Data-driven Science of Science

Yi Bu[1], Guangjian Li[2], Wei Lu[3], Ying Ding[1]
[1]School of Informatics and Computing, Indiana University, Indiana, U.S.A.
[2]Department of Information Management, Peking University, Beijing, China
[3]School of Information Management, Wuhan University, Wuhan, Hubei, China

Abstract
The number of scientific researches has grown exponentially over the past several decades. The surge of data-driven science of science has significantly affected and changed traditional bibliometrics. Here, viewpoints will be provided and discussed to help better understand data-driven science of science. The purpose of this session is to showcase the current data-driven science of science research by highlighting several cutting-edge researches and to construct a community of researchers to explore questions critical to the future of data-driven science of science, especially a community of data-driven science of science in iSchools so as to facilitate collaboration and inspire innovation.
Keywords: Data-driven science of science; data science; information science.
Citation: Editor will add citation
Copyright: Copyright is held by the authors.
Contact: buyi@iu.edu

1   Purpose and Intended Audience

The number of scientific research has grown exponentially over the past decades. Growth in research articles, conference papers, posters, patents, preprints, science and technology reports, and informal content on the Web (Hu et al., 2015) has created "big data", which provide opportunities and challenges to people in both academia and industry.

Compared with traditional bibliometric and scientometrics studies, the current surge in science of science research is distinguished by different features. First of all, large-scale datasets are heavily relied, many of which contain millions of authors, papers, and their citations (ISSS, 2016). Additionally, research themes are diverse, ranging from scientific career to rising star detection, from knowledge diffusion to successful innovation, from computational hypotheses generation to data-driven knowledge discovery; and the applied domains are varied, from social science to natural science, from information science to computer science, from economics to management science, from mathematics to physics, etc.

The purpose of this session is to showcase the current data-driven science of science researches by highlighting several studies and to construct a community of researchers to explore questions critical to the future of data-driven science of science, especially a community of data-driven science of science in iSchools so as to facilitate collaboration and inspire innovation.

2   Proposed Activities

Using a lightning talk format, this SIE will be conducted as follows:

- Part 1: Introduction (Yi Bu, 5 minutes)
  - Content: A brief introduction of data-driven science of science will be given and several examples and frontiers in this field will be mentioned.
  - Bio: Yi Bu is a doctor student and research assistant in School of Informatics and Computing at Indiana University supervised by Dr. Ying Ding. He graduated with a specialty of Information Management and System and a Computer Science Minor from Peking University in China. His research mainly focuses on data-driven science of science, data-driven decision-making, citation analysis, scientific collaboration, scientific career, and network science.

- Part 2: Three lightning talks followed by Q & A (20 minutes each, 60 minutes in total)

  Lightning Talk 1: Knowledge Diffusion and Data-Driven Science of Science
  - Speaker: Dr. Guangjian Li
  - Abstract: Knowledge diffusion is a concept that has been developed and changed with the technological and social environment. It is also closely related to data-driven

science of science. This talk will firstly showcase the concept changes and some frontiers in knowledge diffusion and emerging topics detecting. The relationship between knowledge diffusion and data-driven science of science will further be pointed out with a new theoretical frame. Several interesting algorithms in knowledge diffusion will also be illustrated and their applications in data-driven science of science will be detailed.

○ Bio: Dr. Guangjian Li is the professor and the dean of Department of Information Management, Peking University. His research interests lie in data-driven knowledge diffusion, web information system, competitive intelligence analysis, Web data mining, and the application of Web technologies. He has been involved in various projects funded by China Social Science Fund. He has published 80+ papers in journals and conferences and has also written three books.

Lightning Talk 2: Deep Semantic Mining of Academic Text
○ Speaker: Dr. Wei Lu
○ Abstract: Scholarly big data mining is going towards to more in-depth, fine-grained, and semantic-oriented. This talk will firstly revisit the research development of academic document understanding, and then will present some of our initial work on academic document segmentation, structure function analysis, term function recognition, and their applications. Future directions will also be discussed in the end of this talk.
○ Bio: Dr. Wei Lu is the professor and the vice dean of School of Information Management, Wuhan University. His research interests include information retrieval, data mining, and academic document understanding. He was also recently elected as Youth Yangtze River Scholar by the Ministry of Education of the People's Republic of China. He did his postdoc at City University London, and worked as a visiting researcher at Royal School of Library Science, Denmark. He has published dozens of papers in journals, conferences and workshops.

Lightning Talk 3: Data-Driven Science of Science
○ Speaker: Dr. Ying Ding
○ Abstract: Data have fundamentally changed our daily lives. Data now can be bought and sold, and soon data analytics will become commodity as well. Data have significantly challenged every discipline to revisit its research roadmap to stay competitive. This talk will point out the paradigm shift of bibliometrics (or science of science in a broader way) and illustrate some potential interesting research directions that are data-driven. It will showcase some concrete researches on data-driven discovery and data-driven decision making.
○ Bio: Dr. Ying Ding is an Associate Professor at School of Informatics and Computing and is currently associate director of data science online program at Indiana University. She has been involved in various NIH, NSF and European-Union funded projects. She has published 190+ papers in journals, conferences, and workshops, and served as the program committee member for 180+ international conferences. She is the co-editor of book series called Semantic Web Synthesis by Morgan & Claypool publisher. She is co-author of the book "Intelligent Information Integration in B2B Electronic Commerce" published by Kluwer Academic Publishers, and co-author of the book chapter in "Spinning the Semantic Web" published by MIT Press. She is the co-editor in chief for Journal of Data and Information Science, and serves as the editorial board member for four ISI indexed journals in Information Science and Semantic Web. She is the co-founder of Data2Discovery company advancing cutting edge technologies in data science. Her current research interests include data-driven knowledge discovery, semantic Web, knowledge graph, scientific collaboration, and the application of Web Technology.

● Part 3: Group discussions (25 minutes).
○ Content: We will divide the participants into several small groups to discuss the challenges and potentials for data-driven science of science, with the specific focuses on: research questions, available technologies, and potential collaboration on joint

research including grant proposals and publications. We will conclude the session by 1-3 minutes' presentations from each groups.

3   Relevance to the Conference/Significance to the Field

The theme of iConference 2017 is "Effect, Expand, Evolve: Global Collaborations across the Information Community". It provides a great opportunity to create community and to set up effective scientific collaboration in data-driven science of science. By setting up this session, we hope to bring together leading researchers from various disciplines and to form discussions on challenges and opportunities of data-driven science of science. iConference, as the top conference in Information Science, is the ideal place to create community awareness and to inspire innovative research and collaborations toward data-driven science of science.

4   The Length of My Event

This SIE will last about 90 minutes.

5   References

Hu, B., Dong, X., Zhang, C., Bowman, T., Ding, Y., Yan, E., Milojević, S., Ni., C., & Lariviere, V. (2015). A lead-lag analysis of the topic evolution patterns for preprints and publications. *Journal of the Association for Information Science and Technology*, 66(12), 2643-2656.

The Organizers of International Symposium on Science of Science (ISSS) (2016). http://icss.ist.psu.edu/index.html