Interactions between data science and policy analysis: Evidence from the perspective of bibliometrics

Yi Zhang yi.zhang@uts.edu.au University of Technology Sydney

Denise Chiavetta dchiavetta@searchtech.com Search Technology Inc.

Nils Newman newman@searchtech.com Search Technology Inc.

Scott Cunningham s.cunningham@tudelft.nl Delft University of Technology

Keywords: Bibliometrics; Data science; Policy analysis; Topic analysis; Network analysis.

Extended Abstract

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 integrates ITs with econometric models to handle real-world issues in science, technology, innovation and policy (STIP) research (Funk & Owen-Smith 2016; Tang & Popp 2016), and data scientists or leading IT companies develop systematic toolkits to support decision making in a number of public sectors (Bakshy, Eckles & Bernstein 2014; Shah et al. 2015). 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 (Chen, Chiang & Storey 2012; Kim, Trimi & Chung 2014). However, raised by Athey (2017), theoretical and practical innovation of ITs is still required to stimulate big data to achieve the full potential in informing STIP. Under this circumstance, this paper aims to answer the following two research questions: 1) considering the increasingly in-depth interactions between data science and policy analysis, how can we detect and track such interactions and visualize their dynamics in the past decades? 2) Since the interactions between data science and policy analysis seem like a trend, is there a direct correlation between such interactions and the quality of related research?

A bibliometric study is constructed to address concerns to the two questions above, and the designed research framework is given in Figure 1, in which four functions are mainly involved: search strategy, topic analysis, network analysis, and interaction exploration. An expert panel with both data scientists and policy analysts is formed to support the study.

The expert panel will provide a list of leading journals in the fields of public administration, computer science, information science, business and management, and multidisciplinary studies, and articles published in these journals are collected. Then, we extract the bibliographic information of these articles, including author information and citation statistics, and natural language processing techniques are used to retrieve terms.

Topic analysis and network analysis are conducted in parallel. Topic analysis is designed for the first question. We introduce a term clumping process (Zhang et al. 2014) to remove noise and consolidate synonyms, and a list of core scientific and technological terms will be generated. Then, topic models (Suominen & Toivanen 2016) are used to identify core topics, and a model of scientific evolutionary pathways (Zhang et al. 2017 (to appear)) is to discover the evolutionary relationships between topics (e.g., predecessors and descendants) within a learning process. Oriented to the second question, we construct a co-authorship network and identify the degree of the active interactions of nodes via centrality measures (Yan & Ding 2009). Citation statistics and journal impact factors then act as reference indicators for evaluating the quality of articles, and correlation analysis is used to examine whether a correlation exists between the quality of articles and active cross-disciplinary interactions.

Quantitative results then will be integrated with expert knowledge provided by the expert panel, and an in-depth discussion will be given to explore insights on the interactions between data science and policy analysis.

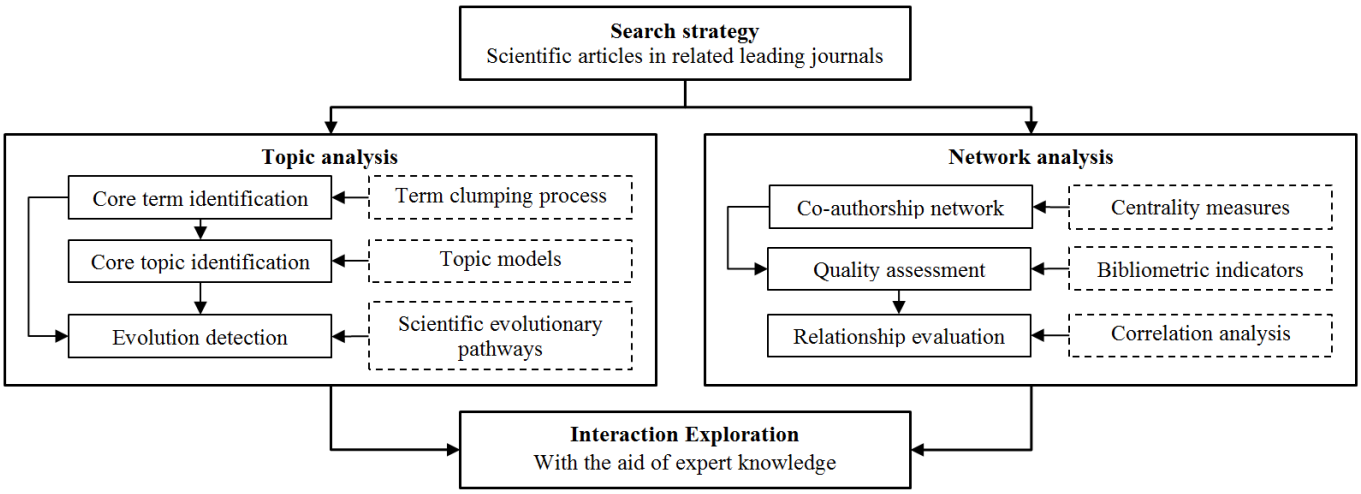

Figure 1. Research framework of the bibliometric study for exploring interactions between data science and policy analysis

References

Athey, S. 2017, 'Beyond prediction: Using big data for policy problems', *Science*, vol. 355, no. 6324, pp. 483-5.

Bakshy, E., Eckles, D. & Bernstein, M.S. 2014, 'Designing and deploying online field experiments', *Proceedings of the 23rd International Conference on World Wide Web*, ACM, pp. 283-92.

Chen, H., Chiang, R.H. & Storey, V.C. 2012, 'Business intelligence and analytics: From big data to big impact', *MIS Quarterly*, vol. 36, no. 4, pp. 1165-88.

Funk, R.J. & Owen-Smith, J. 2016, 'A dynamic network measure of technological change', *Management Science*.

Kim, G.-H., Trimi, S. & Chung, J.-H. 2014, 'Big-data applications in the government sector', *Communications of the ACM*, vol. 57, no. 3, pp. 78-85.

Shah, D.V., Cappella, J.N., Neuman, W.R., Burscher, B., Vliegenthart, R. & De Vreese, C.H. 2015, '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, no. 1, pp. 122-31.

Suominen, A. & Toivanen, H. 2016, 'Map of science with topic modeling: Comparison of unsupervised learning and human‐assigned subject classification', *Journal of the Association for Information Science and Technology*, vol. 67, no. 19, pp. 2464‐76.

Tang, T. & Popp, D. 2016, 'The learning process and technological change in wind power: Evidence from China's CDM wind projects', *Journal of Policy Analysis and Management*, vol. 35, no. 1, pp. 195-222.

Yan, E. & Ding, Y. 2009, 'Applying centrality measures to impact analysis: A coauthorship network analysis', *Journal of the Association for Information Science and Technology*, vol. 60, no. 10, pp. 2107-18.

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

Zhang, Y., Zhang, G., Zhu, D. & Lu, J. 2017 (to appear), 'Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics', *The Journal of the Association for Information Science and Technology*.