Please cite as: Zhang, Y., Porter, A., Cunningham, S., Chiavetta, D. & Newman, N. 2018, How is data science involved in policy analysis? A bibliometric perspective, *2018 Portland International Conference on Management of Engineering and Technology*, Hwaii, the US.

How is Data Science Involved in Policy Analysis? *A Bibliometric Perspective*

Yi Zhang,[1] Alan L. Porter,[2, 3] Scott Cunningham,[4] Denise Chiavetta,[3] Nils Newman[3]

[1]Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia;
[2]Technology Policy and Assessment Centre, Georgia Institute of Technology, USA;
[3]Search Technology Inc., USA;
[4]Faculty of Technology, Policy and Management, Delft University of Technology, Netherlands;

***Abstract*: What are the implications of big data in terms of *big impacts*? Our research focuses on the question, "How are data analytics involved in policy analysis to create complementary values?" We address this from the perspective of bibliometrics. We initially investigate a set of articles published in *Nature* and *Science*, seeking cutting-edge knowledge to sharpen research hypotheses on what data science offers policy analysis. Based on a set of bibliometric models (e.g., topic analysis, scientific evolutionary pathways, and social network analysis), we follow up with studies addressing two aspects: (1) we examine the engagement of data science (including statistical, econometric, and computing approaches) in current policy analyses by analyzing articles published in top-level journals in the areas of political science and public administration; and (2) we examine the development of policy analysis-oriented data analytic models in top-level journals associated with computer science (including both artificial intelligence and information systems). Observations indicate that data science contribution to policy analysis is still an emerging area. Data scientists are moving further than policy analysts, due to technical difficulties in exploiting data analytic models. Integrating artificial intelligence with econometrics is identified as a particularly promising direction.**

I. Introduction

With the rapid development and broad application of Information Technologies (ITs), the interactions between data science and policy analysis have been widely observed (e.g., policy analysts integrate ITs with econometric models to handle real-world issues in Science, Technology, Innovation, and Policy (STIP) research [1, 2]), and data scientists or leading IT companies developed systematic toolkits to support decision making in a number of public sectors [3, 4]. The involvement of big data further increases this process, and terms such as *big data analytics* and *business intelligence* have become more and more popular in STIP studies [5, 6]. However, Athey [7] has raised the point that theoretical and practical innovation of ITs are still required to stimulate big data to achieve full potential in informing the STIP.

Our investigation focuses on the question, "How are data analytics involved in policy analysis to create complementary values?" and is divided into two aspects: (1) how has data science been engaged in current policy analysis; and (2) what is the current progress of the development of policy analysis-oriented data analytic models in IT areas? To our best knowledge, such research has never been conducted, but empirical insights explored by the investigation are expected to provide decision support for science policy, R&D plan, and entrepreneurship. In particular, we consider technology management as the private sector counterpart to policy analysis, and thus the involvement of data science in policy analysis (especially science policy) would provide a reference to the technology management community.

We address this from the perspective of bibliometrics. A set of articles published in *Nature* and *Science* was initially analyzed to seek cutting-edge knowledge to sharpen research hypotheses on what data science offers to policy analysis. Based on a set of bibliometric models (e.g., co-word analysis, bibliographic coupling analysis, citation analysis, and science mapping) [8, 9], we then collected observations from two aspects: (1) the engagement of data science (including statistical, econometric, and computing approaches) in current policy analyses by analyzing articles published in top-level journals in the areas of political science and public administration; and (2) the development of policy analysis-oriented data analytic models in top-level journals associated with computer science (including both artificial intelligence and information systems).

Observations indicated that the contribution of data science to policy analysis is still emerging. On the one hand, data scientists are moving further than policy analysts, due to technical difficulties for policy analysts in exploiting data analytic models. On the other hand, the development of existing decision support systems (in particular those policy analysis-oriented models) by data scientists is relatively blind, in which the connections between the real-world needs of policy analysis and the initial objectives of decision support systems are not as strong as what we imagine. As a particularly promising direction for next-step research, integrating artificial intelligence with econometrics is recommended.

The rest of this paper is organized as follows: Section II presents the data and pre-processing activities and Section III reports findings observed in the empirical study, with

discussion and implications. We finally conclude our research and address concerns for future studies in Section IV.

II. Data and Methodology

Three sets of articles were collected from the Web of Science (WoS)[1] on October 4, 2017. Search strategies, including search terms, target journals, and the number of articles, are given in Table I.

TABLE I. SEARCH STRATEGY

NO	#A[a]	Search Strategy
#1	72	TS[b]= (data AND policy) AND SO[c] = (*Nature* OR *Science*)
#2	1158	TS= (data SAME (big OR analy* OR science)) AND SO = (*American Journal of Political Science* OR *World Politics* OR *Journal of Public Administration Research And Theory* OR *Public Administration Review* OR *Review of International Political Economy* OR *Journal of Policy Analysis And Management* OR *International Organization* OR *Political Analysis* OR *American Political Science Review* OR *British Journal of Political Science* OR *Research Policy*)
#3	455	TS= (policy SAME analy*) AND SO = (*MIS Quarterly* OR *Journal of Information Technology* OR *Information Sciences* OR *IEEE Systems Journal* OR *Journal of Strategic Information Systems* OR *Business & Information Systems Engineering* OR *Information & Management* OR *Decision Support Systems* OR *European Journal of Information Systems* OR *Information Systems* OR *Journal of Management Information Systems* OR *Journal of The Association for Information Science and Technology* OR *ACM Transactions on Information Systems* OR *Journal of the Association for Information Systems* OR *Knowledge-based Systems* OR *Expert Systems with Applications* OR *International Journal of Intelligent Systems* OR *Communications of The ACM*)
Timeline		Between January 1, 2005 to 2017

[a.] Number of articles.

[b.] Topic, including title, abstract, and keywords.

[c.] Publication name.

Dataset 1 focuses on articles addressing both data and policy issues in the two world-leading journals *Nature* and *Science*. Dataset 2 involves the top ten journals (with the highest impact factor in 2016) in the WoS subject categories "political science" and "public administration." Since the research scope of the journal *Research Policy* is highly relevant to our study, we also added it. Dataset 3 initially selected the two WoS subject categories "computer science & information systems" and "computer science & artificial intelligence." However, considering the fact that some journals within these categories are purely technique-emphasized, we asked several IT researchers for help and, based on impact factor, we manually chose 19 journals to represent the target journals of policy analysis-oriented data analytics.

The general design of the search strategy is to seek cutting-edge knowledge to sharpen research hypotheses on what data science offers to policy analysis from Dataset 1. And as a result, we can explore empirical evidence from the two parallel datasets (e.g., Datasets 2 and 3). In particular, the selection of three datasets is based on the following reasons: (1) *Nature* and *Science* provide an interesting selection of potentially high-impact papers; they both treat science policy analyses and have covered some data analytics in STIP issues. It is admittedly a special sample, but one that can offer exploratory work that could connect data analytics and policy analysis. (2) Dataset 2 involves the leading journals in the area of policy analysis, and would be the best source to uncover research frontiers down the line. (3) Journals in Dataset 3 specifically emphasize applications of information systems and artificial intelligence techniques, and the use of data analytics in supporting policy making could be within this area.

VantagePoint[2] was applied to conduct data pre-processing: (1) a light function of name disambiguation was used to consolidate the variations of country/region names; and (2) a function of natural language processing was exploited to retrieve terms from combined titles and abstract fields. A term clumping process [10] was used to remove noise and consolidate technological synonyms.

The main methods exploited in this study include (1) co-word analysis: the relationships among terms were calculated based on co-occurrence statistics (i.e., the more frequently two terms appear together, the closer their relationship is [11]); (2) bibliographic coupling analysis, in which the affiliation information of authors was used to investigate the collaboration among countries and among institutions [12]; (3) citation analysis: our study mostly emphasized the use of the citation linkages between journals (i.e., the more citation linkages two journals had, the closer their relationship [13]; and (4) science mapping: we visualized the above relationships identified in our studies via a manner of science maps, in which the software VoSviewer was used [14].

III. Empirical Study

A. *Something Cutting-Edge: Interactions Observed from* Nature *and* Science

Focusing on Dataset 1, we manually read the 72 articles published in *Nature* and *Science*, and the results precisely matched our search purpose. For example, [7] discussed the gap between machine learning-based prediction and real-world issue-oriented decision making, and [15] investigated the behavior impact of China's one-child policy via econometric approaches. Under this circumstance, we aimed to discover: (1) research disciplines and core topics in this cutting-edge area; (2) related journals cited by the two world-leading journals, which would indicate potential theoretical/technical support for the area; and (3) countries that were spearheading this area. Note that in this study we defined "data science" in a relatively narrow scope (e.g., data analytics with IT support). Even though mathematic, statistical, and econometric approaches could be considered as a part of data science with relatively broad definitions, our study would take those traditional approaches apart from data science.

The term correlation map with top terms retrieved from the 72 *Nature* and *Science* articles is given in Figure 1. Despite clear clues, data science-related terms (the left part) and terms

[1] https://webofknowledge.com/

[2] VantagePoint is commercial software used in text mining and particularly in science, technology, and innovation text analysis. More details can be found on their website: https://www.thevantagepoint.com/

related to economic models (the right part) were relatively isolated, but policy-related terms (the middle part) likely acted as a bridge to link the two parts.

The observation that the interactions between data science and policy analysis are just emerging becomes obvious in the journal correlation map with journals cited by *Nature* and *Science*, given in Figure 2. Certain detailed findings are listed as follows:

1) So far, there are not too many computer science-related journals, but *Communications of the ACM* can be a good bridge.

2) Medical and bioscience journals have already been deeply interacted, resulting from the rapid growth of bioinformatics.

3) Business & management journals (e.g., *Management Science*) and statistical & economic journals (e.g., *American Economic Review*) can be observed here and there, which would be the current mainstreaming analytic instrument for policy analysis.

4) Multidisciplinary journals are playing increasingly crucial roles in this area. Besides *Nature* and *Science* themselves, *Proceedings of the National Academy of Sciences* also has a significant impact.

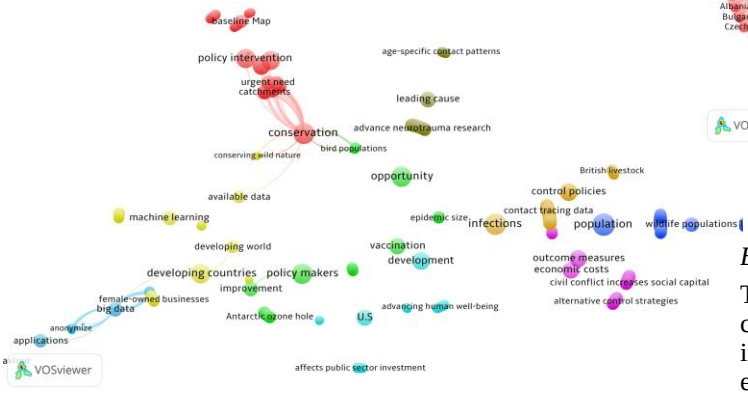

Fig. 1. Term correlation map in *Nature* and *Science*.

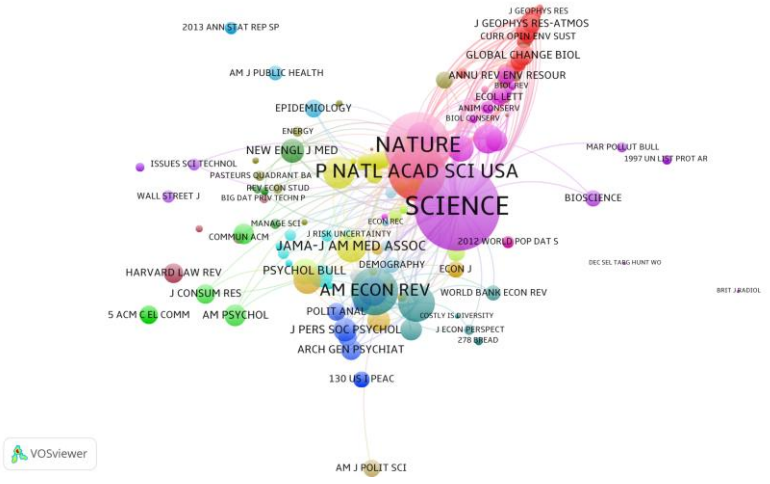

Fig. 2. Journal correlation map with journals cited by *Nature* and *Science*.

The country correlation map with countries retrieved from the 72 articles is given in Figure 3. Observations indicate that developed countries (especially the US, European, and Australasia countries) dominant this cutting-edge area, but Asian countries (e.g., those productive countries in academic research, such as China, Japan, and South Korea) are surprisingly not on the board. However, Germany, Sweden, Switzerland, and France seem like the gate to link Europe with the world.

In general, the contribution of data science to policy analysis is still emerging — even the interactions between the two disciplines are just beginning to develop. Given the circumstances, our empirical study was conducted from two parallel aspects: (1) to explore the engagement of data science in current policy analyses, and (2) to investigate the development of policy analysis-oriented data analytic models in computer science areas.

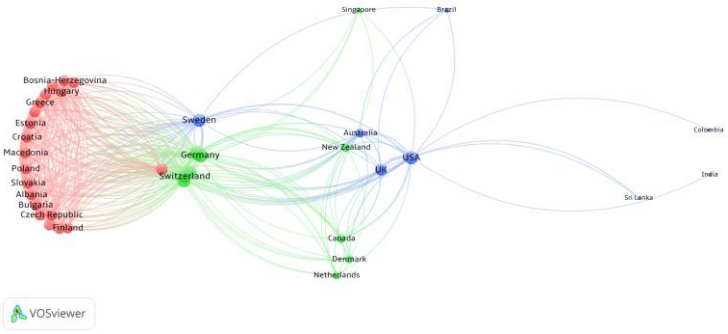

Fig. 3. Country correlation map with authors in *Nature* and *Science*.

B. *Policy Analysis with Data Analytics*

This aspect aims to explore the engagement of data science in current policy analyses, in which Dataset 2 was used. Some interesting articles can also be retrieved from this set. For example, [16] exploited the network analysis to explore insights in international relations, and [17] analyzed the determinants of sensitive nuclear assistance in the United States via data analytics. We followed the main methods used in Section A to explore insights from terms/topics, cited journals, and countries.

The term correlation map with terms retrieved from the 1,158 articles published in the 11 top-level policy analytic journals is shown in Figure 4. It is clear to observe that:

1) Terms related to R&D, strategic management, and innovation management (Red Nodes) occupy a significant proportion, and those topics are certainly emergent issues in STIP research;

2) Policy analysis-related terms (Brown Nodes) that focus on certain specific real-world issues could be another crucial component;

3) Statistical and econometric models (Purple Nodes) are the mainstreaming instrument, which support the findings observed in Section A.

4) The potential of IT techniques has already been gradually raised up—in particular its power in clustering and prediction. Some interesting terms include Monte Carlo simulation, network data, and time series analysis.

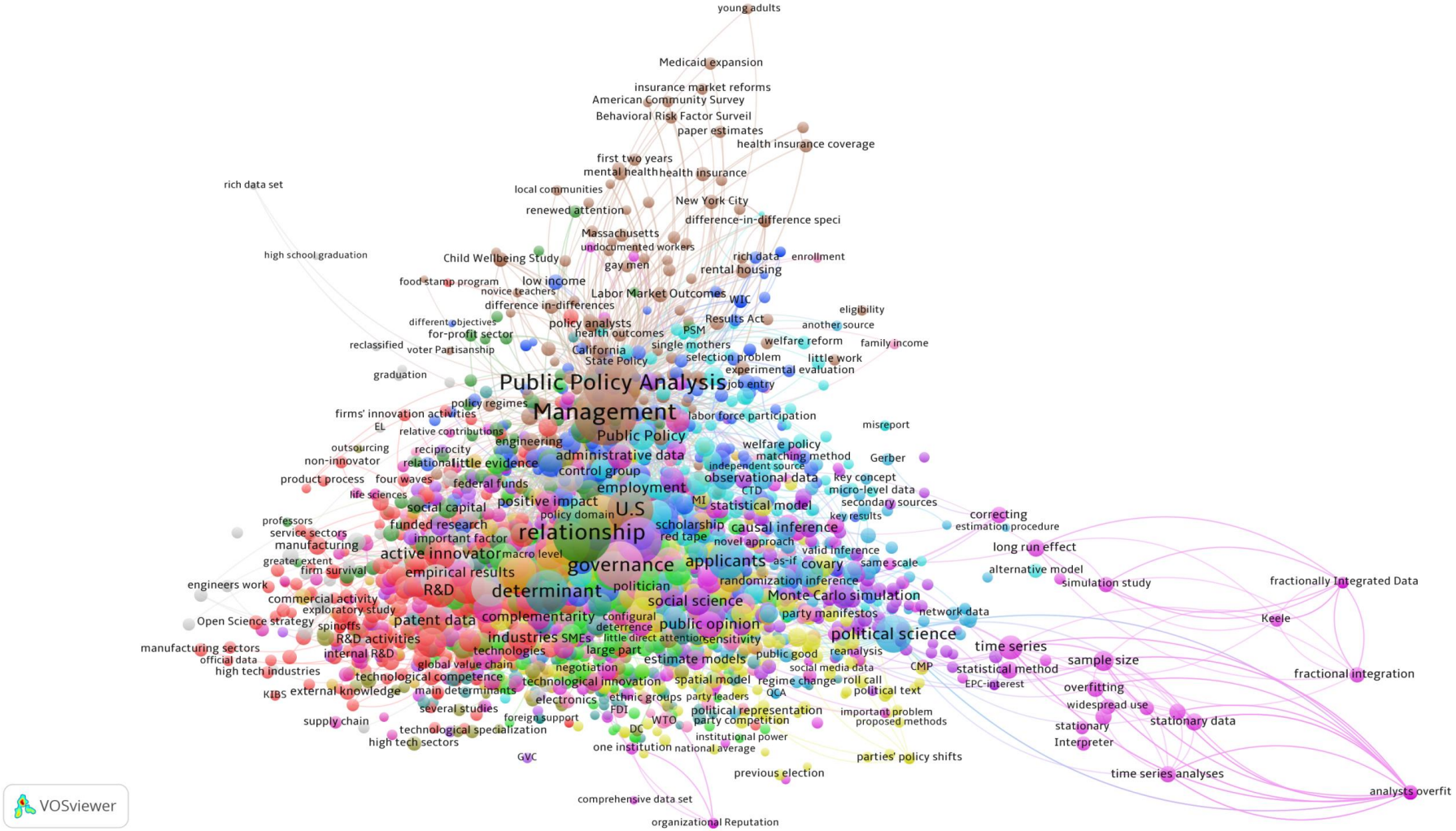

Fig. 4. Term correlation map with top journals in policy analysis.

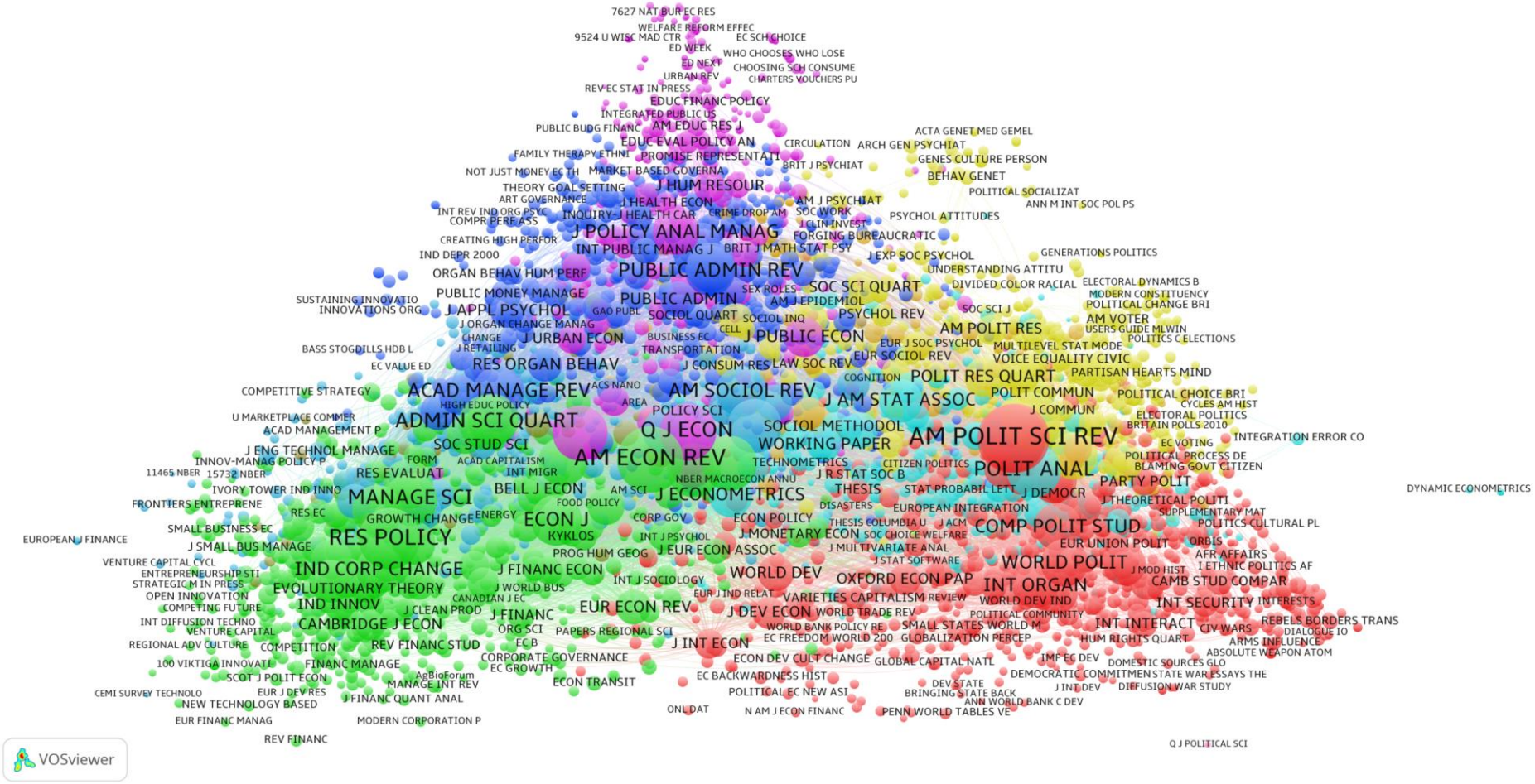

Fig. 5. Journal correlation map with journals cited by top journals in policy analysis.

The journal correlation map with journals cited by the top-level policy analytic journals (given in Figure 5) provides a bird's-eye view to understand the landscape of the current policy analysis. Unsurprisingly, economic-related journals dominate the majority of the list, and those specific journals could be divided into several sub-areas, such as management science (Green Nodes), political science (Red and Yellow Nodes), and public policy (Purple and Blue Nodes). However, it is rare to locate any data science-related journals, indicating that the community of policy analysis seems to be confident on existing approaches (e.g., econometrics). Also, the attempts to involve IT-based data analytic models have not created recognizable effects of scale.

The country correlation map with countries retrieved from Dataset 2 is given in Figure 6. It is obvious that the US and UK dominate policy analysis. Comparably, the UK mostly interacts with European countries, while the US broadly collaborates with Asian, African, and South American countries. In addition, compared to Figure 3, the gap between developed countries and developing countries on policy analysis is relatively small.

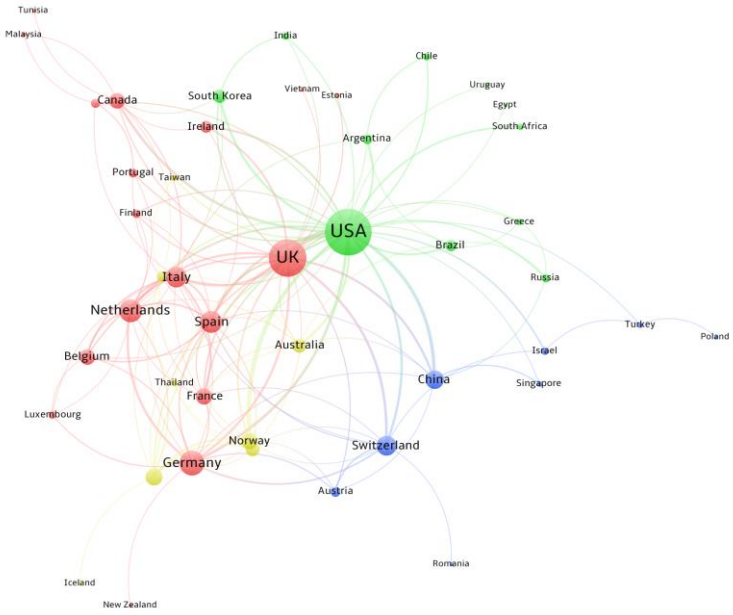

Fig. 6. Country correlation map with authors in top-level policy analysis journals.

Despite the largest dataset among the three candidates, the engagement of data science in the current policy analysis is not as interactive as what we imagined. Traditional statistical and econometric models are still the main instrument for policy analysts. One hidden reason might be the technical difficulties of exploiting, modifying, and developing data analytic models, which could be an unbridgeable gap for policy analysts so far.

C. Data Analytics for Policy Analysis

This aspect aims to explore the development of policy analysis-oriented data analytic models in top-level journals associated with computer science. The 455 articles in Dataset 3 also included interesting applications of data analytics in real-world policy issues. For example, [18] involved social network analysis in understanding terrorist behaviors, and [19] evaluated the immigration and border security of the United States via sentiment analysis and network analysis. We also followed the main methods used in Section A to explore insights from terms/topics, cited journals, and countries.

The term correlation map with the 19 selected top-level journals in data analytics is given in Figure 7. It is interesting to note that except for those mainstream information technologies (Red Nodes) such as classification, fuzzy sets, and controlling, IT researchers have presented their strong interest on real-world issues, such as decision support, governance, and climate change. Note that decision support systems (including recommender systems) could be one important sub-branch of information systems, which emphasize the application of information systems for supporting decision-making in real-world issues. Articles falling in this sub-branch, on the one hand, would use some terms such as decision-making and decision support, but would likely have different meanings with those used in policy analysis-related articles. On the other hand, real-world issues oriented by those decision support systems do cohere with STIP problems, and in those cases, the interactions between data science and policy analysis could be traced.

The journal correlation map with journals cited by the top-level journals in data analytics is given in Figure 8. Surprisingly, compared to either Figure 2 or Figure 5, the interactions between multiple disciplines are well revealed.

1) IT and engineering journals (Red Nodes): those journals which are mostly published by IEEE illustrate resources from which those policy analysis-oriented data analytic models gain technical supports.

2) Information system journals (Dark Blue Nodes): even though these journals would also belong to IT disciplines, those journals raise a trend focusing on real-world issues to develop information systems. Such efforts can be considered as a bridge between data analytics and business disciplines, and some examples could be *MIS Quarterly*, *Communications of the ACM*, and *Journal of the Association for Information Science and Technology* (known as *Journal of the American Society for Information Science and Technology* before).

3) Business & management journals (Light Blue, Pink, Yellow, and Green Nodes): these journals cover a broad range of business disciplines, but most of them could be assigned to the areas of management science and economics. However, despite several appearances (e.g., *Marine Policy* and *Health Policy and Technology*), the weight of pure policy journals is relatively weak.

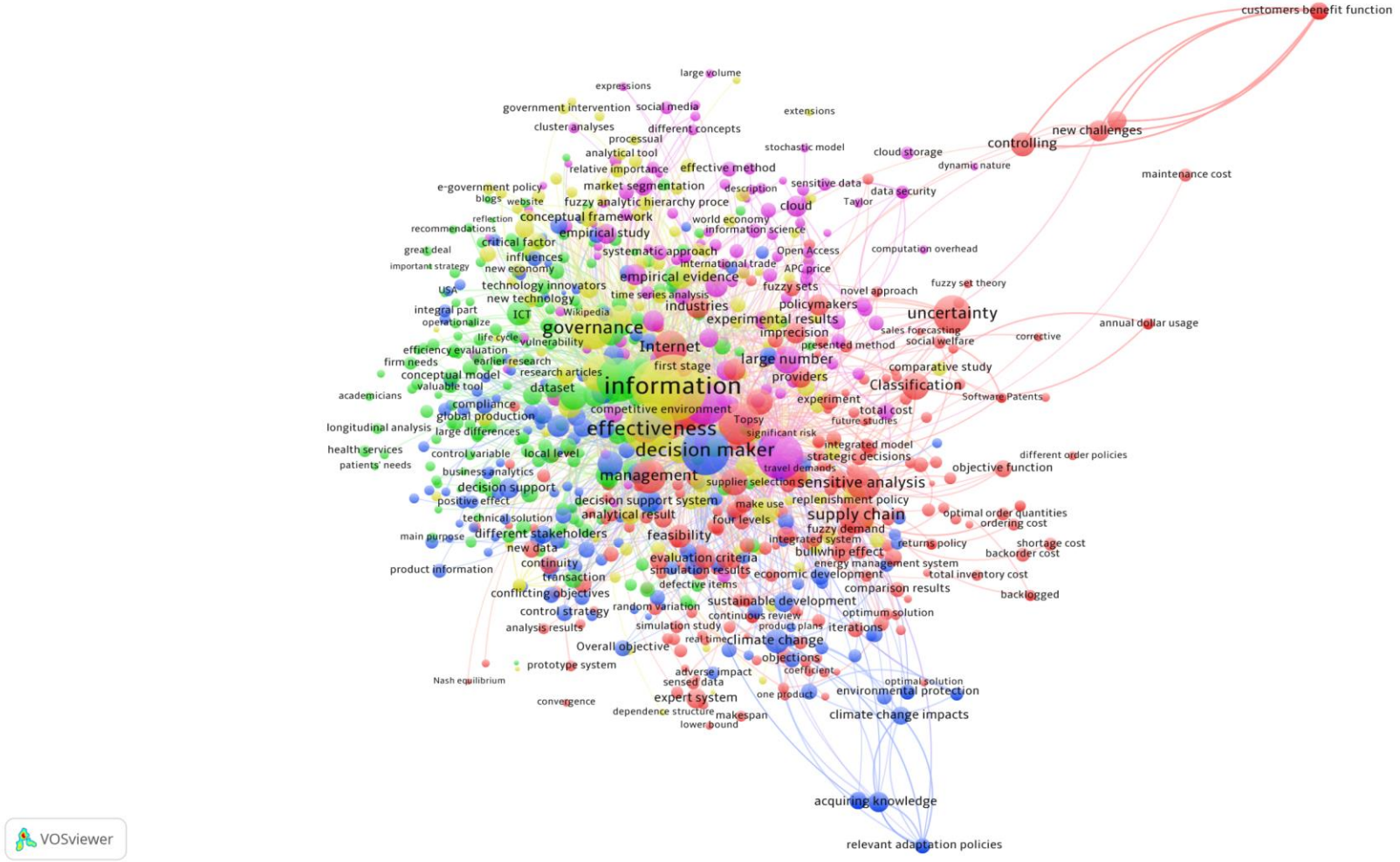

Fig. 7. Term correlation map with top-level journals in data analytics.

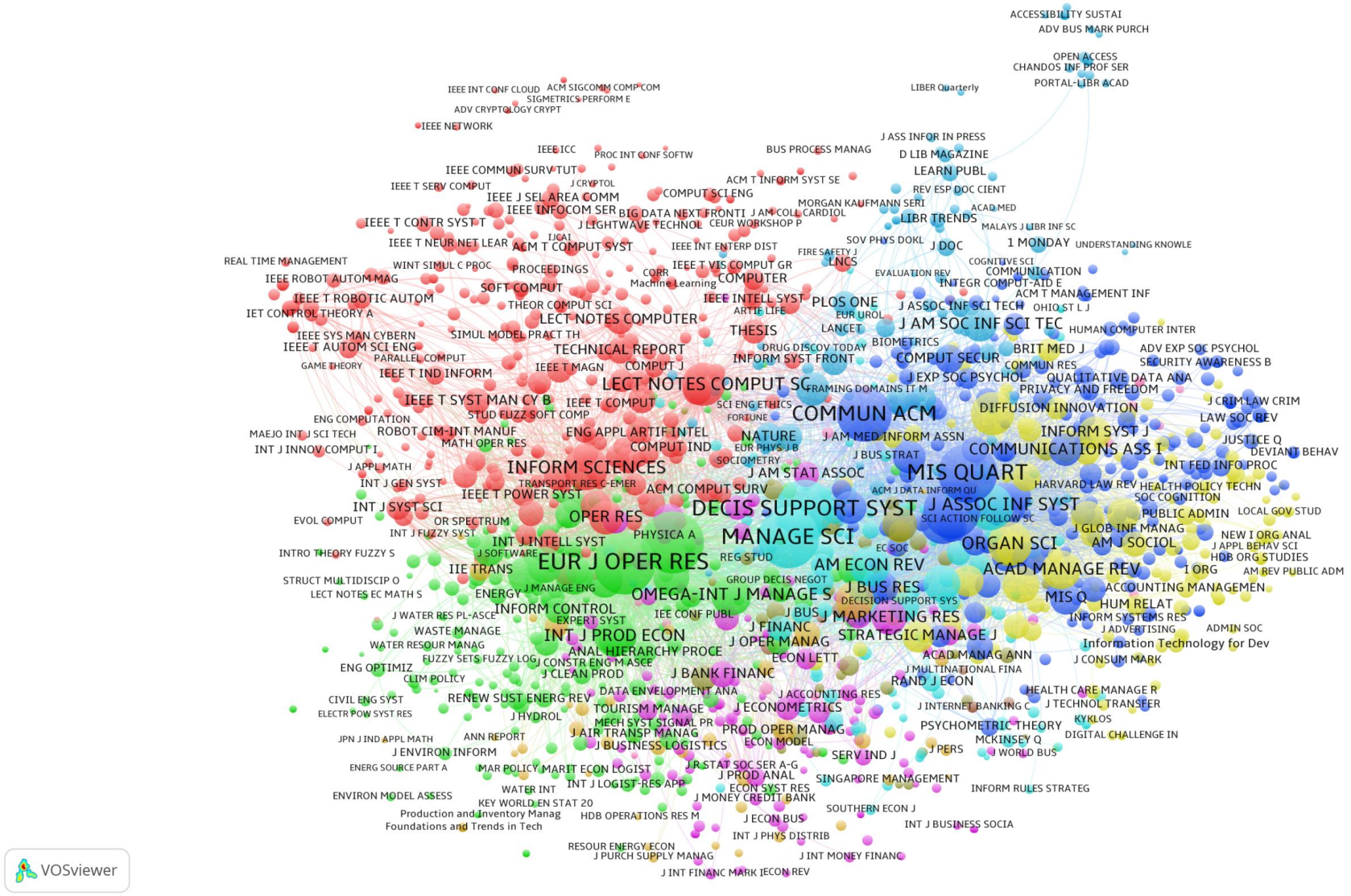

Fig. 8. Journal correlation map with journals cited by top-level journals in data analytics.

The country correlation map with authors in top-level journals in data analytics is given in Figure 9. Apparently, compared to Figures 3 and 6, the area of data science likely has the lowest threshold on language and cultural background. Besides the US and European countries, Asian countries easily gained their global reputation (e.g., China became the country with the second largest number of articles in Dataset 3 and the accomplishment of South Korea and Singapore is also significant). With the findings observed from Datasets 1 and 2, China has strong competitiveness on data analytics, and has presented great interest on gaining reputation in the global community of policy analysis. Thus, it would be reasonable to imagine that China has been equipped with the ability to publish cutting-edge papers involving both data analytics and policy analysis.

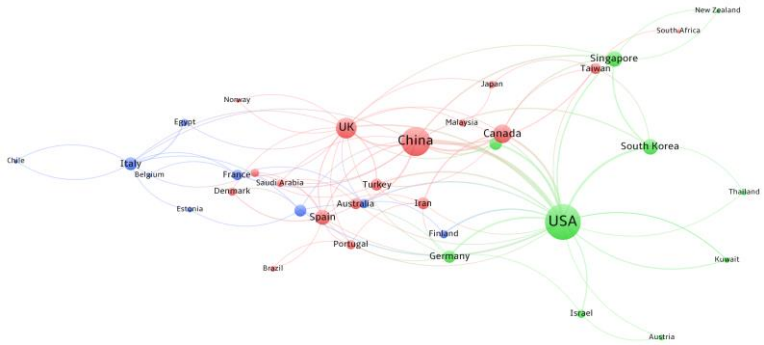

Fig. 9. Country correlation map with authors in top-level journals in data analytics.

Compared to the community of policy analysis, our observations indicated that IT researchers have been more active in pushing the interaction between data analytics and policy analysis forward, even though such efforts at the moment mostly stand on the level of business and management and endeavors on specific policy analytic tasks are still rare. In addition, information systems, in particular decision support systems, are considered as an effective instrument for IT researchers to bridge data analytics and policy analysis.

D. Discussion and Implications

This study investigated the question, "how is data analytics involved in policy analysis to create complementary values?" from the perspective of bibliometrics. Articles published in *Nature* and *Science* indicated the complementation of data analytics for policy analysis could be a cutting-edge research direction for disciplines such as computer science, information technology, business and management. The US and European countries, due to their economic, academic, and language strengths, are spearheading this direction, while Asian and other developing countries are still further from the main stage.

When focusing on the two areas: (1) policy analysts still emphasize the use of traditional approaches, such as statistical and econometric models, and the influence of IT-based data analytic approaches is relatively elusive. The advantages of traditionally developed countries further helped policy analysts easily access the threshold to publish high-quality research papers. (2) Data scientists have presented strong interest on developing and applying data analytic models to handle STIP issues, and decision support systems are used as an effective tool to conduct such research. Based on those cultural and language barriers, it became easier for Asian countries to access the community and gain reputation, but, definitely, developed countries still dominant the stage.

Concentrating on specific topics and routines that bridge data analytics with policy analysis (given in Figure 10), a current routine is to exploit information systems that integrate advanced information technologies (e.g., artificial intelligence) to potentially support economic models (e.g., econometrics) for policy analysis. However, the role of information systems in such routine is not crucial, and mostly those cases use information systems for indicator calculation within an econometric framework.

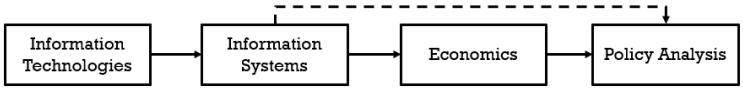

Fig. 10. Routines bridging data analytics with policy analysis.

One interesting comparison between two articles in *Science* coincides with our findings. Ref [20] in Dataset 1 was published in 2015, and machine learning techniques were developed to predict poverty and wealth from mobile phone metadata, which spearheaded a cutting-edge area of conducting prediction through data analytics at that time. This paper was authored by a group of researchers whose expertise aligned with computer science. Ref [21] was published in January 2018, in which machine learning and optimal matching techniques were designed for investigating refugee integration issues, and importantly the authors were all policy analysts. Such changes clearly indicated that intelligent technology-enhanced data analytic models (e.g., information systems) would be used to directly support policy analysis. However, the current gap is that IT researchers are unfamiliar with STIP needs and existing endeavors on decision support systems that might still not exactly match real-world issues, since simulation and refined real-world datasets are widely used.

In general, our study clearly described the trend that involves data analytics in policy analysis that is holding great interest from both communities; however, due to diverse underlying reasons, existing endeavors from the two sides are not equal. In particular, the significance of this study includes: (1) this is the first empirical study that investigates the interaction between data science and policy analysis from a bibliometric perspective; (2) considering that the involvement of data analytics in policy analysis is a cutting-edge direction, we identified core topics, relevant academic journals, and leading countries from the two parallel disciplines. Importantly, compared to expert knowledge, the use of bibliometrics provides a solution to grasp insights in an accurate and objective way and also boosts the case; and (3) we raised one solution to enhance the use of data analytics for analyzing political issues, and a related gap was also provided.

IV. Conclusions

This paper conducted an empirical study to investigate the question, "How are data analytics involved in policy analysis to create complementary values?" Based on several bibliometric approaches (e.g., co-word analysis, bibliographic coupling analysis, and science mapping), we collected three datasets with relevant articles published in (1) *Nature* and *Science*; (2) top-level journals in policy analysis; and (3) top-level journals in data analytics. Insights were identified, which would be beneficial for science policy, entrepreneurships, and strategic management. Despite not having a direct linkage with issues in technology management, it is reasonable to believe that technology management could be a private sector counterpart to policy analysis. Similarly, the involvement of data science in technology management could be in the same situation and the design of the empirical study could be adaptable for the technology management community.

Future studies of this research could go the following ways: (1) extensive bibliometric studies could be involved to provide insights from diverse perspectives (e.g., citation statistics, and advanced topic analysis); and (2) the findings observed by the current exploratory study could still require systematic examinations, and the engagement of econometric models could be such a solution.

Acknowledgment

This work is partially supported by the National Science Foundation of China under Grant 71673024.

References

[1] R. J. Funk and J. Owen-Smith, "A dynamic network measure of technological change," *Management Science*, vol. 63, pp. 791-817, 2016.

[2] T. Tang and D. Popp, "The learning process and technological change in wind power: Evidence from China's CDM wind projects," *Journal of Policy Analysis and Management*, vol. 35, pp. 195-222, 2016.

[3] E. Bakshy, D. Eckles, and M. S. Bernstein, "Designing and deploying online field experiments," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 283-292.

[4] D. V. Shah, J. N. Cappella, W. R. Neuman, B. Burscher, R. Vliegenthart, and C. H. De Vreese, "Using supervised machine learning to code policy issues: Can classifiers generalize across contexts?," *The ANNALS of the American Academy of Political and Social Science*, vol. 659, pp. 122-131, 2015.

[5] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, pp. 1165-1188, 2012.

[6] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Communications of the ACM*, vol. 57, pp. 78-85, 2014.

[7] S. Athey, "Beyond prediction: Using big data for policy problems," *Science*, vol. 355, pp. 483-485, 2017.

[8] Y. Zhang, G. Zhang, H. Chen, A. L. Porter, D. Zhu, and J. Lu, "Topic analysis and forecasting for science, technology and innovation: Methodology and a case study focusing on big data research," *Technological Forecasting and Social Change*, vol. 105, pp. 179-191, 2016.

[9] Y. Zhang, G. Zhang, D. Zhu, and J. Lu, "Science evolutionary pathways: Identifying and visualizing relationships for scientific topics," *Journal of the Association for Information Science and Technology*, vol. 68, pp. 1925-1939, 2017.

[10] Y. Zhang, A. L. Porter, Z. Hu, Y. Guo, and N. C. Newman, ""Term clumping" for technical intelligence: A case study on dye-sensitized solar cells," *Technological Forecasting and Social Change*, vol. 85, pp. 26-39, 2014.

[11] M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Social Science Information*, vol. 2, pp. 191-235, 1983.

[12] M. M. Kessler, "Bibliographic coupling between scientific papers," *American Documentation*, vol. 14, pp. 10-25, 1963.

[13] C. Calero-Medina and E. C. Noyons, "Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field," *Journal of Informetrics*, vol. 2, pp. 272-279, 2008.

[14] L. Waltman, N. J. van Eck, and E. C. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *Journal of Informetrics*, vol. 4, pp. 629-635, 2010.

[15] L. Cameron, N. Erkal, L. Gangadharan, and X. Meng, "Little emperors: behavioral impacts of China's One-Child Policy," *Science*, vol. 339, pp. 953-957, 2013.

[16] E. M. Hafner-Burton, M. Kahler, and A. H. Montgomery, "Network analysis for international relations," *International Organization*, vol. 63, pp. 559-592, 2009.

[17] M. Kroenig, "Exporting the bomb: Why states provide sensitive nuclear assistance," *American Political Science Review*, vol. 103, pp. 113-133, 2009.

[18] S. Tutun, M. T. Khasawneh, and J. Zhuang, "New framework that uses patterns and relations to understand terrorist behaviors," *Expert Systems with Applications*, vol. 78, pp. 358-375, 2017.

[19] W. Chung and D. Zeng, "Social‐media‐based public policy informatics: Sentiment and network analyses of US Immigration and border security," *Journal of the Association for Information Science and Technology*, vol. 67, pp. 1588-1606, 2016.

[20] J. Blumenstock, G. Cadamuro, and R. On, "Predicting poverty and wealth from mobile phone metadata," *Science*, vol. 350, pp. 1073-1076, 2015.

[21] K. Bansak, J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, *et al.*, "Improving refugee integration through data-driven algorithmic assignment," *Science*, vol. 359, pp. 325-329, 2018.