DSI DATA SCIENCE AND INFORMETRICS
VOL.2, NO.3, AUG. 2022

The scientific applications of big data in science of science

Yunwei Chen*, Qiuyang Chen, Lingjing Cao

a. Scientometrics & Evaluation Research Center (SERC), Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu, China
b. Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

The value of big data in science of science for knowledge discovery is that it can reveal deeper information and knowledge, promote knowledge integration in the whole process of scientific research, guide interdisciplinary integration, and provide new ideas and new methods for knowledge discovery research. This paper discusses the value and role of big data in science of science in knowledge discovery from five aspects, including exploring the laws of scientific research, revealing scientific structure, analyzing scientific research activities, supporting technical recognition and prediction, and serving science and technology evaluation.

KEYWORDS

Big Data in Science of Science; Science of Science; Scientometrics; Scientific Structure; Knowledge Discovery

Nowadays, the scientific community is an extremely large and highly connected internal complex ecosystem. A large number of achievements, personnel, institutions, countries, fields and other information intertwined, attracting more and more scientists from computer science, social science into the field of science of science. They carry out knowledge discovery research by using big data in science of science. In recent years, with the rapid progress of digital technology and the accelerated evolution of digital industrialization, data becomes richer, diverse and easier to be analyzed, and big data as a factor of production has become increasingly distinctive (Chen et al., 2021). The digital degree of information about scientific research activities, scientific research behavior and scientific research achievements is constantly improving, and the accessibility of structured digital information with in-depth index is increasingly improved, so more and more data analysis tools can be used (Fortunato et al., 2018). It provides more dimensions for scientific research, makes it possible to reveal more potential information and release deep knowledge, and gives a broader perspective for understanding scientific laws. Science of science is a discipline that reveals the process of scientific discovery and the law of scientific activities from different spatial and time scales. It can analyze the scientific structure, predict the trend of science and technology, discover the frontier of science, and serve the formulation of science and technology policies, strategic planning and deployment, science and technology management and related social problems.

* Corresponding Author: chenyw@clas.ac.cn

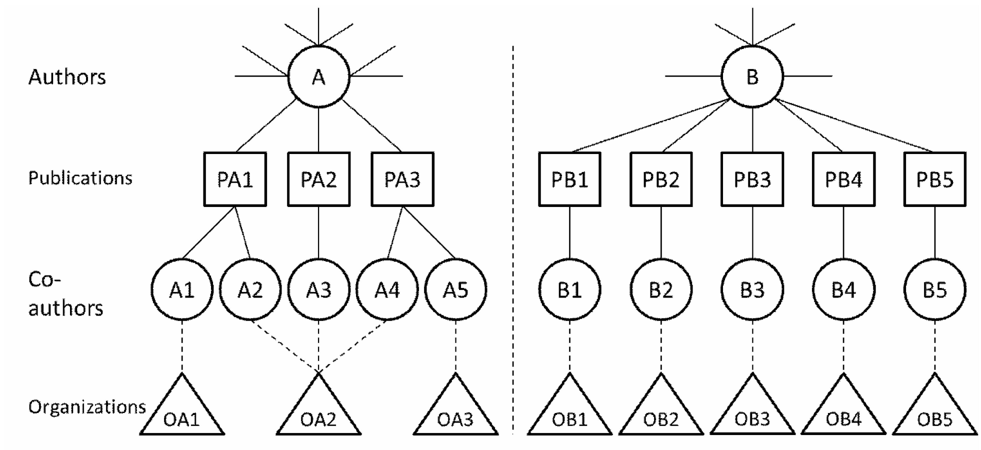
Big data in science of science is a general term for data used in science of science research. It is a data collection characterized by large capacity, multiple types, and high value, which includes scientific and technological literature data such as funds, papers and patents, network relationship data such as cooperation, citation and flow of scientific research subjects, as well as scholar background data, scientific research behavior information, scientific research tool software data, scientific research new media data, etc (Chen & Cao, 2020).

Science of science is able to reveal the interactions between scientific subjects of different spatial and temporal scales, mining the scientific structures and general laws of specific fields, with the goal being to promote scientific research (Nature Index, 2019). Based on this, knowledge discovery of big data in science of science is the process of identifying effective, novel and potentially comprehensible knowledge from massive data sets. So the value of big data in science of science research for knowledge discovery lies in revealing deeper information and knowledge based on multiple big data, promoting knowledge fusion in the whole process of scientific research, guiding interdisciplinary integration, and providing new ideas and methods for knowledge discovery research. Various kinds of scientific and technological literature databases and online analysis platforms, such as Web of Science (WoS), Scopus, Elsevier, Springer, Google Scholar, Nature Index (Nature Index, 2019), Incopat, Relecura, etc. And various pre-prints, research communities, research reports, research reports, funds and other databases, it is possible for researchers to have a deeper understanding of the scientific process.

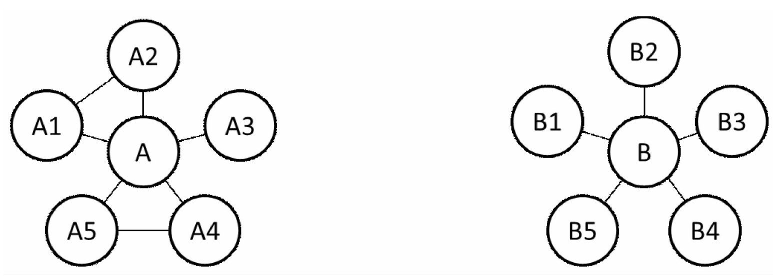
In general, the academic community has not systematically combed and understood the importance of big data in science of science in knowledge discovery research. Therefore, this paper discusses the main ways, values and challenges of big data analysis in supporting knowledge discovery research from the following five aspects, in order to promote the in-depth development of knowledge discovery research.

1 Explore the law of scientific research

There are more and more high-quality data supporting science of science research, and the relevant entity information, concepts and subject words constitute a network of tens of thousands of nodes, containing rich potential information and hidden knowledge. Based on these rich science data, it can reveal and describe the internal laws and characteristics of scientific research, such as reveal the birth and evolution of discipline direction, mining scientific research activities information and innovation law, etc. It also helps to obtain the overall understanding of the various factors affecting the scientific development from the perspective of the whole process of scientific research, so as to promote science and technology management and research and development organization behavior to a better direction. For example, Wolfgang Glänzel found in 2001 that the citation rate of international cooperation papers was higher than that of pure domestic papers (Glänzel, 2001). Many other studies have also reached similar conclusions (Wuchty et al., 2007; Panzarasa & Opsahl, 2014; Abramo et al., 2011; Kato & Ando, 2013), that is, there is a positive correlation between scientists' cooperative structure and their scientific research performance. This law has also become the basis for scientific research evaluation based on cooperation. For example, Yunwei Chen et al. (2015) constructed a composite cooperation intensity Compound Collaboration Strength (CCS) index based on cooperation characteristics to evaluate scientists' scientific research performance. Figure 1 shows the fundamental of CSS, which considers not only the impact of co-author feature, but also the impacts from the co-authors affiliations and their distributions among an author's publications when evaluating authors' performance based on collaborations. Schummer (2004) summarized the existence of symmetric and asymmetric cooperation modes in interdisciplinary studies. Zuo and Kang (2018) found that there was no clear positive correlation between the institutional subject diversity and the level of cooperation. Feng and Zhang (2020) showed that behavioral attitudes, perceived behavioral control, and subjective norms had a significant positive impact on the development of potential interdisciplinary cooperation. These findings suggest that we need to make reasonable boundary limits when using the subject field structure or cooperation structure of institutions to avoid absolute assumptions. Wu et al. (2019) have found that small teams have more potential to make disruptive innovations. This shows that scientific and technological innovation activities need multi-organizational mode, and we need to consider supporting various forms of scientific research organization mechanisms in the process of formulating policy tools and scientific research funding.



(a) Coupled co-author networks of A and B



(b) Ego co-author networks of A and B.

Figure 1 Comparison of the co-author networks of author A and author B (Chen et al., 2015).

Both of A and B have ten publications and five co-authors. They can win the same score by classical counting and co-author networks by counting their nodes degree centrality when only two ego co-author networks have been considered as Fig.1 (b). Furthermore, A's ego network has higher density than B. It seems like that A performs better than B. Indeed, A's five co-authors came from three different organizations and collaborated with A published three papers. While B's five co-authors came from five different organizations and collaborated with A published five papers. From that perspective, it seems like that B performs better than A.

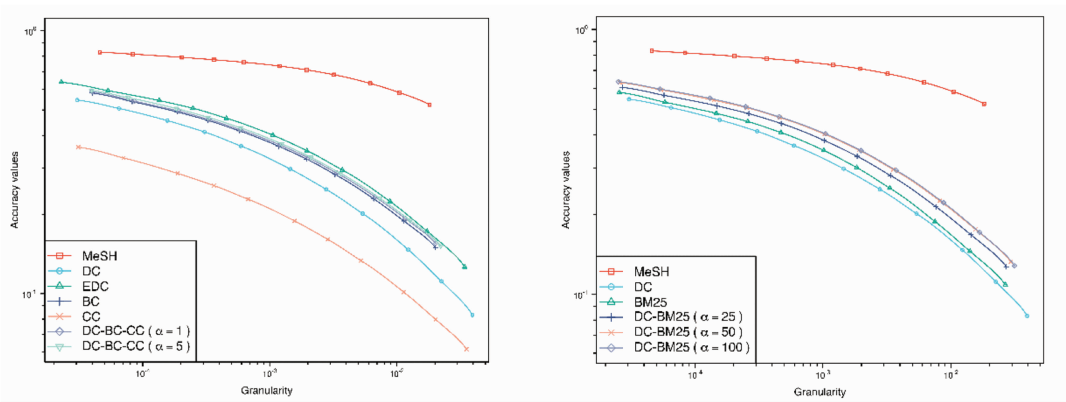
Generally speaking, scientific laws refer to universal laws and representations. However, in terms of scientific research, this general law or characteristics are usually limited to the common characteristics and laws within a certain region or field, and this law arises under the context of a specific culture and a specific field. Therefore, the universality of scientific research laws and characteristics revealed based on big data in science of science needs to be understood dialectically, interpreted under certain conditions, and tested accordingly when necessary. For example, Fangfang Wen found that there are differences in the cooperation modes between China and abroad in the field of information science. Chinese scholars are mainly related to teachers, students or colleagues, while foreign cooperation is mainly based on common research fields or interests (Wen, 2015). Therefore, when discussing domestic and foreign cooperation networks together, because the same data characteristics may imply different internal dynamics, it will be a very thorny practical challenge for information analysis and interpretation.

2 Reveal the scientific structure

In the field of scientometrics, the research on mathematical modeling of science and technology system has a long tradition, which aims to answer the basic mechanism behind the emerging structures such as scientific disciplines, scientific paradigms and interdisciplinary research frontiers. As early as 2005, Boyack et al. (2005) produced a scientific structure map of the whole field based on more than 1 million journal articles in more than 7,000 journals in the fields of natural sciences and social sciences, which can be used to detect the relationship between disciplines. Börner et al. (2011) pointed out that using advanced data mining and modeling methods to reveal scientific structure and perform static or dynamic visual presentation based on high-quality and high coverage data is a frontier research field in the field of scientometrics.

Especially in recent years, thanks to the development of network analysis methods, data mining algorithms, machine learning models and information visualization methods, as well as the emergence of various analysis tools such as CiteSpace, sci2, VosViewer and CitNetExplorer, researchers can analyze and identify the inherent objective scientific structural characteristics from the external characteristics based on the big data (Wei & Wei, 2011). For example, Janssens (1979) confirmed the effectiveness of revealing the scientific structure of the research field based on the data relationship between text and bibliometrics; Junping Qiu (Qiu & Dong, 2013) and Shiji Chen (Chen et al., 2015) found that citation network plays an important role in presenting scientific structure and discipline evolution process; Alexander et al. found that all disciplines has shown a phenomenon of cross growth since the beginning of the 20th century (Gates et al., 2019); We found that the author cooperative network community is an effective means to present the scientific structure in our previous research (Girvan & Newman, 2002; Chen et al., 2016). In terms of specific scientific structure research practices, Yue Ting revealed significant differences in the citation science structure between G7 countries and BRICS countries based on literature data (Yue, 2008), and pointed out that science and technology policies and strategies may have affected the evolution of scientific structure (Yue et al., 2018).

It should be pointed out that the current analysis of scientific structure is mostly based on cooperation or citation network. However, there is no systematic theoretical explanation or clear theoretical basis for this choice. Which type of network is more suitable to reveal the scientific structure still needs to be explored deeply. For example, Ahlgren compared the community division effects of direct citation, co-citation, and document coupling, and found that communities based on co-citation networks commonly used in the library and information science have a relatively weak degree of distinction between topics (Ahlgren et al., 2020), the result of which was showed in Fig.2. In addition, research on community division of multiple heterogeneous networks or hybrid networks has been proven to have better results (Zhang et al., 2019; Jiang & Chen, 2021). Therefore, it is necessary to continue to study the construction of the network in the research of scientific structure. Whether the network is scientific, reasonable and accurate is the most basic premise for effectively revealing the scientific structure.



(a) Comparison of DC, EDC, BC, CC and the two variants of DC-BC-CC.

(b) Comparison of DC, BM25 and the three variants of DC-BM25.

Figure 2 GA plot for comparing different approaches, MeSH used as the evaluation criterion

The GA plot of Figure 2(a) visualizes the accuracy results of enhancing DC by indirect citations. The performance of EDC, the combination of DC with BC and CC, as well as the performance of DC, BC and CC, is shown. CC exhibits the worst performance among the citation-based approaches. EDC has the best performance. In Figure 2(b), a GA plot that shows the results of enhancing DC by BM25, and thereby by textual relations. The plot also shows the performance of DC and BM25. BM25 performs better than DC but is outperformed by all three DC-BM25 variants.

3 Analyze scientific research activities

By analyzing the scientific research output data such as papers and patents, as well as the educational experience and career resume of researchers, we can not only find the regular knowledge of researchers' scientific behavior, employment and career choice, but also reveal the organization operation mode of scientific research teams or institutions, which is of great value for guiding the construction of scientific research teams, talent introduction and training. For the personal research activities of researchers, many scholars analyzed the regularity of the performance and achievements of scientists in their careers based on the output ability of scientific research results and the cited data of the papers. For example, Way et al. (2017) found that the most effective time of scientific research output is the time period within 8 years after scientists become chief researchers through scientific research output data in the field of computer science, and the peak year is usually the year before researchers get their first promotion. Others found that it is easier to produce scientists' personal best scientific achievements in the early to middle of their career (Jones & Weinberg, 2011). And researchers should not change their research direction frequently in the early stage of their career (Zeng et al., 2019). In addition, Sugimoto et al. (2017) found that the free flow of researchers is the most conducive to the maximum value of scientists and the healthy development of the scientific system. Other studies shows that there are obvious differences in the working habits of scientists in different countries (Wang et al., 2012). In terms of patent data analysis, Zhigang Feng found that the number of multinational patents around the world, and formed a transnational patent application network with the United States, the UK and Germany as the cores (Feng et al., 2020).

In terms of research on scientific research activities of scientific research institutions or teams, our team compared the characteristics of scientific research activities of four quantum scientific research teams in the world, mainly finding the differences of organization mode through the scientist cooperation network, discovering the team structure based on the discipline background and age information of talent team, and revealing the overall level of influence of scientific researchers through the citations of each scientist's papers, in order to fully show the competitiveness and scientific research level of scientific research teams (Zhang et al., 2018).

However, whether scientific researchers or scientific research institutions, scientific research output such as papers and patents are only derivatives of their scientific research activities. The information of scientific research activities based on scientific research output is not the whole picture of scientific research activities, so the use of these results should strictly control the boundary conditions.

4 Supporting Technology Identification and Prediction

The desire to predict new scientific developments, such as knowing in advance what a new discovery will be, by whom, when and where, exists today in almost every field of modern science. Scientists often predict interesting, influential, and fundable research directions or topics. Publishers and funding agencies evaluate submissions and project applications by predicting their future impact. Staffing review committees also predict which candidates will make significant scientific contributions over the course of their careers. The more predictable the scientific discovery process is, the more effectively human and financial resources can be used to support worthwhile research (Clauset et al., 2017).

In the field of technology forecasting, earlier research has been dominated by peer review and Delphi methods. However, researchers expect that using this big data in science of science information can produce more objective and accurate predictions than the qualitative predictions of experts in the library and information science field. Researchers continue to be enthusiastic about conducting technology prediction research based on quantitative data. There is a wealth of knowledge in scientific data and related information. Identification of technology frontiers, emerging technologies or disruptive technologies through analysis of existing papers, patents and other information can effectively support technology identification and prediction research. Prediction research based on the literature and its associated data can support at least four areas: prediction of citations of past discoveries, prediction of job bids, prediction of scientific productivity, and prediction of which candidates will make important scientific contributions over the course of ones' careers (Clauset et al., 2017). For example, Buchanan and Corken (2010) developed a science-intensive disruptive technology prediction model based on patent big data. Färber (2016) used semantic wiki for technology forecast and technology monitoring, demonstrating that semantic wikis can be useful in the field of technology forecast and technology monitoring, in addition to being an ideal knowledge management tool. Wang et al. (2021) proposed a concept of a neural network-based nonlinear citation-forecasting combined model for predicting the potential citation frequency of articles, and their results showed that the model has high accuracy and robustness in predicting the potential citation frequency of papers and the ranking of academic articles, which in turn supports the use in potential frontier technology identification studies. Huang et al. (2021) used online medical behavior data such as online medical consultation, online medical appointment, and online medical search to predict the trend of the number of cases using a multivariate vector autoregressive model to provide a reference basis for guiding epidemic prevention and control and resource deployment.

In addition, citation network main path identification studies are considered to be an effective tool for technology prediction studies. It can analyze the history of technology evolution along the main citation path and predict the next development trend. Our team proposed the concept of engine technology in 2017, and four hypotheses related to the structure of patentee cooperation networks to reflect the characteristics of engine technology. Based on this, the analysis of the patentee cooperation network structure on the basis of a large-scale patent dataset was carried out to realize the anticipatory study of engine technology (Chen et al., 2017). The result of this study shows that there are significant differences in the network structure of patentee cooperation between engine and non-engine technologies, and in turn, technology foresight studies can be achieved according to the different characteristics of this network structure.

However, the dramatic growth of big data in science of science is rapidly increasing the demand for analytical methods and tools. Traditional statistical and analytical metrics, algorithms and models are no more capable of meeting the new research needs, especially in technology identification and prediction research methods, for which there are no fully reliable and effective methods and tools. In order to solve this bottleneck, it is necessary to build a scientific and feasible theoretical framework system and methodological approach to propose effective algorithms or models on the one hand, and to develop open-source standardized tools and platforms for users to widely use and continuously upgrade the usefulness of the tools on the other hand. Machine learning and artificial intelligence methods have gained wider use in recent years. For example, Zhou et al. (2020) used data augmentation and deep learning techniques to forecast emerging technologies based on patent data, and the empirical prediction results of this study on Garter's emerging technology hype cycles showed that the method had an accuracy of 77%.

It is important to note that as the predictive efficacy based on big data in science of science increases, the more we need to be careful that such predictive studies based on historical data do not have a dampening effect on future scientific discoveries and the development potential of currently underestimated groups. On the one hand, it is difficult to predict completely new discoveries from historical data, and some potential principles or laws that are consistent with scientific laws may not be easily predicted accurately by the specific time of discovery due to technical constraints. On the other hand, the regular output patterns of individual scientists obtained statistically for group sample data cannot replace the judgment of individual research potential, and avoid denying the future development potential of individuals based on the overall historical statistics.

5 Services for Science and Technology Evaluation

In July 2021, the General Office of the State Council of China issued the "The General Office of the State Council on improving the Evaluation Mechanism of Scientific and Technological Achievements" (hereinafter referred to as "Opinions"). The opinion emphasizes the need to adhere to the correct orientation of science and technology results evaluation, innovative ways to evaluate scientific and technological achievements, improve and perfect the evaluation system of scientific and technological achievements. In order to better play the role of evaluation of scientific and technological achievements (Chen & Zhang, 2020). Therefore, the evaluation of scientific and technological achievements need produce high level scientific and technological outcomes through scientific evaluation programs. In this process, we should fully recognize the diversity of the value of scientific and technological achievements and avoid using single and one-sided indicators. For example, academic papers reflect a large number of theoretical achievements of basic research, and are one of the important carriers to evaluate the scientific value of scientific and technological achievements, but they can't effectively reflect other values. Many key core technology breakthroughs are often just the optimization of specific parameters or indicators, and it is difficult for non-professional to find the significant value behind them. If the achievements in the field of Humanities and social sciences are also measured by impact factors or patents, it is easy to erase their cultural value. Knaapen noted that the plurality of scientific assessments could move toward quantitative performance indicators and qualitative internal evaluations and beyond economic productivity to value the broader societal contributions of science. A more diverse external assessment may not be more accurate than quantitative indicators, but by assessing the infinite contribution of science and the diversity of its producers and users, it promises to be less one-sided and perhaps more impartial (Knaapen, 2021).

Big data in science of science provides a data basis for the evaluation of multiple values. The evaluation of multiple values of science and technology involves the selection of different methods and indicators. For example, Wang Yunhong et al. constructed the panoramic information data evaluation model of scientific and technological talents, in order to more comprehensively reflect the comprehensive strength of scientific and technological talents (Wang et al., 2017). Currently, the common practice that has won academic consensus is to adopt a combination of quantitative and qualitative approaches, where quantitative metrics are often used to support qualitative expert review (The San Francisco Declaration on Research Assessment, 2017; Hicks et al., 2015), and are subject to strict limitations (Moed, 2020; Adams et al., 2019). Article 8 of the "Opinions" explains the principles of indicator utilization. That is, we must resolutely crack the evaluation of scientific and technological achievements in the "only papers, only titles, only education, only awards, only hats" problem, a comprehensive correction of scientific and technological achievements in the evaluation of purely quantitative indicators, light quality contributions and other undesirable tendencies, to encourage science and technology workers to combine research with practical problems in order to promote the solution of major national problems. In the earlier documents of "Opinions" and the Ministry of Science and Technology of China, it should be noted that it is proposed to implement a representative work system for the evaluation of papers and to encourage the publication of "three types of high-quality papers". Therefore, how to improve the existing evaluation index scientifically and reasonably has become a very important concern in the academic circle, and the development of big data in science of science brings an opportunity for science and technology evaluation towards deepening. For example, Liu et al. (2021) used the idea of citation iteration and added the time dimension to construct a dynamic paper impact evaluation Q-index that simultaneously considers the number of citations, the quality of published journals and the impact of cited papers. And based on the Q-index scores of all the authors' papers, an F-index was constructed to reflect the influence of scientists, which can reveal the difference in the influence of papers based on the cited literature of multiple citation levels.

The previous section discusses the value of big data in science of science analyses for knowledge discovery and the ways or means to make a difference in five aspects. We can find that the increasing abundance of big data in science of science provides an unprecedented data base for in-depth knowledge discovery research. However, due to the diverse types and huge amount of Big Data in science, and the very specific and urgent needs for knowledge discovery, there is still a certain degree of gap between supply and demand. Theories, methods and tools for big data in science of science analytics to bridge this gap are still relatively lacking. Thus, big data in science of science analytics for knowledge discovery still faces great challenges and also presents significant opportunities. In the case of predictive research conducted with big data in science of science, for example, little is known about the understanding of the birth of scientific discoveries, despite a great deal of academic discussion. Precise predictions through scientific pathways are rare, and even the scientific community still lacks an accurate understanding of which scientific discoveries can be predicted and which cannot be effectively predicted (Clauset et al., 2017). As can be seen, the forecasting work presented in the previous section also hints at the limitations of data-driven forecasting regarding discovery generation. In terms of science and technology evaluation based on big data in science of science, science needs more diverse external evaluation. While some argue that evaluation and innovation are antagonistic (D'Agostino & Malpas, 2021), evaluation still needs to constantly seek to maximize benefits and minimize negative impacts in promoting innovation, and how to scientifically utilize the vast amount of big data in science of science is a major challenge and a major opportunity for science and technology evaluation. Most of the analyses about research laws, research activities, and scientific structure are based on retrospective generalizations and feature condensation of historical data. On the one hand, the accuracy of such regular features cannot yet be fully verified. On the other hand, when using these laws or features, policy makers and management departments should fully consider their statistical limitations to avoid over-reliance on them for resource allocation and other tasks, which may lead to negative impacts on the future layout of science and technology innovation and talent development.

Acknowledgements

This paper is supported by the National Social Science Foundation of China--Research on the Hybrid Network for Scientific Structure Analysis (Grant No. 19XTQ012)

References

Abramo, G., D´Angelo, C. A., & Solazzi, M.(2011). The relationship between scientists´ research performance and the degree of internationalization of their research. *Scientometrics, 86* (3), 629–643.

Adams, J., McVeigh, M., Pendlebury, D., & Szomszor, M. (2019). *Profiles, not metrics.* Recuperado de: https:// clarivate. com/g/profiles–not–metrics.

Ahlgren, P., Chen, Y., Colliander, C., & van Eck, N. J. (2020). Enhancing direct citations: A comparison of re–latedness measures for community detection in a large set of pubmed publications. *Quantitative Science Studies, 1* (2), 1–17.

Börner, K., Glänzel, W., Scharnhorst, A., & van den Besselaar, P. (2011). Modeling science: studying the struc–ture and dynamics of science. *Scientometrics, 89* (1), 347–348.

Boyack, K. W., Klavans, R., & B?rner, K. (2005). Mapping the backbone of science. *Scientometrics, 64* (3), 351–374.

Buchanan, B., & Corken, R.(2010). *A toolkit for the systematic analysis of patent data to assess a potentially disruptive technology*. Intellectual Property Office United Kingdom, London.

Chen, S., Arsenault, C., Gingras, Y., & Larivière, V. (2015). Exploring the interdisciplinary evolution of a disci–pline: the case of Biochemistry and Molecular Biology. *Scientometrics, 102* (2), 1307–1323.

Chen, S., Qiu, J., & Yu, B.(2021).China´s research contribution in big data. *Data Science and Informetrics, 1* (3), 1–13.

Chen, Y., & Cao, L. (2020). Opportunities and challenges of science of science big data. *Journal of Information Resources Management, 10* (3), 11–17+69.

Chen, Y., Deng, Y., Chen, F., Ding, C., Zheng, Y., & Fang, S. (2017). A technology foresight model used for foreseeing impelling technology. *Library and Information Service, 61* (13), 77–86.

Chen, Y., Deng, Y., Chen, F., et al. (2015). Research on construction and application of ccs index. *Library and Information Service, 59* (13), 96–103.

Chen, Y., Xu, Y., Wang, L., Xu, J., Fang, S. (2016). Network–based analysis on scientific paradigm of scientific research unit. Documentation, *Information & Knowledge, 2016* (5), 53–61.

Chen, Y., & Zhang, Z. (2020). Opinions on new science and technology evaluation methods. *Journal of the China Society for Scientific and Technical Information, 39* (8), 796–805.

Clauset, A., Larremore, D. B., & Sinatra, R. (2017). Data–driven predictions in the science of science. *Science, 355* (6324), 477–480.

D´Agostino, F., & Malpas, J. (2021). Evaluation and innovation: An antagonistic pair?. *Social Science Informa–tion, 60* (3), 345–349.

Färber, M. (2016). Using a semantic wiki for technology forecast and technology monitoring. *Program, 50* (2), 225–242.

Feng, Z., Zhang, Z., & Liu, Z. (2020). Network structure and evolution law of international transnational patent ownership. *Information Science, 38* (7), 153–161.

Feng, Z., & Zhang, Z. (2020). Analysis on the influencing factors of potential interdisciplinary cooperation behav–ior. *Information Studies: Theory & Application, 43* (02), 114–120+149.

Fortunato, S., Bergstrom, C. T., Börner, K., et al. (2018). Science of science. *Science, 359* (6379), eaao0185.

Gates, A. J., Ke, Q., Varol, O., & Barabási, A. L.(2019). Nature´s reach: Narrow work has broad impact. *Nature, 575* (7781), 32–34.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences, 99* (12), 7821–7826.

Glänzel, W. (2001). National characteristics in international scientific co–authorship relations. *Scientometrics, 51* (1), 69–115.

Hicks, D., Wouters, P., Waltman, L. et al. (2015). The Leiden Manifesto for research metrics. *Nature, 520*, 429–431.

Huang, W., Cao, B., Yang, G., Luo, N., & Chao, N. (2021). *Turn to the internet first? Using online medical be–*

havioral data to forecast COVID-19 epidemic trend. *Information Processing & Management, 58* (3), 102486. https://doi.org/10.1016/j.ipm.2020.102486

Janssens, F. (2007). *Clustering of scientific fields by integrating text mining and bibliometrics* [Doctoral dissertation, KU Leuven]. Libis.

Jiang, L., & Chen, Y. (2021). A review of community detection in hybrid networks with multiple nodes and multiple relationships. *Library and Information Service, 65* (19), 142-150.

Jones, B. F., & Weinberg, B. A. (2011). Age dynamics in scientific creativity. *Proceedings of the National Academy Of Sciences, 108* (47), 18910-18914.

Kato, M., & Ando, A. (2013). The relationship between research performance and international collaboration in chemistry. *Scientometrics, 97* (3), 535-553.

Knaapen, L. (2021). Science needs more external evaluation, not less. *Social Science Information, 60* (3), 338-344.

Liu, Q., Gulín, J., & Chen, Y. (2021). A citation iteration method for publications and scientists evaluation. *Data Science and Informetrics, 1* (2), 100-109.

Moed, H. F. (2020). Appropriate use of metrics in research assessment of autonomous academic institutions. *Scholarly Assessment Reports, 2* (1).

Nature Index. (2019). *Frequently asked questions.* https://www.natureindex.com/faq#introduction1.

Panzarasa, P., & Opsahl, T. (2014). *Patterns of scientific collaboration in Business and Management: The effects of network structure and interdisciplinarity on research performance.* http://www.ifr.ac.uk/netsci08/Download/CT_Abstracts/CT314.pdf

Qiu, J. A., & Dong, K. (2013). Methods and empirical research on deep integration of literature in citation network:case study on xml research literature from wos. *Journal of Library Science in China, 22* (1), 10-18.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59* (3), 425-465.

Sugimoto, C. R., Garcia, N. R., Murray, D. S., Yegros-Yegros, A., Costas, R., & Lariviere, V. (2017). *Scientists have most impact when they´re free to move. Nature, 550* (7674), 29-31.

The San Francisco Declaration on Research Assessment. (2017). *https://en.wikipedia.org/wiki/San_Francisco_Declaration_on_Research_Assessment.*

Wang, K., Shi, W., Bai, J., Zhao, X., & Zhang, L.(2021). Prediction and application of article potential citations based on nonlinear citation-forecasting combined model. *Scientometrics, 126* (8), 6533-6550.

Wang, X., Xu, S., Peng, L., et al. (2012). Exploring scientists´ working timetable: Do scientists often work overtime?. *Journal of Informetrics, 6* (4), 655-660.

Wang, Y. H., Pan, Y. T., & Zhao, X. Y. (2017). Evaluation on comprehensive scientific research capability of technology talents and application study based on big data of scientific and technological information. *Journal of Medical Informatics, 38* (12), 7-13.

Way, S. F., Morgan, A. C., Clauset, A., & Larremore, D. B. (2017). The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences, 114* (44), E9216-E9223.

Wei, J., & Wei, J. (2011). Review on analysis methods of science structure and evolution. *Library and Information, 2011* (4), 48-52.

Wen, F. F. (2015). Comparative study of domestic and international co-author patterns of information science based on social network analysis. *Information Science, 2015* (7), 128-133.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566* (7744), 378-382.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316* (5827), 1036-1039.

Yue, H. J. (2008). Prediction and relative dynamic performance on field of research of China. *Studies in Science of Science, 26* (3), 530-538.

Yue, T., Yang, L., Ahlgren, P., Ding, J., Shi, S., & Frietsch, R. (2018). A comparison of citation disciplinary structure in science between the G7 countries and the BRICS countries. *Journal of Data and Information Science, 3* (3), 14-30.

Zeng, A., Shen, Z., Zhou, J. et al. (2019). Increasing trend of scientists to switch between topics. *Nature Communications,10*, 3439.

Zhang, R., Chen, Y., & Deng, Y. (2019). A review of community discovery in hybrid network for science structure analysis. *Library and Information Service, 63 (*4), 135–141.

Zhang, Z., Chen, Y., Tao, C., et al. (2018). Bibliometric analysis on international competitive situation of quantum information research. *World Sci–Tech R & D, 40* (1), 37–49.

Zhou, Y., Dong, F., Liu, Y., Li, Z., Du, J., & Zhang, L. (2020). Forecasting emerging technologies using data augmentation and deep learning. *Scientometrics, 123* (1), 1–29.

Zuo, Z., & Zhao, K. (2018). The more multidisciplinary the better? – The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of Informetrics, 12* (3), 736–756.