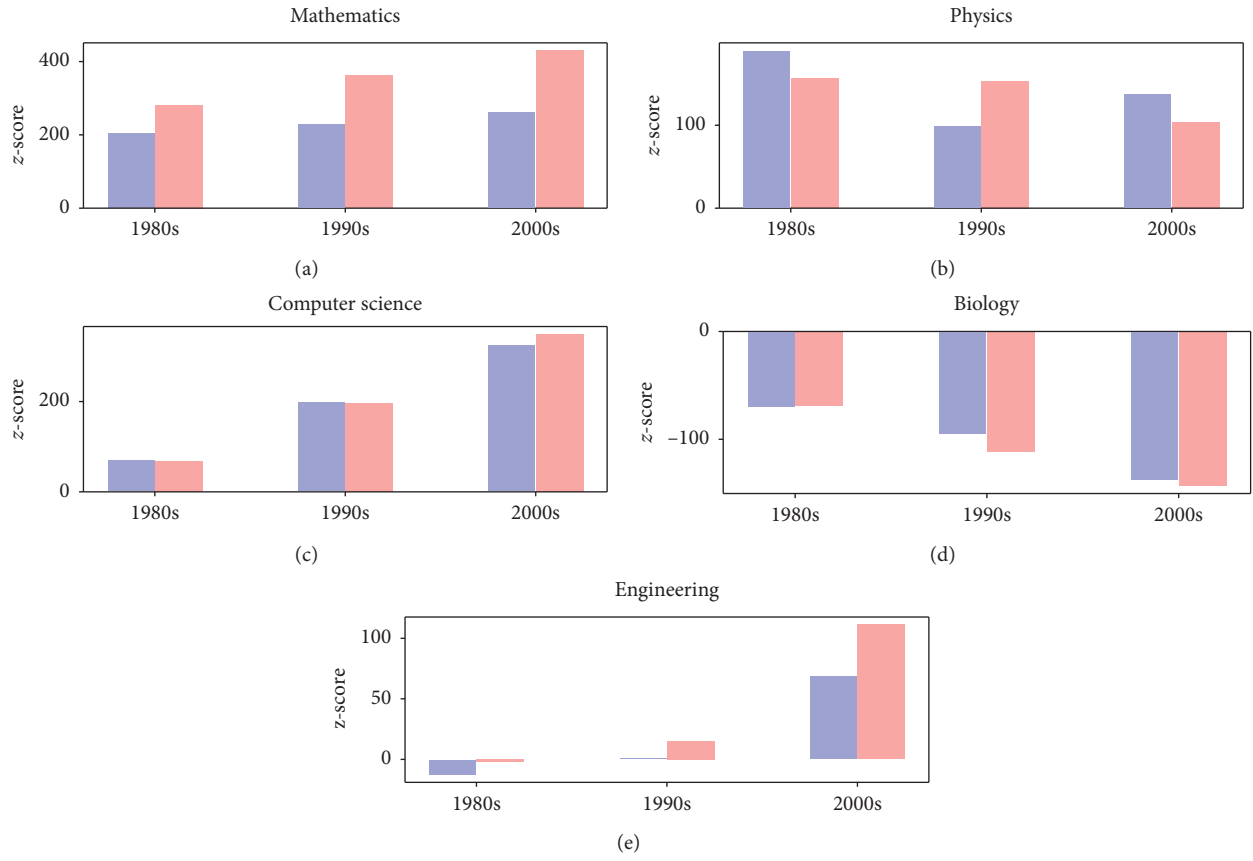
FIGURE 3: The aggregated z-score between complex system research and several selected academic fields. (a) Mathematics, (b) physics, (c) computer science, (d) biology, and (e) engineering. Specifically, blue bars mean how complex system research cites other fields, and red bars represent how other fields cite complex system research.

Interestingly, complex system research shows a relatively diverse feature among all academic fields, as shown in the dash line (Figures 4(a) and 4(b)). Moving beyond a single year, we repeat the same analysis for all academic fields from 1970 to 2018, finding similar results (Figures 4(c) and 4(d)). Comparing the reference or impact broadness of complex system research with the same quantities for nuclear physics, which is often considered as a deeply disciplinary field [19], we find that complex system research consistently shows higher value of reference or impact broadness (Figures 4(c) and 4(d)). Our results thus show a multidisciplinary nature of complex system research, suggesting that complex systems research may offer a language in which many academic fields can interact with each other without border.

2.4. Reference Broadness and Scientific Impact. The multidisciplinary nature of complex system research encourages us to explore the association between field's reference broadness and its scientific impact, which has policy implications for funding agencies. To answer this question, we calculate, for each field, its reference broadness and its impact, which is measured by the average paper citations within 8 years of publication for a specific field (C8). We find a significant positive correlation between field reference broadness and the scientific impact (Figures 5(a) and 5(b)), suggesting that fields with more diverse reference lists are more likely to have more scientific impact in the future. To test the robustness of this association, we repeat the analysis from 1970 to 2008, finding consistent and significant results (Figure 5(c)).

Finally, one might argue that different scientific fields often show different citation patterns. For example, biology and chemistry often have, on average, more citations than mathematics and engineering [35]. Moreover, citation behaviors change over time [11]. In order to eliminate this concern, we use a simple linear regression by focusing on the effect of reference broadness on field impact, while controlling for time, ancestor field categories, the number of publications within each field, and the average number of references. We find consistent positive correlation between the reference broadness and field's scientific impact (Figure 5). To further eliminate the effect of the number of references, we did the same regression analysis for subsamples with similar number of references, finding consistent results. The results can be interpreted as follows: one standard deviation (SD) increase of reference broadness for a scientific field is associated with on average 13.9% increase of future scientific impact.

(a)

(b)

- Complex systems
- Nuclear physics

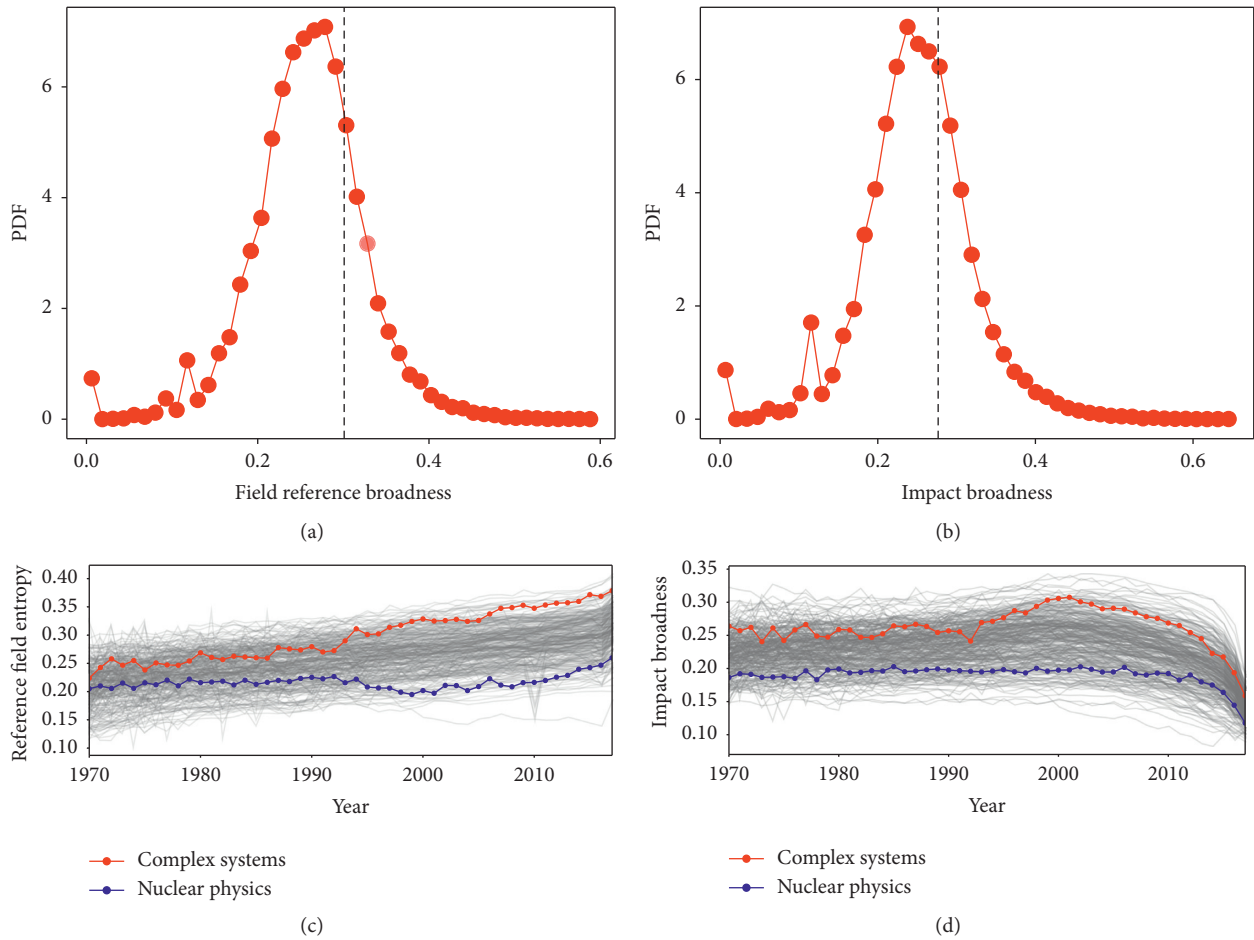(c)

- Complex systems
- Nuclear physics

(d)

FIGURE 4: Field referencing and citation broadness. The distribution of field reference (a) and impact broadness (b) at 1995, with the dashed line representing complex system research. (c) Field's reference broadness as a function of time, with red and blue lines showing the field of complex systems and nuclear physics, respectively. (d) The same as (c) but for impact broadness.
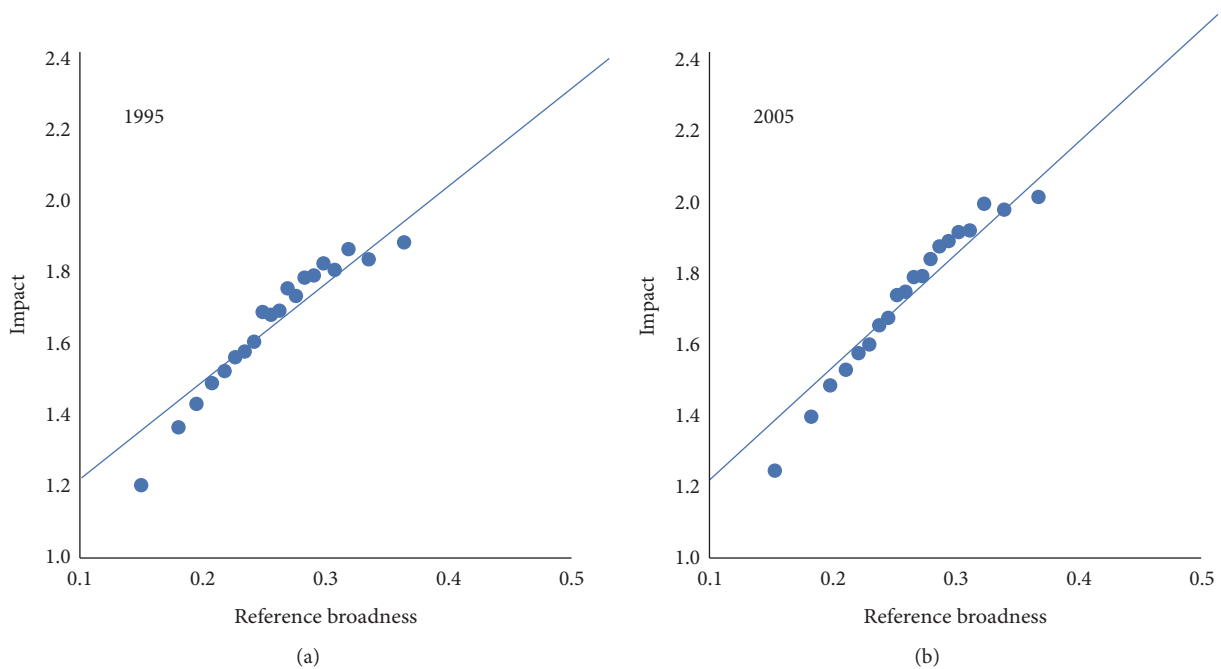

(a)

(b)

FIGURE 5: Continued.

(c)

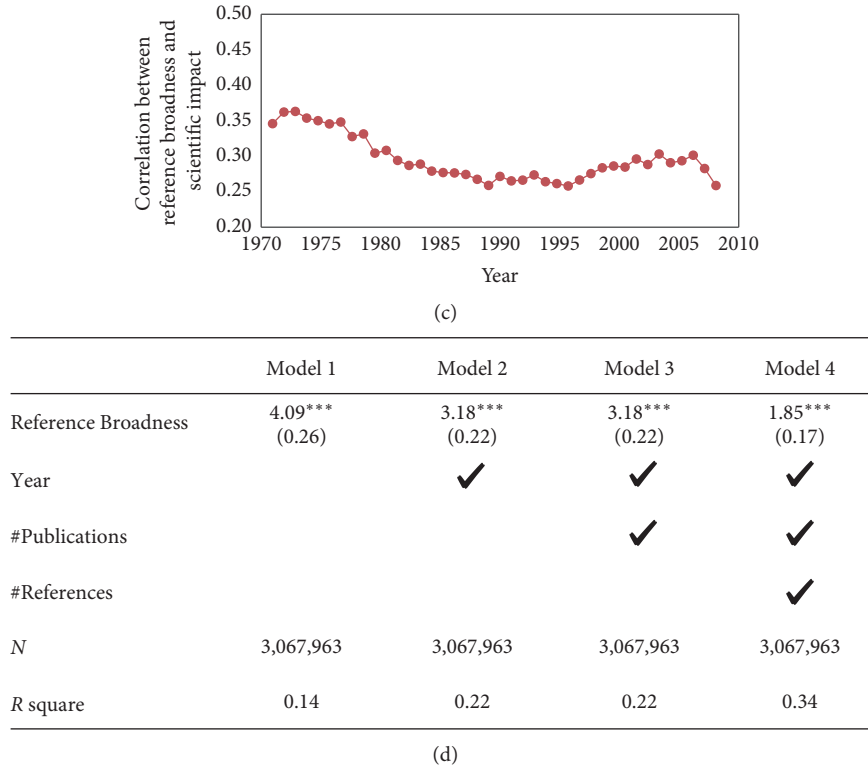|	Model 1	Model 2	Model 3	Model 4
Reference Broadness	4.09*** (0.26)	3.18*** (0.22)	3.18*** (0.22)	1.85*** (0.17)
Year		✔	✔	✔
#Publications			✔	✔
#References				✔
N	3,067,963	3,067,963	3,067,963	3,067,963
R square	0.14	0.22	0.22	0.34

(d)

FIGURE 5: The association between field's reference broadness and its scientific impact. The correlation between field reference broadness and its impact for 1995 (a) and 2005 (b). (c) The correlation as a function of time. (d) Linear regression results with dependent variable the field's scientific impact and the focal independent variable as field's reference broadness. Standard errors are clustered at the MAG level-0 field level. ***$p < 0.001$, **$p < 0.05$, and *$p < 0.1$.

3. Conclusions

Many of us involve with reviewers for funding agencies or publishers may confront with questions like: Can a particular work belong to the framework of complex system research? What is the nature of complex system study? In this paper, we present an anatomy of complex system research spanning from 1960 to 2018 by leveraging a large-scale bibliometric dataset. By investigating its knowledge production, connections with other academic fields, and its reference or impact broadness, our results show the very nature of complex system research, i.e., its multidisciplinary, computational or data driven, and applied nature. The empirical results are in line with many anecdotal evidences from textbooks or review articles [3, 9]. Our results also go beyond case studies, which only focus on referencing/citation behaviors within a specific field. By investigating the connections between fields in the whole knowledge space, we systematically quantify the evolution of the focal field with respect to science.

The significant positive association between fields' reference multidisciplinarity and their future scientific impact may have policy implications, from individual department to funding agencies and government. In general, our results are consistent with prior findings that isolations of certain scientific fields bring significant impact penalties [19]. In order to foster and nurture multidisciplinary science that is key to many real-world problems, funding agencies and government may support fields with multidisciplinary nature. From the perspective of individual scientist [20], our results provide actionable strategies of choosing future research areas [14, 36, 37] and mentors [38]. By choosing research areas that are able to link to their surrounding fields, individual scientists are able to show future scientific impact. Isolation to their environment, however, is destined to extinction [19].

Our understanding of the story that complex system research tends to be inspired by a diverse set of academic areas is not without limitation. There are still several remaining issues for future study. First, while this work considers large-scale citation relationships among scientific fields, there is a lack of other dimensions including scientist mobilities among various fields [20] or text similarities between different papers [39]. Such limitations offer a chance to go beyond our current understanding and ask further questions: how to integrate citations, individual scientists, as well as text feature among academic fields? Given that many important discoveries emerge from the intersections of various fields [16], the analysis of knowledge graph using large-scale dataset could help us uncover or predict emerging/promising research areas. Finally, with the intense connections between science and technology, future work is to investigate the association between fields' multidisciplinarity and their roles in advancing technologies.

Data Availability

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Y. Bar-Yam, "General features of complex systems," in *Encyclopedia of Life Support Systems (EOLSS)*, UNESCO, EOLSS Publishers, Oxford, UK, 2002.

[2] P. Azoulay, J. Graff-Zivin, B. Uzzi et al., "Toward a more scientific science," *Science*, vol. 361, no. 6408, pp. 1194–1197, 2018.

[3] A. Zeng, Z. Shen, J. Zhou et al., "The science of science: from the perspective of complex systems," *Physics Reports*, vol. 714-715, pp. 1–73, 2017.

[4] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[5] C. A. Hidalgo and R. Hausmann, "The building blocks of economic complexity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10570–10575, 2009.

[6] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, "The product space conditions the development of nations," *Science*, vol. 317, no. 5837, pp. 482–487, 2007.

[7] M. E. J. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[8] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proceedings of the National Academy of Sciences*, vol. 101, no. 1, pp. 5200–5205, 2004.

[9] A.-L. Barabási and M. Pósfai, *Network Science*, Cambridge University Press, Cambridge, UK, 2016.

[10] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.

[11] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, 2013.

[12] M. E. J. Newman, "The first-mover advantage in scientific publication,," *EPL (Europhysics Letters)*, vol. 86, no. 6, 2009.

[13] Q. Ke, E. Ferrara, F. Radicchi, and A. Flammini, "Defining and identifying sleeping beauties in science," *Proceedings of the National Academy of Sciences*, vol. 112, no. 24, pp. 7426–7431, 2015.

[14] A. Rzhetsky, J. G. Foster, I. T. Foster, and J. A. Evans, "Choosing experiments to accelerate collective discovery," *Proceedings of the National Academy of Sciences*, vol. 112, no. 47, pp. 14569–14574, 2015.

[15] A. M. Petersen, "Quantifying the impact of weak, strong, and super ties in scientific careers," *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. E4671–E4680, 2015.

[16] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, "Atypical combinations and scientific impact," *Science*, vol. 342, no. 6157, pp. 468–472, 2013.

[17] D. Li, P. Azoulay, and B. N. Sampat, "The applied value of public investments in biomedical research," *Science*, vol. 356, no. 6333, pp. 78–81, 2017.

[18] M. R. Frank, D. Wang, M. Cebrian, and I. Rahwan, "The evolution of citation graphs in artificial intelligence research," *Nature Machine Intelligence*, vol. 1, no. 2, pp. 79–85, 2019.

[19] R. Sinatra, P. Deville, M. Szell, D. Wang, and A.-L. Barabási, "A century of physics," *Nature Physics*, vol. 11, no. 10, pp. 791–796, 2015.

[20] F. Battiston, F. Musciotto, D. Wang, A.-L. Barabási, M. Szell, and R. Sinatra, "Taking census of physics," *Nature Reviews Physics*, vol. 1, no. 1, pp. 89–97, 2019.

[21] Y. Sun and V. Latora, "The evolution of knowledge within and across fields in modern physics," *Scientific Reports*, vol. 10, Article ID 12097, 2020.

[22] E. Yan, "Disciplinary knowledge production and diffusion in science," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, pp. 2223–2245, 2016.

[23] Y. Dong, H. Ma, Z. Shen, and K. Wang, "A century of science: globalization of scientific collaborations, citations, and innovations," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1437–1446, Halifax, Canada, August 2017.

[24] C. Jin, C. Song, J. Bjelland, G. Canright, and D. Wang, "Emergence of scaling in complex substitutive systems," *Nature Human Behaviour*, vol. 3, no. 8, pp. 837–846, 2019.

[25] A. Sinha, Z. Shen, Y. Song et al., "An overview of microsoft academic service (MAS) and applications," in *Proceedings of the 24th International Conference on World Wide Web*, pp. 243–246, Florence, Italy, May 2015.

[26] K. Wang, Z. Shen, C. Huang et al., "A review of Microsoft academic services for science of science studies," *Frontiers in Big Data*, vol. 2, 2019.

[27] S. Fortunato, C. T. Bergstrom, K. Börner et al., "Science of science," *Science*, vol. 359, no. 6397, Article ID eaao0185, 2018.

[28] A. Donald, "Let's not forget plants," *Physical Biology*, vol. 11, Article ID 053008, 2014.

[29] M. A. Serrano, M. Boguná, and A. Vespignani, "Extracting the multiscale backbone of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6483–6488, 2009.

[30] T. S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, IL, USA, 1963.

[31] M. Ahmadpoor and B. F. Jones, "The dual frontier: patented inventions and prior scientific advance," *Science*, vol. 357, no. 6351, pp. 583–587, 2017.

[32] B. Castellani, *Map of the Complexity Sciences*, Art and Science Factory, Cleveland, OH, USA, 2018.

[33] D. Lazer, A. Pentland, L. Adamic et al., "Social science: computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[34] A. J. Gates, Q. Ke, O. Varol, and A.-L. Barabási, "Nature's reach: narrow work has broad impact," *Nature*, vol. 575, no. 7781, pp. 32–34, 2019.

[35] F. Radicchi, S. Fortunato, and C. Castellano, "Universality of citation distributions: toward an objective measure of scientific impact," *Proceedings of the National Academy of Sciences*, vol. 105, no. 45, pp. 17268–17272, 2008.

[36] T. Jia, D. Wang, and B. K. Szymanski, "Quantifying patterns of research-interest evolution," *Nature Human Behaviour*, vol. 1, Article ID 0078, 2017.

[37] A. Zeng, Z. Shen, J. Zhou et al., "Increasing trend of scientists to switch between topics," *Nature Commununications*, vol. 10, pp. 1–11, Article ID 3439, 2019.

[38] Y. Ma, S. Mukherjee, and B. Uzzi, "Mentorship and protégé success in STEM fields," *Proceedings of the National Academy of Sciences*, vol. 117, no. 25, pp. 14077–14083, 2020.

[39] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, http://arxiv.org/abs/1301.3781.